CHEMICAL and BIO-PROCESS CONTROL

Fifth Edition

CHEMICAL and BIO-PROCESS CONTROL

Fifth Edition

James B. Riggs

M. Nazmul Karim

Joesph S. Alford

Ferret Publishing

This book was set in Times New Roman and printed on acid-free paper.

Printed in the United States of America

ISBN 978-09669601-9-8

Ferret Publishing
Austin, Texas 78737 USA
1-512-363-5107

www.ferretpublish.com

Dedicated to the memory of

Naz Karim and Charlie Cutler

Preface

What motivated me to write the first edition of this text in 1999 is still relevant today: To provide a text that presents the fundamentals of process control along with the industrially relevant skills that engineers can use in an industrial setting. In fact, the previous editions of this book were designed based on developing the skill set necessary for industrial control engineers. That is, an industrial control engineer should be able to understand and tune control loops (Chapters 7-9), select the proper type of controller (Chapters 7 and 12-14), troubleshoot control loops (Chapters 2 and 10) and be proficient in the terminology of the field. In addition, process control analysis based on considering the individual dynamics of the actuator, process and sensor is introduced in the dynamic modeling chapter (Chapter 3) and used throughout the text. This overall emphasis remains, but it has been enhanced with the addition of improved examples and with a special emphasis on the use of MATLAB, Python, Simulink and our visual basic process simulator and a new chapter on batch and discrete control.

We are very lucky to have added Joe Alford as a co-author for this edition. His industrial experience in the pharmecuetical industry was very valuable for updating much of the material on control loop hardware and industrial practices as well as developing the new chapter on batch and discrete control.

Our objective for this text has been to offer an approach to teaching process control to undergraduate students that has an industrially relevant component.

James B. Riggs
Austin, Texas
May 2020

About the Authors

James B. Riggs is Professor Emeritus of Chemical Engineering at Texas Tech University. He received his BS and MS degrees in chemical engineering from the University of Texas at Austin and his PhD degree in chemical engineering from the University of California, Berkeley. He co-founded the Texas Tech Process Control and Optimization Consortium in 1992 and served as its director for 16 years and has over 90 technical publications on process modeling, control and optimization. He is the author of five other engineering textbooks. In addition, he has a total of over five years industrial experience.

M. Nazmul Karim was a Professor and the Department Chair of Chemical Engineering at Texas A&M University. He received the BSc. (Honors) degree in Chemical Engineering from the Bangladesh University of Engineering and Technology, Dhaka. He earned his MSc. degree in Control Engineering and Ph.D. degree in Chemical Engineering from the University of Manchester Institute of Science and Technology, U.K. Before joining Texas A&M University in 2011, he served as chair at Texas Tech University and taught at Colorado State University for more than twenty years. Dr. Karim was the Director of the Advanced Industrial Bio-Processing Short Course, which he offered at Colorado State University for over twenty years and more than 130 companies have participated in this course. Dr. Karim has co-authored seventy refereed journal papers and published hundreds of conference papers. He has co-edited a book, *Modeling and Control of Biotechnical Processes 1992,* with Professor Gregory Stephanopoulos (MIT). He was a Fellow of the American Institute of Chemical Engineers.

Dr. Joseph S. Alford, a member of the Process Automation Hall of Fame, is a recognized leader in industrial bioprocess automation, batch process control, computer system validation, on-line data analysis, alarm management, Process Analytical Technologies (PAT), and applications of artificial intelligence. He was educated at Purdue U. (BS-ChE) and at the U. of Cincinnati (M.S. and Ph.D in ChE), after which he served in the US Navy and then a 35 year career at Eli Lilly and Co., spent primarily in automating their bioprocesses and managing plant automation and corporate advanced technology groups. He has received Eli Lilly's top corporate technology awards, national automation technology awards from both AIChE and ISA (International Society of Automation), been recognized with both Purdue U. and the U. of Cincinnati College of Engineering's Distinguished Alumnus awards, and been elected as an AIChE and ISA Fellow. He has authored or co-authored 45 publications, including several invited book chapters and the ISA book "Automation Applications in Bio-Pharmaceuticals." Dr. Alford is a member of both AIChE and ISA journal Editorial Advisory Boards, is a former ABET (Accreditation Board for Engineering and Technology) Program Evaluator, and is a past member of several university Industrial Advisory Boards. He chaired AIChE's New Books Committee for many years and is a co-author of ANSI/ISA-18.2 , which is the national standard on process alarm management. Since retiring as an Engineering Advisor from Eli Lilly in 2006, he has remained active as an automation consultant, in technical committee and editorial advisory work for AIChE and ISA, and in giving guest lectures at universities.

Table of Contents

PART III: PID CONTROL

PART IV: ADVANCED PID CONTROL

Part I

Introduction

Chapter 1

Introduction to Chemical and Bio-Process Control

Chapter Objectives

- Provide background and justification for the use of process control by industry.
- Introduce the terminology of feedback control using everyday examples.
- Present the format and structure for control diagrams and present control diagrams for several simple industrial control loops.
- Overview the general types of controllers and the responsibilities of process control engineers.
- Introduce process optimization and indicate how process control and process optimization work together.

1.1 Chemical and Bio-Process Control

This text addresses the fundamental and generic aspects of process control, but also focuses on process control applications in the **chemical process industries** (**CPI**) and the **biotechnology industries**. That is, examples relevant to the CPI and the biotechnology industries are presented throughout the text as well as a discussion of some of the key operational issues associated with these industries. In addition, the major elements of the control hardware associated with these industries are also presented and their effect on control performance is addressed throughout the text.

The CPI represents a range of industries that use processing units to produce a wide range of products, including hydrocarbon fuels, petrochemical products, concrete, pharmaceutical products, paper products, man-made fibers and films, agrochemical products and ceramics[1]. In addition, the CPI also involves processing technologies that provide environmental protection, refrigeration, air conditioning and electric power generation. Two primary examples of the CPI are refineries and chemical plants. Refineries produce fuel products, such as gasoline, jet fuel and diesel fuel, and produce a wide range of high volume chemical intermediates, such as ethylene, benzene and linear paraffins. On the other hand, chemical plants use these chemical intermediates and other feedstock sources to produce a range of chemical products, including plastics, resins, solvents, synthetic fibers and a large number of final chemical products, e.g., gasoline additives, food additives and preservatives, diapers and detergents.

Biotechnology[2] is the technology that uses **microbial species** or any other living organisms or part of them to produce useful products. Biotechnology is used to produce high-volume products such as ethanol (bio-fuel) and citric acid from renewable resources, e.g., corn. High-value products, such as pharmaceutical and **recombinant drugs** for combating various diseases, are the main emphasis of modern biotechnology. Some portion of certain bio-processes use typical unit operations that are found in the CPI, such as mixers, reactors, heat exchangers, distillation columns, liquid/liquid separation, membrane separation, crystallizers, etc. The difference is that a specific **microorganism** is grown to produce a desired product, which could be the cell mass itself, and contamination may not be allowed in the process. Thus, maintenance of sterility is a key factor that separates many bio-processes from traditional chemical processes. In addition, in certain cases local hydrodynamic shear rates must be kept below specific levels or cell damage to the microbial species can result. Another key factor for human therapeutics is that every process, critical instrumentation and control configurations (i.e., that is those associated with "critical process parameters"), and the computer systems used to produce these products needs to be validated in accordance with current Good Manufacturing Practices (cGMPs) as enforced by the Food and Drug Admistration (FDA). This requires, for example, extensive equipment qualification, formal testing, change control and operator training, compliance to electronic record and signature requirements, and full documentation. Also, the processes that produce pharmaceutical and recombinant drugs have much smaller-scale production rates than those typically used in the CPI and as a result bio-processes typically use batch reactors instead of the continuous processes used extensively in the CPI.

Bio-processes usually involve a bio-reactor and bio-separation systems. Bio-reactors are usually batch reactors, i.e., the reactor is filled with microbial species and food and **nutrients** for the microbial species and after the reactions attain the desired degree of completion, the reactor is emptied and the process is repeated. For "bio-separations", unit operations that are not common in chemical industries, such as chromatographic columns, ultra-filtration and micro-filtration, are used. In summary, **the primary differences between processes from the CPI and biotechnology systems are (1) many bio-systems must maintain sterile conditions throughout the system, (2) stagnant regions need to be eliminated from bio-systems to enhance cleanability, sterilizability and to maintain a uniform environment for the microorganisms, (3) limits on local shear rates occur for many bio-processes, (4) bio-processes that produce products for human consumption must have FDA validation, (5) bio-separations tend to be different separation technologies than used in the CPI and (6) most bio-systems tend to have much smaller production rates than CPI processes.**

Chemical Process Control. Chemical process control (**CPC**) is concerned with operating a processing plant such that the product quality and production rate specifications are met in a safe and reliable manner. To attain these objectives, various flow rates, levels, pressures, temperatures, and compositions are maintained at or near operating target values. CPC is part of the larger field of automatic control, which ranges from controlling aircraft to controlling robots to controlling the operation of the critical systems in a computer.

Automatic control was first applied to refineries and chemical plants due to their large processing rates and complex configurations of unit operations. The first refineries and chemical plants were designed with large holding tanks between processing units, allowing each unit to be operated more or less independently. Because of the holding tanks, the plant operators were able to perform most of the control functions for these early processing systems. Holding tanks are an expensive approach for simplifying the control of a process because these tanks require significant additional capital, require significant space for installation, represent additional safety risks and result in an increase in inventory costs. Intermediate holding tanks have largely been removed from large-scale processing systems as the capability and reliability of process control has improved. That is, modern process control systems have allowed operating companies to remove these intermediate holding tanks for economic reasons while, in fact, improving the overall safety of their operations and the quality of their

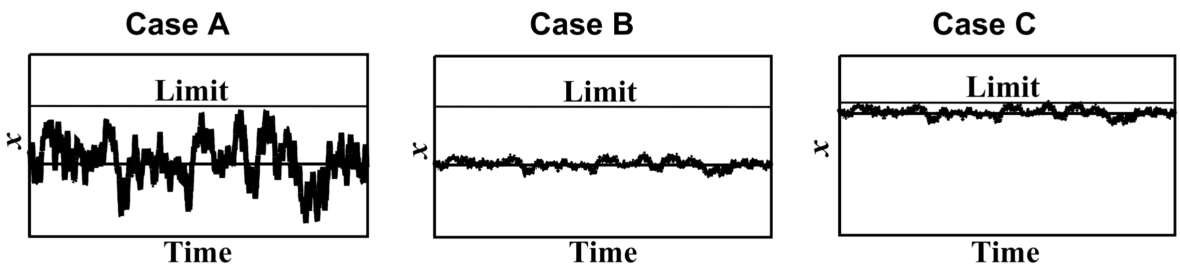


Figure 1.1.1 Comparison between impurity measurements (x) and the upper limit on the impurity in a product for the original control system (case A), the improved control system with the original impurity target (case B) and the improved control system with a new impurity target (case C).

products. In addition, the rapid increase in oil prices in the 1970s resulted in more complex and integrated process designs in an effort to conserve energy. These design modifications produced processes that rely more on process control for effective operation.

Over the past 40 to 50 years, the CPI have been in a transition from a relatively young industry, largely driven by innovation in new products and new processing approaches, to a more mature industry in which the technology of the industry is changing much more slowly. In earlier times, new products such as nylon and Teflon® were developed and new process designs such as fluidized catalytic cracking (FCC) and plastic processing technologies were implemented. These innovative products and processing approaches provided a major economic advantage to their developers. Today, technological breakthroughs are relatively rare in the CPI. On the other hand, companies in the CPI can attain significant economic advantage by optimizing the performance of their existing processes. CPC is an integral part of attaining the most efficient operation of processes in the CPI.

CPC is intimately involved in the effort to meet the operational objectives of the process while striving for the most efficient operation of the plant. Minimizing the **variability** (i.e., magnitude of the deviations from the target) in a product is often a key operational objective and is directly affected by the performance of the process control system. In fact, the performance of an overall process control system is many times expressed in terms of the variability in the products produced by the process. Figure 1.1.1 shows the measurement of the impurity in a product for the original control system (case A). Case B represents the performance of a new and improved control system. The controller corresponding to case B produces a product with less variability in the impurity than for case A; therefore, case B is referred to as producing a lower variability product than case A. For many products, low variability is an important product specification. If a product does not meet its product variability specifications, the resulting product can be low-valued with low demand, while products that meet the variability specification can be high-valued with high demand. Because case B has a lower variability, the target impurity level can be moved closer to the impurity specification (case C), usually allowing greater production rates or lower energy usage, both of which result in more efficient operation of the process. Other types of operational limits are encountered resulting from environmental regulations, capacity limits on equipment and safety limits. In a similar manner, operating close to these limits can also be economically important.

Summarizing, the benefits of improved control in the CPI can be (a) producing a lower variability product, (b) increasing the process throughput and/or (c) reducing the energy usage while satisfying safety and environmental constraints. It should be emphasized that these economic benefits can generally be attained with modest or no additional capital investment by improving the performance of the control systems.

Bio-Process Control. Beer and wine making, which date back to ancient times, were among the earliest bio-processes. Another ancient bio-based process is the production of fermented foods, such as bread products and yogurt. Control of these ancient bio-processes was done manually, without any sensors or feedback control. As sensors, such as thermometers and pressure gauges, and pneumatic controllers were developed and demonstrated in the CPI, they were gradually adopted starting in the 1940s by the bio-process industries which allowed for more complicated process designs and operational scenarios.

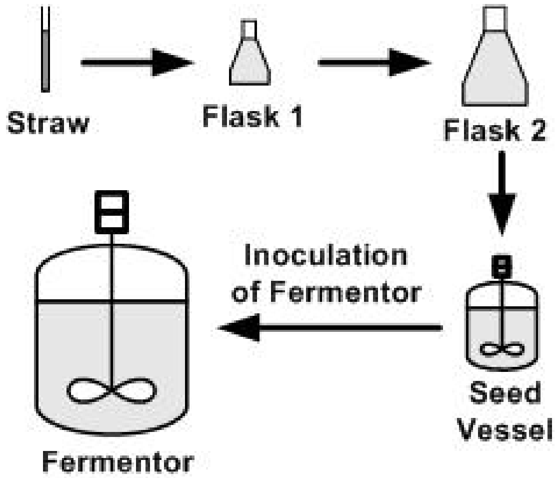


Figure 1.1.2 Processing steps for production of the inoculum for a bio-reactor. After a drawing by Eli Lilly and Company.

A typical biological process uses the following steps: a) raw materials (e.g., media) preparation, b) preparation of the fermentation **inoculum** (microbial **cells**), c) **sterilization** of the process, d) combine the media and the microbial cells in the bio-reactor, which is known as **inoculation**, e) implement the **fermentation** step f) bulk product (i.e., active pharmaceutical ingredient) recovery from the fermentation broth, (g) formulation of the final product (e.g., granulation, milling, blending steps), and (h) preparation of the products for sale (e.g., vials, tablets, capsules, and packaging). In each of these steps, certain process conditions need to be maintained for acceptable operation, and this is accomplished by process control techniques. The preparation of the fermentation inoculum requires a number of scale-up steps from the straw (1 ml volume) to the inoculum (up to 10 % of the fermentor volume) produced by the seed vessel (Figure 1.1.2). Many times a bio-reactor is operated in two distinct modes during the fermentation step. First, the bio-reactor is operated to increase the concentration of the microorganisms of interest. Then, the operating conditions are changed so that the microorganisms produce the desired product. In both of these steps, effective process control is critical. High-volume bio-processes (millions of gallons per year capacity), e.g., processes that produce ethanol from grain, are typically run in a similar process control environment as the CPI. As a result, improved process control for these high-volume bio-process results in improved product uniformity, increased throughput and/or reduced utility usage. The most important difference between the CPI and the biotechnology industries is that in bio-processes live microorganisms are used, and as a result, maintaining sterility of the fermentation vessel (bio-reactor) and the media for fermentation can be critically important. In addition, batch processes are almost always used for relatively low-volume bio-processes, and as a result, the benefits of and requirements for process control are quite different than for the CPI.

Batch bio-processes go through a sequence of systematic events, such as sterilization, filling a vessel, maintaining a certain temperature, pH and dissolved oxygen concentration for a portion of the batch, emptying the vessel and washing the vessel. These steps in the past were done manually; nowadays with the advent of modern control technology, these sequences are automated. For high-value products, such as human therapeutics, which can be sold at millions of dollars per kilogram, the processes are small, the product concentrations can be in Pico-gm/liter (i.e., 10^{-12} g/l) and controls of these processes are FDA regulated. These processes are required by the FDA to operate following what are known as Current Good Manufacturing Practices (cGMPs), which are strict requirements developed by the federal government. They include requirements for development, testing, operations, maintenance, change control, monitoring, and documentation of processes and associated support systems (e.g., purified water, computer systems). cGMP pertains to the basic principles, procedures and resources needed to ensure a manufacturing environment which is suitable for producing bio-pharmaceuticals of acceptable quality. In the past, FDA was very strict about certification of a process for producing human therapeutics, a procedure known as "validation." All steps of the process operation needed to be validated. A company had to demonstrate that a process operated precisely within a very narrow operational band. This required good process sensors (which also need to be validated), and precise book-keeping. The advent of high speed computing and more reliable process control hardware and software has made it easier to document a bio-pharmaceutical process for validation. In 2004, the FDA introduced a procedure known as "Process Analytical Technology" (PAT), which is defined by FDA as "a system for designing, analyzing, and controlling manufacturing through timely measurements (i.e., during processing) of critical quality and performance attributes of raw and in-process materials and processes with the goal of ensuring final product quality."[3] Companies are now in a position to use modern process control technologies to improve their processes within the new FDA guideline. In addition, the FDA guidelines should make it easier to move to larger-scale, continuous bio-processes while maintaining strict quality control. Further, the FDA states that these process control tools, when used within a system "can provide effective and efficient means for acquiring information to facilitate process understanding, develop risk-mitigation strategies, achieve continuous improvement, and share information and knowledge."[3] **The primary benefits of improved process control for the bio-pharmaceutical industry are reduced process variability, improved consistency of product quality, and greater production rates**, all of which results from more available process instrumentation and automation functionality, greater use of that information (e.g., for on-line decision making) and more reliable operation of the process. As a result, process control is likely to become even more important to the success for the bio-pharmaceutical process industries.

Self-Assessment Questions

Q1.1.1 What are the primary objectives of process control?

Q1.1.2 What is improved process control performance? How is it measured?

Q1.1.3 Why is improved process control financially attractive to processing companies?

Q1.1.4 How is a bio-process different from a typical process in the CPI?

Q1.1.5 What is PAT? How can this help the pharmaceutical industries?

Self-Assessment Answers

Q1.1.1 The primary objectives of process control are to operate a plant such that product quality and process specifications are met in a safe, consistent, and reliable manner, compliant with any environmental and/or plant constraints. Process control also plays a key role in optimizing plant operations.

Q1.1.2 Improved control performance for a control loop is indicated by a reduction in the variability from setpoint. Variability can be expressed by the average error from setpoint or, in some cases, the standard deviation from the setpoint. Sometimes the measurements of the CVare plotted as a function of time with upper and lower limits to graphically show the control performance of a loop.

Q1.1.3 Process control is financially attractive to companies in the CPI and bio-tech industries because it can provide significant economic benefit (i.e., CPI: improved product quality, increased production rates and decreased energy usage; bio-tech industries: improved product quality, faster process certification and greater production rates) for a relatively low cost. That is, to implement a new control strategy, it generally requires only the time of the control engineer and some technicians with a relatively small amount of capital investment.

Q1.1.4 Some bio-processes require sterile conditions, must avoid stagnant regions and high shear rates,
FDA validation, different types of separation technology than the CPI and are usually much smaller scale than the process used in the CPI.

Q1.1.5 PAT (Process Analytical Technology) is a FDA/industry continuous improvement initiative, encouraging greater use of new technologies, risk-based management, and increased automation and data analysis to achieve greater process understanding, increased on-line decision making, reduced cycle time, improved process control, and reduced process variation, while maintaining product quality.

1.2 Everyday Examples of Process Control

Process control is commonplace in our everyday life. Several examples are considered here in an effort to introduce the concept of process control as well as some of the terminology of the field. The terminology introduced here will be frequently used throughout the remainder of the text.

Controlling the Water Temperature of a Shower. Everyone is familiar with the problem of controlling the water temperature of a shower when you first get in the shower or after something has changed the water temperature (e.g., water heater starts running out of hot water). Assume that the valve (faucet) on the cold-water line is adjusted to control the temperature of the shower.

Consider the case in which the hot-water heater is starting to run out of hot water, which leads to a decrease in the temperature of the hot water. As a result, the person in the shower begins to close the valve on the cold-water line to maintain the desired temperature of the shower water.

Let's analyze this process from a process control point of view. In this case, the shower is the **process**. The person's skin "senses" the shower water temperature and is, therefore, referred to as the **sensor**. Controlling the shower temperature is the objective of the control operation; therefore, the shower temperature is the **controlled variable (CV)**. The desired shower temperature is called the **setpoint** of the control loop. The flow of cold water into the showerhead is used to control the temperature of the shower and is referred to as the **manipulated variable (MV)**. The valve on the cold-water line and the person's hand are used to change the flow rate of cold water and are referred to as the **actuator**, which is known industrially as the **final control element**. The person in the shower senses the temperature and combines this with past experience with the cold-water valve (i.e., process knowledge) to determine how much to turn the valve; therefore, the person in the shower is acting as the **controller**. Manipulating the flow rate of the cold water based on the sensed temperature of the shower is an example of **feedback control** or **closed-loop control**. The combination of the sensor, setpoint, controller, actuator and process comprise the feedback control loop.

When a change in the valve on the cold-water line is made and the temperature of the shower is allowed to move to a new steady-state condition without any further changes in the cold-water flow, the resulting time behavior of the shower temperature is called the **open-loop response** of the process. Therefore, an open-loop response does not involve feedback from the measured value of the CV.

A changing hot-water temperature is called a **disturbance** (this is known as a disturbance variable, **DV**) to the process because it is not directly controlled and it affects the shower temperature. Another disturbance can occur when someone flushes a nearby toilet if adequate supply pressure is not available for the cold water. After the toilet is flushed, the shower temperature increases sharply. When you hear someone flushing the toilet, if you adjust the cold water flow before the shower temperature increases in an effort to reduce the resulting temperature increase, that is an example of **feedforward control,** which involves compensating for known disturbances before they affect the process. If you step out of the shower to avoid being burned by the hot water when you hear someone flushing a toilet, that is an example of **override control**.

For a poorly designed shower, when the valve on the cold-water line is adjusted, it may take several seconds before the shower temperature changes. This can result when the velocity of water in the cold-water line is low and/or when the piping from the hot- and cold-water mixing point to the showerhead is excessively long. As a result, when a change in the cold-water valve is made, the person in the shower must wait until the effect on the shower temperature has occurred before making another adjustment to the valve. The time delay between a change in the cold-water valve and the start of the resulting change in the shower temperature is the **deadtime** of the process and deadtime makes feedback control much more difficult.

Return to the case in which the hot-water heater is running out of hot water. As the temperature of the hot water decreases, the cold-water flow rate must also decrease to maintain the desired shower temperature. Eventually, the flow rate of cold water is shut off at which point the shower temperature equals the hot-water temperature. But as time goes on, the temperature of the hot water and, as a result, the shower temperature drops below the desired level. Because the cold-water flow rate cannot be negative, a zero flow rate for the cold water represents a **process constraint** or limit to the process.

An old valve on the cold-water line can exhibit significant “sticking”. That is, if the valve is opened a small amount, no change in the flow results. If additional small changes are made, eventually a large change in the flow rate of the cold water results, causing a sharp decrease in the shower temperature. Then, if the valve is closed in small steps, a number of steps is required before the flow rate of cold water decreases and the resulting increase in the shower temperature is relatively large. A sticking valve affects the metering precision of the flow rate that can be attained and is referred to as the **valve deadband.** The more a valve sticks, the larger the deadband it exhibits. Because the valve affects how accurately the MV is controlled, significant valve deadband can affect how precisely the CV is controlled, i.e., variability in the shower temperature.

To this point, we have considered that the flow rate of the cold water is used to control the shower temperature, which is an example of a **Single-Input/Single-Output** (**SISO**) process because one MV (the flow rate of cold water) is used to control one CV (the shower temperature). Assume now that it is desired to control the shower temperature using the flow rate of the cold water and to control the total flow rate of the shower water using the hot-water flow rate. This is an example of a **Multiple-Input/Multiple-Output** (**MIMO**) process because there are two MVs (the flow rates of hot and cold water) and two CVs (the shower temperature and the total flow rate of water for the shower). If the cold-water flow rate is adjusted to control the shower temperature, the shower flow rate also changes. Likewise, if the hot-water flow rate is adjusted to control the shower flow rate, the shower temperature also changes. This is an example of **coupling** or **interaction** between control loops in a MIMO process. Instead of using the cold-water flow rate to control the shower temperature and the hot-water flow rate to control the shower flow rate, the hot-water flow could be used to control the shower temperature and the cold-water flow rate could be used to control the shower flow rate. In addition, other MVs can be used, e.g., the ratio of the cold-water flow rate to the hot-water flow rate. The **pairing of MVs and CVs** has a very significant effect upon the control performance for MIMO processes (Chapter 15).

If a small change in the valve position results in a relatively large change in the shower temperature, the shower is a **high-gain process.** Conversely, if a large change in the valve results in a moderate or small change in the shower temperature, the process is a **low-gain process.** That is, the process gain is the ratio of the change in the CV caused by the change in the MV. At the beginning of a shower when the hot water is at its highest temperature and a relatively small amount of hot water is being used in the shower, small changes to the valve on the hot water can produce relatively large changes in the shower temperature, corresponding to a high-gain process. On the other hand, as the hot-water heater runs out of hot water, the temperature of the hot water is near the desired temperature for the shower and the flow rate of the hot water is high. At these latter conditions, large changes in the valve on the hot water cause relatively small changes in the temperature of the shower, corresponding to a low-gain process. When the gain of a process changes as the operating conditions change, the process is referred to as a **nonlinear process** and most chemical and bio-processes have some degree of nonlinearity.

Controlling the Water Temperature in a Bathtub. Consider filling a bath tub with water while attempting to maintain a specified temperature of the water in the tub. The process is the water in the bathtub. The setpoint is the desired bath-water temperature. The sensor is the person's hand immersed in the bath water to sense the water temperature. Also, the person's hand is used to sense the temperature of the water added to the tub. The MV used to control the temperature of the water in the bathtub is the temperature of the water entering the bathtub and the actuator is the valve on the hot water line and the person's hand. The person is the controller for this process. This process is called a **semi-batch process** because water enters the tub but does not exit. The shower is an example of a **continuous process** because water continuously enters and leaves the shower. A **batch process,** with a few exceptions, has neither feed streams entering nor product streams leaving the system.

For the bathtub example, there are two separate control loops: the control loop for the tub-water temperature and the control loop for the temperature of the water entering the tub. If a person samples the tub water and finds that it is too hot, he or she must reduce the temperature of the water entering the tub by adjusting the valve on the hot-water line. This is an example of **cascade control** because the control loops are applied in tandem, i.e., the setpoint for the temperature controller on the water entering the tub is determined by the temperature controller for the tub water. As the setpoint for the temperature of the water entering the tub is changed, the valve on the hot-water line is adjusted to meet this setpoint.

When adjustments are required for the water temperature in the tub, it takes some time for a correction in the measured temperature to result. The "inertia" of the process is referred to as the **lag** of the process and depends on the volume of water in the tub and the volumetric flow rate of water into the tub. That is, as the size of the tub and the level of water in the tub is reduced for a set water feed rate, the temperature of the water in the tub changes faster than in a tub with more water in it, i.e., the smaller tub exhibits less thermal lag than a larger tank.

Driving a Car. Consider a person driving a car. The objective (i.e., setpoint) is to keep the car in its lane on the road. The sensor is the person's eyes and the CV is the position of the car on the road. The MV is the steering wheel position while the actuator can be thought of as the person's hands and arms that turn the steering wheel. Loose steering (i.e., the steering wheel can be moved back and forth without changing the direction of the car) represents deadband in the actuator/MV system. The controller is the person driving. Curves in the road are disturbances and driver anticipation of a curve is feedforward control. Wear on the tires or a worn suspension represent very slow disturbances to this process. When the driver sees that the car is drifting to the center or the shoulder of the road, the driver takes corrective action, which is feedback control.

Cruise Control on an Automobile. The sensor for the cruise control unit can be two magnets attached to the drive shaft in combination with an electrical coil that generates a current. The current generated by the electrical coil is directly proportional to the revolutions per minute that the drive shaft is turning, which in turn is directly proportional to the speed of the automobile. When the automobile reaches the desired speed, the setpoint for the cruise control unit is specified by pushing the "set button". The setpoint for the controller is actually the electrical current that is generated by the sensor when the set button is pushed. The CV is the generated electrical current from the sensor, which correlates directly with the speed of the automobile. The actuator is the throttle position on the engine while the MV is the flow of gasoline and air to the engine. Hills or a loss of power produced by the engine represent disturbances to the cruise control unit. Because changes to the throttle position made by the cruise control unit are directly proportional to the error from setpoint, the cruise control system acts as a **proportional-only controller**. When a steep hill is encountered, a small but persistent error from the specified automobile speed results. This constant error from setpoint is referred to as **offset**. If the driver pushes the resume button when the automobile is far below the setpoint speed, the throttle fully opens and the maximum flow rate of gasoline and air to the engine results. This is an example of a **saturated manipulated variable** because the MV of the process is at its highest level.

Balancing a Spoon on a Finger. Consider a person balancing a spoon on his/her finger. The objective is to keep the spoon pointing vertically, balanced on the finger. The sensor for this control system is the person's eyes and the MV is the horizontal location of the finger. The actuator is the person's hand and arm. This is an example of an **open-loop unstable process** because if control is not applied, the spoon will fall off the person's finger. For the shower example, if you were to stop adjusting the cold-water valve, the shower temperature will stabilize at some temperature even though it might not be the desired one. The shower process is an example of an **open-loop stable process**.

Self-Assessment Questions

Q1.2.1 In general, what function does a controller perform in a control loop?

Q1.2.2 In general, what function does a sensor perform in a control loop?

Q1.2.3 In general, what function does an actuator perform in a control loop?

Self-Assessment Answers

Q1.2.1 The controller make adjustments to the MV to maintain the CV at its setpoint.

Q1.2.2 The sensor measures the current value of the CV and transmits it to the controller.

Q1.2.3 The actuator receives the output from the controller and accordingly makes changes in the level of the MV.

Self-Assessment Problems

P1.2.1 Consider a person steering a bicycle as they ride down a street. Identify the controller, the actuator, the process, the sensor and the CV. Also, indicate the setpoint and potential disturbances. Remember that the actuator affects the process to change the CV.

Self-Assessment Answers

P1.2.1 The CV is the location of the bicycle and rider on the street. The MV is the orientation of the handlebars in relation to the frame of the bicycle. The driver's hands and arms serve as the actuator for this system. The controller is the driver and the process is the bicycle and rider on the street. The setpoint is the desired path on the street and disturbances include cars (stationary and moving), potholes in the street and turns in road.

1.3 Control Diagrams and P&IDs

Control diagrams, which consist of a **process flow diagram** (**PFD**) along with the primary control loops indicated on it, will be used to represent process control loops in this text. Figure 1.3.1 presents several of the symbols used in this text for control diagrams. Note that there are two different symbols used to represent a shell and tube heat exchanger. Figures 1.3.2a is an example of a control diagram for a temperature control loop applied to a heat exchanger. The PFD is a simplified drawing of each major piece of equipment along with the flow connections between the pieces of equipment. The PFD portion of Figure 1.3.2a shows that the feed stream enters the heat exchanger header and flows on the tubeside through the heat exchanger. Steam enters the exchanger on the shellside, condenses on the outside of the tubes and leaves as condensate. The primary control loops are indicated on the PFD by symbols for the actuator (Figure 1.3.1), the controller and the sensor (Table 1.1). The actuator is represented by a control valve symbol and its location indicates which stream flow is used as the MV for the control loop. The controller is represented by a circle with the type of controller defined by the letters inside the circle while the last letter is a "C" for controller (e.g., TC for temperature controller). Note that the controller compares the measured value of the CV to the setpoint and sends the resulting control signal to the

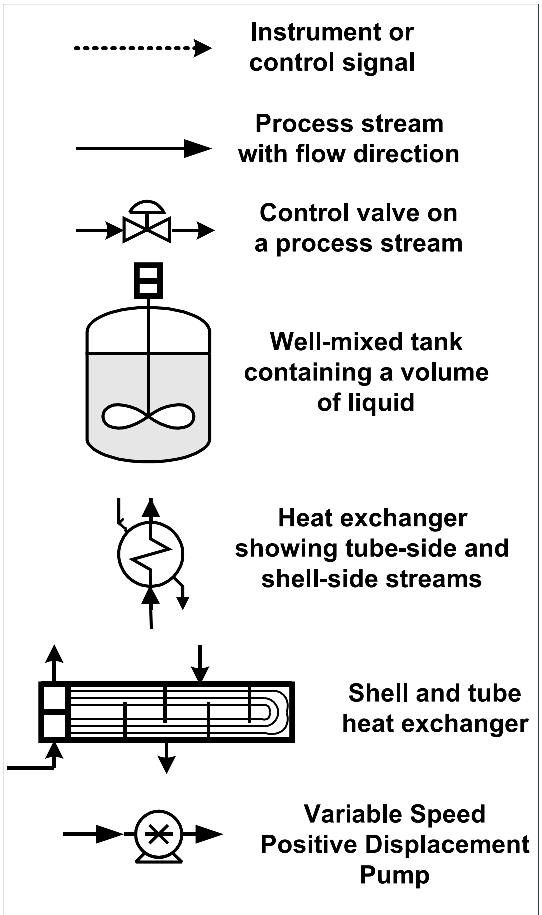


Figure 1.3.1 Definition of symbols used in control diagrams in this text.

actuator (control valve). The sensor is represented by a circle with the type of sensor defined by the letters inside the circle while the last letter is a "T" for transmitter (e.g, TT for a temperature transmitter or in other words a temperature sensor transmitter). The sensor location indicates the CV for the control loop. Note the dashed lines represent signal transmission between elements of the control system (i.e., the measurement signal from the sensor is sent to the controller and the controller output signal is sent to the actuator). Figure 1.3.2a has a setpoint signal entering the controller. For simplicity in future chapters, the setpoint will not be included on control diagrams because it will be assumed that each controller will always have a setpoint provided to it. Table 1.1 list the symbols for the control diagrams used in this text.

Industry also uses control diagrams similar to Figure 1.3.2a to provide simplified descriptions of the primary control loops on a process. In addition, industry documents their control loops using **piping and instrumentation diagrams (P&IDs).** P&IDs provide a very detailed description of the process equipment with all the valves, pumps, piping specifications and sensors indicated on the diagram. P&IDs also provide a detailed layout of all control loops, indicating the connections between the control functions and the sensor and actuators. Figure 1.3.2b is a part of a P&ID for the same heat exchanger and temperature control loop shown in Figure 1.3.2a. Note that all process streams are numbered in a diamond-shaped box and the source or destination of each stream is indicated by a P&ID drawing number. Each controller and sensor in a P&ID has a unique number and each instrument (e.g., a temperature transmitter or a field-mounted pressure gauge) is shown whether or not it is part of a control loop. Process measurements that are available to the

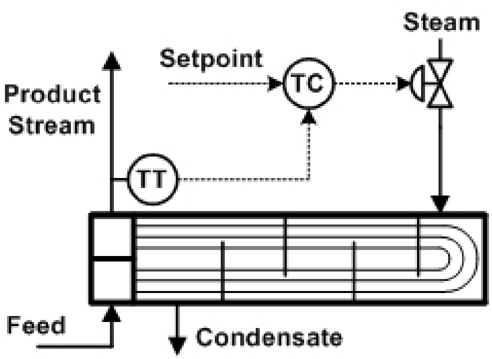


Figure 1.3.2a Control diagram for a temperature control loop applied to a steam-heated heat exchanger.

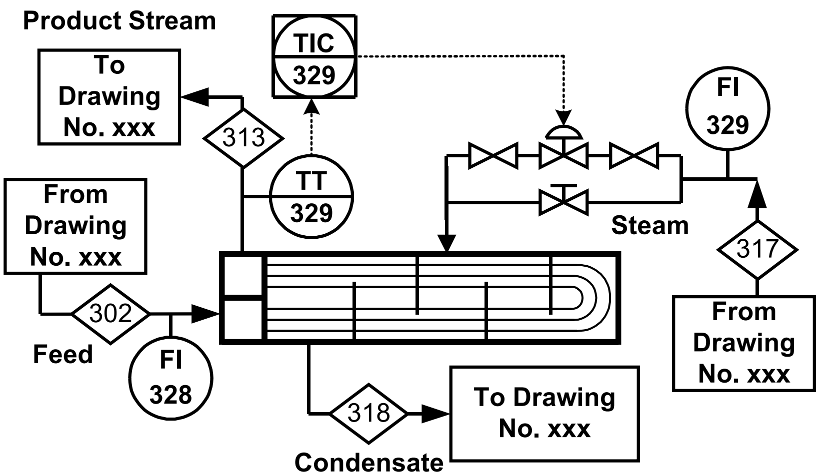


Figure 1.3.2b A P&ID for the same process shown in Figure 1.3.2a.

control computer are designated by a circle with a horizontal line through the middle of the circle with the type of instrument (see Table 1.1) above the line and the instrument number below the line. Field mounted instruments are designated with a circle without a horizontal line but contain the symbol for the instrument and the instrument number. While control diagrams show only the lines, valves, and instruments that are associated with the control loop of interest, P&IDs show all lines, valves and instruments, and indicate whether the instruments provide a reading to the control computer. For example, in Figure 1.3.2b the feed flow and the steam flow rates are measured by FT328 and FT329, respectively, but these measurements are not used by the temperature control loop and do not appear on the control diagram (Figure 1.3.2a). Also, control valves are usually implemented industrially with block valves and a by-pass valve as shown in the P&ID so that the control valve can be removed for repair without shutting down the process, but control diagrams use only the control valve symbol.

Table 1.1 **Definitions of Controller and Instrument Symbols for Control Diagrams.**
AC - analyzer controller (i.e., composition controller)
AT - analyzer transmitter (i.e., composition analyzer/transmitter)
CC - electrical conductivity controller
CT - electrical conductivity sensor/transmitter
DPC - differential pressure controller
DPT - differential pressure sensor/transmitter
FC - flow controller
FF - feedforward controller
FT - flow sensor/transmitter
HS - high select (this element selects the larger of two or more inputs)
LC - level controller
LS - low select (this element selects the smaller of two or more inputs)
LT - level sensor/transmitter
PC - pressure controller
pHC - pH controller
pHT - pH sensor/transmitter
PT - pressure sensor/transmitter
RSP - remote setpoint
S - select controller (i.e., chooses which MV to use)
TC - temperature controller
TT - temperature sensor/transmitter
VPC-valve position controller
- summation block (i.e., for adding or subtracting inputs)
+ - an addition function (i.e., adds two inputs to determine the output)
- multiplication function (i.e., forms the product of two inputs)

P&IDs, due to their detail, are complex and generally are spread over a number of sheets to represent a processing unit. Therefore, the control engineer may have to use several P&IDs to understand the existing controls for a single process unit. Appendix B provides a more thorough introduction to P&IDs.

1.4 Industrial Process Control Examples

Flow Controller. The most common control loop in the CPI is a flow controller (Figure 1.4.1). Many times this control system is used to track the flow rate changes called for by higher-level controllers. The sensor in this system is usually a combination of an orifice plate (a plate with a hole in it that is smaller than the pipe diameter) flanged into the line and a device that measures the pressure drop across the orifice, which is directly related to the flow rate. The actuator is the control valve in the line. The process is the fluid in the pipe and in the control valve. The flow controller (FC) compares the measured flow rate with the specified flow rate (i.e., the flow setpoint) and opens or closes the control valve accordingly. A DV for this process is a change in the upstream pressure for the process stream. A more complete analysis of the operation of a flow controller is presented in Section 7.9.

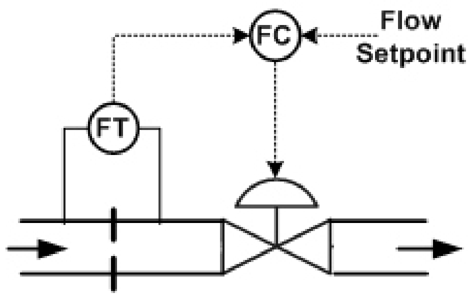


Figure 1.4.1 Control diagram of a flow controller.

Continuous Stirred Tank (CST) Thermal Mixer. Figure 1.4.2 shows a schematic for a CST thermal mixer with two feed streams at different temperatures for which it is desired to maintain a specified temperature for the product. The setpoint is the desired product temperature and the CV is the product temperature. The MV is the flow rate of Feed 1 to the mixer and the actuator is the control valve on that line. The sensor is a temperature sensor/transmitter (TT) located in the product line. The process is the fluid inside the CST thermal mixer. The temperature controller (TC) compares the measured temperature with its setpoint and makes changes to the control valve on feed stream 1. A DV for this process is a change in the temperature of the feed to the CST. The volume of the mixer divided by the total volumetric flow rate to the mixer is the **residence time** or the average time that an element of feed spends in the mixer, assuming that the liquid level in the CST is held constant. For this process, the **time constant** of the process is equal to the residence time. The time constant is a measure of how fast the temperature in the mixer can change after an input change. As a rule of thumb, **you can assume that it takes approximately four time constants to observe the full effect of a step change of an input to a process** under open-loop conditions, provided no other input changes to the process occur. This is explained in more detail in Chapter 6.

Continuous Stirred Tank Composition (CST) Mixer. In a manner similar to the CST thermal mixer, the CST composition mixer combines two streams with different concentrations and the control objective is to maintain a specified composition of the product (Figure 1.4.3). The setpoint is the desired concentration of the CST product and the MV is the flow rate of Feed 1. The actuator is the control valve on the manipulated feed stream. The process is the fluid inside the CST composition mixer. A DV for this process is a change in the composition of a feed stream. The sensor is a composition analyzer that analyzes samples taken from the product line. Many times in the CPI, gas chromatographs (GCs) are used as on-line analyzers and provide new composition measurements typically every three to ten minutes. The time between new measurements is referred to as **analyzer deadtime** or **analyzer delay**. The controller compares the measured composition of the product with its setpoint and makes changes to the control valve on the feed flow to the mixer.

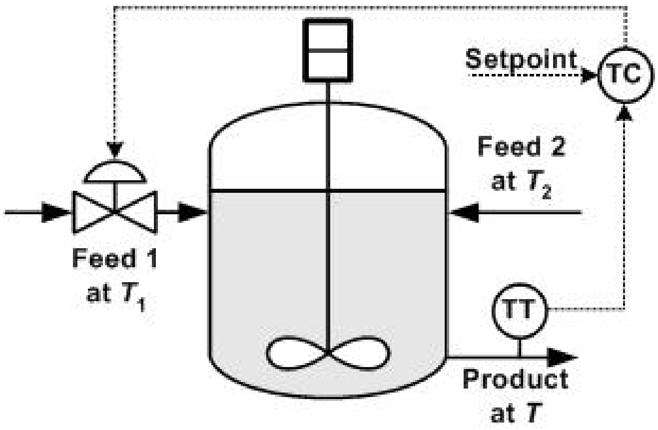


Figure 1.4.2 Control diagram of a CST thermal mixer with temperature controller.

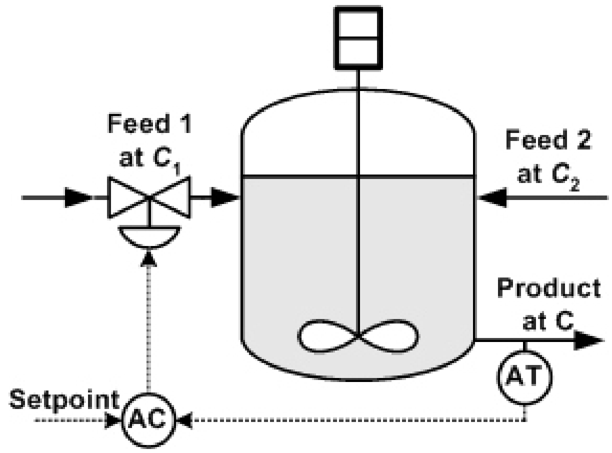


Figure 1.4.3 Control diagram for a CST composition mixer.

Level Control in a Tank. Figure 1.4.4 shows a control diagram for a level controller on a tank. Maintaining levels in tanks is a common objective in the process industries. The setpoint is the desired level in the tank and the CV is the level in the tank. The MV is the exit flow from the tank and the actuator is the control valve on the outflow line. The sensor is the level indicator on the tank (LT) and changes in the inlet flow rate are disturbances to the process. The level controller (LC) compares the measured level with the setpoint for the level in the tank and makes a change to the control valve on the exit flow from the tank.

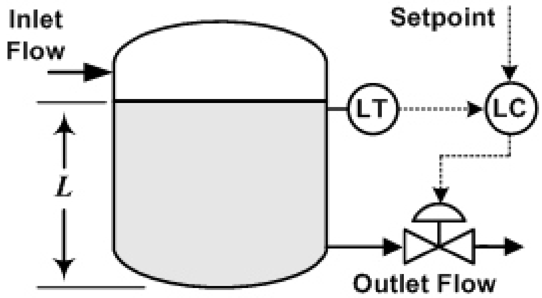


Figure 1.4.4 Control diagram of a tank with a level controller.

Endothermic Continuous Stirred Tank Reactor (CSTR). A control diagram for an endothermic CSTR is shown in Figure 1.4.5. CSTRs are commonly used reactors in the CPI. The control objective for this system is to maintain the temperature of the product stream at a specified level. The CV is the temperature of the product leaving the reactor and the MV is the feed rate of steam to the heat exchanger. The process is the fluid in the CSTR, the heat exchanger and associated lines. The sensor is a temperature sensor/transmitter (TT) that measures the temperature of the product stream leaving the reactor. The temperature controller (TC) compares the measured value of the product temperature with its setpoint and makes changes to the control valve on the steam to the heat exchanger. DVs for this process include changes in the steam pressure, feed composition and feed flow rate.

Aerobic Fermentation Process. A control diagram for a dissolved oxygen controller applied to a batch **aerobic fermentation process** (i.e., with O_2 present) is shown in Figure 1.4.6. The control objective is to maintain a specified dissolved oxygen concentration in the fermentation reactor so that the cells in the process have adequate oxygen levels. The CV is the measured dissolved oxygen concentration in the fermentor and the MV is the air flow rate to the fermentor. In addition, the rotational speed of the mixer can also be used as the MV for a DO control loop. The process is the fluid in the fermentor and the sensor is the dissolved oxygen sensor-transmitter (AT), which is also know as a dissolved oxygen sensor. The dissolved oxygen controller (AC) compares the measured value of the dissolved oxygen from the AT to the desired dissolved oxygen level and sets the air flow rate to the fermentor. The actuator for this control loop is a variable speed air compressor while for

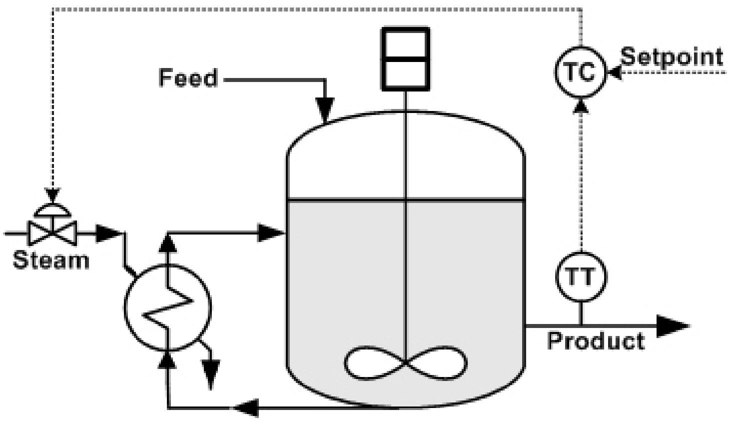


Figure 1.4.5 Control diagram of an endothermic CSTR with a temperature control loop.

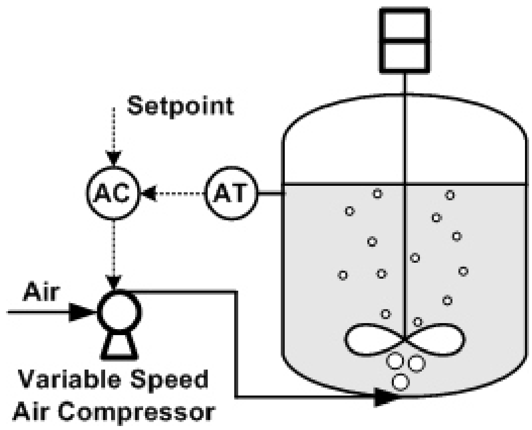


Figure 1.4.6 Schematic of a bio-reactor with a dissolved oxygen controller.

the previous chemical engineering examples the actuator was a control valve. A DV for this process may be changes in the rpm of the mixer impeller.

Fed-Batch Bio-Reactor. A control diagram for a fed-batch bio-reactor is shown in Figure 1.4.7. Fed-batch bio-reactors are used for the production of a wide range of products, e.g., bio-fuels and antibiotics. The control objective is to maintain a specified glucose concentration in the bio-reactor. The CV is the glucose concentration in the bio-reactor and the MV is the feed rate of the glucose and nutrients mixture to the bio-reactor. The process is the liquid broth in the bio-reactor and the sensor is a glucose sensor (AT), which directly measurement the glucose concentration in the broth. The glucose concentration controller (AC) compares the glucose concentration to the specified level and based on this difference chooses the glucose feed rate. The control signal is sent to the actuator, which is a variable speed pump. A DV for this process is a change in the concentration of glucose in the feed.

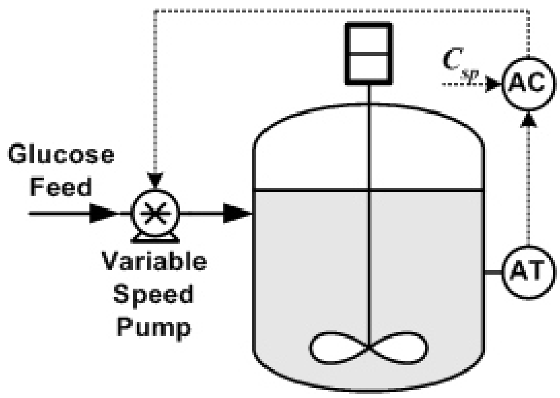


Figure 1.4.7 Control diagram for a glucose concentration controller on a fed-batch reactor.

Bio-Separator. Figure 1.4.8 shows a conductivity control loop applied to a bio-separator that is used to remove salt from the protein products in the holding tank. The CV is the salt concentration in the exit stream from the filtration module, which is measured by a conductivity (analyzer) sensor (CT), and the MV is the circulation rate of the protein/salt solution. The process is the salt in the solution in the reservoir, filtration module and the lines. The conductivity controller compares the measured conductivity in the salt solution to the conductivity setpoint and based on this difference sends a control signal to the variable speed pump, which changes the circulation rate. Note that a level controller adds water to the holding tank to maintain a constant level to compensate for the permeate removed from the process.

Self-Assessment Questions*

Q1.4.1 Using the control diagram shown in Figure 1.4.2, identify the CV for the control loop. Explain how you used this figure to determine the CV.

Q1.4.2 Using the control diagram shown in Figure 1.4.5, identify the MV for the temperature control loop. Explain how you used this figure to determine the MV.

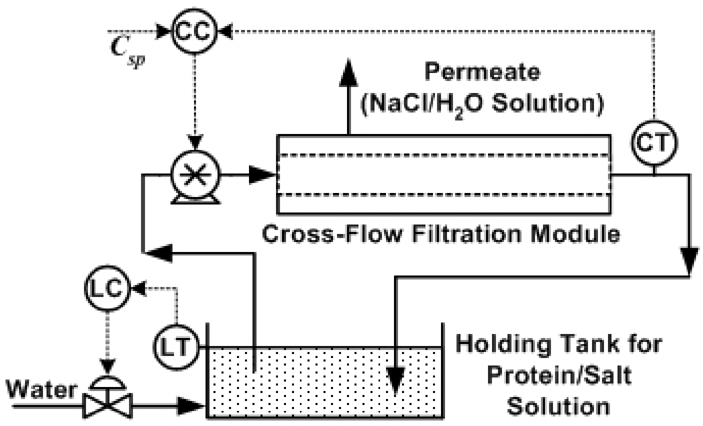


Figure 1.4.8 Control diagram for a composition control loop for a bio-separator for concentrating cells based on a cross-flow filtration module.

Self-Assessment Answers

Q1.4.1 The CV is the temperature of the product stream. The CV can be located by identifying the sensor that sends its output to the controller.

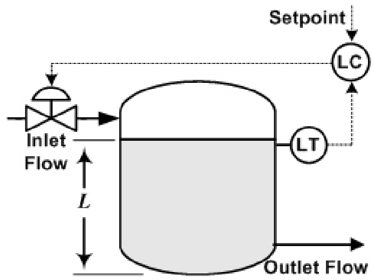
Q1.4.2 The MV is the flow rate of steam to the CSTR. The MV can be located as the stream which is effected by the actuator system, e.g., a control valve or pump, which receives the controller output.

Self-Assessment Problems*

P1.4.1 Draw a control diagram for a level controller on a tank for which the flow rate of the feed to the tank is the MV for the level controller.

Self-Assessment Answers

P1.4.1 The solution is shown in the adjacent figure.



Solution for P1.4.1

1.5 Block Diagram of a General Feedback Control System

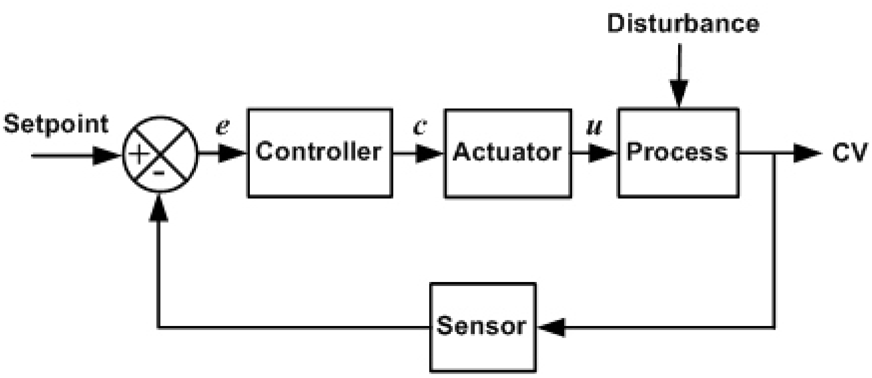


Figure 1.5.1 Block diagram of a general feedback control loop. e is the error from setpoint, c is the controller output and u is the MV.

Figure 1.5.1 shows a block diagram representation of a generalized feedback control system. Block diagrams are also known as logic flow diagrams because they demonstrate the flow of information and the effect of one process element on another. Block diagrams do not represent the physical layout of a process. For example, for the level loop shown in Figure 1.4.4, the outlet flow from the tank is an input for the tank level control system. This diagram can represent each of the previous examples: everyday examples of control and CPI and bio-tech examples of control. That is, each of the previous examples has a controller, an actuator, a process and a sensor, in that order, along with feedback of the measured value of the CV to the controller. In addition, each of the examples is affected by disturbances. The sensor reading is compared with the setpoint to produce the **error from setpoint** (e) and the controller chooses the control action (c) based upon this difference. The actuator system is responsible for making changes in the level of the MV (u) based on changes in c. The "process" for a control loop is only the part of the system that determines the relationship between the inputs (i.e., the MV and disturbances) and the CV. The process considered here can be based upon a number of processing units or one small part of a unit operation.

The symbol in Figure 1.5.1 represents a summation function. The negative sign on the measurement of the CV results in forming the difference between the setpoint and the measured value of the CV, which is the error from setpoint (e). A block diagram of an open-loop process (without feedback control) involves only the actuator, process and sensor.

1.6 Types of Controllers

On-Off Control. An on-off controller applies two modes of control action: full control action and no control action. Consider Figure 1.6.1. Initially, the on-off heater applies heat to the system as long as the controlled temperature is below a specified upper limit. When the upper temperature limit is reached, the addition of heat is stopped. Heat is not added to the process until the controlled temperature becomes less than a specified lower limit. In this manner, the heat is intermittently applied and the controlled temperature cycles between the specified upper and lower limits. Note that the control temperature exceeds the maximum temperature and becomes less than the minimum temperature due to the inertia of the process. In the case of overshoot and undershoot, the upper and lower temperature limits can be adjusted so that the effective maximum and minimum controlled temperatures correspond to the desired maximum and minimum temperatures. A typical room thermostat is an example of an on-off controller.

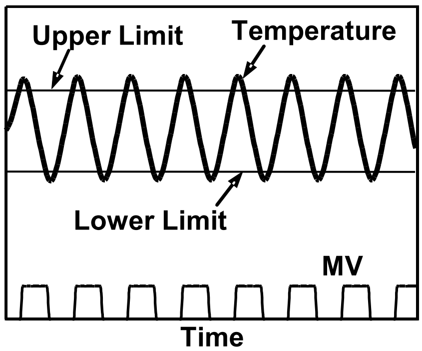


Figure 1.6.1 An example of the operation of an on-off temperature controller.

Manual Control. When process operators act as the controller, this is referred to as manual control. Operators typically make changes in MV levels and wait to see where the process is going to settle before making the next MV change. In this manner, the operator adjusts the MV in a conservative fashion by using a series of steady-state or near steady-state steps. The time between MV changes is set by the **open-loop settling time** of the process, i.e., the time for the process to reach a new steady-state operating point after a single step change in the MV level has been applied to the process.

Consider manual control for a single change in a disturbance. Operators see the initial effect of the disturbance and make their estimate of the needed change in the MV. Then they observe the response of the process (i.e., wait for approximately one open-loop settling time). After the process has settled or nearly settled, they make another adjustment to the level of the MV and observe its effect. It typically takes several adjustments to return the process to the desired setpoint. Obviously, when the process is subjected to frequent disturbance upsets, operators are continually making changes to the process and observing the results in an effort to keep the CV near its setpoint.

Proportional-Integral-Derivative (PID) Controllers. The most common controller in the CPI and bio-tech industries is the PID controller, which is the primary emphasis of this text. The PID controller provides adjustment to the actuator, which is proportional to the error from setpoint, the time integral of the error and the time derivative of the error (Chapter 7). Properly implemented PID controllers provide substantial performance improvement over manual control. For a single disturbance, a properly tuned PID controller typically returns the process to or near setpoint in approximately one open-loop settling time or less. For a series of disturbances, the PID controller significantly outperforms manual control in terms of the resulting variability about setpoint.

Advanced PID Controllers. PID controller performance can be enhanced by a number of techniques including cascade control, ratio control, feedforward control, scheduling of controller tuning, decouplers and anti-windup (Chapters 12-14). These techniques are designed to assist PID controllers with regard to more effectively handling disturbances, deadtime, coupling, etc.

Model-Based Controllers. Controllers that directly use process models to determine the control action can handle process nonlinearity, disturbances, coupling and complex dynamics if the model used in the controller accurately represents these factors. The most common type of model-based controller is model predictive control (Chapter 16), which uses linear process models to control MIMO processes. Model-based controllers can offer significant performance improvement over PID controllers in certain cases when properly implemented and maintained.

1.7 Responsibilities of a Chemical Process Control Engineer

Overview. From an overall point of view, process control engineers are responsible for using their process control skills and process knowledge **to make money** for the operating company. Specifically, process control engineers are responsible for seeing that a plant, which was not generally designed with regard to process control performance, operates safely, reliably and efficiently while producing the desired product quality. Control engineers do this by transferring the variability that would otherwise go into important CVs (e.g., product compositions) to MVs (e.g., steam flow rates) and less important CVs (e.g., byproduct compositions). To meet these objectives, it is essential that process control engineers use a complete knowledge of the process to attain the most desirable performance of the process.

Process Specification: All new plants and major retrofits should begin with development of a Functional Requirements document, approved by all stakeholder groups. This document defines the requirements associated with the process and is the basis of subsequent project activities, such as, design and testing. The control engineer should be a member of the team that develops this document because the requirements of this document need to be practical with respect to process control and should accommodate the relevant constraints of the plant. The control engineers also needs to help determine if any process control customization is required and perhaps be part of the effort to develop such customization.

Process Design. The design of a process can have a dominant effect on the ultimate control performance of the system. That is, the reduction in holdup of material in the process and the application of material recycle and energy recovery can each have a very significant economic advantage from a steady-state design point of view. On the other hand, these steady-state advantages can, in the extreme, render the process impossible to control. For this reason, many companies include a control engineer on their process design teams. Design modifications can sometimes convert an existing plant from a poorly performing process into a well performing system from a control standpoint. The control engineer may also be responsible for specifying the type of sensors and control valves.

Controller Design. The control engineer is responsible for selecting the proper mode of a PID-type controller (P-only, PI or PID) depending upon the characteristics of the process. That is, the P-only controller uses only proportional feedback action, the PI controller uses proportional and integral action and PID uses proportional, integral and derivative action. Each of these controller options has advantages when applied in the proper situation. The control engineer is also responsible for applying advanced PID versions (cascade, ratio, feedforward, etc.) to the cases in which these methods offer significant advantages. Therefore, the control engineer must understand the advantages and disadvantages of standard and advanced PID techniques as well as understand the control-relevant aspects of the process in question.

Control system implementation, testing and tuning. The control engineer is involved in many aspects of implementing and testing control systems, including tuning. Controller tuning is critically important for the proper functioning of a controller. Selecting the values of the controller settings involves considering how a control loop affects the overall process and many times involves a compromise between controller performance and controller reliability (Chapter 9). Implementation also includes for example, ensuring that all sensors are mounted correctly (e.g., flow sensors at sufficient distance from piping bends), and that valves fail closed or open, as specified, that all alarms are configured correctly. More generally, this phase of a project ensures that all functional requirements related to process control have been met.

Controller Troubleshooting. Even a properly designed and tuned controller may not function properly. For example, an erratic sensor or an improperly functioning actuator can seriously undermine the performance of a controller. Excessive disturbance levels can also be the source of unacceptable controller performance. It is the responsibility of the control engineer to identify the reasons for improperly functioning control loops and correct them as much as possible (Chapter 10).

Documentation of Process Control Changes. Any significant change to a process, including changes to the process controls, requires approval by a safety review committee for most companies in the CPI. Before approving changes in the controls for a process, the safety review committee typically requires that a data sheet for process changes, which describes the proposed process changes, be completed. The data sheet for process changes must be approved by the operational and management authorities for the affected area of the plant before the changes can be implemented. The process control engineer is responsible for completing the process change data sheet, obtaining the approval signatures and presenting the process modifications to the safety review committee for process control changes. The process control engineer usually serves on the safety review committee to assess the effects of process modifications to the process control systems. For the bio-tech industries, safety is less of an issue for process control changes, but process control changes usually require approval by a product quality committee.

Types of Process Control Engineers. In general, there are several levels of responsibility to which control engineers are assigned. For example, for an entry-level assignment, process control engineers might have the responsibility for the day-to-day operation of the regulatory control loops (i.e., tuning and troubleshooting) in their portion of the plant. Control engineers with five or more years experience or with graduate training in process control may be responsible for larger, more challenging control projects throughout the plant. These control engineers are involved in designing, tuning and troubleshooting controllers. Finally, corporate level control engineers typically are stationed at corporate engineering headquarters and are involved in advanced control projects at a variety of plant sites. These engineers typically have an advanced engineering degree often focusing on process control within chemical or electrical engineering. For each of these job levels, effective process control engineers use their knowledge and experience on control systems combined with a thorough understanding of the chemical processes with which they are working to solve process control problems. The key point here is that for process control engineers to perform effectively, they must have a thorough knowledge of the process behavior and the operational objectives and constraints of the process, i.e., **process knowledge is a prerequisite for becoming a successful process control engineer**.

1.8 Operator Acceptance

A major issue in effectively functioning as a process control engineer is **operator acceptance**. Operator acceptance depends upon developing process control solutions that work effectively in an industrial setting as well as getting along with the operators. The quickest way to ensure that the operators not accept your control work is to deal with them in an arrogant and condescending manner. For example, if you come into the control room with an arrogant attitude, a significant number of operators will take it upon themselves to make sure that you are not successful in their plant. Regardless of how well your controller functions on the process, if you have offended the operators, they will see that your controller does not stay on-line. Remember that at 2 a.m., they are the “kings” of the process.

That said, the best way to interact with the operators is to give them the respect that they deserve and seek their advice and input whenever you start a control project. Remember that they are observing the process day in and day out. They may not correctly know why certain things happen in the process but you can count on them knowing what happens and under what conditions. As a result, operators are a valuable resource of operating experience.

With regard to implementing controllers that function well in an industrial setting, the control engineer must ensure that the proper controller type (Chapters 7, 12 and 13) has been chosen for the process in question and that the controller is properly tuned (Chapter 9). Bumpless transfer (Chapter 14) and anti-windup (Chapter 14) should be used as well. Validity checks and filtering (Chapters 7 and 9) should also be applied to sensor readings. In this manner, controllers can provide good control performance in a highly reliable fashion. The best controllers are the ones that meet their control objectives, stay in service unless there is a sensor or actuator failure and respond “gracefully” when an actuator or sensor failure occurs. In addition, operators need to be presented with relevant prioritized information about the process, organized in an easy to access way and presented in such a way to avoid information overload for the operators in the event of a major plant upset.

1.9 Process Optimization and Process Control

Process optimization is concerned with operating a plant so that the plant produces the highest rate of profit generation for the operating company consistent with the safety of personnel, equipment and the environment. Most optimization applications result from competing factors. Consider the stripping section of a distillation

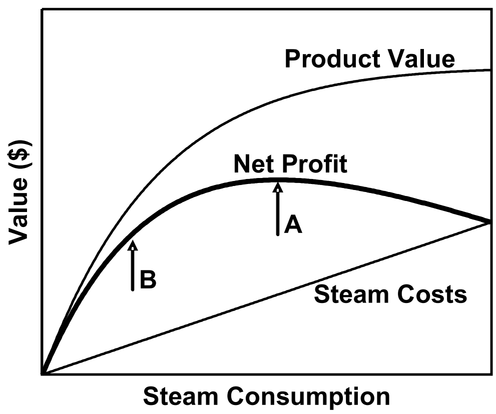


Figure 1.9.1 A plot showing the competing factor for the optimum product recovery for the stripping section of a distillation column.

column. As the steam flow rate to the reboiler is increased, the recovery of the heavy component in the bottom product increases, which corresponds to increased value, but the cost of the steam used obviously increases with steam flow rate. Figure 1.9.1 shows the effect of competing factors in which the value of the increased recovery of the heavy product is combined with the cost of the steam used to yield the net profit. That is,

Net Profit = Product Value - Steam Costs

Note that the net profit goes through a maximum. The maximum in the net profit (Point A in Figure 1.9.1) indicates the optimum operating condition for this process. If the process is operated at this point, the largest rate of profit generation will result. If the steam rate were increased above the value corresponding to Point A, the value of the heavy product in the bottoms would increase, but the increase in value of the bottom product would be less than the cost of the additional steam used. Likewise, if the steam rate were reduced from the optimum rate, the cost of steam usage would decrease, but the reduction in the value of the product recovered would be larger.

Process constraints usually affect the optimum operating conditions for industrial processes. Consider Figure 1.9.1 for which there is a maximum reboiler temperature above which significant polymerization results in the reboiler. A maximum reboiler temperature constraint could limit the reboiler steam flow rate to less than the optimum in Figure 1.9.1 (e.g., point B). For this case, to ensure that the maximum reboiler temperature is not exceeded, the column would be operated at a "safe" distance below the maximum reboiler temperature. The better the control performance for this column, the closer the column can operate to this constraint and the higher the profit for the operation.

Process optimization applied to large-scale systems usually results in optimum operating conditions that require operating the process at a number of process constraints, e.g., maximum equipment throughput, environmental and safety limits. Figure 1.9.2 shows a two-dimensional operating region of a process with a limited feasible operating region. Note that all constraints are satisfied only in the "feasible region" indicated in Figure 1.9.2 and that the optimum operating point is located at the intersection of two constraints. For Figure 1.9.2a the control is quite loose as indicated by the diameter of the circle representing the control region. Figure 1.9.2b shows the same system, but with improved control due to the smaller control region. Note that the superior control shown in Figure 1.9.2b allows closer operation to the optimal operating conditions, and as a result, better economic performance. Therefore, **the performance of the control system that maintains operation at the optimum operating conditions of a process can have a direct effect on the economic performance of the process.**

Figure 1.9.3 shows a logic flow diagram for a typical process optimization application. The optimization process starts with an initial estimate of the optimum operating conditions, which, in the case of the distillation column example, is the initial estimate of the optimum steam flow rate. The numerical optimization algorithm passes the initial estimate of the optimum operating conditions to the model equations where all the operating conditions of the process are calculated. For the distillation example, tray-to-tray material balances that describe the column can be used to determine the recovery of the heavy bottoms product. After the model is solved, the model results

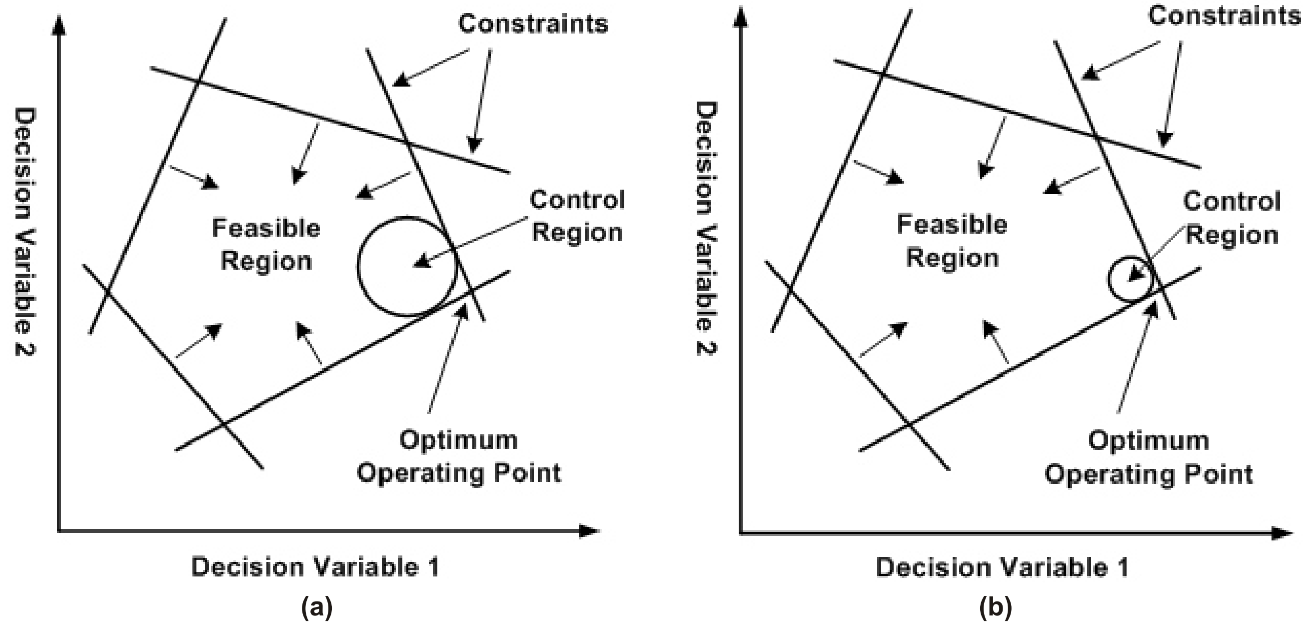


Figure 1.9.2 An example of a 2-dimensional optimization problem with the optimum located at the intersection of two constraints. (a) loose control to the optimum (b) tighter control to the optimum.

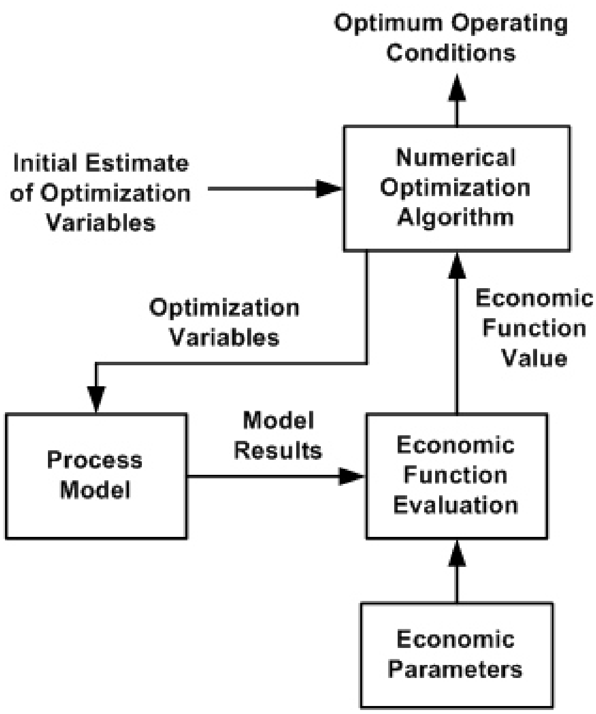


Figure 1.9.3 A logic flow diagram of a process optimization procedure.

are combined with the economic parameters to calculate the rate of profit generation. An economic objective function is typically represented as follows:

$$P_i V_i \qquad F_i V_i \qquad U_i C_i \qquad \textbf{1.9.1}$$

where is the rate of profit generation or the value of the economic objective function, P_i is a production rate of a product or byproduct, V_i is the value or cost per unit mass of a feed, product, or byproduct, F_i is a feed rate, U_i is a utility rate (e.g., steam flow rate) and C_i is the corresponding unit cost for utility usage. V_i and C_i correspond to the economic parameters shown in Figure 1.9.3. The economic objective function given by Equation 1.9.1 is applicable for determining the optimal operation of an existing process and does not consider equipment costs. For the distillation case, the economic objective function would be the value of the heavy product minus the cost of the steam used.

In the optimization procedure shown in Figure 1.9.3, the numerical optimization algorithm calculates values of the optimization variables or the **optimization decision variables** until the optimal value of the economic objective function is identified**.** Once the optimum value of (e.g., the maximum value of is identified the optimization calculations are finished. The corresponding values of the optimum operating conditions (i.e., the optimum values of the optimization decision variables) are typically applied to the process as setpoints for controllers on the process. For most applications, **the process optimizer determines the optimum operating conditions, which are passed to the process controllers as setpoints.**

Example 1.1 Optimization of a CSTR with a Series Reaction

Problem Statement. Figure 1.9.4 shows how optimization can be applied to an endothermic CSTR. Note that the result of the optimizer in Figure 1.9.4 is the optimum reactor temperature while the optimizer is composed of the numerical optimization algorithm, the process model and the evaluation of the economic objective function as shown in Figure 1.9.3. The optimal reactor temperature determined by the optimizer becomes the setpoint for the reactor temperature controller, which is referred to as a **supervisory controller**. The output of the supervisory controller is the setpoint for the flow controller on the steam, which is referred to as a **regulatory controller**. Generally, composition and temperature control loops serve as supervisory control loops while pressure, level and flow control loops are used as regulatory control loops. In summary, the process optimizer determines the setpoints for the supervisory control loops, which in turn select the setpoints for the regulatory control loops, which adjust control valves on the process.

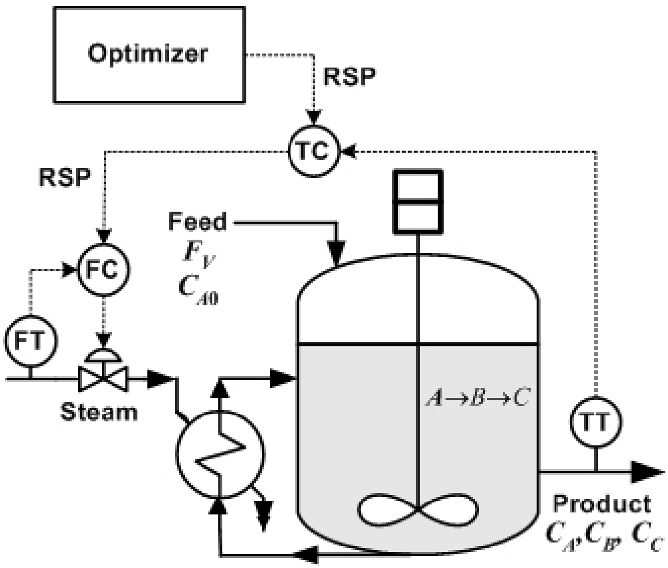


Figure 1.9.4 Schematic of an endothermic CSTR with regulatory and supervisory controls and a process optimizer.

Determine the optimum reactor temperature based on the following economic parameters, operating conditions and reaction parameters for a CSTR (Figure 1.9.4):

Reaction: A ^{r_1} B ^{r_2} C	
$r_1 = k_1\, C_A \exp(-E_1/RT)$	$r_2 = k_2\, C_B \exp(-E_2/RT)$
$k_1 = 3.8604 \times 10^6\ s^{-1}$	$k_2 = 1.8628 \times 10^{13}\ s^{-1}$
E_1®= 5033 K	E_2®= 10,065 K
F_V= 10 l/s	V_r=100 l
C_{A_0} = 1.0 g mol/l	V_{AF}= \$0.15/g mol
V_A=\$0.10/g mol	V_B=\$0.50/g mol
V_C=\$0.20/g mol	

where F_V is the volumetric feed rate, V_r is the volume of the reactor, T is the reaction temperature, V_{AF} is the cost of A in the feed, V_A is the value of A in the product, V_B is the value of the primary product B, V_C is the value of the secondary product C and C_{A_0} is the concentration of A in the feed to the reactor. Assume that the feed contains only A and neglect the utility costs in the optimization analysis of this system. Note that V_A is less than V_{AF}, indicating that the separation costs of recovering A from the product stream are significant. Applying optimization to this process is an attempt to identify the most economically preferred process operation for this system.

Solution. When undertaking an optimization problem, it is best to develop a physical understanding of the process based on the competing factors that form a nontrivial optimization problem. For this reactor, if the reaction temperature is too low, little reaction occurs and only a small amount of product is produced. On the other hand, if the temperature is too high, most of the primary product B reacts to form the lower-valued secondary product C; therefore, there is a temperature between these limits that maximizes the profitability of this reactor, i.e., the optimum reactor operating temperature, T^*. The optimization of this process is analyzed in terms of the logic flow diagram for optimization presented in Figure 1.9.3.

Model development. The model equations for this process are used to calculate all the values of the process operating conditions, which in turn are used to determine the value of the economic objective function given the values of the economic parameters (i.e., V_{AF}, V_A, V_B and V_C). In this case, given the reactor temperature, T, the concentrations of A, B and C in the product (i.e., C_A, C_B and C_C) can be calculated directly using steady-state material balances for A, B and C, respectively. Applying a steady-state material balance for A (i.e., the flow rate of A entering the process minus the flow rate of A leaving the process minus the rate of consumption of A within the process is equal to zero) yields

$$F_V C_{A_0} \quad F_V C_A \quad k_1 \exp[\; E_1 / RT] C_A V_r \quad 0$$

Rearranging and solving for C_A results in

$$C_A \quad \frac{C_{A_0}}{1 \quad \dfrac{k_1 \exp \dfrac{E_1}{RT} \; V_r}{F_V}} \tag{1.9.2}$$

Note that the reaction temperature and the process parameters are used in Equation 1.9.2 to calculate C_A. Similarly applying steady-state mole balances for B and C yield

$$C_B \quad \frac{k_1 \exp \dfrac{E_1}{RT} \; C_A V_r}{F_V \quad k_2 \exp \dfrac{E_2}{RT} \; V_r} \tag{1.9.3}$$

$$C_C \qquad \frac{k_2 \exp \dfrac{E_2}{RT} \quad C_B V_r}{F_V} \qquad \textbf{1.9.4}$$

Equation 1.9.3 requires the value of C_A and Equation 1.9.4 requires the value of C_B; therefore, Equations 1.9.2-1.9.4 can be applied sequentially.

Formation of the objective function. Once the model equations are solved (i.e., Equations 1.9.2-1.9.4), the value of can be calculated by applying Equation 1.9.1 using the values of the economic parameters (i.e., V_A, V_B, V_C and V_{AF}):

$$F_V C_A V_A \quad F_V C_B V_B \quad F_V C_C V_C \quad F_V C_{A0} V_{AF} \qquad \textbf{1.9.5}$$

where has units of \$/s. The solution procedure is as follows: (1) select a reactor temperature, (2) use Equation 1.9.2 to calculate C_A, (3) use Equation 1.9.3 to calculate C_B, (4) use Equation 1.9.4 to calculate C_C, (5) use Equation 1.9.5 to calculate and (6) repeat this process with different reactor temperatures until the maximum profit is identified.

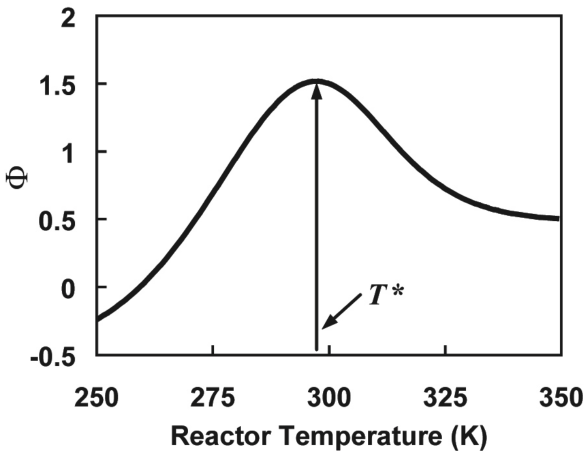


Figure 1.9.5 Graphical Solution to Example 1.1

Results. Figure 1.9.5 shows the effect of the reactor temperature on the economic objective function as well as the optimum reactor temperature, T^*. At reactor temperatures below 260 K, is negative indicating that the feed costs are greater than the product value because of a low reaction conversion. At high temperatures, levels out at the profit level corresponding to converting all the feed to product C. The maximum profitability for the reactor occurs at a reactor temperature of 297 K (i.e., the optimum reactor temperature, T^*). In general, the numerical optimization algorithm as shown in Figure 1.9.5 numerically determines the optimum reactor temperature instead of using a graphical solution as shown in Figure 1.9.3. Because the reactor temperature is the only optimization variable, a plot of the value of versus the reactor temperature is an easy and direct method to identify the optimum reactor temperature. Otherwise, a numerical optimization procedure is required if, for example, there are more than two decision variables, which eliminate the convenient use of graphical techniques.

Periodic evaluation of the optimum reactor temperature may be required because variations in the reactor feed rate, feed composition, feed costs or product values can result in significant variations in the optimum reactor temperature. In this manner, the optimum reactor temperature is calculated on-line to maintain operation at the optimum reactor temperature in the face of variations in process operating conditions and the values of the economic parameters. Finally, according to Figure 1.9.4, the optimum reactor temperature obtained in this procedure becomes the setpoint for the temperature controller on the reactor. This hierarchy of

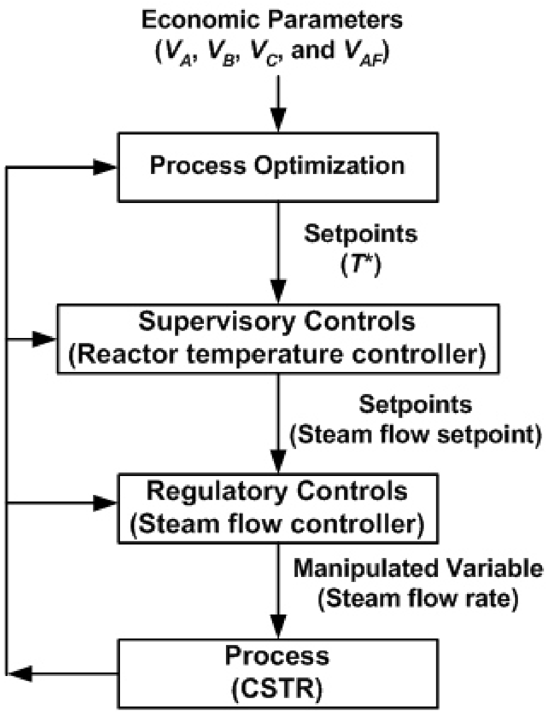


Figure 1.9.6 Logic flow diagram showing the interconnections between process control and optimization in relation to Example 1.1.

optimization/supervisory control/regulatory control is shown schematically in Figure 1.9.6 for this example. The connection from the process to the regulatory controls, supervisory controls and the process optimizer represents the flow of process measurements to each of these functions.

Self-Assessment Questions

Q1.9.1 What is the economic objective function of a process?

Q1.9.2 Why is a process model usually necessary to apply process optimization? When is a model not necessary?

Self-Assessment Answers

Q1.9.1 The economic objective function of a process is usually the rate of profit generation based on the model of the process. In general, the rate of profit generation is equal to the flow rate of the products multiplied by their value minus the feed rates times their costs and the rate of the utility usage times its cost.

Q1.9.2 Optimization usually involves tradeoffs between competing factors; therefore, an accurate model is necessary to locate the optimum tradeoff. At other times, an analysis of the process and the process economics leads to a straightforward determination of the optimum operating conditions; therefore, a complex detailed analysis using a process model is not required in the this case.

Self-Assessment Problems

P1.9.1 Explain how process control is different from process optimization. Give an industrial example of each to substantiate your analysis.

P1.9.2 Identify an example for which you use optimization in your everyday life. List the degrees of freedom (the things that you are free to choose) and clearly define the process and describe the objective function.

Self-Assessment Answers

P1.9.1 For process control, the MVs are adjusted in order to keep the CVs at or near their setpoints. For process optimization, the optimizer chooses the setpoints for the control loops so that the best economic performance of the process results. For the economic optimization of a distillation column, it is usually a compromise between the separation provided by the column and the use of utilities, such as process steam. The optimizer determines the optimum tradeoff between utility usage and product purities and expresses the result as setpoints for the product compositions. These setpoints are then applied by the control loops on the product compositions.

P1.9.2 An everyday example of optimization is minimizing your living expenses. The objective function is the total cost of food, housing, transportation expenses, clothing, insurance, etc. The degrees-of-freedom for this optimization problem are what food you purchase, your selection for housing, the type of transportation you choose (e.g., a car with a 4 cylinder engine or using public transportation), etc. Some individuals may choose to maximize the money that they have available for entertainment, which reduces to minimizing their living expenses.

1.10 Linear versus Nonlinear Process Behavior

Throughout the remainder of this book the difference between linear and nonlinear process behavior will be emphasized in a variety of ways. For example, almost all controllers function in a linear fashion while industrial processes are almost always nonlinear at least to some degree (i.e., a linear system means that a positive or negative input change of the same size will produce the same size change in the output although the sign of the change will be different). The process models developed in Chapter 3 are nonlinear while the analysis tools (Chapters 3-7) that we will apply to these process models are primarily linear. Linear analysis tools can be effectively applied to nonlinear models because, in general, the nonlinearity of these process models is relatively mild. Thus using a linear analysis (e.g., transfer functions, Chapter 5) can provide important understanding and insights. On the other hand, it is important to remember that all real processes are nonlinear to some degree and that linear analysis and linear controllers have practical limits.

1.11 Summary

- The CPI generally uses large-scale processing systems to produce products based on chemical reactions and/or separation while the bio-tech industries are usually small production rate processes that employ living organisms to produce products.
- Chemical and bio-process control are concerned with operating their processing units so that the operational objectives of these units are met in a safe and reliable manner.
- The primary economic benefits of improved control are associated with (1) producing lower variability products, (2) increasing production rates, (3) reducing energy usage (CPI only), (4) improving process safety, and (5) greater compliance to environmental and other regulations and constraints.
- A single control loop combines a controller, an actuator and a sensor with the process to maintain the CV at its assigned setpoint. Disturbances enter the process and require the control system to take action to absorb their effect to keep the CV near its setpoint. Feedback control results when the sensor reading for the CV is compared with the setpoint for the CV and the controller operates on this difference.
- Control diagrams use simplified PFDs with the major control elements while P&IDs show all the process, control and instrumentation hardware along with the signals passed between these devices.
- Process control is concerned with maintaining the CVs at their respective setpoints by adjusting the MVs. Process optimization is concerned with selecting the setpoint of certain control loops so that the process operates at a condition that maximizes the rate of profit generation for the process. Process optimization usually involves operating at the most advantageous set of process constraints and making the optimal tradeoff between competing factors.

1.12 References

1. Austin, G.T., *Shreve's Chemical Process Industries*, 5th ed., McGraw-Hill, New York (1984).
2. Atkinson, B., Mavituna, F, *Biochemical Engineering and Biotechnology Handbook*, 2nd Ed., Stockton Press, New York (1992).
3. http://www.fda.gov/cder/OPS/PAT.htm

1.13 Additional Terminology

Aerobic fermentation process - growth of microbes in an environment containing oxygen.

Actuator - the system that changes the level of the MV. The actuator system usually involves a control valve and associated equipment or a variable speed pump.

Analyzer deadtime - the time from process stream sampling to the availability of the analyzer reading.

Analyzer delay - the time from process stream sampling to the availability of the analyzer reading.

Batch process - a process that has neither streams entering nor product streams leaving the system.

Biotechnology industries - companies (e.g., pharmaceutical companies) that produce products using living organisms.

Broth - the slurry in a fermentor containing microorganism, food for the microorganism (e.g., glucose) and nutrients.

Cell - a fundamental structural unit of a living organism, consisting of a small mass of cytoplasm and usually a nucleus surrounded by a membrane or a rigid cell wall.

CPC - chemical process control.

CPI - chemical processing industries (e.g., chemical plants and refineries).

Cascade control - manipulation of a regulatory controller setpoint by a supervisory controller.

cGMP - current good manufacturing practices, which are set by the FDA for the bio-pharmaceutical products.

Closed-loop control - use of the measured value of the CV to select the MV level.

Constraint - a limit on process operation.

Continuous process - a process for which material continuously enters and leaves.

Control diagram - a PFD of a process along with a simplified diagram of the control loops.

Controlled variable (CV) - the process variable that the control loop is attempting to maintain at its setpoint.

Controller - a unit which adjusts the MV level to keep the CV at or near its setpoint.

Coupling - the interaction between control loops for a MIMO process.

CST - continuously stirred tank.

CSTR - a continuously stirred-tank reactor.

CV - the controlled variable of a control loop.

Deadtime - the time difference between a MV change and significant process change.

Decision variables - the operating conditions of a process that are selected by an optimization algorithm for the optimum operation of the process.

Disturbance - an input to the process that is not a MV.

DO - dissolved oxygen.

DV - a disturbance variable.

Error from setpoint - the difference between the measured value of the CV and its setpoint.

Fed-batch process - see semi-batch process.

Feedback control - use of the sensor reading and the setpoint value to select the level for the MV for a process.

Feedforward control - a controller that makes adjustments to the MV level based upon measured disturbances in an effort to absorb the effect of the disturbance before it affects the process.

Fermentation Process - cultivation of microbial population under anaerobic (absence of oxygen) conditions producing ethanol (strictly speaking). However, this term has been used for any cell cultivation process.

Final control element - the system that changes the level of the MV. The final control element usually involves a control valve and associated equipment or a variable speed pump.

High-gain process - a process for which a relatively small input change causes a relatively large change in the output variable.

Inoculation - the step of combining the inoculum and the media to initiate the fermentation process.

Inoculum - the collection of microorganisms necessary to start the microbial growth in a fermentor.

Interaction - the effect of control loops on each other in a MIMO system.

Lag - the property of a process that keeps it from responding instantaneously to input changes.

Low-gain process - a process for which a relatively large input change causes a relatively small change in the output variable.

Manipulated variable (MV) - the process variable, usually a flow rate, which is adjusted to keep the CV at its setpoint.

Microbial species - microscopic sized organisms such as bacteria or yeast or virus.

Microorganism - microscopic sized organisms such as bacteria or yeast or virus.

MIMO - multiple-input/multiple-output process, i.e., a process with two or more inputs and two or more outputs.
MV- the manipulated variable of a control loop.
Nonlinear process - a process for which the process gain is not constant.
Nutrients - for a bio-reactor are substances other than glucose that the cells required to grow (e.g., salts and yeast extract).
Offset - a persistent error between the measured value of the CV and its setpoint.
Open-loop response - the measured value of the CV as a function of time after a change in the MV value without feedback control.
Open-loop settling time - the time for the CV to attain 95% of its ultimate change after a step change in the MV.
Open-loop stable process - a process that reaches a new steady state after an input change.
Open-loop unstable process - a process that does not reach a new steady state after an input change.
Operator acceptance - having the operators routinely use a controller, i.e., operator trust in reliable and effective controller operation.
Optimization decision variables - the operating conditions of a process that are selected by an optimization algorithm for the optimum operation of the process.
Override control - the arrangement of control loops to prevent the violation of safety, environmental, or operational constraints.
Pairing of MVs and CVs - for MIMO processes, selecting which MV is to be used to control which CV.
PFD - a simplified diagram of a process showing the major pieces of equipment and the process streams.
Process - the system whose outputs are affected by the inputs.
Process constraint - a limit on process operation.
Process flow diagram - a simplified diagram of a process showing the major pieces of equipment and the process streams.
Process optimization - selecting the setpoints for key controllers such that the process produces the highest rate of profit generation.
Proportional-only controller - a feedback controller that makes changes to the MV value which are proportional to the error from setpoint.
P&ID - piping and instrumentation diagram.
Recombinant drug - a drug, typically a protein, which is produced by using recombinant DNA technology, in which isolated genes from one organism are recombined with other DNA that can be propagated in a similar organism.
Regulatory controller - the lowest level of controls, usually flow controllers, pressure controllers and level controllers.
Remote setpoint - a setpoint for a regulatory control loop that is determined by a supervisory control loop.
Residence time - the average time that an element of feed spends in the process.
Saturated MV - an MV that is at its maximum or minimum level.
Semi-batch process - a process with one or more feed streams, but no product streams.
Sensor - the device that measures a process variable.
Setpoint - the desired or specified value for the CV.
SISO - a single-input/single-output process.
Sterilization - the process of removing microorganisms from a process.
Supervisory controller - the controllers that are responsible for meeting the setpoints applied by the optimizer or the operator, usually temperature or composition control loops.
Time constant - a measure of how fast a process changes for a change in an input.
Valve deadband - the maximum positive and negative change in the signal to the final control element that does not produce a measurable change in the flow rate in question.
Variability - the magnitude of the deviations from the setpoint for the CV value.

1.14 Preliminary Questions

1.1 Chemical and Bio-Process Control

Q1.1.1 What is the overall purpose of process control?

Q1.1.2 Years ago large holding tanks were placed between units to make it easier to control these units. Why is this not a viable approach today?

Q1.1.3 What is improved process control performance? How is it measured?

Q1.1.4 How can improved control performance be used to improve the economic performance of a company in the CPI?

Q1.1.5 Why is process control financially attractive to companies in the CPI?

Q1.1.6 Why is process control financially attractive to companies in the bio-tech industries?

1.2 Everyday Examples of Process Control

Q1.2.1 How are a disturbance and a change in a MV alike? How are they different?

Q1.2.2 What is the difference between feedback control and feedforward control?

Q1.2.3 How can you convert a closed-loop process into an open-loop process?

1.3 Control Diagrams and P&IDs

Q1.3.1 In P&IDs, how are pneumatic lines and electrical connections distinguished from each other?

Q1.3.2 What is the difference between PID and P&ID?

Q1.3.3 What is a summation block and how is it used in a feedback control loop?

Q1.3.4 What does a symbol on a control diagram that ends with a “C” (e.g., TC, FC and DPC) represent?

Q1.3.5 What does a symbol on a control diagram that ends with a “T” (e.g., TT, FT and PT) represent?

Q1.3.6 What does the symbol “FY” represent on a P&ID?

Q1.3.7 How are P&IDs different from the control schematics used in this text? How are they alike?

1.5 Block Diagram of a General Feedback Control System

Q1.5.1 From a logic flow point of view, why must a control loop be arranged in the following order: controller, actuator, process and sensor?

Q1.5.2 What is the most common actuator used in the CPI?

Q1.5.3 What is the most common actuator used in the bio-tech industries?

Q1.5.4 What is the difference between a bio-reactor and a reactor in the CPI?

1.6 Types of Controllers

Q1.6.1 Summarize the differences between manual, PID, advanced PID and model-based control. How are they similar?

1.7 Responsibilities of a Chemical Process Control Engineer

Q1.7.1 For applying process controls, why do process control engineers need to understand the processes?

Q1.7.2 What are the primary responsibilities of a process control engineer? Give an example of each one that you identify.

Q1.7.3 What is the difference between process design effects on process control and designing process control systems?

1.15 Analytical Questions and Exercises

1.2 Everyday Examples of Process Control

P1.2.1* Identify a control system with which you interact daily. Identify the controller, actuator, process and sensor and draw a block diagram for the control loop similar to Figure 1.5.1. Choose an example that is different from the ones presented in the text.

P1.2.2* Give an everyday example for each of the following terms: a. disturbance, b. deadtime, c. process constraint, d. coupling and e. lag. Use examples different from those presented in the text.

1.3 Control Diagrams and P&IDs

P1.3.1*** pH in a bio-reactor is a very important parameter. It relates to active cellular metabolism. Typically, in a bacterial or yeast fermentation process, the cells produce acids and it is necessary to neutralize them to control the pH at an optimum value. If the pH is not controlled properly, the pH in a reactor drops and this can inhibit the growth of the cells and may even kill the cells. pH control is accomplished by adding a sterile sodium hydroxide solution if the measured value of the pH is less than the setpoint. Draw a P&ID to accomplish pH control in a bio-reactor using Figure P1.3.1 as the basis. Also, add a board-mounted level indicator so that the operator can tell when the reactor is full.

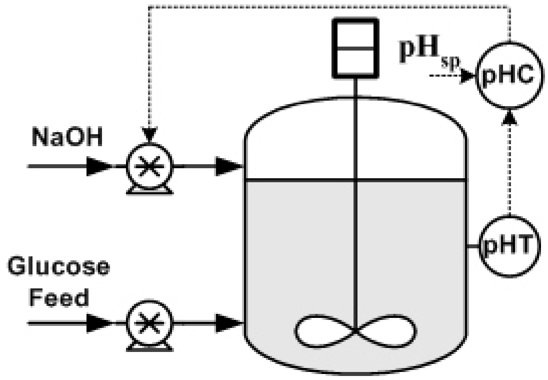


Figure P1.3.1 Schematic of a pH controller applied to a fed-batch reactor.

1.4 Industrial Process Control Examples

P1.4.1* Choose an industrial process control system to which you have been exposed. Identify the controller, actuator, process and sensor and draw a control diagram to illustrate the control loop.

P1.4.2* Identify a SISO process and a MIMO process used in industry.

P1.4.3** Identify what constitutes the process for the control loop shown in Figure P1.4.3.

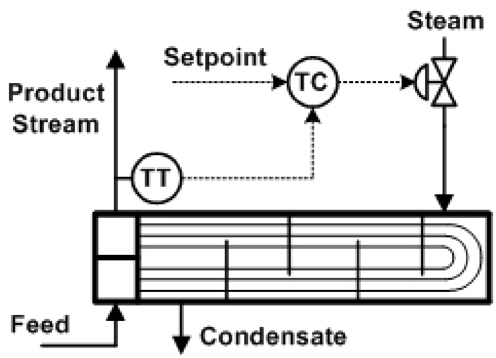


Figure P1.4.3 Control diagram for a temperature control loop applied to a heat exchanger.

P1.4.4** Furnaces are used to heat process streams to high temperatures before further processing, e.g., furnaces are used to provide heat input for crude units in refineries and are used to preheat feed streams before entering a reactor. Consider the temperature control loop for a furnace shown in Figure P1.4.4. Draw a logic flow diagram similar to Figure 1.5.1 for the system shown in Figure P1.4.4, i.e., assemble each element of the temperature loop for the furnace to form a feedback diagram. Also, indicate what are specific disturbances for this control loop.

P1.4.5** Utility boilers are used in the CPI to provide steam for a variety of uses, e.g, as reboiler duty for distillation columns, heat duty for heat exchangers, stripping steam for crude columns, steam tracing lines, vacuum injectors, etc. Utility boilers are insulated tanks that have boiler tubes placed in the bottom of them and hot combustion gases are passed through the boiler tubes, which transfer heat to the water in the boiler producing steam. The steam generated in the boiler is distributed throughout a plant through the steam header. It is important to maintain a proper water level in a utility boiler because if the tank overflows, water will enter the steam header causing damage to the plant and if the level becomes too low, the boiler tubes will be exposed, causing damage to the tubes. Even though the demand for steam can vary widely, it is crucial that the level control for the utility boiler functions properly. Consider the utility boiler shown in Figure P1.4.5. Add

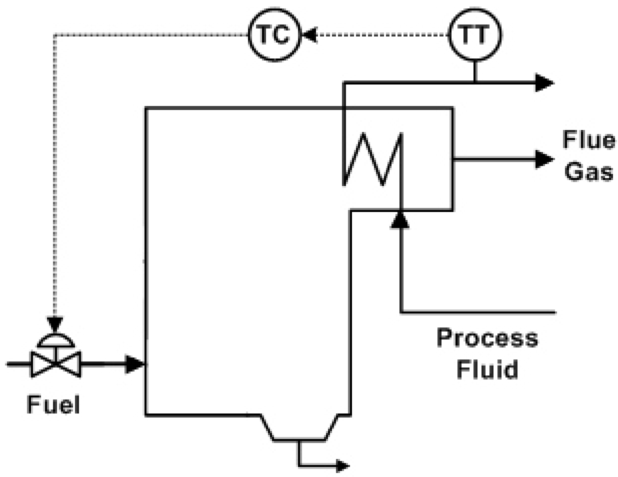


Figure P1.4.4 Schematic of a furnace with a temperature control loop.

the required sensor, actuator and controller to this diagram and connect them for a feedback level controller for a utility boiler that will maintain the water level in the boiler between acceptable limits in the face of varying demand for steam production.

P1.4.6** Separation using membranes is a common practice in biotechnology industries. Typically, different types of membranes are used for different separation needs. A cross-flow filtration device, which separates the cells from the broth without plugging the filter, is used to separate cells from the broth of a bio-reactor. Typically, the transport through the membrane filtration system is maximized, which maximizes the cell concentration in the product stream. Normally, the flow rate through the membrane separator is manipulated depending upon the on-line measurements of the cell concentration in the cell solution. On-line measurements, in the simplest case, can be the conductivity of the process stream (conductivity sensor, CT in Figure P1.4.6) containing the cells. Add the controls to the process flow diagram (Figure P1.4.6) for this process.

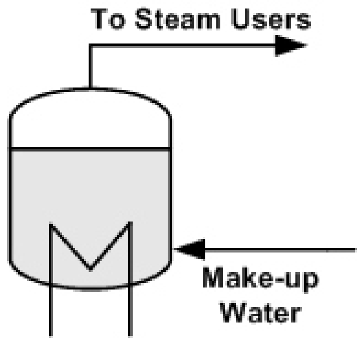


Figure P1.4.5 PFD for a utility boiler.

P1.4.7** Consider level control and temperature control for a swimming pool. If the flow rate of the makeup water is used to control the level in the pool and the fuel flow rate to the pool heater is used to control the temperature of the water, draw the control loops on a figure of a swimming pool. Also, identify the major disturbances for both loops. Assume that a pump circulates the pool water through the pool filter and the pool heater. Also, assume that the makeup water is added directly to the pool.

P1.4.8*** Fed-batch systems, which are used extensively in the biotechnology field, add feed to a reactor until the reactor achieves a maximum predetermined level (e.g., 90% of the tank volume) but the product is not withdrawn until the reaction period is complete**.** During the operation of the fed-batch process, dissolved oxygen (DO) and pH are typically maintained by feedback control loops to ensure conditions that are conducive for cell growth. Consider the batch bio-reactor shown in Figure P1.4.8, which can be converted into a fed-batch system by adding feed (a glucose and nutrient mixture) until the bio-reactor achieves the maximum predetermined level. A feed tank, a variable speed feed pump, a glucose sensor and a level sensor are

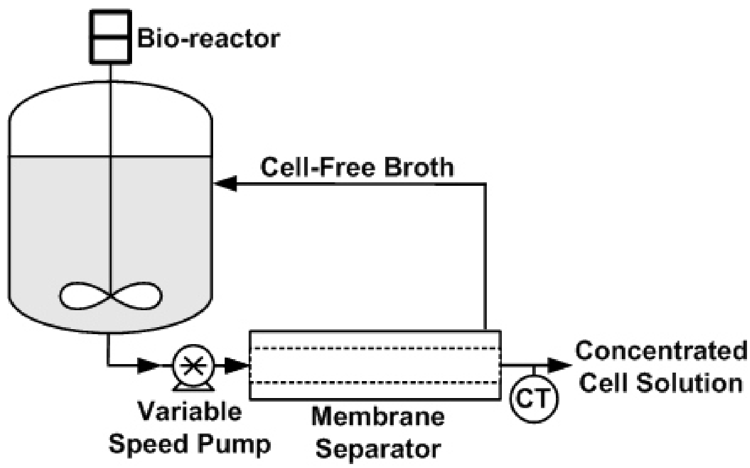


Figure P1.4.6 A PFD for a cell concentrator combined with a bio-reactor.

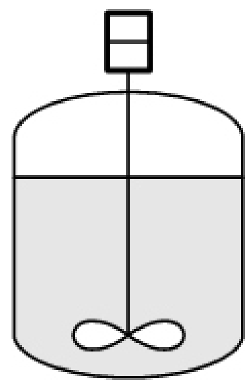


Figure P1.4.8
Batch bio-reactor.

required to deliver the feed to the fed-batch reactor. In addition, pH control is attained by measuring the pH and adding NaOH to the bio-reactor from a NaOH storage tank and DO control is accomplished by adjusting the air feed rate to the reactor. Add the feed, NaOH and air delivery systems to Figure P1.4.8. In addition, add the control systems to Figure P1.4.8 that will maintain the glucose concentration, pH and the DO at the desired operating points until the level in the reactor level reaches its maximum value.

P1.4.9*** Distillation columns are the primary means of separation used in the CPI. The separation derived from distillation is based on separating components based on relative values of their volatilities. Consider a distillation column that separates propane from butane shown in Figure P1.4.9. Note that for simplicity the level senor/transmitter were omitted from this figure. The overhead product (D) from the column, which is largely propane, is used as fuel while the bottoms product (B), which is largely butane, is a much more valuable product because it is a component blended directly into the gasoline pool. The greater the separation produced by the distillation column (i.e., the lower the mole fraction of butane in the overhead propane product and the lower the mole fraction of propane in the butane product), the larger the fraction of the butane recovered in the process, but at the expense of requiring a larger steam usage. Consider how optimization, supervisory control and regulatory control are applied to this column.

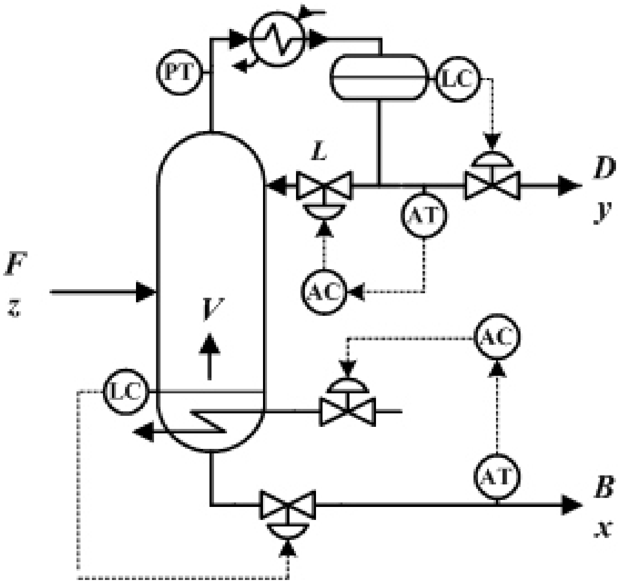


Figure P1.4.9 Schematic of a distillation column with the [L,V] configuration.

a. Write an equation for the economic objective function for this case. Remember that the butane that leaves with the propane in the overhead product is valued as fuel (V_F) and the propane in the butane in the bottoms product does not yield gasoline in the reformer reactors but is valued as fuel grade propane. The butane that is recovered in the bottoms product has a value V_B while the propane in the bottoms product is valued as fuel, V_F. The cost per pound of steam is given as V_S.

b. Draw a control and optimization schematic similar to Figure P1.4.9 showing the connection between optimization, supervisory control and regulatory control for this column. Assume that the reflux flow rate is used to control the butane in the overhead product and the steam flow to the reboiler is used to control the propane in the bottoms product. In addition, assume that the flow rate of the reflux and the steam to the reboiler are controlled by flow control loops.

P1.4.10** Consider the distillation column shown in Figure P1.4.9. Note that for simplicity the level sensor/transmitter were omitted from this figure. Because the overhead product composition is controlled by the reflux flow rate, L and the bottoms product composition is controlled by the boilup rate, V, this configuration is referred to as the [L,V] configuration. Draw the control diagram for this column if the overhead composition is controlled by adjusting the distillate product flow rate and the accumulator level is controlled by the reflux flow rate. Also, the bottom product composition is controlled by the flow rate of the bottoms product and the reboiler level is controlled by the boilup rate. This latter configuration is referred to as the [D,B] configuration.

P1.4.11** Dialysis is used to treat patients with malfunctioning kidneys. Normally functioning kidneys continuously remove waste products from the blood and send these wastes to the bladder for discharge. An individual with malfunctioning kidneys will experience buildup of these toxic wastes, leading to severe toxicity and eventually to death. Dialysis, which is a way for individuals with malfunctioning kidneys to remove the toxic wastes from their blood, is often used for patients that are awaiting a kidney transplant. A stream of blood from the patient is passed through the dialysis unit where the waste materials are removed from the blood and replaced with dialysis medium before the blood is returned to the patient (Figure P1.4.11). It is important to provide the correct flow rate of dialysis medium so that the patient's blood can be properly purified. As a result, most dialysis systems contain an onboard microprocessor that contains a feedback controller. The control system uses the measurement of the waste materials in the patients blood to set the flow rate of dialysis medium to the dialysis unit. Identify the MV, CV, DVs, process and the actuator for the dialysis process shown in Figure P1.4.11. Draw an information flow diagram similar to Figure 1.5.1 which contains this information for this dialysis control loop.

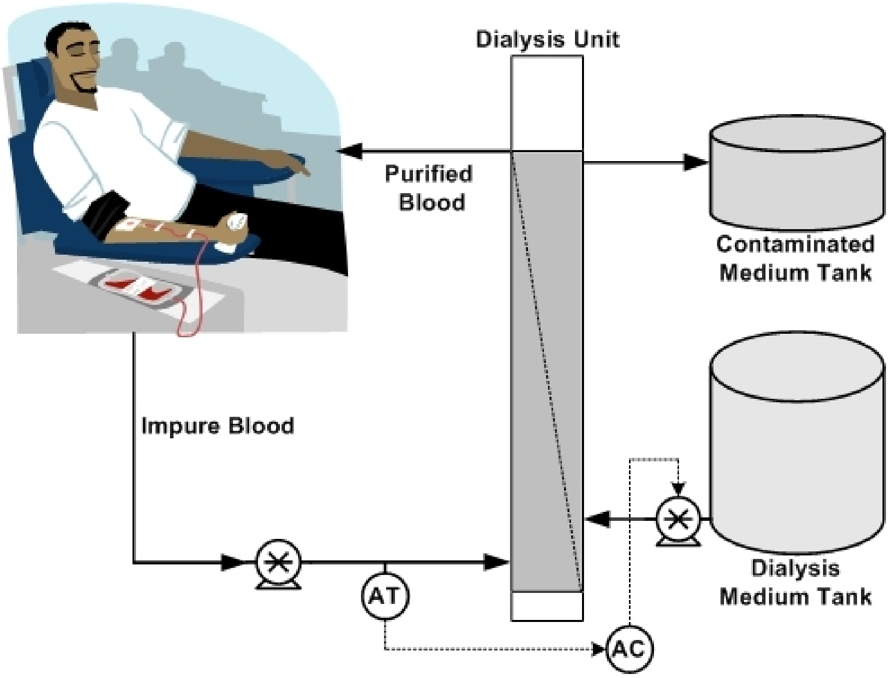


Figure P1.4.11 Schematic of a dialysis process with control system.

1.9 Process Optimization and Process Control

P1.9.1** Determine the optimum reactor temperature for the CSTR in Example 1.1 if the value of C (V_C) is decreased to $-0.20/gmol. What can you conclude? What does a negative value of a byproduct signify?

P1.9.2*** Evaluate the effect of F_V on the optimum reactor temperature in Example 1.1 by determining T^* for F_V in l/s equal to [2, 5, 10, 20, 40]. What can you conclude?

(Degree of difficulty for problems: * least difficult , ** more difficult, * most difficult.**

Chapter 2

Control Loop Hardware

Chapter Objectives

- Show the hardware and signals that are required to implement a typical feedback control loop.
- Overview the evolution of control computers from pneumatic controllers to the present.
- Demonstrate when linear and equal percentage globe valves should be used.
- Present a methodology for selection and sizing of control valves/actuators.
- Present the control relevant aspects of typical process sensors used in the CPI and bio-industries.
- List the expected performance characteristics for typical actuator and sensor systems.

2.1 Introduction

Control loop hardware and software provide the means to control a process at desired conditions. The objectives are several fold, including reducing process variability (e.g., automating functions previously performed manually), and improving the ability of a process to operate within safety, environmental, product quality, and other constraints. Control loop hardware is comprised of mechanical and electrical devices that perform the functions of the actuator, sensor and controller. For example, to implement the control loops shown in Section 1.4, hardware is required. In addition, operators use certain sensors to monitor process operation and use certain actuators to adjust the process (i.e., manual control). Therefore, not all process instrumentation is part of an automatic feedback control loop.

Choosing the proper hardware for a control application and ensuring that it operates effectively are primary responsibilities of control engineers. In addition, when control loops are not functioning properly, and the issue is beyond what direct support process engineers can handle, control engineers (e.g., from the supporting plant automation group) must identify the source of the problem and correct it, which is known as troubleshooting (Chapter 10). To accomplish these tasks, control engineers must understand the control-relevant issues associated with each of the components that make up a control loop. This chapter describes the hardware components that comprise a typical feedback control loop used in the CPI and bio-tech industries by providing an overview of the design approaches and performance measures for these components. Because a complete description of these devices is beyond the scope of this text, the descriptions here focus on their control-relevant aspects.

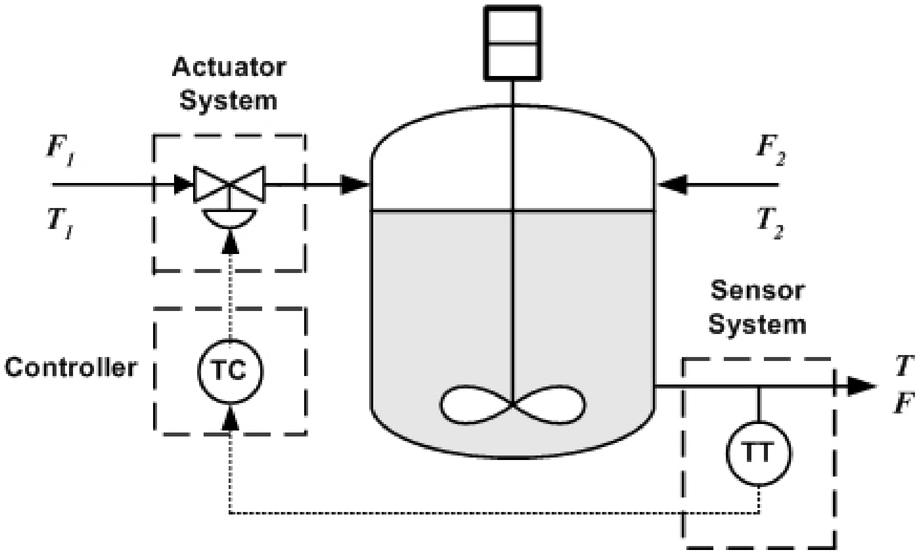


Figure 2.1.1 A control diagram for a temperature control loop applied to a CST thermal mixer.

Figure 2.1.1 is a control diagram for a temperature controller on the CST thermal mixer (Figure 1.4.2). This feedback control loop consists of a controller, a final control element, a process and a sensor. Figure 2.1.2 is a schematic of the hardware that comprises this feedback temperature control loop as well as the signals that are passed between the various hardware components. The sensor system in Figure 2.1.1 corresponds to the thermowell, thermocouple and **transmitter** in Figure 2.1.2 while the actuator system in Figure 2.1.1 corresponds to the control valve, I/P (current-to-pressure) signal converter and instrument air system in Figure 2.1.2. Likewise, the controller in Figure 2.1.1 represents an A/D (analog-to-digital) and a D/A (digital-to-analog) signal converter, the control computer and the operator console in Figure 2.1.2. This arrangement is specific to the **Basic Process Control System** (**BPCS**), which is discussed in detail in the next section. It will be demonstrated in the next section that different types of controllers require different types of hardware to interface with the sensors and actuators.

A thermocouple is used to measure the temperature inside the mixing tank and is placed in thermal contact with the process fluid leaving the mixing tank by means of a thermowell in the product line. Because the process is assumed to be well mixed, the temperature of the product stream should be the same temperature as inside the mixing tank. The temperature transmitter converts the millivolt signal generated by the thermocouple into a 4-20 mA (milliamperes) analog electrical signal that is proportional to the temperature inside the thermowell. When the thermocouple/transmitter system is calibrated properly and when the thermowell is correctly designed and located, the value of the analog signal should correspond closely to the temperature in the mixing tank. The thermocouple/thermowell/temperature transmitter comprises the sensor system shown in Figure 2.1.1.

The 4-20 mA analog signal from the temperature transmitter is converted into a digital reading by the **analog-to-digital (A/D) converter**. (The 4-20 mA analog signal was chosen due to safety considerations, i.e., it is low enough that the energy of a spark caused by a wire failure would unlikely ignite a flammable source yet high enough for accuracy.) The output of the A/D converter is a digital measurement of the temperature, which is used in the control calculations. The operator console shown in Figure 2.1.2 allows the operator or control

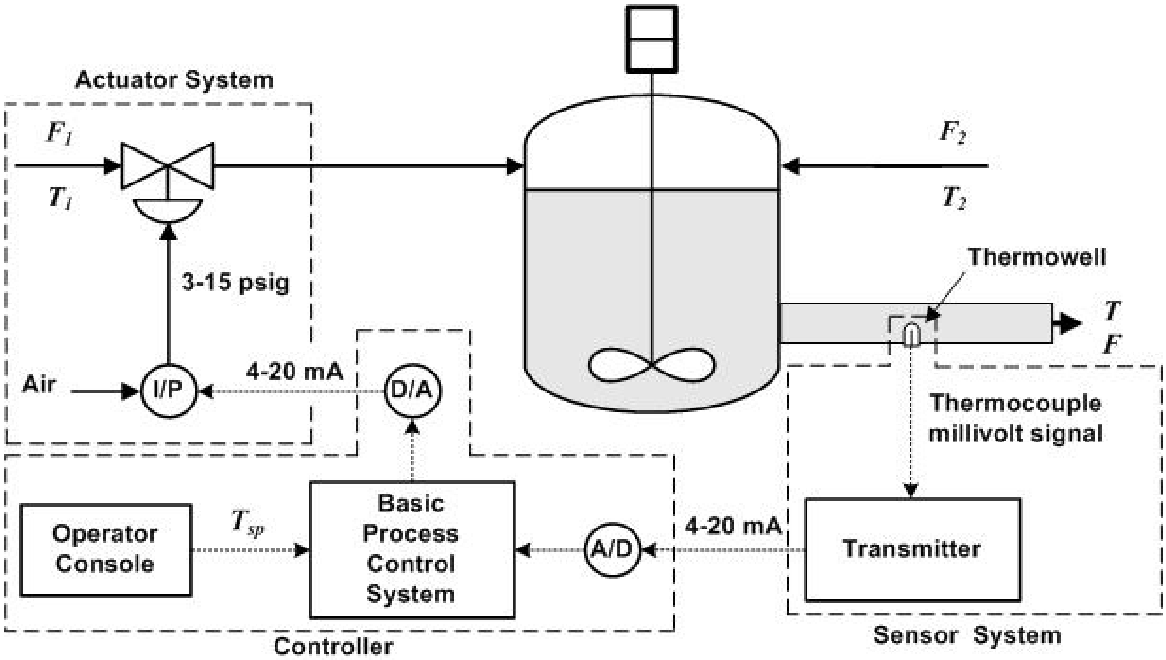


Figure 2.1.2 Schematic of a temperature control loop on a CST thermal mixer showing all the hardware components of the control loop and the signals passed between components.

engineer to observe the performance of the control loop and to change the setpoint, T_{sp}, and controller tuning parameters for this loop. The value of T_{sp} and the digital value of the measured mixer temperature are used by the control algorithm in the DCS (i.e., the control computer) to calculate the control action. The output from the controller is a digital signal that is converted into a 4-20 mA analog signal by the **digital-to-analog (D/A) converter**. The DCS, D/A and A/D converters and the operator consoles, which represent the control system shown in Figure 2.1.1, are typically located in a centralized control room while the remaining equipment associated with the actuator and sensor is normally located in the field near the process equipment.

The 4-20 mA analog signal from the D/A converter goes to the **current-to-air pressure (I/P) converter**. The I/P converter uses a source of instrument air to change the air pressure (3-15 psig) applied to the control valve corresponding to the value of the analog signal. That is, if I is the value of the analog signal and P is the instrument air pressure delivered to the control valve, by proportionality

$$\frac{I \quad 4}{20 \quad 4} \qquad \frac{P \quad 3}{15 \quad 3} \qquad \textbf{2.1.1}$$

because a 4 to 20 mA range in the analog signal corresponds to a 3 to 15 psig range in the instrument air pressure. Changes of instrument air pressure to the control valve cause changes in the stem position of the control valve, which result in changes in the flow rate to the process (i.e., when I=4 mA and P=3 psig, the control valve is closed; when I=20 mA and P=15 psig, the control valve is wide open for an air-to-open control valve in both cases). These changes in the flow rate to the process cause changes in the temperature of the liquid in the mixer, which are measured by the sensor, completing the feedback control loop. The final control element shown as the actuator system in Figure 2.1.1 consists of the I/P converter, the instrument air system and the control valve.

The ability to effectively troubleshoot a control loop requires the knowledge of the components that actually implement the control loop as well as the signals that are passed between these elements. This chapter considers the design and control-relevant aspects of the DCS, the actuator system and a number of commonly encountered sensors.

Example 2.1 Conversion of Signals within a Feedback Loop

Problem Statement. Determine the value of the 4-20 mA signal to the I/P converter and the pneumatic signal to the control valve in Figure 2.1.2 if the controller output is 75% of full range.

Solution. A controller output signal equal to 75% of full range results in a 4-20 mA signal of

$$I \quad 4 \quad 0.75 \;\; 20 \quad 4 \quad 16 \text{ mA}$$

because the minimum controller output corresponds to a 4 mA signal and a maximum controller output corresponds to a 20 mA signal. To calculate the magnitude of the pneumatic signal, rearrange Equation 2.1.1, i.e.,

$$P \quad 3 \quad \frac{12}{16} \; 16 \quad 4 \quad 12 \text{ psig}$$

or use the fact that the pneumatic signal is also 75% of full range, i.e.,

$$P \quad 3 \quad 0.75 \;\; 15 \quad 3 \quad 12 \text{ psig}$$

Self-Assessment Questions

Q2.1.1 For a typical feedback loop, where are 4-20 mA signals used?

Q2.1.2 What hardware comprises a controller when a control loop is implemented using a DCS?

Q2.1.3 For Figure 2.1.2, what hardware is located in the field and what hardware is located in the control room?

Self-Assessment Answers

Q2.1.1 The 4-20 mA signals are used to transmit sensor readings from the sensor to the A/D converter in the DCS and to transmit the controller output from the D/A converter to the I/P converter. For an electronic analog controller, 4-20 mA signals are used to transmit the sensor readings to the controller and to transmit the control signal to the I/P converter.

Q2.1.2 The D/A and A/D converters, the DCS control computer and the operator console comprise the controller when a DCS is used.

Q2.1.3 The control system hardware (the D/A and A/D converters, the DCS control computer and the operator console) are located in the control room. The actuator system (control valve, I/P converter and instrument air system) and the sensor system (sensor, transmitter and sensor/process interface) are located in the field next to the process.

Self-Assessment Problems

P2.1.1 Consider an I/P converter. If the analog signal to the I/P converter is 12 mA, what is the magnitude of the pneumatic signal leaving the I/P converter?

Self-Assessment Answers

P2.1.1 Using Equation 2.1.1, $\frac{12 \quad 4}{20 \quad 4} \quad \frac{P \quad 3}{15 \quad 3}$; $\quad P \quad 3 \quad 0.5 \quad 12 \quad 9$ psig

2.2 Control Systems

Background. Pneumatic PID controllers were introduced in the 1920s and were in widespread use by the mid 1930s. Pneumatic controllers use bellows, baffles and nozzles with a pressurized supply of air to apply PID control action. That is, the pneumatic controller receives pneumatic signals corresponding to the measured value of the CV and the setpoint and forms the error from setpoint. The pneumatic controller acts on the error from setpoint signal with a bellows, baffle and nozzle in conjunction with the instrument air system to produce a pneumatic signal that is sent to the control valve. For the early versions of pneumatic controllers, the controllers were installed in the field near the sensors and control valves. In the late 1930s, transmitter-type pneumatic controllers began to replace the field-mounted pneumatic controllers because of the increase in size and complexity of the processes being controlled. For the transmitter-type pneumatic controllers, the sensor readings were converted into pneumatic signals (i.e., 3-15 psig) that were conveyed by metal tubing into the control room where the pneumatic controller determined the control action. In turn, the control action was pneumatically transmitted to the actuator on the process. Because the transmitter-type pneumatic controllers were typically located in a central control room, operators could conveniently address the overall control of the process using a number of controllers from a centralized location. Figure 2.2.1 shows the hardware and signals associated with a pneumatic controller applied to a temperature control loop. The recording of manipulated and controlled variables associated with such controllers was usually done via local strip chart or circular chart recorders.

In the late 1950s, electronic analog controllers (Figure 2.2.2) became commercially available. These devices use capacitors, resistors and transistor-based amplifiers to implement control action. Because electronic signals are transmitted to and from the control room using electrical wires, the need for long runs of metal tubing used by

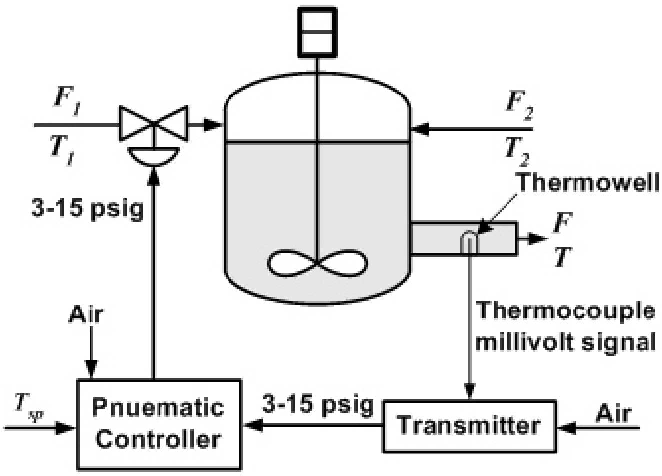


Figure 2.2.1 Hardware and signals associated with a pneumatic temperature controller.

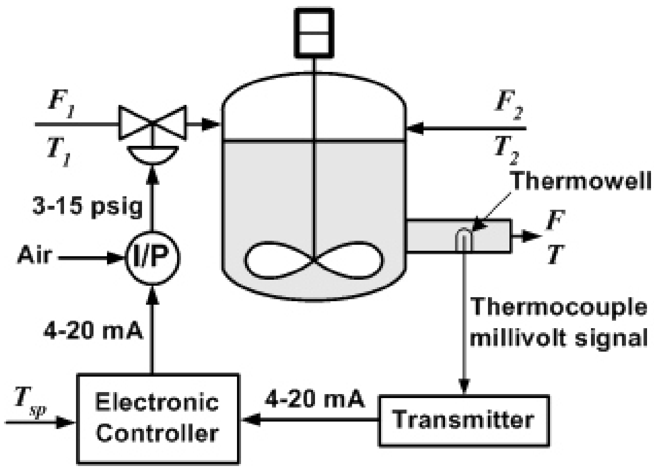


Figure 2.2.2 Hardware and signals associated with an electronic temperature controller.

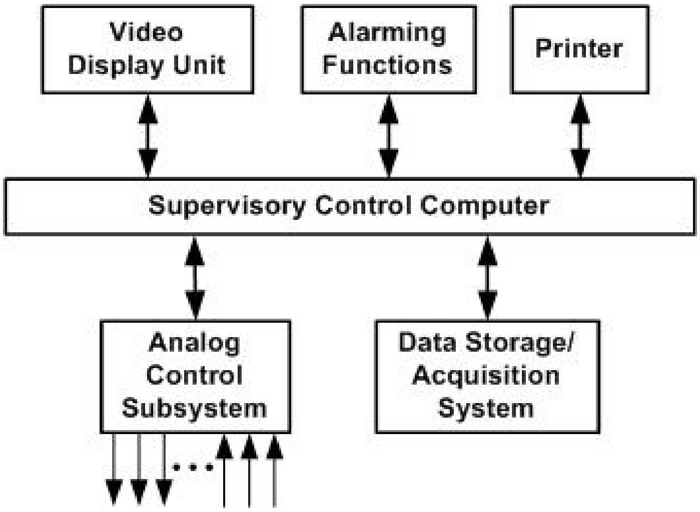


Figure 2.2.3 Logic flow diagram of a direct digital control (DDC) system.

pneumatic controller was eliminated, greatly reducing the installation costs and resulting in faster-responding controllers. By 1970, the sales of electronic controllers exceeded the sales of pneumatic controllers in the CPI[1].

The first supervisory computer control system, which is known as **direct digital control (DDC)** because it performs the control calculations digitally, was installed in a refinery in 1959. A simplified schematic of a DDC control system is shown in Figure 2.2.3. Note that this system offered data storage and acquisition as well as control loop alarms that previous control systems did not offer. In addition, the centralized computer could use the available operating data to determine the setpoints for certain key control loops in an effort to obtain the most efficient operation of the plant (i.e., process optimization).

The biggest disadvantage of the DDC approach was that if the control computer failed, the entire control system was shut down. A redundant control computer was an expensive alternative and not always reliable. Due to the technological breakthroughs in computers and associated systems, a new computer control architecture was developed and introduced by vendors in the late 1970s. This architecture is based on using a number of **local control units (LCUs)**, which have their own microprocessors and are interconnected by **shared communication lines** (i.e., a **data highway**). The LCU network is connected to a data acquisition system, operator/engineer consoles and a general purpose computer. This computer control architecture became known as a **distributed**

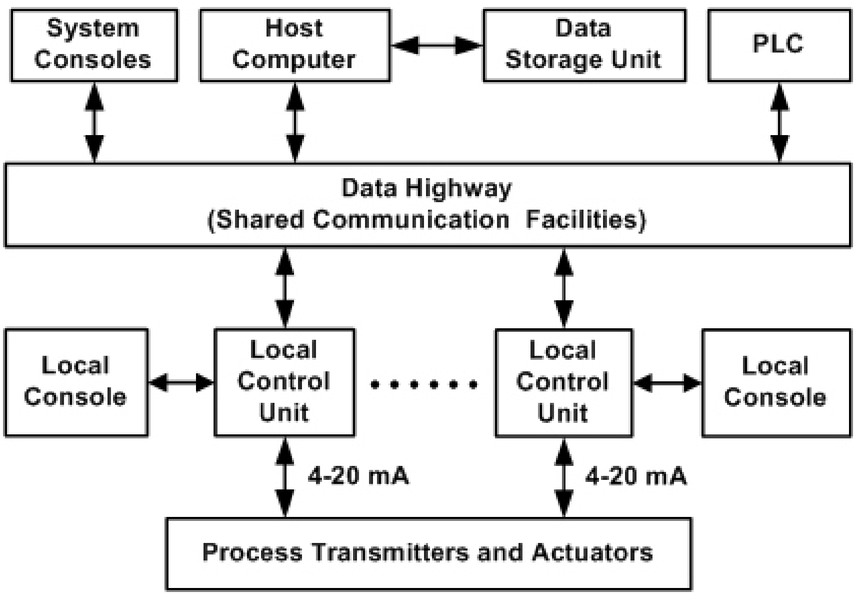


Figure 2.2.4 A generalized logic flow diagram of a DCS

control system (**DCS**; Figure 2.2.4) because it involves a network with various control functions distributed for a variety of users. The PLC indicated in Figure 2.2.4 is usually a programable logic controller, but could be any type of 3rd party control device.

The advantages of a DCS over a centralized control computer result from the use of microprocessors for the local control function. Even if a microprocessor were to fail, only the control loops serviced by that LCU would be affected. A redundant microprocessor that performs the same calculations as the primary microprocessor (i.e., a hot backup) greatly increases the reliability of the system. As a result, the probability that all the control loops fail at the same time, or even a major portion of the control loops fail, is greatly reduced in comparison to a centralized control computer. In addition, the DCS is much easier to expand. To increase the number of control loops serviced by the DCS, only a primary and a redundant LCU need to be added. The modular nature of DCSs can be a major economic advantage for plants that undergo expansion. From a controller performance standpoint, the application of a DCS and an electronic analog controller is generally equivalent. On the other hand, the implementation of a DCS to a large process is much easier and less expensive per loop than for electronic analog controllers. In addition, the DCS offers additional features, such as digital data storage and a host computer for advance control and optimization applications.

Structure of a DCS. A generalized schematic of a DCS is shown in Figure 2.2.4. A number of LCUs, which contain redundant microprocessors, perform the control functions for the process in a distributed fashion. Each LCU has several consoles attached to it. The consoles (**video display units**, **VDUs**), which utilize **cathode ray tubes** (**CRTs**), have video displays that show process schematics with current process measurements. Operators and control engineers use these displays to monitor the behavior of the process, set up control loops and enter setpoints and tuning parameters. A photograph of a typical control room for a DCS is shown in Figure 2.2.5. Normally, the consoles have touch screen capability so that if operators want to make a change to a control loop, they touch the icon on the CRT screen for the controller of interest. Then a screen pops up that allows the operator to make the desired changes. On some DCSs, control loops can be conveniently set up by clicking and dragging

Figure 2.2.5 Photograph of a control room for a DCS. Courtesy of Honeywell.

the tags for the desired sensor readings and the final control elements and connecting them to the type of controller that is selected. Because the LCU is attached to the data highway (i.e., the shared communication facilities), a local display console can view schematics and current operating data for other parts of the plant, but a local display console typically can make changes only to the control loops associated with its LCU. The local console can also be used to display historical trends of process measurements. To do this, the local console must access historical data in the data storage unit by using the data highway.

Data acquisition is accomplished by transferring the process measurements from the LCUs, through the data highway into the host computer where the process data are passed on to the data storage unit, which is usually a computer hard drive. The archived process data can be accessed from one of the system consoles or one of the local consoles. Previously, during the era of electronic analog controllers, data storage for important control loops was typically accomplished using a strip chart recorder, which printed measurements on a small roll of paper in different colors of ink to distinguish different sensor readings. PLCs, which are discussed in more detail later in this section, communicate with the LCUs through the data highway and are responsible for startup and shutdowns as well as safety overrides of the control system.

The data highway holds the entire DCS together by allowing each modular element and each global element to share data and communicate with each other. The data highway is composed of one or more levels of communication hardware and the associated software. System consoles are directly attached to the data highway and can act as a local console for any of the local control units. In addition, system consoles can be used to change linking functions of the distributed elements. Operators have access to the local consoles, but they are usually restricted from modifying certain functions (e.g., altering control configurations). On the other hand, engineers have access to most of the DCS functions. The host computer, which is often a minicomputer, is used to provide a unified information environment for operators (e.g., alarm information), execution of process control recipes affecting multiple unit operations, short term data storage, process optimization calculations and the application of advanced process control software.

DCS Performance and Use. A goal of a DCS is to apply the control calculations for each control loop so fast that the control appears continuous. Because DCSs are based upon sequential processors, each control loop is applied at a discrete point in time and the control action is held constant at that level until the next time the controller is applied. The time between subsequent calls to a controller applied by the DCS is called the **controller cycle time** or the **control interval**. Most DCSs have functionality that permits choosing the frequency of computer calls (i.e., access) to individual controllers which then triggers the control algorithm for the controller to execute its contained PID or other algorithm (i.e., update the controller output). The default standard in industry is a cycle time (i.e., update time) of one second. However, users can choose other cycle times (sometimes as frequently as once per 0.1 second, or a cycle time of several seconds, with users choosing a time depending on the requirements of the control loop. If there are lots of loops needing a high frequency of sampling, often other parts of the automation system may need to make sacrifices on update time [e.g., the human-machine interface (HMI)]. So, situations may exist where controllers in the software execute at a desired high frequency, – but the operator screens showing the status of the controller are updated much less frequently. Some computer systems may accomplish the above by allowing users to set priorities on different computer activities. This is not as clean cut as setting controller access at specific cycle times because it can cause some variability when individual controllers are accessed. For fast changing control loops, it can be important to know the specific controller access update frequency.

Regardless, the vast majority of control loops in industry are set to execute once per second, which most DCSs can easily accommodate along with performing all the other functions assigned to the DCS. While a control computer executing everything quickly was an issue in past decades, Moore's Law (indicating computer execution speeds double about every 2 years due to adding ever more transistors into microchips) has largely eliminated computer execution speed issues, except in unusual situations.

Because DCSs are based on digital calculations, a wide variety of special control options are available in self-contained modular form and can be easily selected by "click and drag" action on most DCSs. In this manner, complex control configurations can be conveniently assembled, interfaced and implemented. In addition, a variety of signal conditioning techniques, including filtering (Section 9.6) and validity checks, can be applied to process measurements.

Process Alarm Systems. To maintain the safety, consistency, and reliability of the operation of a process, processes are associated with alarms to warn the operator of abnormal situations. Many alarms may already come embedded in PLC application software that are provided with vendor supplied unit operation equipment. Other alarms may be added by end users, e.g., to deal with customized applications or monitoring of plant utilities. Alarms are usually designed to identify when (1) a particular piece of equipment is abnormally in or out of service or (2) a process measurement is out of its normal operating range. An example of an equipment service alarm is an alarm that would indicate to the operator that a compressor had been shutdown, perhaps by a power surge in the compressor motor. An example of an out of range process measurement is a high level in a tank or a high temperature in a reactor. For each of these alarm examples, if the operator does not take corrective action (e.g., restart the compressor, reduce the feed to the tank or increase the cooling for the reactor, respectively), serious damage to the process may occur.

Alarms usually use available process measurements from the DCS along with alarm logic to trigger a flashing light on the DCS local console and/or start a buzzer in the control room. If the limits on the alarms are too tight, alarms will routinely trigger during normal operation, creating a nuisance for the operator. Moreover, in this case, it will be more difficult for an operator to distinguish a "nuisance alarm" from an important alarm. On the other hand, if the alarms are applied too loosely, serious operating conditions can occur without the operators knowledge. Therefore, proper alarm design and implementation is a challenging and important problem. See Section 17.4 for additional information regarding alarm systems.

Programmable Logic Controllers (PLCs). **Programmable logic controllers**[2,3] are control computers that are designed to withstand harsh industrial environment and are applied to perform both discrete and continuous control functions. The previous industrial control cases presented in Section 1.4 are examples of continuous control while discrete control involves individual steps. An example of discrete control for a reactor would be to open a valve until the level in the reactor reaches a preset value, then open the valve on the steam to the reactor jacket until the reaction mixture reaches the desired temperature, next start catalyst flow to the reactor and so on. PLCs were first applied by General Motors in the late 1960s for the control of their assembly lines as a replacement for electro-mechanical relay systems. PLCs are more easily reprogrammed than rewiring relay systems when new automobile models were brought on line. Initially, PLCs were designed to replace relay systems and were limited to individual applications of discrete control, such as batch operations and startup and shutdown of processes. Today, PLCs are much more versatile and are also able to apply PID and advanced PID control on a continuous basis as well as discrete control.

PLCs are based on a programming language (**ladder logic**) that is similar to the logic used for communication switching, and as a result, PLCs can usually be programmed by plant electrical personnel. PLCs are relatively low priced, are highly reliable, have low maintenance costs, have relatively small physical dimensions and are able to conveniently communicate with other control systems in a plant.

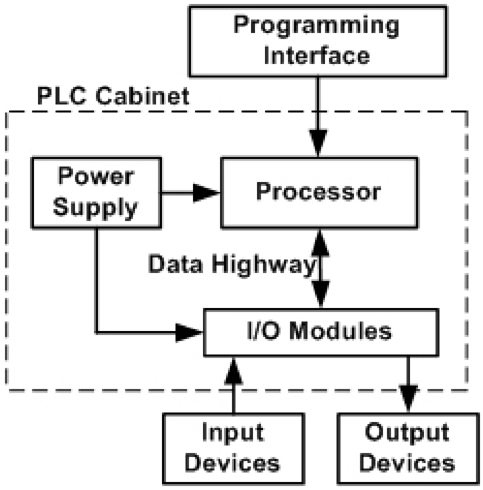


Figure 2.2.6 Architecture of a generic PLC.

The generic architecture for a PLC is shown in Figure 2.2.6. The control logic for a PLC is input to the processor by using the programming interface. Once the PLC is programmed, it receives input measurements from the process through the I/O (input/output) modules and using the programmed instructions determines the appropriate responses, which are sent to the output devices through the I/O modules. PLCs usually contain electric power line conditioning hardware to filter out voltage surges. They also contain a power supply that takes the conditioned 115 V AC and creates other AC and DC voltages needed by different components within the computer system. The PLC processor, like the processor in a PC, is made up of a CPU and a memory device. The CPU is responsible for scanning the inputs and generating the outputs, controlling the I/O data highway and performing self-diagnostics for itself and the other devices that makeup the PLC. The memory device stores the user's application program, the PLC system program and temporarily stores I/O data. Inputs come to the PLC typically as frequency, voltage or current signals. The outputs are usually on-off signals to solenoid valves or analog or digital signals to final control elements (control valves and variable speed pumps). The data highway is a high-speed multiplexer that provides communication between the I/O modules and the CPU.

The two primary types of control computers used in industrial plant automation are PLCs and DCSs. The advantages of PLCs for many applications are 1) the PLC is better able to handle harsh operating environments, such as high temperatures, high humidity, and dusty and/or high vibration conditions, 2) the PLC can have faster control cycles for applications involving a relatively small number of control loops, 3) the total cost of a PLC (capital, maintenance, and down-time costs) can be significantly lower and 4) PLCs may be easier to maintain and update due to their simpler overall software structure (i.e., fewer functions and interfaces to other systems). A very small PLC can have as few as 8 I/O channels while a large PLC can have over 2000 I/O channels. Small PLCs, which typically have between 32 and 512 I/O channels, are typically used for stand-alone unit operations equipment, e.g., compressors, chemical batch reactors, distillation columns. PLCs, while originally developed to replace mechanical relay racks utilizing ladder logic can handle continuous, batch, and/or discrete process control. Large PLCs can be used to control entire plants, especially small plants, although such use is not common. Supervisory control computers overseeing entire plants (in real-time) are usually DCSs.

PC Control Computers. Personal computers (PCs) are commercially available that are capable of interfacing with I/O subsystems and performing process control. A few are used in CPI and bio-technology manufacturing applications although such applications are not common. Due, in part, to their low cost, their most common use is in 1) academic environments and 2) industry for research, bench-scale, and pilot plant experiments and operations. The range of functions available for PC based control computers is significantly reduced compared to PLC and DCS based systems. Also, operating system bugs, lack of redundancy of key components, and other limitations have resulted in system availability (i.e., uptime) of PC based systems being less than that for PLC and DCS based systems. That is, PC based systems, at least historically, are more vulnerable to incidents requiring system rebooting.

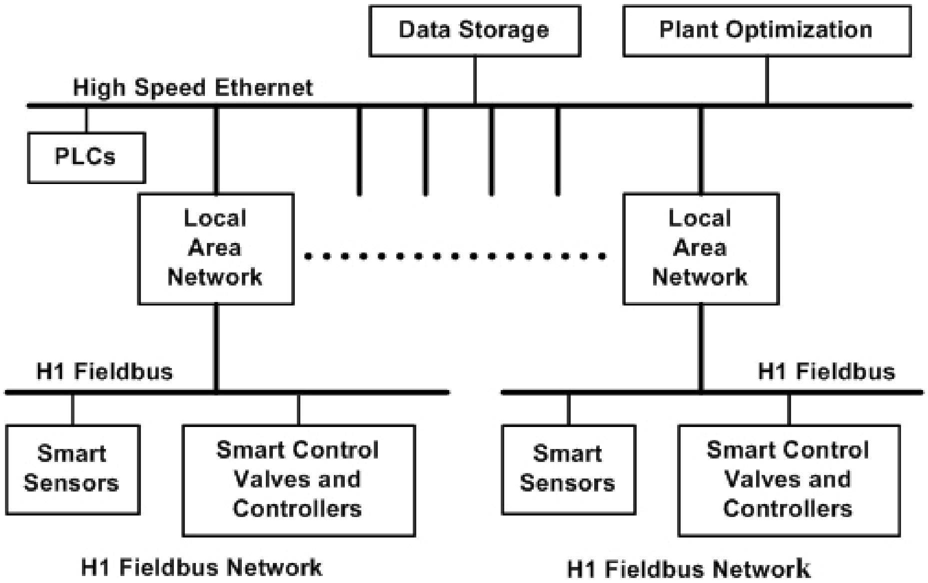


Figure 2.2.7 Schematic of the integration of H1 Fieldbus with a local area network using a high speed ethernet system.

Fieldbus Technology. Fieldbus is a digital network technology used for real-time distributed control, standardized as IEC 61158. The fieldbus approach to distributed control uses H1 fieldbus combined with a high speed Ethernet (HSE) system and is shown schematically in Figure 2.2.7. The H1 fieldbus approach is used or compatible with data highway(s) available with traditional DCSs with DCSs being able to handle any of the multiple standard communication technologies commercially available. Note that Fieldbus (and Profibus) are only digital so do not involve analog signals. H1 fieldbus uses either two wire or fiber optic hardware. The availability of digital networks has enabled the development of smarter sensors, local controllers, and valves which then enables certain control functions to move to (be distributed to) such smart field devices, if desired. For example, the smarter field devices enable greater functionality regarding self diagnostics (since more information is available) than the diagnostics available in host control computers. However, host control computers (DCSs and PLCs) can still be used for all controller logic, if desired, and can also be used to communicate diagnostic messages to operators from smart field devices if such messages are made available by the field devices to the host computer. One of the main differences between Fieldbus and Profibus is that Fieldbus enables peer to peer (e.g., field device to field device) communications, while Profibus only permits a primary/secondary-type relationship regarding communications.

Another major advantage of the fieldbus design is that when a sensor, controller or actuator is connected to the H1 fieldbus, it identifies itself allowing for convenient software connections between the selected sets of actuators, controllers and sensors. This results in a significant reduction in the time and cost associated with system installation compared to a traditional DCS. Fieldbus technology is commercially available and it is rapidly receiving industrial acceptance due to it economic advantages and technological capabilities. Fieldbus is primarily used in the US while Profibus and Device Net are widely used outside the US.

Control System Wiring. The domain of electronic and computer based control is that of an electrical factory with electronics involved in the flow of information from sensors to final element transducers (e.g., I/P converters driving valves), in driving many final control elements (e.g., motors and pumps) and with electronics also involved in powering most of the components in control loops. The installation of both signal and power wiring needs to accommodate manufacturer's recommendations as well as other best practices in order to avoid problems that can interfere with the accurate and reliable collection of measurements and operation of control loops. Often, it is useful to use a trained electrical engineer to design and review electrical aspects of an installation.

One common type of electrical problem is voltage surges, caused by lightning strikes during electrical storms and many other sources. Some components involved in computer process control may include embedded hardware, provided by the equipment vendor, to offset or combat such voltage surges; other components do not have such protection (although external surge protectors can be used). Even with modern technology, it is not unusual that a passing electrical storm in a plant's location will knock out a few pieces of operating equipment. In addition to some hardware components having protection from voltage surges, many plants choose to install UPS (uninterruptible power supply), isolation transformers, or other external equipment on power supply lines to protect computer control systems from voltage surges (and other electrical power line anomalies).

Another common source of electrical problems is ground loops. There are typically several different grounding systems in a manufacturing plant and while they may ultimately be tied together, there are localized areas where the grounds may be at somewhat different potentials. So, any circuit that is grounded in two different places that may be at different voltage levels (e.g., a measurement wire grounded at both the transmitter and computer input ends) has the potential of creating a ground loop which sets up an inappropriate current flow, potentially invalidating the measurement. So care should be taken to ground a particular circuit at only one spot (not multiple locations).

Other common sources of electrical problems are induced electrical noise caused by RFI (radio frequency interferences) and/or EMI (electronic magnetic interferences). RFI interferences can come from sources such as local radar transmitters or walkie talkies used by operators or maintenance personnel walking the production floors. EMI interferences can come from sources such as nearby induction motors starting up or cleaning people operating floor scrubbers on the production floor. These types of interferences can induce spurious currents or voltages in nearby analog signal wiring (especially those carrying low voltage or current signals) and can even cause a "bit" to be added or dropped in certain digital communications. Because of such potential induced interferences, common practice is to ensure that analog signal wiring (e.g., twisted pair) is protected by a metal foil shield (grounded at only one end), which acts as an antenna to pick up spurious electronic signals from the environment and drain them to ground. It has also been reported that any metal housings on sensor transmitters can act as pseudo antennas and pick up extraneous electric signals from the surrounding environment; therefore, use of non-conductive (e.g., plastic) housings is deemed advisable to minimize this possibility.

Another common practice is to ensure that analog signal wiring, especially those carrying low voltage or low current signals, are never installed close to and parallel to power wiring (as such an orientation causes power wiring to induce a current in the signal wiring). If signal and power wiring need to be installed close to one another, it should always be in a perpendicular orientation. There are several other wiring best practices and readers are referred to the Alford/Cline reference[4] on Installation Qualification, to Equipment Vendor recommendations, and to grounding requirements associated with buildings in which the process operates.

Note that the issues noted above embody one of several reasons digital networks (e.g., Fieldbus, Profibus) have been developed; that is, to minimize the use of analog wiring for control systems.

Self-Assessment Questions

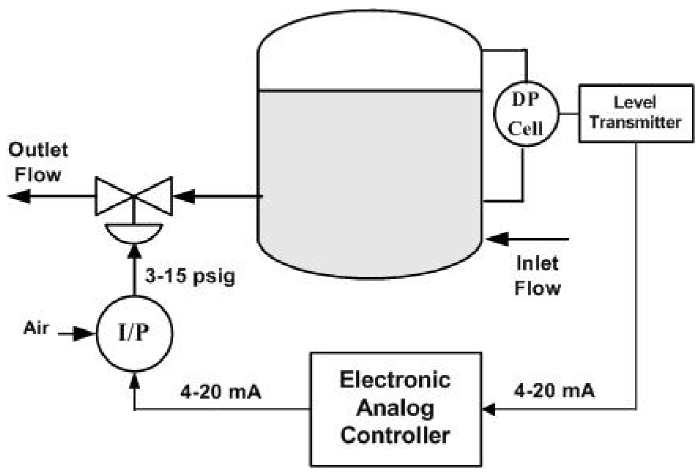
Q2.2.1 For a pneumatic controller, what mechanical devices are used to implement PID control?
Q2.2.2 Why have DCSs replaced electronic analog controllers?
Q2.2.3 How frequently can a DCS execute most regulatory control loops?
Q2.2.4 Using Figure 2.2.4, explain how process data are stored and later displayed on a system console for a DCS.
Q2.2.5 What is a PLC and how is it different from a DCS? How are they alike?
Q2.2.6 What is the difference between a traditional DCS and the fieldbus approach to distributed control?

Self-Assessment Answers

Q2.2.1 Bellows, baffles and nozzles are the mechanical devices that are used in pneumatic controller to implement PID control.
Q2.2.2 DCSs have replaced analog controllers primarily due to lower cost per control loop and because DCS offer additional capabilities, such as more convenient application of advanced PID controller, vastly better data storage and retrieval, better interface for process optimization and better interface with PLCs for startup and emergency shutdown.
Q2.2.3 Most regulatory control loops in a DCS are executed each 0.5 to 1.0 s.
Q2.2.4 Data from the process (i.e., temperature, flow, level, pressure, and composition measurements) as well as controller output signals are transferred as one minute or one hour averages, for example, from the LCU through the data highway to the host computer where they are downloaded to the data storage unit. The local console has process data trending capabilities such that when specified data from a specified time interval is requested, the data is captured from the data storage units and passed to the host computer where it is transferred to the LCU for display through the data highway.
Q2.2.5 A PLC is a programmable logic controller and is generally used for controlling a sequence of discrete operations, such as startup and shutdown operations. The DCS is generally used for continuous control of the process after it has been started-up. A PLC and a DCS are both microprocessor-based systems that are used to control the process. Today many of the functions that were available only on a PLC are becoming available on DCSs and likewise many of the functions that were available only on a DCS are becoming available on PLCs.
Q2.2.6 The DCS performs all the control calculations in the LCU, while the fieldbus approach uses field mounted controllers for regulatory control loops and fieldbus uses only digital communications. More importantly, fieldbus is able to fully utilize the diagnostic capabilities of smart sensors and valves, thus significantly enhancing the reliability of control operations.

Self-Assessment Problems

P2.2.1 For the level control loop shown in Figure 1.4.4, make a drawing similar to Figure 2.1.2 and



Solution for P2.2.1

list all signals on your diagram assuming that an electronic analog controller is used instead of a DCS.

Self-Assessment Answers

P2.2.1 Solution provided in adjacent figure. Notice that the D/A and A/D converters are not required for an analog controller because the input and the output for an analog controller is a 4-20 mA signal.

2.3 Actuator Systems (Final Control Elements)

Actuator systems, or as they are known industrially final control elements, are used to make changes in the MV for a control loop. That is, the actuator receives its input signal from the controller and based on that signal it normally changes flow rate to the process. Control valves and adjustable speed pumps are the two primary types of actuators used in the CPI and bio-processing industry, respectively. Globe control valves are the most common actuator used for the CPI and an actuator system based on a control valve consists of the globe control valve, which includes the valve body and the valve actuator, the I/P transmitter and the instrument air system. For bio-processing systems, final control elements usually need to be easily sterilizable and cleanable, leading to common use of variable speed pumps (e.g., peristaltic pumps), diaphragm valves, and butterfly valves for such use.

Linear and Equal Percentage Globe Control Valves. Control valves are broadly classified as sliding stem or rotary shaft devices. The most common type of control valve body in the CPI is the sliding stem valve shown schematically in Figures 2.3.1 and 2.3.2, which is known as a globe valve. They are called globe valves because of the globe shape of the flow chamber of the valve. Butterfly valves, which are discussed later in this section, are a rotary shaft valve. For globe valves, the closure member is a **valve plug** positioned at the end of the **valve stem**. As the valve stem is lowered, the plug approaches the **valve seat**, restricting the area for flow through the valve. When the plug makes contact with the valve seat, the valve is closed and flow through the valve is shut off. Globe valves are characterized by the fact that the plug travels in a direction perpendicular to the flow through the valve. The valve stem is attached to a spring-and-diaphragm actuator and is moved up or down by the force difference between the instrument air and the valve actuator spring. Consider Figure 2.3.2 for which, as the instrument air pressure is increased, the diaphragm moves against the spring, moving the stem downward, thus moving the valve plug closer to the valve seat, reducing the flow through the valve. Likewise, when the air pressure is decreased, the flow through the valve increases. Therefore, changes in the instrument air pressure coming from the I/P converter directly effect changes in the flow rate through the control valve.

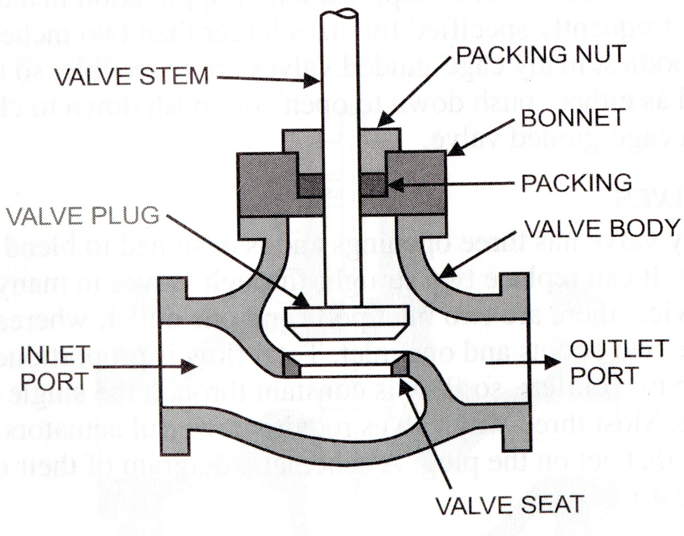


Figure 2.3.1 Schematic of a typical valve body for a globe valve. Printed with permission of ISA.

Figure 2.3.2 shows a detailed cross-section of a globe control valve along with notation indicating some of the key components of the control valve. Note that the valve actuator is mounted by two bolts to the top of the valve

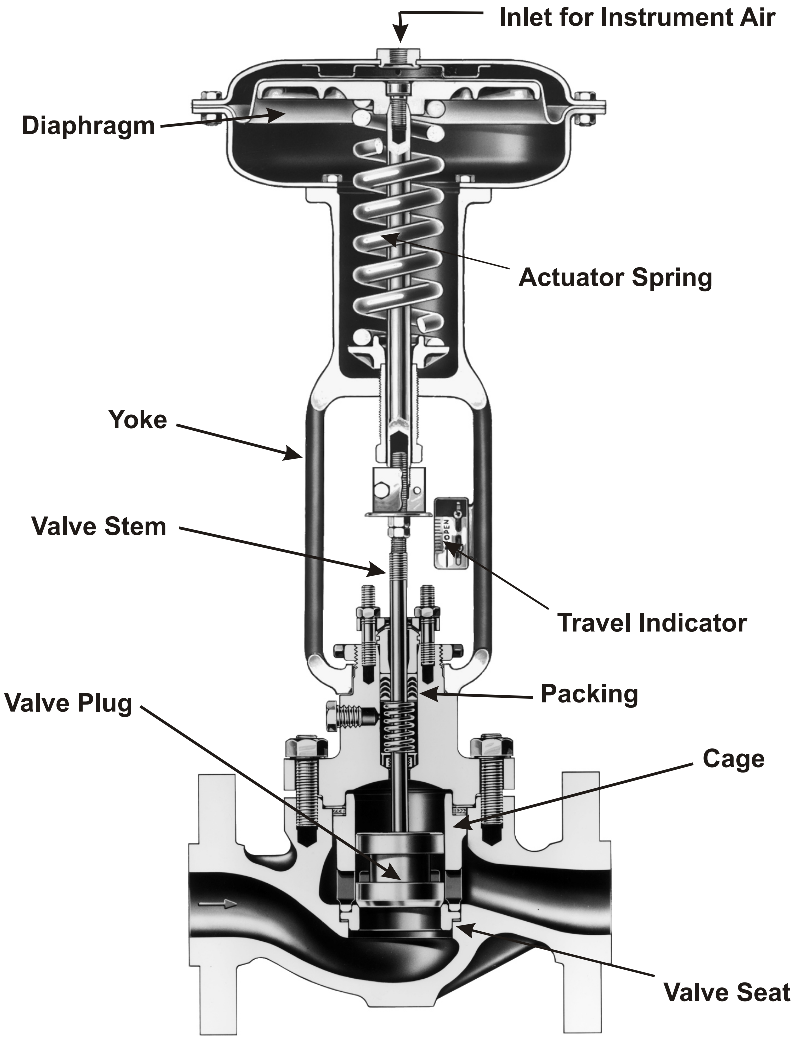


Figure 2.3.2 Cross-section of a globe valve. Courtesy of Fisher-Rosemount.

Figure 2.3.3 Photograph of a valve cage. Courtesy of Fisher-Rosemount.

body. The valve body has a valve plug in a **cage-guided valve** arrangement. Note that the flow direction indicated in the inlet flow channel shows that the flow enters the plug/seat area from the bottom. If the flow direction were opposite to this, the frictional flow forces would tend to prematurely shut off the flow through the valve. The valve cage provides guidance for the plug as the plug moves toward or away from the valve seat. The cage also provides part of the flow restriction produced by the control valve. An example of a valve cage for a globe valve is shown in Figure 2.3.3. The valve packing reduces the leakage of the process stream into the environment but results in resistance to movement of the valve stem and contributes to sticking of the valve. Sticking of a globe valve is reduced by using a stiff actuator spring and a diaphragm with a large surface area, which give the globe valve its characteristic shape. The

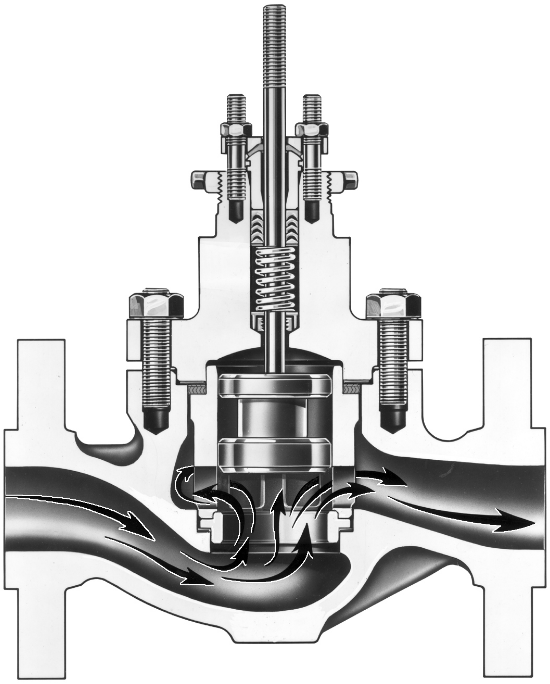


Figure 2.3.4 Cross-section of a globe valve body assembly with an unbalanced plug. Courtesy of Fisher-Rosemount.

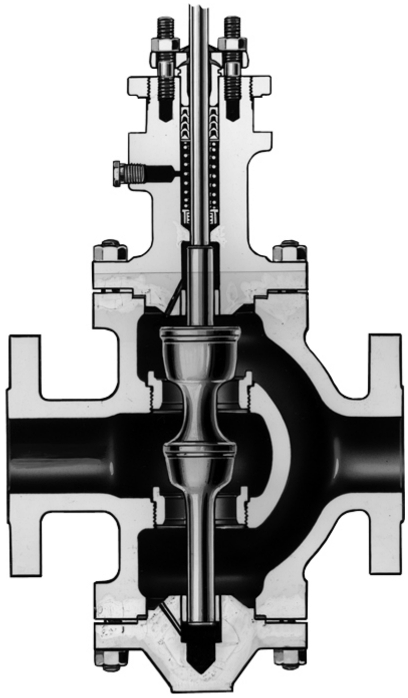


Figure 2.3.5 Cross-Section of a globe valve body assembly with a balance plug. Courtesy of Fisher-Rosemount.

travel indicator provides a visual indication of the valve stem position. Valve sticking, common with many types of control valves, is more commonly known as stiction. Stiction is defined as the static friction that needs to be overcome to enable a change if valve position.

Figure 2.3.4 shows the valve body assembly for a globe valve with an unbalanced plug (compare to Figure 2.3.2). The unbalanced plug is subject to a static force directly related to the pressure drop across the valve and a flow force due to the fluid velocity past the plug. The greater these forces, the more force that is required to close the valve and the less force that is required to open the valve. Figure 2.3.5 shows the valve body assembly for a globe valve with balanced plugs. This valve is referred to as having balanced plugs because the top and bottom of the plug are subjected to the same downstream pressure when the valve is closed but these forces balance each other. Thus, the static force on the balanced valve plug is low. As a result, valves with balanced plugs are preferred because they are faster responding than valves with unbalanced plugs and require smaller valve actuators, but they should be used only with clean liquids and are not effective for tight shut-off of the flow. For example, an unbalanced plug is preferred for service with a liquid that tends to crystallize on the surface of the valve plug because shear forces for an unbalanced plug would prevent the buildup of crystals on the valve plug. A balanced plug is more susceptible to buildup due to the lower velocity past the plug because the balanced plug arrangement uses two separate flow paths.

In looking at Figures 2.3.1 and 2.3.2, readers will note that a closed (i.e., non-flow through) pocket or cavity exists in the valve above the valve plug. This cavity makes it difficult to sterilize the valve and also to clean it of any particulate (e.g., microbes or spores) that may have collected in the cavity. Note that such microbes or spores, if viable, could contaminate the following batch process. For this reason, globe valves are not used very much as final control elements for sterile operations, common in most bioprocesses.

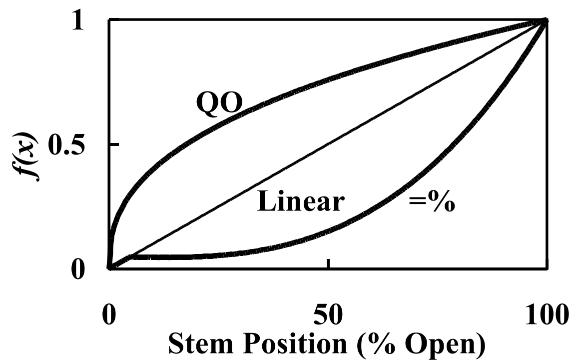


Figure 2.3.6 Inherent valve characteristic for quick opening (QO), linear and equal percentage (=%) valves.

Inherent and Installed Valve Characteristics. The installed valve characteristics determines the useful operating range of a control valve. The installed valve characteristics of a control valve are determined from the inherent valve characteristics of the valve and the available pressure drop across the valve.

Inherent valve characteristic. Figure 2.3.6 shows the **inherent valve characteristics,** which indicate how the flow through a control varies with stem position for a **fixed pressure drop across the valve**, for a quick opening (QO), linear and equal percentage valve (=%). Note that $f(x)$ is a normalized valve coefficient, i.e.,

$$f(x) = \frac{C_v(x)}{C_v^{\max}} = \frac{F_V(x)}{F_V^{\max}} \qquad \textbf{2.3.1}$$

where $C_v(x)$ is the valve coefficient and is proportional to the volumetric flow through the valve, $F_V(x)$, for a fixed pressure drop across the valve. Note that $f(x)$ is equal to zero when the valve is closed (x=0) and is equal to unity when the valve is fully open (x=100%).

The design of the plugs, valve seats and cages (where applied) determine the valve coefficient versus stem position that a control valve provides and is also known as the **valve trim**. That is, the shape of the valve plug and the flow openings in the cage determine the shape of the flow restriction as the valve stem position is changed. For example for a quick-opening valve, as the valve is opened from the closed position, the cross-sectional area of the restriction of the valve increases much faster than the linear or equal percentage valves. Quick-opening valves are not usually used for feedback flow control applications but are used in cases where it is important to start a flow rate as quickly as possible (e.g., coolant flow through a by-pass around a control valve for an exothermic reactor prone to thermal runaway). Linear and equal percentage valves are primarily used for flow control applications. Table 2.1 list the C_vs for various sizes of a specific equal percentage valve as a function of stem position expressed as a percentage of total travel. Other types of equal percentage valves will have different values of C_vs than the ones listed in Table 2.1. Note that the values in Table 2.1 could be used to generate the inherent valve characteristics for the different valve sizes by applying Equation 2.3.1, recognizing that $C_v^{\max}$ is the value of C_v for the stem position at 100% open. The following equation represents C_v for a linear valve as a function of stem position, x.

$$C_v(x) = C_v^{\max} \frac{x}{100} \qquad \textbf{2.3.2}$$

Table 2.1 Representative C_vs for an Equal Percentage Globe Valve.

	Body Size (in)	Stem Position as a Percentage of Total Travel									
		10	20	30	40	50	60	70	80	90	100
	1	0.79	1.25	1.80	2.53	3.63	5.28	7.59	10.7	12.7	13.2
	1.5	0.80	1.23	1.91	2.95	4.30	6.46	9.84	16.4	22.2	28.1
C_v	2	1.65	2.61	4.30	6.62	11.1	20.7	32.8	44.7	50.0	53.8
	3	3.11	5.77	9.12	13.7	21.7	36.0	60.4	86.4	104	114
	4	4.90	8.19	13.5	20.1	31.2	52.6	96.7	140	170	190

Flow equation for a control valve. The installed valve characteristics can be determined from the valve coefficient, $C_v(x)$, and the pressure drop across the valve using the valve flow equation. A simplified valve flow equation based on incompressible flow through a control valve is given by

$$F_V = K\,C_v(x)\sqrt{\frac{\Delta P}{s.g.}} \qquad \textbf{2.3.3}$$

where F_V is the volumetric flow rate through the valve, K is a constant that depends on the units used in this equation, $C_v(x)$ is the valve coefficient, which is dependent upon the stem position (x) [i.e., $C_v(x) = C_v^{max}$ when the valve is fully open (i.e., x equal to 100%) and $C_v(x)$=0 when the valve is closed], $s.g.$ is the specific gravity of the fluid and ΔP is the pressure drop across the valve. **Note that the values of C_v are chosen such that K is equal to unity when the pressure drop is given in psi and the flow rate is expressed in gpm.** A more complex formulation is required for compressible flow through a control valve.

Example 2.2 Flow Rate through a Control Valve

Problem Statement. Calculate the flow rate of water through a 4-inch equal percentage control valve that is 80% open based on stem position with an available pressure drop across the valve of 30 psi. Obtain the C_v for this valve from Table 2.1.

Solution. Using Equation 2.3.3 with C_v equal to 140 from Table 2.1 and the specific gravity and K are equal to unity,

$$F_V = KC_v(x)\sqrt{\frac{\Delta P}{s.g.}} = 1 \times 140\sqrt{30/1} = 767 \text{ gpm}$$

Example 2.3 Pressure Drop Across a Control Valve

Problem Statement. Calculate the required pressure drop across a 2-inch linear control valve that is 40% open for a flow rate of water of 35 gpm. Assume that the maximum C_v for the linear valve is equal to the C_v for an equal percentage valve that is fully open (Table 2.1).

Solution. From Table 2.1, C_v^{max} is equal to 53.8. Therefore, C_v for a linear valve 40% open is equal to 0.4×53.8=21.52. Rearranging Equation 2.3.3 to is solve for P yields

$$P \qquad \frac{s.g.\,F_v^2}{K^2\,C_v^2} \qquad \frac{1}{1^2}\left|\frac{35^2}{21.52^2}\right. \qquad 2.65\,\text{psi}$$

Example 2.4 Determine the Installed Flow Rate for a Known Valve Stem Position

Problem Statement. Determine the flow rate though a 3-inch equal percentage control valve that is 40% open for the C_v data given in Table 2.1 and the installed pressure drop relationship presented in Table 2.2.

Table 2.2 Installed Pressure Drop For a Control Valve versus Flow Rate

F_V (gpm)	P (psi)	F_V (gpm)	P (psi)
50	19.3	74	18.0
54	19.1	78	17.8
58	18.9	82	17.5
62	18.7	86	17.2
66	18.5	90	16.9
70	18.3	94	16.6

Solution. From Table 2.2, P is a function of flow rate; therefore, an iterative solution of Equation 2.3.3 is required to determine the unknown flow rate, F_V. That is,

$$F_V \qquad KC_v(x)\sqrt{\frac{P(F_V)}{s.g.}}$$

From Table 2.1, C_v is 13.7 for a 3-inch valve that is 40% open. From Table 2.2, it is clear that installed pressure drop is not a strong function of flow rate; therefore, a flow rate can be assumed to calculate the pressure drop for the installed valve, and this value used to update the flow rate. Assuming that the flow rate is 94 gpm, the installed pressure drop is equal to 16.6 psi using Table 2.2, and then Equation 2.3.3 yields a flow rate of 55.8 gpm. Using this flow rate in Table 2.2, the pressure drop for the installed valve is 19.0 psi by linear interpolation. Once again, Equation 2.3.3 is applied, yielding a calculated flow rate of 59.7 gpm. Updating P (18.8 psi) yields a flow rate of 59.4 gpm, which is the converged solution for this problem.

Installed Valve Characteristic. The inherent valve characteristics of a control valve are based on a fixed pressure drop across the valve. For most applications, however, the pressure drop across a control valve varies significantly with the flow rate. That is, you have to consider the entire flow system (i.e., pump or pressure source, piping and control valve) and this can be done by applying the Bernoulli Equation[5] for incompressible flow in a flow system with a control valve:

$$\frac{p_{system}}{\quad} \qquad \frac{g \quad h_{system}}{g_c} \qquad W_{pump} \qquad p_{valve} \qquad p_{frictional}$$

where p_{system} is the pressure drop between the inlet and the outlet of the system, is the density of the fluid, h_{system} is the elevation change between the inlet and outlet of the system, W_{pump} is the work provided by a pump, p_{valve} is the pressure drop across the valve and $p_{frictional}$ is the pressure drop in the lines due to frictional losses.

The **installed valve characteristics** give the flow rate through the valve as a function of stem position for a valve in service. In this case, the pressure drop across the valve is a function of the flow rate; therefore, the installed valve characteristics depend on the particular flow system in which the valve is applied. From a process control standpoint, it is desirable to have a control valve that exhibits a linear relationship between the installed flow rate and stem position over a wide range for the installed valve. As a rule of thumb, **the slope of the installed valve characteristic versus stem position should not vary greater than a factor of four over the range of operation for effective flow control**[5]. Otherwise, accurate flow control will not result. For example, when the slope of the installed valve characteristic is too large, the flow through the control valve will tend to oscillate because small changes in the valve position will result in large changes in the flow rate. On the other hand, if the slope is too small, the control of the flow rate will be sluggish because changes in the valve position will produce relatively small flow rate changes.

As a general approximation based on the Bernoulli Equation, the pressure drop available for a control valve, P_{valve}, is given by

$$P_{valve} \qquad P_{driving\ force} \qquad P_{line\ losses} \qquad \textbf{2.3.4}$$

where the pressure drop driving force is the combination of the work provided by a pump and any elevation changes and neglecting kinetic energy changes. The overall pressure driving force for flow typically comes from a hydrostatic head or from a pump. In general, hydrostatic head remains relatively constant with flow rate while the head developed by a pump tends to decrease as the flow rate is increased. The pressure drop associated with line losses comes from frictional losses for straight runs of pipe, elbows and sudden expansions or contractions and these losses tend to vary with the square of the flow rate. That is, as the flow rate increases, the pressure drop due to line losses increases quadratically. Therefore, from Equation 2.3.4 the pressure drop for the valve decreases as the flow rate through the system increases, i.e., the valve must open for an increase in the flow rate.

To determine the installed valve characteristics for a valve, one must first calculate the pressure drop available for the control valve as a function of flow rate using a pressure balance equation similar to Equation 2.3.4. The application of the pressure balance equation requires quantitatively describing the line losses and the driving force for flow in terms of the hydrostatic head or the head provided by a pump as a function of flow rate through the piping system. The line losses can be calculated from the Reynolds Number and the length of piping run and the number of elbows, expansions and contraction in the flow loop applying Bernoulli's Equation[4]. The head developed by the pump as a function of flow rate can be obtained from the pump curve for the pump in question.

Once P_{valve} is determined as a function of flow rate, the installed valve characteristics can be determined directly using Equation 2.3.3 along with the inherent valve characteristics of the control valve. That is, first select a flow rate, F_V, and determine the corresponding available pressure drop, P, using the available pressure drop (e.g., Equation 2.3.4). Then calculate the valve coefficient, $C_v(x)$, using the valve flow equation (Equation 2.3.3).

Finally, the corresponding valve position can be determined from the inherent characteristics of the valve, e.g., Table 2.1 or Equation 2.3.2. In this manner, flow rates are chosen and the corresponding valve position is determined until the complete operating range of the valve is determined, e.g., $10\%<x<90\%$. Instead of this procedure, you can select the valve position and then calculate the corresponding flow rate, but this requires an iterative solution procedure for each assumed valve position as demonstrated by Example 2.4. In certain cases, it is not necessary to determine the installed characteristic of a control valve if the valve is used for a relatively small operating range. On the other hand, when a control valve must be precisely sized for a relatively wide operating range, an accurate determination of the installed valve characteristic is required.

Example 2.5 Comparison between Linear and Equal Percentage Valves for a Pump Driven Flow System.

Problem Statement. Consider the flow system shown schematically in Figure 2.3.7. This schematic represents the flow system that is used to deliver reflux from the accumulator to the top tray of the column. Compare the installed valve characteristics for a 4-inch linear and a 4-inch equal percentage valve for this system. Use the C_vs for the equal percentage valve from Table 2.1. Assume that the lines are 3-inch Schedule 40 piping.

Solution. There are three sources of pressure changes in this flow system: (1) pressure drop through the straight run pipe, elbows and fittings, and the orifice for the flow sensor/transmitter, (2) the head increase provided by the pump and (3) the pressure drop across the control valve. Because, for this example, the pressure above the liquid in the accumulator and the discharge pressure on the top tray of the column are approximately equal, the pressure drop across the control valve (P_{valve}) is set by the difference between the head developed by the pump (P_{pump}) and the pressure drop created by the line devices (P_{line}), i.e.,

$$P_{valve} \quad P_{pump} \quad P_{line}$$

Figure 2.3.8 shows the pressure developed by the pump, the pressure drop due to line losses and the available pressure drop for the control valve as a function of flow rate for the flow system shown in Figure 2.3.7. The

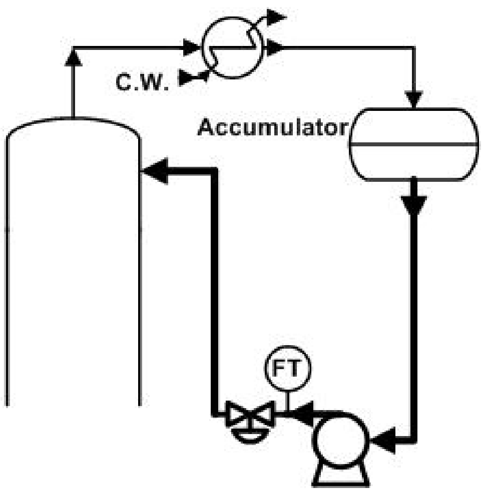


Figure 2.3.7 Schematic of a flow system for delivering reflux from the accumulator to the top tray.

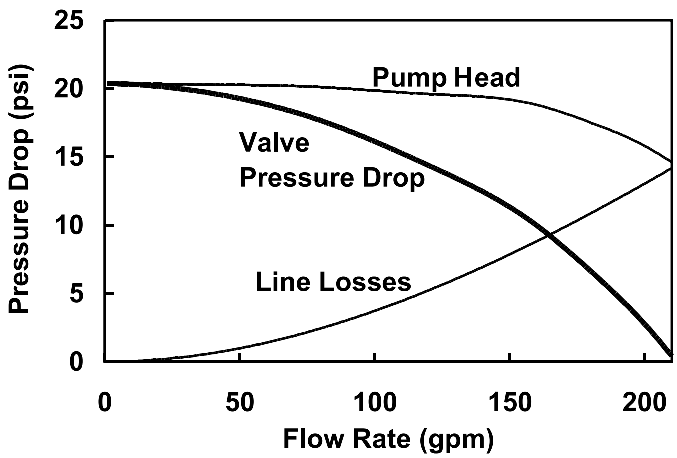


Figure 2.3.8 Pressure drop versus flow rate for the reflux flow system.

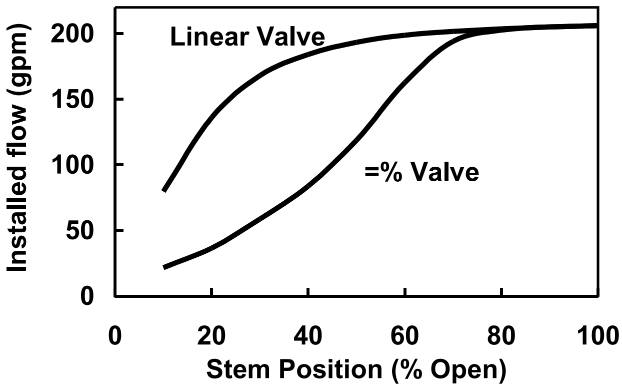


Figure 2.3.9 The installed valve characteristics for a linear and equal percentage valve for the reflux flow system.

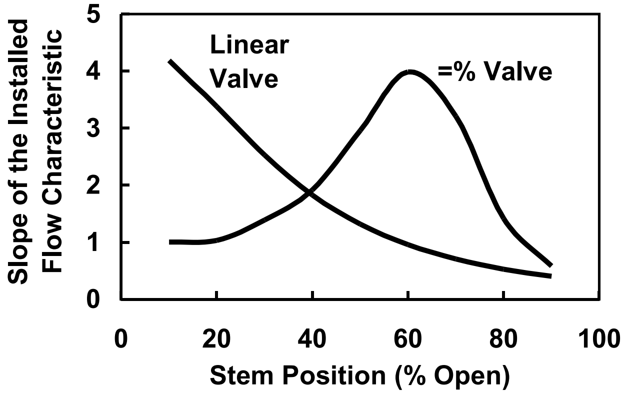


Figure 2.3.10 Slope of the installed valve characteristics for a linear and equal percentage valve.

relationships shown in this figure can be understood by recognizing that, at low flow rates, the pump head is at its largest value while the line losses are at their smallest. At this flow rate, the control valve must provide a relatively large pressure drop to maintain a low flow rate in the flow system. At large flow rates, the pump head is significantly reduced while the line losses are at their highest level. As a result, the control valve must provide a relatively small pressure drop to maintain the large flow rate in the flow system. For this case, the pressure drop across the control valve varies significantly with flow rate.

According to the procedure outlined earlier, flow rates are selected and the corresponding valve positions determined until the full operating range of the valve was calculated. Figure 2.3.9 shows the installed valve characteristics for a linear and equal percentage valve for this example.

While it is clear from Figure 2.3.9 that there are differences between the installed valve characteristics for a linear and an equal percentage valve, at first glance the performance of these two valves may not appear to be significantly different. Because the performance criterion for a valve is based on the change in the slope of the installed valve characteristic, the slope of the installed valve characteristic (Figure 2.3.9) for the linear and equal percentage valves is shown in Figure 2.3.10. Based on the heuristic rule (i.e., the ratio of the maximum to the minimum slope of the installed valve characteristic should be less than 4), the linear valve could operate effectively between 10% and 55% (i.e., slope range is from 4 to 1) or between 55% and 90% (i.e., slope range is from 1 to 0.25). On the other hand, the equal percentage valve can operate effectively between 10% and 80%. **This example demonstrates the advantage of equal percentage valves compared to linear valves for cases in which the pressure drop across the control valve varies significantly with flow rate.**

Example 2.6 Comparison between Linear and Equal Percentage Valves for a Gravity Driven Flow System.

Problem Statement. Consider the flow system shown in Figure 2.3.11. A 5-inch line discharges liquid from the tank to an open reservoir. Compare the installed valve characteristics for a 4-inch linear and a 4-inch equal percentage valve for this system. Use the C_vs for the equal percentage valve from Table 2.1.

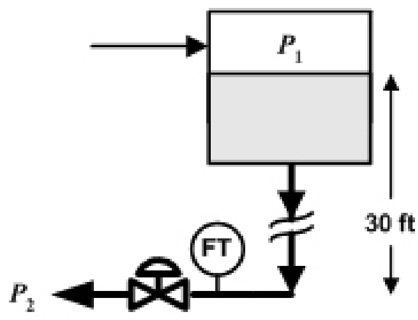


Figure 2.3.11 Schematic of a gravity flow line from a tank with a control valve.

Solution. The hydrostatic head (30 ft) provides the driving force for flow through this system while the line losses and the control valve cause pressure drops. The change in the liquid level in the elevated tank is negligible relative to the 30 ft elevation difference; therefore, the hydrostatic head is assumed to be constant. Similar to the previous example, because the pressure above the liquid (P_1 in Figure 2.3.11) in the tank and the discharge pressure (P_2 in Figure 2.3.11) are assumed equal to atmospheric pressure, the pressure drop required by the control valve is given by

$$P_{valve} \qquad P_{head} \qquad P_{line}$$

Figure 2.3.12 shows the hydrostatic head, the pressure drop due to line losses and the required pressure drop across the control valve as a function of flow rate. In this example, an oversized line was chosen; therefore, the line losses, which were modeled in a manner similar to the approach used for the last example (i.e., using Bernoulli's equation), remain moderate even for the largest flow rate. As a result, the pressure drop for the control valve remains relatively constant for the full range of flow rates considered. On the other hand, if a smaller line were chosen, the pressure drop across the control valve would have been a much stronger function of flow rate resulting in a more nonlinear installed valve characteristic for the linear valve in this case.

The installed valve characteristics for a linear valve and an equal percentage valve (=%) for this case are shown in Figure 2.2.13. These results were generated by choosing flow rates and calculating the corresponding valve position in a manner similar to that used in Example 2.5. The available pressure drop for the control valve versus flow rate shown in Figure 2.3.12 was used in this procedure. The C_vs for the equal percentage value were taken from Table 2.1. The C_vs for the linear valve were determined using Equation 2.3.2 assuming that the maximum value of C_v for the linear valve is equal to the maximum C_v for the equal percentage valve. From Figure 2.3.13, it is clear that the linear valve is the preferred choice in this case because its installed valve characteristic is much more linear than for the equal percentage valve. **This example demonstrates the advantage of linear valves compared to equal percentage valves for cases in which the pressure drop across the control valve does not vary significantly with flow rate.**

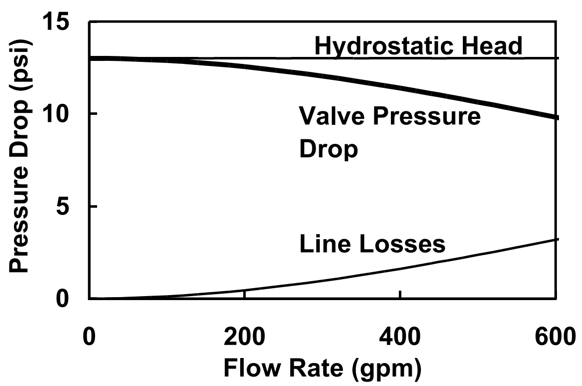


Figure 2.3.12 Pressure drop versus flow rate for discharge from an elevated tank.

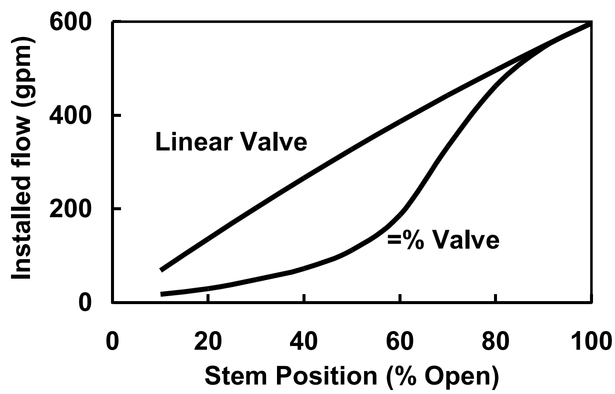


Figure 2.3.13 Installed valve characteristic for an equal percentage and a linear valve applied to the discharge line for an elevated tank.

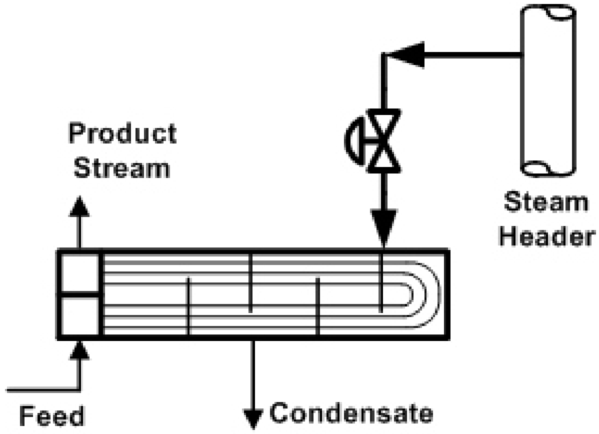


Figure 2.3.14 Schematic of a flow system from a steam header to a heat exchanger.

In approximately 10% of the control valve applications in the CPI[6], the pressure drop across the control valve remains relatively constant and a linear control valve is preferred. Figure 2.3.14 shows a case in which the flow rate of steam is manipulated to heat a process stream. Because the pressure losses in the lines are usually relatively small compared to the pressure difference between the steam header and the steam pressure in the heat exchanger, the pressure drop across the control valve remains relatively constant for a wide range of steam flow rates. Linear control valves are preferred when the pressure drop across a control valve provided by the flow system remains relatively constant.

If the ratio of pressure drop across the control valve for the lowest flow rate to the highest flow rate is greater than 5, an equal percentage valve is recommended[7]. (Remember that the lowest flow rate has the largest pressure drop for the control valve.) For Example 2.5, the maximum flow rate is approximately 200 gpm, which corresponds to 2 psi pressure drop across the valve from Figure 2.3.8 while the minimum flow rate (20 gpm) corresponds to a pressure drop of 20 psi. Therefore, the ratio of maximum pressure drop to the minimum pressure drop is 10 (i.e., 20/2), indicating that an equal percentage valve should be used. For Example 2.6, the maximum flow rate is approximately 600 gpm, which corresponds to a 10 psi pressure drop across the valve from Figure 2.3.12 while the minimum flow rate (60 gpm) corresponds to a pressure drop of approximately 13 psi. Therefore, the ratio of the maximum to minimum pressure drop is 1.3, indicating that a linear valve should be used. Therefore, Examples 2.5 and 2.6 are consistent with this guideline. It should be remembered that **a control valve can be converted to either a linear or equal percentage valve by changing the valve cage**, which is relatively easy to accomplish once the valve is removed from service.

Control Valve Sizing. Sizing of control valves is important because if the valve is oversized or undersized, it can significantly affect the range over which the valve provides accurate flow metering. When the valve is oversized, at times the valve will not be sufficiently open to allow the valve to accurately control the valve plug position just above the valve seat. Typically for valve stem positions less than 10% open, the plug will regularly impact the valve seat while trying to maintain a valve position less than 10% open, causing rapid wear to the valve. When the valve is undersized, the valve will try to operate at valve positions greater than 90% open so that accurate control is not possible or, in certain cases, the required flow cannot be met even when the valve is fully open. As a result, the range of accurate flow rate control for industrial control valves is limited. In general, the stem position for the maximum flow should be less than 85-90% open and the stem position for the minimum flow rate should be at least 10-15% open. Therefore, the **turndown ratio**, which is the ratio of the maximum to minimum controlled flow rate, should, in general, be less than 9 (i.e., 90%/10% assuming a linear installed valve characteristic).

When the turndown ratio is moderate (e.g., less than 3), a detailed valve sizing is not required. When the turndown ratio is moderate, as a rule of thumb the valve size is usually set equal to the line size or one size smaller. For example, for a 4-inch line, a 3-inch or a 4-inch control valve is selected. The key issue here is to ensure that the control valve can handle the maximum flow rate. That is, the control valve should be no more than 85-90% open for the maximum anticipated flow rate.

As the turndown ratio for a control valve increases, the proper sizing of the valve and specification of the valve plug and valve cage geometry becomes a much more challenging problem because the valve must be able to accurately control the flow rate at both the minimum and maximum flow rates. Control valve vendors typically offer software to size control valves, but the control engineer should ensure that the available pressure drop across the control valve for the maximum and minimum flow rate used by the vendor to size the valve adequately represents the process.

Example 2.7 Control Valve Sizing Problem

Problem Statement. Size a control valve for service in a line carrying water with a maximum flow rate of 150 gpm and a minimum flow rate of 30 gpm. Assume an equal percentage valve with the pressure drop versus flow rate shown in Figure 2.3.8. Use the C_vs for equal percentage valves of different sizes that are presented in Table 2.1.

Solution. A simplified design procedure is used to size the valve for this application. The valve will be sized so that the smallest valve is used that will effectively deliver the maximum flow rate. Therefore, the value of C_v is calculated for the maximum flow rate and this value can then be used to select the proper valve size by maintaining the valve position less than 85% open at the maximum flow rate.

From Figure 2.3.8, the available pressure drop is approximately 11 psi at the maximum flow rate of 150 gpm; therefore, using Equation 2.3.3, C_v is equal to 45.2, which corresponds to a valve position of 81% using linear interpolation applied to Table 2.1 for a 2-inch valve. For the minimum flow rate (30 gpm), the available pressure drop is approximately 19 psi; therefore, C_v is equal to 6.9, which corresponds to a valve position 40% open for a 2-inch valve. While this sizing should work, the largest flow rate is near the upper limit for controllability for a 2-inch valve. A 3-inch valve would be only 65% open at the maximum flow rate and 23% open at the minimum flow rate. A 3-inch valve will accommodate larger production rates in the future, but a 3-inch valve costs more than a 2-inch valve. Strictly speaking, the 2-inch valve is the solution to the original problem, but a 3-inch valve can also be a viable choice depending on the specific application.

Cavitation. Cavitation results when the liquid vaporizes and then implodes inside the control valve. As a liquid flows through a control valve, the pressure drops sharply near the restriction between the valve plug and the valve seat due to high velocity in this region. As the liquid passes the valve restriction region and enters a region with a larger cross-section, the pressure increases sharply due to the drop in the fluid velocity (i.e., pressure recovery). If the pressure in the valve restriction region is less than the vapor pressure of the liquid, a portion of the liquid vaporizes and, when the pressure recovers due to a drop in the velocity, the bubbles violently collapse at nucleation sites on the metal surface. Cavitation results in noise, vibration, reduced flow and possibly rapid erosion of the body of the valve. In the field, cavitation sounds like marbles are flowing through the valve and can result in damage to a control valve if allowed to persist. If the operating pressure of a control valve is increased, this can eliminate cavitation. Therefore, increasing the downstream pressure from a control valve can be an effective means to eliminate cavitation in certain cases.

Valve Actuator. The valve actuator provides the force necessary to change the valve stem position and alter the flow rate through the valve. The valve actuator must provide the force necessary to overcome pressure forces,

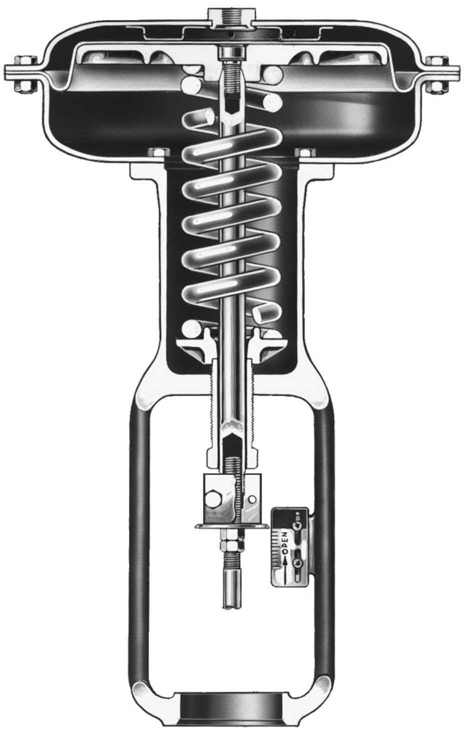


Figure 2.3.15 Cross-section of an air-to-close valve actuator. Courtesy of

flow forces, friction from valve packing (the major contribution) and friction from the plug contacting the valve cage.

Figure 2.3.15 shows a cross-section of a typical **air-to-close actuator** (i.e., a reverse-acting actuator). The pressure of the instrument air acts on the diaphragm/spring system from the top causing the valve to close as the air pressure supplied to the valve actuator is increased. The diaphragm is constructed of an air impermeable, flexible material (typically fabric-reinforced neoprene) that allows the valve plug to move from fully open to closed as the instrument air pressure is increased from 3 to 15 psig. Note that the force generated by the instrument air pressure on the surface of the diaphragm is balanced by the force of the compressed actuator spring (Figures 2.3.2 and 2.3.15). For an **air-to-open actuator** (a direct-acting actuator), the instrument air usually enters below the diaphragm so that, as the air pressure is increased, the valve stem moves upward, opening the valve. Valve actuators are generally chosen to provide a fail-safe function. That is, in the event of a loss of instrument air pressure, the valve actuator causes the valve to either open fully or to close. An actuator with an air-to-open unit fails closed and an air-to-close unit fails fully open. For example, consider the valve on the cooling water to an exothermic reactor. Obviously, an air-to-close actuator is selected so that the loss of instrument air pressure opens the value and, therefore, cools down the reactor preventing a thermal runaway.

Example 2.8 Valve Actuator Selection

Problem Statement. Consider the control valve on the steam for the endothermic CSTR shown in Figure 2.3.16. Determine whether an air-to-open (direct-acting) or air-to-close (reverse-acting) valve actuator should be used in this case.

Solution. If this valve fails (e.g., a loss of instrument air pressure), is it better for the valve to fail open or fail closed? In this case, if the valve were to fail open, excessive steam flow to the heat exchanger would result causing a reactor temperature in excess of the normal operating value. On the other hand, if the valve were to fail closed, the reactor would simply cool off, which is not desirable but is preferred in this case. Therefore, an air-to-open actuator should be selected in this case.

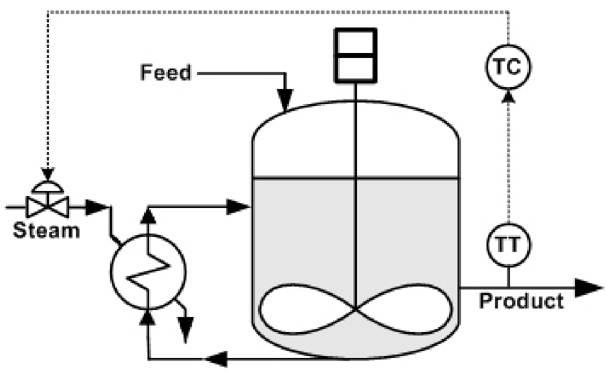


Figure 2.3.16 Control diagram of an endothermic CSTR with a temperature control

Example 2.9 Valve Actuator Selection

Problem Statement. Consider the control valve on the refrigerant for the overhead condenser for the rectifying section of a distillation column. Determine whether an air-to-open or an air-to-close valve actuator should be used for this valve (Figure 2.3.17).

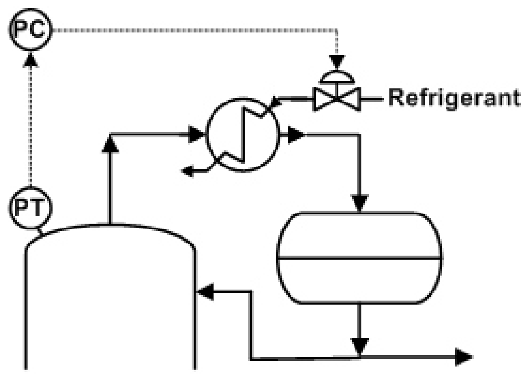


Figure 2.3.17 Rectifying section of a distillation column that uses the flow of a refrigerant to control column pressure.

Solution. If the valve actuator on the refrigerant were an air-to-open actuator and instrument air pressure was lost, the pressure in the column would increase sharply and result in venting of a significant amount of overhead product. On the other hand, if an air-to-close valve actuator were used, during a loss of instrument air pressure, the column would use more refrigerant than necessary, but venting from the column would be avoided. Therefore an air-to-close valve actuator should be selected.

I/P Transmitter. The I/P transmitter is an electro-mechanical device, which converts the 4-20 mA signal from the controller to a 3-15 psig instrument air pressure signal to the valve actuator, which in turn affects the valve stem position.

Figure 2.3.18 Photograph of a globe valve with a pneumatic valve positioner. Courtesy of Fisher-Rosemount.

Valve Positioners. A **valve positioner** is designed to control the valve stem position at a prescribed position in spite of packing friction and other forces on the stem and is usually contained in its own cabinet mounted on the side of the valve actuator (Figure 2.3.18). The valve positioner itself is a controller that compares the measured stem position with the specified stem position and makes adjustments to the instrument air pressure to provide the desired stem position. A diagram of a valve positioner with a general valve position setpoint signal is shown in Figure 2.3.19. The valve position measurement used by a valve positioner is provided by mechanical linkage or an electrical position sensor both of which are connected to the valve stem. The setpoint for a valve positioner can be a pneumatic signal coming from an I/P converter, a 4-20 mA analog signal coming directly from the DCS or an electronic analog controller or a digital signal from a digital controller on a fieldbus network. That is, if a pneumatic input signal is supplied to a valve positioner, the valve positioner is referred to as a pneumatic valve positioner. Similarly, valve positioners that receive an electronic or digital input signals are referred to as electronic or digital valve positioners, respectively.

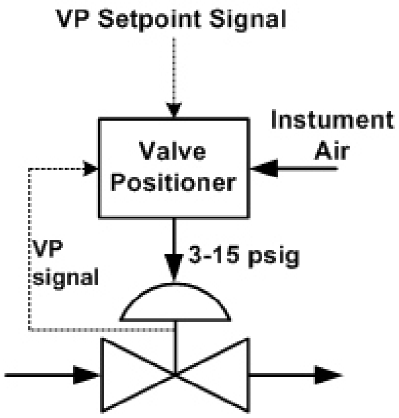


Figure 2.3.19 Schematic of a generalized valve positioner.

Traditional DCSs can use either a pneumatic valve positioner or an electronic valve positioner. Digital positioners, which are used in fieldbus applications, have the advantage that they can be calibrated, tuned and tested remotely, and they can also be equipped with self-tuning capabilities.

Due to the friction from the packing, it is not possible to move the valve stem position to a precise value. As a result, the valve positioner continuously opens and closes the valve bracketing the desired stem position so that the average valve position agrees quite closely with the valve position specified to the valve positioner. The high-frequency feedback provided by the valve positioner can result in precise metering of the **average** flow rate. A valve with a deadband of 25% can provide a repeatability of the flow rate of less than 0.5% for the average flow rate using a valve positioner. Valves with low levels of valve friction can control the average flow rate to a precision approaching 0.1% using a valve positioner. **For flow control loops that are controlled by a DCS, a valve positioner is a necessity because the control interval for a DCS (i.e., 0.5 to 1.0 seconds) is not fast enough for most flow control loops.** A control valve with a positioner is referred to as a **direct-acting final control element** because if the valve position sent to it is increased, the flow rate through the valve increases, regardless of whether an air-to-open or an air-to-close valve actuator is used. If an air-to-open valve actuator is used on a valve without a valve positioner, the final control element is also direct acting. On the other hand, if an air-to-close valve actuator is used without a valve positioner, a **reverse-acting final control element** results because as the pneumatic signal to the valve actuator is increased, the flow through the valve will decrease.

Valve Deadband. An important characteristic of a control valve is the **valve deadband** which is a measure of how precisely a control valve can control the flow rate. The deadband for a steering system on an automobile is the maximum positive and negative turn in the steering wheel that does not result in a noticeable change in direction of the automobile. For a control valve, deadband is the maximum positive or negative change in the signal to a control valve that does not produce a measurable change in the flow rate. Valve deadband is caused by the friction between the valve stem and valve packing and other forces on the valve stem. Typically, industrial control valves without a valve positioner have a deadband of 10 to 25%. That is, for a 25% deadband, a change in the signal to the control valve that is greater than 25% will result in a measured change in the flow rate through the valve. On the other hand, a change that is less than 25% may or may not produce a change in the flow rate. Generally, the larger and older the control valve, the larger the deadband. A properly functioning valve with a valve positioner typically should have a deadband less than 0.5%. Note that deadband is reported in percent and represents the maximum relative change in the signal to the control valve that does not cause a measurable change in the flow rate through the valve. A term related to deadband is hysteresis. Hysteresis is defined as the difference between the valve position on the upstroke (i.e., coming from a lower valve position) and its position on the downstroke (i.e., coming from a higher value position) for a fixed input signal. Hysteresis is most often caused by a high degree of static friction on the value stem, e.g., over tightening of the packing around the value stem.

Booster relays. Booster relays (Figure 2.3.20) are designed to provide extra flow capacity for the instrument air system, which decreases the **dynamic response time** of the control valve (i.e., the time for most of a change to occur). Booster relays are used on valve actuators for large valves that require a large volume of instrument air to

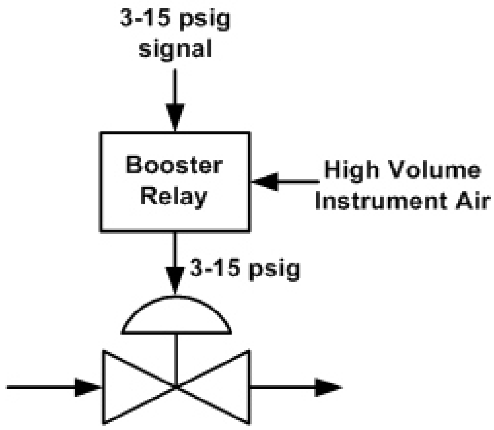


Figure 2.3.20 Schematic of a booster relay.

move the valve stem. Booster relays use the pneumatic signal as input and adjust the pressure of a high flow rate capacity instrument air system that provides air pressure directly to the diaphragm of the valve actuator.

Adjustable Speed Pumps. Adjustable speed pumps are used instead of the control valve systems just discussed for specific applications. A centrifugal pump directly driven by a variable speed electric motor is the most commonly used form of adjustable speed pump in the CPI. For a variable speed motor, an rpm sensor is linked to a speed controller to set the power delivered to the electric motor. Another type of adjustable speed pump is based on using a variable speed electric motor combined with a positive displacement pump based on a piston or a diaphragm. Adjustable speed pumps have the following advantages compared with control valve-based actuators: 1) they use less energy, 2) they provide fast, accurate flow metering without additional device requirements and 3) they do not require an instrument air system. Their major disadvantages are higher capital cost, particularly for large flow rate applications, and lower reliability compared to control valves. Another disadvantage of adjustable speed pumps is that they do not fail open or closed like a control valve with an air-to-close or air-to-open valve actuator, respectively. As a result, the CPI almost exclusively use control valve-based actuators except for low flow applications, such as catalyst addition systems or base injection pumps for wastewater neutralization, which typically use adjustable speed pumps.

Positive displacement adjustable speed pumps are commonly used as an actuator for bio-processing applications. Because it is easier to maintain sterile conditions using a positive displacement variable speed pump and relatively low flow rates are usually associated with bio-processes, positive displacement adjustable speed pumps are preferred over control valves for most bio-processes. In addition, shear-rate limitations are usually placed on the streams in many bio-processes which also favors adjustable speed pumps. High shear rates can damage microbial cells or denature protein products produced in a bio-process.

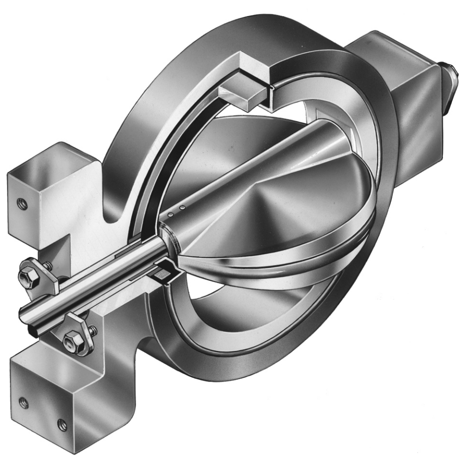


Figure 2.3.21 Partial cutaway view of a butterfly valve. Courtesy of Fisher-Rosemount

Butterfly Valves. Figure 2.3.21 shows a cutaway drawing of a butterfly valve. A disk is attached to a shaft so that, as the shaft is rotated, the restriction to flow is changed. Butterfly valves typically have a installed valve characteristic somewhere between a linear and quick-opening valve. Butterfly valves are flanged into a line and have a motor and positioner attached to move the disk to a specified orientation (i.e., a rotary actuator). Butterfly valves are much less expensive than globe valves, but they usually have a range of accurate flow metering that is about half that of a globe valve. Butterfly valves become economically attractive as control valves for applications with pipe diameters above 6 inches[6].

Self-Assessment Questions

Q2.3.1 What hardware comprises the final control element?

Q2.3.2 Why are globe valves generally used for flow control applications in the CPI?

Q2.3.3 What is the difference between inherent and installed valve characteristics?

Q2.3.4 What determines whether a globe valve is linear, equal percentage, or quick-opening?

Q2.3.5 Why is it important for the installed valve characteristics to be relatively linear over a wide operating range? What happens when they are not sufficiently linear?

Q2.3.6 What is the most commonly used actuator in the biotechnology industries and why is it used so frequently?

Self-Assessment Answers

Q2.3.1 The final control element is made up of the I/P converter, the instrument air system, and the control valve (actuator and valve body assembly). The control valve can also be equipped with a valve positioner or a booster relay.

Q2.3.2 Globe valves are generally used for flow control applications in the CPI because they are the cheapest device available that provides a large range of operation with a linear installed flow characteristic and a small deadband.

Q2.3.3 Inherent valve characteristics are the flow rate versus stem position relationship for a valve with a constant pressure drop across the valve. Installed valve characteristic is the flow rate versus stem position relationship for a valve installed in actual service which generally does not provide a constant pressure drop across the valve, i.e., the pressure drop across the valve tends to decrease with an increase in flow rate.

Q2.3.4 The manner in which the area of the restriction between the valve plug and the value cage and seat changes with stem position determines whether a globe valve is linear, equal percentage or quick-opening.

Q2.3.5 If the installed valve characteristics for a control valve are not sufficiently linear, the range of effective flow control will be limited. The range for effective flow control for a valve is the range over which the slope of the installed valve characteristics changes by less than a factor of 4. When the slope of the installed valve characteristic is too large, flow control will cycle or become unstable. When the slope is too small, the flow control will be sluggish.

Q2.3.6 For bio-processes, variable speed pumps or blowers are preferred over control valves as actuator systems because they are easier to maintain sterile conditions and because of the relatively low processing rates, the costs of variable speed pumps and blowers is not prohibitive.

Self-Assessment Problems

P2.3.1 Calculate the flow rate through a linear control valve with a C_v^{max} equal to 83.3 that has a valve position of 75% open when the pressure drop across the valve is 55 psi. Assume that the liquid flowing through the control valve has a density equal to 55 lb/ft^3.

P2.3.2 Size a control valve for a maximum flow rate of 90 gpm and a minimum flow rate of 30 gpm of water using Tables 2.1 and Figure 2.3.8.

Self-Assessment Answers

P2.3.1 First, C_v is calculated from C_v^{max} by $C_v \quad x C_v^{max} \quad 0.75 \quad 83.3 \quad 62.5$. Then, Equation 2.3.3 can be applied directly recognizing that K=1 for the units used in this problem.

$$Q_f \quad K\,C_v(x)\sqrt{\frac{P}{s.g.}} \quad 62.5\sqrt{\frac{55\,\text{psi}}{55\ \text{lb/ft}^3}\left|\frac{62.4\ \text{lb/ft}^3}{}\right.} \quad 493.5\ \text{gpm}$$

P2.3.2 From Figure 2.3.8, the pressure drop is equal to approximately 20 psi for both 30 and 90 gpm. Solving for $C_v(x)$ from Equation 2.3.3 yields

$$C_v(x) \quad \frac{Q_f}{K}\sqrt{\frac{s.g.}{P}} \quad 30\ \text{gpm}\sqrt{\frac{1}{20\,\text{psi}}} \quad 6.71; \quad C_v(x) \quad 90\text{gpm}\sqrt{\frac{1}{20\,\text{psi}}} \quad 20.1$$

Using Table 2.1, a 1.5-inch valve is too small for this application because it is almost 90% open at 90 gpm. On the other hand, a 2-inch valve is approximately 60% open at 90 gpm and approximately 40% open at 30 gpm. Therefore, a 2-inch valve should be selected.

2.4 Sensor Systems

Sensor systems are composed of the sensor, transmitter and sampling device that samples a process stream or brings the sensor in contact with the process. The sensor measures certain quantities (e.g., voltage, current or resistance) associated with an instrument in contact with the process such that the measured quantities correlate strongly with the variable value that is to be measured. There are two general classifications for sensors: continuous measurements and discrete measurements. Pressure, temperature, level, ion-specific electrodes and flow sensors typically yield continuous measurements while certain composition analyzers (e.g., gas chromatography) provide discrete-time measurements based on withdrawing and analyzing samples periodically.

Several terms are used to characterize the application or performance of a sensor:

- **Zero** is the lowest reading available from the sensor/transmitter, i.e., the sensor reading corresponding to a transmitter output of 4 mA. For example, for Figure 2.4.1 the zero is 100 psia.
- **Span** is the difference between the largest measurement value made by the sensor/transmitter and the smallest. For example, in Figure 2.4.1 the span is 200 psi.
- **Range** is the maximum and minimum sensor reading. For example in Figure 2.4.1, the range of the pressure sensor is 100 psia to 300 psia.
- **Accuracy** is the difference between the value of the measured variable indicated by the sensor and its true value (Figure 2.4.2). The true value is never known; therefore, accuracy is usually estimated by the difference between the sensor value and an accepted standard.

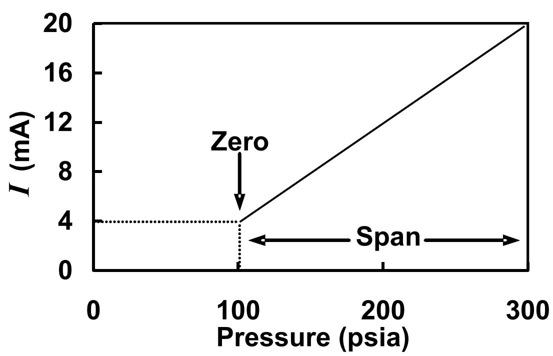
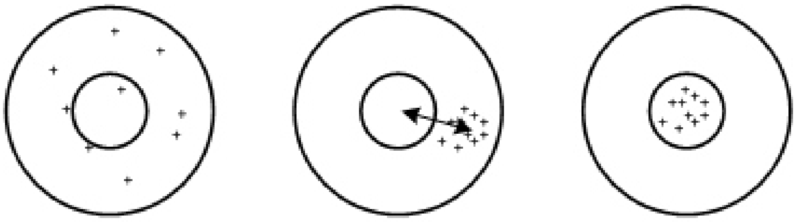


Figure 2.4.1 An example of the linear correspondence between a pressure reading and its analog signal.

- **Bias or systematic error** is the difference between the average of a number of sensor readings and the true valve. See Figure 2.4.2b for a graphical example.
- **Repeatability** is related to the difference between the sensor readings while the process conditions remain constant (Figure 2.4.2). That is, the repeatability is the variation is the sensor reading due to sensor noise.
- **Process measurement dynamics (sensor dynamics)** indicate how quickly the sensor responds to actual changes in the value of the measured variable. Sensor dynamics is normally described by an effective time constant for the sensor and it is usually assumed that four time constants is equal to the time interval necessary for the sensor to respond to a change in the true value of the CV. In other words, the time constant of a sensor is roughly equal to ¼ of the response time of the sensor.
- **Rangeability** is the ratio of the largest accurate sensor reading to the smallest accurate reading.



(a) (b) (c)

Figure 2.4.2 Targets that demonstrate the difference between accuracy and repeatability. (a) Neither accurate nor repeatable. (b) Repeatable but not accurate. Note the arrow indicates the bias error. (c) Accurate and repeatable.

- **Calibration** involves the adjustment of the correlation between the sensor output and the predicted measurement so that the sensor reading agrees with a standard. Most plants calibrate their temperature and pressure sensors once every few months to a year or two depending on the application. Important composition analyzers are many times calibrated every day.

Example 2.10 Sensor Signals

Problem Statement. Determine the temperature reading corresponding to a 10 mA analog signal from a temperature transmitter that has a span of 200ºC and a zero equal to 20ºC.

Solution. The position between the maximum and minimum analog signal is given by

$$\frac{10\,\text{mA} \quad 4\,\text{mA}}{20\,\text{mA} \quad 4\,\text{mA}} \quad 0.375$$

This corresponds to a 37.5% position on the temperature span of the transmitter or 0.375 × 200ºC , i.e., 75ºC. Then the temperature sensor reading is simply 75ºC plus the temperature of the zero (20ºC) for a sensor reading of 95ºC. That is,

$$T \quad 20\ \text{C} \quad 0.375 \quad 200\ \text{C} \quad 95\ \text{C}$$

Example 2.11 Sensor Signals

Problem Statement. Determine the value of the electric analog reading in mA for an actual level of 45% for a level transmitter that has a span of 40% and a zero of 20%.

Solution. A 45% level measurement represents a position between the maximum level measurement (60%) and the minimum level measurement (20%) of

$$\text{Position in span} \quad \frac{45\% \quad 20\%}{60\% \quad 20\%} \quad 0.625$$

Then, the electric analog signal is given by

4 mA 0.625 20 mA 4 mA 14 mA

Overview of Sensor Systems. A wide variety of sensors is available for measuring process variables[8,9]. Choosing the correct sensor for a particular application depends on the measured variable, the properties of the process, accuracy and repeatability requirements and costs, both initial and maintenance. Best practice[10] for instrument selection, for instrument installation and to reduce maintenance costs has been identified for the CPI. The remainder of this chapter will analyze the control-relevant issues associated with some of the most commonly used sensors for feedback control in the CPI and the bio-processing industries.

Smart sensors. **Smart sensors** have built-in microprocessor-based diagnostics. For example, some smart pH sensors are able to identify the buildup of coatings on the electrode surface and trigger a wash cycle to reduce the effect of these coatings. In addition, smart temperature sensors are many times able to predict the imminent failure of the sensor in adequate time to allow for its replacement before failure occurs. Also, smart orifice flow sensors are able to identify plugged lines or partially plugged lines, which can significantly affect the accuracy of these flow sensors. In general, smart sensors are moderately more expensive than conventional sensors but, when they are properly selected and implemented, smart sensors can be an excellent investment due to greater sensor reliability and reduced maintenance. The full utilization of smart sensors is best accomplished with certain communication protocols (e.g., Foundation Fieldbus, Profibus).

Temperature Measurements. The three primary temperature sensing devices used in the CPI and bio-processing industries for feedback control are **thermocouples (TCs)**, **resistance temperature detectors (RTDs)** and **thermistors**. Filled thermometers and bimetallic thermometers are used for field-mounted temperature indicators. Field-mounted sensors are located on the process equipment for operators to read, but do not have transmitters, and as a result, are not connected to the control computer.

Thermocouples. Thermocouples are based on the phenomena that two metal junctions at different temperatures generate a voltage and the magnitude of the voltage is proportional to the temperature difference between the two junctions. Thermocouples (Figure 2.4.3) are constructed of two different types of metal wire that are welded to each other at both ends (i.e., the junctions). The cold junction of a thermocouple is normally at ambient temperature, but is electrically compensated so that it behaves as if it were at a constant temperature. The hot junction is used to measure the process temperature of interest. In general, the voltage generated by the hot junction, which is located inside a thermowell in contact with a process fluid, varies quite linearly with the process temperature. The voltage generated by a thermocouple (e.g., V in Figure 2.4.3) is normally converted to a 4-20 mA signal by its transmitter. Thermocouples are constructed of metal pairs including iron-constantan, copper-constantan, chromel-alumel, and platinum-rhodium. The latter is the most popular material of construction and results in the most accurate thermocouples. (Alumel, chromel and constantan are trade names for alloys that are used to make these thermocouples.) High temperature thermocouple that work up to 2000ºC are constructed of noble metals, such as tungsten, rhenium,

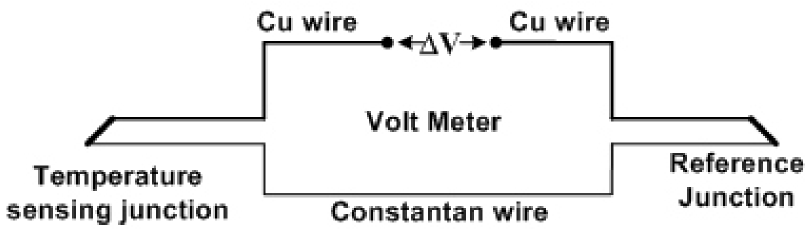


Figure 2.4.3 Circuit drawing for a Cu/Constantan TC.

molybdenum and tantalum and their alloys. Table 2.3 list the repeatability (ºC) for eight types of thermocouples. Note that both Type R and S have relatively low repeatibilities (i.e., ºC) from 0ºC to 1000ºC, which covers most of the normal process operating range. Also, the repeatability in temperature measurement increases for very low temperatures (-200ºC) and for very high temperatures (1600ºC).

Table 2.3 Repeatability (±ºC) for Different Thermocouple Types

	Thermocouple Type (Source National Instruments: http://digital.ni.com)							
Temperature (ºC)	**B**	**E**	**J**	**K**	**N**	**R**	**S**	**T**
-200				3.0	3.0			3.0
-100				2.5	2.5			1.5
0		1.7	1.5	1.5	1.5	1.0	1.0	0.5
200		1.7	1.5	1.5	1.5	1.0	1.0	0.8
400		2.0	1.6	1.6	1.6	1.0	1.0	
600	1.5	3.0	2.4	2.4	2.4	1.0	1.0	
800	2.0	4.0		3.2	3.2	1.0	1.0	
1000	2.5			4.0	4.0	1.0	1.0	
1200	3.0			9.0	9.0	1.3	1.3	
1400	3.5					1.9	1.9	
1600	4.0					2.5	2.5	

RTDs. RTDs are based on the observation that the resistance of certain metals depends strongly on their temperature. A Wheatstone bridge or other type of resistance measuring bridge can be used to measure the resistance of the RTD element and thus estimate the process temperature. Platinum is the most commonly used metal for RTD elements and has a wide useful range (i.e., -200 C to 800 C) while nickel has a more limited operating range (-80 C to 320 C) but is less expensive than platinum. Tungsten is used for high temperature applications because it can be used up to 2700ºC. Each of these metals has a known temperature dependence for its resistance; therefore, calibration requires only applying the RTD to a known temperature condition. Unlike TCs, RTDs require a separate power supply. Also, RTDs have the most linear signal vs. temperature profile of any electronic temperature sensor. This helps maximize practical usable range of the sensor and simplifies calibration.

Figure 2.4.4 Photograph of a thermowell and transmitter housing (left) and several TCs. Courtesy of Fisher-Rosemount.

Thermistors. Thermistors use the resistance of a solid state semiconductor to measure temperature. Thermistors use semiconductors that have a negative coefficient of resistance, i.e., its resistance decreases as the temperature increases. In general, thermistors have a more nonlinear relationship between resistance and temperature than RTDs. As a result, the useful operating temperature range for thermistors is generally smaller than for RTDs. When applied properly to a limited temperature range, thermistors are as accurate as RTDs, but they are much less expensive. Because a thermistor has a much larger sensitivity (change in resistance for a change in temperature), the hardware necessary to make the resistance measurement is less expensive for a thermistor than an RTD.

Thermowells. Thermowells are typically cylindrical metal tubes that are capped on one end and protrude into a process line or vessel to bring a TC, RTD or thermistor in thermal contact with the process fluid (Figure 2.4.4). Thermowells provide a rugged, corrosion resistant barrier between the process fluid and the sensor that allows for removal of the sensor while the process is still in operation. Use of thermowells (which typically contain a RTD) is a standard practice in measuring temperature in bioprocesses such as fermentations. This is because the thermowell provides a smooth easily cleaned and sterilized surface (without cracks, crevices, or pockets) when in contact with fermentor broth and so minimizes the possibility that the temperature well and probe can be a source of contamination. Thermowells that are coated with polymer or other adhering material can significantly increase the lag associated with the temperature measurement, i.e., significantly increase the response time of the sensor. Figure 2.4.4 shows a typical thermowell and housing as well as several thermocouples.

Overall comparison of TCs, RTDs and thermistors. TCs are the least expensive and most rugged compared to RTDs and thermistors, but TCs have an order of magnitude larger repeatability than RTDs and thermistors. Typically, RTDs or thermistors should be used for important temperature control points, such as for reactors temperatures and tray temperatures for distillation columns.

Repeatability, accuracy and dynamic response. TCs typically have a repeatability of approximately C while RTDs and properly implemented thermistors have a repeatability of approximately 0.1 C. Accuracy is a much more complex issue. Errors in the temperature reading can result from heat loss along the length of the thermowell, electronic error, sensor error, error from nonlinearity, calibration errors and other sources[11]. The dynamic response time of a TC, RTD or thermistor sensor within a thermowell can vary over a wide range and is a function of the type of process fluid (i.e., gas or liquid), the fluid velocity past the thermowell, the separation between the sensor and inside wall of the thermowell and the material filling the thermowell (e.g., air or oil). Typical well-designed applications result in time constants of 6-20 seconds for measuring the temperature of most fluids.

Optical pyrometers. Optical pyrometers are used to measure high temperatures (up to 4000ºC) without requiring thermal contact with the temperature point. Optical pyrometers estimate the temperature from the wave length for the radiation of a hot body (e.g., furnace tubes). The repeatability of optimal pyrometers is a strong function of temperature. For temperature about 1000ºC, the repeatability is about 1ºC, for 2000ºC about 3ºC and for 3000ºC about 6ºC [12].

Pressure and Differential Pressure Measurements. Process pressure measurements used for feedback control are usually made using a differential pressure sensor with the low pressure tap exposed to atmospheric conditions. Bourdon tube-type and bellows pressure gauges are typically used for field-mounted pressure measurements but are not used for feedback control. Differential pressure sensors are also used for measuring flow rates by measuring the pressure drop across an orifice plate in a line. Two types of differential pressure sensor are used in the CPI: the balanced bar design and the strain gauge design.

Figure 2.4.5 Photograph of a DP cell. Courtesy of Fisher-Rosemount.

Balanced bar design. The balanced bar design is the most commonly used differential pressure sensor in the CPI. A differential pressure cell (**DP cell,** Figure 2.4.5) uses a balance bar that is deflected based on the pressure differential between two compartments that are in contact with opposite sides of the balance bar. A precision forcing motor is used to maintain the balance bar at a specific position. The measurement of the pressure is directly related to the force used by the forcing motor to balance the bar. In Figure 2.4.5, the transmitter and associated electronics are mounted on the top of the compartment that houses the balanced bar.

Strain gauge design. The strain gauge design uses a strain gauge to measure the pressure difference between two compartments that act on opposite sides of a diaphragm. Strain gauges are based upon the property that when a wire is stretched elastically, its length increases while its diameter decreases, both of which increase the resistance of the wire. Serpentine lengths of elastic resistance wires can be bonded to the surface of an elastic element (diaphragm). When an increase in the deformation of the diaphragm occurs as the result of a differential pressure increase, the wires elongate and, therefore, the resistance of these wires increases, indicating an increased differential pressure reading. The resistance of the strain gauge is usually measured using a Wheatstone bridge. Both the balanced bar design and the strain gauge design of differential pressure sensors are very fast responding (i.e., with time constants less than 0.2 s). Repeatability for these differential pressure sensors is varies between 0.1% to ±1.0%, depending on the magnitude of the differential pressure being measured.

Flow Measurements. Flow measurements are used as part of flow control loops applied to MVs (manipulated variables) and are used to measure disturbances, e.g., a feed flow rate change to a distillation column. The most commonly used flow sensor in the CPI is an orifice meter, which uses the pressure drop across an orifice plate flanged into a line. Orifice meters are simple and inexpensive from a capital cost point-of-view, but they have high maintenance costs associated with them because the pressure taps on the DP cell tend to plug easily. On the other hand, magnetic flow meters and vortex shedding meters have much higher capital costs, but have much lower maintenance costs because they do not use pressure taps. **Flow meters**, whichever type is chosen, **are typically installed upstream of the control valve to provide the most accurate, lowest noise measurement** to avoid the effects of flashing and/or nonuniform flow caused by the throttling of the control valve. Installing the flow sensor downstream of a control valve subjects the sensor to flow fluctuations and even two-phase flow, which reduce the sensor accuracy and increase the measurement noise.

Orifice meter. An orifice meter uses the measured pressure drop across a fixed-area flow restriction (an orifice) to predict the flow rate. An example of a paddle type orifice plate is shown in Figure 2.4.6. The pressure drop across an orifice is usually measured using a **DP cell** (Figure 2.4.5). The volumetric flow rate, F_V, through an orifice is related to the pressure drop across an orifice plate, P, by the following equation:

Figure 2.4.6 Photograph of an paddle-type orifice plate. Courtesy of Thermocouple Instruments,

$$F_V \quad \frac{C_d A_2}{\sqrt{1 \quad (A_2/A_1)^2}} \sqrt{\frac{2g_c \quad P}{}} \qquad \textbf{2.4.1}$$

where A_1 is the pipe cross-sectional area, A_2 is the cross-sectional area of the orifice, is the density of the fluid, g_c is a unit conversion factor (32.2 lb_m-ft/lb_f-s^2) and C_d is the discharge coefficient. C_d is dimensionless and is a function of the Reynolds number and the type of fluid, but is constant at a value of approximately 0.6. **A straight run of pipe preceding the orifice meter is required to develop uniform flow at the orifice meter.** If not, an error as large as 15% can result in the measured flow rate. Because the orifice meter is based upon a measured pressure drop, it is a very fast-responding measurement. Orifice meters typically provide a repeatability in the range of ±0.3 to 1%. For differential pressure-based flow sensors, **it is important that both pressure taps on the process line are at the same elevation** otherwise a systematic error in the differential pressure reading will result.

Example 2.12 Flow Rate through an Orifice Plate

Problem Statement. Calculate the flow rate of water through a 1.5-inch diameter orifice in a Schedule 40 3-inch line if the pressure drop across the orifice is 5 psi.

Solution. The inside diameter of a Schedule 40 3-inch line is 3.068 inches. Then A_2 is calculated to be 1.767 in^2 and A_1 is 7.393 in^2. The density of water is taken as 62.4 lb_m/ft^3. Equation 2.4.1 can be applied directly to solve this problem, but care should be taken with the associated unit conversions. For example, g_c (32.2 lb_m-ft/lb_f-s^2) and the conversion of in^2 to ft^2 is required to cancel the units. Applying Equation 2.4.1 yields

$$F_V \quad \frac{C_d A_2}{\sqrt{1 \quad (A_2/A_1)^2}} \sqrt{\frac{2g_c \quad P}{}} \quad \frac{(0.6)\,(1.767\ \text{in}^2)\,(\text{ft}^2/144\ \text{in}^2)}{\sqrt{1 \quad (1.767/7.393)^2}} \sqrt{\frac{(2)\,(5\ \text{lb}_f/\text{in}^2)\,(32.2\ \text{lb}_m \quad \text{ft}/\text{lb}_f \quad \text{s}^2)}{(62.4\ \text{lb}_m/\text{ft}^3)\,(\text{ft}^2/144\ \text{in}^2)}}$$

$$0.2067 \quad \text{ft}^3/\text{s} \qquad 92.8\ \text{gpm}$$

Example 2.13 Pressure Drop across an Orifice Plate

Problem Statement. For the system described in Example 2.12, calculate the pressure drop across the orifice for a flow rate of 75 gpm.

Solution. Solving for P by rearranging Equation 2.4.1 yields

$$P \qquad \frac{F_V^2 \quad 1 \quad (A_2 / A_1)^2}{2 g_c \, C_d^2 \, A_2^2}$$

Substituting the numerical values and performing the necessary unit conversions yields

$$P \qquad \frac{(75 \text{ gal / min})^2 \ (\text{min / 60 s})^2 \ (\text{ft}^3 \text{ / 7.481 gal})^2 \ (62.4 \text{ lb}_\text{m} \text{ / ft}^3) \ [1 \quad (1.767 / 7.393)^2]}{(2)\ (32.2 \text{ lb}_\text{m} \quad \text{ft / lb}_\text{f} \quad \text{s}^2)\ (0.6)^2 (1.767 \text{ in}^2)^2 (\text{ft}^2 \text{ / 144 in}^2)}$$

$$P \qquad 3.27 \text{ psi}$$

A simpler method to solve this problem is to apply Equation 2.4.1 twice, i.e., once for Example 2.12 and once for this example, and take the ratio, noting that all the parameters of the problem cancel except P and F_V. Then, the pressure drop for 75 gpm can be calculated directly by

$$P_{75 \text{ gpm}} \qquad P_{92.8 \text{ gpm}} \quad \frac{75 \text{gpm}}{92.8 \text{ gpm}} \quad ^2 \qquad 3.27 \text{ psi}$$

Orifice meter sizing. The sizing of an orifice meter is directly related to the differential pressure sensor that is used. Typically, differential pressure sensors are available in various sizes indicated by the maximum pressure drop, e.g., 1 psi, 2 psi, 5 psi and 10 psi sizes. Sizing an orifice meter involves choosing the range of the DP cell and then calculating , the ratio of the diameter of the opening in the orifice to the inside pipe diameter, while satisfying the following three restrictions: (1) should be greater than 0.2 and less than 0.7 for most orifice designs, (2) the pressure drop measured across the orifice should be less than 4% of the line pressure and (3) the Reynolds number for flow in the pipe must be between 10^4 and 10^7 for normal operation[3]. Because the maximum turndown ratio for a differential pressure sensor is about 9 and because there is a square root relation between flow rate and pressure drop, the maximum turndown ratio an orifice flow meter is about 3. On other hand, an orifice flow meter that uses a smart transmitter can provide a turndown ratio of 10:1.

The recommended procedure for sizing an orifice flow meter is as follows: First, calculate the Reynolds number for flow in the line in question to ensure that it is between 10^4 and 10^7. If this is not the case, the line will require resizing for an accurate orifice flow meter. Next, select the range of the differential pressure sensor. Then assume that at the maximum flow rate the differential pressure reading is 67% of full range[3] (i.e., $0.67\times \ P_{full\,range}$). At this point, the second restriction ($P<0.04P_{line}$) can be evaluated. If this restriction is not satisfied, a differential cell with a smaller pressure range must be selected. Next, Equation 2.4.1 is rearranged to calculate based on the maximum flow conditions (i.e., P and flow rate). Then the first restriction (0.2< <0.7) is evaluated. If is greater than 0.7, a DP cell with larger differential pressure range should be selected to decrease . On the other hand, if is less than 0.2, a DP cell with a smaller differential pressure range should be used. Finally, the turndown ratio, which is the ratio of the maximum to minimum flow rates, is evaluated.

Example 2.14 Sizing an Orifice Meter

Problem Statement. Size an orifice meter for service on a 2½-inch Schedule 40 pipe carrying water with a maximum flow rate of 180 gpm and a minimum flow rate of 60 gpm. Assume that the line pressure is 150 psig.

Solution. The Reynolds number for flow in the pipe for this case is 1.9×10^5, which is within the specified range and satisfies the third requirement. Assume that a differential pressure sensor with a 2 psi maximum is used for this application. For the maximum flow, it is assumed that the resulting differential pressure is 1.33 psi (67% of full reading). This result corresponds to 0.9% of the line pressure, which satisfies the second requirement. Recognizing that $D_2 = \ D_1$ and rearranging Equation 2.4.1 to solve for yields

$$\sqrt[4]{\frac{F_V^2}{F_V^2 \quad 2C_d^2A_1^2g_c \quad P/}} \qquad \frac{1}{\sqrt[4]{1 \quad \frac{2C_d^2A_1^2g_c \quad P}{F_V^2}}}$$

This equation, with the diameter of 2½-inch Schedule 40 ferrous pipe (2.469 in), a pressure drop of 1.33 psi, C_d equal to 0.61 and a flow rate of 180 gpm yields a of 0.742. Because this is greater than 0.7, this is not an acceptable design for an orifice meter. Because a larger pressure drop across the orifice is required, the next largest size differential pressure sensor should be used, i.e., a 5 psi differential sensor. For the maximum flow, it is now assumed that the resulting differential pressure is 3.33 psi (67% of full reading). This result corresponds to 2.2% of the line pressure, which satisfies the second restriction. Using the previous equation for a pressure drop of 3.33 psi at a flow rate of 180 gpm yields a of 0.573. The turndown ratio for the flow is 3; therefore, the low flow rate should also be accurately measured. As a result, equal to 0.573 (i.e., an orifice bore equal to 1.41 inches) with a 5 psi differential pressure sensor is a viable orifice meter design for this application.

Figure 2.4.7 Photograph of a vortex shedding meter. Courtesy of Yokogawa

Vortex shedding flow meters. Vortex shedding flow meters (Figure 2.4.7) are based on inserting an unstreamlined obstruction (i.e., a blunt object) inside the pipe and measuring the frequency of downstream pulses created by the flow past the obstruction. The flow rate is directly related to the frequency of the pulses. Vortex shedding meters are recommended for clean, low viscosity liquids and gases and can typically provide a rangeability of about 15:1. Care should be taken to ensure that (1) cavitation does not occur in the measuring zone and (2) the velocity does not become less than its lower velocity limit. Vortex shedding meters are usually accurate at Reynolds number greater than 10,000, which sets the lower velocity limit. Vapor bubbles resulting from cavitation increase the noise and decrease the accuracy of the measurement;

therefore, care should be taken to ensure that the pressure in the line remains above the cavitation limit.

Figure 2.4.8 Photograph of a magnetic flow meter. Courtesy of Sparling, Inc.

Magnetic flow meters. Magnetic flow meters (Figure 2.4.8), which are low-pressure drop devices, can be used to measure the flow rate of electrically conducting fluids. The conductivity of typical tap water, which is not particularly conductive, is sufficient to use a magnetic flow meter. Magnetic flow meters are based on the principle that a voltage is generated by an electronically conducting fluid flowing through a magnetic field. The magnetic flow meter creates a magnetic field using an electromagnet and measures the resulting voltage, which is proportional to the flow rate in the pipe. Magnetic flow meters provide accurate flow measurements over a wide range of flow rates and are especially accurate at low flow rates. Deposition on the electrodes is a limitation of magnetic flow meters in certain applications. Typical applications of magnetic flow meters are for metering the flow rates of viscous fluids, slurries and highly corrosive chemicals[6]. Magnetic flow meters are used extensively in water treatment facilities and in municipal drinking water systems. The low flow rate velocity limit for magnetic flow meters is about 1 ft/s for water, but the low flow rate limit is higher for more viscous fluids.

Level Measurements. The most common type of level measurement is based upon measuring the hydrostatic head in a vessel using a DP cell. This approach typically works well as long as there is a large difference between the density of the light and heavy phases. Because it is based on a pressure measurement, this approach usually has relatively fast measurement dynamics. If the pressure tap connections between the process and the DP cell become partially blocked, the dynamic response time of the sensor can be significantly increased, resulting in slow-responding level measurements. Level measurements typically have a repeatability of approximately 1%.

Figure 2.4.9 shows how a differential pressure measurement can be used to determine the level in a vessel. This approach directly measures the hydrostatic head in the vessel. Because of plugging and corrosion problems, it may be necessary to keep the process fluid from entering the DP cell. In addition, it is important to keep vapor from condensing in the upper tap and collecting in the low pressure side of the DP cell. This can be accomplished by insulating the pressure tap and wrapping it with resistive heating tape or installing steam tracing. Use of differential pressure to measure liquid level in a vessel can be affected by an agitator stirring the contents of the vessel. That is, the motion of fluids near the bottom tap of the differential pressure measurement can cause dynamic fluid flow forces to be present, affecting the pressure at the bottom tap. Differential pressure can still often be used, but a correction to the differential pressure measurement to account for mixing effects may be needed. Nuclear level gauges, ultrasonic level sensors and electrical level sensors are used in special cases. Float activated devices, which are similar to the level measuring approach used in the water reservoir in toilets, are sometimes used in the CPI.

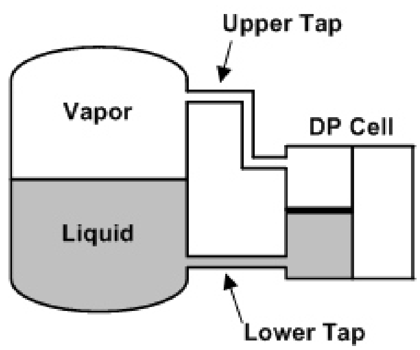


Figure 2.4.9 Schematic of a typical differential pressure level application.

Other types of level sensors are used in special situations. Nuclear level sensors are based on the difference between the gamma ray absorption of the liquid and the vapor space above it. Radar level gauages, which are used for large diameter storage tanks, rely on the reflection of an electromagnetic pulse from the liquid level and the measurement of the travel time of the reflected pulse.

Chemical Composition Analyzers. The most commonly used on-line composition analyzer in the CPI is the **gas chromatograph (GC)** while inroads have been recently made by infrared analyzers and ultraviolet and visible-radiation analyzers. On-line composition measurements are generally much more expensive and less reliable than temperature, pressure, flow rate and level measurements. Due to their large associated cost, the decision to use an on-line composition analyzer is normally based on process economics. For example, for refineries and high volume chemical intermediate plants, on-line analyzers (usually GCs) are used extensively because 1) due to the large flow rates used in these large plants, process improvement due to on-line composition analysis easily economically justifies the application and 2) the measurement techniques are generally well established for this industry. On the other hand, for the specialty chemicals industry, much less use of on-line analyzers is made due to 1) lower production rates and 2) unavailability of reliable analyzers for the complex compounds that are typical to this segment of the CPI.

Gas chromatograph. GCs process a vapor sample, along with a carrier gas, through a small diameter (1/8-3/8 in) packed column. As a result of different affinities of the sample components for the column packing, the various sample components have different residence times in the packed column. As each component emerges from the column, it passes through a detector. The most commonly used detectors are thermal conductivity detectors and hydrogen-flame ionization detectors. Hydrogen-flame ionization detectors are more complicated than thermal conductivity detectors but are much more sensitive for hydrocarbons and organic compounds. Repeatability for GCs can vary over a wide range and is dependent on the particular system being measured. New analyzer readings are typically updated every 3 to 10 minutes for GCs.

Infrared, ultraviolet and visible radiation. These analyzers are based on the property that each compound absorbs specific frequencies of radiation and the greater the concentration, the higher the degree of absorption. To identify a component from among several components, only the absorption frequencies of the component of interest are required.

Sampling systems. The sampling system is responsible for collecting a representative sample of a process stream and delivering it to the analyzer for analysis. Obviously, the reliability of the sampling system directly affects the reliability of the overall composition analysis system. The transport delay associated with the sampling system contributes directly to the overall deadtime for an on-line composition measurement. For example, an improperly designed sampling system can result in a transport time of one hour for the sample to travel from the process to the analyzer while a properly designed system can result in a transport delay of 10 seconds or less. This difference in sampling deadtime can have a dramatic effect on the performance of the control loop that uses the measurement.

Mass spectrometer gas analyzers: Mass spectrometers are used in the CPI for the determination of dryer operation end-point and stack gas monitoring. These have been called near "universal gas analyzers" as they can measure about any volatile compound present in gas streams, from ppm levels to 100 %, with high accuracy. A single mass spectrometer can be interfaced to multiple sources of gases, as an analysis of a particular gas normally only takes a few seconds. Automated calibration checks are also easily performed. There are different technologies used with process mass spectrometers (e.g., magnetic sector, quadrupole), with magnetic sector types having a stability and accuracy advantage. Further discussion of the use of mass spectrometers in fermentation applications is provided in the section on sensors in common use in bioprocesses.

Sensors in Common Use in Bioprocesses. Figure 2.4.10 shows some typical bio-sensors that are used for continuous control and periodic monitoring of bio-processes. For a continuous feedback control example, a temperature sensor (TT) can be used to control the amount of heating or cooling necessary to maintain a bio-reactor at a preset operating temperature. For a periodic process monitoring example, in certain applications the power consumption of the mixer motor, which varies directly with the viscosity of the broth, can be observed to identify when a bio-reaction is complete. In this manner, some sensors are used for continuous feedback control while others are used by the operator to periodically monitor the performance of the process. A sensor that has a transmitter (e.g., an RTD) can be used for either continuous feedback control or for process monitoring while a sensor not equipped with a transmitter (e.g., rotameter) can only be used for process monitoring.

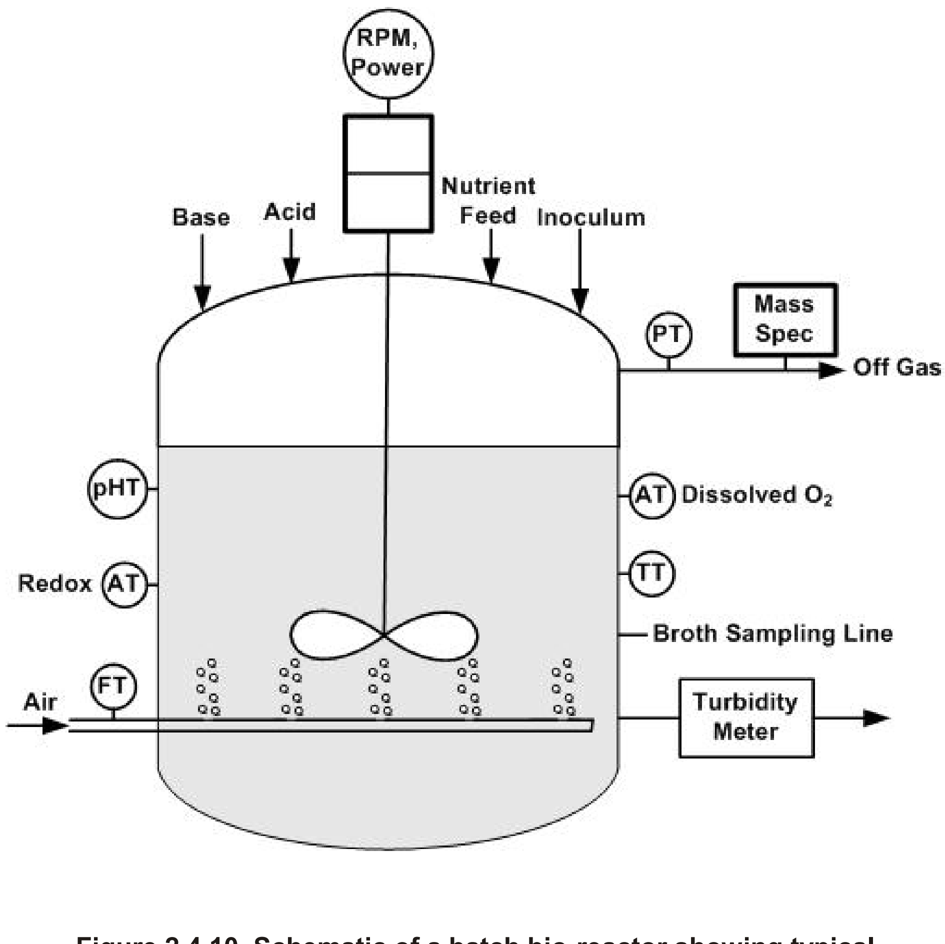


Figure 2.4.10 Schematic of a batch bio-reactor showing typical instrumentation.

One of the key issues associated with bio-sensors is that many of them must be sterilized before use in a bio-process. Otherwise, unwanted microbial species can grow in the bio-process which will usually contaminate the products. Certain bio-processes, such as beer making, are not as susceptible to contamination as bio-pharmaceutical processes, for example. To sterilize a bio-process, it is typical to wash the process, including sensors, and then maintain the internal portions of the process at an elevated temperature for period of time. The temperature and the duration of the sterilization process dictate the death rate of the undesirable microbial species. The sterilization process may not kill all the unwanted microbial species, but if done properly, it reduces them to the point where there is less than one chance in a million that an unwanted microorganism will survive. It is common in the bio-industries to sterilize a process by maintaining all parts of the process that will come in contact with live culture at 120ºC for 40 minutes. Typically an entire bio-process is sterilized before each batch is started or for a continuous process the process is sterilized before startup. The media (i.e., food for the microbial species) may be sterilized separately either by thermal treatment or by passing it through a 0.1-0.2 micron filter. Typical microorganisms are 1-10 microns in diameter.

Flow measurements. In the bio-tech industries, the most commonly used flow sensor is the Coriolis meter. In general, Coriolis meters are applied with each actuator (e.g., variable speed pump or air compressor) so that accurate flow control can be maintained. Coriolis meters send the process fluid through a U-shaped tube and the resulting angular deflection in the tube is directly proportional to the mass flow rate. Coriolis meters are particularly well suited to the bio-tech industries because they use smooth flow conduits that are relatively easy to clean and maintain sterile and because they have low maintenance requirements and do not require frequent calibrations. In addition, Coriolis meters do not subject the fluid to high shear stresses, which can be important in certain cases, e.g., mammalian cell cultures.

Referring to Figure 2.4.10, air is injected into the bio-reactor using an air compressor followed by a filter that maintains sterile conditions in the bio-reactor. Air enters the bio-reactor through a sparger, which provides uniform contacting between the reaction mixture and the injected air. The flow rate of air, especially for industrial fermentations, is usually measured using a Coriolis meter such that an automated electronic indication of air flow is available in the control computer for automated control of the air (several air flow rate changes are typically required as the cell mass increases during a batch) and for use in several on-line calculated variables (e.g., oxygen uptake). For smaller fermentors (e.g., bench scale), sometimes rotameters (a manual non-electronic device) are used to control airflow. Note that certain process monitoring parameters and automated control functions are not available in real-time when using such devices. Rotameters, which employ a float in a glass housing to indicate the flow rate, are used for monitoring while Coriolis meters can also be used for feedback control. In a similar fashion, Coriolis meters and rotameters are used to measure the flow rate of reactor feed streams.

Turbidity meter. **Turbidity** is the measure of the amount of light scattering caused by suspended solids in a sample. The cell concentration in a sample from a fermentation process has been shown to vary linearly with its turbidity up to a certain cell concentration limit if the reaction broth is clear except for cells. As a result of this latter restriction, most bio-reactors in the bio-tech industries are unable to use a turbidity meter for the on-line measurement of cell concentration. Turbidity meters are continuous, but they are sensitive to gas in the sample and deposits on the measurement cells of the instrument. Even when the reaction broth is clean except for cells, the turbidity measurement is not a measure of live cells. Also, many industrial fermentations contain initially insoluble nutrients (e.g., soy bean meal) that may slowly break down and solubilize during the fermentation but which would, nevertheless, represent an unwanted non-cell contribution to turbidity measurements while existing as insoluble particles.

Ion-specific analyzers. The pH, **redox potential** and **dissolved oxygen (DO)** analyzers each use an electrode to preform their measurements. Figure 2.4.11 is a cross-section of a typical ion-specific electrode. The voltage generated by this electrochemical cell is directly related to the concentration of a specific ion or dissolved species at the measuring electrode. Therefore, electrodes can be designed so that the electromotive force (voltage) generated by the electrode correlates directly to the concentration of an ionic or dissolved species. The use of selective membranes (see Figure 2.4.11) can ensure that only the desired ions come in contact with the measuring electrode, otherwise the measurement may be compromised. Electrode-based sensors provide continuous readings and are fast responding (2-5 s time constants) to process changes when they are clean but become slower responding as their membrane fouls. They are generally quite reliable because they have been used industrially for many years, but they can be susceptible to excessive membrane fouling in certain applications. In addition, if they are located in contact with gas bubbles, noisy readings result. The sterilization of an entire bio-process may result in excessive pressure for an electrode. In that case, removable electrodes or electrodes that are designed for higher pressures can be used. Another potential problem for electrodes is electrical interference. In general, a metal vessel should be grounded, and ideally, the transmitter used to send the electrode measurement signal to the control computer should be electrically isolated.

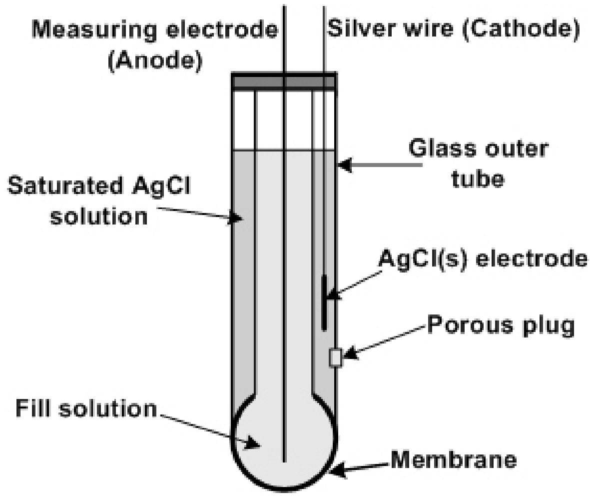


Figure 2.4.11 Cross-section of an electrode with a built-in Ag/AgCl reference electrode.

<u>pH sensors.</u> pH has a significant effect on most bio-processes; therefore, pH is controlled continuously or at least monitored for almost all bio-processes and most processes in the CPI that involve water. pH electrodes are sensitive to temperature changes and may require compensation for temperature in certain applications. pH sensors use a glass membrane.

<u>DO sensors.</u> Many bio-reactors use oxygen as a reactant and, as a result, it may be important to tightly control the DO concentration in the reactor broth. Remember that the equilibrium DO concentration is a function of temperature and pressure in the process. Therefore, to maintain a consistent DO concentration you must first have good temperature and pressure control. DO sensors typically use a Teflon membrane, which is replaced frequently. Fouling of a DO electrode will usually result in a slow-responding DO reading.

<u>Redox sensors.</u> The redox potential is an indication of the equilibrium between the reducing species (i.e., electron donors) and the oxidizing species (i.e., electron receptors) in a fermentation mixture and is a function of temperature. The redox potential gives information about the DO levels at values less than 5% of saturation, which are below the lower measurement limit of a DO electrode. In general, the redox potential is a good indicator of the "health" of an anaerobic or mildly aerobic (i.e., low O_2 concentration) fermentation process. Redox potential electrodes typically use a platinum measurement electrode.

Off-gas analyzers. Mass spectrometers are commonly used to monitor the gas streams entering and leaving a fermentor to measure the concentration of nitrogen, oxygen, carbon dioxide, and other compounds and sending the results to the control computer where this information is combined with other process measurements (e.g., gas flow rates) to perform material balances in which cell mass oxygen uptake, carbon dioxide evolution, gas transfer constants (i.e., k_la) and other calculated variables are determined. This information is updated and available in near real-time so as to be useful for on-line process diagnostics and more complex real-time models (such as estimation of cell mass)[13].

Process mass spectrometers are also sometimes used to monitor other compounds in the exhaust gas steam (e.g., alcohols) that may be of environmental interest or as useful markers in noting certain metabolic activity by the cells. Automatic sampling by process mass spectrometers for fermentation gas analysis is conveniently applied because the pressure difference between the fermentor (i.e., usually operating at a few psig) and the mass spectrometer (i.e., operating at high vacuum) can be used to drive gas samples from the fermentor to the mass spectrometer so that no sample transfer pumps are needed. Also, while process mass spectrometers can be expensive (> $ 100K), analysis is fast enough (i.e., a few seconds per sample) that such instruments are usually interfaced, via a multiplexing valve provided with the instrument, to up to 32 different fermentors. Therefore, the cost per fermentor is relatively low.

In addition, mass spectrometers are very fast (i.e., analyzer deadtime < 3s), highly accurate (i.e., 3-4 significant figures accuracy) and highly reliable. The off-gas from a bio-reactor is often monitored for the concentration of CO_2, O_2, N_2, H_2S and specific organic compounds (e.g., methanol and ethanol) and this data is used to form a software-based sensor for the metabolic rate of the microbial species in the bio-reactor, i.e., a measure of the resources used for cell growth and cell maintenance, using material balances. The CO_2 production rate and the O_2 consumption rate by a bio-reactor are measures of the consumption of the substrate (e.g., glucose) for cell maintenance and for cell growth (Example 13.3). In addition, the ratio of the CO_2 production divided by the O_2 consumption, which is known as the **respiration quotient (RQ)** is typically used to indicate the primary substrate being utilized by the cell culture. For example, for a fermentation utilizing glucose ($C_6H_{12}O_6$) as the primary substrate (i.e., carbon source), it can be seen from the elemental balance (i.e, equal number of C and O atoms) that RQ should be about 1.0. For fermentations utilizing a lipid as the primary substrate, the RQ is about 0.9 or possibly less, depending upon the specific lipid used. For an operation that has only a few bio-reactors, it may be more cost effective to use individual O_2 and CO_2 analyzers instead of a mass spectrometer although they are not as accurate. CO_2 in the off-gas can be measured using a infrared spectrometer after water vapor has been removed from a side stream sample by passing the sample through a column of silica gel. Also, an O_2 electrode can be used to measure O_2 concentration in the off gas.

Some process mass spectrometers have been reported as interfaced with probes containing semipermeable membranes, inserted in the liquid fermentation broth, permitting real-time monitoring of compounds in the liquid of sufficient vapor pressure to transfer to the gas side of the membrane in measurable quantities. This technology is known as MIMS (membrane introduction mass spectrometry).

Fermentation products analyzer. Some quality attributes of the final product or intermediate composition of a fermentation mixture can sometimes be determined using **high pressure liquid chromatography** (**HPLC**) or a **flow injection analyzer** (**FIA**). For both analyzers, a small sample stream is removed from the fermentor or the product stream, passed through a filter and sent to the HPLC or FIA instrument for analysis. HPLC is similar to a GC in that it uses the variation in the mobility of components to separate them from a liquid mixture using a chromatographic column. HPLC is very flexible because columns can be changed to modify the components that are analyzed, but it typically requires 20-30 minutes per analysis. HPLC is used for both online and off-line monitoring of bio-processes. FIA is based on mixing several portions of the sample with colorimetric solutions and measuring the resulting colors of the mixtures. This is accomplished by using several mixing cylinders with solenoid valves to implement the filling and draining actions required. FIA analyzers are not very flexible, but FIA is quite accurate and has a relatively fast analysis cycle time (2-3 min). FIA analyzers are primarily used for monitoring. Filtering the sample and preparing it for injection into a HPLC or FIA analyzer typically requires human intervention and represents the primary obstacle to the use of these analyzer for feedback control applications.

Transmitters. The transmitter (**transducer**) converts the output from the sensor (e.g., a millivolt signal, a differential pressure, a displacement, etc.) into a 4-20 mA analog signal that represents the measured value of the CV. Transmitters are typically designed with two knobs that allow for independent adjustment of the span and the zero of the transmitter. Properly functioning and implemented transmitters are fast and do not normally contribute to the dynamic lag of the process measurement. Modern transmitters have features that, if not applied properly, can reduce the effectiveness of the control loop. For example, excessive filtering (see Section 9.6) of the measurement signal by the transmitter can add extra lag to the feedback loop, thus degrading control loop performance. Table 2.3 summarizes the dynamic characteristics, repeatability and rangeability or turndown ratio of several final control elements and several different types of sensors for the CPI and bio-tech industries.

Self-Assessment Questions

Q2.4.1 What is the bias error of a sensor?

Q2.4.2 What are the two most important differences between TCs and RTDs ?

Q2.4.3 Why are flow measurement devices usually located upstream of the control valve?

Q2.4.4 How are differential pressure sensors used to measure system pressure?

Q2.4.5 Name three composition measurements that are performed using an ion-specific electrode.

Self-Assessment Answers

Q2.4.1 The bias or systematic error is the difference between the average of a number of readings and the true value.

Q2.4.2 Cost and repeatability. RTDs are more expensive than TCs, but have a repeatability that is about ±0.1 ºC compared to ±1.0 ºC for a TC.

Q2.4.3 If the flow indicator is located downstream of the control valve, it will be susceptible to flashing and turbulence caused by the control valve thus affecting its repeatability.

Q2.4.4 Differential pressure sensors are used to measure the pressure of a system by exposing the low pressure tap to the atmosphere, thus measuring the gauge pressure of the system.

Q2.4.5 Dissolved oxygen concentration, pH and redox measurements are each based on using ion-specific electrodes.

Self-Assessment Problems

P2.4.1 Determine the pressure reading corresponding to a 5.5 mA analog signal from a pressure transmitter that has a span of 150 psi and a zero of 25 psig.

P2.4.2 Calculate the flow rate of water through a 1-inch diameter orifice in a line with a 2-inch internal diameter if the pressure drop across the orifice is 1 psi.

Table 2.4
Summary of Control-Relevant Aspects of Actuators and Sensors

	Time Constant (sec)	Valve Deadband or Sensor Repeatability	Turndown Ratio, Rangeability or Range
Control valve *	3 - 15	10 - 25%	9:1
Control valve w/valve positioner*	0.5 - 2	0.1 - 0.5%	9:1
Variable speed positive displacement pumps	<0.1	0.1 - 0.5%	9:1
TC w/ thermowell	6 - 20	1.0 ºC	-200ºC to 1300ºC
RTD w/ thermowell	6 - 20	0.1 ºC	-200ºC to 800ºC
Thermistor w/thermowell	6 - 20	0.1 ºC	
Magnetic flow meter	<1	0.1%	20:1
Vortex shedding meter	<0.1	0.2%	15:1
Orifice flow meter	<0.2	0.3 - 1%	3:1
Orifice meter w/smart transmitter	<0.2	0.3 - 1%	10:1
Differential pressure level indicator	<1	1%	9:1
Pressure sensor	<0.2	0.1%	9:1
pH electrode	2-5	±0.1 pH units	4-8 pH units
DO analyzer	2-5	±0.1-±0.5%	20-70% of saturation
Coriolis meter	<0.1	±0.1-±0.5%	10:1
Turbidity meter	3-10	±2.5%	0-60 g/l

* Based on globe valves.

Self-Assessment Answers

P2.4.1

$$P \quad 25\text{ psig} \quad \frac{5.5\text{ mA} \quad 4\text{ mA}}{20\text{ mA} \quad 4\text{ mA}}(150\text{ psig}) \quad 39.06\text{ psig}$$

P2.4.2 Based on the problem statement, A_1=0.785 in^2 and A_2=3.14 in^2**.** Then, applying the flow equation for an orifice,

$$Q \quad \frac{(0.6)\,(0.785\text{ in}^2)\,(\text{ft}^2/144\text{ in}^2)}{\sqrt{1 \quad (0.5)^2}}\sqrt{\frac{(2)\,(1\text{ lb}_f/\text{in}^2)\,(32.2\text{ lb}_m\text{ ft}/\text{lb}_f\text{ s}^2)}{(62.4\text{ lb}_m/\text{ft}^3)\,(\text{ft}^2/144\text{ in}^2)}} \quad 0.0460\text{ ft}^3/\text{s} = 20.7\text{ gpm}$$

2.5 Summary

- An industrial feedback control loop consists of a controller, an actuator system, a process and a sensor system. The sensor generates an output that is related to the CV and the transmitter converts this reading into a 4-20 mA analog signal. For DCSs, PLCs, or PCs, the A/D converter converts the analog signal into a digital value for the sensor reading. The controller (e.g., DCS) compares the digital sensor reading to the setpoint and calculates the digital value of the controller output. The D/A converter converts this digital reading into a 4-20 mA analog signal which, in turn, is converted to a 3-15 psig instrument air pressure by the I/P converter for a control valve. The instrument air pressure acts on the control valve, which causes the manipulated flow to the process to change. This change to the process, as well as other input changes, causes the value of the CV to change. The sensor reading changes and the control loop is complete.

- The evolution of controllers from pneumatic controllers to analog controllers to computer based controllers and the evolution of network wiring from analog to digital (e.g., use of Fieldbus) has been driven by economics, functional performance and reliability.
- DCSs were developed and adopted because of a lower cost per control loop, their ability to implement all standard control technologies and network communication protocols, their ability to oversee, control, and optimize entire manufacturing plants (containing many thousands of I/O points), and their use to consolidate/centralize operational functions (e.g., MMI, alarm system) in one system, rather than having operators deal with different HMIs and alarm systems in different vendor supplied control packages.
- PLCs offer significant advantages over DCSs for smaller-scale systems, such as bio-processes, because PLCs are easier to maintain due to their modular nature, have a lower total cost and are able to better withstand harsh operating environments.
- The primary influence of the controller on the performance of a control loop comes from the settings for controller tuning, the control interval and possibly the resolution of the A/D and D/A converters.
- Control valves are the most commonly used actuator in the CPI. Equal percentage control valves, which should be used when the pressure drop required by the valve is a strong function of flow rate, are used in approximately 90% of the cases in the CPI. Otherwise, linear control valves are used.
- Valve positioners are routinely used to overcome the inherently large deadband associated with industrial control valves.
- For line sizes greater than 6 inches in diameter, butterfly valves are preferred over globe valves due to a significantly lower capital cost.
- Most bio-processes use positive displacement adjustable speed pumps as actuators for feed loops (e.g., feeding glucose, acid, base) because the scale of the processes is relatively small, the adjustable speed pumps are fast responding with low levels of deadband and they are easier to clean and maintain sterile.
- The primary influence of the actuator system on the performance of a control loop comes from the dynamic response and the deadband of the actuator.
- The sensor system is composed of the sensor, the transmitter and the sample system, which brings the sensor in contact with the process.
- For the CPI, the primary sensors are for measuring temperature (TC, RTD and thermistor), differential pressure (strain gauge and balance bar), flow rate (orifice meter, magnetic flow meter and vortex shedding flow meters), level (differential pressure) and composition analysis (GC).
- For the bio-tech industries, the primary sensors are for measuring temperature (TC and RTD), flow rate (Coriolis meter and rotameter), cell concentration (turbidity meter), pH, redox potential and DO analysis (ion specific electrodes), agitator RPM, pressure, off-gas analysis (mass spectrometers) and fermentation product analysis (HPLC and FIA).

- The primary influence of the sensor system on the performance of a control loop comes from the dynamic response, the repeatability and the accuracy of the sensor.

2.6 References

1. Lucas, M.P., *Distributed Control Systems*, Van Reinhold Company, New York, p. 4 (1986).

2. Liptak, B.G., *Instrument Engineers Handbook*, Chilton, Philadelphia, PA, pp. 721-764 (1995).

3. Coggan, D.A., *Fundamentals of Industrial Control*, 2nd Ed., ISA, Research Triangle Park, NC, pp. 605-702 (2005).

4. Alford, J. and F. Cline, Computer System Validation - Installation Qualification, *BioPharm*, Sept. 1990, and *Pharmaceutical Technology*, Sept. 1990.

5. Bird, R.B., W.E. Stewart, and E.N. Lightfoot, Transport Phenomena, Wiley and Sons, New York, NY, pp. 211-212 (1960).

6. Junk, Ken., Fisher Rosemount, personal communication (2001).

7. *Perry's Chemical Engineers Handbook*, 7th edition, R. H Perry and D. W. Green, editors, McGraw-Hill, New York, pp. 8-64 (1997).

8. Coggan, D.A., *Fundamentals of Industrial Control*, 2nd Ed., ISA, Research Triangle Park, NC, pp.1-200 (2005)

9. *Perry's Chemical Engineers Handbook*, 7th edition, R. H Perry and D. W. Green, editors, McGraw-Hill , pp. 8-47 to 8-53 (1997).

10. McMillan, G. K., G. E. Mertz and V. L. Trevathan, "Troublefree Instrumentation", *Chemical Engineering*, pp. 80-88 (Nov 1998).

11. McMillan, G.K. *Advanced Temperature Control*, Instrument Society of America, pp. 133-155(1995).

12. Coggan, D.A., *Fundamentals of Industrial Control*, 2nd Ed., ISA, Research Triangle Park, NC, pp. 138 (2005)

13. Alford, J., *Bioprocess Mass Spectrometry: A PAT Application*, Journal of Process Analytical Technology, May-June., 2006

2.7 Additional Terminology

A/D Converter - analog-to-digital converter. Converts a 4-20 mA electrical analog signal into a digital reading that can be processed by the DCS.

Accuracy - the difference between the true value and the measurement.

Air-to-close actuator - a valve actuator that causes the valve to close as its instrument air pressure is increased (i.e., fails open).

Air-to-open actuator - a valve actuator that causes the valve to open as its instrument air pressure is increased (i.e., fails close).

Basic Process Control System - a variety of control systems that are used to provide control function, including DCSs and PLCs.

BPCS - a variety of control systems that are used to provide control function, including DCSs and PLCs.

Bias - is the difference between a sensor reading and its true value.

Cage-guided valve - a valve with a cage around the valve plug that guides the plug toward the valve seat.

Calibration - an adjustment of the correlation between the sensor output and the predicted measurement so that the sensor reading agrees with the standard.

CRT - cathode ray tube. A computer console that allows the operators and engineers to access process operating conditions and adjust the process control activities of a DCS.

Control interval - the time period between adjacent calls to a controller from a DCS.

Controller cycle time - the time period between adjacent calls to a controller from a DCS.

D/A converter - digital-to-analog converter. Converts a digital value from the DCS into a 4-20 mA electrical analog signal.

Data highway - communication hardware and the associated software in a DCS that allows the distributed elements of a DCS to exchange data with each other.

Deadband - the maximum percentage change in the input that can be implemented without an observable change in the output.

DCS - distributed control system. A control computer that is made up of a number of distributed elements that are linked together by the data highway.

DDC - direct digital control. A single mainframe control computer. One of the first attempts at using computers for process control.

Direct-acting final control element - a final control element with an air-to-open valve actuator or a control valve with a valve positioner.

DO - dissolved oxygen concentration in the broth of a bio-reactor.

DP cell - a differential pressure sensor/transmitter.

Dynamic response time - the time for a system to make most of its change after an input change has occurred.

FIA - flow injection analyzer. A composition analyzer used to measure composition of fermentation products in the bio-tech industries.

GC - gas chromatography. A composition analyzer that is based on separating the components of a mixture in a small diameter packed column.

HPLC - high pressure liquid chromatography. A composition analyzer used to measure composition of fermentation products in the bio-tech industries.

I/P converter - an electro-mechanical device that converts a 4-20 mA electrical signal to a 3-15 psig pneumatic signal, i.e., a current-to-pressure converter.

Inherent valve characteristics - the flow rate versus stem position for a fixed pressure drop across the valve.

Installed valve characteristics - the flow rate versus stem position for a valve installed in service.

LCU - local control unit. A microprocessor in a DCS that is responsible for performing control functions for a portion of a plant.

Ladder logic - a programming language used in PLCs to implement a sequence of actions.

LAN - local area network.

PLC - programmable logic controller. A process computer typically used to apply a sequence of control actions, e.g., startup, shutdown and batch operations.

Process measurement dynamics - a measure of the speed with which a sensor responds to a change in the process.

Range - the maximum and minimum sensor reading.

Rangeability - the ratio of the largest accurate sensor reading to the smallest accurate reading.

Redox potential - a measure of the relative reducing/oxidizing character of the broth in a bio-reactor, which provide a measure of the health of an aerobic fermentation process.

Repeatability - the variation in a sensor reading not due to process changes. It provides an indication of consistency of the sensor reading.

Respiration quotient (RQ) - the ratio of the CO_2 produced divided by the O_2 consumed for a bio-reactor.

Reverse-acting final control element - a final control element with an air-to-close valve actuator if the valve does not have a valve positioner.

RTD - resistance temperature device. A temperature sensor that is based on the known temperature dependence of a pure metal resistor.

Shared communication facility - communication hardware and the associated software in a DCS that allows the distributed elements of a DCS to exchange data with each other.

Smart sensor - a sensor that is equipped with a microprocessor that provides onboard diagnostics and/or calibration.

Span - the difference between the maximum and the minimum value of a measurement that can be made by a sensor/transmitter.

Stiction - the static friction that needs to be overcome to enable relative motion of stationary objects in contact.

Systematic error - is the difference between a sensor reading and the true value.

TC - thermocouple. A temperature sensor that is based upon the fact that metal junctions at different temperatures generate an electrical voltage.

Thermistor - a temperature sensing device that is based on temperature dependence of the resistance of a solid state semi-conductor.

Transmitter - a device that is part of a sensor system that converts the output of the sensor to a 4-20 mA signal.

Transducer - a device that is part of a sensor system that converts the output of the sensor to a 4-20 mA signal.

Turbidity - a measure of the amount of light scattered by suspended solids, which is used to measure cell concentration.

Turndown ratio - the ratio of the maximum to minimum controllable flow rates for a control valve.

VDU - video display unit. A computer console that allows the operators and engineers to access process operating conditions and adjust the process control activities of a DCS.

Valve deadband - the maximum percentage change in the input to the valve that can be implemented without an observable change in the flow rate through the valve.

Valve plug - the device in a valve that is responsible for restricting flow through the valve.

Valve positioner - a device that adjusts the instrument air pressure to a control valve to maintain a specified value for the stem position.

Valve seat - the portion of the valve against which the valve plug rests when the valve is fully closed.

Valve stem - a rod that connects the diaphragm in the valve actuator with the valve plug so that, as the air pressure acts on the diaphragm, the plug provides more or less restriction to flow through the valve.

Valve trim - the valve flow characteristics that are determined by the geometry of the valve plug, seat and cage.

Zero - the lowest sensor/transmitter reading possible, i.e., the sensor reading corresponding to a transmitter output of 4 mA.

2.8 Preliminary Questions

2.1 Introduction

Q2.1.1 For a typical control loop, where are 3-15 psig signals used?

Q2.1.2 For a typical control loop, where are A/D and D/A converters used?

Q2.1.3 What hardware comprises a typical sensor system?

Q2.1.4 For a pneumatic controller, what mechanical devices are used to implement PID control?

2.2 Control System

Q2.2.1 Why have electronic analog controllers replaced pneumatic controllers?

Q2.2.2 Why have DCSs replaced electronic analog controllers?

Q2.2.3 What type of device is used as a local control unit?

Q2.2.4 What system in a DCS allows a user on a local console to observe operation of the plant controlled by other LCUs?

Q2.2.5 Based on Figure 2.2.4, where are CRTs used in a DCS?

Q2.2.6 Using Figure 2.2.4, explain how process data are stored and later displayed on a system console.

Q2.2.7 What is a PLC and how is it different from a DCS? How are they alike?

Q2.2.8 What is the purpose of a process alarm system?

Q2.2.9 For what type of control function were PLCs originally designed?

Q2.2.10 Why do some processes use personal computers as the control computer?

Q2.2.11 What is the difference between a DCS and the fieldbus approach to distributed control?

Q2.2.12 For a fieldbus system (Figure 2.2.7), in what devices are control calculations performed?

2.3 Actuator Systems (Final Control Elements)

Q2.3.1 What is the difference between the actuator system and the final control element? What is the difference between the actuator system and the valve actuator?

Q2.3.2 How do you choose between selecting a globe valve with an unbalanced plug and one with a balanced plug?

Q2.3.3 Why are equal percentage valves generally selected over linear and quick-opening valves?

Q2.3.4 Using Figure 2.3.2, indicate how you determine the stem position of a valve in operation.

Q2.3.5 Why is the pressure drop across a control valve in a flow system usually a strong function of flow rate?

Q2.3.6 In general, is the valve position constant for a control valve with a valve positioner?

Q2.3.7 Explain how cavitation in control valves occurs and what it causes.

Q2.3.8 What physical characteristic of the process determines whether a valve actuator is air-to-open or air-to-close?

Q2.3.9 Identify a case where an air-to-close valve actuator should be used and explain your reasoning.

Q2.3.10 Why has an increased usage of DCSs resulted in a greater use of valve positioners?

Q2.3.11 Under what conditions are adjustable speed pumps preferred over a control valve?

Q2.3.12 Why are adjustable speed pumps used for most bio-processes?

2.4 Sensor Systems

Q2.4.1 From a process point of view, which is generally more important for a sensor, accuracy or repeatability? Explain your reasoning.

Q2.4.2 When is the dynamic response time of a sensor important to a process control system and when is it not important?

Q2.4.3 What is a smart sensor?

Q2.4.4 What are the two most important differences between TCs and RTDs ?

Q2.4.5 What are the two most common types of differential pressure sensors?

Q2.4.6 Why is a straight run of pipe preceding an orifice meter required?

Q2.4.7 What advantages and disadvantages does an orifice flow meter have compared to magnetic flow meters and vortex shedding meters?

Q2.4.8 What determines whether an on-line analyzer should be installed?

Q2.4.9 What special considerations are required for bio-sensors compared to sensors applied to the CPI?

Q2.4.10 What is the difference between a sensor that is used for process monitoring and a sensor that is used for feedback control?

Q2.4.11 What is the turbidity and what property of a bio-reactor is it used to measure?

Q2.4.12 How are the measurement of the concentration of O_2 and CO_2 in the off-gas from a fermentation process used to monitor the operation of bio-reactor?

2.9 Analytical Questions and Exercises

2.1 Introduction

P2.1.1* Determine the 4-20 mA signal and the pneumatic signal if the controller output is 72% of its full-scale reading for a system corresponding to Figure 2.1.2.

P2.1.2* Determine the 4-20 mA signal and the controller output in percent of full range, if the pneumatic signal is 10 psig for a system corresponding to Figure 2.1.2.

P2.1.3** For the control loop shown in Figure P2.1.3, make a drawing similar to Figure 2.1.2 and list all signals on your diagram. Assume that a pneumatic valve positioner and a DCS control computer are used in this application.

P2.1.4** For the control loop shown in Figure P2.1.4, make a drawing similar to Figure 2.1.2 and list all signals on your diagram. Assume that a pneumatic valve positioner and an pneumatic controller are used in this application.

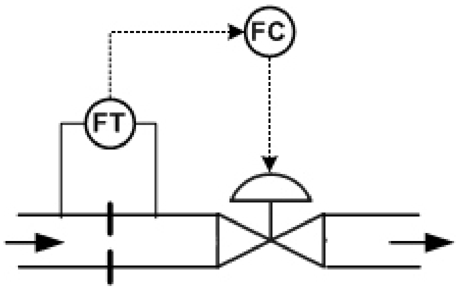


Figure P2.1.3 Control diagram of a flow controller.

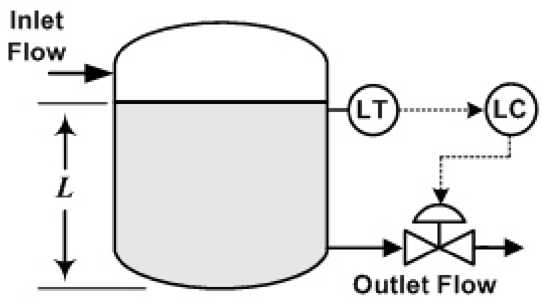


Figure P2.1.4 Control diagram of a tank with a level controller.

P2.1.5** For the control loop shown in Figure P2.1.5, make a drawing similar to Figure 2.1.2 and list all signals on your diagram. Assume that an electronic valve positioner and an electronic analog controller are used in this application.

P2.1.6*** For the control loop shown in Figure P2.1.6, make a drawing similar to Figure 2.1.2 and list all signals on your diagram. Assume that a Fieldbus control computer are used in this application.

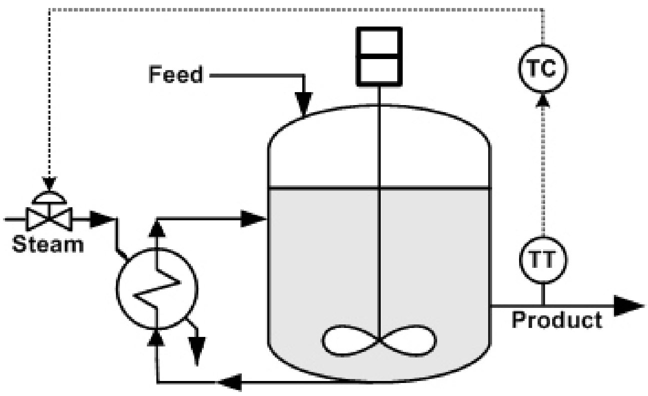


Figure P2.1.5 Control diagram of an endothermic CSTR with a temperature control loop.

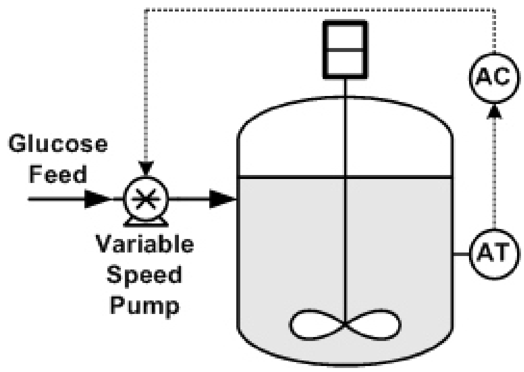


Figure P2.1.6 Control diagram for cell concentration controller on a fed-batch reactor.

P2.1.7** Consider Figure 2.4.10. Assume that a variable speed air blower is used to deliver air to the sparger and that it is adjusted by a control algorithm in a DCS to maintain a specified dissolved oxygen level in the reactor broth. Draw a figure similar to Figure 2.1.2 with all the relevant signals assuming that the variable speed blower takes a 4-20 mA control signal as its input.

2.2 Control System

2.2.1* What are the key advantages of Fieldbus Technology over DCS**?**

2.2.2** Why have DCSs replaced analog controllers and supervisory control computers in the CPI? Why is Fieldbus technology likely to begin replacing DCSs in the future? Can you identify a pattern?

2.3 Actuator Systems (Final Control Elements)

2.3.1* Explain why a valve with a 10% deadband may not produce a change in the flow rate through the valve for a 9% change in the signal to the control valve. Explain why a 2% change in the signal to the control valve may, in certain situations, produce a change in the flow rate through the same control valve.

2.3.2* Calculate the flow rate in gpm of a hydrocarbon stream with a density of 44 lb/ft^3 through a 3-inch linear control valve that is 80% open based on stem position with a pressure drop across the valve equal to 82 psi. Assume that the C_v for this valve when it is fully open (100% valve position) is equal to the C_v for a fully open 3-inch equal percentage valve whose valve characteristics are listed in Table 2.1.

2.3.3* Calculate the pressure drop across a 3-inch equal percentage valve that is 63% open based on stem position for a flow of 95 gpm of a hydrocarbon stream with a density of 45 lb/ft^3. Obtain the C_v for this valve from Table 2.1.

2.2.4* Calculate the pressure drop across a 2-inch linear valve that is 80% open based on stem position for a flow of 35 gpm of a hydrocarbon stream with a density of 45 lb/ft^3. Assume that the C_v for this valve when it is fully open (100% valve position) is equal to the C_v for a fully open 2-inch equal percentage valve whose valve characteristics are listed in Table 2.1.

2.3.5** Determine the flow rate of water in gpm though a 3-inch linear valve that is 15% open based on stem position with a maximum value of C_v equal to 100 and the installed pressure drop presented in Table 2.2.

2.3.6** Determine the flow rate of a hydrocarbon liquid (specific gravity equal to 0.65) though a 3-inch equal percentage valve that is 30% open based on stem position for the C_v given in Table 2.1 and the installed pressure drop presented in Table 2.2.

2.3.7* Determine the valve stem position of a 2-inch equal percentage valve if 64 gpm of hydrocarbon liquid (specific gravity equal to 0.65) flow through the system. Use Table 2.1 to determine the C_v and Table 2.2 for the available pressure drop across the control valve.

2.3.8* Determine the valve stem position of a 2-inch linear valve if 74 gpm of hydrocarbon liquid (specific gravity equal to 0.65) flow through the system. Use Table 2.2 for the available pressure drop across the control valve. Assume that the C_v for this valve when it is fully open (100% valve position) is equal to the C_v for a fully open 2-inch equal percentage valve whose valve characteristics are listed in Table 2.1.

2.3.9** Size a control valve for service on a line carrying water with a maximum flow rate of 200 gpm and a minimum flow rate of 20 gpm. Assume an equal percentage valve with the pressure drop versus stem position shown in the table for this problem. The C_v's for equal percentage valves of different sizes are presented in the Table 2.1.

2.3.10** Size a control valve for service on a line carrying a hydrocarbon liquid (specific gravity equal to 0.65) with a maximum flow rate of 400 gpm and a minimum flow rate of 50 gpm. Assume an equal percentage valve with the pressure drop versus stem position shown in the table for this problem. The C_v's for equal percentage valves of different sizes are presented in the Table 2.1.

2.3.11** Size a control valve for service on a line carrying a hydrocarbon liquid (specific gravity equal to 0.75) with a maximum flow rate of 300 gpm and a minimum flow rate of 100 gpm. Assume an equal percentage valve with the pressure drop versus stem position shown in the table for this problem. The C_v's for equal percentage valves of different sizes are presented in the Table 2.1.

Pressure Drop versus Flow Rate for an Installed Valve for Problems 2.3.9-2.3.11

Q (gpm)	P (psi)	Q (gpm)	P (psi)
0	30.6	220	22.9
20	30.5	240	21.5
40	30.2	260	20.2
60	29.9	280	18.7
80	29.5	300	17.0
100	28.9	320	15.0
120	28.2	340	12.6
140	27.4	360	9.9
160	26.5	380	8.6
180	25.4	400	7.4
200	24.2	420	6.1

2.4 Sensor Systems

2.4.1* Determine the flow rate reading corresponding to a 12.2 mA analog signal from a flow transmitter that has a span of 10,000 lb/hr and a zero of 1,000 lb/hr.

2.4.2* Determine the level reading corresponding to a 6.8 mA analog signal from a level transmitter that has a span of 75% and a zero of 10%.

2.4.3* Determine the value of the electric analog reading in mA from a flow transmitter that has a span of 100,000 lb/hr and a zero of 15,000 lb/hr corresponding to a measured flow rate of 66,732 lb/hr.

2.4.4* Determine the value of the electric analog reading in mA from a level transmitter that has a span of 65% and a zero of 12% corresponding to a measured level of 47%.

2.4.5* Consider a flow sensor/transmitter that reads 10,000 lb/h when the transmitter output is 7 mA and reads 15,000 lb/h when the transmitter output is 12 mA. What are the zero and span of this flow sensor/transmitter?

2.4.6* Consider a pressure sensor/transmitter that reads 180 psig when the transmitter output is 6 mA and reads 250 psig when the transmitter output is 10 mA. What are the zero and span of this pressure sensor/transmitter?

2.4.7** Calculate the flow rate of water through an orifice with equal to 0.87 in a Schedule 40 4-inch line if the pressure drop across the orifice is 2 psi.

2.4.8** Calculate the pressure drop across an orifice with equal to 0.59 in a Schedule 40 2-inch line for a flow rate of a hydrocarbon liquid (specific gravity equal to 0.65) of 40 GPM.

2.4.9*** Size an orifice meter for service on a 2-inch Schedule 40 pipe carrying a hydrocarbon liquid (specific gravity equal to 0.65) with a maximum flow rate of 75 GPM and a minimum flow rate of 25 GPM. Assume that the line pressure is 150 psig.

2.4.10*** Size an orifice meter for service on a 3-inch Schedule 40 pipe carrying a hydrocarbon liquid (specific gravity equal to 0.65) with a maximum flow rate of 175 GPM and a minimum flow rate of 75 GPM. Assume that the line pressure is 150 psig.

2.10 Projects

Project 2.1 Specify the instruments and design the orifice meters and control valves for each of the following cases. Size the lines (i.e., to available cast iron Schedule 40 pipe sizes) based on assuming that the linear velocity in the lines is approximately 7 ft/s. For each case, assume that water is the process fluid and the line pressure is 150 psig.

a. CST temperature mixer (Example 3.2) shown in Figure Project 2.1a. The operating conditions for this process are given as

- F_1 - mass flow rate of stream 1 (5 kg/s)
- F_2 - mass flow rate of stream 2 (5 kg/s)
- M - mass of liquid in the mixer (100 kg)
- T - temperature of the mixed liquid (50 C)
- T_1 - temperature of stream 1 (25 C)
- T_2 - temperature of stream 2 (75 C

b. Level in a tank (Example 3.4) is shown in Figure Project 2.1b. The operating conditions for this process are given as:

- A_c - cross-sectional area (0.3 m^2)
- L - the level of liquid in the tank (2 meters)
- F_{in} - the mass flow rate of liquid into the tank (1.0 kg/s)
- F_{out} - the mass flow rate of liquid leaving the tank (1.0 kg/s).
- - the fluid density (1 kg/l)

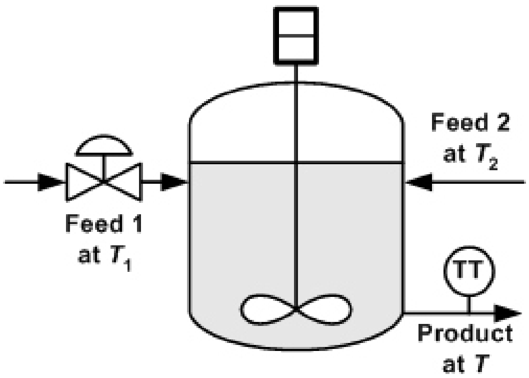


Figure Project 2.1a Schematic of a CST thermal mixing process.

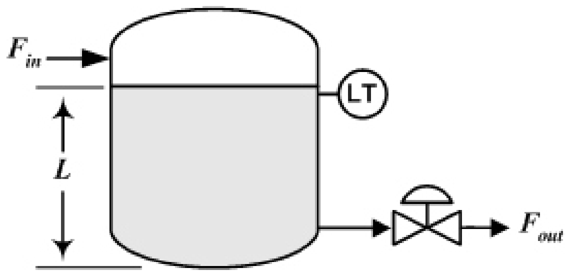


Figure Project 2.1b Schematic of a level in a tank process.

Project 2.2 Grains, such as corn and sorghum, are used to produce ethanol for use in motor fuel (i.e., gasohol). A process for producing ethanol from grain is shown in Figure Project 2.2. Crushed grain, water and enzymes are added to the enzyme reactor and heated to approximately 140ºF and held at this temperature for 60-80 h so that the enzymes can break down the carbohydrates of the grain into glucose. Steam is fed to a set of submerged coils in the tank to heat the grain/enzyme solution up to temperature and to maintain this temperature for the desired processing period. After the conversion to glucose is complete, the contents of the enzyme reactor are pumped into the fermentor, including the solids. Initially, a portion of the glucose stream is pumped into the fermentor and the inoculum (active yeast) are added. After an incubation period of approximately 2-3 h, the remainder of glucose from the enzyme reactor is pumped into the fermentor. Then the fermentor is operated in the batch mode while maintaining pH by adding NaOH solution and temperature by using cooling water for a total fermentor processing time of 15-20 h until the desired EtOH concentration is attained. After the fermentation reactions are complete, the contents of the fermentor, including the solids, are pumped to a distillation column, which concentrates the ethanol (EtOH) in the overhead product to a 80 volume per cent product. A tray temperature near the top of the column is maintained at a fixed level to ensure that the overhead product is the desired composition by adjusting the reflux flow rate. The accumulator level is maintained by adjusting the overhead product rate while the reboiler level is maintained using the bottom product flow rate. The bottoms product of the column contains the solids and a weak ethanol solution.

a. Place the necessary actuators on PFD for this system.

b. Add the necessary sensors for the proper operation of this process.

c. Add feedback controllers where necessary assuming that an operator will serve as the supervisory controller for this process. That is, only flow, level, pressure, pH and temperature controllers should be applied as feedback controllers.

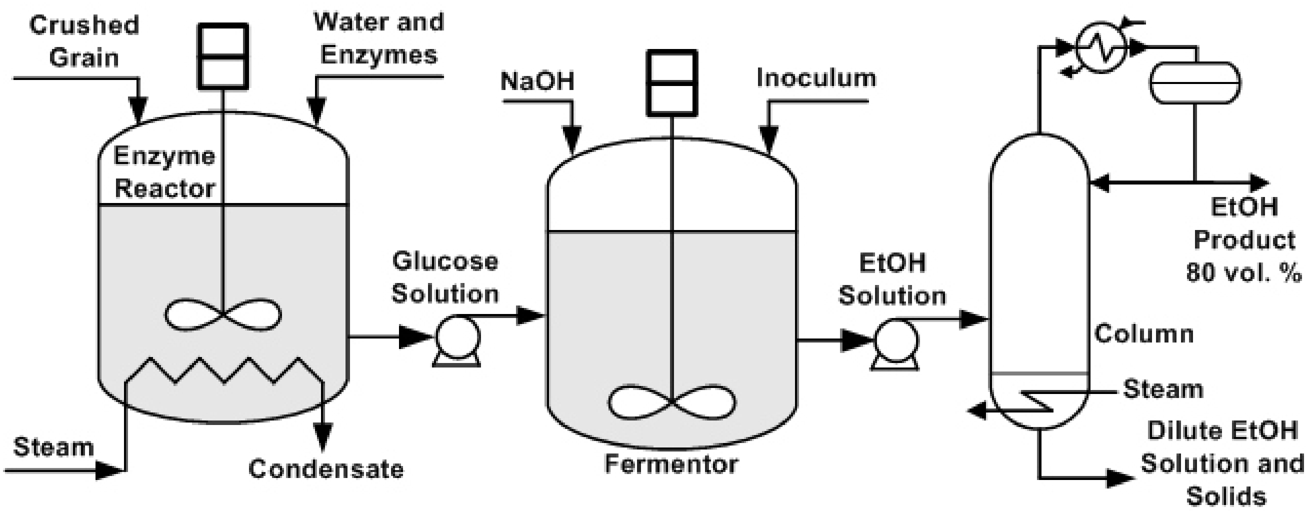


Figure Project 2.2 PFD of a ethanol (EtOH) from grain process.

Part II

Process Dynamics

Chapter 3

Dynamic Modeling

Chapter Objectives

- Define lumped and distributed parameter models.
- Present the general unsteady-state balance equations that are used to develop dynamic process models.
- Introduce empirical dynamic models for actuator and sensor systems.
- Present the numerical solution of actuator/process/sensor models using MATLAB and Python for a number of simple processes relevant to the CPI and bio-tech industries.

3.1 Introduction

Most chemical engineering courses are concerned with the steady-state aspects of chemical and biological systems. Due to almost continuous fluctuations in feed flow rate and composition to a process and other disturbances, such as steam pressure and cooling water temperature changes, most processes are in a constant state of flux. Also, the control systems for these units respond to these disturbances using changes in the MVs to drive the CVs toward their setpoints. As a result, generally there are few periods, if any, that resemble purely steady-state operation. Therefore, dynamic behavior is an integral part of industrial operations, including process control operations. The first step in becoming knowledgeable in the field of process control is to develop an understanding of process dynamics.

Understanding the dynamic behavior of chemical and biological processes depends first on understanding the steady-state behavior of these processes. For example, if an input is implemented on a process (e.g., a change in the MV), a steady-state analysis of the process using the new input level indicates where the process settles after a sufficient period of time. Then the dynamic characteristics of the process (e.g., the time constant and deadtime) can be used to determine how long it takes to approach the new steady-state and what path the process takes. Understanding the dynamic behavior of a complex process made up of a number of separate unit operations (heat exchangers, reactors, separation devices, holding tanks, etc.) requires combining the dynamic behavior of all the unit operations. Using a dynamic understanding of each unit in the process, the control engineer can anticipate the transient behavior of a complex process by considering the effect of an input as its effect propagates through the entire process.

Dynamic models for process control applications, which are usually in the form of ordinary differential equations (ODEs), should have the outputs (e.g., CVs) expressed as a function of the inputs (e.g., MVs and DVs). Dynamic models can be as simple as a SISO system or as complex as a dynamic model of a major portion of a process plant. Several dynamic models relevant to the CPI and the biotechnology industries are used in this chapter to illustrate the key features of process modeling and dynamics.

3.2 Uses of Dynamic Models

Process Design. Dynamic models can be used to design batch processing systems. For example, the reaction time of a batch bio-reactor can be determined from a dynamic model of the reactor so that production rate and product quality specifications are met.

Analysis of Process Control Approaches. A range of potential process control configurations can be compared directly using dynamic models. Each control approach can be implemented using a dynamic model of the process and the resulting control performance can be calculated for a standard disturbance test. In this manner, a controlled comparison between different control approaches can be quantitatively assessed, i.e., each controller can be applied to exactly the same process with exactly the same disturbances. As processes become more highly integrated (i.e., using material recycle and heat recovery) in an effort to improve the economic performance of the process, the use of dynamic models for process control evaluation is becoming increasingly important to ensure that these highly integrated processes produce on-specification products with safe and reliable operation.

Operator Training. A dynamic simulation of a process can be interfaced with the same type of control computer that controls the actual process and the resulting system can be used to train operators. In this manner, the operator can be exposed to a wide range of major upsets and potentially dangerous operational scenarios without upsetting or endangering the process.

Start-up/Shutdown Strategy Development. Viable process startup and shutdown strategies can be identified using dynamic process simulators. This class of dynamic simulators must be able to model process behavior over a much wider range of operation than the dynamic simulators used for process control evaluation.

3.3 Classifications of Phenomenological Models

There are two general classifications of phenomenological process models (i.e., models based on conservation of mass, energy and momentum): **lumped parameter models** and **distributed parameter models**. A lumped parameter model assumes that the dependent variables of the process are not functions of spatial location within the process. Consider the schematic of a reactor shown in Figure 3.3.1. A lumped parameter model of this reactor assumes that the composition of the chemical species present in the reactor and the temperature in the reactor are uniform throughout the reactor volume. If the mixer on the reactor is properly designed, this assumption can be quite good. **Macroscopic balances** are used to model lumped parameter systems and consider what enters or leaves the process and what is occurring inside the process as a whole. For Figure 3.3.1, feed enters the reactor, product is removed, heat is added by the heat exchanger and the same reaction rate occurs throughout the reactor volume. A model developed from macroscopic balances is called a **macroscopic model**.

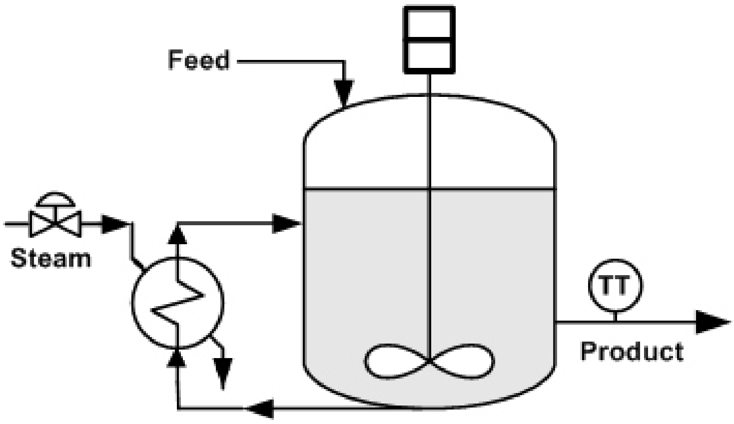


Figure 3.3.1 PFD of a CSTR.

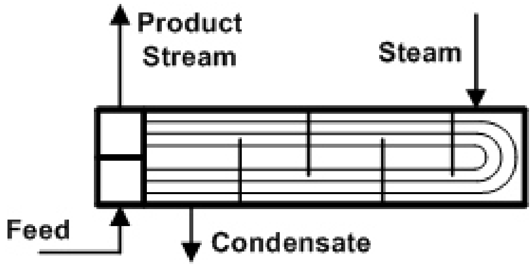


Figure 3.3.2 PFD of a heat exchanger.

Models that consider the spatial variation in the dependent variables are referred to as **distributed parameter models**. If the mixer on the reactor in Figure 3.3.1 is not functioning properly, significant spatial variation in the concentrations of the reactor species and reactor temperature can result, requiring a distributed parameter model. The heat exchanger shown in Figure 3.3.2 is an example of a distributed parameter system because the tube-side temperature of the process fluid changes continuously as it flows through the heat exchanger. Microscopic balances, i.e., balances based on differential volumes, are typically used to model distributed parameter processes. These microscopic balances are used to derive differential equations, which can be applied over the full spatial region to develop a model that describes the entire process. A model developed by applying microscopic balances is called a **microscopic model**.

3.4 Dynamic Balance Equations

This section presents unsteady-state balance equations based on conservation of material and energy that can be applied to develop dynamic process models. Depending upon whether the balances are applied around an overall process or to a differential element within the process, these equations can be applied to develop either macroscopic or microscopic models, respectively.

Total or Component Mass Balance Equation. Mass balance equations relate the rate of accumulation of mass of a component in the system of interest to the rate of the mass of the component entering and leaving the system:

rate of accumulation of mass within the system	flow rate of mass entering the system	flow rate of mass leaving the system	**3.4.1**

Note that this equation can be applied to the total mass of a system or the mass of a component in the system. The units of each term in Equation 3.4.1 are mass per time. The accumulation term can be evaluated in several ways. For example, for a total mass balance, the rate of accumulation of mass within the system can be expressed in terms of the total mass within the boundaries of the system (m) or in terms of the density ( ) and volume (V) of the material in the system as

$$\frac{dm}{dt} \text{ or } \frac{d(\ V)}{dt}$$

On the other hand, for a component mass balance, the rate of accumulation of mass of a component within the system can be expressed in terms of the total mass of component i (m_i) or the mass fraction of the component of interest (x_i) and the total mass in the system (m) as

$$\frac{dm_i}{dt} \text{ or } \frac{d(x_i m)}{dt}$$

For a total mass balance, the flow rate of mass entering or leaving the system is simply the flow rate of each stream that enters or leaves the system. For a component mass balance, the flow rate of a component entering or leaving the system is equal to the product of mass fraction (x_i) of the component and the corresponding stream mass flow rate. **Equation 3.4.1** applies to the total mass of a system whether chemical reactions are occurring or not, but **applies to the mass of a component only if no chemical reactions involving that component occur**.

Component Mole Balance Equation. The following equation represents the conservation of the number of moles of a component in a reacting system.

rate of accumulation of moles within in the system	flow rate of moles entering the system	flow rate of moles leaving the system	
rate of generation of moles by reaction	rate of consumption of moles by reaction		**3.4.2**

This equation applies to the total number of moles or the moles of a particular component in the system, but is usually applied as a component mole balance. The units of each term in Equation 3.4.2 are moles per unit time. The accumulation term can be expressed in terms of the total moles of component i (n_i) or in terms of the volume of material in the system (V) and the concentration of component i (C_i) as

$$\frac{dn_i}{dt} \text{ or } \frac{d(VC_i)}{dt}$$

The flow rate of a component entering or leaving a system is usually expressed as the product of the mole fraction (x_i) times the total molar flow rate of the stream (N) or the product of the concentration times the volumetric flow rate of the stream (F_V), i.e.,

$$x_i N \text{ or } C_i F_V$$

For a single reaction, the rate of generation or consumption of a component by reaction is given by

$$\nu_i r V$$

where ν_i is the stoichiometric coefficient for component i in the reaction, r is the reaction rate and V is the volume of material undergoing reaction.

Energy Balance Equation. For processes such as reactors, heat exchangers and distillation columns, potential energy changes, kinetic energy changes, mechanical work and the heat generated by frictional losses are typically negligible compared with convective heat transfer (i.e., the energy carried with streams entering or leaving the process), heat exchanged across the boundaries of the system and the energy generated or consumed by reaction. Therefore, the energy balance equation typically used for process systems is given by

$$\begin{bmatrix}\text{rate of}\\ \text{accumulation of}\\ \text{thermal energy}\end{bmatrix} = \begin{bmatrix}\text{rate of convective}\\ \text{heat transfer}\\ \text{entering the system}\end{bmatrix} - \begin{bmatrix}\text{rate of convective}\\ \text{heat transfer}\\ \text{leaving the system}\end{bmatrix} + \begin{bmatrix}\text{net rate of energy}\\ \text{generation by}\\ \text{chemical reaction}\end{bmatrix} + \begin{bmatrix}\text{net rate of heat transfer}\\ \text{through the boundaries}\\ \text{of the system}\end{bmatrix} \qquad \textbf{3.4.3}$$

The units for each term in this equation are energy per unit time. The thermal energy of the material within the boundaries of a system can be expressed as $MC_v(T - T_{ref})$, where M is the total mass in the system, C_v is the heat capacity at constant volume of the material in the system, T is the temperature of the material in the system and T_{ref} is the reference temperature. Therefore, the rate of accumulation of thermal energy is

$$\frac{d[MC_v(T - T_{ref})]}{dt} = MC_v\frac{dT}{dt}$$

when the mass and heat capacity are assumed to be constant. The rate of convective heat transfer can be expressed as $FC_p(T_S - T_{ref})$, where F, C_p and T_S are the flow rate, heat capacity at constant pressure and the temperature of a stream that enters or leaves the system. The rate of heat generation by reaction is given as $[-\Delta H_{rxn}(T)]rV$, where $\Delta H_{rxn}(T)$ is the heat of reaction at temperature T, r is the reaction rate and V is the volume of reacting material. For liquid systems, C_v can be assumed to be equal to C_p.

It should be pointed out that for fluid flow through a piping system, kinetic energy, mechanical work, potential energy and frictional losses are important and a mechanical energy balance, e.g., Bernoulli's equation[1], should be used to model these systems. Note that each of **the previous dynamic equations (i.e., Equations 3.4.1 to 3.4.3) can be converted into steady-state equations by setting the accumulation terms equal to zero.**

Constitutive Relationships. A number of physical relationships known as constitutive relationships are required to implement the equations that result from the application of Equations 3.4.1 to 3.4.3. Examples of **constitutive relations** include:

- Gas law equations
- Vapor/liquid equilibrium relationships
- Heat-transfer correlations

- Expressions for rates of reaction
- Correlations for pressure drop as a function of flow rate
- Physical property correlations (e.g., enthalpy functions)

For example, modeling the dynamic behavior of the temperature of a reactor requires rate expressions for the major reactions and a dynamic distillation column model uses vapor/liquid equilibrium relationships.

Degrees of Freedom Analysis. Developing a dynamic model of a complex process can involve a large number of differential equations and constitutive relationships, which are usually in the form of algebraic equations. A degrees of freedom analysis involves counting the total number of unknowns, N_v, i.e., variables that are unknown and must be calculated and the total number of independent equations, N_e, both differential and algebraic. The degrees of freedom, N_f, of a model is given by

$$N_f \quad N_v \quad N_e$$

When N_f is equal to zero, the model is referred to as **exactly determined** or **exactly specified**. For a solvable model, the number of degrees of freedom must be zero. When N_f is less than zero, the model is referred to as **overdetermined** or **overspecified**. For this case, there are more equations than unknowns, which can result from redundant equations or an improperly formulated model. When N_f is greater than zero, the model is referred to as **underdetermined** or **underspecified**. For this case, there are more unknowns than equations. To solve such a model, additional equations must be identified or unknown variables eliminated. The variables (e.g., time, flow rates, compositions and temperatures) that are specified are the **independent variables** while the variables computed from the solution of the equations are the **dependent variables**. To make this distinction clearer, the independent and dependent variables are identified in the examples that are presented in Section 3.7. Constants used in the model equations, such as densities, heat capacities, heats of reaction and gas constants, are called **process parameters**.

Self-Assessment Questions

Q3.4.1 What is the difference between a component mass balance (Equation 3.4.1) and a component mole balance (Equation 3.4.2)? When do they yield the same result?
Q3.4.2 What are constitutive relationships and how are they used for process models?
Q3.4.3 What does it mean when a set of model equations is overspecified?

Self-Assessment Answers

Q3.4.1 Equation 3.4.1 is a balance on the mass of a component entering and leaving a system, but does not consider the effect of chemical reactions. Equation 3.4.2 is a balance on the moles of a component entering, leaving and generated or consumed by reaction. Therefore, Equation 3.4.1 and 3.4.2 yield the same result if the component under consideration is not involved in any chemical reactions. In general, mass balances are applied to systems that do not involve chemical reactions while mole balances are used for systems that involve chemical reactions.
Q3.4.2 Constitutive equations take the form of algebraic equations and they are used to define relationships for dependent variables in a model.
Q3.4.3 A system is referred to as overspecified if there are more equations than unknown variables.

3.5 Numerical Solution of Dynamic Models using MATLAB

Dynamic models developed from dynamic material and energy balances take the form of a set of ordinary differential equations (ODEs), i.e., a separate ODE for each dependent variable. MATLAB offers a convenient built-in function[2] for solving this class of problems: `ode45`, which is based on a fourth-order Runge-Kutta integrator.

An ODE is a differential equation which contains derivatives with respect to only one independent variable. A single first-order ODE has the following general form:

$$\frac{dy}{dt} \quad f(t,y) \text{ where } y_0 \quad y(t_0)$$

where y is the dependent variable, t is the independent variable and $y(t_0)$ is the initial conditions for the dependent variable.

The general problem for a set of n-coupled first-order ODEs constituting a dynamic model can be represented as

$$\frac{dy_1}{dt} \quad f_1(t,\mathbf{y})$$

$$\frac{dy_2}{dt} \quad f_2(t,\mathbf{y})$$

$$\vdots$$

$$\frac{dy_n}{dt} \quad f_n(t,\mathbf{y})$$

where $\mathbf{y}$ is a vector containing each of the dependent variable values.

MATLAB's `ode45` is designed to solve a set of ODEs to a specified accuracy by adjusting the integration step size during the integration process. The call statement for applying `ode45` to solve a set of ODEs I

```
soln=ode45(%FunctionName,[t0,dt,tf],[y0],odeset('RelTol', 1.E-4, AbsTol, 1.E-6))
```

where '**`soln`**' is a matrix of dependent variable values calculated by `ode45` with each row containing each of the dependent variable values and corresponding to a different time, '`FunctionName`' is the name of the function (i.e., a subroutine in the program) that calculates the derivatives of the dependent variables with respect to the independent variable, '`[to,tf]`' is the range of integration of the ODEs from the initial conditions '`to`' to the final conditions '`tf`', `and` '`[`**`y0`**`]`' is a column vector of the initial conditions for the dependent variables. The function `odeset` allows the user to specify the accuracy requirement, i.e., `RelTol`, the relative error tolerance and `AbsTol`, the absolute error tolerance. In general, you should execute your code with smaller values of `RelTol` until you have an accurate solution. For the models that follow in Section 3.7 `RelTol` was set equal to 1.E-6, which more than meets the desired accuracy requirements.

Example 3.1 ODE Integration using MATLAB's `ode45`

Problem Statement. Consider the following set of ODEs and initial conditions:

$$\frac{dy_1}{dx} \quad 1 \quad y_1 \exp y_2 \qquad\qquad y_1(0) \quad 1$$

$$\frac{dy_2}{dx} \quad 1 \quad y_1 y_2 \qquad\qquad y_2(0) \quad 1$$

Using `ode45` develop a solution for these ODEs from x=0 to x=1.

Solution. Following is a listing of the MATLAB program and a plot of the solution for this problem.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                          NOMENCLATURE
%
%  dydt - the vector of functions of the derivative of y with respect to t
%  soln - the solution matrix for the ODE problem
%   t - a vector containing the values of the independent variable time
%  t0 - the initial value of t (0)
%  tf - the final value of t (1)
%  y1 - a vector containing the calculated value of the 1st dependent variable
%  y2 - a vector containing the calculated value of the 2nd dependent variable
%  y0 - the initials value of y (1 1)
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                             PROGRAM
function Ex3_1ode45intro
clear; clc;
t0=0; tf=1;  y0=[1;1];        % Input problems specifications
    % Call ode45 and store solution in soln
soln=ode45(@f,[t0,tf],y0,odeset('RelTol', 1.E-6, AbsTol, 1.E-6));
t=linspace(0,tf,100);         % Generate the values of t using linspace
y1=deval(soln,t,1);           % Retrive value of y(1) from soln
y2=deval(soln,t,2);           % Retrive value of y(2) from soln
plot(t,y1,‘k-’,t,y2,‘k—’,‘LineWidth’,2);     % Plot t/y data
legend(‘y1’,‘y2’);
grid on;                      % Add grid to plot
title(‘ode45 Solution for Ex3.1’);% Add title to plot
end
%
% User specified function for dydt for each dependent variable
%
function dydt=f(t,y)
dydt(1)=1-y(1)*exp(y(2));
dydt(2)=1+y(1)*y(2);
dydt=dydt’                    % Return dydt as column vector
end
%                             PROGRAM END
```

Note that this code can easily be used to solve another set of ODEs. Simply change the inlet problem specifications (i.e., the values of `t0`, `tf` and the initial conditions for **`y`**) and change the ODEs in function `dydt`.

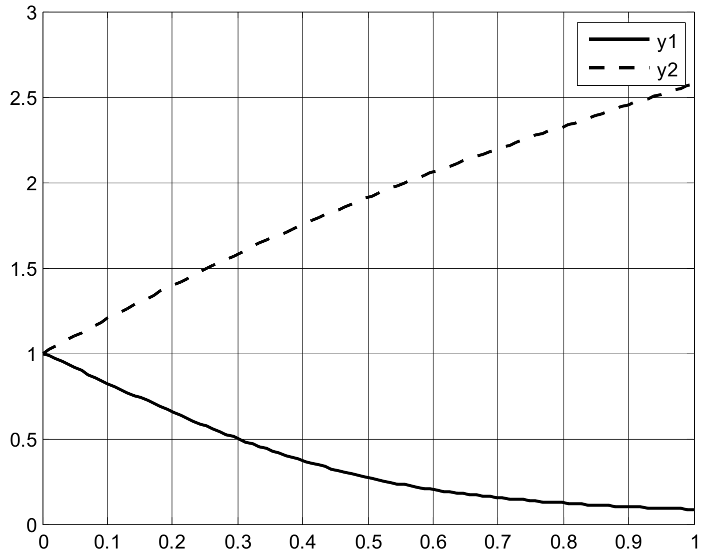


Figure 3.5.1 Results for Example 3.1

As the number of ODEs change, you will also need to change the retrival and plot statements.

3.6 Numerical Solution of Dynamic Models using Python

Dynamic models developed from dynamic material and energy balances take the form of a set of ordinary differential equations (ODEs), i.e., a separate ODE for each dependent variable. Python offers a convenient built-in function for solving this class of problems: solve_ivp, which can be used to apply a fourth-order Runge-Kutta integrator.

An ODE is a differential equation which contains derivatives with respect to only one independent variable. A single first-order ODE has the following general form:

$$\frac{dy}{dt} \quad f(t,y) \text{ where } y_0 \quad y(t_0)$$

where y is the dependent variable, t is the independent variable and $y(t_0)$ is the initial conditions for the dependent variable.

The general problem for a set of n-coupled first-order ODEs constituting a dynamic model can be represented as

$$\frac{dy_1}{dt} \quad f_1(t,\mathbf{y})$$

$$\frac{dy_2}{dt} \quad f_2(t,\mathbf{y})$$

$$\vdots$$

$$\frac{dy_n}{dt} \quad f_n(t,\mathbf{y})$$

where **y** is a vector containing each of the dependent variable values.

Python's solve_ivp is designed to solve a set of ODEs to a specified accuracy by adjusting the integration step size during the integration process. The call statement for applying solve_ivp to solve a set of ODEs is

soln=scipy.integrate.solve_ivp(fun, tspan, y0, method='RK45', t_eval=tvalues, rtol=1.E-4, atol=1.E-6)

where '**soln**' is a matrix of dependent variable values calculated by solve_ivp with each row containing each of the dependent variable values and corresponding to a different time, '**fun**' is the name of the user-specified function that calculates the derivatives of the dependent variables with respect to the independent variable, '**tspan**' is a vector that contains the range of integration of the ODEs from the initial conditions to the final conditions, '**y0**' is a column vector of the initial conditions for the dependent variables, '**method**' specifies the ODE integration method to be used (in this case the Runge Kutta 45 is used), '**t_eval**', a vector that contains the values of the independent variable for which dependent variable values are calculated, '**rtol**' is the relative error specification and '**atol**' is the absolute error specification. In general, when integrating ODEs, you should reduce rtol until the desire accuracy is attained. For the models that follow in Section 3.7, rtol was set equal to 1.E-6, which more than met the desired accuracy requirements.

Example 3.2 ODE Integration using Python's solve_ivp

Problem Statement. Consider the following set of ODEs and initial conditions:

$$\frac{dy_1}{dx} \quad 1 \quad y_1 \exp y_2 \qquad y_1(0) \quad 1$$

$$\frac{dy_2}{dx} \quad 1 \quad y_1 y_2 \qquad y_2(0) \quad 1$$

Using scipy.integrate.solve_ivp develop a solution for these ODEs from x=0 to x=1.

Solution. Following is a listing of the Python program and a plot of the solution for this problem.

```
###################################################################################
#                                   NOMENCLATURE
#
# dy1 - the derivative of y1 with respect to t
# dy2 - the derivative of y2 with respect to t
# fun - the name of the user-defined function that calculates the derivatives of the dependent variables
# soln - a matrix that contains the results of the numerical integration
# tspan- a vector containing the initial and final values of the independent variable
# time - a vector contain the independent variable at specified locations
# y0 - the initials value of y
# y1 - a vector containing the values of y1 at specified locations
# y1 - a vector containing the values of y1 at specified locations
###################################################################################
#                                   PROGRAM
###################################################################################
# Import required modules
import scipy.integrate
import numpy as np
import matplotlib.pyplot as plt
#
# Define ODEs
#
def fun(t, y):
    dy1=1-y[0]*np.exp(y[1])
    dy2=1+y[0]*y[1]
    return [dy1, dy2]
tspan = [0.0, 1.0]              # Set range of integration
y0=[1, 1]                       # Define initial conditions for dependent variables
tvalues=np.linspace(0, 1, 50)   # Define where variable will be calculated
#
# Call the ODE integrator
#
soln=scipy.integrate.solve_ivp(fun, tspan, y0, method='RK45', t_eval=tvalues, rtol=1.E-4, atol=1.E-6)
time=soln.t
y1=soln.y[0]                    # Retrieve calculated values for y1
y2=soln.y[1]                    # Retrieve calculated values for y2
#
# Plot results
#
plt.figure(figsize=(6,4.5))
plt.plot(time, y1, 'k-', time, y2, 'k—', linewidth=2)
plt.axis([0, 1, 0, 3])
plt.title('Solution for Ex3.2', fontsize=16)
plt.xlabel('t', fontsize=14)
plt.ylabel('y', fontsize=14)
plt.legend(('y1', 'y2'), loc='upper left', fontsize=14)
plt.show
plt.savefig("FigC3Ex2 Intro.jpg")
#                                   PROGRAM END
###################################################################################
```

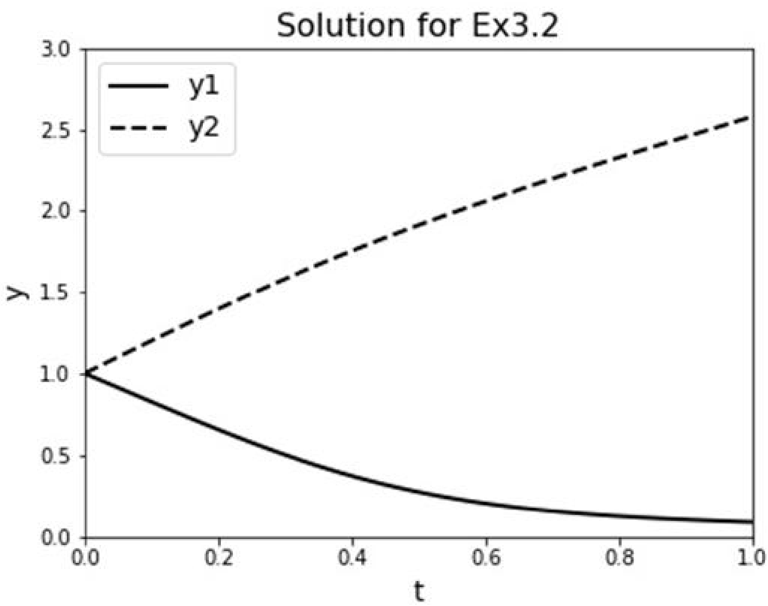


Figure 3.6.1 Plot of the solution for Ex3.2.

Note that this code can easily be used to solve another set of ODEs. Simply change the inlet problem specifications (i.e., the values of **tspan** and the initial conditions for **y**) and change the ODEs in user-defined function fun. As the number of ODEs change, you will also need to change the retrival and plot statements.

3.7 Modeling Examples

In this section, dynamic models are developed and results from these models are presented. Riggs[3] presents a systematic approach to process modeling, which includes a detailed model validation procedure. Bequette[4] analyzes dynamic process behavior and presents the development of a number of dynamic models.

The representation of a process for a control application involves the combination of models of the actuator (e.g., a control valve), process (e.g., a reactor) and sensor (e.g., thermocouple/transmitter) as shown in Figure 3.7.1. To affect a change in a process, the signal to the actuator, c, must be changed, which results in a change in the flow rate of the input variable, u, to the process, which, in turn, causes the process to change. Process changes cause

Figure 3.7.1 Block diagram for a dynamic control model.

the actual value of the output variable, y, to change, resulting in a change in the value of y measured by the sensor as y_s.

The actuator, process and sensor each have their own dynamic behavior. In this section, we first develop dynamic models for an actuator system and for several common types of sensors. Then, we will use these results to develop actuator/process/sensor models for a number of different processes. The actuator and sensor models are based on empirical approximations while the process models are phenomenological models, which are based on the conservation of mass and/or energy.

Actuator System Models. The actuator system (final control element) based on a control valve consists of the I/P converter, the instrument air system and the control valve. When a change in the analog signal to the I/P converter is made, the instrument air pressure to the valve changes. This causes the diaphragm in the control valve to expand or contract, which, in turn, causes the valve stem position to change, which affects the flow rate through the control valve. After the valve stem position changes, the flow rate through the valve changes very quickly. The dynamic response of the valve actuator to changes in the instrument air pressure applied to the valve is usually considerably slower than either the response of the I/P instrument air system or the flow through the valve. Because the dynamic response of the actuator is based on the instrument air working against a spring in a first-order manner, the dynamic behavior of the control valve can be represented as a linear first-order process (Section 6.3). Therefore, the flow rate, F, through a control valve can be represented by the following equation

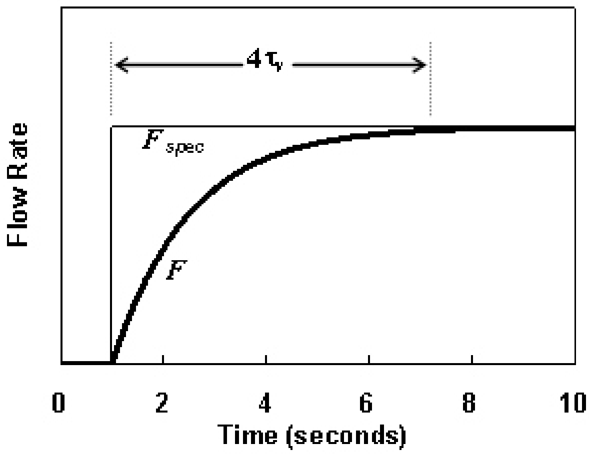


Figure 3.7.2 The dynamic response of the flow through a flow control loop in response to a step change in the specified flow rate, F_{spec}.

$$\frac{dF}{dt} = \frac{1}{\tau_v}(F_{spec} - F) \qquad \textbf{3.7.1}$$

where F_{spec} is the specified flow rate (input) and τ_v is the time constant of the valve. For this case, F is a dependent variable and t and F_{spec} are independent variables. τ_v is a process parameter and typically ranges between 3 and 15 seconds (Table 2.3) for cases in which a valve positioner is not used or the valve is not applied in a flow control loop. The dynamic behavior of a flow control loop or a control valve with a positioner can also be effectively modeled using Equation 3.7.1 where the time constant, τ_v, typically ranges between 0.5 and 2 seconds (Table 2.3). Usually the larger and older the control valve, the larger the value of τ_v. Figure 3.7.2 shows the resulting dynamic behavior for a step change in F_{spec} for a flow control loop with a τ_v of 1.5 sec. In fact, the flow rate changes for a control valve exhibit some deadtime and are not perfectly first-order responses, but the assumed first-order form shown is reasonably accurate and sufficient for the modeling examples considered here. The effective response time for a first-order process is $4\tau_p$ (i.e., time equal to 7s in Figure 3.7.2). In Section 6.3, it will be demonstrated that for a first-order process 98% of the total change occurs in $4\tau_p$ after the input change has been implemented.

An actuator system based on a variable speed pump is typically faster responding than a control valve-based actuator. From Table 2.3, the time constant for a variable speed pump is less than 0.1 s. Therefore, when modeling a variable speed pump, which are primarily used as actuators for bio-processes, it is reasonable to assume that the flow rate responds instantaneously to a change in the input to the variable speed pump, i.e.,

$$F \quad F_{spec} \tag{3.7.2}$$

because the dynamic response of most bio-processes is much slower than the response of a variable speed pump.

For certain systems, it is convenient to represent the combined effect of flow rate and a heat-transfer process as a lumped system. For example, when an input variable to a process is the rate of heat transfer (e.g., a reboiler for a distillation column), the actuator system for such a process usually involves the flow rate of a heat-transfer fluid (e.g., steam) and the transfer of heat through a contacting device (e.g., a heat exchanger). As a result, modeling the dynamic behavior of the actuator of such a process involves considering the dynamics of the flow control of the heat-transfer fluid and the dynamics of the heat-transfer process. For the flow control of a heat-transfer fluid, a flow controller is typically used and the previous analysis is valid. Therefore, the time constant for the flow controller is expected in the range of 0.5 to 2 seconds. The dynamics of the heat-transfer process are affected by the dynamics of heating or cooling the metal that passes the heat from the hot to the cold source and the transport delay for the process fluid to flow through the tubes of the heat exchanger. To increase the rate of heat transfer in a heat exchanger, the temperature of the metal tubes in the heat exchanger must be increased and the fluid must flow through the heat exchanger tubes. The thermal lag associated with changing the temperature of the metal tubes provides the dynamic lag associated with heat transfer. The time constant for changing the temperature of the heat-exchanger tubes is typically in the range of 1 to 6 seconds, while the transport delay is in the range of 5 to 30 seconds. Modeling the combined dynamics of the actuator and the heat-transfer system can also be represented as a first-order process given by

$$\frac{dQ}{dt} \quad \frac{1}{_{H}}(Q_{spec} \quad Q) \tag{3.7.3}$$

where Q is the rate of heat transfer, Q_{spec} is the specified heat-transfer rate, which is the input to the system, and $_{H}$ is the effective time constant for heat transfer (6 - 40 seconds). Q is a dependent variable and t and Q_{spec} are independent variables. $_{H}$ is a process parameter. Clearly, if a more accurate dynamic model of a heat-transfer system is required, detailed models of each element (e.g., flow control loops, heat transfer from the hot fluid to the metal of the tubes, accumulation of thermal energy in the metal tubes and heat transfer from the metal tubes to the cold fluid) should be used.

Sensor Models. Dynamic models for the primary sensors used in the CPI and biotechnology industries are considered here. The dynamic behavior of temperature sensors (e.g., a thermocouple or RTD) and level sensors (e.g., a differential pressure sensor) are well represented as linear first-order models similar to Equation 3.7.1. For example, for a temperature sensor, the following equation can be used to model the measured temperature, T_s,

$$\frac{dT_s}{dt} \quad \frac{1}{_{Ts}}(T \quad T_s) \tag{3.7.4}$$

where T is the actual process temperature and $_{Ts}$ is the time constant for the temperature sensor. Considering the sensor by itself, T_s is a dependent variable, T and t are independent variables, and $_{Ts}$ is a process parameter. $_{Ts}$ typically ranges from 6 to 20 seconds depending on the mass of metal in the thermowell, the thermal resistance between the temperature sensor and the inner wall of the thermowell and the velocity of process fluid past the thermowell.

Similarly, the model for a level sensor is given by

$$\frac{dL_s}{dt} = \frac{1}{\tau_{Ls}}(L - L_s) \tag{3.7.5}$$

where L_s is the measured level, L is the actual process level and τ_{Ls} is the time constant for the level sensor. Because a differential pressure measurement is typically used to determine the level and the dynamics of differential pressure measurements are relatively fast, τ_{Ls} is typically less than one second. As a result, it is usually reasonable to neglect the dynamics of the level sensor and assume that the level sensor makes an instantaneous measurement. Differential pressure sensors are used to measure flow rates, levels and system gauge pressure, each of which usually can also be assumed to have instantaneous sensor dynamics. For level sensors, improperly designed or partially plugged liquid lines from the pressure taps to the differential pressure cell can significantly increase the value of τ_{Ls}. Flow indicators, such as orifice, vortex shedding and magnetic flow meters, have time constants that are small enough that their dynamic response can be assumed instantaneous in most cases.

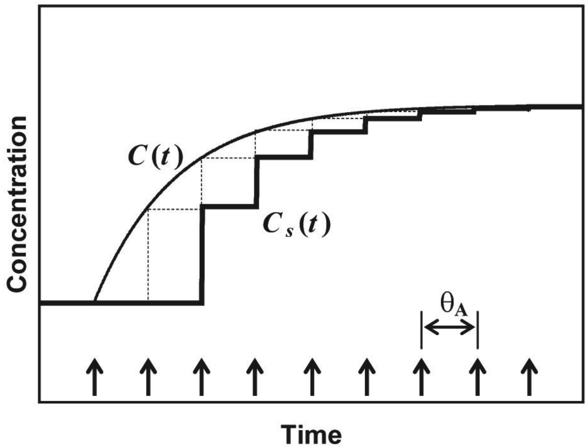


Figure 3.7.3 The dynamic response of the actual composition and the measured composition.

The most common composition analyzer in the CPI is the gas chromatograph (GC). Because a GC is a packed column, to obtain an analysis it requires time for the sample to migrate, by plug flow, the length of the separation column. Therefore, the dynamic model of a GC is a pure time delay, i.e.,

$$C_s(t) = C(t - \theta_A) \tag{3.7.6}$$

where $C_s(t)$ is the measured composition and $C(t - \theta_A)$ is the actual composition in the process θ_A time units earlier. θ_A is the cycle time for the analyzer or the time for the sample to flow through the packed column before it is measured by the detector. Figure 3.7.3 shows a plot of $C_s(t)$ and $C(t)$. $C(t)$ is the actual composition that responds to an input change and $C_s(t)$ is the measured composition, which is subject to analyzer delay. The arrows above the x-axis indicate when samples were injected into the GC. Note that the value $C_s(t)$ remains constant for θ_A time units, indicating the cycle time of the sensor. Also, note that in Figure 3.7.3 the measured value of $C(t)$ is delayed a time equal to $2\theta_A$ after the $C(t)$ begins to change. This initial delay before a change in the measured value of $C(t)$ is noted can range from θ_A to $2\theta_A$ depending on exactly when $C(t)$ starts to change in relation to when the samples are sent to the analyzer. For example, if the sample were taken shortly after $C(t)$ starts to change, the next analyzer result would indicate a small change in the measured value of $C(t)$. For the case shown in Figure 3.7.3, the initial delay is $2\theta_A$ and a large initial change in $C(t)$ results.

Table 2.4 also includes the time constants for several bio-sensors; therefore, the dynamic response of these sensors can be effectively modeled using an equation of the form of Equation 3.7.4. Electrode-based sensors (DO, Redox and pH) have time constants that typically range from 3 to 5 s while turbidity meters also have time constants in the range of 3-10 s. HPLC and FIA composition analyzer, which perform a batch analysis of a sample of a process stream, can be modeled as an analyzer delay using Equation 3.7.6 with the appropriate sensor deadtime. The following examples combine these models for sensors and actuators with process models based on balance equations to produce dynamic models for several processes commonly found in the CPI and bio-tech industries.

Example 3.3 CST Thermal Mixing Tank

Problem Statement. Develop the dynamic model equations and a MATLAB and a Python program that implements the model for the **continuous stirred tank (CST)** thermal mixer process shown in Figure 3.7.4. The process parameters and variables are defined as

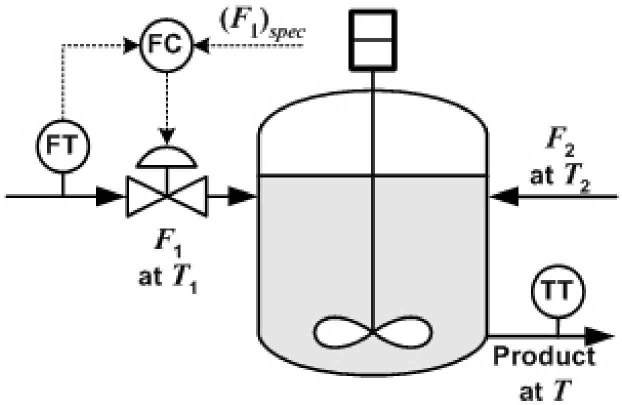


Figure 3.7.4 Schematic of a CST thermal mixing process.

- F_1 - mass flow rate of stream 1 (initially 5 kg/s)
- F_2 - mass flow rate of stream 2 (5 kg/s)
- M - mass of liquid in the mixer (100 kg)
- T - temperature of mixed liquid (initially 50°C)
- T_1 - temperature of stream 1 (25 C)
- T_2 - temperature of stream 2 (75 C
- t - time (s)
- $_v$ - the time constant for the flow controller on stream 1 (2 s).
- $_{Ts}$ - the time constant for the temperature sensor on the product stream (6 s).

At time equal to 10 seconds, a step change in the specified flow rate for stream 1 is made from 5 kg/s to 4 kg/s.

Solution. Assuming that the mixer volume is perfectly mixed, this process can be treated as a lumped parameter process, suitably represented by a macroscopic model. Applying Equation 3.4.3, noting that there are no chemical reactions and no heat transfer, yields

rate of accumulation of thermal energy	rate of convective heat transfer entering the system	rate of convective heat transfer leaving the system

The application of the balance equation to the CST thermal mixer yields

$$\frac{d[MC_v(T \quad T_{ref})]}{dt} \quad F_1C_p(T_1 \quad T_{ref}) \quad F_2C_p(T_2 \quad T_{ref}) \quad (F_1 \quad F_2)C_p(T \quad T_{ref})$$

Assuming perfect level control (i.e., $F_T \quad F_1 \quad F_2$ and M is constant), the heat capacity of each liquid stream is the same and the heat capacities at constant volume are equal to the heat capacities at constant pressure yields

$$M\frac{dT}{dt} \quad F_1T_1 \quad F_2T_2 \quad (F_1 \quad F_2)\,T \qquad \textbf{3.7.7}$$

To determine the initial product temperature, this equation is set equal to zero (i.e., steady-state conditions) and the initial process conditions are applied, i.e.,

$$M\frac{dT}{dt} \quad 0 \quad (5\,\mathrm{kg/s})(25^{\circ}\mathrm{C}) \quad (5\,\mathrm{kg/s})(75^{\circ}\mathrm{C}) \quad (10\,\mathrm{kg/s})\,T$$

Therefore, the initial product temperature is 50ºC.

The actuator is modeled using Equation 3.7.1 and the sensor is modeled using Equation 3.7.4. Therefore, the model equations used to represent the CST thermal mixer are

Actuator
$$\frac{dF_1}{dt} \quad \frac{1}{\;_{v}}(F_{1,spec} \quad F_1) \qquad \textbf{3.7.1}$$

Process
$$M\frac{dT}{dt} \quad F_1T_1 \quad F_2T_2 \quad (F_1 \quad F_2)\,T \qquad \textbf{3.7.7}$$

Sensor
$$\frac{dT_s}{dt} \quad \frac{1}{\;_{Ts}}(T \quad T_s) \qquad \textbf{3.7.4}$$

Figure 3.7.5 shows the resulting dynamic behavior of the measured temperature of the mixed liquid for this process. The process model is affected by changing $F_{1,spec}$, which changes F_1, resulting in a change in T, which is measured by the sensor as T_s. After a change in the input, the process reaches a new steady-state condition; therefore, this process is referred to as a **self-regulating**

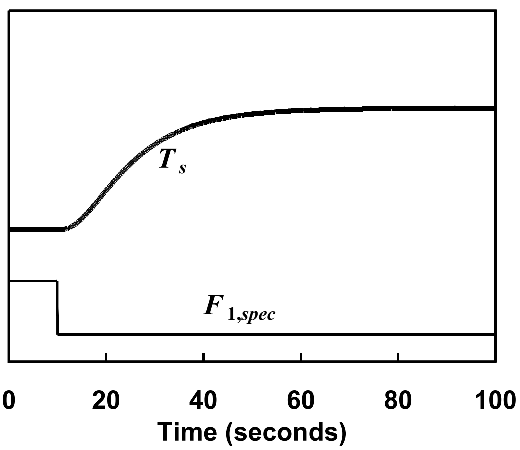


Figure 3.7.5 Dynamic response of the CST thermal mixer to a step change in $(F_1)_{spec}$.

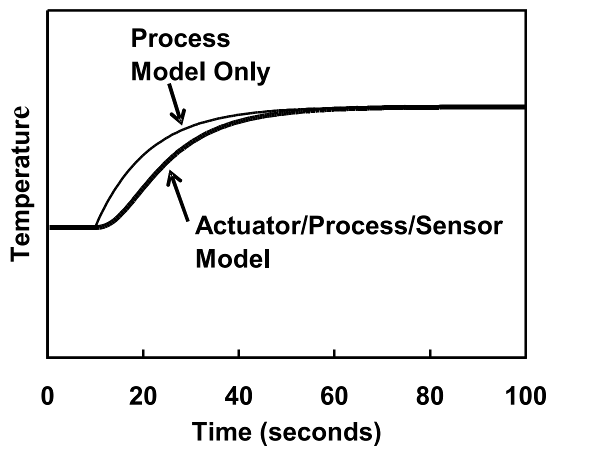


Figure 3.7.6 Comparison between the dynamic response of the model of the process by itself and a model of the actuator/process/sensor system.

process. These results were obtained by simultaneously integrating Equations 3.7.1, 3.7.7 and 3.7.4, which represent the combined effects of the flow controller, the thermal mixing process and the temperature sensor. Figure 3.7.6 shows a comparison between the actuator/process/sensor model and a model of the mixer process by itself. For the mixer model without the actuator and sensor dynamics included, the temperature of the product changes as soon as the specified value of F_1 changes while there is a noticeable time period before a significant change in the product temperature results for the actuator/process/sensor model. If the dynamics of the actuator and sensor are not included, the model has a very different dynamic response, particularly when the initial responses of the two models are compared. Therefore, when dynamic models of a process are developed for process control purposes, the dynamics of the actuator and sensor should be considered unless their time constants are significantly less than that of the process.

F_1, T and T_s are dependent variables and t, $F_{1,spec}$, F_2, T_1 and T_2 are independent variables. $_{v,}$ M and $_{Ts}$ are process parameters. This problem is exactly specified because the number of dependent variables (3) is equal to the number of equations (3). The visual basic simulation (Section 3.9) software that accompanies this text contains a simulator of this process based on the model developed here. The executable MATLAB and Python codes that solves this problem is listed below. Note from the codes below that an if-statement is used to change $F_{1,spec}$ after 10 s.

MATLAB Code:

```
function Ex3_3Tmixer
clear; clc;
t0=0; tf=100; y0=[5;50;50]; %Input specs(y(1)-F1spec; y(2)-T; y(3)-Ts)
% Call ode45 and store solution in soln
soln=ode45(@f,[t0,tf],y0,odeset('RelTol', 1.E-6, AbsTol, 1.E-6));
t=linspace(0,tf,100);        % Generate the values of t using linspace
y1=deval(soln,t,3);          % Retrive value of y from soln

plot(t,y1,'k-','LineWidth',2); % Plot t/y data and specify labels
xlabel('t'); ylabel('Outlet Temp (deg C)');
end
% User specified function for dydt for each dependent variable
function dydt=f(t,y)
tauv=2; F1spec=5; row=1; F2=5; T1=25; T2=75; tauTs=6; M=100;
if t>=10; F1spec=4; end
dydt(1)=(F1spec-y(1))/tauv;
dydt(2)=(y(1)*T1+F2*T2-(y(1)+F2)*y(2))/M;
dydt(3)=(y(2)-y(3))/tauTs;
dydt=dydt'                   % Return dydt as column vector
end
```

Python Code:

```
import scipy.integrate
import numpy as np
import matplotlib.pyplot as plt
def fun(t, y):                                              # User-defined function
    # Input specs(y[0]-F1spec; y[1]-T; y(3)-Ts)
    tauv, F1spec, F2, T1, T2, tauTs, M = 2, 5, 5, 25, 75, 6, 100     # Input data for problem
    if t>= 10:
```

```
        F1spec=4
    dy1=(F1spec-y[0])/tauv
    dy2=(y[0]*T1+F2*T2-(y[0]+F2)*y[1])/M
    dy3=(y[1]-y[2])/tauTs
    return [dy1, dy2, dy3]
tspan = [0, 100]
y0=[5, 50, 50]                    #  Input specs(y[0]-F1spec; y[1]-T; y[2]-Ts)
tvalues=np.linspace(0, 1, 100)
#  Call solve_ivp to integrate the model equations
soln=scipy.integrate.solve_ivp(fun, tspan, y0, method='RK45', t_value=tvalues, rtol=1.E-6, atol=1.E-6)
time=soln.t                       # Retrive independent variable
yResult=soln.y[2]                 # Retrive dependent variable
#  Plot results
plt.figure(figsize=(6,4.5))
plt.plot(time, yResult, 'k-', linewidth=2)
plt.axis([0, 100, 50, 54])
plt.title('Solution for Ex3.3', fontsize=16)
plt.xlabel('t', fontsize=14)
plt.ylabel('Ts', fontsize=14)
plt.show
```

Example 3.4 CST Composition Mixing Tank

Problem Statement. Develop a dynamic model for the CST composition mixing process shown in Figure 3.7.7. The process parameters and variables are given by

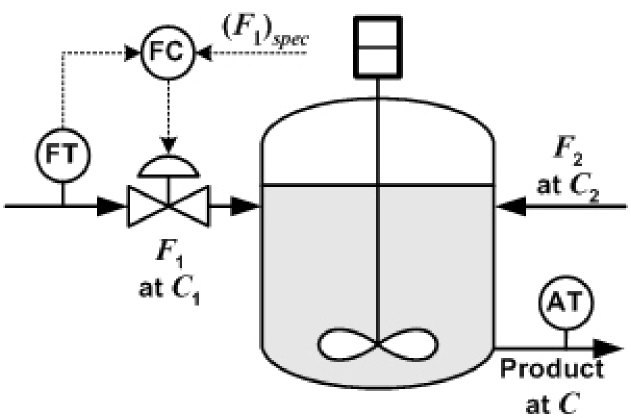


Figure 3.7.7 Schematic of a CST composition mixing process.

- C - the concentration of the component in the mixed stream (g mol/l)
- C_1 - the concentration of the component in stream 1 (0.5 g mol/l)
- C_2 - the concentration of the component in stream 2 (1.0 g mol/l)
- F_1 - the mass flow rate of stream 1 (initially 500 kg/min)
- F_2 - the mass flow rate of stream 2 (500 kg/min)
- t - time (min)
- V - the volume of the mixer (1000 l)
- - the constant density of the feed and product streams (1 kg/l)
- $_v$ - the time constant for the flow controller on stream 1 (2 s).
- $_A$ - analyzer deadtime (5 min).

At time equal to 5 minutes, a step change in the specified feed rate for stream 1 is made from 500 kg/min to 400 kg/min.

Solution. Assuming that the mixer volume is perfectly mixed and assuming that there are no chemical reactions occurring, a lumped parameter model based on a component mole balance can be used for this case.

rate of accumulation of moles within in the system	flow rate of moles entering the system	flow rate of moles leaving the system

Assuming a constant volume (V) in the mixer and applying the component mass balance equation in terms of concentration yields

$$V\frac{dC}{dt} \quad C_1(F_V)_1 \quad C_2(F_V)_2 \quad C[(F_V)_1 \quad (F_V)_2]$$

Because the mass flow rates are specified in the problem statement, it is necessary to convert from volumetric flow rates to mass flow rates assuming that the density is not affected by the concentration (i.e., $(F_V)_1 \quad F_1 /$ and $(F_V)_2 \quad F_2 /$). Converting the previous equation to mass flow rates yields

$$V\frac{dC}{dt} \quad F_1 C_1 \quad F_2 C_2 \quad (F_1 \quad F_2)C \qquad \textbf{3.7.8}$$

The model for the actuator is given by Equation 3.7.1 and the model for the composition analyzer is given by Equation 3.7.6. The equations that represent the CST composition mixer are as follows:

Actuator $$\frac{dF_1}{dt} \quad \frac{1}{\ _v}(F_{1,spec} \quad F_1) \qquad \textbf{3.7.1}$$

Process $$V\frac{dC}{dt} \quad F_1 C_1 \quad F_2 C_2 \quad (F_1 \quad F_2)C \qquad \textbf{3.7.8}$$

Sensor $$C_s(t) \quad C(t \quad \ _A) \qquad \textbf{3.7.6}$$

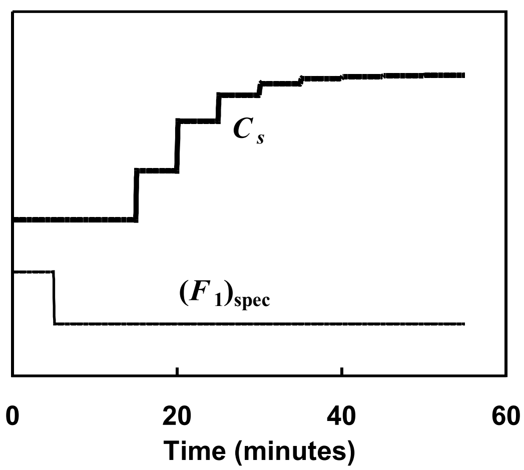


Figure 3.7.8 Dynamic response of the CST composition mixer for a step change in the specified value of F_1.

Figure 3.7.8 shows the simulated measured mixer composition as a function of time. This process is self-regulating because the open-loop response moves to a new steady-state condition after an input change. These results were obtained by simultaneously integrating Equations 3.7.1 and 3.7.8 while applying Equation 3.7.6. These equations represent the combined effects of the flow controller, mixer and the composition analyzer. F_1, C and C_s are dependent variables and t, $F_{1,spec}$, F_2, C_1 and C_2 are independent variables. $_v$, , V and $_A$ are process parameters. This problem is exactly specified because the number of dependent variables (3) is equal to the number of equations (3). The visual basic simulation (Section 3.9) simulation software that accompanies this text contains a simulator of this process based on the model developed here. A MATLAB and Python solution of these equations are not presented here because of the complexity of implementing analyzer deadtime using either MATLAB or Python.

Example 3.5 Level in a Tank

Problem Statement. Develop a dynamic model of a level in the tank shown in Figure 3.7.9. The process parameters and variables are given by

- A_c - cross-sectional area of the tank (0.3 m^2)
- L - the level of liquid in the tank (initially 2 meters)
- F_{in} - the mass flow rate of liquid into the tank (1.0 kg/s)
- F_{out} - the mass flow rate of liquid leaving the tank (initially 1.0 kg/s)
- t - time (s)
- - the fluid density (1 kg/l)
- $_v$ - the time constant for the control valve on F_{out} (5 s).

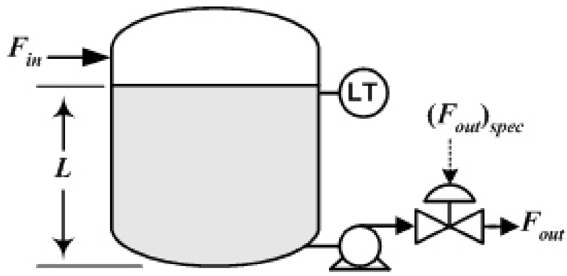


Figure 3.7.9 Schematic of a level in a tank process.

At time equal to 10 seconds, the specified value of F_{out} is changed from 1.0 to 0.9 kg/s. The dynamics of the level sensor are assumed instantaneous.

Solution. Applying a macroscopic mass balance to this process yields

rate of accumulation of mass within the system	flow rate of mass entering the system	flow rate of mass leaving the system

Expressing the total mass in the tank as $A_c L$ and recognizing that the mass flow rate into the tank is F_{in} and the mass flow rate out of the tank is F_{out}, the mass balance equation becomes

$$A_c \frac{dL}{dt} \quad F_{in} \quad F_{out} \tag{3.7.9}$$

Using Equation 3.7.1 to model the dynamic behavior of the actuator, neglecting the dynamics of the level sensor and using the process model for this system (Equation 3.7.9) yields the following system of ODEs that represent the dynamic model of this system.

$$\textbf{Actuator} \quad \frac{dF_{out}}{dt} \quad \frac{1}{_v} \quad F_{out,spec} \quad F_{out} \tag{3.7.1}$$

$$\textbf{Process} \qquad A_c \frac{dL}{dt} \quad F_{in} \quad F_{out} \tag{3.7.2}$$

$$\textbf{Sensor} \qquad L_s \quad L$$

Figure 3.7.10 shows the level in the tank as a function of time. This process does not move to a new steady-state condition; therefore, this system is referred to as a **non-self-regulating process**. These results were obtained by simultaneously integrating Equations 3.7.1 and 3.7.9, which represent the combined effects of the flow controller and the level process. F_{out}, L and L_s are dependent variables and t, $F_{out,spec}$ and F_{in} are independent variables. A_c and are process parameters. This is an example of an integrating process. This problem is exactly specified because the number of dependent variables (3) is equal to the number of equations (3). The simulation software that accompanies this text contains a simulator of this process based on the model developed here. The executable MATLAB code that solves this problem is listed below. Note from the code below that an if-statement is used to change $F_{1,spec}$ after 10 s. Because the dynamics of the level sensor are neglected, the measured tank level is equal to y(2).

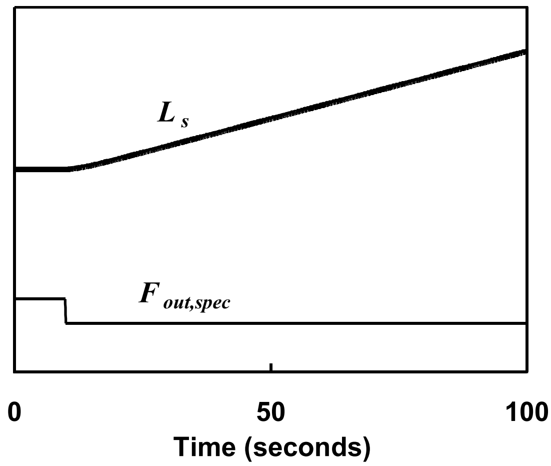


Figure 3.7.10 Dynamic response of a level in a tank to a step change in the flow rate leaving the tank.

MATLAB Code:

```
function Ex3_5Level
clear; clc;
t0=0; tf=100; y0=[1;2];
% Input specs (y(1)-Fout; y(2)-L;)
% Call ode45 and store solution in soln
soln=ode45(@f,[t0,tf],y0,odeset('RelTol', 1.E-6, AbsTol, 1.E-6));
t=linspace(0,tf,100);              % Generate the values of t using linspace
y1=deval(soln,t,2);                % Retrive value of y from soln
plot(t,y1,'k-','LineWidth',2);     % Plot t/y data and specify labels
xlabel('t (s)'); ylabel('Tank Level (m)');
end
% User specified function for dydt for each dependent variable
function dydt=f(t,y)
tauv=2; Foutspec=1; Fin=1; den=1; Ac=0.3; tauv=5;
if t >= 10; Foutspec=0.9; end
dydt(1)=(Foutspec-y(1))/tauv;
dydt(2)=(Fin-y(1))/den/Ac;
dydt=dydt'                         % Return dydt as column vector
end
```

Python Code:

```
import scipy.integrate
```

```
import numpy as np
import matplotlib.pyplot as plt
def fun(t, y):                                          # User-defined function
    tauv, Foutspec, Fin, den, Ac = 2, 1, 1, 1, 0.3      # Input data for problem
    if t>= 10:
        Foutspec=0.9
   dy1=(Foutspec-y[0])/tauv
   dy2=(Fin-y[0])/den/Ac
   return [dy1, dy2]
tspan = [0, 100]
y0=[1,2]                    % Input specs (y[0]-Fout; y[1]-L;)
tvalues=np.linspace(0, 100, 100)
# Call solve_ivp to integrate the model equations
soln=scipy.integrate.solve_ivp(fun, tspan, y0, method='RK45', t_value=tvalues, rtol=1.E-6, atol=1.E-6)
time=soln.t                                             # Retrive independent variable
yResult=soln.y[1]                                       # Retrive dependent variable
# Plot results
plt.figure(figsize=(6,4.5))
plt.plot(time, yResult, 'k-', linewidth=2)
plt.axis([0, 100, 0, 40])
plt.title('Solution for Ex3.5', fontsize=16)
plt.xlabel('t', fontsize=14)
plt.ylabel('L', fontsize=14)
plt.show
```

Example 3.7 Endothermic CSTR with First-Order Reaction

Problem Statement. Develop a dynamic model for the endothermic **CST**R shown in Figure 3.7.11. The process parameters and variables are given by

- C_A - reactant concentration (initially 0.25 g mol/l)
- C_{A_0} - feed composition (1.0 g mol/l)
- C_p - heat capacity of the reactor feed and product (1 cal/g-K)
- C_v - assumed equal to C_p
- E/R- normalized activation energy (20,000 K)
- F - mass feed rate and product rate (10 kg/s)
- k_0 - first-order rate constant (1.968 s^{-1})
- Q - heat addition rate (initially 1.20 cal/s)
- T - reactor temperature (initially 350 K)
- T_0 - feed temperature (400 K)

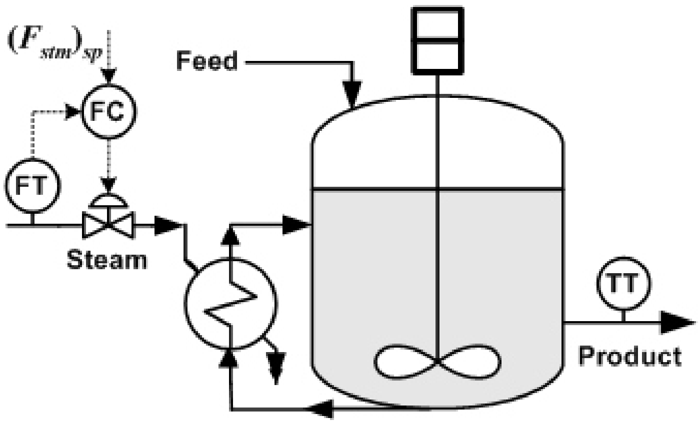


Figure 3.7.11 Schematic of an endothermic CSTR.

- t - time (s)
- V_r - reactor volume (100 l)
- - the constant density of the reactor feed and product (1 kg/l)
- H_{rxn}- heat of reaction (160,000 cal/g mol)
- $_H$ - the time constant for the combined heat transfer from the steam to the CSTR (10 s).
- $_{Ts}$- the time constant for the temperature sensor on the product stream (6 s).

These conditions represent the initial steady-state conditions. At time equal to 10 seconds, a step change in the specified heat addition rate (Q_{spec}) is made from 1.20 cal/s to 1.32 cal/s. Because heat addition by the heat exchanger involves a change in the flow rate of the steam as well as heat transfer in the heat exchanger, the dynamics of heat addition is modeled as a first-order process with a time constant, $_{H.}$ Develop a MATLAB and Python program that solves this problem.

Solution. Assuming that the CSTR is perfectly mixed, the process can be modeled as a lumped parameter process. For this model, the concentration of A and the temperature of the reactor are modeled using macroscopic mass and energy balances, respectively. The concentration of A can be modeled using Equation 3.4.2 by recognizing that the generation term is zero.

rate of accumulation of moles within in the system	flow rate of moles entering the system	flow rate of moles leaving the system	rate of consumption of moles by reaction

The total moles of A in the reactor is given by $V_r C_A$ which yields $V_r dC_A/dt$ as the accumulation because V_r is constant. Using a first-order irreversible reaction and Arrhenius temperature dependence for the reaction rate constant (i.e., $r \quad k_0 e^{\ E/RT} C_A$), the previous material balance equation becomes

$$V_r \frac{dC_A}{dt} \quad F_V (C_{A_0} \quad C_A) \quad V_r k_0 C_A e^{\ E/RT}$$

The reactor feed rate is given as a mass flow rate; therefore, $F_V \quad F/$ is used to convert the equation from a mass flow rate to a volumetric flow rate resulting in

$$V_r \frac{dC_A}{dt} \quad \frac{F}{\ }(C_{A_0} \quad C_A) \quad V_r k_0 C_A e^{\ E/RT} \qquad \textbf{3.7.10}$$

Equation 3.4.3 is applied to model the reactor temperature:

rate of accumulation of thermal energy	rate of convective heat transfer entering the system	rate of convective heat transfer leaving the system

net rate of energy generation by chemical reaction	net rate of heat transfer through the boundaries of the system

The total mass in the reactor is V_r ; therefore, the rate of accumulation of thermal energy is $V_r \quad C_v \frac{dT}{dt}$. The energy generated by the reaction is $V_r \quad H_{rxn} r$. As a result, Equation 3.4.3 yields

$$V_r \quad C_v \frac{dT}{dt} \quad F C_p (T_0 \quad T) \;\; V_r \quad H_{rxn} C_A \, k_0 \, e^{\;E/RT} \quad Q \qquad \textbf{3.7.11}$$

Using Equation 3.7.3 to represent the actuator, Equation 3.7.4 to model the temperature sensor and Equations 3.7.10 and 3.7.11 to model the process, the actuator/process/sensor model for this process is given as:

Lumped Actuator $$\frac{dQ}{dt} \quad \frac{1}{\quad}(Q_{spec} \quad Q) \qquad \textbf{3.7.3}$$

Process $$V_r \frac{dC_A}{dt} \quad \frac{F(C_{A_0} \quad C_A)}{\quad} \;\; V_r \, k_0 \, C_A \, e^{\;E/RT} \qquad \textbf{3.7.10}$$

$$V_r \quad C_p \frac{dT}{dt} \quad F C_p (T_0 \quad T) \;\; V_r \quad H_{rxn} C_A \, k_0 \, e^{\;E/RT} \quad Q \qquad \textbf{3.7.11}$$

Sensor $$\frac{dT_s}{dt} \quad \frac{1}{_{Ts}}(T \quad T_s) \qquad \textbf{3.7.4}$$

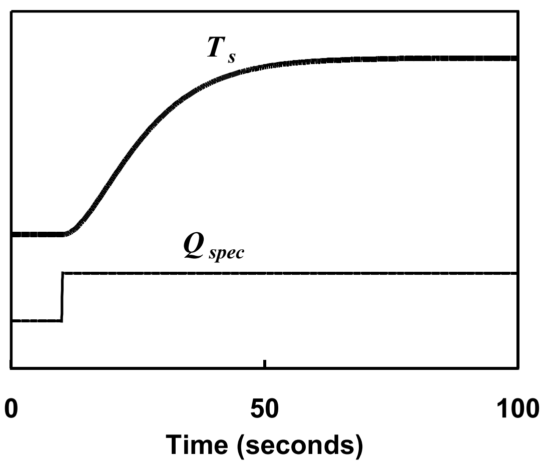


Figure 3.7.12 Dynamic response of the endothermic CSTR for a step increase in the heat addition rate to the reactor.

Figure 3.7.12 shows the measured reactor temperature as a function of time for a step change in heat addition rate. These results were obtained by simultaneously integrating Equation 3.7.3, Equation 3.7.10, Equation 3.7.11 and Equation 3.7.4. Note that the reactant concentration must be modeled because C_A appears in Equation 3.7.11. Q, C_A, T and T_s are dependent variables and t, Q_{spec}, T_0, C_{A_0}, F and V_r are independent variables. $_H$, , C_p, H, k_0, E, R and $_{Ts}$ are process parameters. This problem is exactly specified because the number of dependent variables (4) is equal to the number of equations (4). The visual basic simulation (Section 3.9) simulation software that accompanies this text contains a simulator of this process based on the model developed here. The executable MATLAB and Python codes that solve this problem are listed below. Note from the code below that an if-statement is used to change $F_{1,spec}$ after 10 s.

MATLAB Code:

```
function Ex3_6CSTR
clear; clc;
 %Input specs(y(1)-Q; y(2)-CA; y(3)=T; y(4)=Ts)
```

```
t0=0; tf=100; y0=[1.2e6;0.25;350;350];
% Call ode45 and store solution in soln
soln=ode15s(@f,[t0 tf],y0,odeset('RelTol', 1.E-6, AbsTol, 1.E-6));
x=linspace(0,tf,100);        % Generate the values of t using linspace
y1=deval(soln,x,3);          % Retrive value of y from soln
plot(x,y1,'k-','LineWidth',2); % Plot t/y data and specify labels
xlabel('t (s)'); ylabel('Reactor Temp(K)');
end
% Use specified function for dydx for each dependent variable
function dydt=f(t,y)
tauH=10; Qspec=1.2e6; den=1; tauTs=6; F=10; CA0=1; Cp=1; ER=2e4; T0=400;
Vr=100; k0=1.96765e24; DHr=160000;
if t >= 10; Qspec=1.32e6; end
dydt(1)=(Qspec-y(1))/tauH;
dydt(2)=(F*(CA0-y(2))/den-Vr*k0*y(2)*exp(-ER/y(3)))/Vr;
dydt(3)=(F*Cp*(T0-y(3))-Vr*DHr*y(2)*k0*exp(-ER/y(3))+y(1))/(Vr*den*Cp);
dydt(4)=(y(3)-y(4))/tauTs;
dydt=dydt';
end
```

Python Code:

```
import scipy.integrate
import numpy as np
import matplotlib.pyplot as plt
def fun(t, y):       # User-defined function
   tauH, Qspec, den, tauTs, F, CA0, Cp = 10, 1.2E6, 1, 6, 10, 1, 1      # Input data for problem
   ER, T0, Vr, k0, DHr = 2E4, 400, 100, 1.96765E24, 160000
   if t>= 10:
      Qspec=1.32E6
   dy1=(Qspec-y[0])/tauH
   dy2=(F*(CA0-y[1])/den-Vr*k0*y[1]*np.exp(-ER/y[2]))/Vr
   dy3=(F*Cp*(T0-y[1])/den-Vr*DHr*y[1]*k0*np.exp(-ER/y[2])+y[0])/(Vr*den*Cp)
   dy4=(y[2]-y[3])/tauTs
   return [dy1, dy2, dy3, dy4]
tspan = [0, 100]
y0=[1.2E6, 0.25, 350, 350]           # Input specs(y[0]-Q; y[1]-CA; y[2]=T; y[3]=Ts)
tvalues=np.linspace(0, 100, 100)
# Call solve_ivp to integrate the model equations
soln=scipy.integrate.solve_ivp(fun, tspan, y0, method='RK45', t_value=tvalues, rtol=1.E-6, atol=1.E-6)
time=soln.t                                  # Retrive independent variable
yResult=soln.y[3]                            # Retrive dependent variable
# Plot results
plt.figure(figsize=(6,4.5))
plt.plot(time, yResult, 'k-', linewidth=2)
plt.axis([0, 100, 350, 360])
plt.title('Solution for Ex3.5', fontsize=16)
plt.xlabel('t', fontsize=14)
plt.ylabel('L', fontsize=14)
plt.show
```

Example 3.7 Exothermic CSTR

Problem Statement. Develop a dynamic model for an exothermic CSTR. The exothermic CSTR considered here is identical to the endothermic CSTR shown in Figure 3.7.11 except that an exothermic reaction occurs in the reactor and heat is removed in the heat exchanger by cooling water. The process parameters are also the same except that Q_{spec} = -1,173,540 cal/s; C_A = 0.6415 g mol/l; T = 340 K; and H = -160,000 cal/g mol.

Solution. The equations used to model this process are all exactly the same as the equations used to model the endothermic CSTR with several of the model parameters and independent variables set to different values. Figure 3.7.13 shows the open-loop response of the exothermic CSTR to a 1% increase and a 1% decrease in the specified heat removal rate for the heat exchanger. This process is an example of an **open-loop unstable process** because a decrease in the heat removed by the heat exchanger causes a temperature runaway and an increase extinguishes the reaction. Not all exothermic reactors are open-loop unstable processes. For the case in which the reaction is extinguished (i.e., an increase in Q_{spec}), the outlet temperature is determined almost exclusively by the inlet feed temperature and the heat removal rate because the reaction rate is essentially zero. This process is another example of a non-self-regulating process.

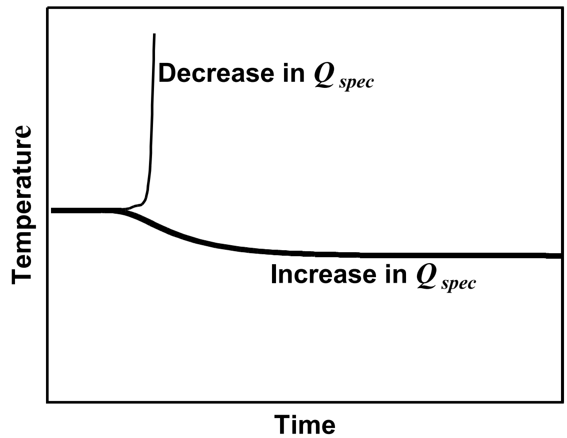


Figure 3.7.13 Open-loop response of the exothermic CSTR to a 1% increase and decrease in the heat removal rate.

Example 3.8 Fed-Batch Reactor for Ethanol Production

Problem Statement. Develop a dynamic model for the fed-batch fermentation process used to produce ethanol (EtOH) shown in Figure 3.7.14. Grains, such as corn, are converted to glucose by enzyme reactions and provide the feed for the fed-batch reactor modeled here. EtOH is recovered from the product produced by the fed-batch reactor by distillation. EtOH is added to gasoline to levels of approximately 10 percent by volume to produce gasohol, which is a common type of motor gasoline.

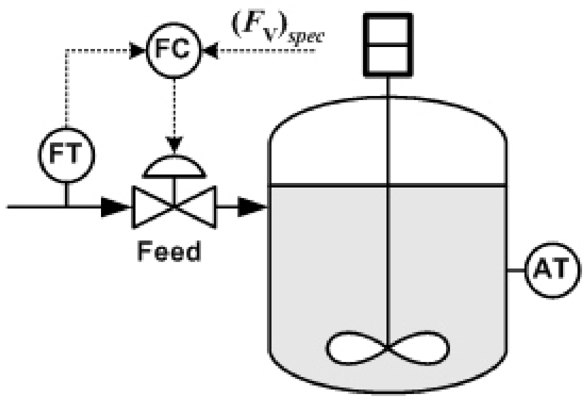


Figure 3.7.14 Schematic of a fed-batch reactor.

The process variables and parameters for this system are

- F_V - volumetric feed rate to the fermentor (l/h)
- k_p - inhibition constant (16.0 g/l)
- k_p - inhibition constant (71.5 g/l)
- k_s - Monod constant (0.22 g/l)

- k_s - Monod constant (0.44 g/l)
- m_P - total mass of EtOH in the fermentor (g).
- m_S - total mass of glucose in the fermentor (g).
- m_x - total mass of yeast cells in the fermentor (g).
- P - concentration of EtOH in the fermentor (g/l)
- P_i - initial concentration of EtOH in the fermentor (0.0 g/l)
- S - concentration of glucose in the fermentor (g/l).
- S_i - initial concentration of glucose in the fermentor (150 g/l).
- S_0 - concentration of glucose in the feed to the fermentor (150 g/l)
- t - time (h)
- V - volume of reaction mixture in the fermentor (l)
- V_i - initial volume of reaction mixture in the fermentor (5000 l)
- x - yeast cell concentration (g/l)
- x_i - initial yeast cell concentration in the fermentor (1.0 g/l)
- Y_{xS} - yield coefficient (0.1 g-cell produced per g of glucose consumed)
- $_0$ - maximum specific productivity (1.0 h^{-1})
- - specific productivity (h^{-1})
- $_0$ - maximum specific growth rate [0.408 g-cells produced/(g-cell h^{-1})]
- - specific growth rate [g-cells produced/(g-cells h^{-1})]
- $_v$ - the time constant for the flow control loop on feed stream (2 s)
- $_s$ - the time constant for the EtOH analyzer (density measurement) on the product (20 s)

The duration of this fed-batch process is 35 h. For the initial 13 h, feed is not added to the fed-batch fermentor (V=5000 l) to allow the yeast to grow to a sufficient level. After 13 h, the feed is added at a constant rate of 4,000 l/h for an additional 17 h. After 30 h, the feed rate is set to zero, holding the volume constant from that point. The flow controller on the feed to the fermentor has an effective time constant of 2 s. An on-line density measurement is taken to infer the EtOH concentration in the fermentor and this sensor has an effective time constant of 20 s. Develop a MATLAB and Python program that solves this problem.

Solution. Nutrients containing glucose and other growth medium for the yeast are fed to the reactor. The yeast consumes the glucose and nutrients to produce ethanol. To model this process, it is necessary to develop models for the concentration of yeast cells, glucose (substrate) and ethanol (product), but we will not apply balance equations for the other nutrients. In the case of bio-reactors, even though biological reactions are occurring in these systems, it is conventional to perform material balances instead of mole balances because in most cases the stoichiometric coefficients of the reactions are not known. As a result, conversion constants are used to express the relationship between the cell growth, the consumption of the substrate and the generation of the product.

Let's consider the balances for cell growth in this bio-reactor assuming that the reactor is well mixed so that we can apply a lumped parameter model. Because after the inoculum is added to the batch, cells are not added or removed from the system, the mass balance for cells is

rate of accumulation of cell mass within in the system	rate of generation of cell mass by growth

During the incubation period, the volume (V) is constant. The rate of accumulation of cell mass during the incubation periods is equal to $V\dfrac{dx}{dt}$, where x is the cell concentration. Therefore, the balance for the cell mass becomes

$$V\frac{dx}{dt} \quad xV$$

where is the specific growth rate, i.e., the growth rate of the cell mass in mass per time is xV.

After the initial incubation period, substrate is added to the reactor; therefore, based on an overall mass balance equation during this period, $dV/dt=F_V$, which gives $V \quad F_V t \quad V_0$ when F_V is constant. The rate of accumulation of cell mass during this period is equal to $\dfrac{d(xV)}{dt}$. Therefore, the material balance for cell mass is given by the following equation, which includes the effect of generation of cell mass and changes in the volume of material in the fermentor due to feed addition.

$$\frac{dm_x}{dt} \quad \frac{d}{dt}(Vx) \quad V\frac{dx}{dt} \quad xF_V \qquad xV$$

Note that as before no cells are added to or removed from the bio-reactor. Solving for dx/dt yields

$$\frac{dx}{dt} \quad x \quad \frac{F_V x}{V}$$

The fermentation reactions are modeled using the specific growth of the yeast cells ( ) using Monod kinetics[7] modified for the effect of product inhibition

$$\frac{_0}{1 \quad P/k_P} \quad \frac{S}{k_S \quad S}$$

Now consider the balance for the glucose (substrate). During the incubation period, glucose is consumed by cell growth and the mass balance for glucose becomes

rate of accumulation of mass within in the system - rate of consumption of mass by reaction

The accumulation term for this case is $V\dfrac{dS}{dt}$ because the volume is constant. The rate of glucose consumption by the cells is related to the cell growth by xV / Y_{xS}, where Y_{xS} is a proportionality constant between cell growth and glucose consumption. Therefore, the mass balance equation for glucose is given as

$$\frac{dm_S}{dt} \quad V\frac{dS}{dt} \qquad \frac{xV}{Y_{xS}}$$

After the incubation period, because glucose is added to the bio-reactor and consumed by the cells, the following balance equation applies in this case.

rate of accumulation of mass within in the system	flow rate of mass entering the system	- rate of consumption of mass by reaction

The accumulation term for this case is $\frac{d(VS)}{dt}$, the flow rate of glucose entering the reactor is $F_V S$ and the rate of glucose consumption by the cells is related to the cell growth by xV / Y_{xS}. Therefore, the material balance for glucose is given by the following equation.

$$\frac{dm_S}{dt} \quad \frac{d}{dt}(VS) \quad V\frac{dS}{dt} \quad S\frac{dV}{dt} \quad F_V S_0 \quad \frac{xV}{Y_{xS}}$$

Using the dynamic description of the volume of the fermentor and rearranging results in

$$\frac{dS}{dt} \quad \frac{F_V (S_0 \quad S)}{V} \quad \frac{x}{Y_{xS}}$$

For the product, the specific production rate of EtOH ( ) is given by

$$\frac{{}_0}{1 \quad P / k_P} \quad \frac{S}{k_S \quad S}$$

During the incubation period, EtOH is produced by the cells. Therefore, the mass balance for EtOH is given by

rate of accumulation of mass within in the system	rate of generation of mass by reaction

The accumulation term for this case is $V\frac{dP}{dt}$ because the volume is constant. The rate of EtOH generation by the cells is related to the cell growth by xV. Therefore, the mass balance equation for EtOH is given as

$$\frac{dm_P}{dt} \quad \frac{d}{dt}(VP) \quad V\frac{dP}{dt} \quad xV$$

After the incubation period, the previous equation applies although the volume is a function of time. The accumulation term for this case is $\frac{d(VP)}{dt}$. Note that EtOH is neither added to or removed from the bio-reactor. Therefore, the material balance for the product (EtOH) is given by the following equation.

$$\frac{dm_P}{dt} \quad \frac{d}{dt}(VP) \quad V\frac{dP}{dt} \quad P\frac{dV}{dt} \quad xV$$

Using the dynamic description of the volume of the fermentor and rearranging results in

$$\frac{dP}{dt} \quad x \quad \frac{F_V P}{V}$$

Because the process is slow responding, the actuator and sensor dynamics based on a density measurement can be neglected in this case (time constants of the order of seconds can be neglected when the time scale of the simulation is 33 h). Therefore, the model equations that are used to represent the fed-batch fermentor are given by

Actuator $$\frac{dF_V}{dt} \quad (F_{V,spec} \quad F_V)/ \quad {}_V$$

Process $$\frac{dx}{dt} \quad x \qquad \frac{dS}{dt} \quad \frac{x}{Y_{xS}} \qquad \frac{dP}{dt} \quad x \qquad (t \quad 13\,\text{h};\; t \quad 30\,\text{h})$$

$$\frac{dV}{dt} \quad F_V \qquad \frac{dx}{dt} \quad x \quad \frac{F_V x}{V} \qquad \frac{dS}{dt} \quad \frac{F_V (S_0 \quad S)}{V} \quad \frac{x}{Y_{xS}} \qquad \frac{dP}{dt} \quad x \quad \frac{F_V P}{V} \qquad (13\text{ h} > t > 30\text{ h})$$

where

$$\frac{{}_0}{1 \quad P/k_P} \quad \frac{S}{k_S \quad S} \qquad \text{and} \qquad \frac{{}_0}{1 \quad P/k_P} \quad \frac{S}{k_S \quad S}$$

Sensor $$\frac{dP_S}{dt} \quad \frac{P \quad P_S}{{}_S}$$

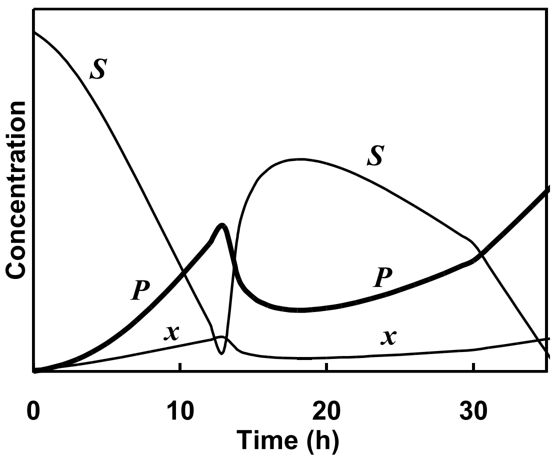


Figure 3.7.15 Dynamic behavior of the fed-batch reactor where *S* is the glucose concentration, *P* is the EtOH concentration and *x* is the concentration of yeast cells.

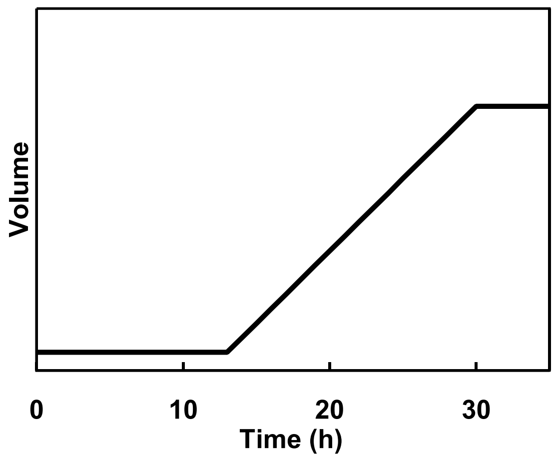


Figure 3.7.16 Volume of the fed-batch reactor as a function of time.

Figure 3.7.15 shows the resulting dynamic behavior of the glucose concentration, the yeast cell concentration and the EtOH concentration in the fermentor during the 33 h batch. These results were obtained by simultaneously integrating the model equations that describe this fed-batch reactor. Initially, the fermentor has a small volume when the yeast is added. After 13h, feed is added at a constant rate until 30 h (Figure 3.7.16). From 30 to 33 h, the process is operated as a batch process. The concentrations of yeast cells and EtOH drop sharply when the feed to the fermentor is started (13 h), but both recover afterwards. This problem is exactly specified because the number of dependent variables (6) is equal to the number of equations (6). The executable MATLAB and Python codes that solves this problem are listed below. Note from the code below that an if-statement is used to change the ODEs when t is greater than 17 h and less than 30 h.

MATLAB Code:

```
function Ex3_8BatchFerment
clear; clc;
t0=0; tf=33; y0=[0;5000;1;150;0;0];
% Input specs (y(1)-FV; y(2)-V; y(3)=x; y(4)=S; y(5)-P; y(6)-Ps)
    % Call ode45 and store solution in soln
soln=ode45(@f,[t0,tf],y0,odeset('RelTol', 1.E-6, AbsTol, 1.E-6));
t=linspace(0,tf,100);            % Generate the values of t using linspace
y1=deval(soln,t,6);              % Retrive value of y from soln
plot(t,y1,‘k-’,‘LineWidth’,2);   % Plot t/y data and specify labels
xlabel(‘t (h)’); ylabel(‘EtOH concentration (g/l)’);
end
% User specified function for dydt for each dependent variable
function dydt=f(t,y)
tauv=2/3600;mu0=0.408;kp=16;ks=0.22;esp0=1;kpp=71.5;ksp=0.44;
    Ysx=0.1;tauS=20/3600.;S0=150;
mu=(mu0/(1+y(5)/kp))*(y(4)/(ks+y(4)));
esp=(esp0/(1+y(5)/kpp))*(y(4)/(ksp+y(4)));
dydt(1)=0;
dydt(2)=0;
dydt(3)=mu*y(3);
dydt(4)=-mu*y(3)/Ysx;
dydt(5)=esp*y(3);
dydt(6)=(y(5)-y(6))/tauS;
if t >= 13 && t<30;
dydt(1)=(4000-y(1))/tauv;
dydt(2)=y(1);
dydt(3)=mu*y(3)-y(1)*y(3)/y(2);
dydt(4)=y(1)*(S0-y(4))/y(2)-mu*y(3)/Ysx;
dydt(5)=esp*y(3)-y(1)*y(5)/y(2);
end
dydt=dydt’;        % Return dydt as column vector
end
```

Python Code:

```
import scipy.integrate
import numpy as np
import matplotlib.pyplot as plt
```

```
def fun(t, y):        #  User-defined function
tauv, mu0, kp, ks, esp0, kpp, ksp = 2/3600, 0.408, 16, 0.22, 1, 71.5, 0.44
Ysx, tauS, S0 = 0.1, 20/3600, 150
mu=(mu0/(1+y[4]/kp))*(y[3]/(ks+y[3]))
esp=(esp0/(1+y[4]/kpp))*(y[3]/(ksp+y[3]))
dy1, dy2 = 0, 0
dy3=mu*y[2]
dy4=-mu*y[2]/Ysx
dy5=esp*y[2]
dy6=(y[4]-y[5])/tauS
if t>= 13 and t<=30:
   dy1=(4000-y[0])/tauv
   dy2=y[0]
   dy3=mu*y[2]-y[0]*y[2]/y[1]
   dy4=y[0]*(S0-y[3])/y[1]-mu*y[2]/Ysx
   dy5=esp*y[2]-y[0]*y[4]/y[1]
return [dy1, dy2, dy3, dy4, dy5, dy6]
tspan = [0, 33]
# Input specs (y[0]-FV; y[1]-V; y[2]=x; y[3]=S; y[4]-P; y[5]-Ps)
y0=[0, 5000, 1, 150, 0, 0]
tvalues=np.linspace(0,33, 100)
#  Call solve_ivp to integrate the model equations
soln=scipy.integrate.solve_ivp(fun, tspan, y0, method='RK45', t_value=tvalues, rtol=1.E-6, atol=1.E-6)
time=soln.t                                        # Retrive independent variable
yResult=soln.y[5]                                  # Retrive dependent variable
#  Plot results
plt.figure(figsize=(6,4.5))
plt.plot(time, yResult, 'k-', linewidth=2)
plt.axis([0, 33, 0, 150])
plt.title('Solution for Ex3.8', fontsize=16)
plt.xlabel('t', fontsize=14)
plt.ylabel('Ps', fontsize=14)
plt.show
```

Example 3.9 Dissolved Oxygen (DO) Process

Problem Statement. Certain bio-processes operate under anaerobic or near anaerobic conditions, i.e., no oxygen or a small level of oxygen in the broth, while for other bio-processes the cells depend on a supply of oxygen to produce the desired products (aerobic conditions). Develop a dynamic model for the dissolved oxygen process in the aerobic batch bio-reactor shown in Figure 3.7.17. Air is injected in the bottom of the bio-reactor through a sparger by a variable speed air compressor (blower). The sparger creates a large number of relatively small bubbles and the mixer further increases the number of bubbles and reduces their size, increasing the transport of oxygen to the broth. The transport of oxygen to the broth is controlled by the transport from the bubble surface to the bulk liquid in the broth. In addition, oxygen is consumed by the cells as they consume the substrate. Assume the same Monod kinetics used in Example 3.8. Determine the DO concentration as a function of time for a 10% increase in the air feed rate. Following are the process variables and parameters for this system. Develop a MATLAB and Python program that solves this problem.

- C_{O_2} - concentration of O_2 in the reaction broth (initially 1.10×10^{-4} g-moles/l)
- $C^*_{O_2}$ - saturated concentration of O_2 in the broth (2.20×10^{-4} g-moles/l)
- $C_{O_2,s}$ - the measurement of the O_2 concentration in the broth (initially 1.1×10^{-4} g-moles/l)
- F_{air} - the volumetric flow rate of air to the bio-reactor (initially 500 cfm)
- K_{O_2} - cellular uptake of O_2 (1.98 g-moles O_2/g-cells)
- k_La - the overall liquid phase mass transfer coefficient for transport from the bubble surface to the bulk broth (initially 0.25 s^{-1})
- T - broth temperature (35ºC)
- t- time (s)
- V - the volume of broth in the bio-reactor (1000 l)
- x - constant cell concentration in the bio-reactor (0.25 g/l)
- $_{max}$ - maximum specific growth rate (5.56×10^{-5} s^{-1})
- $_s$ - the time constant of the DO sensor (30 s)

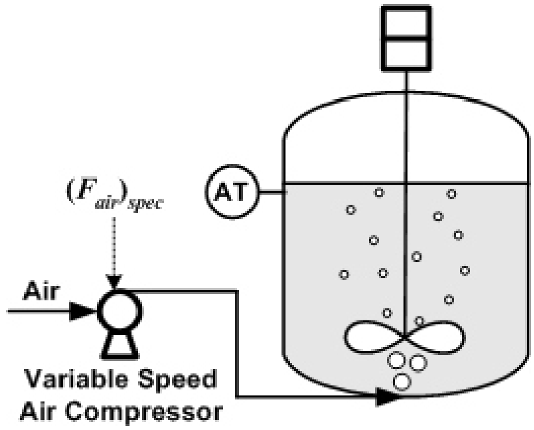


Figure 3.7.17 Schematic of a DO process.

Solution. The DO in the broth in this bio-reactor can be modeled by applying a dynamic mole balance on O_2. For this process, oxygen is transferred from the injected air to the broth and consumed from the broth by the cells in the bio-reactor. Therefore, the mole balance equation for this system is

rate of accumulation of moles within in the system | flow rate of moles entering the system | rate of consumption of moles by reaction

Based on a constant volume, the accumulation term is $V_r dC_{O_2} / dt$. Therefore, the unsteady-state O_2 mole balance on the broth in the reactor can be expressed in terms of moles O_2 per unit time as

$$V \frac{dC_{O_2}}{dt} \quad f_{air,O_2} \quad Vr_{O_2}$$

where f_{air,O_2} is the rate of mass transfer of O_2 from the air into the broth and Vr_{O_2} is the rate of consumption of O_2 by the cells. The mass transfer of O_2 from the air to the broth is controlled by the liquid phase transport from the surface of the air bubbles to the bulk broth (k_La). Here $C^*_{O_2}$ is the equilibrium concentration of O_2 at the surface of the bubbles and is a function of the temperature, pressure and composition of the broth. Therefore, the transport of O_2 from the surface of the bubbles into the bulk broth is a function of the concentration driving force, the volume of broth, and the overall liquid phase mass transfer coefficient, i.e.,

$$f_{air,O_2} \quad Vk_L a \ C^*_{O_2} \quad C_{O_2}$$

The liquid phase mass transfer coefficient (k_La) is a function of the air flow rate. It is assumed here that over a limited operating range, k_La is a linear function of the air feed rate, i.e.,

$$k_L a \quad 0.25 \quad 0.001(F_{air} \quad 500)$$

where F_{air} is expressed in standard cubic feet per minute. Although $C^*_{O_2}$ is specified for this problem, it can be estimated using the Henry's Law constant, which is primarily a function of temperature but also a function of pressure and broth composition.

The consumption of O_2 is proportional to the cell growth rate, i.e.,

$$r_{O_2} \quad K_{O_2} \quad {}_{\max} x$$

This cell growth expression is based on assuming that the substrate concentration is relatively high. In addition, it is assumed that x is constant because the dynamics for x are much slower than for dissolved oxygen (i.e., a pseudo-steady-state assumption for x). Substituting these relationship into the O_2 balance equation and simplifying yields

$$\frac{dC_{O_2}}{dt} \quad k_L a \; C^*_{O_2} \quad C_{O_2} \quad K_{O_2} \quad {}_{\max} x$$

Note that this formulation assumes that the cell concentration (x) is constant. Because dynamics of the DO concentration are so much faster than the dynamics of cell growth, this is valid assumption. The dynamics of the variable speed air compressor are much faster than the process in this case, and as a result, are assumed to be instantaneous. The sensor is a DO electrode ($_s$=30 s). The model equations for the actuator, process and sensor are given by

Actuator $$F_{air} \quad (F_{air})_{spec}$$

Process $$k_L a \quad 0.25 \quad 0.001(F_{air} \quad 500)$$

$$\frac{dC_{O_2}}{dt} \quad k_L a \; C^*_{O_2} \quad C_{O_2} \quad K_{O_2} \quad {}_{\max} x$$

Sensor $$\frac{dC_{O_2,s}}{dt} \quad \frac{1}{{}_s} \; C_{O_2} \quad C_{O_2,s}$$

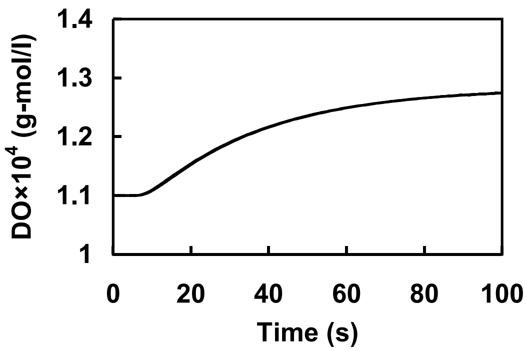


Figure 3.7.18 Response of the dissolved oxygen concentration for the DO process to a 10% step increase in the air feed rate.

Figure 3.7.18 shows the response of the measured value of the dissolved oxygen concentration for a 10% increase in the feed rate at time equal to 6 s. These results were obtained by simultaneously integrating the model equations for the actuator, process and sensor from the initial steady-state conditions forward in time. Note that the dissolved oxygen concentration is self-regulating. F_{air}, $k_L a$, C_{0_2} and $C_{O_2,s}$ are the dependent variables and x, t and $(F_{air})_{spec}$ are the independent variables. $C^*_{O_2}$, K_{O_2}, $_{\max}$ and $_s$ are process parameters. The executable MATLAB and Python codes that solves this problem are listed below. Note from the codes below that an if-statement is used to increase F_{air} by 10% at t=10 s.

MATLAB Code:

```
function Ex3_9DOProcess
clear; clc;
% (y(1)-O2 conc; y(2)-Measured O2 conc)
t0=0; tf=100; y0=[1.1e-4;1.1e-4];% Input specs
% Call ode45 and store solution in soln
soln=ode45(@f,[t0,tf],y0,odeset('RelTol', 1.E-6, AbsTol, 1.E-6));
t=linspace(0,tf,100);            % Generate the values of t using linspace
y1=deval(soln,t,2);             % Retrive value of y from soln
plot(t,y1,'k-','LineWidth',2);  % Plot t/y data and specify labels
xlabel('t (s)'); ylabel('Measured O2 concentration (gmol/l)');
end
function dydt=f(t,y)
Fair=500; CO2_sat=2.2e-4; KO2=1.98; x=0.25; mu_max=5.56e-5; tauS=30;
if t>=10; Fair=1.1*500; end;
kLa=0.25+0.001*(Fair-500);
dydt(1)=kLa*(CO2_sat-y(1))-KO2*mu_max*x;
dydt(2)=(y(1)-y(2))/tauS;
dydt=dydt';                      % Return dydt as column vector
end
```

Python Code:

```
import scipy.integrate
import numpy as np
import matplotlib.pyplot as plt
def fun(t, y):      # User-defined function
   Fair, CO2_sat, KO2, x, mu_max, tauS = 500, 2.2E-4, 1.98, 0.25, 5.56E-5, 30
   if t>= 10:
      Fair=1.1*500
   kLa=0.25+0.001*(Fair-500)
   dy1=kLa*(CO2_sat-y[0])-KO2*mu_max*x
   dy2=(y[0]-y[1])/tauS
   return [dy1, dy2]
tspan = [0, 100]
# (y(1)-O2 conc; y(2)-Measured O2 conc)
y0=[1.1E-4, 1.1E-4]
tvalues=np.linspace(0, 100, 100)
# Call solve_ivp to integrate the model equations
soln=scipy.integrate.solve_ivp(fun, tspan, y0, method='RK45', t_value=tvalues, rtol=1.E-6, atol=1.E-6)
time=soln.t                          # Retrive independent variable
yResult=soln.y[1]                    # Retrive dependent variable
# Plot results
plt.figure(figsize=(6,4.5))
plt.plot(time, yResult, 'k-', linewidth=2)
```

```
plt.axis([0, 100, 1.E-4, 1.4E-4])
plt.title('Solution for Ex3.9', fontsize=16)
plt.xlabel('t', fontsize=14)
plt.ylabel('DOs', fontsize=14)
plt.show
```

Self-Assessment Questions

Q3.7.1 Why are there two model equations for Example 3.6 while Examples 3.3-3.5 have only one model equation?
Q3.7.2 How can you determine if a process is self-regulating or non-self-regulating from its open loop response?

Self-Assessment Answers

Q3.7.1 For Examples 3.2-3.5, only one dependent is required for the process model because the output variable in each case (i.e., temperature, composition and level, respectively) is defined by one balance equation (energy balance, component balance and overall mass balance, respectively). On the other hand, for Example 3.5 an energy balance and a component mole balance is required to model the reactor temperature. In other words, the reactor temperature is a function of C_A and T; therefore, an energy balance and a component mole balance are required to model this process.
Q3.7.2 The output variable of a self-regulating process will move to a new stable steady-state operating point after an input to the process while the output variable of a non-self-regulating process will theoretically increase or decrease without limit.

Self-Assessment Problems

P3.7.1 Consider the level process shown in Figure 3.7.9 except that the inlet flow is the input and the outlet flow is given by $F_{out} \quad K\sqrt{L}$ where K is a constant and L is the level in the tank. Develop an actuator/process/sensor model for this process. Assume that the flow controller on the feed stream has a time constant of 2 seconds. Identify all process variables and process parameters along with their dimensional units. Identify the dependent variables, the independent variables and the parameters for this model.

Self-Assessment Answers

P3.7.1 The model equations are the same as Example 3.3 except that $F_{out} \quad K\sqrt{L}$ and F_{in} is determined by a control valve. Making this substitution yields the following actuator/process/sensor model equations.

Actuator $\frac{dF_{in}}{dt} \quad \frac{1}{2s} \quad F_{in,spec} \quad F_{in}$ **Process** $A_c \frac{dL}{dt} \quad F_{in} \quad K\sqrt{L}$ **Sensor** $L_s \quad L$

All flow rates have units of kg/s, density is given in kg/l, A_c is given in m^2, L is given in m and K is given in kg/m$^{1/2}$-s. The dependent variables are F_{in}, L and L_s. The independent variables are $F_{in,spec}$ and time and the parameters are r, A_c, t_v and K.

3.8 Sensor Noise

Modeling of sensor noise is important for the realistic dynamic modeling of control loops applied to industrial processes. To reduce the effect of noise on feedback control, the process measurement is routinely filtered for industrial applications. Filtering of the process measurement adds lag to the overall dynamic response of a feedback loop; therefore, neglecting sensor noise, when simulating a control loop, results in a dynamic model that responds faster than the corresponding process that uses filtering of sensor readings.

Sensor noise can usually be modeled by assuming that it is **Gaussian-distributed white noise**. To model Gaussian-distributed white noise, you have to choose only the standard deviation of the noise, . Then the following equation can be used to approximate a "bell-shaped" Gaussian distribution.

$$y_n \qquad \frac{1.961 \quad (x_n \quad 0.5)}{[(x_n \quad 0.002432)(1.002432 \quad x_n)]^{0.203}} \qquad \textbf{3.8.1}$$

where y_n is the noise contribution to the measured value of the controlled variable and x_n is a random number between 0 and 1. Because $0 < x_n < 1$ with a mean of 0.5, y_n has approximately the same number of positive and negative values. Appendix C presents a simple algorithm for generating numbers that are very nearly random. Also, MATLAB and Python offer functions that generate random numbers.

The procedure for modeling a sensor reading with noise is as follows:

1. Select the standard deviation of the noise. Remember that 4 (i.e., 2) should contain 95% of the readings, which is directly related to the repeatability of the sensor.
2. Identify the noise-free sensor reading, y_{nf}, by applying the model equations. For example, Figure 3.7.5 shows the noise-free sensor readings for a step change for the CST thermal mixer (Example 3.3).
3. Generate a random number with values between 0 and 1, x_n (Appendix C).
4. Apply Equation 3.8.1 to calculate y_n.
5. Then calculate the sensor reading, y_s, as

$$y_s \qquad y_{nf} \qquad y_n$$

Each time a new sensor reading is required, apply steps 2 to 5. The repeatability of the sensor reading of interest (Table 2.4) can be used to estimate the value of to use to model the sensor noise. For example, consider a level indicator. From Table 2.4, the repeatability is about ±1% corresponding to 4 ; therefore, is equal to 0.5%.

Figure 3.8.1 shows the results of adding sensor noise to the CST thermal mixer for a step change in the flow rate of stream 1 while Figure 3.7.5 shows the noise free results. The noise on the temperature sensor reading is modeled using a of 0.05 C for a thermocouple. The noise free temperature results are shown in Figure 3.7.5.

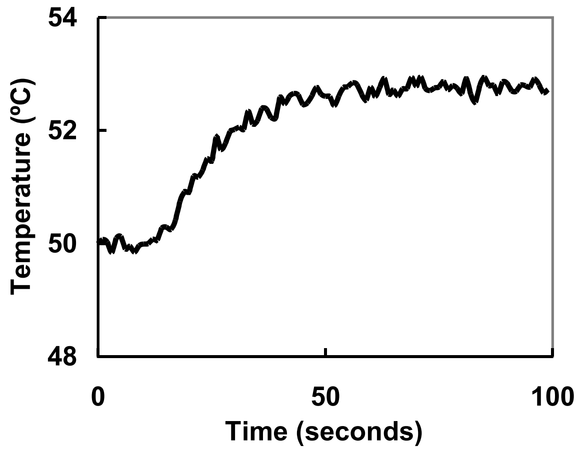


Figure 3.8.1 Dynamic response of the CST thermal mixer with noise modeled on the sensor reading.

Example 3.10 Noise Model for an Orifice Flow Meter

Problem Statement. Develop a noise model for an orifice flow meter assuming that it has a repeatability of ±0.8% where the nominal flow measurement is 10,000 lb/h.

Solution. Because the repeatability (i.e., ±0.8%) is assumed equal to 4 (i.e., ±2), is equal to 0.4% or 40 lb/h at the nominal flow rate. The procedure for modeling noise on a flow measurement from an orifice flow meter is as follows.

1. Determine the noise-free value of the flow, F_{nf}.

2. Determine a random number (0,1) x_n.

3. Calculate as = 0.004 F_{nf}.

4. Apply Equation 3.8.1 to calculate F_n.

5. Calculate the sensor reading as

$$F_s \quad F_{nf} \quad F_n$$

3.9 Visual Basic Simulator

A visual basic simulator (VBS) is provided with this text and is based on several of the models developed in this chapter: the CST thermal mixer, the CST composition mixer, the level in a tank and the CSTR. In addition, a model of a first-order plus deadtime process (Section 6.8) and a model of a heat exchanger[6] are included.

The VBS runs under MS Excel using Visual Basic (Figure 3.9.1). In order to execute this program, you will need to use macros that require that you select the lowest setting for macro security under MS Excel. In order to do this, under File select Options and then click on the Trust Center category. In the Trust Center, select Trust Center Settings and click on Macro Settings. After the security settings are set, reboot your computer and you can start the VBS by clicking on the model (Visual Basic Simulators.xls), which should open with the "Process Control Simulator" page shown in Figure 3.9.1. Simply click on the "Input Form" button and the input form opens (also shown in Figure 3.9.1).

For open loop tests, first select the type of model that you want to execute by selecting from the drop down menu next to "Simulator". Next, select the open loop test (i.e., MV step test, Disturbance step test and ATV test), set the magnitude of the open loop test and click the "Run Simulation" button. A new window opens with Excel plots for the MV, CV and DV as well as the raw data for these variables as a function of time.

In order to return to the simulator, click the "input" tab in the lower left corner (Figure 3.9.2) to the right of the "Temp" tab, which will return you to the page shown in Figure 3.9.1. Once again click on the "Input Form" button and apply another simulation.

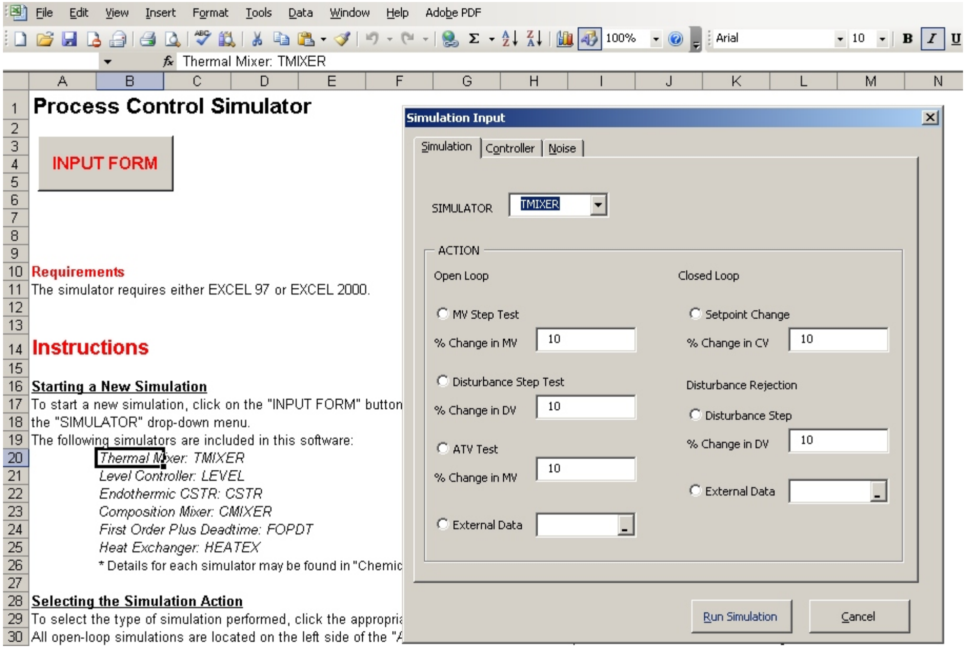


Figure 3.9.1 Screen shot of VBS with input form.

For closed loop tests, select the simulator and the type of closed loop test that you would like to perform (i.e., setpoint change or disturbance rejection). Next, click on the "Controller" tab at the top of the simulation input form and select the "Controller Type" and "Action" from the drop down menus. Then, provide the tuning parameters for the controller selected (e.g., "Controller gain" and "TAU I" for a PI controller). Once the controller type and tuning parameters have been selected, click on the "Run Simulation" button and the results for the MV, CV and DV will appear as plots and as raw data.

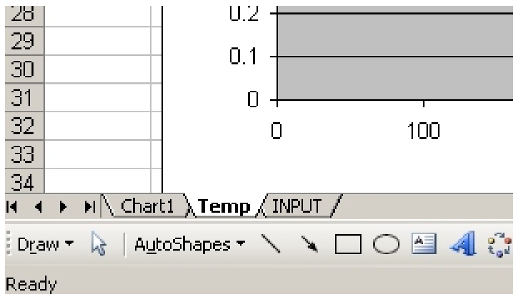


Figure 3.9.2 The "INPUT" tab.

The last tab in the simulation input form is the "Noise" tab which allows the user to add sensor noise and sensor filtering to the CV value. The standard deviation of the noise and the filter factor *f* can be specified by the user. The VBS can be used to observe dynamic behavior and for tuning exercises (Section 9.13).

3.10 Summary

- For dynamic models for process control applications, it is essential to consider the combined dynamics of the actuator, process and sensor.
- The component mass balance equation is given by

rate of accumulation of mass within the system	flow rate of mass entering the system	flow rate of mass leaving the system

- The component mole balance equation is given by

rate of accumulation of moles within in the system	flow rate of moles entering the system	flow rate of moles leaving the system
rate of generation of moles by reaction	rate of consumption of moles by reaction	

- The energy balance equation for process systems is given by

rate of accumulation of thermal energy	rate of convective heat transfer entering the system	rate of convective heat transfer leaving the system
net rate of energy generation by chemical reaction	net rate of heat transfer through the boundaries of the system	

- The dynamic model for actuators and certain sensors can be well-approximated as a first-order system, i.e., for an actuator, $dF/dt \quad (F_{spec} \quad F)/\ _v$ where $_v$ is the time constant for the actuator, F_{spec} is the specified flow rate provided to the actuator and F is the actual flow rate to the process; for a sensor, $dT_s\ /\ dt \quad (T \quad T_s)/\ _s$ where $_s$ is the time constant for the sensor, T is the actual value in the process that the sensor is attempting to measure and T_s is sensor reading.
- When sensor noise is significant, it is important to add sensor noise to a dynamic process model to take into account the effect of sensor noise and sensor filtering on the dynamic behavior.

3.11 References

1. Bird, R.B., Stewart, W.E. and Lightfoot, E.N., *Transport Phenomena*, John Wiley and Sons, New York, p. 216 (1960).

2. Riggs, J.B., *Computational Methods for Engineers with MATLAB Applications*, Ferret Publishing, Austin, Texas (2013), pp. 289-294.

3. Riggs, J.B., *An Introduction to Numerical Methods for Chemical Engineers,* Second Edition, Texas Tech University Press, Lubbock, Texas, pp. 387-420 (1994).

4. Bequette, B.W., *Modeling and Analysis of Dynamic Systems*, Prentice Hall, Englewood, New Jersey, (1998).

5. Aiba, S., M. Shoda and M. Nagatani, *Biotechnol. Bioeng.*, 10, p. 845 (1968).

6. Riggs, J.B. and M.N. Karim, *Chemical and Bio-Process Control*, 3rd Ed., Ferret Publishing, Austin, TX (2006).

3.12 Additional Terminology

Constitutive relations - algebraic equations such as kinetic expressions or gas laws that are necessary to solve model equations.

CST - continuous stirred tank, i.e., a vessel in which the temperature and composition throughout the tank are uniform.

CSTR - continuous stirred tank reactor, i.e., a reactor in which the composition and temperature are uniform throughout the reactor volume because the fluid in the reactor is well mixed.

Dependent variable - a process variable that is determined by the operation of the process or the solution of the model equations**.**

Diafiltration - a process for removing salts from products by allowing salt plus diluting medium to permeate through a membrane.

Distributed parameter model - a process model that can be applied to a system for which the dependent variables vary with spatial location within the process.

Exactly determined - a system of equations with the same number of unknowns as equations.

Exactly specified - a system of equations with the same number of unknowns as equations.

Gaussian-distributed white noise - noise that follows an equal probability-based Gaussian distribution.

Independent variable - a process variable that is specified independently of the process model results.

Lumped parameter model - a process model that assumes that the dependent variables are uniform throughout the spatial region of the process.

Macroscopic balances - balances based on what enters or leaves the process boundaries and treats the process in a lumped manner.

Macroscopic model - a model that is based upon macroscopic balances.

Microscopic balances - balances that consider that the dependent variables vary with spatial location within the process.

Microscopic model - a model that is based upon microscopic balances.

Non-self-regulating process - a process that does not move to a new steady-state condition after a change in a process input is made.

Open-loop unstable process - a process for which an input change can cause an unbounded increase in an output variable.

Overdetermined - a system of equations with more equations than unknowns.

Overspecified - a system of equations with more equations than unknowns.

Process parameter - a constant appearing in a model equation, e.g., physical parameters, rate constants, gas constants, etc.

Self-regulating process - a process that moves to a new steady-state after a change in a process input is made.

Stiff ODE - an ODE with a (Equation 3.21) that is equal to or greater than 500.

Underdetermined - a system of equations with more unknowns than equations.

Underspecified - a system of equations with more unknowns than equations.

3.13 Preliminary Questions

3.1 Introduction

Q3.1.1 Why are undergraduate chemical engineering students generally unfamiliar with process dynamics before studying process control?

3.2 Uses of Dynamic Models

Q3.2.1 Why are dynamic models for control and startup/shutdown analysis not generally the same?

Q3.2.2 What advantages does training operators on process simulators have compared to training them on the process?

3.3 Classifications of Phenomenological Models

Q3.3.1 Why are macroscopic models generally used for lumped parameter processes and microscopic models used for distributed parameter processes?

3.4 Dynamic Balance Equations

Q3.4.1 Explain what the accumulation term is for mole balance on a system.

Q3.4.2 When does the application of a mass balance and a mole balance yield the same equation?

Q3.4.3 What form do constitutive equations take and how are they used in models?

Q3.4.4 Does an overspecified system of equations have more equations than unknowns or more unknowns than equations?

3.6 Modeling Examples

Q3.6.1 Why are actuator systems based on control valves represented as a linear first-order system?

Q3.6.2 For what type of sensors is pure deadtime an appropriate model?

Q3.6.3 What individual physical processes are modeled in a lumped fashion by Equation 3.6.3?

3.7 Sensor Noise

Q3.7.1 If sensor measurements with noise average out to the sensor reading without noise, why should you model sensor noise when dynamically modeling a feedback control loop?

Q3.7.2 What are the sources of sensor noise?

Q3.7.3 Why is a random number used to model sensor noise?

3.14 Analytical Questions and Exercises

3.7 Modeling Examples

P3.7.1** When can the dynamics of a sensor or actuator be assumed instantaneous?

P3.7.2* Estimate a reasonable range for the standard deviation of the noise for a pressure sensor.

P3.7.3** Consider a vertical cylindrical tank that is 3 ft in diameter and has a water level of 6 feet (Figure P3.7.3). Assume that electrical heaters are placed inside the tank and can provide an instantaneous change in the heat addition rate to the tank. A thermocouple and thermowell with a dynamic time constant of 10 seconds are used to measure the temperature and the contents of the tank can be assumed to be well mixed. Develop the dynamic model equations for this batch process that describe the measured temperature of the water in the tank assuming that the water is initially at 50ºF and that a fixed heating rate, Q (200 Btu/s), is applied to the tank by the electrical heaters. Assume that the specific gravity is unity and the heat capacity per unit mass is 1 Btu/ºR·lb_m and the density is 62.4 lb_m/ft^3. Identify the dependent variables, the independent variables and the parameters for this model. Using a MATLAB program to implement this model, determine the time required to raise the meassure value of the water temperature to 100ºF.

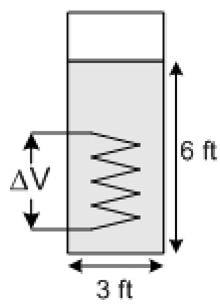


Figure P3.7.3 Electrically heated cylindrical tank.

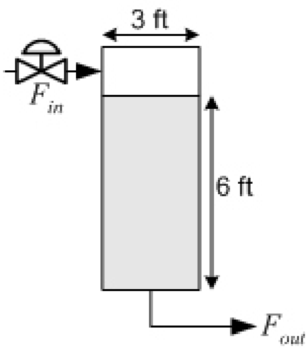


Figure P3.7.4 Self-regulating level in a tank.

P3.7.4** Consider a vertical cylindrical tank that is 3 feet in diameter and has a water level of 6 feet (Figure P3.7.4). Assume that there is flow into the tank (F_{in}) and that the flow out of the tank (F_{out}) is given by

$$F_{out} \quad k\sqrt{H}$$

where H is the height of the liquid in the tank. Initially the level is constant when the inlet feed rate is 30 gal/min. Develop the dynamic equations for this system assuming that initially the system is at steady state and then F_{in} is increased by 20% at time equal to zero. Assume that a control valve with a time constant of $_v$ of 5 s is used to implement the change in F_{in}. Identify the dependent variables, the independent variables and the parameters for this model. Using a MATLAB program to apply your model, determine the level in the tank for 100 min after the increase in the feed to the tank is implemented.

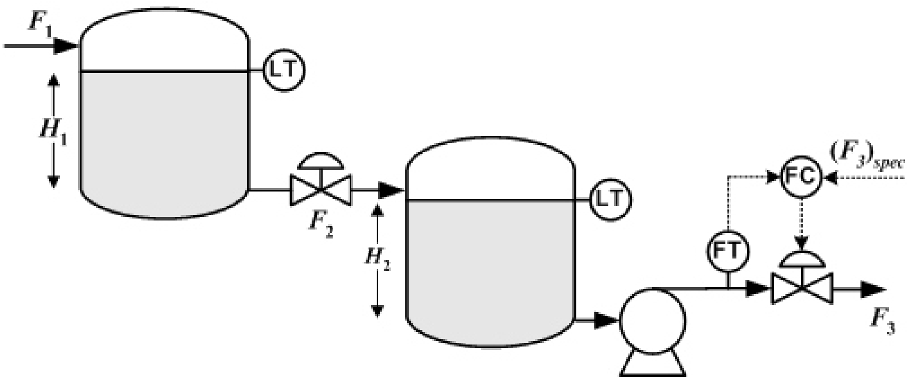


Figure P3.7.5 Schematic of two non-interacting tank levels.

P3.7.5** Consider the two noninteracting tank levels shown in Figure P3.7.5 where the Fs represent volumetric flow rates. Develop the dynamic models for H_1 and H_2 assuming that F_2 is given by

$$F_2 \quad k\sqrt{H_1}$$

which represents the effect of the valve with a constant valve position in the line between the first and second tank. A_1 (15 ft^2) and A_2 (20 ft^2) are the cross-sectional areas of the first and second tanks, respectively. Because this valve has a fixed stem position, which determines the value of k, the flow dynamics associated with this valve can be assumed instantaneous. Also assume that the dynamics of the level sensors are instantaneous. The changes in F_3 have a time constant $_v$ (3 s) and F_1 (30 gpm) is assumed constant. Initially, this system is at steady-state with H_1=H_2=5 ft. Identify the dependent variables, the independent variables and the parameters for this model. Using a MATLAB program to apply your model, determine the two levels in the tanks for 200 min after a 10% increase in the $(F_3)_{spec}$ is implemented.

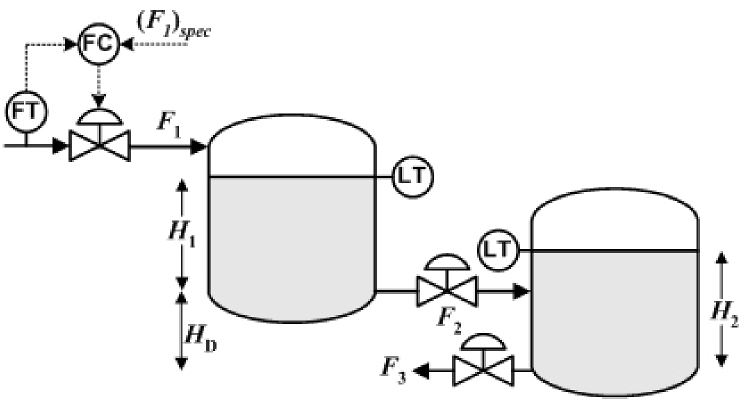


Figure P3.7.6 Schematic of two interacting tank levels.

P3.7.6** Consider the two interacting tank levels shown in Figure P3.7.6 in which the Fs represent volumetric flow rates. Develop the dynamic models for H_1 and H_2 assuming that F_2 and F_3 are given by

$$F_2 \quad k_1\sqrt{H_1 \quad H_D \quad H_2}$$

$$F_3 \quad k_2\sqrt{H_2}$$

which represent the effect of the valves in the exit lines and the hydrostatic head for the two tanks. A_1 (15 ft^2) and A_2 (20 ft^2)are the cross-sectional areas of the first and second tanks, respectively. Because the valves have a fixed stem position, which determines the value of k, the flow dynamics associated with these valves can be assumed instantaneous. Also assume that the dynamics of the level sensors are instantaneous. The changes in F_1 have a time constant $_v$ (3 s). Initially, this system is at steady-state with H_1=H_2=5 ft and H_D=2 ft for F_1=30 gpm. Identify the dependent variables, the independent variables and the parameters for this model. Using a MATLAB program to apply your model, determine the two levels in the tanks for 400 min after a 10% increase in the $(F_1)_{\text{spec}}$ is implemented.

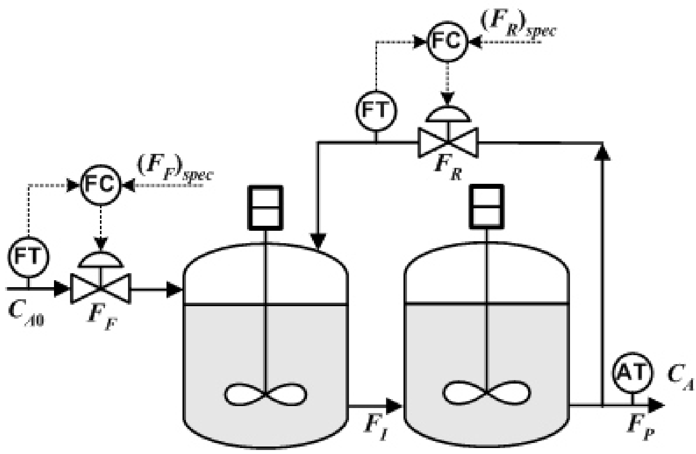


Figure P3.7.7 Schematic of two tanks with a recycle.

P3.7.7* ** Develop the dynamic modeling equations for the concentration of A in the product stream as a function of time for the specified flow rates for the feed and the recycle stream for the two mixing tanks shown in the Figure P3.7.7. Note that all the Fs are expressed as volumetric flow rates and both flow controllers have the same dynamic response, which is described by $_v$(5 s). Assume perfect level control for both tanks and assume that both tanks are perfectly mixed and have the same holdup of liquid of volume V (100 l). The analyzer on the product stream are IR based analyzers and have a time constant of 30 s for their measurements. Identify the dependent variables, the independent variables and the parameters for this model. Initially, the feed composition and the composition in both tanks have a concentration of 1 gmol/l. At time equal to zero, the $(F_F)_{spec}$ is decreased from 100 l/min to 90 l/min, $(F_R)_{spec}$ is decreased from 50 l/min to 45 l/min and the feed composition increases to 1.2 gmol/l. Using a MATLAB program to apply your model, determine the composition measurement of the product and the composition in the two tanks from time equal to zero to 10 min.

P3.7.8* Develop the model equations that can be used to represent the dynamic behavior of a stirred tank cooler considering the combined models of the actuator, process and sensor. Assume that the stirred tank is identical to the endothermic CSTR given in Example 3.6 except that no reactions take place. Identify the dependent variables, the independent variables and the parameters for this model. Assume that the system is initially at steady-state with a measured product temperature of 350 K. At time equal to 10 s, the specified heat removal rate is reduced by 10%. Using a MATLAB or Python program to apply your model, determine the measured temperature of the product from time equal to zero to 100 s.

P3.7.9** Develop the set of dynamic equations that describe an isothermal CSTR (Figure P3.7.9) in which series reactions occur using an actuator/process/sensor modeling approach. Assume that one feed stream enters the reactor and one product stream leaves the reactor. The reaction scheme is given by

$$A \quad ^{r_1} \quad B \quad ^{r_2} \quad C$$

where $r_1 = k_1\ C_A$ (k_1=1 s^{-1}) and $r_2 = k_2\ C_B$ (k_2=1 s^{-1}) . The time constant for the control valve for the feed rate has a value of 5 seconds and the feed stream contains only component A at a concentration C_{A_0} . The volume of the reactor is 100 l, the density of the feed and reaction mixture is 1 kg/l, the feed rate to the reactor is

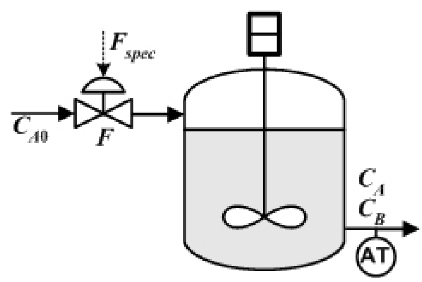


Figure P3.7.9 PFD of an isothermal CSTR.

constant at 10 l/s and the feed composition C_{A0} is equal to 3 gmol/l. The output variable is the concentration of B in the reactor product. Assume that an IR analyzer on the product stream has a time constant of 40 s. Also assume perfect level control in the reactor (i.e., the flow rate out of the reactor equals the flow rate into the reactor). Identify all process variables and process parameters along with their dimensional units. Identify the dependent variables, the independent variables and the parameters for this model. Initially, the reactor is filled with a liquid at the feed composition with no component B present. At time equal to zero, a catalyst is added to the reactor and the reactions begin. Using a MATLAB or Python program to apply your model, determine the measured concentration of component B in the product from time equal to zero to 200 s.

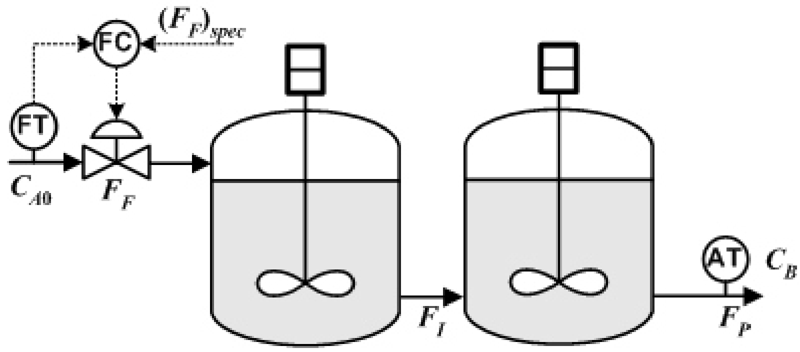


Figure P3.7.10 Schematic of two isothermal CSTRs in series.

P3.7.10** Develop the set of dynamic equations that describe two isothermal CSTR's in series shown in Figure P3.7.10 using an actuator/process/sensor modeling approach. Assume that a single irreversible reaction occurs in this system given by

$$A \xrightarrow{r} B$$

where $r = k\,C_A^2$ where k=1 l/gmol-s. Assume that the feed rate to the first reactor is the input variable and that the flow controller on the feed has a time constant of 2 seconds. Assume that the Fs in Figure P3.5.10 represent volumetric flow rates and each has a value of 10 l/s while the volume of each reactor is 100 l. The inlet feed contains only component A at a concentration of C_{A_0} (3 gmol/l). The output variable of this process is the concentration of B in the outlet product stream and the IR analyzer on this stream has a time constant of 30 s. Also assume perfect level control in each reactor (i.e., the flow rate out of the reactor equals the flow rate into the reactor). Identify the dependent variables, the independent variables and the parameters for this model. Initially, each reactor has a concentration of A equal to 3 gmol/l with no product B present and at time equal to zero a catalyst is added to both reactors that allows the reactions to occur. Using a MATLAB or Python program to apply your model, determine the measured concentration of component B in the product leaving the second reactor from time equal to zero to 200 s.

P3.7.11*** Develop a macroscopic model of a steam-heated heat exchanger using an actuator/process/sensor modeling approach. The output variable is the outlet temperature of the process fluid and the input variable is the specified steam pressure. For a macroscopic model of a heat exchanger (Figure P3.7.11), the metal of the heat exchanger is at one temperature and the temperature of the process stream used for heat-transfer calculations is the average between the inlet temperature and the exit temperature for the heat exchanger, i.e., the heat-transfer rate from the metal to the process fluid and the heat transfer rate from the steam to the metal are given by

$$Q = U A(\overline{T}_m - \overline{T}) \qquad Q_{stm} = U_{stm} A(T_{stm} - \overline{T}_m)$$

respectively, where U is the overall heat-transfer coefficient between the process stream and the metal tubes (500 W/m^2-K), A is the surface area for heat transfer (100 m^2), $\overline{T}$ is the average temperature for the process fluid inside the heat exchanger, $\overline{T}_m$ is the average temperature for the metal tubes in the heat exchanger, T_{stm} is the steam temperature and U_{stm} is the heat-transfer coefficient between the steam and the metal tubes (2000 W/m^2-K). The result for the heat-transfer rate can be used to calculate

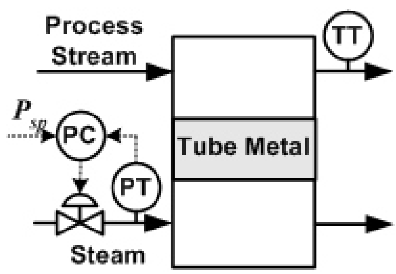


Figure P3.7.11 Schematic of a one-dimensional representation of a heat exchanger.

the outlet temperature of the process fluid. Assume that the steam is saturated so that the steam pressure controller sets the steam temperature. Assume that the pressure controller on the steam has a time constant of 10 seconds and the outlet temperature sensor for the process fluid has a time constant of 6 seconds. Identify the dependent variables, the independent variables and the parameters for this model. The feed flow rate of the process fluid is 25 kg/s, feed temperature of the process fluid is 25°C, the heat capacity of the process fluid is 4.18 kJ/kg-°C, the heat capacity of copper is 0.385 kJ/kg-°C and the mass of copper in the tubes used in the heat exchanger is 200 kg. Initially, the heat exchanger is filled with cold water at 25°C when the steam is added to the heat exchanger with a pressure setpoint of 300 kPa. Assume that the initial pressure of steam in the heat exchanger is 50 kPa. Use the following equation that relates the steam temperature in °C to the steam pressure in kPa:

$$T_{stm} \quad \frac{1810.95}{8.14018 \quad \log 10(7.5025 P_{stm})} \quad 244.485$$

Using a MATLAB or a Python program to apply your model, determine the measured outlet temperature of the water from time equal to zero to 30 s.

P3.7.12** Develop the set of dynamic equations that describe an isothermal semi-batch reactor as shown in Figure P3.7.12 using an actuator/process/sensor modeling approach. For a semi-batch reactor feed is added to the reactor, but product is not removed. Assume the same reaction scheme and rate expression as used in Problem 3.7.10. Assume that a flow controller with a time constant of 2 seconds controls the feed to the reactor and that the product B composition is measured by an IR analyzer with a analyzer time constant of 30 s. The density of the feed and reaction mixture is 1 kg/l and the rate constant for the chemical reaction is 1.0 l/gmol-s. Identify all process variables and process parameters along with their dimensional units. (Hint: Remember to model the volume of the reactor as a function of time. Also, use the number of moles of A and B as dependent variables.) Identify the dependent variables, the independent variables and the parameters for this model. Initially, the reactor is empty and at time equal to zero the specification for the feed flow controller is increased from zero to 10 kg/s. Using a MATLAB or Python program to apply your model, determine the measured concentration of component B in the product in reactor from time equal to zero to 100 s.

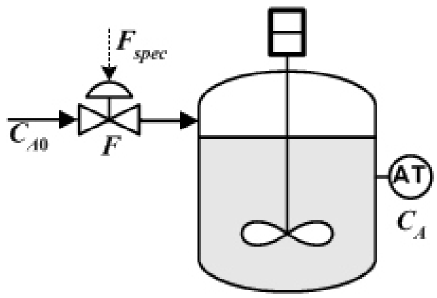


Figure P3.7.12 PDF of a semi-batch reactor.

P3.7.13** Consider the moonshine still shown in Figure P3.7.13. This batch process consists of putting sour mash (i.e., largely water and ethyl alcohol) into the still and adding heat. The ethyl alcohol is more volatile than water; therefore, the distillate from the still is rich in alcohol. Write a set of model equations for this process using the actuator/process/sensor approach, assuming that the charge is already heated to the boiling point of the mixture. The heat input rate, Q, is has a constant value of 25,000 Btu/h while the output variable is the concentration of ethyl alcohol in the product stream. Because the composition analysis is done by using a continuos density sensor, the analyzer delay has a time constant of 60 seconds. Assume that both ethyl alcohol and water have the same constant heat of vaporization per unit mass (H_{vap}=1000 Btu/lb_m) and that the mass fraction of the alcohol in the vapor leaving the still, y, is related to the mass fraction of alcohol remaining in the still, x, by the following equation

$$y \quad \frac{x}{1 \quad (\quad 1)x}$$

where is equal to 5. Identify the dependent variables, the independent variables and the parameters for this model. Initially, the still has a charge of 200 lb_m with 15 weight percent alcohol. Using a MATLAB or Python program to apply your model, determine the measured concentration of alcohol in the product from time equal to zero to 3 h.

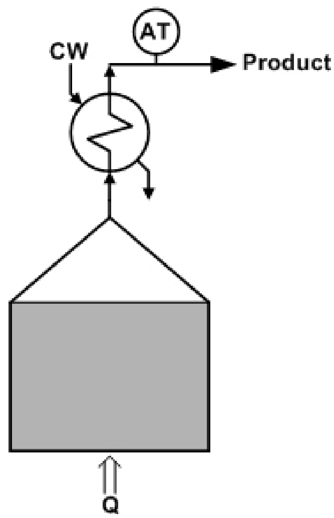


Figure P3.7.13 Schematic of a moonshine still.

P3.7.14** Consider a pressure vessel (1000 l) for which there is a fixed gas flow into the vessel and an exit line that has a control valve on it. Assume that the input to this process is the signal to the control valve on the exit line and the output

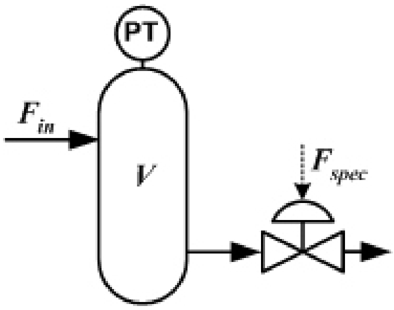


Figure P3.7.14 Schematic of a compressed air tank with an inlet and outlet flow.

variable is the pressure of the gas in the vessel. Assume that $_{v}$ is equal to 10 seconds and that the temperature inside the vessel is 300 K. Develop a dynamic model for this process using the combined models of the actuator, process and sensor. Identify the dependent variables, the independent variables and the parameters for this model. (Hint: Use the ideal gas law to relate pressure to the number of moles of gas in the vessel.) Initially, the system is at steady-state with the flow into the tank is 10 gmol/s and the vessel pressure at 2 atm. At time equal to zero, the flow rate through the control valve is decreased by 10%. Using a MATLAB or a Python program to apply your model, determine the vessel pressure from time equal to zero to 100 s.

P3.7.15** Develop the model equations for a hot-water heater. Consider that the hot-water heater is well mixed and has a water holdup of 30 gallons. Assume that initially the water temperature is 120ºF. At time equal to zero, hot water is withdrawn at rate of 5 gallons per minute and 5 gallons per minute of cold water at 60ºF are simultaneously added. Assume that the heat addition rate to the hot-water heater is constant at 6×10^4 BTU/h and is applied at the instant the hot water is removed from the tank. For this case, neglect actuator and sensor dynamics. Also, determine how long it takes for the hot-water temperature to drop to less than 90ºF. Develop an analytical solution for this problem.

P3.7.16** Develop a dynamic model for the level in a cone-shaped bin. The equation for volume of fluid in a cone is given by

$$V \quad \frac{r^2 h}{3}$$

where h is the height of the fluid in the cone and r is the radius of the cone at the fluid level in the cone. Note that r is a function of h given by

$$h \quad \sqrt[3]{\frac{3M}{\tan^{2}}}$$

where M is the mass in the bin, is the angle of the cone (30^{o}) and is the density of the fluid. Consider the flow into and out of the cone, the actuator dynamics and the level sensor dynamics. Assume that the density of the fluid is 1 kg/l and that the flow rate leaving the tank is given by

$$F_{out} \quad k\sqrt{h}$$

Initially, the tank is at steady-state with a feed rate of 100 l/s and a level in the cone-shape bin of 3 m. At time equal to zero, the feed rate to the tank is increased by 10% and the valve has a time constant of 5 s. Using a MATLAB or Python program to apply your model, determine the height of the level in the bin from time equal to zero to 1000 s.

Chapter 4

Laplace Transforms

Chapter Objectives

- Review the definitions of Laplace transformations.
- Apply partial fraction expansions and the Heaviside method for evaluating the unknown constants from partial fraction expansions.
- Demonstrate the solution of the three general types of dynamic differential equations using Laplace transforms: (1) individual real poles, (2) repeated real poles and (3) complex poles.

4.1 Introduction

Chapter 3 demonstrated how to develop dynamic models for process systems and solve them numerically and this chapter deals with how to solve dynamic differential equations using Laplace transforms. Laplace transforms represent a relatively simple means to develop analytical solutions of **linear differential equations**. In addition, Laplace transforms can be used to develop transfer functions (Chapter 5), which conveniently and meaningfully represent the input/output behavior of a linear process. Laplace transforms and transfer functions provide insight into the fundamental behavior of dynamic systems while introducing important terminology relevant to the process control field.

4.2 Laplace Transforms

From Example 3.5, the process model represented by Equation 3.7.9 is given as

$$A_c \frac{dL}{dt} = F_{in} - F_{out} \qquad \text{at } t = 0, L = L_0$$

where L is the dependent variable and F_{in} and F_{out} are independent variables. The solution of this initial-value problem yields the tank level, L, as a function of time, i.e., $L=f(t)$ based on the initial conditions and specified values for F_{in} and F_{out}. The next sections will show how to use Laplace transforms to solve dynamic equations like this for $f(t)$.

The Laplace transform of a function, $f(t)$, e.g., $L(t)$ for Equation 3.7.9, is defined by

$$\mathscr{L}\left[f(t)\right] \equiv \int_0^\infty f(t)e^{-st}\,dt \equiv F(s) \qquad \textbf{4.2.1}$$

where $f(t)$ is a relatively general function of time, t, for $t \geq 0$, $\mathscr{L}$ is the Laplace operator, s is a complex independent variable and $F(s)$ is the symbol for the Laplace transform of *f(t)*. Equation 4.2.1 can be rearranged to yield the expression for the inverse Laplace transform.

$$\mathscr{L}^{-1}\left[F(s)\right] = f(t) \qquad \textbf{4.2.2}$$

Laplace transforms are useful for solving linear dynamic equations. The time-domain equations are transformed into the Laplace domain (Equation 4.2.1) where they are solved algebraically, yielding an equation for the output variable in the Laplace domain. The output variable in the Laplace domain can be then be transformed back to the time domain using inverse Laplace transforms (Equation 4.2.2).

Table 4.1 lists the Laplace transforms of several commonly encountered functions. Table 4.1 can also be used to apply inverse Laplace transforms by going from the Laplace transform to the corresponding time function, *f(t)*.

The Laplace transformation is a linear operation. That is, the Laplace transform of a sum of two functions is the sum of Laplace transform of the individual functions, i.e.,

$$\mathscr{L}\left[a f_1(t) + b f_2(t)\right] = a\,\mathscr{L}\left[f_1(t)\right] + b\,\mathscr{L}\left[f_2(t)\right] \qquad \textbf{4.2.3}$$

where a and b are constants. On the other hand,

$$\mathscr{L}\left[f_1(t)f_2(t)\right] \neq \mathscr{L}\left[f_1(t)\right]\mathscr{L}\left[f_2(t)\right]$$

You can understand this by comparing the Laplace transform of $e^{-at}\sin\omega t$ with the Laplace transforms of e^{-at} and $\sin\omega t$ in Table 4.1.

Example 4.1 Derivation of a Laplace Transform

Problem Statement. Using the definition of a Laplace transform (Equation 4.2.1), derive the Laplace transform for e^{-at}.

Solution. Applying Equation 4.2.1 for $f(t)=e^{-at}$ yields

$$\mathscr{L}\left[e^{-at}\right] = \int_0^\infty e^{-at}e^{-st}\,dt = \int_0^\infty e^{-(s+a)t}\,dt = -\frac{1}{s+a}e^{-(s+a)t}\Big|_{t=0}^{t=\infty} = -\frac{1}{s+a}(0-1) = \frac{1}{s+a}$$

Table 4.1 Laplace transforms for some common functions.

f(*t*)	*F*(*s*)	*f*(*t*)	*F*(*s*)
Unit impulse at $t = 0$	1	$\sin \quad t$	$\frac{\quad}{s^2 \quad ^2}$
Step input A at $t = 0$	A/s	$\cos \quad t$	$\frac{s}{s^2 \quad ^2}$
Ramp input, $a\ t$	a / s^2	$e^{\ at} \sin \quad t$	$\frac{\quad}{(s \quad a)^2 \quad ^2}$
t^2	$\frac{2}{s^3}$	$e^{\ at} \cos \quad t$	$\frac{s \quad a}{(s \quad a)^2 \quad ^2}$
t^n	$\frac{n!}{s^{n \quad 1}}$	$\frac{d\, f(t)}{dt}$	$s\, F(s) \quad f(0)$
$e^{\quad at}$	$\frac{1}{s \quad a}$	$\frac{d^2 f(t)}{d\, t^2}$	$s^2\, F(s) \quad s\, f(0) \quad \frac{df}{dt}_{t \quad 0}$
$\frac{1}{\ } e^{\quad t/}$	$\frac{1}{s \quad 1}$	$\frac{d^n f(t)}{dt^n}$	$s^n F(s) \quad \overset{n}{\underset{k \quad 1}{\ }} s^{n \quad k} f^{\,k \quad 1}(0)$
$A(1 \quad e^{\ t/} \ )$	$\frac{A}{s(\ s \quad 1)}$	$K\, f(t)$	$K\, F(s)$
$t^n e^{\quad at}$	$\frac{n!}{(s \quad a)^{n \quad 1}}$	$f(t \qquad)$ (Time delay)	$F(s)\, e^{\quad s}$
$\frac{1}{a_1 \quad a_2} \quad e^{\ a_2 t} \quad e^{\ a_1 t}$	$\frac{1}{s \quad a_1 \quad s \quad a_2}$	$\overset{t}{\underset{0}{\ }} f(\ )\, d$	$\frac{1}{s} F(s)$

Example 4.2 Derivation of a Laplace Transform

Problem Statement. Using the definition of a Laplace transform (Equation 4.2.1), derive the Laplace transform for step function.

Solution. A unit step function is given by

$$
\begin{aligned}
f(t) \quad 0 \qquad\qquad t \quad 0 \\
f(t) \quad A \qquad\qquad t \quad 0
\end{aligned}
$$

Applying Equation 4.2.1,

$$\mathscr{L}\ f(t) \qquad {}_{0}\ Ae^{\ st}dt \quad \frac{A}{s}e^{\ st}\Big|_{0} \quad \frac{A}{s}(0 \quad 1) \quad \frac{A}{s}$$

Example 4.3 Derivation of a Laplace Transform

Problem Statement. Using the definition of a Laplace transform (Equation 4.2.1), derive the Laplace transform for sin t.

Solution. Applying Equation 4.2.1 using integration by parts[1],

$$\mathscr{L}\ \sin \quad t \qquad {}_{0}\ e^{\ st}\sin \quad t\,dt \quad \frac{e^{\ st}(\ s\sin \quad t \qquad \cos \quad t)}{s^{2} \qquad ^{2}}\Bigg|_{0} \quad 0 \quad \frac{}{s^{2} \qquad ^{2}} \quad \frac{}{s^{2} \qquad ^{2}}$$

Example 4.4 The Laplace Transform of a Series of Functions

Problem Statement. Determine the Laplace transform of $f(t)$ if

$$f(t) \quad t^{2} \quad t^{2}e^{\ 3t}$$

Solution. By defining

$$f_{1}(t) \quad t^{2}$$
$$f_{2}(t) \quad t^{2}e^{\ 3t}$$

and using the property of a linear operation (Equation 4.2.3),

$$\mathscr{L}\ f(t) \qquad F(s) \quad \mathscr{L}\ t^{2} \qquad \mathscr{L}\ t^{2}e^{\ 3t}$$

Then, using Table 4.1,

$$F(s) \quad \frac{2}{s^{3}} \quad \frac{2}{(s \quad 3)^{3}}$$

The Final-Value Theorem. The **final-value theorem** states that

$$\lim_{t} \{f(t)\} \qquad \lim_{s \quad 0} \{sF(s)\} \tag{4.2.4}$$

if the limit of $f(t)$ as t exists. The final-value theorem can be used to determine the steady-state conditions of a process if a model of the process is available in the Laplace domain, assuming that the process is stable (i.e., there are no positive poles; see Section 5.3).

The Initial-Value Theorem. The **initial-value theorem** is given by

$$\lim_{t \to 0} \{f(t)\} = \lim_{s \to \infty} \{sF(s)\} \qquad \textbf{4.2.5}$$

The initial-value theorem can be used to determine the initial conditions of a function if the Laplace transform of the function is known. Note that when time becomes large, s approaches zero and *vice-versa*.

Example 4.5 Application of the Initial-Value and Final-Value Theorems

Problem Statement. Apply the initial-value and final-value theorems to the Laplace transform of

$$f(t) = e^{-at} \sin \omega t$$

Solution. From Table 4.1, the Laplace equation for this function is

$$F(s) = \frac{\omega}{(s+a)^2 + \omega^2}$$

Applying Equation 4.2.5 yields

$$\lim_{t \to 0}[f(t)] = \lim_{s \to \infty} \frac{s\omega}{(s+a)^2 + \omega^2} = 0$$

Note that when t equal to zero is substituted into the time-domain version of the function, the same result is obtained.

Applying Equation 4.2.4 yields

$$\lim_{t \to \infty}[f(t)] = \lim_{s \to 0} \frac{s\omega}{(s+a)^2 + \omega^2} = 0$$

which is also consistent with the time-domain function.

Self-Assessment Questions

Q4.2.1 What are $f(t)$ and $F(s)$ in Equation 4.2.1?

Q4.2.2 What restrictions are placed on $f(t)$ so that its Laplace transform, F(s), can be determined?

Self-Assessment Answers

Q4.2.1 $f(t)$ is a general function of time and $F(s)$ is the Laplace transform of $f(t)$.

Q4.2.2 $f(t)$ must be defined and continuous for $t > 0$.

Self-Assessment Problem

P4.2.1 Using Table 4.1, determine the Laplace transform of $y(t) = e^{-2t} \sin 5t$.

Self-Assessment Answer

P4.2.1 The Laplace transform of $e^{\ at} \sin \ \ t$ is equal to $\frac{}{(s \quad a)^2 \qquad ^2}$. In this case, a=2 and =5; therefore, the Laplace transform of $e^{\ 2t} \sin 5t$ is equal to $\frac{5}{(s \quad 2)^2 \quad 25}$.

4.3 Laplace Transform Solutions of Linear Differential Equations

Laplace transforms can be used to solve linear dynamic equations. Figure 4.3.1 schematically shows this process. A linear differential equation in the time domain is transformed into the Laplace domain by taking the Laplace transform of each term on both sides of the equation, which is illustrated in Figure 4.3.1 by the vertical arrow between the upper and lower left-hand blocks. That is, from the linear operation property of Laplace transforms (Equation 4.2.3), the equality of the equation can be maintained by applying Laplace transforms to the entire equation. For example, consider the following equation

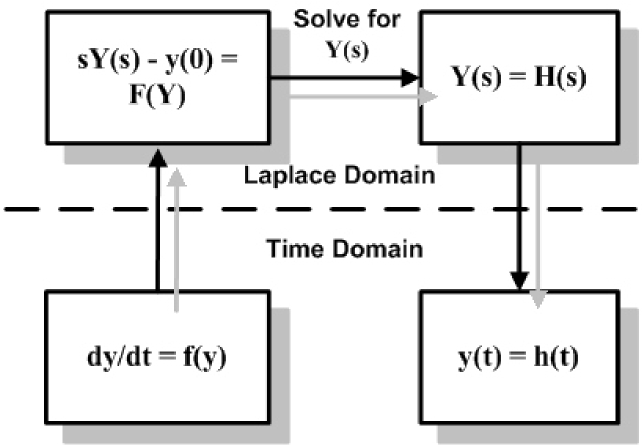


Figure 4.3.1 Schematic that shows how Laplace transforms can be used to solve linear differential equations.

$$\frac{d\,y(t)}{dt} \qquad y(t) \qquad \text{at } t \quad 0 \quad y(t) \quad y(0)$$

If Laplace transformations are applied to each term in the equation, the following equation in the Laplace transform domain results.

$$sY(s) \quad y(0) \qquad Y(s)$$

Next, the transformed equation is rearranged algebraically to solve explicitly for the dependent variable in the Laplace domain, $Y(s)$, which is represented by the horizontal arrow between the upper left- and right-hand blocks in Figure 4.3.1. Even though the previous equation represents the differential equation in the Laplace domain, it contains $y(0)$, which is the initial condition (i.e., a constant) from the time domain. For the example under consideration,

$$Y(s) \quad \frac{y(0)}{s \quad 1} \quad H(s)$$

Finally, the time-dependent behavior of the dependent variable, $y(t)$, is obtained by applying the inverse Laplace transform to $Y(s)$, as shown by the vertical arrow between the upper and lower right-hand blocks in Figure 4.3.1. That is, $H(s)$ must be converted into one of the $F(s)$ forms listed in Table 4.1. Then the time-domain form corresponding to this entry sets the functional form of the time-domain solution, and the values for its parameters come directly from the parameter values in the Laplace domain. For the example under consideration, $Y(s)$ corresponds to $1/(s+a)$ in Table 4.1, which in turn corresponds to e^{-at} in the time domain. Because $y(0)$ is a constant and a is equal to 1 in this case, the time-domain solution is given by

$$y(t) \quad y(0)e^{\ \ t}$$

The key advantage of using Laplace transforms to solve linear differential equations is that, when the differential equation is transformed to the Laplace domain, it becomes an algebraic equation, which is generally easier to solve. Note that this simple example can also be solved in the time domain by separation of variables applied $\frac{dy}{y} \quad dt$ and integration $\ln\frac{y(0)}{y(t)} \quad t$ and solving yields $y(t) \quad y(0)e^{\ \ t}$.

Example 4.6 Solution of a General First-Order Equation

Problem Statement. Consider the following first-order differential equation

$$\frac{dy}{dt} \quad \frac{1}{\ }(u \quad y)$$

where $y(0) = 0$ and   is a constant. Here y is the dependent variable and u is an independent variable. u undergoes a step change from zero to A at time equal to zero. Determine $y(t)$ using Laplace transforms.

Solution. The first step is to apply Laplace transforms to each term in the differential equation. Applying the Laplace transforms listed in Table 4.1 to each term of the differential equation yields

$$sY(s) \quad y(0) \quad \frac{A}{s} \quad \frac{Y(s)}{\ }$$

where $U(s)=A/s$. Rearranging and solving for $Y(s)$ yields

$$Y(s) \quad \frac{A/\ }{s(s \quad 1/\ )} \tag{4.3.1}$$

To apply the inverse Laplace transform, this equation must be converted into the following form

$$Y(s) \quad \frac{C_1}{s} \quad \frac{C_2}{s \quad 1/\ } \tag{4.3.2}$$

where C_1 and C_2 are constants. Combining the two terms on the right-hand side of Equation 4.3.2 yields

$$Y(s) \quad \frac{C_1 s \quad C_1/\ \quad C_2 s}{s(s \quad 1/\ )} \tag{4.3.3}$$

For Equations 4.3.1 and 4.3.3 to be equivalent, the coefficients of s in the numerators of both equations must be equal, i.e., equating the coefficients of s yields

$$C_1 \quad C_2 \quad 0 \tag{4.3.4}$$

Likewise, the constant terms in both numerators must be equal, i.e.,

$$C_1 \quad A$$

Using Equation 4.3.4, $$C_2 \quad A$$

Thus,

$$Y(s) \quad \frac{A}{s} \quad \frac{A}{s \quad 1/}$$

Now Table 4.1 can be used to convert each term in this equation to the time domain (i.e., applying inverse Laplace transforms to each term) resulting in

$$y(t) \quad A(1 \quad e^{\,t/\,})$$

Self-Assessment Questions

Q4.3.1 Explain how Laplace transforms are used to analytically solve differential equations.

Q4.3.2 What is the advantage in using Laplace transforms to solve linear differential equations?

Q4.3.3 After Laplace transforms have been applied to a linear differential equation, what steps are required so that the inverse Laplace transforms can be applied to obtain the time-domain solution?

Self-Assessment Answers

Q4.3.1 Laplace transforms can be used to solve linear differential equations by taking the Laplace transform of each term in the differential equation. Next, the equation is algebraically solved for the dependent variable in the transform space. That is, the Laplace transform of the dependent variable, $F(s)$, is expressed as a function of s and the initial conditions. Finally, the inverse Laplace transform is applied to the previous equation, on the left hand side of the equation, $F(s)$ becomes $f(t)$ and the inverse Laplace transform of the right hand side of the equation is the analytical expression of $f(t)$ in terms of t which represents the time solution of the original differential equation.

Q4.3.2 The primary advantage of using Laplace transforms to solve linear differential equations is that this method is relatively easy to implement, i.e., each step requires standard procedures that can be directly applied. In general, Laplace transforms use algebraic equations to solve differential equations.

Q4.3.3 After Laplace transforms have been applied to a linear differential equation, the equation for $F(s)$ must be converted into a series of terms that allow the application of inverse Laplace transforms. That is, for example $F(s)$ must be converted into terms consistent with the Laplace transforms listed in Table 4.1.

Self-Assessment Problem

P4.3.1 Using Laplace transforms, determine $y(t)$ for the following differential equation.

$$\frac{dy}{dt} \quad y \qquad\qquad y(0) \quad y_0$$

Self-Assessment Answers

P4.3.1 Taking the Laplace transform of each term in the differential equation yields $sY(s) \quad y_0 \quad Y(s)$. Solving for $Y(s)$ yields $Y(s) \quad \frac{y_0}{s \quad 1}$. Applying an inverse Laplace transform yields $y(t)=y_0\, e^{-t}$.

4.4 Individual Real Poles

In this section, the solution of Laplace transforms with individual real poles is considered. In the previous example (Example 4.6), it was necessary to convert Equation 4.3.1 into the form of Equation 4.3.2 so that inverse Laplace transforms could be easily applied. That is, each term in the right hand side of Equation 4.3.2 has a convenient inverse Laplace transform. Consider a general Laplace transforms function, $Y(s)$, expressed as the ratio of two functions of s:

$$Y(s) = \frac{N(s)}{D(s)}$$

The roots of $D(s)$ are referred to as the **poles** of $Y(s)$. That is, the values of s that render $D(s)=0$ are the poles of $Y(s)$. **Partial fraction expansions** is formed by factoring the denominator, $D(s)$, into its roots and expressing $Y(s)$ as a series of terms each of which contains one of the roots in its denominator (i.e., poles), i.e.,

$$Y(s) = \frac{N(s)}{D(s)} = \frac{N(s)}{(s+a_1)(s+a_2)\ldots(s+a_n)} = \frac{C_1}{s+a_1} + \frac{C_2}{s+a_2} + \ldots + \frac{C_n}{s+a_n} \qquad \textbf{4.4.1}$$

For this case, the poles of $Y(s)$ are $-a_1, -a_2, \ldots -a_n$. The coefficients, C_1 to C_n, are calculated using one of several methods available. One method is by equating the terms that contain like powers of s as shown in the previous example. Another method of solving for the coefficients is the **Heaviside method** for partial fraction expansions, which is easier to apply by hand. To determine the value of C_n for the general case just presented (Equation 4.4.1), multiply both sides of the equation by $(s+a_n)$ and evaluate each term at $s=-a_n$, i.e.,

$$\left.\frac{N(s)(s+a_n)}{(s+a_1)(s+a_2)\ldots(s+a_n)}\right|_{s=-a_n} = \left.\frac{C_1(s+a_n)}{s+a_1}\right|_{s=-a_n} + \left.\frac{C_2(s+a_n)}{s+a_2}\right|_{s=-a_n} + \ldots + \left.\frac{C_n(s+a_n)}{s+a_n}\right|_{s=-a_n}$$

Simplifying by taking the limit $s \to -a_n$ results in canceling the $(s+a_n)$ term on the left-hand side of the equation and causes each term on the right hand side of this equation to equal zero except for the term containing C_n,

$$C_n = \left.\frac{N(s)}{(s+a_1)(s+a_2)\ldots(s+a_{n-1})}\right|_{s=-a_n}$$

For the approach used in Example 4.6 to solve for the coefficients, C_i, for the general case, the solution of a system of n equations and n unknowns would be required. For the Heaviside method, a single explicit equation is evaluated for each unknown coefficient. A more convenient equation for coefficient, C_i, is given by

$$C_i = \left.\frac{N(s)}{D(s)/(s+a_i)}\right|_{s=-a_i} \qquad \textbf{4.4.2}$$

where $N(s)$ and $D(s)$ are the numerator and denominator, respectively, of the Laplace transform of the function in question.

Example 4.7 Solution of a General First-Order Equation using the Heaviside Method

Problem Statement. Consider the following first-order differential equation

$$\frac{dy}{dt} \quad \frac{1}{\ }(u \quad y)$$

where $y(0) = 0$ and   is a constant. Here y is the dependent variable and u is an independent variable. u undergoes a step change from zero to A at time equal to zero. Determine $y(t)$ using Laplace transforms with partial fraction expansions and the Heaviside method.

Solution. The first step is to apply Laplace transforms to each term in the differential equation. Applying the Laplace transforms listed in Table 4.1 to each term of the differential equation yields

$$sY(s) \quad y(0) \quad \frac{A}{s} \quad \frac{Y(s)}{\ }$$

Rearranging and solving for $Y(s)$ yields

$$Y(s) \quad \frac{A/\ }{s(s \quad 1/\ )} \tag{4.4.3}$$

Applying partial fraction expansions

$$Y(s) \quad \frac{C_1}{s} \quad \frac{C_2}{s \quad 1/\ }$$

Using the Heaviside method (Equation 4.4.2), C_1 is given by

$$C_1 \quad \left. \frac{\frac{A/\ }{s(s \quad 1/\ )}}{s} \right|_{s \quad 0} \quad \frac{A/\ }{1/\ } \quad A$$

Using the Heaviside method (Equation 4.4.2), C_2 is given by

$$C_2 \quad \left. \frac{\frac{A/\ }{s(s \quad 1/\ )}}{(s \quad 1/\ )} \right|_{s \quad 1/\ } \quad \frac{A/\ }{1/\ } \quad A$$

which agrees with the results obtained from Example 4.6. Therefore, as before

$$Y(s) \quad \frac{A}{s} \quad \frac{A}{s \quad 1/}$$

Now Table 4.1 can be used to convert each term in this equation to the time domain (i.e., applying inverse Laplace transforms to each term) resulting in

$$y(t) \quad A(1 \quad e^{\;t/}\;)$$

Example 4.8 Solution of the CST Thermal Mixer Equation

Problem Statement. The dynamic equation for a CST thermal mixing tank is given by Equation 3.7.7, i.e.,

$$M\frac{dT}{dt} \quad F_1 T_1 \quad F_2 T_2 \quad (F_1 \quad F_2)T$$

where T is the dependent variable of the process and F_1, F_2, T_1 and T_2 are inputs to the process. Determine the time behavior of T, assuming that at time equal to zero, T is equal to T_0 and a step change in T_1 of magnitude A is implemented at time equal to zero while all other inputs remain constant (i.e., F_1, F_2, and T_2 are constant).

Solution. The first step is to take the Laplace transform of each term in the previous equation using Table 4.1.

$$M[sT(s) \quad T_0] \quad \frac{F_1[T_1 \quad A] \;\; F_2 T_2}{s} \quad (F_1 \quad F_2)T(s)$$

Algebraically solving for $T(s)$ yields

$$T(s) \quad \frac{M T_0 \quad \dfrac{F_1 A \quad F_1 T_1 \quad F_2 T_2}{s}}{M s \quad (F_1 \quad F_2)}$$

Rearranging into the form $N(s)/D(s)$ yields

$$T(s) \quad \frac{MT_0 s \quad F_1 A \quad F_1 T_1 \quad F_2 T_2}{s[Ms \quad (F_1 \quad F_2)]}$$

To apply the inverse Laplace transforms, we need to use partial fraction expansions, i.e.,

$$T(s) \quad \frac{C_1}{s} \quad \frac{C_2}{s \quad (F_1 \quad F_2)/M}$$

Applying the Heaviside method,

$$C_1 \quad \left.\frac{MT_0 s \quad F_1 A \quad F_1 T_1 \quad F_2 T_2}{\dfrac{s[Ms \quad F_1 \quad F_2]}{s}}\right|_{s \quad 0} \quad \frac{F_1 A \quad F_1 T_1 \quad F_2 T_2}{F_1 \quad F_2}$$

$$C_2 \quad \left.\frac{MT_0 s \quad F_1 A \quad F_1 T_1 \quad F_2 T_2}{\dfrac{s[Ms \quad F_1 \quad F_2]}{[Ms \quad F_1 \quad F_2]}}\right|_{s \quad \frac{F_1 \quad F_2}{M}} \quad MT_0 \quad \frac{M[F_1 A \quad F_1 T_1 \quad F_2 T_2]}{F_1 \quad F_2}$$

Then $$T(s) \quad \frac{F_1 A \quad F_1 T_1 \quad F_2 T_2}{s(F_1 \quad F_2)} \quad T_0 \quad \frac{F_1 A \quad F_1 T_1 \quad F_2 T_2}{F_1 \quad F_2} \quad \frac{1}{s \quad (F_1 \quad F_2)/M}$$

Applying the inverse Laplace transform yields:

$$T(t) \quad \frac{F_1 A \quad F_1 T_1 \quad F_2 T_2}{F_1 \quad F_2} \quad T_0 \quad \frac{F_1 A \quad F_1 T_1 \quad F_2 T_2}{F_1 \quad F_2} \quad e^{\;t \;\, F_1 \;\, F_2 \;\, /M}$$

Note that at $t = 0$, $T \quad T_0$ and as t

$$T \quad \frac{F_1 A \quad F_1 T_1 \quad F_2 T_2}{F_1 \quad F_2} \quad T$$

The solution can also be written as

$$T \quad T \quad (T \quad T_0)e^{\;t(F_1 \;\, F_2)/M}$$

Self-Assessment Questions

Q4.4.1 What are the poles of a transfer function?

Q4.4.2 Why are partial fraction expansions applied to Laplace transforms?

Q4.4.3 How are partial fraction expansions applied to Laplace transforms?

Q4.4.4 What is the connection between the values of the real poles of a Laplace transform and the time-domain solution?

Self-Assessment Answers

Q4.4.1 When a Laplace transform is expressed as $F(s)=N(s)/D(s)$, the value of s that render $D(s)=0$ (i.e., the roots) are the poles of $F(s)$.

Q4.4.2 Partial fractions are applied to Laplace transforms to put the Laplace transform into a form that can be conveniently transformed to the time domain.

Q4.4.3 Partial fraction expansions are formed by factoring the denominator of a Laplace transform into it roots (poles of the Laplace transform) and expressing the Laplace transform as a series of terms each of which contains one of the roots in its denominator:

$$Y(s) \quad \frac{N(s)}{D(s)} \quad \frac{N(s)}{(s \quad a_1)(s \quad a_2)\ldots(s \quad a_n)} \quad \frac{C_1}{s \quad a_1} \quad \frac{C_2}{s \quad a_2} \quad \ldots \quad \frac{C_n}{s \quad a_n}$$

Q4.4.4 The values of the real poles of a Laplace transform are converted into exponential terms when the Laplace transform is converted to the time domain. That is, if a is a pole, when $F(s)$ converted to the time domain, the time-domain solution will contain an e^{at} term.

Self-Assessment Problems

P4.4.1 Apply partial fraction expansions to the following Laplace transform

$$Y(s) \quad \frac{1}{s^2 \quad 5s \quad 6}$$

P4.4.2 Determine the time-domain solution for the Laplace transform given in the previous problem.

Self-Assessment Answers

P4.4.1 Factoring the denominator and applying a partial fraction expansion yields

$$Y(s) \quad \frac{1}{s^2 \quad 5s \quad 6} \quad \frac{1}{(s \quad 2)(s \quad 3)} \quad \frac{C_1}{s \quad 2} \quad \frac{C_2}{s \quad 3}$$

Applying the Heaviside method yields

$$C_1 \quad \left.\frac{N(s)}{D(s)/(s \quad 2)}\right|_{s \quad 2} \quad \left.\frac{1}{s \quad 3}\right|_{s \quad 2} \quad 1 \qquad C_2 \quad \left.\frac{N(s)}{D(s)/(s \quad 3)}\right|_{s \quad 3} \quad \left.\frac{1}{s \quad 2}\right|_{s \quad 3} \quad 1$$

Finally, $$Y(s) \quad \frac{1}{s \quad 2} \quad \frac{1}{s \quad 3}$$

P4.4.2 Applying inverse Laplace transforms (Table 4.1) to the result from P4.4.1 yields $y(t) \quad e^{\;2t} \quad e^{\;3t}$

4.5 Repeated Real Poles

In this section, the solution of Laplace transforms with repeated real poles is considered. The poles of a Laplace transform function are not always individual real poles as shown in Equation 4.4.1. For example, some functions have repeated poles or complex poles. Consider a general Laplace transform function with n repeated poles,

$$Y(s) \quad \frac{N(s)}{D(s)} \quad \frac{N(s)}{(s \quad a)^n}$$

Applying partial fraction expansions for this case creates a series of terms with the full range of factors in the denominator from $(s+a)^n$ to $(s+a)$, i.e.,

$$Y(s) \quad \frac{N(s)}{(s \quad a)^n} \quad \frac{C_1}{(s \quad a)} \quad \frac{C_2}{(s \quad a)^2} \quad \ldots \quad \frac{C_n}{(s \quad a)^n} \qquad \textbf{4.5.1}$$

While C_n can be found using Equation 4.4.2, the determination of C_1-C_{n-1} requires a special formulation. The values of C_1-C_{n-1} can be determined by multiplying Equation 4.5.1 by $(s+a)^n$ and then taking the derivative of the resulting equation the appropriate number of times. First, Equation 4.5.1 is multiplied by $(s+a)^n$

$$N(s) = C_1(s+a)^{n-1} + C_2(s+a)^{n-2} + \ldots + C_{n-2}(s+a)^2 + C_{n-1}(s+a) + C_n$$

Next differentiate both sides of this equation with respect to s

$$\frac{dN(s)}{ds} = (n-1)C_1(s+a)^{n-2} + (n-2)C_2(s+a)^{n-3} + \ldots + 2C_{n-2}(s+a) + C_{n-1} \qquad \textbf{4.5.2}$$

Evaluating this expression at s=-a yields

$$\left.\frac{dN(s)}{ds}\right|_{s=-a} = C_{n-1}$$

Likewise, Equation 4.5.2 can be differentiated again with respect to s yielding the value for C_{n-2}, when s is set equal to -a, i.e.,

$$\left.\frac{d^2N(s)}{ds^2}\right|_{s=-a} = 2C_{n-2}$$

The general equation for C_i is

$$C_{n-i} = \left.\frac{1}{i!}\frac{d^i}{ds^i}\left[\frac{(s+a)^n N(s)}{D(s)}\right]\right|_{s=-a} \qquad \textbf{4.5.3}$$

Repeated poles are not very common in process systems because the two or more poles must be exactly equal. Repeated roots occur for two identical tanks in series and for sections of a distillation column when the vapor and liquid traffic through that section is constant. In addition, repeated poles can result under certain conditions for feedback systems, which will be addressed in Chapter 8.

Example 4.9 Application of Partial Fraction Expansions for a Case with Repeated Poles

Problem Statement. Consider the following function in the Laplace domain:

$$Y(s) = \frac{1}{(2s+1)^2(s+1)}$$

Determine the values of the coefficients of the partial fraction expansions for $Y(s)$.

Solution. First, the numerator and denominator of $Y(s)$ are divided by 4 to convert it into a form for which the coefficients of s in the factors are unity, i.e.,

$$Y(s) = \frac{\tfrac{1}{4}}{(s+\tfrac{1}{2})^2(s+1)}$$

Now $Y(s)$ is expressed as a series of individual factors using Equations 4.4.1 and 4.5.1.

$$Y(s) = \frac{\tfrac{1}{4}}{(s+\tfrac{1}{2})^2(s+1)} = \frac{C_1}{(s+\tfrac{1}{2})} + \frac{C_2}{(s+\tfrac{1}{2})^2} + \frac{C_3}{(s+1)}$$

Applying Equation 4.4.2 to determine C_2 and C_3 yields

$$C_2 = \left.\frac{\tfrac{1}{4}}{(s+1)}\right|_{s=-\tfrac{1}{2}} = \tfrac{1}{2}$$

$$C_3 = \left.\frac{\tfrac{1}{4}}{(s+\tfrac{1}{2})^2}\right|_{s=-1} = 1$$

Note that you cannot find C_1 using Equation 4.4.2 because dividing the denominator by $(s+\tfrac{1}{2})$ still leaves a $(s+\tfrac{1}{2})$ factor in the denominator, which when evaluated at $s=-\tfrac{1}{2}$ yields an indeterminate result. However, C_1 can be determined by applying Equation 4.5.3. That is, for $I=1$ in Equation 4.5.3

$$C_1 = \frac{d}{ds}\left.\left[\frac{(s+a)^n N(s)}{D(s)}\right]\right|_{s=-a} = \frac{d}{ds}\left[\frac{(s+\tfrac{1}{2})^2 \tfrac{1}{4}}{(s+\tfrac{1}{2})^2(s+1)}\right]_{s=-\tfrac{1}{2}} = \tfrac{1}{4}\frac{d}{ds}\left[\frac{1}{s+1}\right]_{s=-\tfrac{1}{2}} = -\tfrac{1}{4}\left[\frac{1}{(s+1)^2}\right]_{s=-\tfrac{1}{2}} = -1$$

Finally, $Y(s)$ is given by

$$Y(s) = \frac{-1}{s+\tfrac{1}{2}} + \frac{\tfrac{1}{2}}{(s+\tfrac{1}{2})^2} + \frac{1}{s+1}$$

Using Table 4.1, you can easily find $y(t)$.

Self-Assessment Questions

Q4.5.1 How are partial fraction expansions for repeated poles different than partial fraction expansions for individual poles?

Q4.5.2 How do you evaluate the constants that result from partial fraction expansions of a Laplace transform that has repeated poles?

Self-Assessment Answers

Q4.5.1 The application of partial fraction expansion for repeated poles results in a series of terms equal to the number of repeated poles with a full range of factors ranging from $(s+a)$ to $(s+a)^n$, i.e,

$$Y(s) = \frac{N(s)}{(s+a)^n} = \frac{C_1}{(s+a)} + \frac{C_2}{(s+a)^2} + \cdots + \frac{C_n}{(s+a)^n}$$

Q4.5.2 C_n is determined using the Heaviside method and the remainder of the constants are determined using Equation 4.5.3.

Self-Assessment Problem

P4.5.1 Determine the time-domain solution for the following Laplace transform

$$Y(s) = \frac{s+3}{(s^2+2s+1)(s+2)}$$

Self-Assessment Answers

P4.5.1 Factor the denominator and apply partial fraction expansion.

$$Y(s) = \frac{s+3}{(s^2+2s+1)(s+2)} = \frac{s+3}{(s+1)^2(s+2)} = \frac{C_1}{s+1} + \frac{C_2}{(s+1)^2} + \frac{C_3}{s+2}$$

C_2 and C_3 can be evaluated using Equation 4.4.2.

$$C_2 = \left.\frac{s+3}{s+2}\right|_{s=-1} = 2 \qquad C_3 = \left.\frac{s+3}{(s+1)^2}\right|_{s=-2} = 1$$

To determine C_1, apply Equation 4.5.3 with i=1

$$C_1 = \left.\frac{d}{ds}\left(\frac{(s+1)^2(s+3)}{(s+1)^2(s+2)}\right)\right|_{s=-1} = \left.\frac{d}{ds}\frac{(s+3)}{(s+2)}\right|_{s=-1} = \left.\left(\frac{1}{s+2} - \frac{s+3}{(s+2)^2}\right)\right|_{s=-1} = 1 - \frac{2}{1} = -1$$

Then,

$$Y(s) = -\frac{1}{s+1} + \frac{2}{(s+1)^2} + \frac{1}{s+2}$$

Applying inverse Laplace transforms results in $y(t) = -e^{-t} + 2t\,e^{-t} + e^{-2t}$

4.6 Complex Poles

In certain cases, the poles of a Laplace transform function can have complex components. Complex poles can result from quadratic factors in s in the denominator of $Y(s)$ when the roots of the quadratic term have imaginary roots. For example, consider the following case

$$Y(s) = \frac{1}{s^2 + as + b}$$

The poles of $Y(s)$ in this case are

$$s = \frac{-a \pm \sqrt{a^2 - 4b}}{2}$$

When $a^2<4b$, imaginary components for the poles of $Y(s)$ result. Consider a case for which, $a^2<4b$. By completing the square of the terms (s^2+as) by adding and subtracting $\frac{1}{4}a^2$ results in

$$Y(s) = \frac{1}{(s^2 + as + \tfrac{1}{4}a^2) + b - \tfrac{1}{4}a^2} = \frac{1}{\sqrt{b - \tfrac{1}{4}a^2}} \frac{\sqrt{b - \tfrac{1}{4}a^2}}{(s + \tfrac{1}{2}a)^2 + (b - \tfrac{1}{4}a^2)}$$

Note that because $a^2<4b$, the last term in the denominator ($b-\tfrac{1}{4}a^2$) is positive. Also, note that from Table 4.1, the inverse Laplace transform of this Laplace transform function is a damped sinusoidal time-domain response, i.e., it corresponds to $e^{-\alpha t} \sin \omega t$ in the time domain. Therefore, applying inverse Laplace transforms to $Y(s)$ yields

$$y(t) = \frac{1}{\sqrt{b - \tfrac{1}{4}a^2}} e^{-at/2} \sin\left(\sqrt{b - \tfrac{1}{4}a^2}\, t\right)$$

The open-loop behavior of most processes in the process industries are not oscillatory, and hence, do not have complex poles. On the other hand, feedback systems usually exhibit oscillatory behavior, and therefore, these feedback systems have complex poles. This point will be illustrated in more detail in Chapter 8.

The stem position of a globe control valve or a shock absorber can exhibit second-order behavior. The application of Newton's second law to the stem position of a globe control valve or to a shock absorber results in a general second-order differential equation, i.e.,

$$m\frac{d^2x}{dt^2} + B\frac{dx}{dt} + kx = f(t)$$

Here, x is the stem position, m is the mass of the stem, B is the coefficient of friction for movement of the stem past the valve packing and k is the spring constant for the spring in the valve actuator. $f(t)$ is the air pressure forcing function applied to the valve actuator. Assuming zero initial conditions, i.e., $x(0)=0$ and $dx(0)/dt=0$, and applying Laplace transforms results in

$$ms^2X(s) + BX(s) + kX(s) = F(s)$$

Solving for $X(s)$

$$X(s) = \frac{F(s)}{ms^2 + Bs + k}$$

This system will have complex poles (i.e., oscillatory behavior) if $4mk>B^2$. That is, when the product of the mass and spring constant out weighs the friction on the valve stem, oscillatory behavior can result. Also, note that when $B=0$ (i.e., no friction on the valve stem), a harmonic oscillator results, i.e., the stem will oscillate in a sustained fashion without damping.

Example 4.10 Solution of Second-Order Differential Equations

Problem Statement. Consider a general second-order differential equation

$$\frac{d^2y}{dt^2} + a\frac{dy}{dt} + by = u(t)$$

where

$$\left.\frac{dy}{dt}\right|_{t\quad 0} \qquad 0$$
$$y(0) \qquad 0$$

and $u(t)$ is a step change from 0 to 1 at $t = 0$. Evaluate the solutions for this differential equation. Note that because this is a second-order differential equation, two initial conditions are required to solve it.

Solution. Using Table 4.1 to apply Laplace transforms to each term in the differential equation results in

$$s^2 Y(s) \quad s\,y(0) \quad \left.\frac{dy}{dt}\right|_{t\quad 0} \quad a[s\,Y(s) \quad y(0)] \quad b\,Y(s) \quad 1/s$$

Simplifying and solving for $Y(s)$,

$$Y(s) \quad \frac{1}{s(s^2 \quad as \quad b)}$$

Before this equation can be converted back to the time domain by applying inverse Laplace transforms, the roots of the denominator must be calculated so that partial fraction expansions can be used. Applying the quadratic formula to factor the denominator into its roots yields

$$Y(s) \quad \frac{1}{s\left(s \quad \dfrac{a \quad \sqrt{a^2 \quad 4b}}{2}\right)\left(s \quad \dfrac{a \quad \sqrt{a^2 \quad 4b}}{2}\right)} \tag{4.6.1}$$

Three cases can result for this system:

Case 1	$a^2 \quad 4b \quad 0$
Case 2	$a^2 \quad 4b \quad 0$
Case 3	$a^2 \quad 4b \quad 0$

Case 1 Consider the case in which $a = 5$ and $b = 6$ (i.e., $a^2 \quad 4b \quad 1$).

$$Y(s) \quad \frac{1}{s(s^2 \quad 5s \quad 6)} \quad \frac{1}{s(s \quad 2)(s \quad 3)}$$

Applying standard partial fraction expansions

$$Y(s) \quad \frac{1}{s(s \quad 2)(s \quad 3)} \quad \frac{C_1}{s} \quad \frac{C_2}{s \quad 2} \quad \frac{C_3}{s \quad 3}$$

Applying the Heaviside method (Equation 4.4.2),

$$C_1 \quad \left.\frac{1}{(s \quad 2)(s \quad 3)}\right|_{s \quad 0} \quad 1/6 \qquad C_2 \quad \left.\frac{1}{s(s \quad 3)}\right|_{s \quad 2} \quad 1/2 \qquad C_3 \quad \left.\frac{1}{s(s \quad 2)}\right|_{s \quad 3} \quad 1/3$$

Therefore,

$$Y(s) \quad \frac{1}{6s} \quad \frac{1}{2(s \quad 2)} \quad \frac{1}{3(s \quad 3)}$$

Applying the inverse Laplace transformation to each term yields

$$y(t) \quad \mathscr{L}^{-1}[Y(s)] = \mathscr{L}^{-1} \frac{1/6}{s} - \mathscr{L}^{-1} \frac{1/2}{s \quad 2} + \mathscr{L}^{-1} \frac{1/3}{s \quad 3}$$

$$y(t) \quad \frac{1}{6} \quad \frac{1}{2} e^{\quad t} \quad \frac{1}{3} e^{\quad 3t}$$

Case 2 Consider the case in which $a = 6$ and $b = 9$ (i.e., $a^2 \quad 4b \quad 0$).

$$Y(s) \quad \frac{1}{s(s^2 \quad 6s \quad 9)} \quad \frac{1}{s(s \quad 3)(s \quad 3)}$$

Applying partial fraction expansions using Equations 4.4.1 and 4.5.1,

$$Y(s) \quad \frac{1}{s(s \quad 3)(s \quad 3)} \quad \frac{C_1}{s} \quad \frac{C_2}{s \quad 3} \quad \frac{C_3}{(s \quad 3)^2}$$

Using the Heaviside method (Equation 4.4.2) to determine the values of C_1 and C_3,

$$C_1 \quad \left.\frac{1}{(s \quad 3)(s \quad 3)}\right|_{s \quad 0} \quad 1/9$$

$$C_3 \quad \left.\frac{1}{s}\right|_{s \quad 3} \quad 1/3$$

To determine the value of C_2, both sides of Y(s) are multiplied by $(s+3)^2$ and differentiated with respect to s. Then the limit is taken as $s \quad -3$

$$\frac{d}{ds}\frac{1}{s}_{\;s \quad 3} \quad \frac{d}{ds}\left[\frac{C_1(s \quad 3)^2}{s} \quad C_2(s \quad 3) \quad C_3\right]_{s \quad 3} \quad C_2$$

$$C_2 \quad 1/9$$

Therefore, $$Y(s) = \frac{1}{9s} - \frac{1}{9(s+3)} - \frac{1}{3(s+3)^2}$$

Applying the inverse Laplace transformation to each term,

$$y(t) = \frac{1}{9} - \frac{1}{9}e^{-3t} - \frac{1}{3}te^{-3t}$$

$$y(t) = \frac{1}{9}[1 - e^{-3t}(1+3t)]$$

This result corresponds to a critically damped response, which will be discussed in Section 5.3.

Case 3 Consider the case in which $a = 2$ and $b = 5$ (i.e., $a^2 - 4b = -16$).

$$Y(s) = \frac{1}{s(s^2+2s+5)}$$

Applying partial fraction expansions,

$$Y(s) = \frac{C_1}{s} + \frac{C_2 s + C_3}{s^2+2s+5}$$

From the Equation 4.6.1, the poles of the second term are ($-1\pm2j$).

Solving for C_2 and C_3 can be done using the Heaviside method, but it involves the use of complex algebra[2]. Instead, lets solve for the unknown coefficients by equating like powers of s similar to the approach used in Example 4.6. By combining the equation for $Y(s)$ that was obtained from the partial fraction expansion,

$$Y(s) = \frac{C_1}{s} + \frac{C_2 s + C_3}{s^2+2s+5} = \frac{C_1(s^2+2s+5) + C_2 s^2 + C_3 s}{s(s^2+2s+5)} = \frac{1}{s(s^2+2s+5)}$$

Equating terms that contain s^0, s and s^2 yields

$$\begin{array}{lll} [s^0] & 5C_1 = 1 & \text{Therefore, } C_1 = 1/5 \\ [s] & 2C_1 + C_3 = 0 & \text{Therefore, } C_3 = -2/5 \\ [s^2] & C_1 + C_2 = 0 & \text{Therefore, } C_2 = -1/5 \end{array}$$

Thus $$Y(s) = \frac{1}{5s} - \frac{s/5 + 2/5}{s^2+2s+5} = \frac{1}{5s} - \frac{1}{5}\frac{s+2}{s^2+2s+5}$$

Completing the square for (s^2+2s) in the denominator of the second term by adding 1 and subtracting 1 yields

$$Y(s) = \frac{1}{5s} - \frac{1}{5}\frac{s+2}{(s+1)^2+2^2}$$

Rearranging this equation into the forms consistent with Table 4.1 yields

$$Y(s) = \frac{1}{5s} - \frac{1}{5}\frac{s+1}{(s+1)^2+2^2} - \frac{1}{5}\frac{\frac{1}{2}(2)}{(s+1)^2+2^2}$$

The inverse Laplace transforms from Table 4.1 yield the following

$$y(t) = \frac{1}{5}[1 - e^{-t}\cos(2t) - \frac{1}{2}e^{-t}\sin(2t)]$$

This equation can be simplified by using the following trigonometric identity

$$A\sin\omega t + B\cos\omega t = \sqrt{A^2+B^2}\,\sin(\omega t+\phi)$$

where $$\phi = \tan^{-1}(B/A)$$

Then $$y(t) = \frac{1}{5} - \frac{1}{5}e^{-t}\frac{\sqrt{5}}{2}\sin(2t+\phi) = \frac{1}{5}\left[1 - \frac{\sqrt{5}}{2}e^{-t}\sin(2t+\phi)\right]$$

$$\phi = \tan^{-1}(2) = 63.43^\circ$$

This result corresponds to damped sinusoidal behavior, which will be described in more detail later in Section 5.3. It is damped because the term (e^{-2t}) becomes small as time increases and the "sin" term creates periodic behavior.

Self-Assessment Questions

Q4.6.1 What type of time-domain behavior results when poles have an imaginary component?

Q4.6.2 Explain why imaginary poles occur as complex conjugate pairs.

Self-Assessment Answers

Q4.6.1 Poles with imaginary components correspond to oscillatory time-domain behavior.

Q4.6.2 A single pole can only result in exponential time behavior. To obtain an imaginary component there must be a quadratic term (as^2+bs+c) with (b^2-4ac)<0.

Self-Assessment Problem

P4.6.1 Determine the time-domain solution for the following Laplace transform

$$Y(s) = \frac{1}{s^2+s+1}$$

Self-Assessment Answers

P4.6.1 Rearranging $Y(s)$ into a standard form yields

$$Y(s) \quad \frac{1}{s^2 \quad s \quad 1} \quad \frac{1}{s^2 \quad s \quad 0.25 \quad 0.75} \quad \frac{1}{(s \quad 0.5)^2 \quad (\sqrt{0.75})^2} \quad \frac{1}{\sqrt{0.75}} \frac{\sqrt{0.75}}{(s \quad 0.5)^2 \quad (\sqrt{0.75})^2}$$

Therefore, taking the inverse Laplace transform yields $y(t) \quad \frac{1}{\sqrt{0.75}} e^{\quad 0.5t} \sin \sqrt{0.75}\, t$.

4.7 Symbolic Solution of ODEs using MATLAB

MATLAB offers a fast and convenient way to analytically solve ordinary differential equations (ODEs) using its symbolic math codes. The MATLAB built-in function `dsolve` analytically solves ODEs by defining the ODE in terms of the unknown function [`y`], the first derivative of y with respect to t [`diff(y)`] and the second derivative of y with respect to t [`diff(y,2)`]. As an example, consider the ODE that was solved using Laplace transforms for Example 4.10 Case 1:

$$\frac{d^2y}{dt^2} \quad 5\frac{dy}{dt} \quad 6y \quad 1 \qquad y(0) \quad \frac{dy}{dt}(0) \quad 0$$

The following MATLAB code solves this problem using `dsolve`:

```
function OdeSoln
syms y(t)
Dy=diff(y);
S=dsolve(diff(y,2)+5*diff(y)+6*y==1,y(0)==0,Dy(0)==0)
end

>> OdeSoln
S =
exp(-3*t)/3 - exp(-2*t)/2 + 1/6
```

The first line of the program defines y as a symbolic function of t [i.e., `syms y(t)`]. Using the terminology specifed above, the differential equation is inserted as the first argument for `dsolve` (Note that the double equal sign is used to define an equality for symbolic equations). The second argument for `dsolve` is the initial condition $y(0)=0$ and the third argument is the initial condition for the first derivative ($dy/dt=0$ *at* $t=0$). Note that in order to apply the first derivative (dy/dt) at $t=0$, `Dy` is defined as the first derivative of y and then `Dy(0)` is used to define the initial condition for the first derivative in `dsolve`. Finally, note that this solution agrees with the solution derived in Example 4.10. It should be pointed out here that many differential equations resulting from modeling of industrial processes are nonlinear and, therefore, cannot be solved analytically. For these cases, a numerical solution is required (see Sections 3.5 and 3.6).

4.8 Summary

- Laplace transforms are a convenient means for solving linear differential equations. The dynamic equation is transformed to the Laplace domain where the dependent variable can be solved for algebraically. Finally, the dependent variable in the Laplace domain is transformed back into the time domain, yielding the time-domain solution to the original differential equation.
- Partial fraction expansions involve expressing a Laplace transform of the form $N(s)/D(s)$ as a series of terms each of which is the inverse of a factor of $D(s)$. The roots of $D(s)=0$ are the poles of the Laplace transform.
- After partial fraction expansions, associated with individual real poles can be evaluated using the Heaviside method in the following equation

$$C_i = \left. \frac{N(s)}{D(s)/(s-a_i)} \right|_{s=a_i}$$

where a_i is individual real pole of interest.

- After partial fraction expansions, associated with repeated real poles can be evaluated using the Heaviside method in the following equation

$$C_{n-i} = \frac{1}{i!}\frac{d^i}{ds^i} \left. \frac{(s-a)^n N(s)}{D(s)} \right|_{s=a}$$

where a is repeated real pole of interest.

- Complex poles correspond to oscillatory time-domain behavior.

4.9 References

1. Finney, R.L., G.B. Thomas, *Calculus*, 2nd ed., Addison-Wesley, New York, NY, pp. 517-8 (1990).
2. Stephanopoulos, G., *Chemical Process Control*, McGraw-Hill, New York, NY, pp. 145-157 (1984)

4.10 Additional Terminology

Final-value theorem - An application of Laplace transforms that yields the long-term behavior of a dependent variable.
Heaviside method - a method for evaluating the coefficients for partial fraction expansions.
Initial-value theorem - An application of Laplace transforms that yields the initial conditions of a dependent variable.
Partial fractional expansions - expansions of a transfer function into a sum of terms, each of which contains one of the factors of the denominator of the transfer function.
Poles - the roots of the denominator of a Laplace transform.

4.11 Preliminary Questions

4.2 Laplace Transforms

Q4.2.1 What is the definition of the Laplace transform of a function, $g(t)$?

Q4.2.2 Give an example of a nonlinear equation not given in the text.

Q4.2.3 How can the final-value theorem be used?

Q4.2.4 How can the initial-value theorem be used?

4.3 Laplace Transform Solutions of Linear Differential Equations

Q4.3.1 What is the major limitation of using Laplace transforms to solve ODE problems?

Q4.3.2 What advantages do Laplace transforms offer for solving linear ODEs?

4.4 Individual Real Poles

Q4.4.1 What are partial fraction expansions?

Q4.4.2 What is the form of the time-domain solution of a Laplace transform that contains individual real poles?

4.5 Repeated Poles

Q4.5.1 What is the form of the time-domain solution of a Laplace transform that contains repeated real poles?

4.6 Complex Poles

Q4.6.1 What is the form of the time-domain solution of a Laplace transform that contains complex poles?

4.12 Analytical Questions and Exercises

4.2 Laplace Transforms

P4.2.1* Derive the Laplace transform for the following function using Equation 4.2.1.

$$f(t) \quad \frac{1}{a_1 \quad a_2} \quad e^{a_2 t} \quad e^{a_1 t}$$

P4.2.2** Derive the Laplace transform of $f(t)=at$ using Equation 4.2.1.

P4.2.3** Derive the Laplace transform of $f(t) \quad e^{at} \sin \quad t$ using Equation 4.2.1. (Hint: Use the result of the integration by parts presented in Example 4.3.)

P4.2.4* Determine the Laplace transform of $f(t)$ for the following functions using Table 4.1

a. $f(t) \quad e^{at} \quad e^{at} \sin \quad t$

b. $f(t) \quad \frac{d\,g(t)}{dt} \quad e^{2t} \cos 5t$

c. $f(t) \quad \int_0^t g(t)\,dt \quad t^3 e^{5t}$

d. $f(t) \quad 7g(t \quad 10) \quad \sin 10t$

P4.2.5* Apply the initial- and final-value theorems to $f(t)$ for the following functions.

a. $f(t) \quad e^{at} \sin \quad t$

b. $f(t) \quad t^2 e^{5t} \quad \sin 5t$

c. $f(t) \quad e^{2t}\sin 2t \quad e^{2t}$ d. $f(t) \quad \sin \quad t \quad \cos \quad t$

4.3-4.6 Laplace Transform Solutions of Linear Differential Equations

P4.3.1** Find the final value of $y(t)$ for a unit step input [$u(t)$=1] using the following Laplace transform. Check you answer by developing the time-domain solution and take the limits as t .

$$Y(s) \quad \frac{1}{(s \quad 5)(s \quad 5)}U(s)$$

P4.3.2* Apply partial fraction expansions to the following functions.

a. $F(s) \quad \dfrac{2}{(s \quad 2)(s \quad 3)}$ b. $F(s) \quad \dfrac{2}{s^2 \quad 11s \quad 30}$

c. $F(s) \quad \dfrac{7}{(s \quad 1)(s \quad 2)(s \quad 6)}$ d. $F(s) \quad \dfrac{3}{s^2 \quad 3s \quad 2}$

P4.3.3* Determine $y(t)$ by applying partial fraction expansions and inverse Laplace transforms for the following cases

a. $Y(s) \quad \dfrac{s \quad 1}{(s \quad 2)(s \quad 3)}$ b. $Y(s) \quad \dfrac{1}{(s \quad 1)(s \quad 2)}$

c. $Y(s) \quad \dfrac{s \quad 3}{(s \quad 1)^2}$ d. $Y(s) \quad \dfrac{s \quad 2}{(s \quad 1)(s \quad 6)(s \quad 7)}$

P4.3.4* For Problem 4.3.3, apply the initial-value theorem and the final-value theorem.

P4.3.5* Solve the following differential equation using Laplace transforms.

$$\frac{dy}{dt} \quad t^2 \qquad\qquad y(0) \quad 0$$

P4.3.6* Determine $y(t)$ for each of the following differential equations, assuming that

$$y(0) \quad \left.\frac{dy}{dt}\right|_{t \quad 0} \quad 0$$

a. $\dfrac{d^2y}{dt^2} \quad 5\dfrac{dy}{dt} \quad 6y \quad 2$ b. $\dfrac{d^2y}{dt^2} \quad 3\dfrac{dy}{dt} \quad 5$

c. $\dfrac{d^2y}{dt^2} \quad 4\dfrac{dy}{dt} \quad 5y \quad 4$ d. $\dfrac{d^2y}{dt^2} \quad 4\dfrac{dy}{dt} \quad 4y \quad 1$

e. $2\dfrac{d^2y}{dt^2} \quad 11\dfrac{dy}{dt} \quad 12y \quad 5$ f. $3\dfrac{d^2y}{dt^2} \quad 7\dfrac{dy}{dt} \quad 2y \quad 6t$

P4.3.7** For Problem 4.3.2a-d, indicate the expected time-domain behavior based on the given Laplace transform.

P4.3.8** For Problem 4.3.2a-d, determine the differential equation that corresponds to these Laplace transforms assuming that an impulse input was used.

P4.3.9** For the following differential equation, determine its Laplace transform and indicate the character of the time-domain response without solving for $y(t)$.

$$\frac{d^2y}{dt^2} \quad 5\frac{dy}{dt} \quad 6y \quad \sin 2t$$

P4.3.10** Consider the isothermal CSTR shown in Figure P4.3.10. A single irreversible first-order reaction occurs in the reactor, i.e., $A \quad B$ where $r \quad kC_A$. The unsteady-state model balance for this process is

$$V\frac{dC_A}{dt} \quad F_V(C_{A_0} \quad C_A) \quad VkC_A$$

where

- C_A - the concentration of reactant A in the reactor (initially 1.0 gmol/l)
- C_{A_0} - the concentration of reactant A in the feed to the CSTR (1.0 gmol/l)
- F_V - the volumetric feed rate and product flow rate for the CSTR (100 l/min)
- k - the rate constant for the first-order reaction (0.1 min^{-1})
- V - the volume of liquid in the CSTR (1000 l)

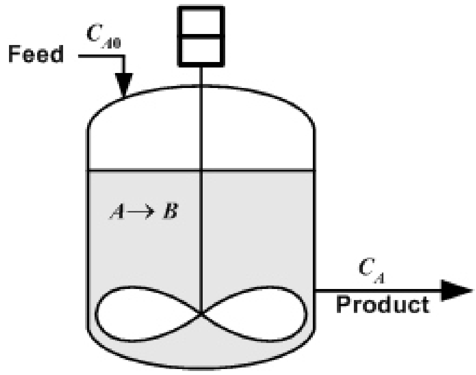


Figure P4.3.10 Schematic of an isothermal CSTR with a single irreversible reaction.

Using Laplace transforms, develop an analytical expression for C_A as a function of time (min).

P4.3.11** Consider the tank level process shown in Figure P4.3.11. A mass balance for this process yields the following process model.

$$A_c\frac{dL}{dt} \quad F_{in} \quad F_{out}$$

Initially, the level is constant and at t=0 the inlet feed rate (F_{in}) is decreased by 10%. Develop an analytical expression for the level as a function of time in hours using Laplace transforms assuming that F_{out} remains constant. Following are the process variables and parameters

- A_c - cross-sectional area (0.3 m^2)
- L - the level of liquid in the tank (initially 2 meters)
- F_{in} - the mass flow rate of liquid into the tank (initially 1.0 kg/s)
- F_{out} - the mass flow rate of liquid leaving the tank (1.0 kg/s).
- - the fluid density (1 kg/l)

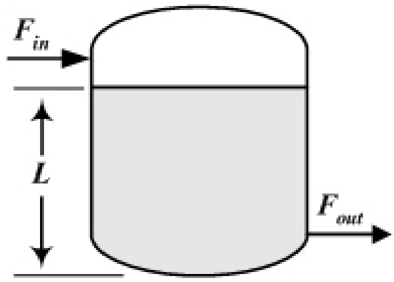


Figure P4.3.11 PFD of a tank level.

P4.3.12** Consider the batch bio-reactor shown in Figure P4.3.12. Consider that the cells in the bio-reactor are in the exponential growth phase described by the specific growth rate, . Applying a mass balance for the cells in the batch bio-reactor yields

$$\frac{dx}{dt} \quad x$$

where the rate of generation of cell is equal to $V \quad x$. The process variables and parameters are

- x - cell concentration (initially 1.0 g/l)
- V - the reactor volume
- - specific growth rate [0.5 g-cells produced/(g-cell h^{-1})]

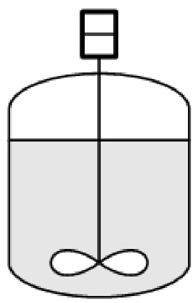


Figure P4.3.12 Schematic of a batch bio-reactor.

Using Laplace transforms, develop an analytical solution for cell concentration as a function of time in h.

P4.3.13* ** For the following set of ODEs

$$\frac{dy_1}{dt} \quad 2y_1 \quad y_2 \quad 2 \qquad\qquad y_1(0) \quad 0$$

$$\frac{dy_2}{dt} \quad y_2 \quad y_1 \quad 0 \qquad\qquad y_2(0) \quad 0$$

a. Solve for $Y_1(s)$ by eliminating $Y_2(s)$. Then determine $Y_2(s)$.

b. Solve for $y_1(t)$ and $y_2(t)$.

4.7 Symbolic Solution of ODEs using MATLAB

P4.7.1* Determine $y(t)$ for each of the following differential equations using MATLAB's `dsolve`, assuming that

$$y(0) \quad \frac{dy}{dt}_{t \quad 0} \quad 0$$

a. $\frac{d^2y}{dt^2} \quad 5\frac{dy}{dt} \quad 6y \quad 2$

b. $\frac{d^2y}{dt^2} \quad 3\frac{dy}{dt} \quad 5$

c. $\frac{d^2y}{dt^2} \quad 4\frac{dy}{dt} \quad 5y \quad 4$

d. $\frac{d^2y}{dt^2} \quad 4\frac{dy}{dt} \quad 4y \quad 1$

e. $2\frac{d^2y}{dt^2} \quad 11\frac{dy}{dt} \quad 12y \quad 5$

f. $3\frac{d^2y}{dt^2} \quad 7\frac{dy}{dt} \quad 2y \quad 6t$

P4.7.2** Consider the isothermal CSTR shown in Figure P4.3.10. A single irreversible first-order reaction occurs in the reactor, i.e., $A \quad B$ where $r \quad kC_A$. The unsteady-state model balance for this process is

$$V\frac{dC_A}{dt} \quad F_V(C_{A_0} \quad C_A) \quad VkC_A$$

where

- C_A - the concentration of reactant A in the reactor (initially 1.0 gmol/l)
- C_{A_0} - the concentration of reactant A in the feed to the CSTR (1.0 gmol/l)
- F_V - the volumetric feed rate and product flow rate for the CSTR (100 l/min)
- k - the rate constant for the first-order reaction (0.1 min^{-1})
- V - the volume of liquid in the CSTR (1000 l)

Using MATLAB's built-in function `dsolve`, develop an analytical expression for C_A as a function of time (min).

Chapter 5

Transfer Functions and State Space Models

Chapter Objectives

- Introduce transfer functions and demonstrate how the structure of a transfer function defines the resulting dynamic behavior of the system.
- Illustrate how to combine the transfer functions of a block diagram.
- Demonstrate how to develop a transfer function from a nonlinear dynamic equation by linearization.
- Show how to develop state space models from a set of nonlinear dynamic models.

5.1 Introduction

In the previous chapter, Laplace transforms were used to solve linear differential equations based on specified inputs. In this chapter, transfer functions, which are the Laplace transform of the output divided by the Laplace transform of the input for a process, are introduced and shown to represent the behavior of a process independent of the process input or the initial conditions applied to the process. The primary advantage of transfer functions is that they provide a concise representation of the dynamic behavior of a process. Transfer functions can also be derived for nonlinear process models. Using the techniques presented in this chapter, transfer functions will be used in Chapters 7 to 9 to analyze the behavior of feedback systems. In addition, this chapter will introduce state space models and how to convert them to a transfer function form.

5.2 General Characteristics of Transfer Functions

In this section, Laplace transforms will be used to develop transfer functions, which relate the output (dependent variable) to the input (independent variable). A **transfer function**, $G(s)$, is defined as

$$G(s) \quad \frac{Y(s)}{U(s)} \qquad \textbf{5.2.1}$$

where $Y'(s)$ is the Laplace transform of the output variable and $U'(s)$ is the Laplace transform of the input variable, both written in **deviation variable** form. Deviation variables [i.e., $y'(t)$ and $u'(t)$] represent changes in a variable from initial steady-state conditions. The **procedure for converting a differential equation**, which represents the effect of $u(t)$ on $y(t)$, **into transfer function is as follows**:

1. Define $y'(t)$ and $u'(t)$ as $y'(t) = y(t) - \bar{y}$ and $u'(t) = u(t) - \bar{u}$, where $\bar{y}$ and $\bar{u}$ are the initial steady-state conditions. By definition, **the initial conditions for $y'(t)$ and $u'(t)$ are zero**.

2. Using the definitions for $y'(t)$ and $u'(t)$, eliminate $y(t)$ and $u(t)$ from the differential equation, resulting in a differential equation in terms of $y'(t)$ and $u'(t)$.

3. Eliminate $\bar{y}$ and $\bar{u}$ from the differential equation, which may require the application of the initial conditions to the original differential equation.

4. Take the Laplace transform of each term in the model equation, which is in deviation variable form.

5. Solve for the ratio of $Y(s)/U(s)$.

Example 5.1 Conversion of a Differential Equation to Deviation Variable Form

Problem Statement. Convert the process model equation for a level (Equation 3.7.9) into deviation variable form.

Solution. The model equation for the level in a tank (Equation 3.7.9 in Example 3.5) is given as

$$A_c \frac{dL}{dt} = F_{in} - F_{out} \qquad \textbf{3.6.9}$$

L, F_{in} and F_{out} require conversion to deviation variable form; therefore, equilibrium values (e.g., the initial steady-state conditions) should be selected for the level ($\bar{L}$) and the flow rate ($\bar{F}$). Then, the deviation variable can be defined as

$$L' = L - \bar{L}$$
$$F'_{in} = F_{in} - \bar{F}$$
$$F'_{out} = F_{out} - \bar{F}$$

Note that F_{in} and F_{out} have the same value ($\bar{F}$) at the initial steady-state conditions. Solving for the original variables (L, F_{in} and F_{out}) and substituting into Equation 3.6.9 yields

$$A_c \frac{d(L' + \bar{L})}{dt} = (F'_{in} + \bar{F}) - (F'_{out} + \bar{F})$$

Simplifying this equation, remembering that $\bar{L}$ and $\bar{F}$ are constants, results in the deviation variable form of this differential equation,

$$A_c \frac{dL}{dt} \quad F_{in} \quad F_{out}$$

Example 5.2 Conversion of a Differential Equation to Deviation Variable Form

Problem Statement. Convert the process model for the CST composition mixer (Equation 3.7.8) into deviation variable form where C and C_1 are the only variables in the equation and the system is initially at steady state at

$$C \quad \overline{C} \quad C_1 \quad \overline{C}_1$$

Solution. The model equation for the CST composition mixer (Equation 3.7.8 in Example 3.4) is given as

$$V \frac{dC}{dt} \quad F_1 C_1 \quad F_2 C_2 \quad (F_1 \quad F_2) C \qquad \textbf{3.5.8}$$

The deviation variables are written as

$$C \quad C \quad \overline{C} \qquad\qquad C_1 \quad C_1 \quad \overline{C}_1$$

Rearranging the equations for the deviation variables to solve for C and C_1 and substituting into Equation 3.7.8 yields

$$V \frac{d(C \quad \overline{C})}{dt} \quad F_1 (C_1 \quad \overline{C}_1) \quad F_2 C_2 \quad (F_1 \quad F_2)(C \quad \overline{C})$$

Simplifying and rearranging results in

$$V \frac{dC}{dt} \quad F_1 C_1 \quad (F_1 \quad F_2) C \quad F_1 \overline{C}_1 \quad F_2 C_2 \quad (F_1 \quad F_2) \overline{C} \qquad \textbf{5.2.2}$$

Because the process is initially at steady state, i.e.,

$$V \frac{d\overline{C}}{dt} \quad F_1 \overline{C}_1 \quad F_2 C_2 \quad (F_1 \quad F_2) \overline{C} \quad 0$$

the sum of the last three terms in Equation 5.2.2 is zero. Therefore,

$$V \frac{dC}{dt} \quad F_1 C_1 \quad (F_1 \quad F_2) C \qquad \textbf{5.2.3}$$

Example 5.3 Derivation of a Transfer Function

Problem Statement. Convert the differential equation for a first-order process given in Example 4.6 into deviation variable form and determine the transfer function for the process.

Solution. The model equation for a first-order process is given by

$$\frac{dy}{dt} \qquad \frac{1}{\ }(u \quad y)$$

where y and u are equal to $\bar{y}$ at time equal to zero. First, write each variable (y - output variable and u - input variable) in deviation variable form

$$\begin{aligned} y &\quad y \quad \bar{y} \\ u &\quad u \quad \bar{y} \end{aligned}$$

Rearranging yields

$$\begin{aligned} y &\quad y \quad \bar{y} \\ u &\quad u \quad \bar{y} \end{aligned}$$

Substituting these equations into the differential equation for the model results in

$$\frac{dy}{dt} \qquad \frac{1}{\ }(u \quad y\ )$$

Taking the Laplace transform of this equation using the initial conditions $y \quad u \quad 0$ at $t \quad 0$ yields

$$sY(s) \qquad \frac{1}{\ }[U(s) \quad Y(s)]$$

Solving for $Y(s)$ yields

$$Y(s) \qquad \frac{1}{s \quad 1}U(s)$$

Then the transfer function of this system is given by

$$G(s) \quad \frac{Y(s)}{U(s)} \quad \frac{1}{s \quad 1}$$

Note that the transfer function for a first-order system contains s to the first power in the denominator.

Example 5.4 Derivation of the Transfer Function for a PID Controller

Problem Statement. Develop the transfer function for a PID controller in which the input is $e(t)$ [the error from setpoint] and the output is $c(t)$ [the controller output], i.e.,

$$c(t) \quad \bar{c} \quad K_c \;\; e(t) \quad \frac{1}{{}_I} \;\; {}^{t}_{0}\, e(t)\,dt \qquad {}_D \frac{d\,e(t)}{dt}$$

where K_c, ${}_I$ and ${}_D$ are tuning constants and $\bar{c}$ is the initial value of $c(t)$.

Solution. $e(t)$ is already in deviation variable form and the deviation variable for $c(t)$ is given as

$$c\,(t) \quad c(t) \quad \bar{c}$$

Applying Laplace transforms using Table 4.1,

$$C(s) \quad K_c \;\; E(s) \quad \frac{E(s)}{{}_I\, s} \qquad {}_D\, s\, E(s)$$

Factoring out $E(s)$ and solving for $C(s)/E(s)$ yields

$$G_c(s) \quad \frac{C(s)}{E(s)} \quad K_c \;\; 1 \quad \frac{1}{{}_I\, s} \qquad {}_D\, s$$

Example 5.5 Predicting the Dynamic Behavior of a Process Using a Transfer Function.

Problem Statement. Determine the time-domain behavior of the first-order system considered in Example 5.3 for (a) a step input change of magnitude A, (b) a unit impulse input change and (c) a ramp input of slope B.

Solution. (a) For a step input change of magnitude A, from Table 4.1, the Laplace transform for the input variable $U(s)$ is given as

$$U(s) \quad \frac{A}{s}$$

Multiplying $G(s)$ by $U(s)$ yields $Y(s)$, i.e.,

$$Y(s) \quad G(s)U(s) \quad \frac{Y(s)}{U(s)}\; U(s) \quad \frac{A}{s(\;\; s \quad 1)} \quad \frac{A\,/\;\;}{s(s \quad 1/\;\;)}$$

To develop a time-domain solution for this equation, partial fraction expansions and evaluation of the resulting constants are required, i.e.,

$$Y(s) \quad \frac{C_1}{s} \quad \frac{C_2}{s \quad 1/\;\;} \quad \frac{A}{s} \quad \frac{A}{s \quad 1/\;\;}$$

Using Table 4.1 to apply inverse Laplace transforms results in

$$y\,(t) \quad A(1 \quad e^{\;\;t/\;\;}\;)$$

(b) If a unit impulse input is used, $U(s)$ becomes

$$U(s) \quad 1$$

Then
$$Y(s) \quad G(s)U(s) \quad \frac{1}{s \quad 1} \quad \frac{1/}{s \quad 1/}$$

and applying an inverse Laplace transform yields the time-domain solution,

$$y(t) \quad \frac{1}{\ } e^{\ \ t/}$$

(c) If a ramp input is used, $\qquad U(s) \quad B / s^2$

Forming the product of $G(s)$ and $U(s)$, applying partial fraction expansions and evaluating the unknown constants

$$Y(s) \quad G(s)U(s) \quad \frac{B}{s^2(\ s \quad 1)} \quad \frac{B/}{s^2(s \quad 1/\ )} \quad \frac{C_1}{s^2} \quad \frac{C_2}{s} \quad \frac{C_3}{s \quad 1/} \quad \frac{B}{s^2} \quad \frac{B}{s} \quad \frac{B}{s \quad 1/}$$

The time-domain solution can be obtained by the application of inverse Laplace transforms applied to each term, resulting in

$$y(t) \quad Bt \quad B \ (1 \quad e^{\ \ t/}\ )$$

Thus, **transfer functions can be used with a wide variety of inputs to determine the output behavior**. Moreover, a transfer function indicates the general dynamic behavior of the process it represents. For this example, each of the time-domain solutions had exponential time dependence (i.e., $e^{-t/}$), which is indicated by a $\frac{1}{s \quad 1/}$ term in the denominator of the transfer function.

Following are some of the **properties of transfer functions**, i.e., $G(s)=N(s)/P(s)$:

1. The steady-state gain of the process represented by the transfer function (i.e., $y/ \ u$) can be determined directly from the transfer function by evaluating the transfer function at s equals to zero.

2. The order of the denominator [$P(s)$] of the transfer function is the same order of the differential equation from which it was derived. Moreover, the roots of $P(s)$ (i.e., the poles which are discussed in the next section) have a major effect of the dynamics of the process represented by the transfer function.

3. The roots of $N(s)$ (the zeros of the transfer function which are discussed in Section 5.5) can also affect the dynamic response.

4. The order $P(s)$ (i.e., the largest power of s in the polynomial) should be greater than or equal to the order of $N(s)$ for **physical realizability**, i.e., for the transfer function to be physically realistic.

Self-Assessment Questions

Q5.2.1 In general terms, what does a transfer function represent?

Q5.2.2 What are deviation variables and what are they based on?

Q5.2.3 In general terms, how is a transfer function derived from a dynamic model equation?

Self-Assessment Answers

Q5.2.1 A transfer function is a linear representation of the input/output characteristics of a process in the Laplace domain. More specifically, the transfer function of a system is the ratio of the Laplace transform of the output variable in deviation variable form divided by the Laplace transform of the input variable also in deviation variable form. The transfer function contains information about the steady-state gain of the system as well as its dynamic characteristics.

Q5.2.2 Deviation variables represent a change in a variable from the initial steady-state conditions. They are based on the initial steady-state conditions.

Q5.2.3 1. Define the deviation variables for the input variables [$u(t)$] and the output variables [$y(t)$] based on the initial steady-state conditions of the system. 2. Convert the process model equation (linear ODE) into deviation variable form. 3. Take the Laplace transform of each term in the model equation in deviation variable form. 4. Solve for the ratio of $Y(s)/U(s)$.

Self-Assessment Problem

P5.2.1 Convert the following equation into deviation variable form where $y(t)$ is the dependent variable and $u(t)$ is the input variable and solve for the transfer function for this system.

$$\frac{dy(t)}{dt} = -y(t) + u(t) \qquad y(0) = \bar{y};\ u(0) = \bar{y}$$

Self-Assessment Answers

P5.2.1 Define the deviation variables for $y(t)$ and $u(t)$, i.e., $y'(t) = y(t) - \bar{y}$; $u'(t) = u(t) - \bar{u}$. Solve for $y(t)$ and $u(t)$, i.e., $y(t) = \bar{y} + y'(t)$; $u(t) = \bar{u} + u'(t)$ and substitute them into the differential equation, i.e.,

$$\frac{dy(t)}{dt} = \frac{d(\bar{y} + y'(t))}{dt} = -y(t) + u(t) = -\bar{y} - y'(t) + \bar{u} + y'(t)$$

Simplifying this expression recognizing that $\bar{y} = \bar{u} = 0$ from the initial conditions, the differential equation in deviation variable form is $\frac{dy'(t)}{dt} = -y'(t) + u'(t)$ where $y'(0) = u'(0) = 0$. Applying Laplace transforms to this equation yields $sY(s) = -Y(s) + U(s)$. Rearranging yield the transfer function for this system: $Y(s)/U(s) = 1/(1+s)$.

5.3 Poles of a Transfer Function

Consider the general form of a transfer function where $N(s)$ and $D(s)$ are polynomials in s

$$G(s) = \frac{N(s)}{D(s)} \tag{5.3.1}$$

The roots of $D(s)=0$ [i.e., the values of s that render $D(s) = 0$] are called the **poles of the transfer function**. The poles of the Laplace transform of a function, $Y(s)$, were introduced in Section 4.4. Because $Y(s)=G(s)U(s)$, the poles of $Y(s)$ are the same as the poles of $G(s)$, except that $U(s)$ can add additional poles to $Y(s)$. Assume that $D(s)$ can be factored into a series of real poles, p_i.

$$G(s) = \frac{N(s)}{(s - p_1)(s - p_2)\ldots(s - p_n)} = \frac{Y(s)}{U(s)} \qquad \textbf{5.3.2}$$

Assume that a unit step input [$U(s) = 1/s$] is applied to the process in question.

$$Y(s) = \frac{N(s)}{s(s - p_1)(s - p_2)\ldots(s - p_n)} \qquad \textbf{5.3.3}$$

By partial fraction expansions assuming that the order of $N(s)$ is less than or equal to the order of $D(s)$, this transfer function can be written as

$$Y(s) = \frac{C_0}{s} + \frac{C_1}{s - p_1} + \frac{C_2}{s - p_2} + \ldots + \frac{C_n}{s - p_n} \qquad \textbf{5.3.4}$$

Taking the inverse Laplace transform yields $\quad y(t) = C_0 + C_1 e^{p_1 t} + C_2 e^{p_2 t} + \ldots + C_n e^{p_n t} \qquad$ **5.3.5**

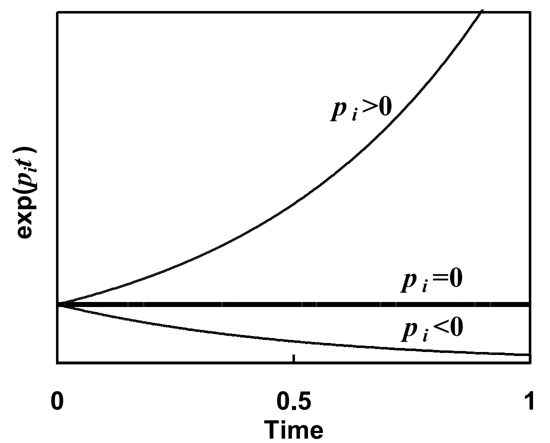


Figure 5.3.1 Different types of exponential behavior, i.e., exponential growth (p_i>0); constant value (p_i=0); and exponential decay (p_i<0).

Although the values of the constant terms are not known at this point, the dynamic behavior of the process can be determined directly from the poles of the transfer function. Figure 5.3.1 shows the time behavior of $e^{p_i t}$ for positive, zero and negative values of p_i. Note that positive values of p_i result in exponential growth with time and negative values of p_i result in exponential decay approaching zero.

Now assume that one of the factors of $D(s)$ is $(s^2 + as + b)$. After partial fraction expansions, a term with the following form results

$$\frac{C}{s^2 + as + b} \qquad \textbf{5.3.6}$$

Now factoring yields

$$\frac{C}{\left(s - \frac{-a + \sqrt{a^2 - 4b}}{2}\right)\left(s - \frac{-a - \sqrt{a^2 - 4b}}{2}\right)} \qquad \textbf{5.3.7}$$

For Equation 5.3.7:

If $a^2 - 4b > 0$, then real roots exist and the previous results for real roots apply.

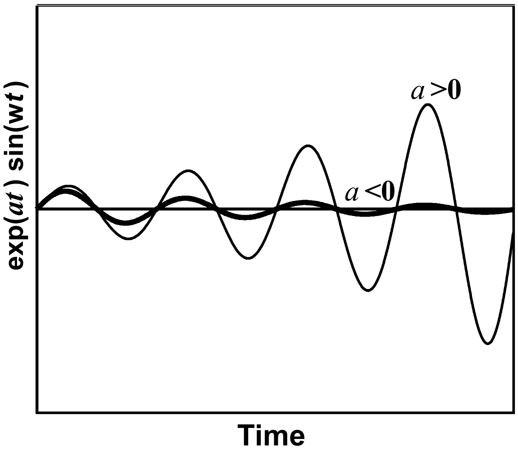


Figure 5.3.2 Exponentially growing (*a* < 0) and exponentially damped (*a* >0) sinusoidal behavior.

If $a^2 \quad 4b \quad 0$, there are two real roots that are equal. Under these conditions, repeated poles result, i.e., $s \quad\quad a/2$. After partial fraction expansions, a term with the following form results

$$\frac{C}{(s \quad a/2)^2}$$

which corresponds to the time-domain solution using Table 4.1, i.e., $y(t) \quad C_1 te^{\quad at/2} \quad C_2 e^{\quad at/2}$.

If $a^2 \quad 4b \quad 0$, imaginary roots result for Equation 5.3.7. For this case, return to Equation 5.3.6, completing the square for the denominator.

$$\frac{C}{s^2 \quad as \quad b} \quad \frac{C}{s^2 \quad as \quad \tfrac{1}{4}a^2 \quad b \quad \tfrac{1}{4}a^2} \quad \frac{C}{(s \quad \tfrac{1}{2}a)^2 \quad (\sqrt{b \quad \tfrac{1}{4}a^2})^2}$$

This set of complex roots ($s \quad\quad \tfrac{1}{2}a \quad i\sqrt{b \quad \tfrac{1}{4}a^2}$) is called a **complex conjugate pair** of poles. The inverse Laplace transform yields

$$\frac{C}{\sqrt{b \quad \tfrac{1}{4}a^2}} e^{\quad at/2} \sin \sqrt{b \quad \tfrac{1}{4}a^2} \quad t$$

Figure 5.3.2 shows the time behavior of this term for $a < 0$ and $a > 0$, which show exponentially growing sinusoidal behavior and damped sinusoidal behavior, respectively. When $a > 0$, the larger the magnitude of a, the faster the sinusoidal response will damp out with time, i.e., approach zero. Likewise, when $a < 0$, the larger the magnitude of a, the faster the sinusoidal response will grow.

For this case, if a=0, one of the factors of $D(s)$ is $(s^2 \quad b)$. After partial fraction expansions, a term with the following form results

$$\frac{C}{s^2 \quad b}$$

The roots of this term are $s \quad i\sqrt{b}$ and $s \quad\quad i\sqrt{b}$. The inverse Laplace transform yields

$$\frac{C}{\sqrt{b}} \sin \sqrt{b}\, t$$

which corresponds to sinusoidal behavior with a fixed amplitude of $C/\sqrt{b}$. This type of dynamic behavior results in fixed amplitude oscillations, which is known as **sustained oscillations** or "harmonic oscillations".

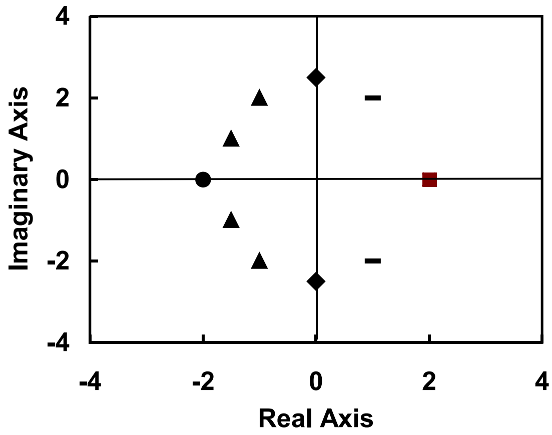


Figure 5.3.3 A complex plane with different types of poles: (●) exponential decay; (▲) damped sinusoidal; ( ) sustained oscillations; (-) unbounded sinusoidal exponential growth; (■) unbounded exponential growth.

Figure 5.3.3 shows a plot of poles on the complex plane, which plots the real and imaginary parts of each pole. The pole represented by a circle symbol (●) corresponds to a real negative pole; therefore, this pole results in exponential decay (Figure 5.3.4a-b). The complex conjugate poles represented by the triangle symbols (▲) in Figure 5.3.3 have a negative real component with equal magnitude positive and negative imaginary parts; therefore, these complex conjugate poles result in damped oscillatory behavior (Figure 5.3.4c-f). Because the poles in Figure 5.3.4e have a real part with a smaller magnitude, the oscillations of the response in Figure 5.3.4f do not damp out as fast as the ones in Figure 5.3.4d. Oscillations that are slow to damp out (e.g., Figure 5.3.4f) are referred to industrially as **ringing**. The poles represented by diamond symbols ( ) correspond to sustained oscillations (Figures 5.3.4g-h). The complex conjugate poles represented by dash symbols (-) have a positive real part; therefore, these poles result in oscillatory behavior that grows exponentially in amplitude (Figures 5.3.4i-j).

The real pole represented by a square symbol (■) indicates unbounded exponential growth (Figures 5.3.4k-l). Note that poles in the right-half plane (i.e., the real part of the pole is positive) of this figure have exponential behavior that grows without bound as time increases, referred to as **unstable behavior**. That is, a process is unstable when bounded (i.e., limited) input changes result in unbounded growth in the dependent variable.

These results show that **the poles of a transfer function [i.e., the roots of $D(s)$] indicate very specifically the type of dynamic behavior for the systems that the transfer function represents**. In other words, by simply determining the poles of a transfer function, you will automatically know the general dynamic behavior of the process represented by the transfer function. Table 5.1 shows different types of transfer functions and their corresponding dynamic behavior. This table shows that the general dynamic behavior of a process can be determined directly by examining the denominator of the transfer function of the process.

Self-Assessment Questions

Q5.3.1 Why do the poles of a transfer function represent the dynamic behavior of the process represented by the transfer function?

Q5.3.2 What type of pole indicates dynamic behavior as an exponential decay?

Q5.3.3 What kind of dynamic behavior is referred to as "unstable"?

Q5.3.4 What type of pole indicates oscillatory dynamic behavior?

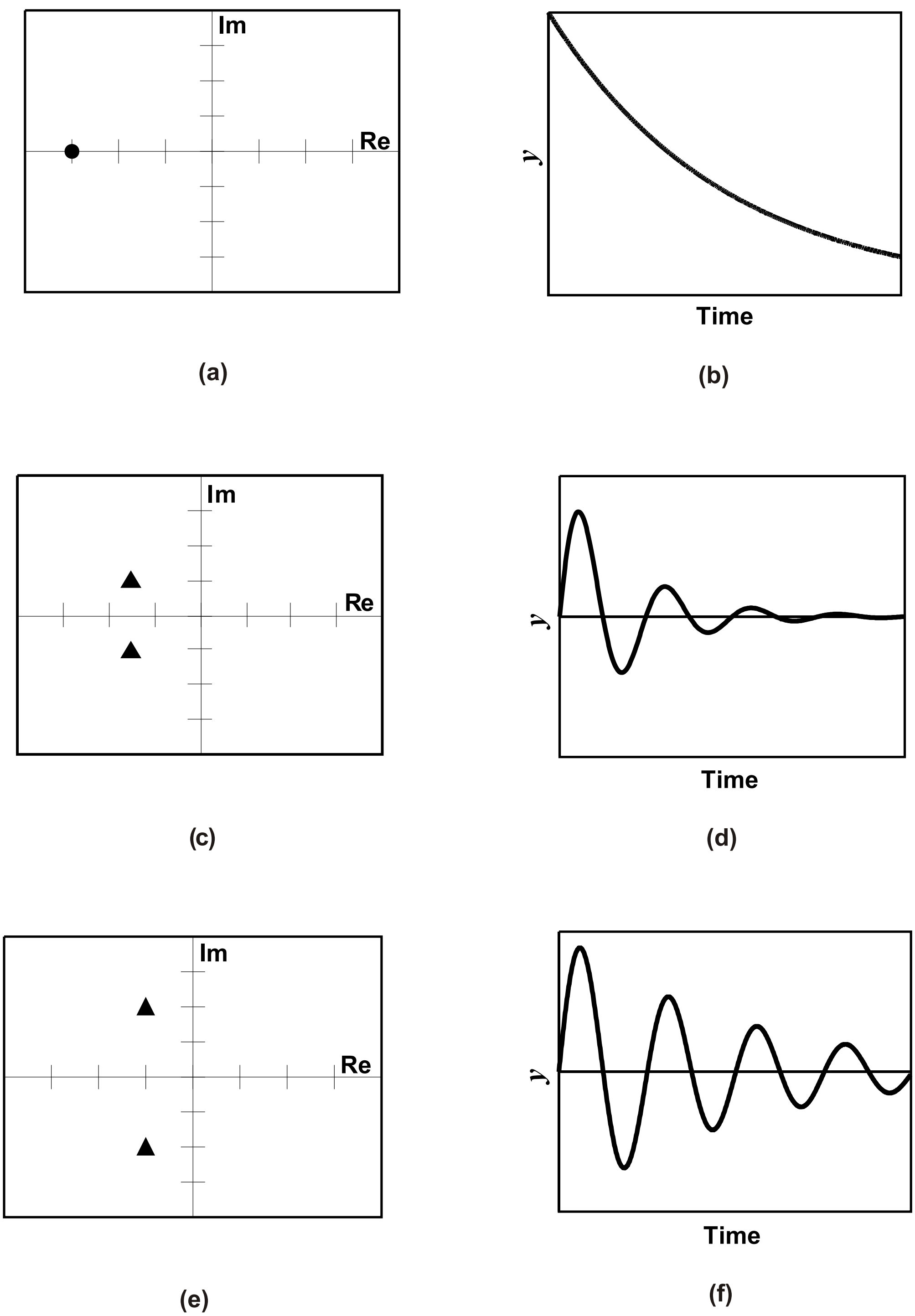


Figure 5.3.4 The correspondence between poles on a complex plane and dynamic behavior in the time domain. (a) and (b) exponential decay; (c) and (d) damped sinusoidal behavior; (e) and (f) damped sinusoidal behavior (ringing).

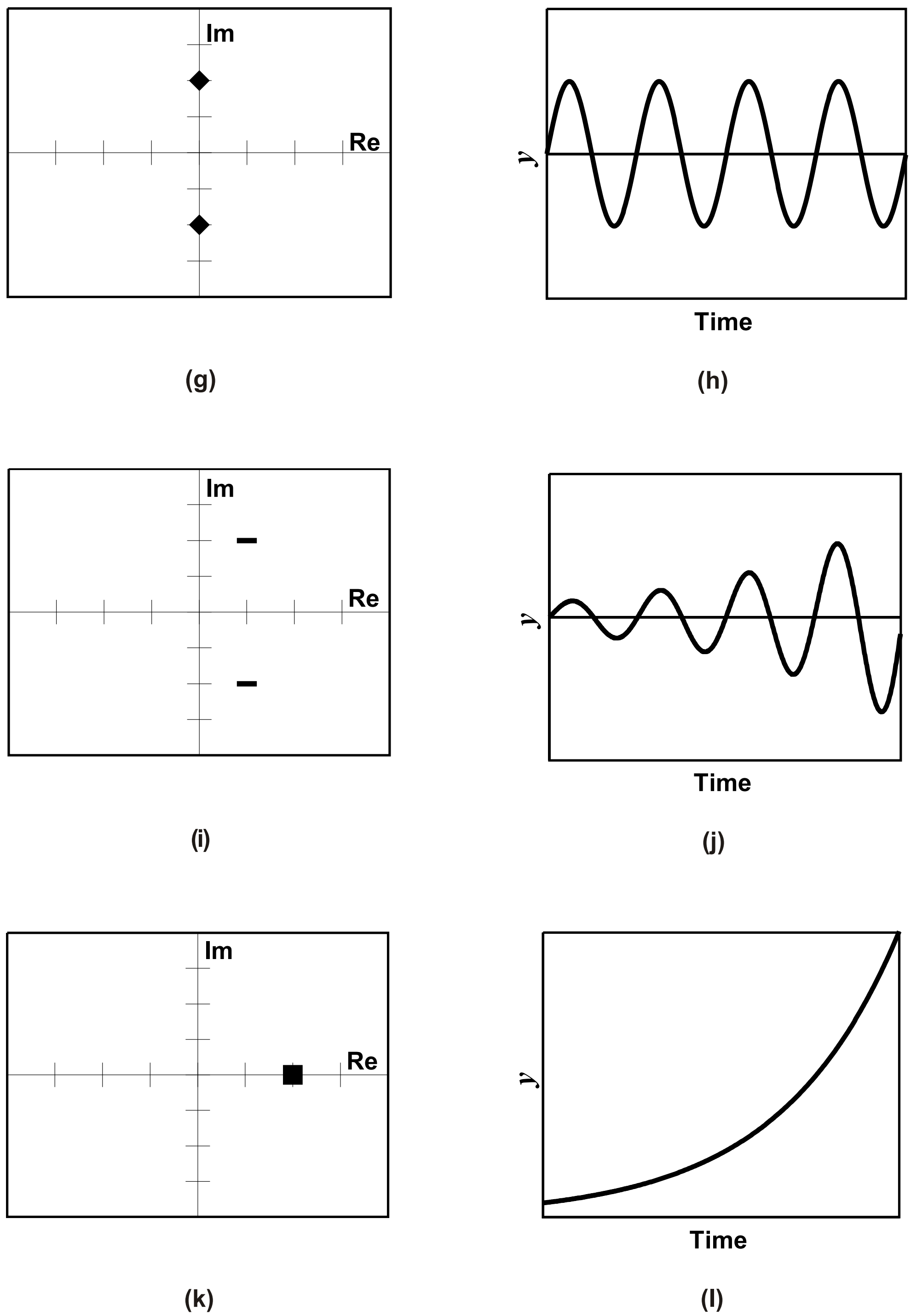


Figure 5.3.4 The correspondence between poles on a complex plane and dynamic behavior in the time domain. (g) and (h) sustained oscillations; (i) and (j) exponential growing sinusoidal behavior; (k) and (l) unbounded exponential growth.

Self-Assessment Answers

Q5.3.1 The poles of a transfer function are the roots of the denominator, $D(s)$. After application of partial fraction expansions of a transfer function times $U(s)$, the poles are retained in the denominator of each term resulting from partial fraction expansions. Therefore, the application of inverse Laplace transform results in time-domain terms that depend directly on the poles of the transfer function. As a result, the poles of a transfer function indicate the form of the time-domain solution of $G(s)U(s)$, but this analysis of the poles of a transfer function does not determine the relative contribution of each pole, which requires a complete solution of the constants of the partial fraction expansions.

Q5.3.2 Real negative poles indicate exponential decay.

Q5.3.3 A positive real portion of a pole (i.e., a right-half plane pole) indicates unbounded exponential growth (unstable behavior).

Q5.3.4 A complex conjugate pair ($p_j \quad a_j \quad ib_j$) indicates oscillatory behavior.

Table 5.1

Different Types of Transfer Functions and their Corresponding Dynamic Behavior

$G(s) \quad \frac{3}{s \quad 1}$	Exponential decay of the form $e^{\quad t}$
$G(s) \quad \frac{3}{s \quad 1}$	Unbounded exponential growth of the form e^{t} (i.e., unstable behavior)
$G(s) \quad \frac{3s \quad 1}{(s \quad 1)\,(s \quad 3)}$	Exponential decay involving terms of the form e^{-t} and e^{-3t}
$G(s) \quad \frac{4}{s^2 \quad 9}$	Sinusoidal behavior of the form sin $3t$
$G(s) \quad \frac{2}{(s \quad 1)^2 \quad 4}$	Damped sinusoidal behavior of the form e^{-t} sin $2t$
$G(s) \quad \frac{2}{(s \quad 1)^2 \quad 4}$	Exponentially growing sinusoidal behavior of the form e^{t} sin $2t$ (i.e., unstable behavior)

5.4 Stability Analysis Using the Routh Array

It has just been shown that the roots of the denominator of a transfer function determine the dynamic response of the process and that if any one of the these roots has a positive real component, unstable behavior results. The Routh array[1], which is an analytical method to determine if any of the roots of a real polynomial have positive real parts, can be used to determine the stability of a transfer function. The Routh array is based on the following equation:

$$a_n s^n \quad a_{n \;\; 1} s^{n \;\; 1} \quad \ldots \quad a_1 s \quad a_0 \quad 0 \qquad \textbf{5.4.1}$$

with each a_i assumed to be positive. If any of the a_is in Equation 5.4.1 are negative, one or more of the roots will have a positive real component; therefore, if Equation 5.4.1 represents the denominator of a transfer function and any of the a_is is negative, the transfer function represents unstable behavior. If all of the a_is are negative, the polynomial can be converted into the form for Equation 5.4.1 by simply multiplying the polynomial equation by -1. That is, when all the a_is are negative, it does not mean that the system is unstable.

The Routh array is formed as follows:

Row	**Routh Array**			
1	a_n	a_{n-2}	a_{n-4}	. . .
2	a_{n-1}	a_{n-3}	a_{n-5}	. . .
3	A_1	A_2	A_3	. . .
4	B_1	B_2	B_3	. . .
.	.	.	.	. . .
.	.	.	.	. . .
$n+1$	Z_1	0	0	. . .

Note that the Routh array has n+1 rows, where n is the order of the polynomial given in Equation 5.4.1. The first and second rows of the Routh array are simply the coefficients of different powers of s in Equation 5.4.1. For example, the first row is comprised of the coefficients a_n, a_{n-2}, a_{n-4}, etc. while the second row is comprised of the coefficients a_{n-1}, a_{n-3}, a_{n-5}, etc. The third row is calculated from the elements of the first two rows as follows:

$$A_1 = \frac{a_{n-1}a_{n-2} - a_n a_{n-3}}{a_{n-1}}$$

$$A_2 = \frac{a_{n-1}a_{n-4} - a_n a_{n-5}}{a_{n-1}}$$

.

.

.

The fourth row is calculated from the values on the second and third rows:

$$B_1 = \frac{A_1 a_{n-3} - a_{n-1} A_2}{A_1}$$

$$B_2 = \frac{A_1 a_{n-5} - a_{n-1} A_3}{A_1}$$

.

.

.

For example, consider a third-order polynomial:

$$a_3 s^3 \quad a_2 s^2 \quad a_1 s \quad a_0 \quad 0$$

The Routh array for this case is

$$\begin{array}{cc} a_3 & a_1 \\ a_2 & a_0 \\ \dfrac{a_2 a_1 \quad a_3 a_0}{a_2} & 0 \\ a_0 & 0 \end{array} \qquad \textbf{5.4.2}$$

Note that terms in the Routh array involving a_{-1}, a_{-2}, ... are set equal to zero. For a fourth-order polynomial the Routh array is

$$\begin{array}{ccc} a_4 & a_2 & a_0 \\ a_3 & a_1 & 0 \\ \dfrac{a_3 a_2 \quad a_4 a_1}{a_3} & a_0 & 0 \\ a_1 \quad \dfrac{a_3^2 a_0}{a_3 a_2 \quad a_4 a_1} & 0 & 0 \\ a_0 & 0 & 0 \end{array} \qquad \textbf{5.4.3}$$

The **Routh stability criterion** is given as: **A necessary and sufficient condition for all roots of a real polynomial to have negative real parts is that all the elements of the first column of the Routh array are positive.** From Equation 5.4.2, using the Routh stability criterion, the necessary and sufficient conditions for all negative real roots for a cubic polynomial is that $a_2a_1>a_3a_0$ because it is assumed that all the coefficients of the polynomial are non-negative. If any of the a_is are zero, the Routh array can still be used, but it must be modified[2].

Example 5.6 Application of the Routh Stability Criterion.

Problem Statement. Using the Routh stability criterion, determine whether the following transfer function represents a stable or unstable system.

$$G(s) \quad \frac{s^2 \quad 100s \quad 1}{s^4 \quad 10s^3 \quad 100s^2 \quad 10s \quad 1}$$

Solution. The roots of the denominator will determine the dynamics of the process represented by this transfer function. Therefore, applying the Routh stability criterion to the polynomial in s in the denominator will determine the stability of the transfer function. For this fourth-order polynomial, a_4=1, a_3=10, a_2=100, a_1=10 and a_0=1. Substituting these values into Equation 5.4.3 yields

1	100	1
10	10	0
99	1	0
9.9	0	0
1	0	0

Because each coefficient of the polynomial is positive and each element in the first column of the Routh array is positive, this transfer function is stable by the Routh stability criterion.

Self-Assessment Questions

Q5.4.1 How can the Routh stability criterion be used to assess the stability of a system represented by a transfer function?

Q5.4.2 When applying the Routh stability criterion to a polynomial, what assumption is made?

Self-Assessment Answers

Q5.4.1 The Routh stability criterion is applied to the polynomial denominator of a transfer function. From the coefficients of this polynomial, the Routh stability criterion indicates whether there are any positive real roots. Because any positive real roots indicates unstable dynamic behavior, the Routh stability criterion is able to determine the stability of a transfer function.

Q5.4.2 The Routh stability criterion assumes that all the coefficients of the polynomial are positive.

Self-Assessment Problem

P5.4.1 Using the Routh stability criterion, determine the stability of the process represented by the following transfer function

$$G(s) \qquad \frac{2s \quad 1}{10s^3 \quad 2s^2 \quad s \quad 1}$$

Self-Assessment Answers

P5.4.1 The coefficients of the denominator polynomial for the transfer function are $a_0 \quad 1$; $a_1 \quad 1$; $a_2 \quad 2$; $a_3 \quad 10$. Applying Equation 5.4.2

10	1
2	1
4	0
1	0

Due to the negative element in the first column, the Routh stability criterion determines that this system is unstable.

5.5 Zeros of a Transfer Function

The zeros of a transfer function are the roots of the numerator of $G(s)$, i.e., $N(s)$. Although the zeros of a transfer function do not affect the stability of the system, they can have a dramatic effect on the dynamic response. Consider a simple generic transfer function

$$G(s)=\frac{Y(s)}{U(s)}=\frac{N(s)}{D(s)}=\frac{(s+c)}{(s+a)(s+b)} \qquad \textbf{5.5.1}$$

Here the poles are $s=-a$ and $s=-b$ and there is one finite zero at $s=-c$. The dynamic response of the system represented by $G(s)$ can change quite significantly depending on whether c is positive or negative and the relative magnitude of c compared to the values of a and b. It is assumed here that $a>0$ and $b>0$ and that $b>a$. The first assumption guarantees that a stable response will result and the second assumption is made for convenience. Consider the response for a unit step response, i.e., $U(s)=1/s$. Then

$$Y(s)=\frac{1}{s}\,\frac{(s+c)}{(s+a)(s+b)}$$

Using partial fraction expansions (Equation 4.4.1) and the Heaviside method (Equation 4.4.2),

$$Y(s)=\frac{c}{ab}\,\frac{1}{s}+\frac{a-c}{a(b-a)}\,\frac{1}{(s+a)}+\frac{b-c}{b(a-b)}\,\frac{1}{(s+b)}$$

Finally, applying the Laplace inverse to each term yields

$$y(t)=\frac{c}{ab}+\frac{a-c}{a(b-a)}e^{-at}+\frac{b-c}{b(a-b)}e^{-bt} \qquad \textbf{5.5.2}$$

Using the Equation 5.5.2, the system response [$y(t)$] can be analyzed for various values of a, b and c (i.e., a parametric study). If a and b are positive, the response of $y(t)$ will be stable (i.e., bounded). Using the time-domain solution, it can be shown that

$$y(0)=0; \qquad y(\infty)=\frac{c}{ab} \qquad \textbf{5.5.3}$$

$$\left.\frac{dy}{dt}\right|_{t=0}=-\frac{a-c}{b-a}-\frac{b-c}{a-b}=\frac{-a+b}{b-a}=1 \qquad \textbf{5.5.4}$$

Under certain conditions, $y(t)$ will pass through a maximum. To determine the value of t at this maximum, set $dy/dt=0$ and solve for t. That is, by differentiating the time-domain solution (Equation 5.5.2) and setting the result equal to zero yields

$$\frac{dy}{dt}=\frac{a-c}{a-b}e^{-at}-\frac{b-c}{a-b}e^{-bt}=0$$

And by rearranging

$$e^{(a \quad b)t} \quad \frac{b \quad c}{a \quad c}$$

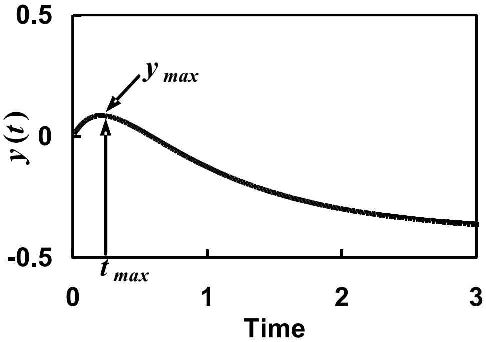


Figure 5.5.1 Response of a system with a positive zero (i.e., c<0).

Taking the natural logarithm of both terms and solving for t yields

$$t_{\max} \quad \frac{\ln \dfrac{b \quad c}{a \quad c}}{b \quad a} \qquad \textbf{5.5.5}$$

Then the maximum value of $y(t)$ can be determined by substituting $t_{\max}$ into the time-domain solution for $y(t)$. Equations 5.5.3-5.5.5 can be used to analyze the effect of zeros on the dynamic behavior of a process.

Systems with Positive Zeros (a>0, b>0, c<0). These cases are referred to as transfer functions with **right-half plane zeros** because the value of the zero in this case is positive (i.e., -c), which is located in the right-half plane. Note that c<0, $y(\quad)$<0 and the initial slope is unity; therefore, the response must pass through a maximum before it settles to a negative value (Figure 5.5.1). If a negative unit step change is applied (i.e., $U(s)$= -1/s), the response will be the mirror image (about the time axis) of Figure 5.5.1. The right-half plane zero results in initial responses that go in the opposite direction to where the long-time solution settles, which is referred to as an **inverse response**. Inverse responses are considered in more detail in Chapter 6.

Example 5.7 Determination of the Response of a Transfer Function with c<0.

Problem Statement. Determine the response of the following transfer function for a unit step input.

$$G(s) \quad \frac{(s \quad 0.2)}{(s \quad 0.9)(s \quad 1)}$$

Solution. For this transfer function based on Equation 5.5.1, c= -0.2, a=0.9 and b=1. Using Equation 5.5.5, $t_{\max}$ is given by

$$t_{\max} \quad \frac{\ln \dfrac{b \quad c}{a \quad c}}{b \quad a} \quad \frac{\ln \dfrac{1.0 \quad (\quad 0.2)}{0.9 \quad (\quad 0.2)}}{1.0 \quad 0.9} \quad 0.870 \text{ time units}$$

$$y_{\max} \quad y(0.667) \quad 0.335$$

And

$$y(\quad) \quad \frac{0.2}{0.9 \quad 1.0} \quad 0.222$$

Figure 5.5.2 shows $y(t)$ [Equation 5.5.2] for this example. Note that this system also exhibits an inverse response and that t_{max}, y_{max} and $y(\quad)$ are consistent with the values calculated.

Systems with Negative Zeros (a>0, b>0, c>0). These cases are referred to as transfer functions with **left-half plane zeros**. The relative position of c compared to a and b can affect the shape of the response to a step change. For some of these cases, a maximum will occur that produces an overshoot in the response.

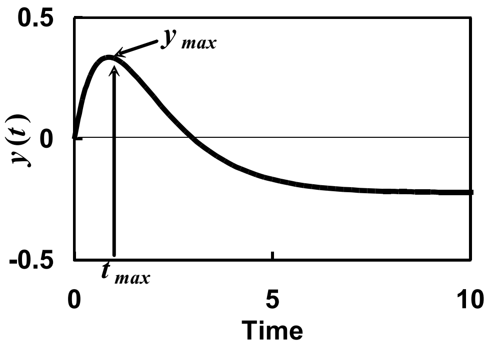


Figure 5.5.2 Response of a system with a right-half plane zero (i.e., a=0.1; b=1.0; c=-1.0).

Case A: $c<a<b$. For this case, a maximum in $y(t)$ results in "**overshoot**". Overshoot results because the maximum value must be greater than the resting value of $y(t)$ [i.e., c/ab] at large times. Figure 5.5.3 shows the results for this case with c=0.05, a=0.1 and b=1.0. The location of the maximum can be determined using Equation 5.5.5:

$$t_{\max} = \frac{\ln\left(\frac{b-c}{a-c}\right)}{b-a} = \frac{\ln\left(\frac{1.0-0.05}{0.1-0.05}\right)}{1.0-0.1} = 3.27 \text{ time units}$$

and the value of $y_{\max}$ can be calculated as before using Equation 5.5.2.

Case B: $b>c>a$. From Equation 5.5.5, it is clear that $t_{\max}$ is indeterminate for this case because the argument for the natural logarithm is negative. This means that there is not a maximum (i.e., no overshoot) for this case for a step input change. $y(t)$ for this case is shown graphically in Figure 5.5.4. Note the monotonic increase from the initial condition to the final resting value (i.e., c/ab).

Case C: $c>b>a$. This case is similar to Case B because $t_{\max}$ is negative in this case. As a result, for t>0, $y(t)$ also shows a monotonic response for this case, except that the final resting value is larger than for Case B because c is larger.

Even though these results were based on a very specific form for a transfer function, a couple of general statements can be made with regard to the zeros of a transfer function. **(1) A positive value of a zero (right-half**

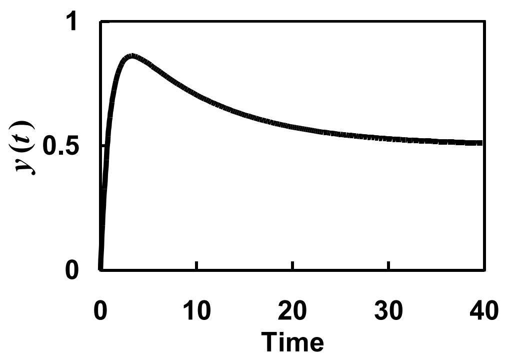


Figure 5.5.3 Response of a system with a negative zero and a>c (i.e., a=0.1; b=1; c=0.05).

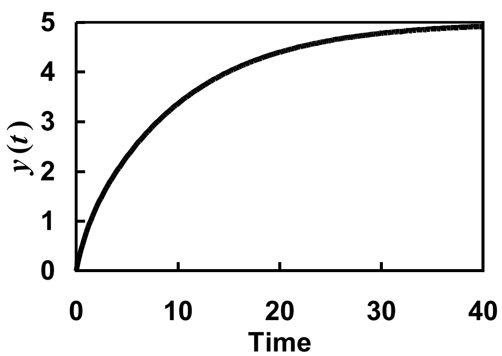


Figure 5.5.4 Response of a system with a left-half plane zero and b>c>a (i.e., a=0.1; b=1; c=0.5).

plane zeros) indicates a transfer function with inverse-acting behavior. (2) Overshoot is indicated if the zero of a transfer function is negative (left-half plane zero) and smaller than the two poles.

Self-Assessment Questions

Q5.5.1 What form of a transfer function indicates an inverse response?

Q5.5.2 What are left-half plane zeros and what type of dynamic behavior do they indicate?

Self-Assessment Answers

Q5.5.1 If the zero of a transfer function is positive, an inverse responding process is indicated.

Q5.5.2 Left-half plane zero indicate either a monotonic (e.g., Figure 5.5.4) or an overshoot response (Figure 5.5.3) will result.

Self-Assessment Problem

P5.5.1 What type of dynamic behavior does the following transfer function indicate?

$$G(s) \quad \frac{3s \quad 2}{s^2 \quad 3s \quad 1}$$

Self-Assessment Answers

P5.5.1 Because this transfer function has a right-half plane zero, an inverse responding process is indicated. Also, the poles are real and negative; therefore, the process should exhibit dynamics with exponential decay.

5.6 Block Diagrams using Transfer Functions

Developing overall transfer functions from block diagrams using transfer function algebra can be useful in understanding a variety of control approaches. The basis of block diagram algebra is related to the properties of a sequence of transfer functions and some simple signal functions.

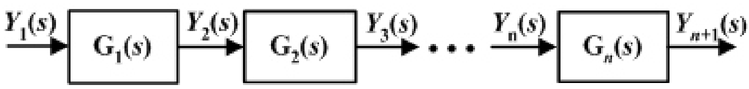


Figure 5.6.1 Schematic of a general series of transfer functions.

A Series of Transfer Functions. Consider the generalized sequence of transfer functions shown in Figure 5.6.1. From the definition of a transfer function, the following equation represents the input/output relationship for any of the transfer functions in this sequence

$$G_i(s) \quad \frac{Y_{i \quad 1}(s)}{Y_i(s)}$$

To determine the relationship between $Y_3(s)$ and $Y_1(s)$, the previous equation will be applied for $G_1(s)$ and $G_2(s)$, i.e.,

$$G_1(s) = \frac{Y_2(s)}{Y_1(s)}$$

$$G_2(s) = \frac{Y_3(s)}{Y_2(s)}$$

Note that by simply multiplying $G_1(s)$ and $G_2(s)$, the relationship between $Y_3(s)$ and $Y_1(s)$ is obtained because the output of $G_1(s)$ cancels the input of $G_2(s)$ in this case.

$$G_1(s)G_2(s) = \frac{Y_2(s)}{Y_1(s)}\frac{Y_3(s)}{Y_2(s)} = \frac{Y_3(s)}{Y_1(s)}$$

When considering a longer sequence of transfer functions, the product of the sequence of transfer functions eliminates the intermediate values so that **the overall transfer function for a sequence of transfer functions is simply the product of transfer functions in the sequence.** Therefore, the overall transfer function for the general sequence of individual transfer functions shown in Figure 5.6.1 is

$$G_{overall}(s) = \frac{Y_{n+1}(s)}{Y_1(s)} = G_1(s)G_2(s)\ldots G_n(s) \qquad \textbf{5.6.1}$$

where the overall transfer function is defined as the ratio of the output of the sequence divided by the input to the sequence. This property of transfer functions in series is particularly useful when analyzing the block diagrams of feedback control loops and other control related systems.

Example 5.8 Transfer Function of an Actuator, Process and Sensor System

Problem Statement. Determine the overall transfer function of an actuator, process and sensor (Figure 5.6.2). Assume that the actuator, process and sensor each exhibit first-order dynamics, i.e.,

$$G_a(s) = \frac{K_a}{\tau_a s + 1}$$

$$G_p(s) = \frac{K_p}{\tau_p s + 1}$$

$$G_s(s) = \frac{K_s}{\tau_s s + 1}$$

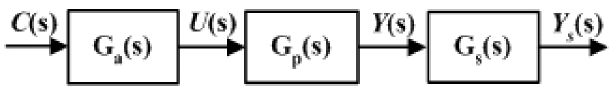
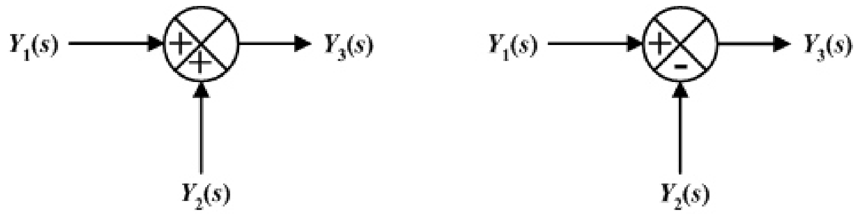


Figure 5.6.2 Schematic of a transfer function representation of an actuator/process/sensor system.

Solution. Applying Equation 5.6.1 yields

$$G_{oa}(s) = \frac{U(s)}{C(s)}\frac{Y(s)}{U(s)}\frac{Y_s(s)}{Y(s)} = \frac{Y_s(s)}{C(s)} = \frac{K_a}{\tau_a s + 1}\frac{K_p}{\tau_p s + 1}\frac{K_s}{\tau_s s + 1} = \frac{K_a K_p K_s}{(\tau_a s + 1)(\tau_p s + 1)(\tau_s s + 1)}$$

Note that this result shows that the combined system of the actuator/process/sensor behaves as a third-order process if the actuator, process and the sensor each have first-order dynamics (see Chapter 6).



(a) (b)

Figure 5.6.3 Examples of summation functions for block diagrams. (a) addition (b) subtraction.

Block Diagram Algebra. The properties of block diagrams and transfer functions can be used to develop general input/output relationships for process control systems. The input/output relationship for a series of transfer functions just developed can be combined with the properties of a summation function (Figure 5.6.3a and Figure 5.6.3b) and a divider (Figure 5.6.4). For the summation function in Figure 5.6.3a, the following relationship results

$$Y_1(s) \quad Y_2(s) \quad Y_3(s)$$

For the summation function in Figure 5.6.3b, the following relationship results

$$Y_1(s) \quad Y_2(s) \quad Y_3(s)$$

The following relationship holds for the divider shown in Figure 5.6.4.

$$Y_1(s) \quad Y_2(s) \quad Y_3(s)$$

Using these rules with the properties of a series of transfer functions, input/output relationships can be derived for a wide range of systems.

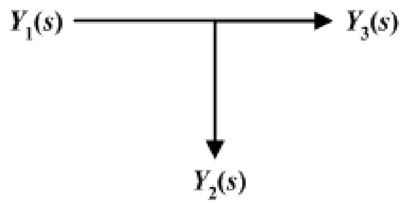


Figure 5.6.4 Schematic of a signal divider.

Example 5.9 Derivation of an Overall Transfer Function from a Block Diagram

Problem Statement. Develop a transfer function for the effect of $E(s)$ on $C(s)$ for the algorithm for a conventional PID controller shown schematically in Figure 5.6.5.

Solution. Using the properties of the summation function in Figure 5.6.5,

$$C(s) \quad A(s) \quad B(s) \quad D(s) \qquad \textbf{5.6.2}$$

Noting that the output of the first transfer function (K_c) goes to a divider that produces three separate signals, resulting in the following equations

$$A(s) \quad K_c\, E(s) \qquad \textbf{5.6.3}$$

$$B(s) = \frac{K_c}{\tau_I s} E(s) \qquad \textbf{5.6.4}$$

$$D(s) = K_c \tau_D s E(s) \qquad \textbf{5.6.5}$$

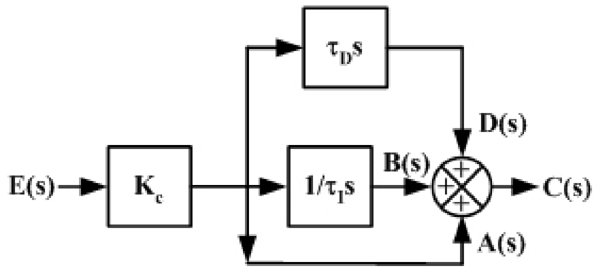


Figure 5.6.5 Block diagram of a conventional PID algorithm (Example 5.9).

Substituting Equations 5.6.3, 5.6.4 and 5.6.5 into Equation 5.6.2 results in

$$C(s) = K_c E(s) + \frac{K_c}{\tau_I s} E(s) + K_c \tau_D s E(s)$$

Factoring out $E(s)$ and solving for $C(s)/E(s)$ yields

$$G_c(s) = \frac{C(s)}{E(s)} = K_c\left(1 + \frac{1}{\tau_I s} + \tau_D s\right)$$

which is the transfer function for a PID controller. Note that this agrees with the results from Example 5.4.

Self-Assessment Question

Q5.6.1 Explain why the overall transfer function for a series of transfer functions is simply the product of each transfer function in the series.

Self-Assessment Answer

Q5.6.1 The overall transfer function for a series of individual transfer functions is simply the product of the individual transfer functions. This results because as each transfer function is added to the sequence, it cancels the previous output variable with its input variable. That is, because a transfer function is defined as the output over the input, when two transfer functions are in series, the output of the first transfer function is the input for the second. Therefore, by multiplying the two transfer functions, after the output of the first transfer function cancels the input of the second, you are left with the input for the first divided by the output of the second, which is the overall transfer function of the two transfer function system. As additional transfer functions are added to the sequence, their inputs cancel the outputs of the preceding transfer function, leaving their output as the output of the overall sequence.

5.7 Linearization of Nonlinear Differential Equations

The use of Laplace transforms to solve differential equations and to develop transfer functions is restricted to linear differential equations. But all real processes have some degree of nonlinearity. Changes in process gains and time constants represent examples of process nonlinearity. A number of processes are inherently nonlinear: (a) temperature behavior of a reactor with reaction rates that are exponential functions of temperature, (b) composition behavior for a distillation column due to limits on product purity (< 100%) and (c) neutralization of an acid with a base due to an S-shaped titration curve.

When the model equations of a process are nonlinear, the equations can be **linearized** about an operating point, thus converting the nonlinear equations into a linear form that allows the application of Laplace transforms. Consider a nonlinear dynamic equation (ODE),

$$\frac{dy}{dt} = f(y,u)$$

$$\text{at } t = 0,\ y = \bar{y},\ u = \bar{u} \text{ and } f(\bar{u},\bar{y}) = 0$$

where y is the output and u is the input variable. $\bar{y}$ and $\bar{u}$ represent an initial steady-state condition and are the basis for deviation variables. The linear approximation of $f(y,u)$ about $(\bar{u},\bar{y})$ can be obtained by applying a Taylor series expansion[3] to this function and truncating the terms that are second-order and higher, i.e.,

$$\frac{dy}{dt} = f(y,u) \approx f(y,u)\Big|_{\bar{y},\bar{u}} + (y - \bar{y})\left.\frac{\partial f(y,u)}{\partial y}\right|_{\bar{y},\bar{u}} + (u - \bar{u})\left.\frac{\partial f(y,u)}{\partial u}\right|_{\bar{y},\bar{u}} \qquad \textbf{5.7.1}$$

Converting to deviation variables (Section 5.2) results in

$$\frac{dy}{dt} = y\left.\frac{\partial f(y,u)}{\partial y}\right|_{\bar{y},\bar{u}} + u\left.\frac{\partial f(y,u)}{\partial u}\right|_{\bar{y},\bar{u}} \qquad \textbf{5.7.2}$$

This approximation is accurate in the vicinity of $(\bar{u},\bar{y})$. Taking the Laplace transform of each term in Equation 5.7.2 yields

$$sY(s) = \left.\frac{\partial f(y,u)}{\partial y}\right|_{\bar{y},\bar{u}} Y(s) + \left.\frac{\partial f(y,u)}{\partial u}\right|_{\bar{y},\bar{u}} U(s)$$

Rearranging yields the transfer function based on linearization of the process at the operating point $(\bar{u},\bar{y})$.

$$G(s) = \frac{Y(s)}{U(s)} = \frac{\left.\dfrac{\partial f(y,u)}{\partial u}\right|_{\bar{y},\bar{u}}}{s - \left.\dfrac{\partial f(y,u)}{\partial y}\right|_{\bar{y},\bar{u}}} \qquad \textbf{5.7.3}$$

Note that if more than one input is present, linearization terms of the form $(u_i - \bar{u}_i)\left.\dfrac{\partial f(y,u)}{\partial u_i}\right|_{\bar{y},\bar{u}}$ must be added to Equation 5.7.1 for each input, but this will not significantly change Equation 5.7.3 because **a transfer function is based on only one input/output pair**. Even though **Equation 5.7.3** was derived for use with nonlinear ODEs, it **can be used to determine transfer functions for both linear or nonlinear first-order ODEs**. Equation 5.7.3 is easier to use than the derivation of a transfer function from the model equations (e.g., Example 5.3) because the use of Equation 5.7.3 already has most of the derivation steps integrated into it (e.g., application of deviation variables and Laplace transforms).

Example 5.10 Linearization of a Quadratic Function

Problem Statement. Linearize the following function

$$f(t) \quad a\,t^2 \quad b\,t \quad c$$

about $\bar{t}$.

Solution. Applying Equation 5.7.1 yields

$$f(t) \quad f(\bar{t}) \quad (t \quad \bar{t})\frac{d\,f(t)}{dt}\Big|_{\bar{t}}$$

Evaluating the derivative of $f(t)$ and simplifying results in

$$f(t) \quad a\,\bar{t}^2 \quad b\bar{t} \quad c \quad (t \quad \bar{t})(2a\,\bar{t} \quad b) \quad a\bar{t}(2t \quad \bar{t}) \quad bt \quad c$$

Example 5.11 Transfer Function for the CST Thermal Mixer

Problem Statement. Use the process model equation for the CST thermal mixer (Equation 3.7.7) to determine the transfer function for the effect of changes of F_1 on T using (a) a complete derivation using linearization and deviation variables and (b) Equation 5.7.3.

Solution. (a) The model equation for the CST thermal mixer (Equation 3.7.7 in Example 3.3) is given by

$$M\frac{dT}{dt} \quad F_1 T_1 \quad F_2 T_2 \quad (F_1 \quad F_2)\,T \quad g(T, F_1) \qquad \mathbf{3.7.7}$$

Note that in Example 5.2, the feed streams to the mixer (F_1 and F_2) were assumed constant, yielding a linear ODE. In this case, F_1 is considered an input variable, resulting in nonlinear (**bilinear**) terms for the ODE. As a result, this equation must be linearized using Equation 5.7.1 before the transfer function can be developed. The linear approximation for Equation 3.6.7 about $\overline{F}_1$ and $\overline{T}$ is given by

$$M\frac{dT}{dt} \quad g(\overline{T}, \overline{F}_1) \quad (T \quad \overline{T})\frac{g(T,F_1)}{T}\bigg|_{\overline{T},\overline{F}_1} \quad (F_1 \quad \overline{F}_1)\frac{g(T,F_1)}{F_1}\bigg|_{\overline{T},\overline{F}_1}$$

where

$$g(\overline{T}, \overline{F}_1) \quad \overline{F}T_1 \quad F_2 T_2 \quad (\overline{F}_1 \quad F_2)\overline{T} \qquad \frac{g(T,F_1)}{T}\bigg|_{\overline{T},\overline{F}_1} \quad (\overline{F}_1 \quad F_2) \qquad \frac{g(T,F_1)}{F_1}\bigg|_{\overline{T},\overline{F}_1} \quad T_1 \quad \overline{T}$$

Then $$M\frac{dT}{dt} \quad \overline{F}_1 T_1 \quad F_2 T_2 \quad (\overline{F}_1 \quad F_2)\overline{T} \quad (T \quad \overline{T})(\overline{F}_1 \quad F_2) \quad (F_1 \quad \overline{F}_1)(T_1 \quad \overline{T})$$

Define the deviation variables as follows.

$$F_1 \quad F_1 \quad \overline{F}_1$$
$$T \quad T \quad \overline{T}$$

Note that in this case we are assuming that T and F_1 are changing; therefore, deviation variables for only these terms are required. Substituting into the linearized model equation yields

$$M\frac{d(T \quad \overline{T})}{dt} \quad \overline{F}_1 T_1 \quad F_2 T_2 \quad (\overline{F}_1 \quad F_2)\overline{T} \quad (\overline{F}_1 \quad F_2)T \quad (T_1 \quad \overline{T})F_1$$

This equation becomes

$$M\frac{dT}{dt} \quad (\overline{F}_1 \quad F_2)T \quad (T_1 \quad \overline{T})F_1 \qquad \textbf{5.7.4}$$

because the following equation holds

$$M\frac{d\overline{T}}{dt} \quad \overline{F}_1 T_1 \quad F_2 T_2 \quad (\overline{F}_1 \quad F_2)\,\overline{T} \quad 0$$

because of the initial steady-state conditions.

The application of Laplace transforms to the Equation 5.7.4 yields

$$M\,sT(s) \quad (\overline{F}_1 \quad F_2)T(s) \quad (T_1 \quad \overline{T})F_1(s)$$

Solving for $T(s)/F_1(s)$ and rearranging this equation results in the desired form for the transfer function.

$$G_p(s) \quad \frac{T(s)}{F_1(s)} \quad \frac{\dfrac{T_1 \quad \overline{T}}{M}}{s \quad \dfrac{\overline{F}_1 \quad F_2}{M}}$$

(b) Now consider the application of Equation 5.7.3 to this problem. First, Equation 3.7.7 must be rearranged into the standard form for an ODE [i.e., $\frac{dy}{dt} \quad f(y,u)$]

$$\frac{dT}{dt} \quad \frac{1}{M}[F_1 T_1 \quad F_2 T_2 \quad (F_1 \quad F_2)\,T] \quad f(T,F_1)$$

Next, the two partial derivatives are determined.

$$\left.\frac{f(T,F_1)}{F_1}\right|_{\bar{T},\bar{F}_1} \quad \frac{T_1 \quad \bar{T}}{M} \qquad\qquad \left.\frac{f(T,F_1)}{T}\right|_{\bar{T},\bar{F}_1} \quad \frac{\bar{F}_1 \quad F_2}{M}$$

Finally, substituting into Equation 5.7.3 yields

$$G_p(s) \quad \frac{T(s)}{F_1(s)} \quad \frac{\dfrac{T_1 \quad \bar{T}}{M}}{s \quad \dfrac{\bar{F}_1 \quad F_2}{M}}$$

Example 5.12 Transfer Function for the CSTR Model Equation with Exponential Temperature Dependence

Problem Statement. Develop a transfer function for the effect of Q on T for Equation 3.7.11 evaluated at $T \quad \bar{T}$. Assume that C_A is constant.

$$V_r \quad C_v \frac{dT}{dt} \quad F C_p (T_0 \quad T) \; V_r \quad H C_A k_0 e^{\;E/RT} \quad Q \qquad \textbf{3.7.11}$$

Solution. To apply Equation 5.7.3, $\frac{f}{T}$ and $\frac{f}{Q}$ must be calculated and evaluated at $T \quad \bar{T}$ and $Q \quad \bar{Q}$.

$$\left.\frac{f}{T}\right|_{\bar{T},\bar{Q}} \quad \frac{F C_p \quad \dfrac{V_r \quad H C_A k_0 E}{R\bar{T}^2} \exp \quad \dfrac{E}{R\bar{T}}}{V_r \quad C_v}$$

$$\left.\frac{f}{Q}\right|_{\bar{T},\bar{Q}} \quad \frac{1}{V_r \quad C_v}$$

Applying Equation 5.7.3 yields

$$G(s) \quad \frac{T(s)}{Q(s)} \quad \frac{\dfrac{1}{V_r \quad C_v}}{s \quad \dfrac{F C_p \quad \dfrac{V_r \quad H C_A k_0 E}{R\bar{T}^2} \exp \quad \dfrac{E}{R\bar{T}}}{V_r \quad C_v}}$$

This approximation shows that the pole of this transfer function is a strong function of $\overline{T}$, which is the temperature about which the nonlinear equation is linearized.

Example 5.13 Analysis of the Linearization of a Level in a Tank

Problem Statement. Consider the level in the tank shown in Figure 5.7.1. The tank is 10 ft in height and 6 ft in diameter. The discharge flow from the tank is given by

$$F_{out} \quad C_v \sqrt{h}$$

where C_v is 447 lb/h-ft^{½}. Initially, the feed rate of water to the tank (F_{in}) is 1000 lb/h, which corresponds to a steady-state liquid level of 5 ft. Develop a nonlinear dynamic model for this process and linearize the nonlinear model about the initial conditions. Compare the steady-state change for the linearized model with the nonlinear model for a step increase in F_{in} to 1450 lb/h.

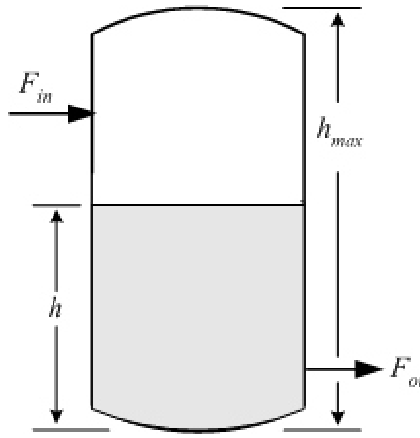


Figure 5.7.1 Schematic of the tank in Example 5.13.

Solution. Applying an unsteady-state macroscopic mass balance to this process results in the following nonlinear equation

$$A \frac{dh}{dt} \quad F_{in} \quad C_v \sqrt{h} \qquad \textbf{5.7.5}$$
$$t \quad 0 \qquad h \quad \bar{h} \quad 5\text{ft}$$

where A is the cross-sectional area of the tank, is the density of the water in the tank and h is the height of water in the tank. The analytical solution of Equation 5.7.5 can be developed by using a variable transformation and separation of variables followed by integration, but is not presented here due to its complexity.

Equation 5.7.5 can be linearized by linearizing the square root of h, which is the only nonlinear term in the equation. That is,

$$\sqrt{h} \quad \sqrt{\bar{h}} \quad \frac{1}{2\sqrt{\bar{h}}} \quad h \quad \bar{h}$$

where $\bar{h}$ is the operating condition about which the equation was linearized (i.e., a liquid level of 5 ft). Then, the linearized equation for this process is given by

$$A \frac{dh}{dt} \quad F_{in} \quad C_v \quad \sqrt{\bar{h}} \quad \frac{1}{2\sqrt{\bar{h}}}(h \quad \bar{h}) \qquad \textbf{5.7.6}$$

The steady-state liquid level in the tank after the change in the inlet flow can be determined for the linear and nonlinear models of this system by setting dh/dt equal to zero for the differential equations (i.e., Equation 5.7.5 and 5.7.6) and solving for the liquid level, h, in each case. The linear representation of the process with F_{in} equal to 1450 lb/h and $\bar{h}$ equal to 5 ft yields a steady-state liquid level in the tank of 9.51 ft. On the other hand, if these values are applied to the nonlinear equation, the steady-state liquid level in the tank corresponding to a feed rate 1450 lb/h is 10.52 ft. Therefore, the linearized model predicts that the liquid level was still within the 10 ft height of the tank while the nonlinear model indicates that the tank would overflow for this feed rate. In this case a 45% increase in inlet feed rate resulted in a one foot error in the steady-state level for the linear model, which, in this case, would cause the actual process to overflow while the linear model would predict only a high level. Clearly, the larger the deviation from the initial conditions, the larger the error between the nonlinear model and the linear approximation. Figure 5.7.2 shows the dynamic behavior of the nonlinear and linearized model which is based on the numerical integration of Equations 5.7.5 and 5.7.6, respectively, for a step change in the inlet flow rate from 1000 to 1450 lb/h at time equal to 20 s. In addition, this figure also indicates the maximum tank level, h_{max} (10 ft).

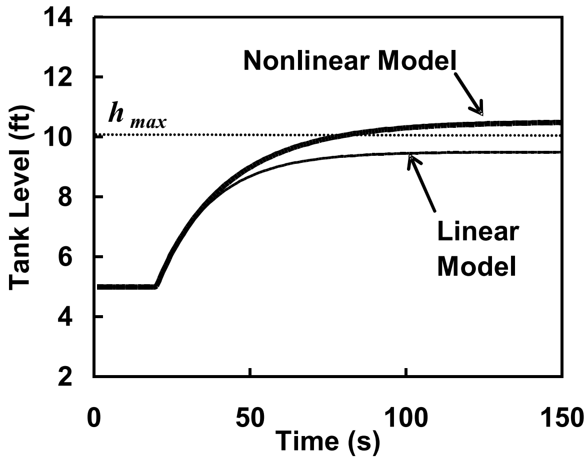


Figure 5.7.2 Comparison between the dynamic response of a nonlinear and a linear model of the level in a tank (Example 5.13). Note that h_{max} is the height of liquid in the tank when the tank is full.

Example 5.14 Overall Transfer Function for the CST Composition Mixer

Problem Statement. For the composition mixing tank (Example 3.4) shown in Figure 5.7.3, develop the transfer function for the effect of the specified flow rate for F_1 on the measured value of the product composition, $C_s(t)$. The process model and data for this process are given below:

Actuator $$\frac{dF_1}{dt} \quad \frac{1}{_v}(F_{1,spec} \quad F_1)$$

Process $$V\frac{dC}{dt} \quad F_1 C_1 \quad F_2 C_2 \quad (F_1 \quad F_2)C$$

Sensor $$C_s(t) \quad C(t \quad _A)$$

- C - the concentration of the component in the mixed stream (initially 0.75 g mol/l)
- C_1 - the concentration of the component in stream 1 (0.5 g mol/l)
- C_2 - the concentration of the component in stream 2 (1.0 g mol/l)
- F_1 - the mass flow rate of stream 1 (initially 500 kg/min)
- F_2 - the mass flow rate of stream 2 (500 kg/min)
- t - time (min)
- V - the volume of the mixer (1000 l)
- - the constant density of the feed and product streams (1 kg/l)
- $_v$ - the time constant for the flow controller on stream 1 (2 s).

- $_A$ - analyzer deadtime (300 s).

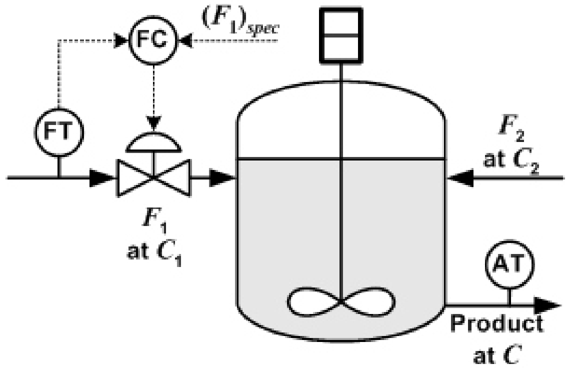


Figure 5.7.3 Schematic of a CST composition mixing process.

Solution. The overall transfer function for the effect of the input to the actuator on the measured value of the production composition is equal to the products of the transfer functions for the actuator, process and sensor, i.e.,

$$G_{oa}(s) \quad \frac{C_s(s)}{F_{1,spec}(s)} \quad G_a(s)G_p(s)G_s(s)$$

Applying Equation 5.7.3 to the equation for the actuator yields

$$G_a(s) \quad \frac{1}{{}_s s \quad 1} \quad \frac{1}{2s \quad 1}$$

Applying Equation 5.7.3 to the model for the process yields

$$G_p(s) \quad \frac{\frac{C_1 \quad \overline{C}}{V}}{s \quad \frac{\overline{F}_1 \quad F_2}{V}} \quad \frac{\frac{C_1 \quad \overline{C}}{\overline{F}_1 \quad F_2}}{\frac{V}{\overline{F}_1 \quad F_2} s \quad 1} \quad \frac{2.5 \quad 10^{\;\;4}}{4000s \quad 1}$$

Finally, because the sensor behaves as a pure deadtime, the transfer function for the sensor is

$$G_s(s) \quad e^{\;\;{}_s s} \quad e^{\;\;300s}$$

Therefore, $$G_{oa}(s) \quad \frac{C_s(s)}{F_{1,spec}(s)} \quad G_a(s)G_p(s)G_s(s) \quad \frac{1}{2s \quad 1} \frac{2.5 \quad 10^{\;\;4}}{4000s \quad 1} \quad e^{\;\;300s}$$

Self-Assessment Questions

Q5.7.1 How can a transfer function be developed for a process model that is nonlinear?

Q5.7.2 When a nonlinear process model consisting of a single first-order differential equation is used to develop a transfer function, what general form results.

Q5.7.3 Why is it more convenient to use Equation 5.7.3 to develop a transfer function from a model equation than to derive it using deviation variables and Laplace transforms?

Self-Assessment Answers

Q5.7.1 A transfer function for a nonlinear process model can be developed by linearizing the nonlinear equation about the initial conditions using a truncated Taylor series expansion. Then, the equation is converted to deviation variable form. Next, Laplace transforms are applied to each term in the deviation variable form of the model equation and the ratio of the output to the input variable yields the transfer function.

Alternatively, you can apply Equation 5.7.3.

Q5.7.2 When a nonlinear process model is used to develop a transfer function, a first-order transfer function (Equation 5.7.3) results.

Q5.7.3 Because you can use a single equation (Eq. 5.7.3) instead of a multi-step derivation process.

Self-Assessment Problem

P5.7.1 Develop the transfer function for the following nonlinear differential equation [Hint: First, determine $u(0)$]:

$$\frac{dy}{dt} \quad y \quad Au^{½} \qquad y(0) \quad \bar{y}; \quad \frac{dy(0)}{dt} \quad 0$$

Self-Assessment Answers

P5.7.1 Based on the steady-state initial condition, $\bar{u} \quad (A/\bar{y})^2$. The applying Equation 5.7.3,

$$\frac{(y \quad Au^{½})}{y}\bigg|_{\bar{y},\bar{u}} \quad 1 \qquad \frac{(y \quad Au^{½})}{u}\bigg|_{\bar{y},\bar{u}} \quad ½Au^{\quad ½}\Big|_{\bar{y},\bar{u}} \quad \frac{A^2}{2\bar{y}}; \qquad \text{Then} \quad G(s) \quad \frac{Y(s)}{U(s)} \quad \frac{\frac{A^2}{2\bar{y}}}{s \quad 1}$$

5.8 State Space Models^Advanced Topic (AT)^

In the last section, it was shown that it is easier to use Equation 5.7.3 to develop a transfer function for a single nonlinear model equation than to derive the transfer function using local linearization, deviation variables and Laplace transforms. This results because Equation 5.7.3 was derived using local linearization, deviation variables and Laplace transforms for a general nonlinear model equation (first-order ODE). Therefore, the details of the derivation of a transfer function have already been implemented in Equation 5.7.3, resulting in a convenient form that can be used to determine a transfer function from a linear or nonlinear first-order ODE. In this section, the previous approach is extended to a system of nonlinear first-order ODEs to develop what are known as **state space models**. To demonstrate state space models, the process model equations for the endothermic CSTR from Example 3.6 (Equations 3.7.10 and 3.7.11) are considered here. Equations 3.7.10 and 3.7.11 describe the temperature and composition changes as a function of time, but note that both equations are nonlinear.

$$V_r \frac{dC_A}{dt} \qquad \overline{F(C_{A_0} \quad C_A)} \quad V_r \, k_0 \, C_A \, e^{\quad E/RT} \tag{3.7.10}$$

$$V_r \quad C_p \frac{dT}{dt} \quad F C_p (T_0 \quad T) \quad V_r \quad H_{rxn} C_A \, k_0 \, e^{\quad E/RT} \quad Q \tag{3.7.11}$$

In this example, V_r, C_{A_0}, , k_0, E, R, C_p and H_{rxn} are assumed constant, F and Q are the independent variables, and C_A and T are the dependent variables. The nonlinearity of Equation 3.7.10 comes from the exponential term ($e^{-E/RT}$), the product of C_A and the exponential term and the product of F and C_A. For Equation 3.7.11, the nonlinearity comes from the exponential term, the product of C_A and the exponential term and the product of F and T. These equation represent a 2-input/2-output system. Equations 3.7.10 and 3.7.11 can be rewritten in the following general form for a set of first-order ODEs:

$$\frac{dC_A}{dt} \quad f_1(C_A,T,F,Q) \quad \frac{F(C_{A_0} \quad C_A)}{V_r} \quad k_0 C_A e^{\quad E/RT}$$
$$\frac{dT}{dt} \quad f_2(C_A,T,F,Q) \quad \frac{F(T_0 \quad T)}{V_r} \quad \frac{H_{rxn} C_A k_0 e^{\quad E/RT}}{C_p} \quad \frac{Q}{V_r \quad C_p} \tag{5.8.1}$$

Functions f_1 and f_2 can be linearized about steady-state conditions, $\bar{C}_A$, $\bar{T}$, $\bar{F}$ and $\bar{Q}$, i.e.,

$$f_1(C_A,T,F,Q) \quad f_1(\bar{C}_A,\bar{T},\bar{F},\bar{Q}) \quad \frac{f_1}{C_A}\Bigg|_{\bar{C}_A,\bar{T},\bar{F},\bar{Q}}(C_A \quad \bar{C}_A) \quad \frac{f_1}{T}\Bigg|_{\bar{C}_A,\bar{T},\bar{F},\bar{Q}}(T \quad \bar{T})$$
$$\frac{f_1}{F}\Bigg|_{\bar{C}_A,\bar{T},\bar{F},\bar{Q}}(F \quad \bar{F}) \quad \frac{f_1}{Q}\Bigg|_{\bar{C}_A,\bar{T},\bar{F},\bar{Q}}(Q \quad \bar{Q}) \tag{5.8.2}$$

$$f_2(C_A,T,F,Q) \quad f_2(\bar{C}_A,\bar{T},\bar{F},\bar{Q}) \quad \frac{f_2}{C_A}\Bigg|_{\bar{C}_A,\bar{T},\bar{F},\bar{Q}}(C_A \quad \bar{C}_A) \quad \frac{f_2}{T}\Bigg|_{\bar{C}_A,\bar{T},\bar{F},\bar{Q}}(T \quad \bar{T})$$
$$\frac{f_2}{F}\Bigg|_{\bar{C}_A,\bar{T},\bar{F},\bar{Q}}(F \quad \bar{F}) \quad \frac{f_2}{Q}\Bigg|_{\bar{C}_A,\bar{T},\bar{F},\bar{Q}}(Q \quad \bar{Q}) \tag{5.8.3}$$

Using the definitions of deviation variables,

$$\begin{array}{ccc} C_A & C_A & \bar{C}_A \\ T & T & \bar{T} \\ F & F & \bar{F} \\ Q & Q & \bar{Q} \end{array} \tag{5.8.4}$$

Because the values of $\bar{C}_A$ and $\bar{T}$ are constant,

$$\frac{dC_A}{dt} \quad \frac{dC_A}{dt} \qquad \text{and} \qquad \frac{dT}{dt} \quad \frac{dT}{dt} \tag{5.8.5}$$

Combining Equations 5.8.2-5.8.5, the nonlinear differential equations for C_A and T can be written as linear approximations:

$$\frac{dC_A}{dt} \quad f_1(\bar{C}_A,\bar{T},\bar{F},\bar{Q}) \quad \frac{f_1}{C_A}\Bigg|_{\bar{C}_A,\bar{T},\bar{F},\bar{Q}} C_A \quad \frac{f_1}{T}\Bigg|_{\bar{C}_A,\bar{T},\bar{F},\bar{Q}} T \quad \frac{f_1}{F}\Bigg|_{\bar{C}_A,\bar{T},\bar{F},\bar{Q}} F \quad \frac{f_1}{Q}\Bigg|_{\bar{C}_A,\bar{T},\bar{F},\bar{Q}} Q$$

$$\frac{dT}{dt} \quad f_2(\bar{C}_A,\bar{T},\bar{F},\bar{Q}) \quad \frac{f_2}{C_A}\Bigg|_{\bar{C}_A,\bar{T},\bar{F},\bar{Q}} C_A \quad \frac{f_2}{T}\Bigg|_{\bar{C}_A,\bar{T},\bar{F},\bar{Q}} T \quad \frac{f_2}{F}\Bigg|_{\bar{C}_A,\bar{T},\bar{F},\bar{Q}} F \quad \frac{f_2}{Q}\Bigg|_{\bar{C}_A,\bar{T},\bar{F},\bar{Q}} Q$$

Assuming that the system is at steady-state $[f_1(\overline{C}_A, \overline{T}, \overline{F}, \overline{Q}) \quad f_2(\overline{C}_A, \overline{T}, \overline{F}, \overline{Q}) \quad 0]$ at t=0,

$$\frac{dC_A}{dt} \quad \left.\frac{f_1}{C_A}\right|_{\overline{C}_A,\overline{T},\overline{F},\overline{Q}} C_A \quad \left.\frac{f_1}{T}\right|_{\overline{C}_A,\overline{T},\overline{F},\overline{Q}} T \quad \left.\frac{f_1}{F}\right|_{\overline{C}_A,\overline{T},\overline{F},\overline{Q}} F \quad \left.\frac{f_1}{Q}\right|_{\overline{C}_A,\overline{T},\overline{F},\overline{Q}} Q$$

$$\frac{dT}{dt} \quad \left.\frac{f_2}{C_A}\right|_{\overline{C}_A,\overline{T},\overline{F},\overline{Q}} C_A \quad \left.\frac{f_2}{T}\right|_{\overline{C}_A,\overline{T},\overline{F},\overline{Q}} T \quad \left.\frac{f_2}{F}\right|_{\overline{C}_A,\overline{T},\overline{F},\overline{Q}} F \quad \left.\frac{f_2}{Q}\right|_{\overline{C}_A,\overline{T},\overline{F},\overline{Q}} Q$$

The above two equations are linearized forms of the original nonlinear equations for this CSTR system (Equation 5.8.1). The state space models for the CSTR can be expressed in a more general form using the following definitions:

State Variables: $x_1 \quad C_A \quad x_2 \quad T$

as a result $\frac{dx_1}{dt} \quad \frac{dC_A}{dt} \quad \frac{dx_2}{dt} \quad \frac{dT}{dt}$

Input Variables: $u_1 \quad F \quad u_2 \quad Q$

Output Variables: $y_1 \quad C_A \quad y_2 \quad T$

The state variables of a process are the minimum independent set of variables necessary to define the state of the process at a point in time. Knowing the state of a process will allow you to determine all the other properties of the system. The state variables for the CSTR case are C_A and T. An output variable is a process variable that is measured. An output variable is measured and can be a state variable or a linear combination of all or some of the state variables. For example, for a distillation column the liquid compositions for each tray are state variables, but usually the product compositions are the output variables because they are typically measured.

In addition, lets define a_{ij} and b_i as

$$a_{ij} \quad \left.\frac{f_i}{x_j}\right|_{\underline{\overline{x}},\underline{\overline{u}}} \qquad\qquad b_{ij} \quad \left.\frac{f_i}{u_j}\right|_{\underline{\overline{x}},\underline{\overline{u}}}$$

where $\underline{\overline{x}}$ is a vector that contains the initial conditions of all the state variables [i.e., $\underline{\overline{x}} \quad (\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_n)$] and $\underline{\overline{u}}$ is a vector that contains the initial conditions for the input variables [i.e., $\underline{\overline{u}} \quad (\overline{u}_1, \overline{u}_2, \ldots, \overline{u}_m)$]. Therefore, for the CSTR case

$$a_{11} \quad \left.\frac{f_1}{x_1}\right|_{\underline{\overline{x}},\underline{\overline{u}}} \quad \left.\frac{f_1}{C_A}\right|_{\overline{C}_A,\overline{T},\overline{F},\overline{Q}} \qquad a_{12} \quad \left.\frac{f_1}{x_2}\right|_{\underline{\overline{x}},\underline{\overline{u}}} \quad \left.\frac{f_1}{T}\right|_{\overline{C}_A,\overline{T},\overline{F},\overline{Q}}$$

$$a_{21} \quad \left.\frac{f_2}{x_1}\right|_{\underline{\overline{x}},\underline{\overline{u}}} \quad \left.\frac{f_2}{C_A}\right|_{\overline{C}_A,\overline{T},\overline{F},\overline{Q}} \qquad a_{22} \quad \left.\frac{f_2}{x_2}\right|_{\underline{\overline{x}},\underline{\overline{u}}} \quad \left.\frac{f_2}{T}\right|_{\overline{C}_A,\overline{T},\overline{F},\overline{Q}}$$

$$b_{11} \quad \frac{f_1}{u_1}\bigg|_{\underline{x},\underline{u}} \quad \frac{f_1}{F}\bigg|_{\bar{C}_A,\bar{T},\bar{F},\bar{Q}} \qquad\qquad b_{12} \quad \frac{f_1}{u_2}\bigg|_{\underline{x},\underline{u}} \quad \frac{f_1}{Q}\bigg|_{\bar{C}_A,\bar{T},\bar{F},\bar{Q}}$$

$$b_{21} \quad \frac{f_2}{u_1}\bigg|_{\underline{x},\underline{u}} \quad \frac{f_2}{F}\bigg|_{\bar{C}_A,\bar{T},\bar{F},\bar{Q}} \qquad\qquad b_{22} \quad \frac{f_2}{u_2}\bigg|_{\underline{x},\underline{u}} \quad \frac{f_2}{Q}\bigg|_{\bar{C}_A,\bar{T},\bar{F},\bar{Q}}$$

Based on these definitions for the a_{ij}s and b_{ij}s and the definitions for the state variables, the general form of the state space equations for the CSTR example are given as

$$\frac{dx_1}{dt} \quad a_{11}x_1 \quad a_{12}x_2 \quad b_{11}u_1 \quad b_{12}u_2$$

$$\frac{dx_2}{dt} \quad a_{21}x_1 \quad a_{22}x_2 \quad b_{21}u_1 \quad b_{22}u_2$$

$$y_1 \quad x_1$$
$$y_2 \quad x_2$$

The above equation can be written in the matrix vector form:

$$\begin{matrix} \dfrac{dx_1}{dt} \\ \dfrac{dx_2}{dt} \end{matrix} \quad \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{matrix} \quad \begin{matrix} x_1 \\ x_2 \end{matrix} \quad \begin{matrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{matrix} \quad \begin{matrix} u_1 \\ u_2 \end{matrix}$$

$$\begin{matrix} y_1 \\ y_2 \end{matrix} \quad \begin{matrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{matrix} \quad \begin{matrix} x_1 \\ x_2 \end{matrix}$$

Note for the CSTR case, y_1=x_1 and y_2 = x_2 so c_{11}=1, c_{22}=1, c_{12}=0 and c_{21}=0.

These equations, which are based on the following form of the nonlinear model equations

$$\frac{dx_i}{dt} \quad f(\underline{x},\underline{u}) \qquad\qquad (i \quad 1,2,\ldots,n)$$

can be expressed as matrix equations:

$$\frac{\mathbf{dx}}{\mathbf{dt}} \quad \underline{\underline{A}}\underline{\boldsymbol{x}} + \underline{\underline{B}}\underline{\boldsymbol{u}} \qquad \textbf{5.8.6}$$
$$\underline{\boldsymbol{y}} \quad \underline{\underline{C}}\underline{\boldsymbol{x}}$$

where for a case similar to the CSTR case with two inputs, two state variables and two outputs:

$$\underline{x}\quad \begin{matrix} x_1 \\ x_2 \end{matrix} \qquad \underline{u}\quad \begin{matrix} u_1 \\ u_2 \end{matrix} \qquad \underline{y}\quad \begin{matrix} y_1 \\ y_2 \end{matrix}$$

$$\underline{\underline{A}}\quad \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{matrix} \qquad \underline{\underline{B}}\quad \begin{matrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{matrix} \qquad \underline{\underline{C}}\quad \begin{matrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{matrix}$$

Remember that
$$a_{ij} \quad \left.\frac{f_i}{x_j}\right|_{\underline{\bar{x}},\underline{\bar{u}}} \qquad b_{ij} \quad \left.\frac{f_i}{u_j}\right|_{\underline{\bar{x}},\underline{\bar{u}}}$$

and the c_{ij}s relate the state variables to the output variables. $\underline{\underline{A}}$, which is a matrix formed by all the possible combinations of partial derivatives of the nonlinear functions, $f_i(\underline{x})$, with respect to state variables, x_i, is known as the **Jacobian matrix** of $f_i(\underline{x})$. $\underline{\underline{A}}$ is an $n \times n$ matrix, where n is the number of state variables. $\underline{\underline{B}}$ is an $n \times m$ matrix, where m is the number of input variables and $\underline{\underline{C}}$ is an $p \times n$ matrix, where p is the number of output variables. **State space models are based on (1) a local linearization of the nonlinear dynamic model equations (first-order ODEs), (2) steady-state initial conditions and (3) deviation variables.** In certain cases, output variables ($\underline{y}$) can also be functions of one or more of the input variables ($\underline{u}$).

State space models in the CPI and biotechnology industries are usually associated with unsteady-state mass, energy and momentum balances. For the CSTR case just considered, the state space models were developed from a component mole balance equation and an energy balance equation.

Example 5.15 State Space Model for a Continuous Bio-Reactor

Problem Statement. Develop a state space model for the model portion of the continuous bio-reactor[4] described below based on the initial conditions provided (Figure 5.8.1).

The process variables and process parameters for this examples are

- $F_{V,spec}$ - the specified feed rate to the bioreactor (1050 l/hr at t=13 h)
- F_V - feed rate to the reactor (initially 1000 l/hr)
- K_s - Monod's saturation constant (0.1 g/l)
- P - product concentration in the reactor (initially 1.25 g/l)
- S - substrate concentration in the reactor (initially 25 g/l)
- S_F - substrate concentration in the feed to the reactor (50 g/l)
- t - time (h)
- V - volume of the reactor (5000 l)
- x - cell concentration in the bioreactor (initially 0.25 g/l)

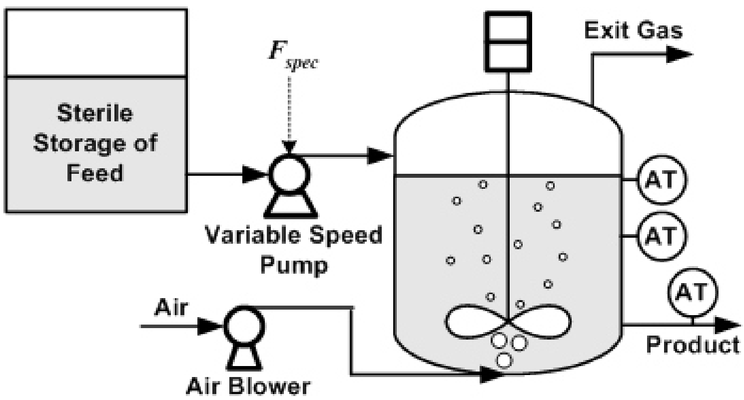


Figure 5.8.1 Schematic of a continuous bio-reactor.

- Y_{xP} - yield factor (0.2 g-cells/g-product)
- Y_{xS} - yield coefficient (0.01 g-cells/g-substrate)
- ${}_{max}$ - maximum specific growth rate (0.2 hr^{-1})
- ${}_{s}$ - the sensor deadtime for a HPLC analyzer (30 min)
- ${}_{TM}$ - the time constant for the turbidity meter used to measure the cell concentration (20 s)

The resulting model equations for this process are

Process

$$\frac{dx}{dt} \quad \frac{F}{V}x \quad {}_{max}x \tag{5.8.7}$$

$$\frac{dS}{dt} \quad \frac{F}{V}S_F \quad \frac{F}{V}S \quad \frac{1}{Y_{xS}} \quad {}_{max}x \tag{5.8.8}$$

$$\frac{dP}{dt} \quad \frac{F}{V}P \quad \frac{1}{Y_{xP}} \quad {}_{max}x \tag{5.8.9}$$

where F is the only independent variable for this process. In this case, x, S and P are both state variables and output variables. For this case, n=3, m=1 and p=3.

Solution. Using the definitions of the elements of the matrices in a state space model given by Equation 5.8.6, the $\underline{\underline{A}}$ matrix is given by

$$\underline{\underline{A}} \quad \begin{matrix} \frac{\bar{F}}{V} \quad {}_{max} & 0 & 0 \\ \frac{{}_{max}}{Y_{xS}} & \frac{\bar{F}}{V} & 0 \\ \frac{{}_{max}}{Y_{xP}} & 0 & \frac{\bar{F}}{V} \end{matrix}$$

which is based on defining x_1 as x, x_2 as S, and x_3 as P while u_1 is F. For example, a_{22} is given by

$$a_{22} \quad \frac{}{S} \quad \frac{F}{V}S_F \quad \frac{F}{V}S \quad \frac{1}{Y_{xS}} \quad {}_{max}x \quad {}_{\bar{F},\bar{x},\bar{S},\bar{P}} \quad \frac{\bar{F}}{V}$$

Likewise,

$$a_{23} \quad \frac{}{P} \quad \frac{F}{V}S_F \quad \frac{F}{V}S \quad \frac{1}{Y_{xS}} \quad {}_{max}x \quad {}_{\bar{F},\bar{x},\bar{S},\bar{P}} \quad 0$$

Because F is the only input variable, $\underline{\underline{B}}$ is actually a vector in this case, i.e.,

$$\underline{B} \quad \begin{array}{c} \frac{\bar{x}}{V} \\ \frac{S_F}{V} \quad \frac{\bar{S}}{V} \\ \frac{\bar{P}}{V} \end{array}$$

For example,

$$b_2 \quad \frac{}{F} \quad \frac{F}{V} S_F \quad \frac{F}{V} S \quad \frac{1}{Y_{xS}} \quad {}_{\max} x \quad {}_{\bar{F}, \bar{x}, \bar{S}, \bar{P}} \quad \frac{S_F}{V} \quad \frac{\bar{S}}{V}$$

The $\underline{\underline{C}}$ matrix relates the state variables (x, S and P) to the output variables. Therefore,

$$\underline{\underline{C}} \quad \begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}$$

Note that $\underline{\underline{C}}$ is an **identity matrix** because the outputs are the state variables themselves in this case. Therefore, the state space model for this case is

$$\begin{array}{c} \frac{dx}{dt} \\ \frac{dS}{dt} \\ \frac{dP}{dt} \end{array} \quad \begin{array}{ccc} \frac{\bar{F}}{V} \quad {}_{\max} & 0 & 0 \\ \frac{{}_{\max}}{Y_{xS}} & \frac{\bar{F}}{V} & 0 \\ \frac{{}_{\max}}{Y_{xP}} & 0 & \frac{\bar{F}}{V} \end{array} \quad \begin{array}{c} x \\ S \\ P \end{array} \quad \begin{array}{c} \frac{\bar{x}}{V} \\ \frac{S_F}{V} \quad \frac{\bar{S}}{V} \\ \frac{\bar{P}}{V} \end{array} \quad F$$

Using the numerical values of these variables yields

$$\begin{array}{c} \frac{dx}{dt} \\ \frac{dS}{dt} \\ \frac{dP}{dt} \end{array} \quad \begin{array}{ccc} 0 & 0 & 0 \\ 10.53 & 0.2 & 0 \\ 0.04 & 0 & 0.2 \end{array} \quad \begin{array}{c} x \\ S \\ P \end{array} \quad \begin{array}{c} 0.00005 \\ 0.005 \\ 0.00025 \end{array} \quad F$$

Self-Assessment Questions

Q5.8.1 What is the difference between a state variable and an output variable? How are they connected?

Q5.8.2 What advantages does Equation 5.8.6 have over deriving the transfer functions directly from a system of nonlinear equations?

Self-Assessment Questions

Q5.8.1 The state variables define the operation of a process while output variables are what is measured on a process. State space models assume that output variables are a linear function of certain state variables.

Q5.8.2 The state space equations (Equation 5.8.6) are derived for the general case and already has the derivation steps integrated into it. To apply Equation 5.8.6, you only have to generate the Jacobian matrix, the partial of the nonlinear equations with respect to the inputs and the linear correspondence between the outputs and the state variables. Therefore, Equation 5.8.6 is much easier to apply than to derive the state space equations directly from the nonlinear model equations.

5.9 Transfer Functions from State Space Models[AT]

Laplace transforms can be applied to the generalized state space model equations (Equation 5.8.6) to determine the transfer functions for the system represented by the state space model. Consider the application of Laplace transforms to Equation 5.8.6 for a SISO case.

$$
\begin{aligned}
sX(s) - x'(0) &= aX(s) + bU(s) \\
Y(s) &= cX(s)
\end{aligned}
$$

where a, b and c are constants. Because x' is a deviation variable, $x'(0)$ is equal to zero. Solving for $X(s)$, substituting into the equation for $Y(s)$ and rearranging yields

$$
\frac{Y(s)}{U(s)} = \frac{c\,b}{s - a}
$$

which indicates that a is the pole of the system.

Now consider a MIMO system. Applying the Laplace transform of Equation 5.8.6 yields

$$
\begin{aligned}
s\underline{\underline{\boldsymbol{I}}}\,\underline{\boldsymbol{X}}(s) - \underline{\boldsymbol{x}}'(0) &= \underline{\underline{\boldsymbol{A}}}\,\underline{\boldsymbol{X}}(\mathrm{s}) + \underline{\underline{\boldsymbol{B}}}\,\underline{\boldsymbol{U}}(\mathrm{s}) \\
\underline{\boldsymbol{Y}}(\mathrm{s}) &= \underline{\underline{\boldsymbol{C}}}\,\underline{\boldsymbol{X}}(\mathrm{s})
\end{aligned}
$$

where $\underline{\underline{\boldsymbol{I}}}$ is an identity matrix. Remembering that the initial conditions are zero and collecting terms yields

$$
(s\underline{\underline{\boldsymbol{I}}} - \underline{\underline{\boldsymbol{A}}})\underline{\boldsymbol{X}}(s) = \underline{\underline{\boldsymbol{B}}}\,\underline{\boldsymbol{U}}(s)
$$

Pre-multiplying both sides of this equation by the matrix inverse of $(s\underline{\underline{\boldsymbol{I}}} - \underline{\underline{\boldsymbol{A}}})$ yields

$$
\underline{\boldsymbol{X}}(s) = (s\underline{\underline{\boldsymbol{I}}} - \underline{\underline{\boldsymbol{A}}})^{-1}\underline{\underline{\boldsymbol{B}}}\,\underline{\boldsymbol{U}}(s)
$$

Multiplying both sides of this equation by $\underline{\underline{\boldsymbol{C}}}$ yields

$$
\underline{\underline{\boldsymbol{C}}}\,\underline{\boldsymbol{X}}(s) = \underline{\boldsymbol{Y}}(s) = \underline{\underline{\boldsymbol{C}}}(s\underline{\underline{\boldsymbol{I}}} - \underline{\underline{\boldsymbol{A}}})^{-1}\underline{\underline{\boldsymbol{B}}}\,\underline{\boldsymbol{U}}(s)
$$

Therefore, the transfer functions for the MIMO case is a matrix that contains the individual transfer functions for each input/output pair and is given by

$$\underline{\underline{\boldsymbol{G}}}_p(s) \quad \underline{\underline{\boldsymbol{C}}}(s\underline{\boldsymbol{I}} \quad \underline{\underline{\boldsymbol{A}}})^{\ 1}\,\underline{\underline{\boldsymbol{B}}} \tag{5.9.1}$$

Note that because $\underline{\underline{\boldsymbol{B}}}$ and $\underline{\underline{\boldsymbol{C}}}$ are constant matrices in the numerator, the poles of a MIMO state space model are determined from $(s\underline{\boldsymbol{I}} \quad \underline{\underline{\boldsymbol{A}}})^{\ 1}$. A numerical example will be presented to demonstrate how these equations can be used.

Example 5.16 State Space Analysis for a Continuous Bio-Reactor

Problem Statement. Using the state space model developed in Example 5.15, determine the poles for the process represented by this state space model and $\underline{\underline{\boldsymbol{G}}}_p(s)$ for this system.

Solution. The poles of a state space model are determined from the $\underline{\underline{\boldsymbol{A}}}$ matrix. In this case,

$$\underline{\underline{\boldsymbol{A}}} \qquad \begin{matrix} 0 & 0 & 0 \\ 10.53 & 0.2 & 0 \\ 0.04 & 0 & 0.2 \end{matrix}$$

The poles are given by the **eigenvalues** of the matrix $\underline{\underline{\boldsymbol{A}}}$. The eigenvalues of $\underline{\underline{\boldsymbol{A}}}$ are the values of s that satisfy

$$\det \ s\underline{\boldsymbol{I}} \quad \underline{\underline{\boldsymbol{A}}} \quad 0$$

That is,

$$\det \begin{matrix} s & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & s \end{matrix} \qquad \begin{matrix} 0 & 0 & 0 \\ 10.53 & 0.2 & 0 \\ 0.04 & 0 & 0.2 \end{matrix} \qquad 0$$

Simplifying

$$\det \begin{matrix} s & 0 & 0 \\ 10.53 & s \quad 0.2 & 0 \\ 0.04 & 0 & s \quad 0.2 \end{matrix} \qquad 0$$

The determinant (det) of this matrix is

$$s(s \quad 0.2)^2 \quad 0$$

Therefore, the poles are 0, -0.2 and -0.2. The s=0 pole indicates that this is an integrating process (see Section 6.5). Applying Equation 5.9.1 for this example yields

$$\underline{\underline{G}}_p(s) = \underline{C}(s\underline{\underline{I}} - \underline{\underline{A}})^{-1}\underline{\underline{B}}$$

$$\underline{\underline{G}}_p(s) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s & 0 & 0 \\ 10.52 & s+0.2 & 0 \\ 0.04 & 0 & s+0.2 \end{bmatrix}^{-1} \begin{bmatrix} 0.00005 \\ 0.005 \\ 0.00025 \end{bmatrix}$$

Using standard matrix inversion techniques[5]

$$\begin{bmatrix} s & 0 & 0 \\ 10.52 & s+0.2 & 0 \\ 0.04 & 0 & s+0.2 \end{bmatrix}^{-1} = \begin{bmatrix} \dfrac{1}{s} & 0 & 0 \\ \dfrac{-10.52}{s(s+0.2)} & \dfrac{1}{(s+0.2)} & 0 \\ \dfrac{-0.004}{s(s+0.2)} & 0 & \dfrac{1}{(s+0.2)} \end{bmatrix}$$

Then by matrix multiplication of $\underline{\underline{C}}$, $(s\underline{\underline{I}} - \underline{\underline{A}})^{-1}$ and $\underline{\underline{B}}$ yields,

$$\underline{\underline{G}}_p(s) = \begin{bmatrix} \dfrac{X(s)}{F(s)} \\ \dfrac{S(s)}{F(s)} \\ \dfrac{P(s)}{F(s)} \end{bmatrix} = \begin{bmatrix} \dfrac{0.00005}{s} \\ \dfrac{0.005}{s+0.2} \\ \dfrac{-0.00025}{s+0.2} \end{bmatrix}$$

Note that in this case, the transfer function for x could have been determined directly using Equation 5.7.3. Likewise, the transfer functions for S and P can be determined by applying Equation 5.7.3 using the transfer function for x.

Self-Assessment Questions

Q5.9.1 Which parts of the state space model equations (Equation 5.8.6) determine the poles of the system?

Q5.9.2 What advantage does Equation 5.9.1 have over Equation 5.8.6?

Self-Assessment Answers

Q5.9.1 The Jacobian matrix (i.e., the $\underline{\underline{A}}$ matrix in Equation 5.8.6) determines the poles of a state space model.

Q5.9.2 Once Equation 5.9.1 is evaluated, you can tell by inspection the dynamic behavior of each input/output pair using your knowledge of transfer functions. On the other hand, Equation 5.8.6 contains the same information, but in a different form that does not lend itself to such a simple analysis.

5.10 Summary

- A transfer function, $G(s)$, is defined as the ratio of the Laplace transform of the output variable, $Y(s)$, divided by the Laplace transform of the input variable, $U(s)$, both in deviation variable form, i.e., $G(s) = Y(s)/U(s)$.
- Transfer functions are independent of the input applied to the system and the initial conditions.
- Deviation variables are defined as $y'(t) = y(t) - \bar{y}$ and $u'(t) = u(t) - \bar{u}$, where $\bar{y}$ and $\bar{u}$ are the initial steady-state conditions of output and input variables, respectively.
- Poles of a transfer function [i.e., the values of s that render $D(s) = 0$ for $G(s) = N(s)/D(s)$] determine a major portion of the dynamic characteristics of the system represented by a transfer function. If any of the poles of a transfer function have positive real components, unstable behavior is indicated.
- Routh stability criterion is a convenient means to determine the stability of a system represented by a transfer function.
- Zeros of a transfer function [i.e., the values of s that render $N(s) = 0$ for $G(s) = N(s)/D(s)$] can indicate characteristics of the dynamic response of a system represented by a transfer function. If any of the zeros of a transfer function are positive, an inverse-responding system is indicated.
- Block diagram algebra can be used to derive the overall transfer functions of a block diagram representation of a system.
- The following equation can be conveniently used to determine the transfer function for a process from either a linear or nonlinear model equation of the form, $dy/dt = f(y,u)$.

$$G(s) = \frac{Y(s)}{U(s)} = \frac{\left.\dfrac{\partial f(y,u)}{\partial u}\right|_{\bar{y},\bar{u}}}{s - \left.\dfrac{\partial f(y,u)}{\partial y}\right|_{\bar{y},\bar{u}}}$$

- State space models can be developed using the following matrix equation

$$\frac{d\underline{\boldsymbol{x}}}{dt} = \underline{\underline{\boldsymbol{A}}}\,\underline{\boldsymbol{x}} + \underline{\underline{\boldsymbol{B}}}\,\underline{\boldsymbol{u}}$$
$$\underline{\boldsymbol{y}} = \underline{\underline{\boldsymbol{C}}}\,\underline{\boldsymbol{x}}$$

  where the elements of $\underline{\underline{\boldsymbol{A}}}$ and $\underline{\underline{\boldsymbol{B}}}$ are given by $a_{ij} = \partial f_i/\partial x_j$ and $b_{ij} = \partial f_i/\partial u_j$, respectively. $\underline{\underline{\boldsymbol{C}}}$ indicates the correspondence between the output variables ($\underline{\boldsymbol{y}}$) and the state variables ($\underline{\boldsymbol{x}}$).

- The poles of a MIMO process are equal to the values of s that satisfy the following equation

$$\det\left(s\underline{\underline{\boldsymbol{I}}} - \underline{\underline{\boldsymbol{A}}}\right) = 0$$

5.11 References

1. Routh, E.J., *Dynamics of a System of Rigid Bodies*, Part 11, MacMillan, London (1905).
2. D'Azzo, J.J. and C.H. Houpis, *Linear Control System Analysis and Design: Conventional and Modern,* McGraw-Hill, Inc., N.Y. (1975).
3. Riggs, J.B., *Computational Methods for Engineers*, Ferret Publishing, Austin, Tx, p. 222 (2013).
4. Riggs, J.B. and M.N. Karim, Chemical and Bio-Process Control, 3rd ed., Ferret Publishing, Austin, Texas, p. 113.
5. Ogata, K., *Modern Control Engineering, 2nd Edition*, Prentice-Hall, Englewood, NJ, p. 898 (1990).

5.12 Additional Terminology

Bilinear terms - the product of two variables.

Complex conjugate pairs - poles of a transfer function that indicate oscillatory time behavior.

Deviation variable - a variable that results from subtracting the initial steady-state conditions from the dependent or independent variable value.

Eigenvalue - of a matrix $\underline{\underline{A}}$ are the values of s that satisfy the equation $\det\left|s\underline{\underline{I}} - \underline{\underline{A}}\right| = 0$.

Inverse response - response of a system that initially moves in a direction that is opposite to the change corresponding to the long-time settling value.

Identity matrix - a square matrix with each element equal to zero except for unity values along the diagonal of the matrix.

Jacobian matrix - all the combinations of the partial derivatives of ODE functions with respect to state variables.

Left-half plane poles - poles with a negative real component.

Left-half plane zeros - zeros with a negative real component.

Linearization - developing a linear approximation of a nonlinear equation.

Overshoot - when the response of a system goes through a maximum or minimum before settling at it steady-state value.

Physical realizability condition - for a transfer function in which the order of the denominator is less than or equal to the order of the numerator.

Poles of a transfer function - the roots of the equation $D(s) = 0$ where $D(s)$ is the denominator of the transfer function.

Right-half plane poles - poles that have a positive real component, which indicate unstable behavior.

Right-half plane zeros - zeros that have a positive real component, which indicate inverse-acting behavior.

Ringing - an industrial term referring to a response that exhibits slow damping of the oscillations.

Routh stability criterion - the necessary and sufficient conditions for stability of a system is that all the elements of the first column in the Routh array must be positive.

State space model - a compact formulation of the linearized form of a set of nonlinear dynamic model equations.

State space variables - the minium number of process variable values necessary to determine all the conditions of a system.

Sustained oscillations - oscillation with a fixed amplitude.

Transfer function - the ratio of the Laplace transform of the output variable divided by the Laplace transform of the input variable each given in deviation variable form.

Unstable behavior - unbounded growth of a dependent variable for a bounded input.

Zeros of a transfer function - the roots of the equation $D(s) = 0$ where $D(s)$ is the numerator of the transfer function.

5.13 Preliminary Questions

5.2 General Characteristics of Transfer Functions

Q5.2.1 How are Laplace transforms related to transfer functions?

Q5.2.2 How can the form of a transfer function be used to determine the dynamic behavior of the process it represents?

5.3 Poles of a Transfer Function

Q5.3.1 What are the poles of a transfer function and what are their significance?

Q5.3.2 What factor in the denominator of a transfer function indicates damped exponential behavior?

Q5.3.3 What factor in the denominator of a transfer function indicates damped oscillatory behavior?

Q5.3.4 What factor in the denominator of a transfer function indicates unbounded oscillatory growth?

Q5.3.5 What is a complex plane and how is it used to represent the dynamic behavior of a process?

5.5 Zeros of a Transfer Function

Q5.5.1 Under what conditions will a transfer function represent overshoot and inverse action?

Q5.5.2 Summarize the dynamic response of the transfer function given by Equation 5.5.1 based on the value of c relative to a and b.

5.6 Block Diagrams using Transfer Functions

Q5.6.1 What are the three elements of block diagram algebra?

Q5.6.2 What is the difference between a summation function and a signal divider?

5.7 Linearization of Nonlinear Differential Equations

Q5.7.1 What two ways can you obtain a transfer function of a nonlinear process model?

Q5.7.2 Explain why Equation 5.7.3 can be used to determine the transfer function for a linear ODE process model?

5.8 State Space Models

Q5.8.1 What three factors are state space models based on?

Q5.8.2 What is the Jacobian matrix and how is it related to state space models?

Q5.8.3 What is an output variable and how is it related to state variables?

Q5.8.4 State space models are based on what type of nonlinear model?

Q5.8.5 When does Equation 5.9.1 produce a vector of transfer functions and when does it produce a matrix of transfer functions?

5.14 Analytical Questions and Exercises

5.2 General Characteristics of Transfer Functions

P5.2.1* Develop the transfer function for the effect of u on y for the following differential equations, assuming $u(0)$=0, $y(0)$=0 and $y'(0)$=0.

a. $\frac{d^2y}{dt^2} \quad 5\frac{dy}{dt} \quad 6y \quad u$ b. $\frac{d^2y}{dt^2} \quad 3\frac{dy}{dt} \quad u$

c. $\frac{d^2 y}{dt^2} \quad 4\frac{dy}{dt} \quad 5y \quad u$

d. $\frac{d^2 y}{dt^2} \quad 4\frac{dy}{dt} \quad 4y \quad u$

e. $2\frac{d^2 y}{dt^2} \quad 11\frac{dy}{dt} \quad 12y \quad 5u$

f. $3\frac{d^2 y}{dt^2} \quad 7\frac{dy}{dt} \quad 2y \quad u$

P5.2.2** Determine $y(t)$ for a unit impulse input and a step input of unit magnitude for

a. Problem 5.2.1 a

b. Problem 5.2.1 b

5.3 Poles of a Transfer Function

P5.3.1* Describe the dynamic behavior indicated by each of the following transfer functions.

a. $G(s) \quad \frac{2}{2s \quad 1}$

b. $G(s) \quad \frac{3}{(s \quad 1)(s \quad 4)}$

c. $G(s) \quad \frac{1}{s^2 \quad s \quad 1}$

d. $G(s) \quad \frac{1}{s^2 \quad s \quad 1}$

e. $G(s) \quad \frac{1}{s^2 \quad 9}$

f. $G(s) \quad \frac{1}{s^2 \quad 2s \quad 4}$

5.4 Stability Analysis using the Routh Array

P5.4.1* Determine the stability of the following transfer function using the Routh stability criterion.

$$G(s) \quad \frac{2s^2 \quad s \quad 1}{3s^3 \quad 3s^2 \quad 4s \quad 5}$$

P5.4.2** Using the Routh stability criterion evaluate the stability of a general second-order process, i.e.,

$$G(s) \quad \frac{b_1 s \quad b_0}{a_2 s^2 \quad a_1 s \quad a_0}$$

5.5 Zeros of a Transfer Function

P5.5.1 Determine the type of dynamic response for each of the following transfer function: monotonic, overshoot or inverse response.

a.* $\frac{s \quad 2}{(s \quad 9)(s \quad 3)}$

b.* $\frac{s \quad 2}{(s \quad 9)(s \quad 3)}$

c.* $\frac{s \quad 5}{(s \quad 4)(s \quad 3)}$

d.** $\frac{5s \quad 4}{3s^2 \quad 7s \quad 3}$

e.** $\frac{3s \quad 1}{3s^2 \quad 5s \quad 2}$

f.** $\frac{8s \quad 7}{8s^2 \quad 10s \quad 3}$

5.6 Block Diagrams using Transfer Functions

P5.6.1** For the block diagram shown in Figure P5.6.1, develop the transfer function between $Y(s)$ and $U(s)$.

P5.6.2** For the block diagram shown in Figure P5.6.2, develop the transfer function between $C(s)$ and $E(s)$.

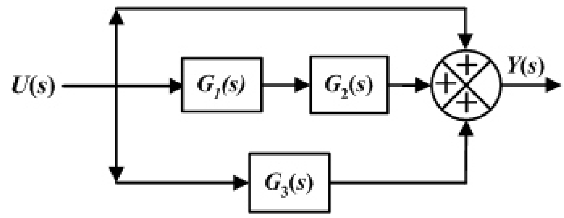


Figure P5.6.1 Block diagram for Problem 5.6.1.

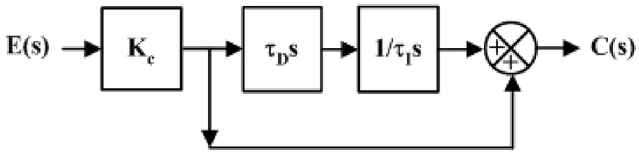


Figure P5.6.2 Block diagram for Problem 5.6.2.

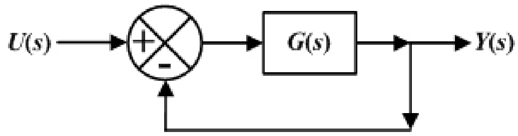


Figure P5.6.3 Block diagram for Problem 5.6.3.

P5.6.3*** For the block diagram shown in Figure P5.6.3, develop the transfer function between $Y(s)$ and $U(s)$.

P5.6.4*** Show that the two block diagrams shown in Figure P5.6.4 are equivalent, i.e., the overall transfer functions are the same.

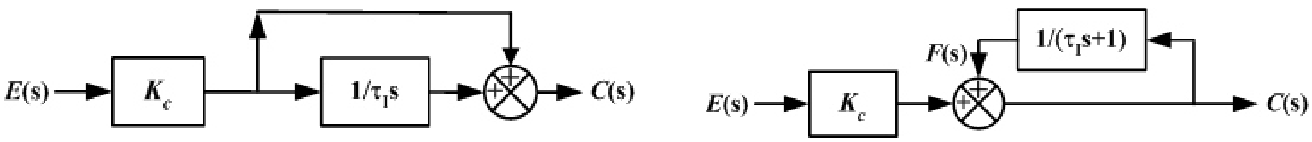


Figure P5.6.4 Block diagrams for Problem 5.6.4

5.7 Linearization of Nonlinear Differential Equations

P5.7.1** Consider the stirred-tank heater shown Figure P5.7.1. The model equation for this process is

$$C_p V \frac{dT}{dt} \quad C_p F(T_{in} \quad T) \quad Q$$

where (feed density), C_p (heat capacity of the feed), V (volume in the tank), T_{in} (inlet feed temperature) and F (feed rate to the system) are assumed constant, T (product temperature) is the output variable and Q (rate of heat addition) is the input variable. Assume that initially the system is at steady state at $(\bar{T}, \bar{Q})$. Determine the transfer function for this process by (a) deriving it using deviation variables and Laplace transforms and (b) using Equation 5.7.3.

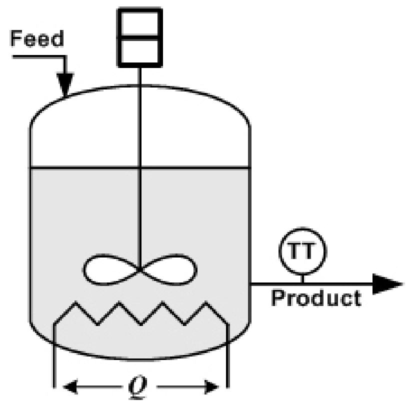


Figure P5.7.1 Schematic of a CST heater.

P5.7.2*** Consider a tray in a distillation column shown in Figure P5.7.2. Liquid from the tray above (L) enters the down comer and flow across the tray over the weir and into the down comer below the tray. Vapor from the tray below (V) passes through the liquid on the tray extracting the light component as it leaves the tray. The mole balance for the light component is given by

$$M\frac{dx_i}{dt} \quad x_{i\ 1}L \quad x_iL \quad y_{i\ 1}V \quad y_iV$$

where M is the molar holdup on the tray, x_i is the mole fraction of the light component in the liquid on the tray i and y_i is the mole fraction of the light component in the vapor leaving tray i. The relationship between x_i and y_i is in terms of the relative volatility ( ) is given by

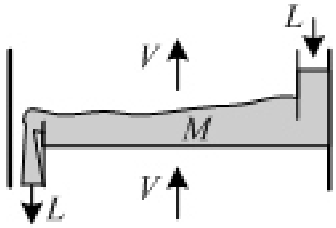


Figure P5.7.2 Cross-section of a tray in a distillation column.

$$y_i \quad \frac{x_i}{1 \ \ (\quad 1)x_i}$$

Assuming that M, x_{i+1}, y_{i-1} and are constant, determine the transfer functions for the effect of changes in L and V on x_i.

P5.7.3** For the compressed air tank process (Figure P5.7.3), determine the transfer function for the effect of stem position on the pressure in the tank using Equation 5.7.3. The process model and data are given as

$$\frac{dm}{dt} \quad 180x\sqrt{P_{\text{supply}} \quad \frac{mRT}{V_{\text{tank}}M}}$$

$$P_{tank} \quad \frac{mRT}{V_{tank}M}$$

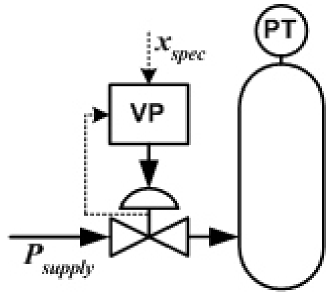


Figure P5.7.3 Diagram of a compressed air tank attached to an air supply.

- $C_v(x)$ - valve coefficient (2.67 when the control valve is fully open)
- F - the flow rate into the compressed air tank (lb/h)
- M - the molecular weight of air (29 lb/lb-mol)
- m - mass of air in the compressed air tank (initially 14.9 lb)
- P_{tank} - the pressure in the compressed air tank (initially 14.7 psia)
- P_{supply} - the pressure of the supply of compressed air (350 psia)
- R - gas law constant (10.73 psia-ft^3/lb-mol-ºR)
- T - temperature of the air in the tank (75ºF)
- V_{tank} - volume of the compressed air tank (200 ft^3)
- x - stem position for the control valve (initially 0 %)

5.7.4** For the DO process shown in Figure P5.7.4, determine the transfer function for the effect of air flow rate on the oxygen concentration in the broth using Equation 5.7.3. The process model and data are given as

$$\frac{dC_{O_2}}{dt} \quad k_La\ C^*_{O_2} \quad C_{O_2} \qquad K_{O_2} \quad {}_{\max}x$$

$$k_La \quad 0.25 \quad 0.001(F_{air} \quad 500)$$

- C_{O_2} - concentration of O_2 in the reaction broth (initially 1.1×10^{-4} g-mol/l).
- $C^*_{O_2}$ - saturated concentration of O_2 in the broth (2.20×10^{-4} g-mol/l).
- $C_{O_2,s}$ - the measurement of the O_2 concentration in the broth (initially 1.1×10^{-4} g-mol/l)

- F_{air} - the volumetric flow rate of air to the bio-reactor (500 scfm)
- K_{O_2} - cellular uptake of O_2 (1.98 g-mol O_2/g-cells)
- k_La - the overall liquid phase mass transfer coefficient for transport from the bubble surface to the bulk broth (initially 0.25 s^{-1}).
- T - broth temperature (35ºC)
- V - the volume of broth in the bio-reactor (1000 l).
- x - constant cell concentration in the bio-reactor (0.25 g/l).
- $_{max}$ - maximum specific growth rate (5.56×10^{-5} s^{-1}).

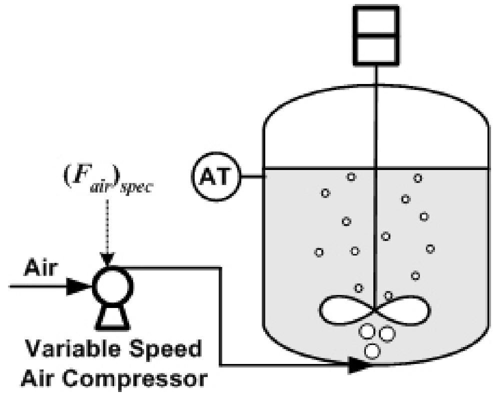


Figure P5.7.4 Schematic of a DO process.

5.7.5** For the CST thermal mixing process shown in Figure P5.7.5, develop the transfer function for the effect of the specified flow rate for F_1 on the **measured** value of T. The process model and data are given as

Actuator $$\frac{dF_1}{dt} \quad \frac{1}{_v}(F_{1,spec} \quad F_1)$$

Process $$M\frac{dT}{dt} \quad F_1T_1 \quad F_2T_2 \quad (F_1 \quad F_2)\,T$$

Sensor $$\frac{dT_s}{dt} \quad \frac{1}{_{Ts}}(T \quad T_s)$$

- F_1 - mass flow rate of stream 1 (initially 5 kg/s)
- F_2 - constant mass flow rate of stream 2 (5 kg/s)
- M - mass of liquid in the mixer (100 kg)
- T - temperature of the mixed liquid (initially 50 C)
- T_1 - temperature of stream 1 (25 C)
- T_2 - constant temperature of stream 2 (75 C
- $_v$ - the time constant for the flow controller on stream 1 (2 s).
- $_{Ts}$ - the time constant for the temperature sensor on the product stream (6 s).

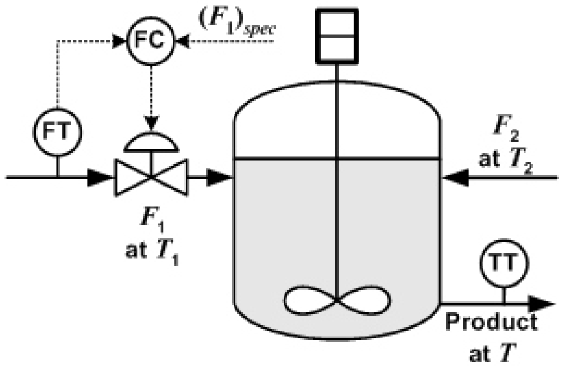


Figure P5.7.5 Schematic of a CST thermal mixing process.

5.8 State Space Models[AT]

P5.8.1** Develop a state space model for the following set of model equations that describe a second-order reaction in a semi-batch reactor (Figure P5.8.1). The first equation is an overall mass balance equation and the second equation is based on a component balance on component A.

$$A_c \frac{dL}{dt} \quad F_{in}$$

$$A_c L \frac{dC_A}{dt} \quad F_{in} C_{A_0} \quad k\,L\,A_c C_A^2$$

where , A_c, C_{A_0} and k are constants or process parameters, F_{in} is the input variable and L and C_A are output (measured) variables. Initially, C_A is equal to $\bar{C}_A$ and L is equal to $\bar{L}$. What assumptions is this state space model based on?

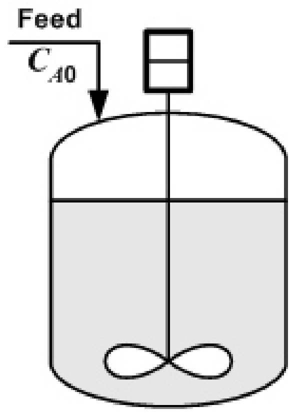


Figure P5.8.1 PFD of a semi-batch reactor.

P5.8.2*** Determine the poles of the system and develop the transfer functions for Problem 5.8.1 using Equation 5.9.1.

P5.8.3** Develop the complete state space model for the process model for the CSTR given in Example 3.5 (Equations 3.6.10 and 3.6.11).

Chapter 6

Dynamic Behavior of Ideal Systems

Chapter Objectives

- Specify several idealized forms for inputs.
- Present the functional form of a variety of idealized dynamic models along with their model parameters.
- Demonstrate how to use process measurements or observations to estimate the key parameters for certain idealized process models.
- Introduce process integration and illustrate how it affects process behavior.

6.1 Introduction

In the previous two chapters, Laplace transform solutions and transfer functions were developed for a number of process models. In this chapter, the dynamic behavior and transfer functions for a variety of idealized systems, e.g., a first-order process, a second-order process, an integrating process, etc., are presented. From this chapter, it should become clear that a full range of process behavior can be represented using these idealized representations. For example, a CST composition mixer, a CSTR with a first-order reaction, flow through a control valve, a temperature measurement from a thermocouple, and charge storage in a capacitor are all well represented with a simple first-order dynamic model. Also, most level systems can be represented as integrating processes and feedback control systems are often described in terms of second-order process parameters. Therefore, the parameters of these idealized models can concisely describe the dynamic behavior of industrial processes, e.g., the speed of the response and the sensitivity of the process to changes in the input. Even though idealized models are not normally used to quantitatively represent industrial processes, control engineers routinely describe the qualitative behavior of their processes using the terminology associated with idealized dynamic models. Therefore, idealized models are critically important to the understanding of process dynamics and the terminology of the process control profession.

6.2 Idealized Process Inputs

Process inputs include manipulated variables, measured disturbances and unmeasured disturbances. Each of the input types considered here can be applied using a manipulated variable, while only ramp and sinusoidal inputs are usually used to describe the effect of disturbances. By understanding how a process responds to one or more of these idealized inputs, you should be able to understand the general dynamic behavior of the process in question.

Impulse input. A unit impulse input has an infinite height for an infinitesimal duration so that the area under the impulse input is unity. An **impulse input** is shown graphically in Figure 6.2.1a. From Table 4.1, the Laplace transform of a unit impulse input applied at t=0 is

$$U(s) \quad 1 \tag{6.2.1}$$

While it is not possible to implement a true impulse input, in certain cases it is possible to approximate an impulse input using a short duration, large magnitude change in $u(t)$. This approach to approximating an impulse input is unlikely to yield an integral of the input equal to unity. For this case,

$$U(s) \quad A \tag{6.2.2}$$

where A is the integral of the change in $u(t)$ from the initial conditions over the duration of the change. In addition, an injection of a concentrated dye, tracer or salt can be used to approximate an impulse input where A in Equation 6.2.2 is equal to the mass of dye, tracer or salt injected into the system.

Step input. One of the easiest input changes to implement is the **step change**, which is a sudden and sustained change. A step change of magnitude A at t=t_0 can be represented as

$$\begin{array}{llll} u(t) & 0 & \qquad t & t_0 \\ u(t) & A & \qquad t & t_0 \end{array} \tag{6.2.3}$$

From Table 4.1, the Laplace transform for a step change of A applied at t_0=0 is given by

$$U(s) \quad \frac{A}{s} \tag{6.2.4}$$

An idealized step change is shown graphically in Figure 6.2.1b. In chemical and bio-process control, manipulated variables are normally flow rates; therefore, due to the dynamics of the final control element and other factors, the actual flow rate does not change instantaneously. If the specified flow rate or the signal to the final control element is considered as the input to the process, virtually instantaneous step changes in inputs can be implemented on industrial processes.

Rectangular pulse. A **rectangular pulse** is similar to a step change except that the input is returned to its original value after a specified amount of time. Thus, a rectangular pulse can be considered as a series of two step changes. A rectangular pulse is given by

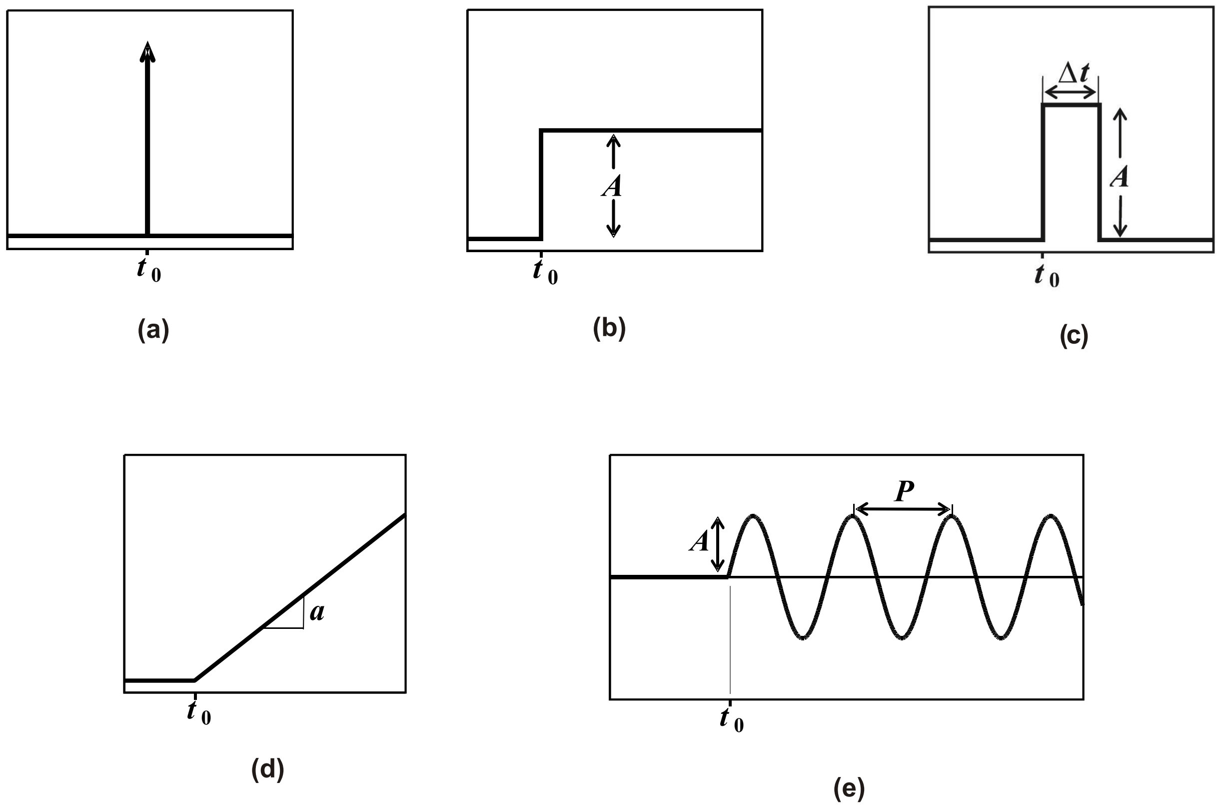


Figure 6.2.1 Idealized inputs (a) impulse, (b) step, (c) pulse, (d) ramp and (e) sinusoidal.

$$\begin{array}{lll} u(t) \quad 0 & t \quad t_0 & \\ u(t) \quad A & t_0 \quad t \quad t_0 \quad t & \\ u(t) \quad 0 & t \quad t_0 \quad t & \end{array} \qquad \textbf{6.2.5}$$

The rectangular pulse is said to have a strength of $A \quad t$, which is equal to $\int_0^t u(t)dt$. Figure 6.2.1c graphically shows a rectangular pulse. From Table 4.1, the Laplace transform of a rectangular pulse is given by

$$U(s) \quad \frac{A}{s}[1 \quad e^{\quad t\, s}] \qquad \textbf{6.2.6}$$

for t_0 is equal to zero (i.e., the pulse input begins at t=0).

Ramp input. Certain types of disturbances can be reasonably represented as ramps. For example, air temperature or cooling-water temperature can change steadily with a relatively constant slope during certain portions of the day. Ramp inputs are also common in batch processing, e.g., a pulp digester in a paper plant, where the temperature in the digester may be increased at a constant slope. A **ramp input** is given by

$$u(t) \quad 0 \qquad t \quad t_0$$
$$u(t) \quad a\,t \qquad t \quad t_0 \tag{6.2.7}$$

and is illustrated in Figure 6.2.1d. From Table 4.1, the Laplace transform of a ramp input is given by

$$U(s) \quad \frac{a}{s^2} \tag{6.2.8}$$

for t_0 equal to zero. Although the ramp described here (Equations 6.2.7 and 6.2.8) is unbounded, in practice ramps are applied for a limited period of time.

Sinusoidal inputs. The time scale over which inputs change can have an important effect on feedback control performance (see Chapter 11). For example, air temperature disturbances have a 24 hour period due to day-to-night variations. On the other hand, feed flow rate changes to a process can have a period of minutes or seconds. Because the frequency is directly related to the time scale of the input, one way to evaluate the effect of different time scales for inputs is to use sinusoidal inputs with different frequencies.

A **sinusoidal input** is given by

$$u(t) \quad 0 \qquad t \quad t_0$$
$$u(t) \quad A\sin \quad t \qquad t \quad t_0 \tag{6.2.9}$$

where   is the radian frequency (rad/time) and A is the amplitude of the sinusoidal input. An example of a sinusoidal input is given in Figure 6.2.1e. The period, P, is equal to 2 / . From Table 4.1 for t_0=0, the Laplace transform for a sinusoidal input is

$$U(s) \quad \frac{A}{s^2 \qquad ^2} \tag{6.2.10}$$

Example 6.1 Derivation of the Laplace Transform for a Non-Standard Input

Problem Statement. Determine the Laplace transform for the input profile shown in Figure 6.2.2.

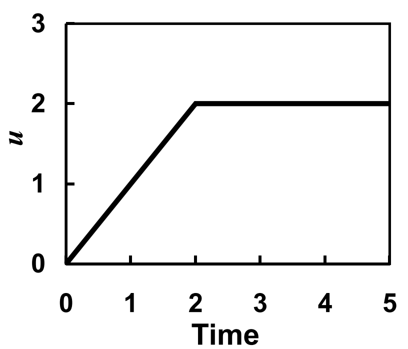


Figure 6.2.2 A non-standard input profile.

Solution. This input profile can be represented in the time domain as

$$u(t) \quad 0 \qquad t \quad 0$$
$$u(t) \quad t \qquad 0 \quad t \quad 2$$
$$u(t) \quad 2 \qquad t \quad 2$$

Until t=2, the input [$u(t)$] behaves as a simple ramp. After t=2, the input is constant at a value of 2. Figure 6.2.3 shows how two ramps can be combined to produce the desired response. Because the first portion of the response is a simple ramp, u_1 in Figure 6.2.3 is given by

$$u_1 \quad t$$

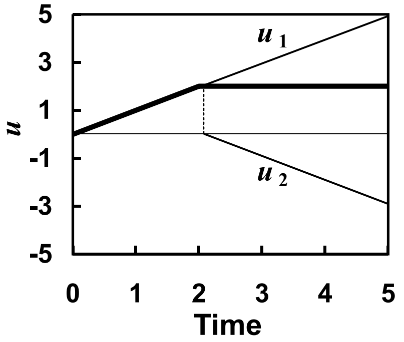


Figure 6.2.3 Combination of ramps for the solution of Example 6.1.

For t 2, a negative ramp that starts at t 2 is required, i.e.,

$$u_2 \quad (t \quad 2) \qquad t \quad 2$$

u_1 and u_2 are added together to provide the desired response. Taking the Laplace transforms of u_1 and u_2 yield

$$U_1(s) \quad \frac{1}{s^2} \qquad\qquad U_2(s) \quad \frac{e^{\ \ 2t}}{s^2}$$

Then, the sum of $U_1(s)$ and $U_2(s)$ yields

$$U(s) \quad \frac{1}{s^2} \quad \frac{e^{\ \ 2s}}{s^2} \quad (1 \quad e^{\ \ 2s})\frac{1}{s^2}$$

Self-Assessment Questions

Q6.2.1 What are the primary parameters of a rectangular pulse input?

Q6.2.2 What aspects of an ideal impulse input are difficult to attain for a real system?

Self-Assessment Answers

Q6.2.1 From Figure 6.2.1c, the height of the pulse (A) and the duration (t) are the primary parameters of a pulse input.

Q6.2.2 The aspect of an impulse input that is difficult to implement industrially is an infinite input amplitude over an infinitesimal duration. Instead, a relatively large amplitude over a finite, but small, duration can be applied to approximate an impulse input.

6.3 First-Order Processes

A CST composition mixer, a CST thermal mixer and an isothermal CSTR with a first-order reaction are examples of **first-order processes**. The differential equation for a first-order process written in the standard form is given by

$$\ _p \frac{dy(t)}{dt} \quad y(t) \quad K_p\, u(t) \qquad \textbf{6.3.1}$$

where y is the output variable, u is the input variable, $\boldsymbol{K_p}$ is the steady-state process gain and $\ _p$ is the process time constant. The process gain is the steady-state change in y divided by the corresponding change in u (Figure 6.3.1), i.e.,

$$K_p \quad \frac{y}{u}$$

The time constant for the CST thermal mixer and the CST composition mixer is the volume of liquid in the CST mixer divided by the total volumetric feed rate, i.e., the residence time of the CST mixer.

The transfer function for a first-order process is given by

$$G_p(s) = \frac{K_p}{\tau_p s + 1} \qquad \textbf{6.3.2}$$

Because the differential equation for a thermal mixer [Equation 3.7.7] can be rearranged into the same form as Equation 6.3.1, considering T_1 as the only input, the thermal mixer is a first-order process. Likewise, the transfer function of any first-order process can be rearranged into the same form as Equation 6.3.2. In each case, the process gain and process time constant can be identified directly. The standard form for a first-order differential equation (Equation 6.3.1) requires that the coefficient of y be unity while the standard form for the transfer function of a first-order process (Equation 6.3.2) requires that the constant term in the denominator be unity. Note that K_p and τ_p are directly available once Equations 6.3.1 and 6.3.2 are in the standard form.

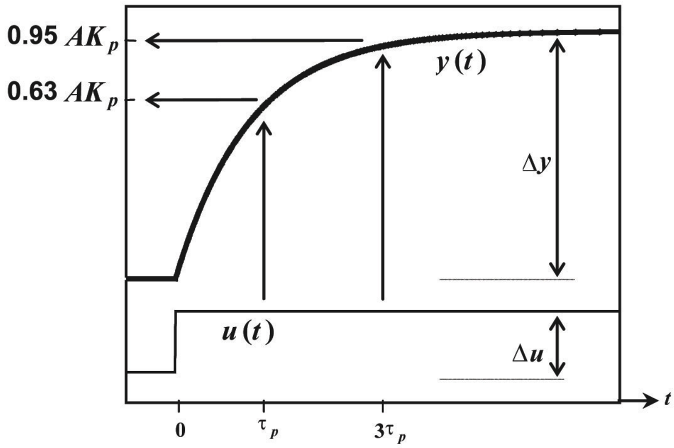


Figure 6.3.1 Dynamic response of a first-order process to a step input change.

Response to a Step Input. Figure 6.3.1 shows the response of a first-order process [$y(t)$] to a step change in $u(t)$. Assuming that $y(0)=0$, the analytical solution of Equation 6.3.1 for a step change, A, in u is given by

$$y(t) = A\,K_p(1 - e^{-t/\tau_p}) \qquad \textbf{6.3.3}$$

where K_p and τ_p are first-order process parameters. The process gain, K_p and the size of the step change, A, determine the new steady-state value of y. The time constant, τ_p, determines the dynamic path the process takes as it approaches the new steady-state, i.e., how long it takes to approach the new steady-state. Note that **63.2% of the final change occurs in one time constant after the input change; 95% of the change occurs in three time constants** (Figure 6.3.1); and **98% of the change occurs in four time constants** (not shown in Figure 6.3.1).

Example 6.2 Characteristics of a First-Order Process

Problem Statement. For the following first-order transfer function, calculate the process gain and the time constant. Also, determine the time required after a step input change for 98% of the change in the output variable to occur.

$$G_p(s) = \frac{16}{s + 2}$$

Solution. Rearrange the transfer function into the standard form for a first-order process by dividing the numerator and denominator by 2 resulting in

$$G_p(s) = \frac{8}{0.5s + 1}$$

Therefore, the process gain, K_p, is 8 and the process time constant, τ_p, is 0.5 time units. Finally, 98% of the total change occurs in four time constants, which corresponds to 2 time units.

Example 6.3 Estimation of a First-Order Model from Plant Observations

Problem Statement. After observing a process, the operator indicates to the control engineer that an increase of 1,000 lb/h of steam (the input) to a reactor produces a 5°F increase in the reactor temperature (the output). When a change in the steam flow rate is made, it takes approximately 40 minutes for the full effect on the reactor temperature to be observed. Using this process information, develop a first-order model for this process.

Solution. The gain of the process, K_p, can be estimated based on the steady-state changes observed for the process, i.e.,

$$K_p = \frac{\Delta y}{\Delta u} = \frac{5^{\circ}\mathrm{F}}{1000\ \mathrm{lb/h}} = 0.005\ \ {}^{\circ}\mathrm{F\text{-}h/lb}$$

The open-loop settling time is 40 minutes, which is approximately equal to 4 τ_p. Therefore, the time constant is 10 minutes. The first-order transfer function for this process is

$$G(s) = \frac{0.005}{10s + 1}$$

where time is given in minutes and the gain has units of (°F-h/lb).

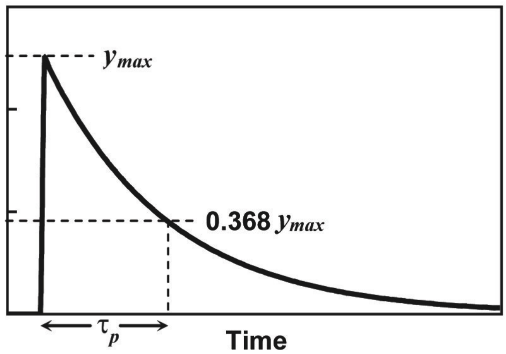


Figure 6.3.2 Response of a first-order process to an impulse input.

Response to an Impulse Input. Figure 6.3.2 shows the response of a first-order process to an impulse input. Assuming that $y(t)$ is initially equal to zero, the analytical solution of Equation 6.3.1 for an A magnitude impulse input is given by

$$y(t) = \frac{AK_p}{\tau_p} e^{-t/\tau_p} \qquad \mathbf{6.3.4}$$

where K_p and τ_p are the parameters for a first-order process. Note that y_{max} in Figure 6.3.2 is equal to the immediate process response after the impulse input is implemented (t=0) and is given by

$$y_{max} = \frac{AK_p}{\tau_p} \qquad \mathbf{6.3.5}$$

In addition, the time required for $y(t)$ to decrease to 36.8% of y_{max} is equal to τ_p from Equation 6.3.4 as shown in Figure 6.3.2.

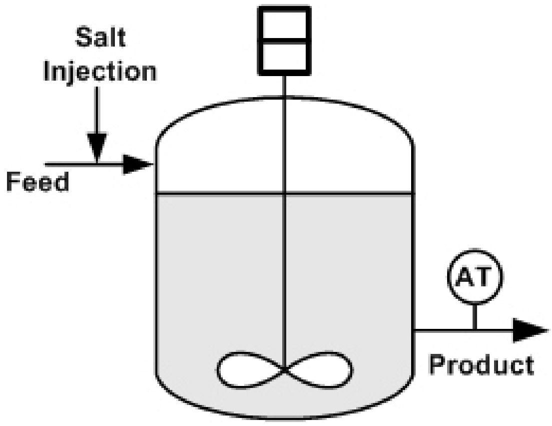


Figure 6.3.3 Schematic of the application of a tracer study to a reactor.

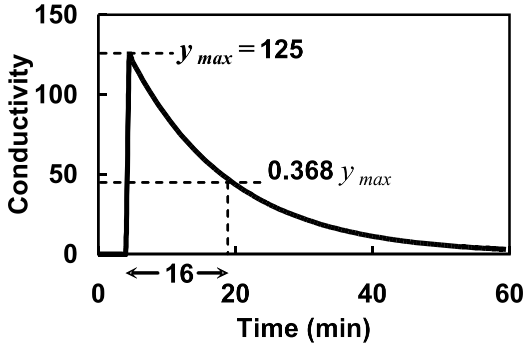


Figure 6.3.4 Response of the reactor to an impulse tracer test.

Example 6.4 Analysis of the Response of a First-Order Process to an Impulse Input

Problem Statement. A tracer study is a common way to analyze the mixing performance for a reactor. A concentrated dye or salt is injected into the feed to the reactor. Then the electrical conductivity (for salt) or color intensity (for dye) are measured at the outlet of the reactor (Figure 6.3.3). Figure 6.3.4 shows an impulse test that was performed for this reactor system by injecting 5 kg of salt in solution. Assuming a first-order process, determine the time constant and the process gain for this system.

Solution. First, the time constant must be determined using y_{max} and the response curve. Then using the value of the time constant, the gain can be calculated from y_{max}. The time constant is equal to the time after the salt injection when the conductivity reduces to 46 siemens/m (i.e., 0.368×125). (Note that siemens are units of electrical conductivity.) From Figure 6.3.4, the time constant is equal to 16 min. From Figure 6.3.4, y_{max} is equal to 125 siemens/m. Rearranging Equation 6.3.5 to solve for the process gain yields

$$K_p = \frac{\tau_p y_{\max}}{A} = \left.\frac{16\,\text{min}}{5\,\text{kg salt}}\right|\frac{125\,\text{siemens}}{\text{m}} = 400\frac{\text{siemens}}{\text{kg salt}}\frac{\text{min}}{\text{m}}$$

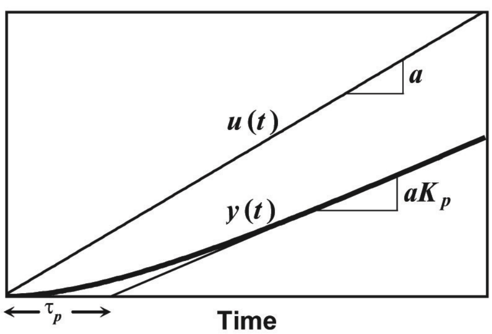


Figure 6.3.5 Response of a first-order process to a ramp input.

Response to a Ramp Input. Figure 6.3.5 shows the response of a first-order process to a ramp input. Assuming that $y(t)$ is initially equal to zero, the analytical solution of Equation 6.3.1 for an ramp input with a ramp slope of a is given by

$$y(t) = K_p \tau_p a(e^{-t/\tau_p} - 1) + K_p a t \qquad \textbf{6.3.6}$$

where K_p and τ_p are the parameters for a first-order process. Note that for this equation at large times,

$$y(t) \approx aK_p(t - \tau_p)$$

After the first-order dynamics have dissipated, $y(t)$ becomes a ramp with a slope equal to aK_p. The projection of the large time slope of $y(t)$ to y=0, yields $_p$, which is shown in Figure 6.3.5 and can be derived from the previous equation.

Example 6.5 Analysis of the Response of a First-Order Process to a Ramp Input

Problem Statement. Determine the first-order model parameters for a process based on its response to a ramp input (Figure 6.3.6). Assume that the ramp input was $u(t)$=0.5t, where t is in minutes.

Solution. From Figure 6.3.6, the slope of $y(t)$ after the initial transients is equal to 20. From Equation 6.3.6, the long time slope is equal to aK_p. Therefore, K_p is equal to 40 because a in this case is equal to 0.5. By extrapolation of the steady-state slope to $y(t)$=0, the time constant for the process is equal to 7 min. The transfer function for this process is

$$G_p(s) \quad \frac{40}{7s \quad 1}$$

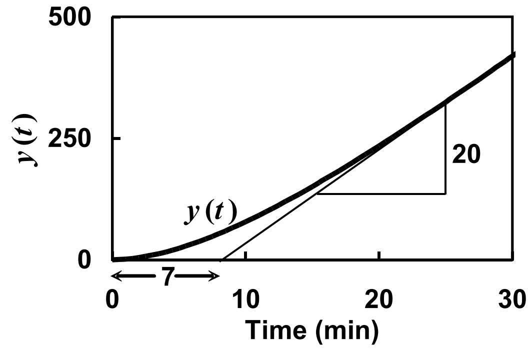


Figure 6.3.6 Response of a process to a ramp input (Example 6.5).

Self-Assessment Questions

Q6.3.1 What are the characteristics of a first-order process subjected to a step input change?

Q6.3.2 What does K_p represent?

Q6.3.3 What does the time constant for a process represent?

Self-Assessment Answers

Q6.3.1 A first-order process subjected to a step input change will monotonically approach a new steady-state condition. For a pure first order process, as soon as the step change is implemented, the process starts moving toward the new steady-state condition. After one time constant of time, 63% of the ultimate change occurs. After three time constants of time, 95% of the change occurs. After four time constants of time, 98% of the change occurs.

Q6.3.2 K_p represents the steady-state process gain, i.e., the $y/\ u$ where y is the resulting steady-state change in the output variable for an input change u.

Q6.3.3 The time constant of a process represents how fast the process reacts to input changes. In particular, it represents how long it takes to reach 63.2% of the ultimate change after a step input has been applied. As an approximation, four time constants is approximately equal to the response time for a process.

Self-Assessment Problem

P6.3.1 Consider the following first-order transfer function for a process:

$$G_p(s) \quad \frac{10}{15s \quad 5}$$

Determine the process gain and the time constant for this process from its transfer function.

Self-Assessment Answer

P6.3.1 Dividing the numerator and denominator of the transfer function by 5 yields

$$G(s) = \frac{10}{15s+5} = \frac{2}{3s+1}$$

Because the transfer function is in the standard form, the process gain (2) and the process time constant (3) can be determined directly by comparing this result with Equation 6.3.2.

6.4 Second-Order Processes

A series of two first-order processes or a first-order process with a PI feedback controller behaves as a **second-order process**. The differential equation for a second-order process written in the standard form is given by

$$\tau_n^2 \frac{d^2y(t)}{dt^2} + 2\zeta\tau_n \frac{dy(t)}{dt} + y(t) = K_p u(t) \qquad \textbf{6.4.1}$$

where $y(t)$ is the output variable, $u(t)$ is the input, K_p is the steady-state process gain, τ_n is the natural period and ζ is the **damping factor**, which determines the general shape of the dynamic response. The natural period is not strictly speaking the time constant for the process. It is, however, a measure of the speed of response of the process and is equal to the inverse of the frequency of the oscillations (radians/time), ω_n,

$$\tau_n = \frac{1}{\omega_n}$$

The transfer function for a second-order process is given by

$$G_p(s) = \frac{K_p}{\tau_n^2 s^2 + 2\zeta\tau_n s + 1} \qquad \textbf{6.4.2}$$

Note that the denominator of the transfer function of a second-order process contains s^2 as the highest power of s. Similar to a first-order process, either the second-order differential equation or the transfer function can be put into these standard forms to directly determine K_p, τ_n and ζ. To put the differential equation into the standard form corresponding to Equation 6.4.1, the coefficient of y is unity. Likewise, the standard form for the transfer function of a second-order process requires that the constant term in the denominator be unity.

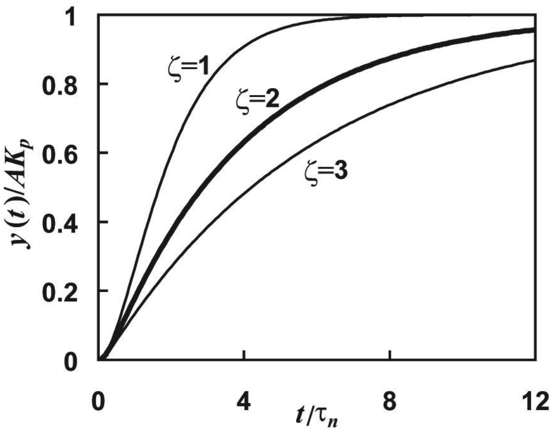


Figure 6.4.1 Dynamic response of an overdamped second-order process (ζ >1).

Response to a Step Input. Figure 6.4.1 shows the response of a second-order process to a step change, A, in the input for several cases for which $\zeta > 1$ (**overdamped** behavior) and $\zeta = 1$ (**critically damped**). Note that for each of these cases, $y(t)$ monotonically approaches the steady-state resting value after the step input change. Figure 6.4.2 shows the step response for $\zeta = 1$ (critically damped) and $\zeta < 1$ (**underdamped behavior**). When ζ is less than one, oscillatory behavior results. Note that a critically damped response ($\zeta = 1$) marks the boundary between overdamped and underdamped behavior. Figure 6.4.3 shows the step response for $\zeta=0$ (sustained

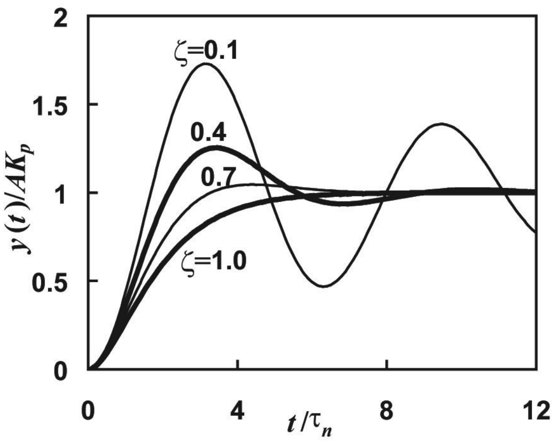


Figure 6.4.2 Dynamic response of an underdamped second-order process (0.1< <1.0).

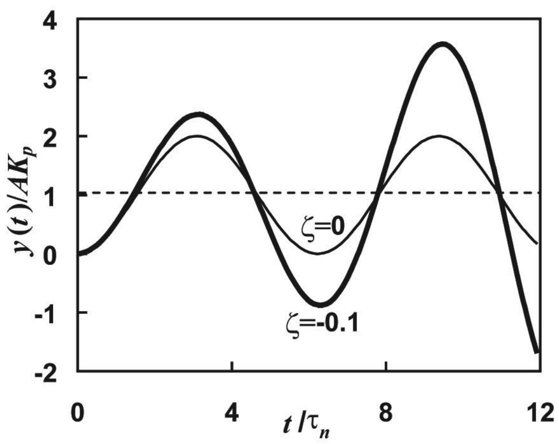


Figure 6.4.3 Dynamic response of an underdamped second-order process (=0; =-0.1).

oscillations) and for =-0.1 (exponentially growing oscillations). In general, the value of determines the shape of the dynamic behavior of a second-order process, $_n$ indicates the time scale of the response, and K_p indicates the steady-state sensitivity to input changes. In Chapter 5, it was shown that the poles of the transfer function determine the dynamic behavior of the process. In the case of a second-order process, real distinct roots correspond to an overdamped system, repeated real roots correspond to a critically damped system and complex roots correspond to underdamped behavior. Figure 6.4.4 shows a typical second-order underdamped response to a step input along with its key characteristics:

1. Rise time, t_{ris}, is the time required for $y(t)$ to first cross its new steady-state value and is given by the following analytical expression

$$t_{ris} \quad {}_n \frac{}{\sqrt{1 \quad {}^2}}$$

where $$\tan^{1} \frac{\sqrt{1 \quad {}^2}}{}$$

2. Overshoot, *B/D*, is the maximum amount by which the response exceeds the new steady-state resting value of *y*. The analytical expression for overshoot is

$$\text{Overshoot} \quad \frac{B}{D} \quad \exp \frac{}{\sqrt{1 \quad {}^2}} \qquad \textbf{6.4.3}$$

Therefore, you can estimate by measuring the overshoot and algebraically solving Equation 6.4.3 for .

3. Decay ratio, *C/B*, is the ratio of the height of successive peaks in the response. The analytical expression for decay ratio is

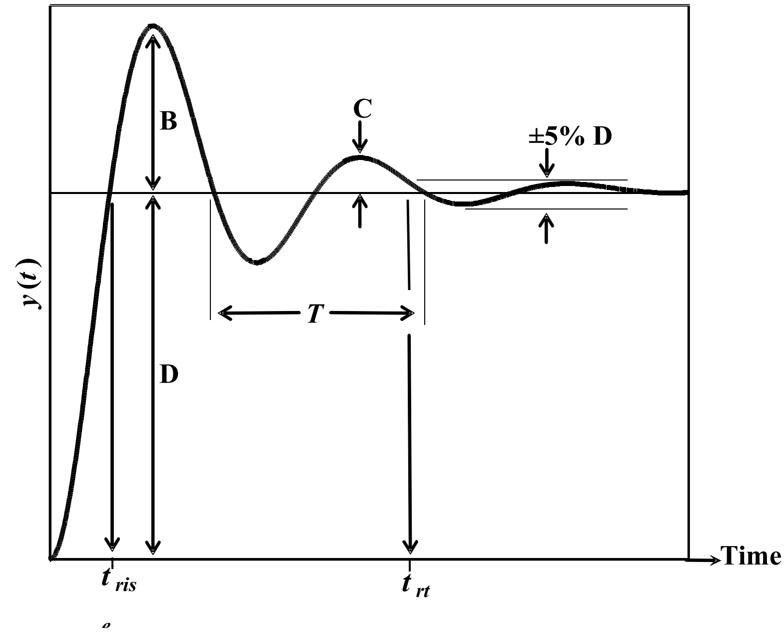


Figure 6.4.4 The key characteristics of an underdamped second-order response.

$$\text{Decay Ratio} \quad \frac{C}{B} \quad \exp \quad \frac{2}{\sqrt{1 \quad ^{2}}} \quad \text{Overshoot}^{2} \tag{6.4.4}$$

Again, you can estimate by measuring the decay ratio and algebraically solving Equation 6.4.4 for .

4. Period of oscillations, T, is the time for a complete cycle. The analytical expression for the period of oscillation is

$$T \quad \frac{2 \quad _{n}}{\sqrt{1 \quad ^{2}}} \tag{6.4.5}$$

5. Response time or **settling time**, t_{rt}, is the time required for the response to remain within a 5% band, based upon the steady-state change in y. That is, the ±5% band corresponds to $D \pm 0.05D$ for Figure 6.4.3.

The above expressions are strictly valid only for second-order processes. Decay ratio, overshoot, settling time, and damping factor ( ) can each be used as a basis for tuning. For example, a decay ratio of ¼ (i.e., quarter-amplitude damping) is a well-known tuning criterion. Selecting a damping factor specifies the overshoot and decay ratio for a second-order process (Equations 6.4.3 and 6.4.4). Tuning based on minimum response time can also be used.

Example 6.6 Characteristics of a Second-Order Process

Problem Statement. For the following second-order transfer function, calculate the process gain, the natural period and the damping factor. In addition, determine the overshoot and the decay ratio.

$$G_p(s) \qquad \frac{1}{2s^2 \quad 1.5s \quad 0.5}$$

Solution. Rearranging the second-order transfer function into the standard form yields

$$G_p(s) \qquad \frac{2}{4s^2 \quad 3s \quad 1}$$

Because the coefficient of s^2 is equal to $\;^2_n$, the natural period is equal to 2 time units. Also, because the coefficient of s is equal to 2 $\;_n$, the damping factor, , is equal to 0.75. Finally, the process gain is 2. Using Equation 6.4.3 with equal to 0.75, the overshoot is equal to 0.028; using Equation 6.4.4, the decay ratio is equal to 0.000805 or 1/1242.

Example 6.7 Determine the Decay Ratio and Damping Factor from the Overshoot

Problem Statement. Consider an underdamped second-order process that results in a 0.20 overshoot for a step input change. Determine the damping factor and the decay ratio for this second-order system.

Solution. Applying Equation 6.4.3 in the text results in

$$0.20 \qquad \exp \frac{}{\sqrt{1 \quad ^2}}$$

Taking the natural logarithm of both sides yields the following after rearranging

$$\sqrt{1 \quad ^2} \qquad \frac{}{1.6094}$$

Squaring both sides of the equation and solving for yields $= 0.4559$. Applying Equation 6.4.4, a decay ratio is the square of the overshoot, 0.04, or 1 / 25.

Example 6.8 Determine Second-Order Transfer Function Parameters from Process Information

Problem Statement. Consider a temperature control loop that exhibits second-order underdamped behavior. The temperature control loop is tuned for a 1/6 decay ratio and the period for oscillations is 20 minutes. The control loop is tuned properly so that, after the process reaches steady-state operation, there is no noticeable offset between the setpoint and the measured value of the temperature. Develop a second-order transfer function for this closed-loop system assuming that the input is the temperature setpoint and the output is the measured value of the process temperature.

Solution. Because there is no offset for this temperature controller, the process gain is one. The damping factor can be determined from Equation 6.4.4 using the specified decay ratio, and the natural period can be determined from Equation 6.4.5. Because Equation 6.4.5 depends on , Equation 6.4.4 should be solved first to determine the value for . Rearranging Equation 6.4.4 and using the specified decay ratio yields equal to 0.274. Rearranging Equation 6.4.5 and using the value of T and yields $_n$ equal to 3.06 minutes; therefore, the transfer function for this closed-loop process is

$$G_{CL}(s) \quad \frac{T_s(s)}{T_{sp}(s)} \quad \frac{1}{9.38s^2 \quad 1.69s \quad 1}$$

where time is in minutes and the gain is in deg/deg.

Response to a Ramp Input. The following equation is the analytical solution for a ramp input applied to an overdamped second-order process assuming $y(0)=0$.

$$y(t) \quad aK_p \quad t \quad (\ _1 \quad _2) \quad \frac{1}{_1 \quad _2}(\quad e^{t/ \ _1} \quad e^{t/ \ _2})$$

where $u(t)=at$ and the real poles of the second-order overdamped system are (-1/ $_1$, -1/ $_2$). Figure 6.4.5 shows the response of an overdamped second-order process to a ramp input. Note that the time required to settle to a response with a constant slope (aK_p) is longer than for a first-order process subject to a ramp input.

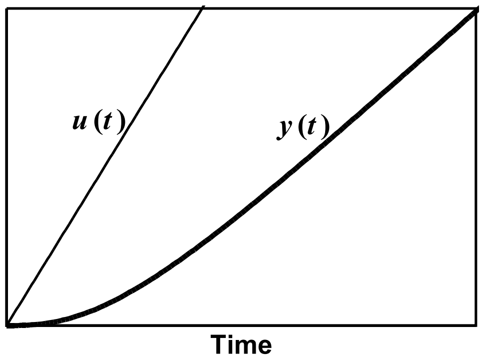


Figure 6.4.5 Response of an overdamped second-order process to a ramp input.

Self-Assessment Questions

Q6.4.1 What is the difference between an underdamped response with =0.1 and =0.5 ?

Q6.4.2 What is the difference between an overdamped response with =2 and =6?

Q6.4.3 Qualitatively describe what the decay ratio is.

Self-Assessment Answers

Q6.4.1 For a damping factor of 0.1, the rise time is faster, the overshoot is larger, the decay ratio is smaller and the settling time is larger compared to a damping factor of 0.5. Overall, the response for a damping factor of 0.1 tends to oscillate and very slowly damp out while the case for a damping factor of 0.5, the oscillations will damp out much faster.

Q6.4.2 The damping factor equal to 6 will be much more sluggish and will have a longer settling time and effective deadtime than for a damping factor equal to 2.

Q6.4.3 The decay ratio is the amount of the second overshoot C in Figure 6.4.4) divided by the first overshoot (B in Figure 6.4.4). Because the decay ratio is less than one, it is expressed as a "one to *x*" decay ratio. That is, instead of referring to a decay ratio as 0.2 it is referred to as a 1/5 decay ratio.

Self-Assessment Problem

P6.4.1 Consider following second order transfer function for a process:

$$G_p(s) \quad \frac{25}{1000s^2 \quad 100s \quad 10}$$

Determine the gain, the natural period and the damping factor for this process from its transfer function.

Self-Assessment Answer

P6.4.1 To convert this transfer function into the standard form, divide the numerator and denominator by 10

$$G(s)=\frac{25}{1000s^2+100s+10}=\frac{2.5}{100s^2+10s+1}=\frac{K_p}{\tau_n^2 s^2+2\zeta\tau_n s+1}$$

Therefore, the process gain is 2.5, the natural period is 10 and the damping factor is 0.5.

6.5 Integrating Processes

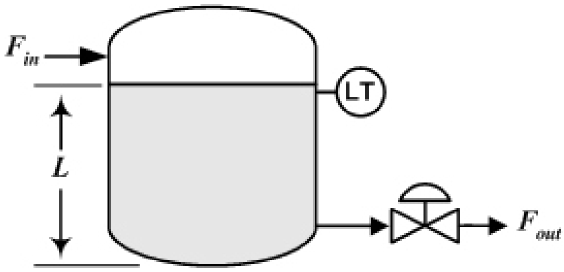


Figure 6.5.1 Schematic of a self-regulating level in a tank, which is not an integrating process.

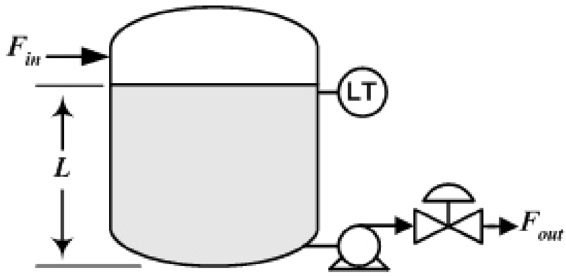


Figure 6.5.2 Schematic of a non-self-regulating level in a tank, which is an integrating process.

The most common type of **integrating process** is the level in a tank for which the outflow and inflow are set independently of the level. Figure 6.5.1 shows a non-integrating level process, which is a self-regulating process because the flow rate through the control valve is dependent on the level in the tank. Figure 6.5.2 shows an integrating process, which is a non-self-regulating process because F_{out} is independent of the level in the tank. Most levels in industry are non-self-regulating similar to Figure 6.5.2. The differential equation describing the dynamic behavior of a level in a tank is given by

$$A_c\rho\frac{dL}{dt}=F_{in}-F_{out} \qquad \textbf{6.5.1}$$

where A_c is the cross-sectional area of the tank, L is the height of liquid in the tank, ρ is the density of the feed and product, F_{in} is the mass flow rate into the tank and F_{out} is the mass flow rate out of the tank. Assuming the inflow, F_{in}, is constant and the outflow, F_{out}, is set independently of the level in the tank, the transfer function for this process is

$$G_p(s)=\frac{L(s)}{F_{out}(s)}=\frac{-1}{A_c\rho\, s} \qquad \textbf{6.5.2}$$

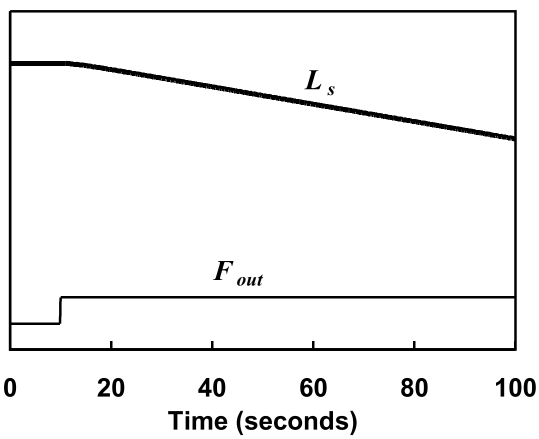


Figure 6.5.3 Dynamic response of an integrating process to a step input change.

The s factor in the denominator of the transfer function indicates that this process is integrating in behavior. Using the final-value theorem (Equation 4.2.4), you can easily determine that this process is non-self-regulating. Figure 6.5.3 shows the response of an integrating process (Equation 6.5.1) to a step increase in F_{out}.

Example 6.9 Developing a Transfer Function for an Integrating Process

Problem Statement. An open-loop test is applied to a tank level process. Initially, the level is constant at a value of 40%. The outflow from the tank is increased by 1000 lb/h. After 5 minutes the level reading is 30%. Develop a transfer function for this process for which the input is a change in F_{out} and the output is the change in the level in the tank.

Solution. For a tank level initially at steady state, a change in the outflow rate produces a constant slope similar to Figure 6.5.3. In this case, the slope is calculated as the change in the level per time or 2% min^{-1}, which corresponds to a flow rate change of 1000 lb/h; therefore, the transfer function is

$$G(s) = \frac{L(s)}{F_{out}(s)} = \frac{2\%}{\text{min}}\left|\frac{\text{h}}{1000\,\text{lb}}\right|\frac{1}{s} = \frac{0.002}{s}\ \frac{\%\ \text{h}}{\text{min - lb}}$$

Self-Assessment Question

Q6.5.1 How does the dynamic response of an integrating process differ from the dynamic response of a first-order or second-order process?

Self-Assessment Answer

Q6.5.1 First-order and second-order processes are self-regulating processes while an integrating process is a non-self-regulating process. Therefore, an input to a first-order or second-order process will cause the process to settle at a steady-state condition while it will cause an integrating process to continuously increase or decrease.

6.6 High-Order Processes

Staged separation devices, such as distillation and absorption columns, can be represented as a series of first-order processes. For example, for a distillation column, each tray can be considered a first-order process. Because the overall transfer function for a process composed of a number of transfer functions in series is the product of each individual transfer function (Equation 5.6.1), the transfer function for a series of first-order processes with equal time constants is given by

$$G_p(s) = \frac{K_p}{(\tau_p s + 1)^n} \qquad \textbf{6.6.1}$$

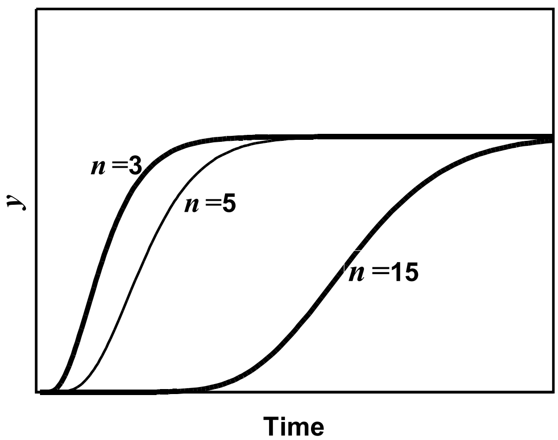


Figure 6.6.1 Dynamic response of three high-order systems (n=3, 5, and 15) to a step input change.

Because the largest power of s in the denominator is n, Equation 6.6.1 represents an nth-order process. Figure 6.6.1 shows the response to a step input for an nth-order process corresponding to Equation 6.6.1 for various values of n (i.e., n = 3, 5, 15). As n becomes larger, the response becomes more **sluggish**, i.e., the slope of the initial response becomes smaller and the response time is longer. For larger values of n, there is a period of time before a noticeable change in the output variable can be observed, and this period of time (deadtime) increases as n increases. A first-order plus deadtime (FOPDT) model (Section 6.8) can provide a good approximation of a high-order system, as shown later in this chapter. The response for n=3 is similar to the open-loop response of the CST thermal mixer, which is also a third-order linear process. This results because the actuator, process and sensor were each modeled as first-order processes although in the case of the CST thermal mixer, the time constants are not equal.

Self-Assessment Question

Q6.6.1 How does the dynamic response of a distillation column change as more trays are added to the column?

Self-Assessment Answer

Q6.6.1 As trays are added to a distillation column, this will increase the order of the process, resulting in more sluggish behavior with a large effective deadtime.

6.7 Deadtime

Deadtime or **transport delay** can result from plug flow transport through a pipe or from transport of solids by a conveyor belt. Figure 6.7.1 shows a CSTR with a product line attached. It is assumed that reaction occurs only in the reactor and the product flows by plug flow a length, L, at which point an on-line analyzer measures the product composition. Turbulent flow through a pipe is well represented as plug flow. The time, , that it takes the reaction mixture to flow by plug flow from the CSTR to the analyzer is

$$\frac{L A_c}{F} \qquad \textbf{6.7.1}$$

where A_c is the cross-sectional area of the product line, is the density of the fluid and F is the mass flow rate of the product through the product line. If the composition measurement is fast compared to , the measured composition, $C_s(t)$, is the reactor composition, $C(t)$, time units before, i.e.,

$$C_s(t) \quad C(t \quad) \qquad \textbf{6.7.2}$$

where is the deadtime or transport delay for this process. From Table 4.1, the transfer function for deadtime is

$$G_p(s) \quad e^{\quad s} \qquad \textbf{6.7.3}$$

Gas chromatographs (GCs) exhibit analyzer deadtime or delay. That is, the sample enters the GC and must flow through a separation column before the analysis is complete. The analyzer delay for a dedicated GC typically ranges between 3 and 10 minutes.

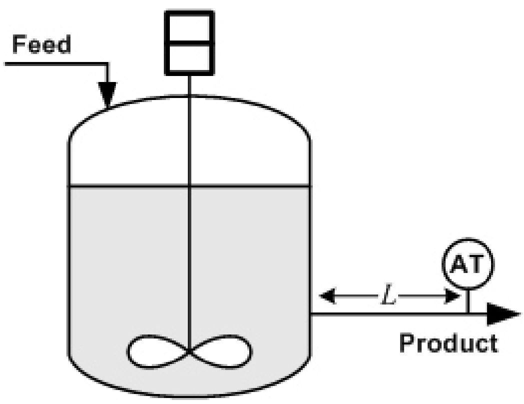


Figure 6.7.1 Schematic of a CSTR with transport delay.

Process deadtime and/or analyzer deadtime can have a significant effect on feedback controller tuning and control performance when the deadtime is significant compared to the time constant of the process. A five minute analyzer delay does not significantly affect feedback control performance for a large distillation column with a time constant of three hours for composition dynamics. On the other hand, a five minute analyzer delay dramatically affects a column with a time constant of five minutes for its composition dynamics.

Deadtime is usually combined with other models to take into account the effect of process and analyzer deadtime as well as the initial response of a highly overdamped process (e.g., a first-order plus deadtime model or an integrator plus deadtime model).

Self-Assessment Question

Q6.7.1 Why would significant analyzer deadtime reduce the performance of a feedback control loop?

Self-Assessment Answer

Q6.7.1 Analyzer delay affects the performance of a controller because a controller, which uses the analyzer reading to choose control action, must wait for a period of time equal to the analyzer delay before taking corrective action for measured deviations from setpoint.

6.8 First-Order Plus Deadtime (FOPDT) Model

A **FOPDT** model is the combination of a first-order model with deadtime:

$$G_p(s) = \frac{K_p e^{-\theta_p s}}{\tau_p s + 1} \qquad \textbf{6.8.1}$$

A step test can be conveniently used to develop a FOPDT model. Figure 6.8.1 shows one such approach. First, identify the resulting change in y (i.e., Δy) and the step change in the input, Δu. Then, from the step response, identify the time required for one-third of the total change in y to occur, $t_{1/3}$. Next, identify the time required for two-thirds of the total change in y to occur, $t_{2/3}$. Then the following formulas can be used to estimate the FOPDT parameters.

$$\begin{aligned} \tau_p &= \frac{t_{2/3} - t_{1/3}}{0.7} \\ \theta_p &= t_{1/3} - 0.4\,\tau_p \\ K_p &= \frac{\Delta y}{\Delta u} \end{aligned} \qquad \textbf{6.8.2}$$

Note that $t_{1/3}$ and $t_{2/3}$ are based upon assuming that time is equal to zero when the step change in u is implemented. This modeling approach is particularly well suited for modeling high-order processes, which makes it useful for representing industrial processes because most industrial processes exhibit high-order behavior. Figure 6.8.2 shows a fifth-order process and a FOPDT model that was selected to match this high-order process. There is

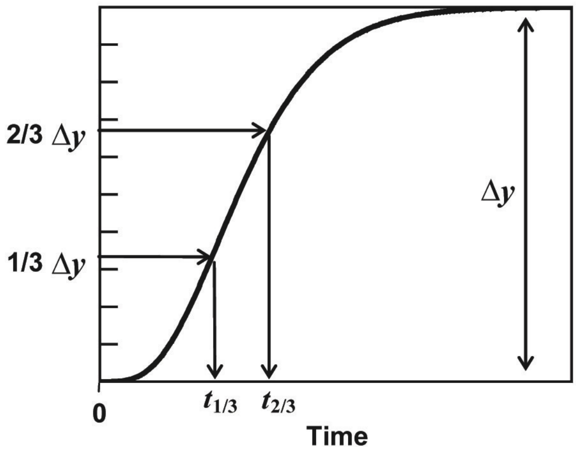


Figure 6.8.1 Graphical representation of an approach for determining the parameters of a FOPDT model.

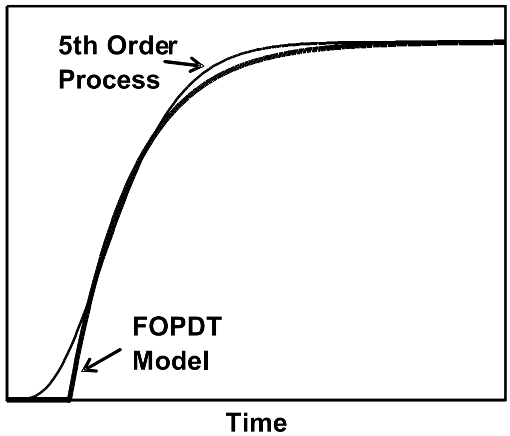


Figure 6.8.2 Comparison between a FOPDT model and an overdamped 5th order process.

slight mismatch initially between the FOPDT model and the high-order process model, but overall the FOPDT model provides a good approximation for overdamped process behavior.

Example 6.10 Estimation of FOPDT Parameters

Problem Statement. Estimate the FOPDT parameters from the following step test results.

Time (s)	Input (u)	Output (y)	Time (s)	Input (u)	Output (y)
0	1	5.5	7	2	4.5
1	1	5.5	8	2	4.3
2	1	5.5	9	2	4.2
3	2	5.5	10	2	4.1
4	2	5.5	11	2	4.0
5	2	5.4	12	2	4.0
6	2	5.0	13	2	4.0

Solution. The process gain is calculated from the steady-state change in y divided by the change in u, i.e.,

$$K_p \quad \frac{y(11) \quad y(3)}{u(11) \quad u(2)} \quad \frac{4.0 \quad 5.5}{2 \quad 1} \quad 1.5$$

Because $t_{1/3}$ and $t_{2/3}$ are based on when the input change is applied, time equal to zero for their calculation is t=3. Because the total change in y is -1.5, $y_{1/3}$ is 5.0 (i.e., 5.5-0.5) and $y_{2/3}$ is 4.5 (i.e., 5.5-1.0); therefore, from the open-loop response, $t_{1/3}$ is equal to 3 (i.e., 6-3) and $t_{2/3}$ is equal to 4 (i.e., 7-3). Applying the formula for ${}_p$

$${}_p \quad \frac{4 \quad 3}{0.7} \quad 1.43 \text{ s}$$

Then $${}_p \quad 3.0 \quad (0.4)\,(1.43) \quad 2.43 \text{ s}$$

The FOPDT transfer function is given by

$$G_p(s) \quad \frac{1.5\ e^{\ \ 2.43\,s}}{1.43\,s \quad 1}$$

Self-Assessment Question

Q6.8.1 Why are FOPDT models so effective for modeling industrial processes?

Self-Assessment Answer

Q6.8.1 Industrial process are usually high-order processes and higher-order processes are well represented by FOPDT models because a higher-order system has a response that is "S" shaped. An "S" shaped response has an initial response that is well represented by deadtime. Therefore, a FOPDT model captures the initial response with deadtime and the rapidly changing portion of the response with a first-order model. In addition, a FOPDT model only requires three model parameters.

6.9 Inverse-Acting Processes

An **inverse-acting process** can occur when opposing factors are acting within the process: one that is faster responding but with less steady-state gain than the other, e.g.,

$$G_p(s) \quad G_1(s) \quad G_2(s) \quad \frac{K_p}{{}_p s \quad 1} \quad \frac{K_p}{{}_p s \quad 1} \qquad \textbf{6.9.1}$$

where $|K_p| \quad |K_p|$ and ${}_p \qquad {}_p$. The response of an inverse-acting process to a step change input is shown in Figure 6.9.1. For a short time the second term [$y_2(t)$ in Figure 6.9.1] is controlling, causing an initial negative response but, as time proceeds, the first term [$y_1(t)$ in Figure 6.9.1], due to its larger steady-state gain, dominates the response causing a net positive response. Combining the terms in Equation 6.9.1 results in

$$G_p(s) \quad \frac{(K_p \ {}_p \quad K_p \ {}_p)s \quad (K_p \quad K_p)}{(\ {}_p s \quad 1)(\ {}_p s \quad 1)}$$

The zero of this transfer function is

$$s \quad \frac{K_p \quad K_p}{K_{p} \;_{p} \quad K_{p} \;_{p}}$$

When this equation is positive (i.e., a right-half-plane zero), inverse action results as shown in Section 5.5.

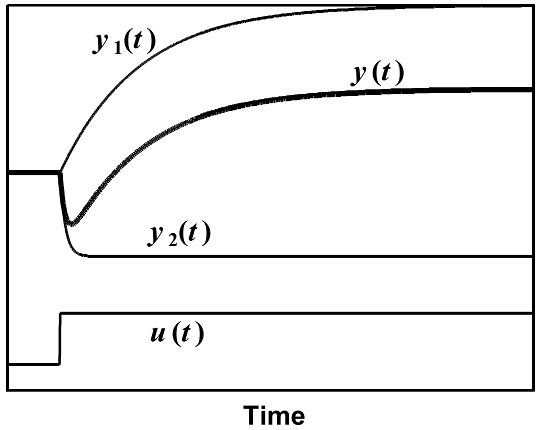


Figure 6.9.1 Dynamic response of an inverse-acting process to a step input change.

Consider a mercury-in-glass thermometer, initially at ambient temperature, that is submersed in hot water. The glass can be viewed as a container for the mercury. The height of the mercury column is used to measure temperature. After the thermometer is put into the hot water, the temperature of the glass around the mercury column increases before the temperature of the mercury. Because glass expands when heated, the inside **diameter** of the glass container increases slightly, causing the height of the mercury column to decrease slightly. Soon after this decrease in the height of the mercury column, the temperature of the mercury begins to rise, causing the overall height of the mercury column to increase rapidly. The effect of the expansion of the glass on the measured temperature is the low-gain, fast-responding behavior while the expansion of the mercury column due to a temperature increase is the high-gain, slower-responding behavior. Together, these two competing factors result in an inverse response.

Certain types of reboilers can exhibit inverse-acting behavior for a step change in the steam flow rate to the reboiler[1]. An increase in steam flow to the reboiler causes the number and volume of bubbles produced on the shell side of the reboiler to immediately increase. This "swell" effect can cause the measured level to show an initial increase after a steam rate increase. Of course, because the reboiler duty is increased, the vapor boilup rate increases, which eventually results in a decrease in the level in the reboiler. The swell is the low-gain, quick-responding behavior while the material balance effect of the increase in vapor rate from the reboiler is the high-gain, slow-responding behavior, which together cause an inverse response. When the heat addition rate to the reboiler is used to control the level in a reboiler that exhibits significant inverse action, a much more challenging control problem is encountered than is usually observed for conventional level control systems.

Example 6.11 Identifying Inverse Action from $G_p(s)$

Problem Statement. Determine whether the following transfer function exhibits inverse action.

$$G_p(s) \quad \frac{2s^2 \quad s \quad 1}{4s^3 \quad 3s^2 \quad 5s \quad 1}$$

Solution. This system representation exhibits inverse action if there are any right-half plane zeros. Therefore, setting the numerator of the transfer function equal to zero,

$$2s^2 \quad s \quad 1 \quad 0$$

Because the roots of this equation are -0.5 and 1.0 and a right-half plane zero is determined, inverse action is indicated.

Self-Assessment Question

Q6.9.1 Explain in your own words why the difference between two first-order models can produce an inverse response.

Self-Assessment Answer

Q6.9.1 Inverse action can result from the difference between two first-order models if one of the first-order models has a smaller process gain and a smaller process time constant. See Figure 6.9.1 as an example.

6.10 Lead-Lag Element

A system can have a lead-lag response if the system input involves a time derivative of the input, which is used for feedforward control (Chapter 13). Consider a first-order process subject to the following input

$$b_1 u(t) \quad b_2 \frac{du(t)}{dt}$$

Combining this input with the ODE for a first-order system (Equation 6.3.1)

$$_p \frac{dy(t)}{dt} \quad y(t) \quad K_p \quad b_1 u(t) \quad b_2 \frac{du(t)}{dt}$$

Applying Laplace transforms assuming steady-state initial conditions yields

$$_p sY(s) \quad Y(s) \quad K_p \quad b_1 U(s) \quad b_2 sU(s)$$

Rearranging and solving for $Y(s)/U(s)$ yields

$$G(s) \quad \frac{Y(s)}{U(s)} \quad \frac{K_p(b_1 \quad b_2 s)}{_p s \quad 1}$$

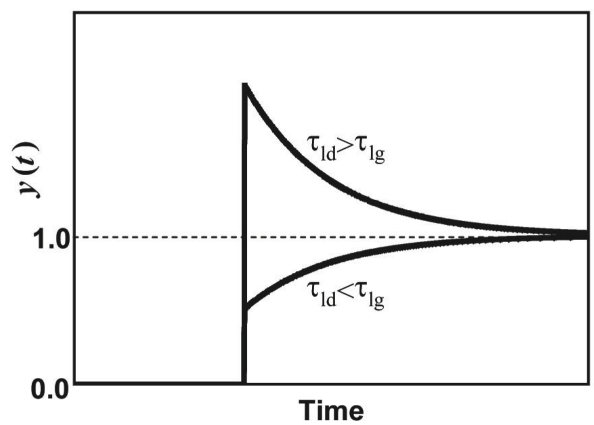


Figure 6.10.1 The effect of the ratio of $_{ld}$ to $_{lg}$ on the dynamic response of a lead-lag element. ($K_{\text{lead-lag}}$=1)

This equation can be rearranged into the general form for the transfer function for a lead-lag element

$$G_{\text{lead} \quad \text{lag}}(s) \quad K_{\text{lead} \quad \text{lag}} \quad \frac{_{ld} s \quad 1}{_{\lg} s \quad 1}$$

where $_{ld}$ is equal to b_2/b_1, $_{lg}$ is equal to $_p$, and $K_{lead\text{-}lag}$ is equal to $K_p b_1$ from the previous example. $_{ld}$ is referred to as the "lead" and $_{lg}$ is called the "lag". The time-domain behavior of a lead-lag element for a unit step input change is given by

$$y(t) \quad K_{\text{lead} \quad \text{lag}} \quad \frac{_{ld}}{_{\lg}} \quad 1 \quad e^{\,t/ \quad _{\lg}} \quad 1 \qquad \textbf{6.10.1}$$

This result is obtained by combining the transfer function for a lead-lag element with a unit step input (i.e., $U(s)=1/s$) and applying partial fraction expansions before it is converted to the time domain. Figure 6.10.1 shows the output of a lead-lag element for a unit step input change for $K_{\text{lead-lag}}$ equal one when $_{ld}$ is larger than $_{lg}$ and when $_{ld}$ is less than $_{lg}$. The dynamic behavior of a lead-lag element can be understood by examining the terms inside the bracket in Equation 6.10.1. When $_{ld}$ is greater than $_{lg}$, the terms in the bracket have a positive result, and initially $y(t)$ is larger than one, but monotonically approaches one. If $_{ld}$ is less than $_{lg}$, the terms inside the bracket have a negative result, and the initial response is less than one, but monotonically approaches one. Lead-lag elements are used to provide dynamic compensation for feedforward control (Chapter 12) and decouplers (Chapter 15).

Self-Assessment Question

Q6.10.1 What is the effect of a lead-lag element when the lead has a larger value than the lag?

Self-Assessment Answer

Q6.10.1 When the lead is larger than the lag, the output of a lead-lag element will overshoot the resulting steady-state response. When the lag is larger than the lead, the output moves to the new steady-state level without overshoot.

6.11 Recycle Processes

Recycle processes are used industrially to improve the economic performance of a process based on a steady-state analysis. Material recycle is used to recover unreacted reactants and recycle them to the reactor as feed to increase overall conversion, and energy recycle or heat recovery is used to recycle heat within a process and thus reduce overall energy usage. Even though processes with recycle can have significant economic advantages compared to the corresponding processes without recycle, the process control problems associated with recycle processes are usually much more challenging. Recycling (also termed **process integration**) can have a dominant effect on the overall process dynamics and resulting control performance. That is, a process with recycle has similar fast dynamics to the corresponding process without recycling, and it has the slow dynamics associated with the recycle stream. For example, consider the recycle system shown in Figure 6.11.1. A change in an input to the reactor will have an immediate effect on the temperature and composition of the product stream leaving the reactor, but it will also affect the reactor through the recycle which is a much slower effect.

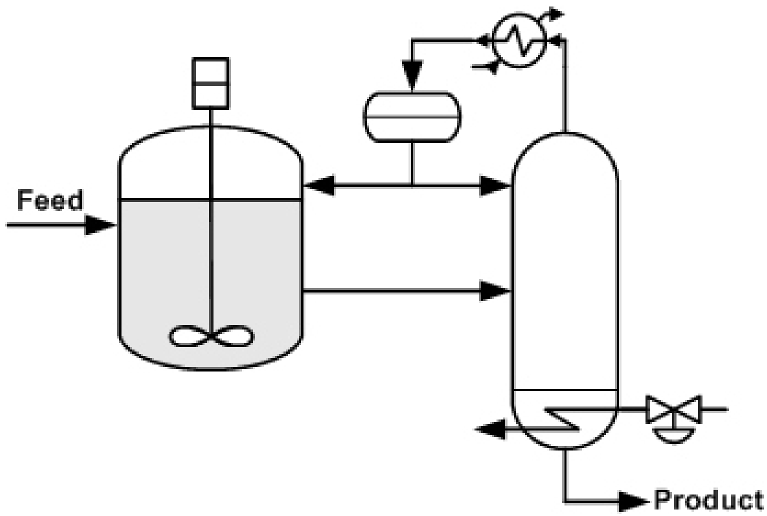


Figure 6.11.1 Schematic of a reactor/stripper process with material recycle.

Material Recycle. Due to economic driving forces that require operating companies to produce as much product as possible from every pound of feed, material recycle is quite common in the CPI. Figure 6.11.1 shows a simple reactor/stripper recycle system. For certain systems it can be very important to operate at low conversion levels in the reactor to minimize the production of waste products. That is, the yield for the undesirable by-product increases with the conversion in the reactor; therefore, it is economically preferred to operate at low conversions. As a result, it is essential to recover the unreacted feed and recycle it back to the reactor. For the configuration shown in Figure 6.11.1, the reactor may operate at a relatively low conversion per pass while, for the overall process with recycle, the conversion can approach 100%. Material recycle can also increase the process gain and the time constant of the overall response and, therefore, complicate the application of process control.

Energy Recycle. Energy recycle or energy recovery is based on recovering thermal energy from streams that otherwise would remove thermal energy from the process. Waste heat recovery is a common example of energy recycle. For example, in an ethylene plant, the high-temperature product gases from the cracking furnace are cooled by passing them through a spray quench column before the products are recovered from the product gas stream by distillation. If the quench water (i.e., the water circulated through the quench column) is cooled by discharging its excess thermal energy to the cooling water system, the thermal energy removed from the product gas stream is lost by the process. On the other hand, if the quench water is operated as a closed-system and used to provide reboiler duty for some of the distillation columns in the distillation train, some of the heat removed from the product gas stream is recycled back to the process reducing the demand for process steam and reducing the load on the cooling water system.

Another example of energy recycle is thermally-coupled distillation columns. In this case, one column is operated at a significantly higher pressure so the overhead condenser duty of the high pressure column can be used to provide reboiler duty for the lower pressure column. The following example analyzes the effect of using a hot product stream to preheat the feed to a process.

Example 6.12 Analysis of an Energy Recycle Case

Problem Statement. Analyze the dynamics of the CSTR with heat integration shown in Figure 6.11.2.

Solution. This process is an exothermic CSTR for which the product stream is used to preheat the feed to the reactor[2]. The temperature of the feed entering the reactor is T_f, which is given by

$$T_f(s) \quad T_o(s) \quad T_f(s) \qquad \textbf{6.11.1}$$

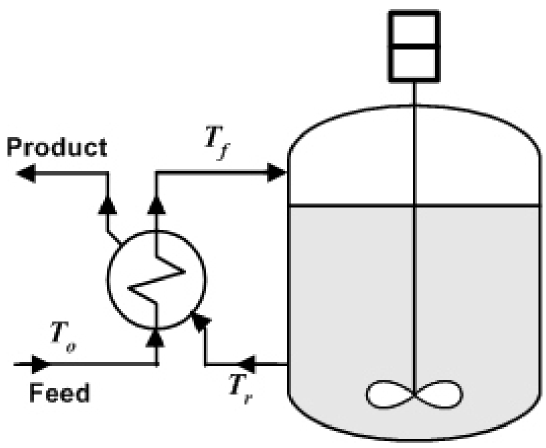


Figure 6.11.2 Schematic of an exothermic CSTR with heat integration.

where $T_f(s)$ represents the change in the temperature of the reactor feed provided by the heat exchanger. Note that because transfer functions are used, each term in Equation 6.11.1 is written in deviation variable form. Heat transfer to the feed depends on the temperature of the reactor, T_r. Using a transfer function model of the heat exchanger yields

$$T_f(s) \quad T_r(s)\, G_H(s) \qquad \textbf{6.11.2}$$

Likewise, the temperature of the reactor, T_r, is a function of the reactor feed temperature, T_f. This relationship can be expressed using a transfer function, $G_R(s)$, as

$$T_r(s) \quad T_f(s)\,G_R(s) \qquad \textbf{6.11.3}$$

Substituting Equation 6.11.1 into Equation 6.11.3 results in

$$T_r(s) \quad G_R(s)\,[T_o(s) \quad T_f(s)] \qquad \textbf{6.11.4}$$

Then Equation 6.11.2 is used to eliminate $T_f(s)$ from Equation 6.11.4.

$$T_r(s) \quad G_R(s)\,[T_o(s) \quad G_H(s)\,T_r(s)]$$

The overall transfer function, $G_{overall}(s)$, for this process becomes

$$G_{overall}(s) \quad \frac{T_r(s)}{T_o(s)} \quad \frac{G_R(s)}{1 \quad G_R(s)\,G_H(s)}$$

Assuming the following forms for $G_R(s)$ and $G_H(s)$

$$G_R(s) \quad \frac{K_R}{{}_R\, s \quad 1}$$

$$G_H(s) \quad K_H$$

results in the following after rearranging into the standard form for a transfer function for a first-order process

$$G_{overall}(s) \quad \frac{\dfrac{K_R}{1 \quad K_H K_R}}{\dfrac{{}_R}{1 \quad K_H K_R}\, s \quad 1}$$

Assuming the following numerical values

$$K_H \quad 0.5$$

$$K_R \quad 1.9$$

$${}_R \quad 1.0$$

yields

$$G_p(s) \quad \frac{1.9}{s \quad 1} \qquad \text{(without heat integration)}$$

$$G_{overall}(s) \quad \frac{38}{20\,s \quad 1} \qquad \text{(with heat integration)}$$

The gain and the time constant have both increased by a factor of 20 for the system with recycle. This is an extreme example, but you can easily see that material recycle and/or heat integration increases the process gain while significantly slowing the overall process response.

Self-Assessment Question

Q6.11.1 When energy recovery or material recycle is applied to a process, what happens to the dynamic response and the process gain of the resulting system?

Self Assessment Answer

Q6.11.1 When energy recovery or material recycle is applied to a process, the time constant and the gain for the overall process usually increase significantly.

6.12 Simulink

MATLAB's Simulink is a convenient software package for generating the time domain response from a transfer function based on a graphical programming environment. Using standardized functional blocks, Simulink combines inputs with transfer functions using a block diagram format to produce time-domain responses that are graphically displayed.

You can start Simulink by opening MATLAB and clicking on "Simulink Library" button, selecting a block, right-click on the block and select "Add block to new model". You can also start Simulink by opening a Simulink model (a file with an "slx" extension). The Simulink Library Browser contains the elements that you can use to build a Simulink model, which are discussed below.

In order to define the input for a Simulink model, select "Simulink" under the Simulink Library Browser and select "Sources". Sources contains a wide range of input functions including a step change, a sinusoidal input and a ramp. By right-clicking on an input block, you can add it to your model or start a new model. Once the block has been added to your model, you can double click on it and modify it attributes. For example, for a step input change, you can define the step time, the initial value, the final value and the sample time.

Transfer functions can be defined for a Simulink model by using the block "Transfer Fcn" which is located in "Continuous" under "Simulink" for the Simulink Library Browser. A transfer function is defined as the ratio of two polynomial in s where the order of the polynomial in the numerator is less than or equal to the order of the polynomial in the denominator. The polynomials are defined by the coefficients of the polynomials. For example, the polynomial s^2-4s+5 would be represented by the vector [1 -4 5]. In addition, the polynomial s3+s+1 is represented by the vector [1 0 1 1]. You can also add deadtime to a model by adding the "Transport Delay" block to the model.

The time-domain results can be displayed by using a "Scope" block, which is located in "Sinks" under "Simulink" for the Simulink Library Browser. The Scope block can plot up to two variables by using a "Mux" block, which is a multiplexer, and is located in the "Most Commonly Used" section of the Simulink Library Browser. A Mux block is indicated by a vertical thick black line with two inputs and one output and is used to deliver the two variables that are to be plotted to the Scope block.

These various blocks can be directly connected to develop a Simulink model similar to the a block diagrams for transfer functions (Section 5.6). Assume that you have arranged the various block functions in the desired order. Then to provide the linkage between two blocks, simply click on the outlet of a block and drag it to the inlet to the next block and a solid black line with an arrow will appear indicating the linkage. If you want to branch from one of these connecting lines, depress the control button on your keyboard and move the cursor to the desired location on the line and click and drag this line to it desired location. In addition, you can use a "Sum" block, which is denoted as a circle with two inputs and one output and can be found in the "Most Commonly Used" section of the Simulink Library Browser, to add or subtract signals. Sum blocks can also be used as part of a Simulink model to simulate feedback control loops as shown in Section 9.13.

As the following examples will demonstrate, Simulink is a convenient means of generating the time-domain solutions of a variety of idealized dynamic models. Moreover, Simulink can be used to explore the effect of parameters of idealized dynamic models.

Example 6.13 Simulink Solution for FOPDT Model

Problem Statement. Develop a Simulink model for a first-order plus deadtime model (K_p=1; $_p$=2; $_p$=2) subject to a 4 unit step change at t=1 time units and apply it to determine the time-domain response**.**

Solution. Figure 6.12.1 shows the Simulink model for this example. Note that the step size and the time delay are entered into the respective blocks and are not shown on the Simulink schematic for the model. The step input is connected to the Mux block so that it can be displayed by the Scope. Figure 6.12.2 shows the results for the Simulink model provided by the Scope block.

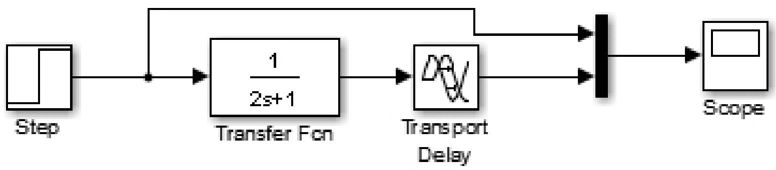


Figure 6.12.1 Simulink model for Example 6.13.

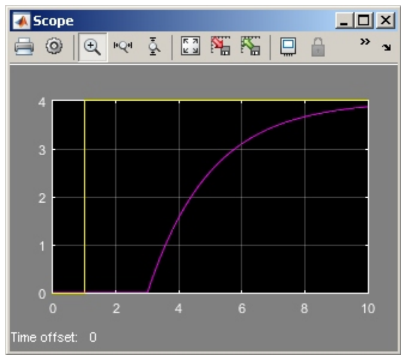


Figure 6.12.2 Simulink Results

Example 6.14 Simulink Solution for an Inverse-Acting Model

Problem Statement. Develop a Simulink model for an inverse-acting model based on two first-order responses (K_1=2; $_1$=2; K_2=-1; $_2$=0.5) subject to a 4 unit step change at t=1 time units and apply it to determine the time-domain response**.**

Solution. Figure 6.12.3 shows the Simulink model for this example. Note that a Sum block is used to combine the two first-order response to produce an inverse response. The Simulink results for this inverse-acting model to a step input change are shown in Figure 6.12.4.

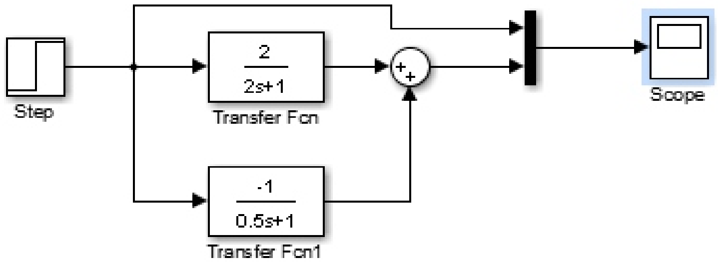


Figure 6.12.3 Simulink model for Example 6.12.2.

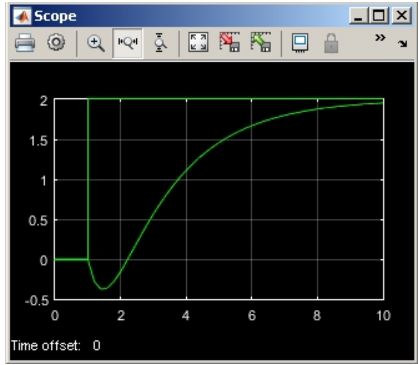


Figure 6.12.4 Simulink Results

Example 6.15 Simulink Model for the CST Thermal Mixer

Problem Statement. Develop a Simulink model for the CST mixing process described in Example 3.3 and apply a 10% increase in the MV and predict the behavior of the measured temperature of the outlet stream.

Solution. Because the actuator and sensor for this system are represented by a first-order model, these models can be applied directly using Simulink. On the other hand, the model for the mixing process itself is nonlinear. Therefore, we will linearize it to develop a first-order model using the procedure described in Section 5.7. In fact, the results from Example 5.11 can be used directly for this problem:

$$G_p(s) = \frac{T(s)}{F_1(s)} = \frac{\dfrac{T_1 - \bar{T}}{M}}{s + \dfrac{\bar{F}_1 + F_2}{M}} = \frac{\dfrac{T_1 - \bar{T}}{\bar{F}_1 + F_2}}{\dfrac{M}{\bar{F}_1 + F_2}\, s + 1}$$

Applying the numerical values from Example 3.3 (i.e., T_1=25°C; $\bar{T}$ =50°C; $\bar{F}_1$=5 kg/s; F_2=5 kg/s; M=100 kg),

$$G_p(s) = \frac{2.5}{10s + 1}$$

The Simulink model for this case is shown in Figure 6.12.5 and the results for a 10% increase in the MV (i.e., F_1 increased to 5.5 kg/s) are shown in Figure 6.12.6. For the step input block, the step time was set equal to 10, the initial value was set equal to zero and the final value was set equal to 0.5. By comparing the results of the nonlinear model (Figure 3.7.5), the results shown in Figure 6.12.6, which are based on a linear analysis, are quite close to the nonlinear results. In fact, the nonlinear model had a 1.20°C steady-state temperature change while the Simulink model had a 1.25°C temperature change because the CST thermal mixer model is only mildly nonlinear.

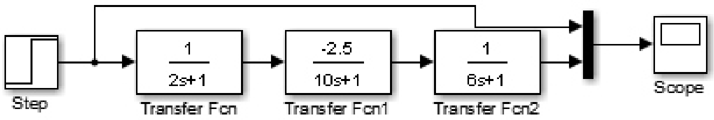


Figure 6.12.5 Simulink model for Example 6.15

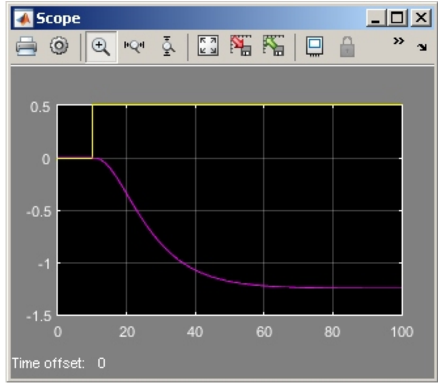


Figure 6.12.6 Results for Ex6.15.

6.13 Summary

- The dynamic behavior of real processes can be described using idealized dynamic models.
- The transfer function for a first-order process is given by

$$G_p(s) = \frac{K_p}{\tau_p s + 1}$$

where K_p is the steady-state process gain and τ_p is the process time constant.

- The transfer function for a second-order process is given by

$$G_p(s) = \frac{K_p}{\tau_n^2 s^2 + 2\zeta\tau_n s + 1}$$

where K_p is the steady-state process gain, τ_n is the natural period and ζ is the damping factor.

- The damping factor, ζ, determines the type of dynamic response of a second-order system. For $\zeta > 1$, an overdamped response results. For $\zeta=1$, a critically damped response occurs. For $1 > \zeta > 0$, an underdamped response results. For $\zeta=0$, sustained oscillations result. For $\zeta<0$, unstable oscillations occur.
- For a second-order underdamped system, the overshoot, decay ratio and response time can be used to describe the dynamic response of a system.
- The transfer function for an integrating system is given by

$$G(s) = \frac{K_p}{s}$$

- A high-order process, which describes most industrial processes, exhibits sluggish behavior and has a characteristic "S" shaped dynamic response for a step input change.
- The transfer function for a FOPDT model is given as

$$G_p(s) \qquad \frac{K_p\, e^{\quad _p s}}{{}_p s \quad 1}$$

where $_p$ is the effective deadtime.

- Inverse-acting responses occur as the result of competing factors: one that is fast acting with a relatively small process gain and another that is slow responding with a larger process gain.
- The transfer function for a lead-lag element is given by

$$G_{\text{lead} \quad \text{lag}}(s) \quad K_{\text{lead} \quad \text{lag}} \frac{{}_{ld}\, s \quad 1}{{}_{\text{lg}}\, s \quad 1}$$

where $K_{lead\text{-}lag}$ is the gain of the lead-lag element, $_{ld}$ is the lead and $_{lg}$ is the lag.

- Process integration (material and energy recycle) provide significant steady-state economic advantages, but increase the difficulty of the resulting control problem.
- Simulink is a convenient program for developing time-domain solutions for idealized dynamic models.

6.14 References

1. Munsif, H.P. and Riggs, J.B. “An Analysis of Inverse Acting Column Levels”, *Ind & Eng Chem Res,* Vol. 35, p. 2640 (1996).

2. Marlin, T.E., *Process Control*, McGraw-Hill, pp. 173-176 (1995).

6.15 Additional Terminology

Critically damped - which corresponds to the transition point between overdamped and underdamped behavior.

Damping factor ( ) - the characteristic of a second-order process that determines the general shape of the dynamic response.

Deadtime - the time delay associated with a composition analyzer or the delay a process experiences after input changes.

Decay ratio - the ratio of the successive peaks for a second-order underdamped response.

First-order process - a process with a transfer function that has unity as the highest power of s in the denominator.

FOPDT model - A First-Order Plus DeadTime model. A model combining a first-order process with deadtime.

Impulse input - an input with an infinite height and an infinitesimal duration.

Integrating process - a process that accumulates or depletes mass or energy.

Inverse-acting process - a process that has two competing factors: one that is faster responding but with less steady-state gain than the other.

K_p - the steady-state process gain ($y/\ u$).

Overdamped process - a second- or higher-order process which does not exhibit oscillatory behavior.
Overshoot - the magnitude of the overshoot divided by the steady-state change.
Period of oscillation - the time required for a complete cycle for an underdamped response.
Process integration - a term for using mass and energy recycle to make a process more economically efficient.
Ramp input - an input that increases or decreases at a constant rate.
Rectangular pulse - a step increase followed after a time by a step decrease that returns the input to its original value.
Recycle process - a process that recovers mass or energy from a process stream before it leaves the process.
Response time - the time after an input change for the process to settle to within 5% of the steady-state change.
Rise time - the time after a step input change for an underdamped response to cross the ultimate steady-state condition for the first time.
Second-order process - a process with a transfer function that has two as the highest power of s in the denominator.
Settling time - the time after a step change for the process to settle to within of the steady-state change.
Sinusoidal input - a process input that is sinusoidal.
Sluggish - behavior of a process for which the output variable is slow to respond to input changes.
Step change - a sudden and sustained change.
Transport delay - the time for material to move from one point in the process to another.
Underdamped process - a second- or higher-order process, which exhibits oscillatory behavior.
$_n$ **-** the natural period of a second-order response.
$_p$ **-** process time constant, i.e., determines the speed of dynamic response.
$_p$ **-** process deadtime.

6.16 Preliminary Questions

6.2 Idealized Process Inputs

Q6.2.1 Why are rectangular pulse inputs more feasible to apply to industrial processes than impulse inputs?

Q6.2.2 What is the parameter for a ramp input?

Q6.2.3 What are the two parameters for a sinusoidal input?

6.3 First-Order Processes

Q6.3.1 For a first-order process, how many time constants are required after an input change to observe 95% of the total change?

Q6.3.2 How can you estimate the time constant for a first-order process from its response to an impulse input?

Q6.3.3 Consider a ramp input $u(t)=at$ applied to a first-order process. What is the slope of $y(t)$ at large times?

6.4 Second-Order Processes

Q6.4.1 When the damping factor of a second-order process is negative, what type of dynamic behavior do you expect?

Q6.4.2 What is the decay ratio of a second-order process?

Q6.4.3 What does the value of the natural period indicate about the response of a second-order process?

Q6.4.4 What are the units of the gain of a second-order process? How does it compare to the gain of a first-order process?

Q6.4.5 Qualitatively describe what the rise time is. Does an overdamped second-order process have a rise time? Why or why not?

6.5 Integrating Processes

Q6.5.1 What is the most common example of an integrating process in the CPI and bio-tech industries?

6.6 High-Order Processes

6.6.1 What characteristic behavior does a high-order process exhibit?

6.7 Deadtime

6.7.1 What is the difference between analyzer delay and transport delay? How are they similar?

6.7.2 In general terms, describe why analyzer deadtime can affect the performance of a controller.

6.8 First-Order Plus Deadtime Model

Q6.8.1 Is the process gain, K_p, for a FOPDT model different from the K_p for a first-order process model or the K_p for a second-order process model?

Q6.8.2 Why are high-order processes well represented by FOPDT models?

6.9 Inverse-Acting Process

Q6.9.1 What causes inverse action?

Q6.9.2 How can a thermometer exhibit inverse action?

6.11 Recycle Processes

Q6.11.1 What is energy recycle?

6.17 Analytical Questions and Exercises

6.2 Idealized Process Inputs

P6.2.1** Consider each of the input profiles in Figure P6.2.1. Develop an expression for $U(s)$ for each case.

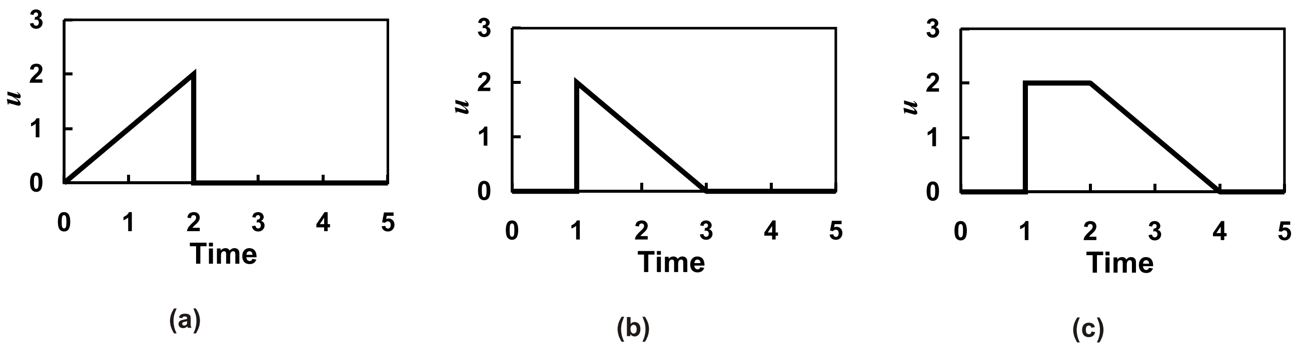


Figure P6.2.1 Input profiles.

P6.2.2** Block sine waves (Figure P6.2.2) are used to determine the deadband and time constant of an actuator (Chapter 10). Express $U(s)$ for the block sine wave shown in Figure P6.2.2.

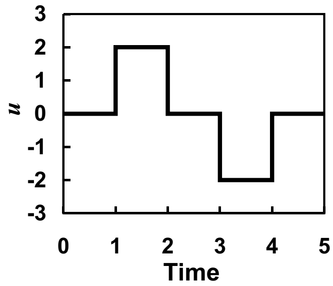


Figure P6.2.2 Input profile for a block sine wave.

6.3 First-Order Processes

P6.3.1* Consider the following first-order transfer functions. What characteristics of the process can you identify from these transfer functions

a. $G_p(s) \quad \frac{16}{456s \quad 100}$ b. $G_p(s) \quad \frac{1}{s \quad 0.1}$

c. $G_p(s) \quad \frac{300}{600s \quad 10}$ d. $G_p(s) \quad \frac{0.1}{3s \quad 0.01}$

P6.3.2** A lumped parameter model of a thermocouple (Figure P6.3.2) yields the following differential equation

$$M\,C_p \frac{dT_s}{dt} \quad A\,h\,(T_p \quad T_s)$$

where M is the mass of the thermocouple per unit length (1 g/cm), C_p is the heat capacity of the thermocouple (0.1 cal/g ºC) , T_s is the temperature of the thermocouple, A is the surface area per unit length (4 cm^2/cm), h is the heat transfer coefficient between the thermocouple and the process fluid (25 cal/cm^2-h-ºC), and T_p is the temperature of the process fluid. Calculate the time constant, in seconds, for this thermocouple system based on this lumped parameter model.

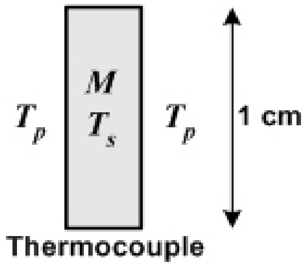


Figure P6.3.2 Schematic of a 1 cm section of a thermocouple.

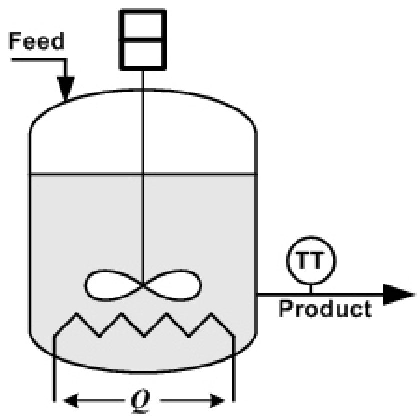


Figure P6.3.3 PFD for a CST heater.

P6.3.3** Consider the stirred-tank heater shown in the Figure P6.3.3. The process model for this system is

$$C_p V \frac{dT}{dt} \quad C_p F (T_{in} \quad T) \quad Q$$

where the liquid density has a specific gravity of 0.9, the heat capacity (C_p) is equal to 0.8 cal/g-ºC, F is 100 kg/min, T_{in} is 25ºC, V is 200 l and initially T is 75ºC. Determine the time constant and gain for this process.

P6.3.4** Determine the process gain and the time constant for the dissolved oxygen process shown in Figure P6.3.4 based on the following process model and initial conditions assuming that the air flow rate (F_{air}) is the input variable.

$$\frac{dC_{O_2}}{dt} \quad k_L a \; C^*_{O_2} \quad C_{O_2} \qquad K_{O_2} \quad {}_{\max} x$$

$$k_L a \quad 0.25 \quad 0.001(F_{air} \quad 500)$$

- C_{O_2} - concentration of O_2 in the reaction broth (initially 1.1×10^{-4} g-moles/l)
- $C^*_{O_2}$ - saturated concentration of O_2 in the broth (2.20×10^{-4} g-moles/l)
- K_{O_2} - cellular uptake of O_2 (1.98 g-moles O_2/g-cells)
- $k_L a$ - the overall liquid phase mass transfer coefficient for transport from the bubble surface to the bulk broth (initially 0.25 s^{-1})

- T - broth temperature (35ºC)
- x - constant cell concentration in the bioreactor (0.25 g/l)
- $_{max}$ - maximum specific growth rate (5.56×10^{-5} s^{-1})
- F_{air} - air feed rate (initially 500 cfm)

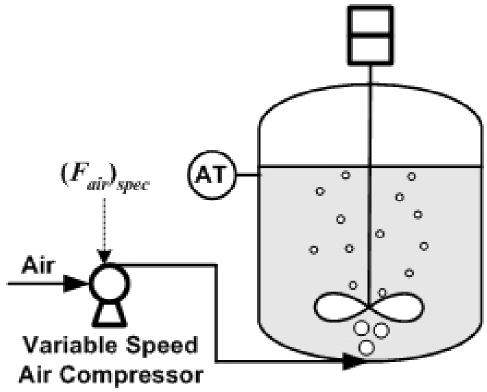


Figure P6.3.4 Schematic of a DO process.

P6.3.5* By observing a process, an operator indicates that an increase of 5,000 lb/h of steam (input) to the reboiler on a distillation column produces a 3% (absolute composition percentage) decrease in the impurity level in the bottoms product (output). When a change in the steam flow rate is made, it takes approximately 120 minutes for the full effect on the product composition to be observed. Using this process information, develop a first-order model for this process.

P6.3.6* By observing a process, an operator indicates that an increase of 2,000 lb/h in cooling water flow rate (input) to a heat exchanger produces a 10ºF decrease in the temperature of the process stream leaving the heat exchanger (output). In addition, when a change in the cooling water flow rate is made, it takes 18 minutes for the full effect on the outlet temperature of the process fluid to be observed. Using this process information, develop a first-order model for this process.

P6.3.7** Consider the continuous bio-reactor shown in Figure P6.3.7. The model for the cell concentration (Example 3.9) is given by

$$\frac{dx}{dt} \quad x \quad \frac{F_V x}{V}$$

Based on the following conditions, develop a transfer function for this process assuming that F_V is the input variable and determine how long it should take for the cell concentration to reach one-half of it initial value if F_V is increased to 50 l/h.

- F_V - volumetric feed rate to the fermentor (30 l/h)
- V - volume of reaction mixture in the fermentor (10 l)
- x - yeast cell concentration (0.9 g/l)
- - specific growth rate [0.5 g-cells produced/(g-cells h^{-1})]

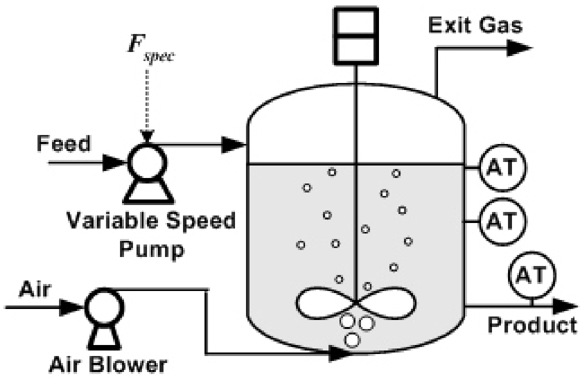


Figure P6.3.7 Schematic of a continuous bio-reactor.

6.4 Second-Order Processes

P6.4.1* For the following second-order transfer functions, determine K_p, $_n$, and .

a. $G_p(s) \quad \dfrac{6}{4s^2 \quad s \quad 4}$ b. $G_p(s) \quad \dfrac{3}{5s^2 \quad 10s \quad 10}$

c. $G_p(s) \quad \dfrac{10}{s^2 \quad 2s \quad 0.5}$ d. $G_p(s) \quad \dfrac{0.5}{0.1s^2 \quad 6s \quad 0.1}$

P6.4.2* Tuning controllers such that the resulting dynamic response has a decay ratio of ¼ has been proposed for some time. Assuming a second-order response for the feedback system, what value of corresponds to a decay ratio of ¼?

P6.4.3* Determine the decay ratio of a second-order process if the overshoot is 0.05.

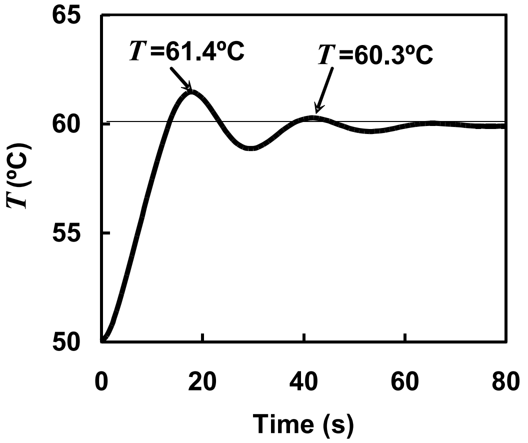


Figure P6.4.6 The closed-loop response of the CST thermal mixer to a setpoint change.

P6.4.4* Consider a pressure control loop that exhibits underdamped behavior. The pressure control loop is tuned for 0.10 overshoot, and the time period for oscillations is 1 minute. The control loop is tuned properly so that after the process reaches steady-state operation, there is no noticeable offset between the setpoint and the measured value of the pressure. Develop a second-order transfer function for this closed-loop system assuming that the input is the pressure setpoint and the output is the measured value of the process pressure.

P6.4.5* Consider a temperature control loop that exhibits underdamped second-order behavior. The temperature control loop is tuned for a 1/8 decay ratio and the time period for oscillations is 14 minutes. The control loop is tuned properly so that after the process reaches steady-state operation, there is no noticeable offset between the setpoint and the measured value of the temperature. Develop a second-order transfer function for this closed-loop system assuming that the input is the temperature setpoint and the output is the measured value of the process temperature.

P6.4.6*** The closed-loop response shown in Figure P6.4.6 was generated using the CST thermal mixer simulation. Using these results, determine as many characteristics of a second-order response as you are able. Then compare and analyze these results.

P6.4.7** Consider the following differential equation

$$\frac{d^2y}{dt^2} \quad K\frac{dy}{dt} \quad y \quad u$$

Discuss the dynamic behavior of this system for $5 \quad K \quad 5$.

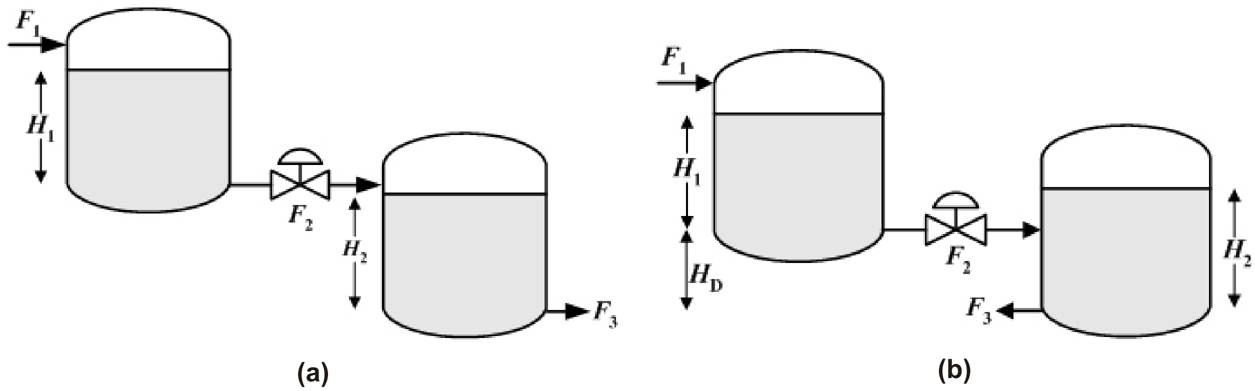


Figure P6.4.8 Schematic for flow through (a) two non-iteracting tank levels and (b) two interacting tank levels.

P6.4.8*** For two non-interactive tank levels shown in Figure P6.4.8a and for the two interacting tank levels shown in Figure P6.4.8b, develop overall transfer functions [$F_3(s)/F_1(s)$] and compare them. For the non-interacting case, the flow from the first to the second tank is given by $F_2 \quad kH_1$ while the discharge flow from the second tank is equal to $F_3 \quad kH_2$. For the interacting case, the flow from the first to the second tank is given by $F_2 \quad k(H_1 \quad H_D \quad H_2)$ while the discharge flow from the second tank is equal to $F_3 \quad kH_2$. Assume that each tank for both cases has the same cross-sectional area (A) and that the density of the liquid is in both cases. Develop a first-order model for each tank separately and combine them for the overall transfer function. Compare the gain, natural period and damping factors for both cases.

6.5 Integrating Processes

P6.5.1* An open-loop test is applied to a tank level process. Initially, the level is constant at a value of 35%. The inflow to the tank is increased by 500 lb/h. After 3 minutes the level reading lines out at 45%. Develop a transfer function for this process for which the input is the change in the inflow rate and the output is the change in the level in the tank.

P6.5.2* An open-loop test is applied to a tank level process. Initially, the level is constant at a value of 35%. The outflow from the tank is increased by 1500 lb/h. After 16 minutes the level reading lines out at 26%. Develop a transfer function for this process for which the input is the change in the outflow rate and the output is the change in the level in the tank.

6.7 Deadtime

P6.7.1** Consider the response of an open-loop underdamped process to a step input change of 500 lb/h shown in Figure P6.7.1. Based on this response, develop a second-order plus deadtime transfer function model for this process. (Hint: Develop the second-order approximation first then determine the deadtime.)

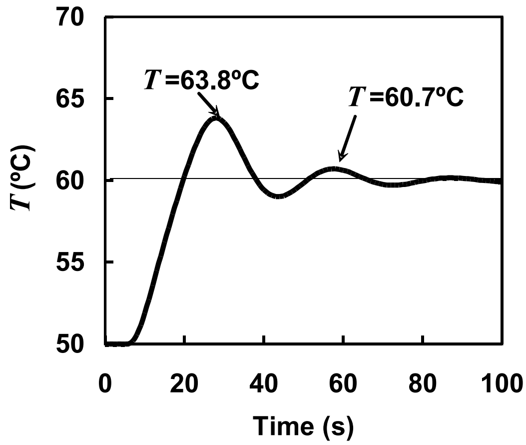


Figure P6.7.1 The response of an open-loop underdamped process to a step input.

P6.7.2* Consider the following transfer function. Indicate as many characteristics of the process corresponding to this transfer function as possible.

$$G_p(s) \quad \frac{323\, e^{\;3s}}{4s^2 \quad 16s \quad 4}$$

6.8 First-Order Plus Deadtime Model

P6.8.1* Consider the following set of input/output data. Develop a FOPDT model for this input/output system and plot your approximation against the data.

Time	Input	Output	Time	Input	Output
0	0	1.0	7	1	1.6
1	0	1.0	8	1	1.8
2	1	1.0	9	1	1.9
3	1	1.05	10	1	1.95
4	1	1.1	11	1	2.0
5	1	1.2	12	1	2.0
6	1	1.4	13	1	2.0

P6.8.2** Section 1.3 states “As a rule of thumb, you can assume that it takes approximately four time constants to observe the full effect of a step change of an input to a process under open-loop conditions”. Evaluate this statement using a FOPDT model.

P6.8.3* Consider four CSTR’s in series. If there is a single irreversible first-order reaction occurring in each reactor, what is the dynamic response of the process?

6.9 Inverse-Acting Process

P6.9.1* Determine whether the following transfer functions exhibit inverse action.

a. $G_p(s) \quad \frac{s^2 \quad 9.99s \quad 0.01}{3s^3 \quad 2s^2 \quad 7s \quad 1}$ b. $G_p(s) \quad \frac{s^2 \quad 4s \quad 3}{s^3 \quad 7s^2 \quad 2s \quad 1}$

c. $G_p(s) \quad \frac{s^3 \quad 3s \quad 2}{s^4 \quad 5s^3 \quad 3s^2 \quad 2s \quad 1}$ d. $G_p(s) \quad \frac{s^3 \quad 3s^2 \quad s \quad 3}{s^4 \quad 4s^3 \quad 3s^2 \quad 5s \quad 1}$

6.12 Simulink

P6.12.1* Using Simulink determine the time-domain behavior of a second-order process (K_p=1, $_n$=0.5) for a two unit step increase in input at time equal to 1 for

(a) =-0.5 **(b)** =0 ©) =0.5 **(d)** =0.9 **(e)** =1.0 **(f)** =2

P6.12.2* Using Simulink determine the time-domain behavior of the following first-order processes using a unit step input at time equal one time unit. Note that the first and third have a right-half plane zero and the second and fourth have a left-half plane zero.

(a) $G_p(s) \quad \frac{s \quad 1}{0.5s \quad 1}$ **(b)** $G_p(s) \quad \frac{s \quad 1}{0.5s \quad 1}$

©) $G_p(s) \quad \frac{s \quad 1}{0.5s \quad 1} e^{\quad 2s}$ **(d)** $G_p(s) \quad \frac{s \quad 1}{0.5s \quad 1} e^{\quad 2s}$

P6.12.3* Using Simulink determine the time-domain response of each of the following process models to a unit step-change input at time equal one time unit:

(a) $G_p(s) \quad \frac{1}{0.5s \quad 1}$ **(b)** $G_p(s) \quad \frac{1}{(0.5s \quad 1)^2}$ **(b)** $G_p(s) \quad \frac{1}{(0.5s \quad 1)^3}$

(d) $G_p(s) \quad \frac{1}{(0.5s \quad 1)^5}$ **(e)** $G_p(s) \quad \frac{1}{(0.5s \quad 1)^7}$ **(f)** $G_p(s) \quad \frac{1}{(0.5s \quad 1)^9}$

P6.12.4* Using Simulink determine the time-domain response of each of the following process models to a unit step input change at time equal one time unit and compare with Figure 6.10.1:

(a) $G_p(s) \quad \frac{4s \quad 1}{3s \quad 1}$ **(b)** $G_p(s) \quad \frac{2s \quad 1}{3s \quad 1}$

P6.12.5* Using Simulink determine the time-domain response of the following transfer function for a unit step change at time equal to one time unit:

$$G_p(s) \quad \frac{3}{3s \quad 1} \quad \frac{2}{0.5s \quad 1}$$

6.18 Projects

Develop FOPDT models using the simulations that accompany the text. Step tests should be applied for ±3% and ±10% changes in the manipulated variable while starting at the nominal, lined out conditions. Average each of the FOPDT parameters obtained from four open-loop step tests. Using the results of the individual step tests, compare the variation in the FOPDT model parameters.

a. CST thermal mixer[S]
b. CST composition mixer[S]
c. Level in a tank[S]
d. CSTR[S]
e. Heat exchanger[S]

A superscript "S" indicates that the visual basic simulators are required to perform these problems.

Part III

PID Control

Chapter 7

PID Control: Characteristics, Forms and Modes

Chapter Objectives

- Derive the characteristic equation and show how it defines the dynamic behavior of a feedback loop.
- Present the various forms of the PID algorithm.
- Demonstrate how to determine whether a controller should be direct or reverse acting.
- Evaluate the fundamental characteristics of proportional, integral and derivative control.
- Present guidelines for selecting the proper mode of a PID controller.
- Compare the relative dynamics of the actuator, process and sensor for several commonly encountered control loops in the CPI and the biotechnology industries to determine the proper PID mode.

7.1 Introduction

Feedback control compares the measured value of the CV to its setpoint and adjusts the MV in an effort to drive the CV to its setpoint. For the everyday examples of feedback control cited in Chapter 1 (i.e., the shower, bathtub, driving a car and balancing a spoon), a human serves as the feedback controller. The primary objectives of feedback control are to:

- Minimize the settling time of the closed-loop process.
- Maintain reliable operation.
- Control to setpoints, i.e., reduce deviations from setpoint and eliminate offset.
- Reject disturbances.

Proportional, integral and derivative action each affect feedback control performance with regard to these objectives. Moreover, controller tuning (selecting the individual levels of each of these feedback components), which is addressed in Chapters 8-9, has a dominant effect on the performance of a feedback controller.

In the CPI and the biotechnology industries, the **Proportional-Integral-Derivative (PID) controller** is the most commonly used controller and is used almost exclusively for flow control loops, pressure control loops and level control loops, as well as most composition and temperature control loops. Operators also serve as controllers (i.e., manual control) for certain loops, e.g., an operator managing a bio-reactor that is generating biomass based on periodic broth samples taken manually and assayed off-line.

As you will see later in this chapter, PID controllers are simple to implement and they are extremely flexible as evidenced by the fact that PID controllers have been applied to almost any conceivable process ranging from refineries to spacecraft to electronic devices to power plants to bio-processes. The PID algorithm is quite computationally efficient and much of the flexibility of a PID controller comes from the unique characteristics of proportional, integral and derivative action.

7.2 Closed-Loop Transfer Functions

The transfer functions for a feedback control loop are derived in this section and used in a later section to determine the fundamental characteristics of proportional action, integral action and derivative action as well as to examine the dynamic properties of feedback loops. Consider the block diagram for a feedback loop shown in Figure 7.2.1 where $G_c(s)$, $G_a(s)$, $G_p(s)$, $G_d(s)$ and $G_s(s)$ are the transfer functions for the controller, actuator, process, disturbance and sensor, respectively. Applying the properties of transfer functions and summation blocks (Section 5.6), the following relationships can be formulated:

(Process) $$Y(s) \quad G_d(s)\,D(s) \quad G_p\,U(s) \qquad \textbf{7.2.1}$$

(Actuator) $$U(s) \quad G_a(s)\,C(s) \qquad \textbf{7.2.2}$$

(Controller) $$C(s) \quad G_c(s)\,E(s) \qquad \textbf{7.2.3}$$

(Summation) $$E(s) \quad Y_{sp}(s) \quad Y_s(s) \qquad \textbf{7.2.4}$$

(Sensor) $$Y_s(s) \quad G_s(s)\,Y(s) \qquad \textbf{7.2.5}$$

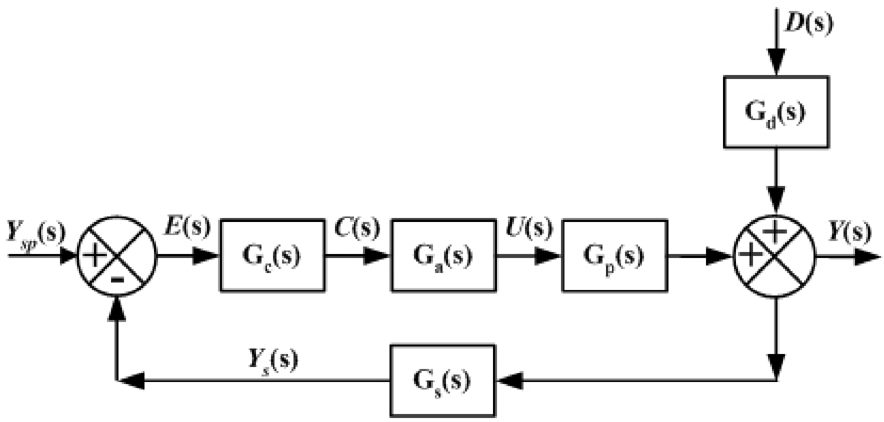


Figure 7.2.1 Block diagram of a general feedback control loop.

Because $G_c(s)$ and $G_a(s)$ represent a series of transfer functions, Equation 5.6.1 can be used to represent the effect of $E(s)$ on $U(s)$, i.e.,

$$\frac{U(s)}{E(s)} \quad G_c(s)G_a(s)$$

Using this relationship to eliminate U(s) from Equation 7.2.1 yields

$$Y(s) \quad G_d(s)D(s) \quad G_c(s)\ G_a(s)G_p(s)E(s)$$

Then, substituting Equation 7.2.4 into this equation to eliminate $E(s)$ gives

$$Y(s) \quad G_d(s)D(s) \quad G_c(s)\ G_a(s)G_p(s)[Y_{sp}(s) \quad Y_s(s)]$$

Finally, substituting Equation 7.2.5 into this equation eliminates $Y_s(s)$ and produces

$$Y(s) \quad G_d(s)D(s) \quad G_c(s)\ G_a(s)G_p(s)[Y_{sp}(s) \quad G_s(s)Y(s)] \tag{7.2.6}$$

Collecting terms and solving for $Y(s)$ results in

$$Y(s) \quad \frac{G_d(s)D(s) \quad G_c(s)\ G_a(s)\ G_p(s)\ Y_{sp}(s)}{G_c(s)\ G_a(s)\ G_p(s)\ G_s(s) \quad 1}$$

Assuming $D(s)$ is zero (i.e., no change in the disturbance level is occurring), the closed-loop transfer function for **setpoint tracking**, i.e., controlling for setpoint changes, is given by

$$\frac{Y(s)}{Y_{sp}(s)} \quad \frac{G_c(s)\ G_a(s)\ G_p(s)}{G_c(s)\ G_a(s)\ G_p(s)\ G_s(s) \quad 1} \tag{7.2.7}$$

Setpoint tracking is also referred to as **servo control**.

Assuming $Y_{sp}(s)$ is zero (i.e., a fixed setpoint is being applied because $Y_{sp}(s)$ is a deviation variable), the closed-loop transfer function for **disturbance rejection**, i.e., controlling to a fixed setpoint in the face of disturbance upsets, is given by

$$\frac{Y(s)}{D(s)} \quad \frac{G_d(s)}{G_c(s)\ G_a(s)\ G_p(s)\ G_s(s) \quad 1} \tag{7.2.8}$$

Controlling for disturbance rejection is also called **regulatory control**.

Note that the denominator of the closed-loop transfer function for setpoint tracking and disturbance rejection is the same. If the denominator of the closed-loop transfer function is set equal to zero, the following equation results,

$$G_c(s)\ G_a(s)\ G_p(s)\ G_s(s) \quad 1 \quad 0 \tag{7.2.9}$$

which is known as the **characteristic equation** of the feedback loop, results. The roots of the characteristic equation are the poles of the feedback process and, therefore, determine the majority of the dynamic behavior of the closed-loop process. For example, if all the roots of the characteristic equation are real negative values, the closed-loop dynamic behavior is overdamped. Further, if there are complex roots, oscillatory closed-loop behavior results. Finally, if any of the roots have positive real parts, the closed-loop system is unstable. Although the poles of a closed-loop system remain the same for servo and regulatory control, different dynamic behavior can result, especially if $G_d(s)$ has a right-half plane zero.

Note that Equations 7.2.7 and 7.2.8 can be represented by

$$G_{CL}(s) = \frac{\Pi(\text{Input to Output})}{\Pi(\text{Loop})+1} \qquad \mathbf{7.2.10}$$

where $G_{CL}(s)$ is the closed-loop transfer function, $\Pi(\text{Input to Output})$ is the product of the transfer functions from the input to the output [e.g., for Equation 7.2.7, the product of the transfer functions from the input to the output is $G_c(s)G_a(s)G_p(s)$] and $\Pi(\text{Loop})$ is the product of the transfer functions around the feedback control loop [e.g., for this case, the product around the loop is $G_c(s)G_a(s)G_p(s)G_s(s)$]. This equation can also be applied to other closed-loop block diagram arrangements.

Example 7.1 Dynamic Behavior of a P-only Controller Applied to a Second-Order Process

Problem Statement. Determine the dynamic behavior of a P-only controller (i.e., $G_c(s) = K_c$) with K_c equal to 2 applied to a second-order process (K_p=1; τ_n=5; ζ=1.5). Assume that the second-order process model represents the combined effect of the actuator, process and sensor.

Solution. Because the roots of the characteristic equation are the poles of the closed-loop system and determine its dynamic behavior, the characteristic equation is used here to determine the dynamic behavior of this closed-loop system. Using the specified properties of the second-order process model, $G_p(s)$ is given as

$$G_p(s) = \frac{K_p}{\tau_n^2 s^2 + 2\tau_n\zeta s + 1} = \frac{1}{25s^2 + 15s + 1}$$

Substituting the specifications of the problem into the characteristic equation (Equation 7.2.9) and setting the result equal to zero yields

$$G_c(s)G_p(s) + 1 = (2)\left(\frac{1}{25s^2 + 15s + 1}\right) + 1 = 0$$

Rearranging yields $$25s^2 + 15s + 3 = 0$$

Putting this equation into the standard form for a second-order transfer function results in

$$8.33s^2 + 5s + 1 = 0$$

which indicates that the closed-loop second-order natural period, τ_n, is 2.89 and the closed-loop damping factor, ζ, is 0.866, which corresponds to underdamped behavior with a small amount of overshoot. That is, the poles of the closed-loop transfer function are a complex conjugate pair with a negative real part, indicating damped oscillatory behavior. Comparing the closed-loop dynamic behavior to the open loop by considering the natural periods for the open- and closed-loop systems demonstrates that the dynamic response of this system is faster for the closed-loop process than the open-loop. This characteristic will be derived for the general case in a later section.

Example 7.2 Analysis of a System with Positive Feedback

Problem Statement. The feedback system shown in Figure 7.2.1 subtracts the measurement of the CV from the setpoint, which is known as **negative feedback**. Analyze the behavior of **positive feedback**, i.e., the measurement of the CV is added to the setpoint, for the closed-loop process. Consider the P-only controller used in Example 7.1.

Solution. Modifying Equation 7.2.6 recognizing that for positive feedback $G_s(s)Y(s)$ is added to $Y_{sp}(s)$,

$$Y(s) \quad G_d(s)D(s) \quad G_p(s)\ G_a(s)G_c(s)[Y_{sp}(s) \quad G_s(s)Y(s)]$$

Solving for Y(s),

$$Y(s) \quad \frac{G_d(s)D(s) \quad G_p(s)\ G_a(s)\ G_c(s)\ Y_{sp}(s)}{G_c(s)\ G_a(s)\ G_p(s)\ G_s(s) \quad 1}$$

Therefore, the characteristic equation for positive feedback is

$$G_c(s)\ G_a(s)\ G_p(s)\ G_s(s) \quad 1 \quad 0$$

Applying this equation to the process and controller used in Example 7.2.1 yields

$$G_c(s)G_p(s) \quad 1 \quad 2\frac{1}{25s^2 \quad 15s \quad 1} \quad 1 \quad 0$$

Rearranging yields $$25s^2 \quad 15s \quad 1 \quad 0$$

which has poles equal to (-0.661, 0.0606), indicating unstable behavior because of a right-half plane pole (i.e., a positive real portion of a pole). The fact that this polynomial has a negative coefficient also indicates unstable behavior. Note that this system can be converted to the stable form as in Example 7.1 by using a negative controller gain (i.e., K_c=-2).

Example 7.3 Feedback Control of an Open-Loop Unstable Process

Problem Statement. Consider an open-loop unstable second-order system

$$G_p(s) \quad \frac{1}{25s^2 \quad 15s \quad 1}$$

which has poles equal to (-0.661, 0.0606). Evaluate the closed-loop performance of this process for a P-only controller with a controller gain, K_c.

Solution. Applying the characteristic equation for negative feedback (Equation 7.2.9)

$$K_c \frac{1}{25s^2 \quad 15s \quad 1} \quad 1 \quad 0$$

Rearranging and collecting terms $$25s^2 \quad 15s \quad 1 \quad K_c \quad 0$$

The closed-loop poles are given by $$s \quad \frac{15 \quad \sqrt{225 \quad 100(K_c \quad 1)}}{50}$$

which results in negative real portions of the poles for K_c>1 while a positive real portion for the poles results for K_c<1. Therefore, when K_c>1, the closed-loop system is stable, but the system is unstable for K_c<1. That is, this open-loop unstable process is stabilized when the controller gain is large enough. Exothermic CSTRs can be open-loop unstable.

Example 7.4 Stability Limit for Feedback Control

Problem Statement. Most chemical and bio-chemical processes are open-loop stable [e.g., exceptions are an exothermic CSTR (Example 3.6) or a nuclear reactor]. However, open-loop stable processes can become unstable systems under feedback control. A stable open-loop industrial process will become unstable if the wrong controller settings are used. Besides determining the values of the poles of a process, one of the ways to determine the stability of a system is to apply the Routh stability criterion.

Consider an overdamped second-order process (K_p=10; $_n$=10; =2) and its sensor ($_s$=1). Find the values of the controller gain for a P-only controller that render this feedback system unstable using the Routh stability criterion. Assume $G_a(s)$=1.

Solution. The characteristic equation for this feedback system is

$$G_c(s)G_p(s)G_s(s) \quad 1 \quad K_c \frac{10}{100s^2 \quad 40s \quad 1} \frac{1}{s \quad 1} \quad 1 \quad 0$$

Rearranging this equation into a polynomial form yields

$$100s^3 \quad 140s^2 \quad 41s \quad 1 \quad 10K_c \quad 0$$

The Routh array for this cubic equation is given by Equation 5.4.2. Substituting the numerical values yields

$$\begin{array}{cc} 100 & 41 \\ 140 & 1 \quad 10K_c \\ 40.3 \quad 7.14K_c & 0 \\ 1 \quad 10K_c & 0 \end{array}$$

From the Routh stability criterion, each element in the first column must be positive for stability. From the fourth row of the first column, K_c>-0.1. From the third element in the first column, for stability K_c<5.64. Therefore, for stable closed-loop behavior,-0.1<K_c<5.64. From a practical standpoint, only positive controller gains should be considered. Therefore, this closed-loop system becomes unstable when K_c exceeds 5.64. In general, a feedback control loop for any real process can be rendered unstable by using a large enough value for the controller gain.

Self-Assessment Questions

Q7.2.1 What is the characteristic equation of a feedback loop and what is its significance?

Q7.2.2 What do the transfer functions for the closed-loop response indicate about the dynamics of feedback systems for setpoint changes and disturbance rejection?

Q7.2.3 What is the difference between servo control and regulatory control?

Q7.2.4 Can a negative feedback system stabilize an open-loop unstable process? Why?

Self Assessment Answers

Q7.2.1 The characteristic equation of a feedback loop is the denominator of the closed-loop transfer function set equal to zero. More specifically, for a typical loop it is the product of the transfer function of the actuator, the process, the sensor, and the controller plus one and set equal to zero, i.e.**,**

$$G_a(s)G_c(s)G_s(s)G_p(s)+1=0$$

The roots of this equation (i.e., the poles) determine the majority of the dynamic behavior of the closed-loop process.

Q7.2.2 The closed-loop transfer functions indicate that a closed-loop process has the same poles for setpoint changes and disturbance upsets. That is, if a linear process exhibits overdamped closed-loop behavior for setpoint changes, it will be overdamped for disturbance upsets. On the other hand, if the disturbance [D(s)] adds positive zeros to the closed-loop transfer function for disturbance rejection, the dynamic behavior will be significantly different than the closed-loop response for setpoint changes due to the inverse response associated with right-half plane zeros.

Q7.2.3 Servo control is feedback control for setpoint changes while regulatory control, which is based on a fixed setpoint, is concerned with rejecting the effects of disturbances.

Q7.2.4 Negative feedback can stabilize an open-loop unstable process. The gain of the controller must be greater than the minimum and less than the maximum gain to stabilize an open-loop unstable process. Feedback control can change the dynamic response of the open-loop response. For example, Example 7.1 showed that a P-only controller applied to an open-loop overdamped process produced a closed-loop underdamped process. Therefore, under certain circumstances, feedback control can stabilize an open-loop unstable process by using the corrective action of the MV to compensate for the instability of the open-loop system.

Self-Assessment Problems

P7.2.1 Using the characteristic equation, determine the closed-loop dynamic behavior for a first-order process (K_p=1; τ_p=1) with a P-only controller [$G_c(s)$=2]. Assume that $G_a(s)$=$G_s(s)$=1.

P7.2.2 Find the values of K_c for a P-only controller that stabilize the following open-loop unstable system: $G_p(s)$=1/(s-5). Assume $G_a(s)$=$G_s(s)$=1.

Self-Assessment Answers

P7.2.1 Applying the characteristic equation

$$K_c\left(\frac{K_p}{\tau_p s+1}\right)+1=(2)\left(\frac{1}{s+1}\right)+1=0$$

Rearranging into the standard form for a first-order system, $0.333s+1=0$, which indicates that the closed-loop response is first-order with a time constant equal to 0.333.

P7.2.2 Applying the characteristic equation

$$K_c\left(\frac{1}{s-5}\right)+1=0$$

Rearranging into the standard form for a first-order system, $s+K_c-5=0$; therefore, for stability, $K_c>5$.

7.3 Position Forms of the PID Algorithm

The ISA (Instrument Society of America) standard for the PID algorithm in the **position form** is given as

$$c(t) = \bar{c} + K_c\left[e(t) + \frac{1}{\tau_I}\int_0^t e(t)\,dt + \tau_D\frac{d\,e(t)}{dt}\right] \qquad \textbf{7.3.1}$$

where K_c, τ_I, and τ_D are the user-selected tuning parameters, $c(t)$ is the output from the controller and $e(t)$ is $[y_{sp} - y_s(t)]$. **Note that $\bar{c}$ is the value of the controller output when the controller is turned on**. Proportional action is provided by the first term inside the bracket [i.e., $e(t)$]; the second term provides integral action; the third term provides derivative action. K_c is the controller gain and should not be confused with the process gain K_p. K_p has units of $\Delta y_s/\Delta c$ while K_c has units corresponding to $\Delta c/\Delta y_s$. The controller gain, K_c, can have a variety of units depending on the units used for c and y. For example, if the controller output is expressed as a 0 to 100% signal and the CV is temperature in ºF, the controller gain has units of %/ºF. On the other hand, if the controller output is expressed as a flow rate (i.e., lb/h) and the CV is a pressure in psi, the controller gain has units of lb/h-psi.

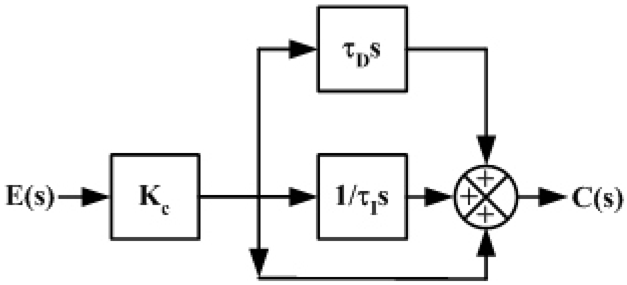


Figure 7.3.1 Block diagram of a conventional PID algorithm.

The transfer function for a PID controller (Example 5.9) is given by

$$G_c(s) = \frac{C(s)}{E(s)} = K_c\left(1 + \frac{1}{\tau_I s} + \tau_D s\right) \qquad \textbf{7.3.2}$$

A block diagram for the PID algorithm is shown in Figure 7.3.1. Note that proportional, integral and derivative action act on $e(t)$ in parallel with each other.

Example 7.5 Dynamic Behavior of a PI Controller Applied to a Second-Order Process

Problem Statement. Determine the dynamic behavior of a PI controller with K_c equal to 1 and τ_I equal to 1 applied to a second-order process (K_p=1, τ_n=5 and ζ=2). Assume that the second-order process model represents the combined effect of the actuator, process and sensor.

Solution. Substituting this information into the characteristic equation (Equation 7.2.9) yields

$$G_c(s)G_p(s) + 1 = 1\left(1 + \frac{1}{s}\right)\frac{1}{25s^2 + 20s + 1} + 1 = 0$$

Rearranging yields $$25s^3 + 20s^2 + 2s + 1 = 0$$

By numerical root finding, a real root of this cubic equation is determined (i.e., s=-0.764). Factoring out this root by long division yields

$$25s^2 + 0.905s + 1.309 = 0$$

Rearranging into the standard form for a second-order function,

$$19.1s^2 + 0.692s + 1 = 0$$

which indicates that for the closed-loop system τ_n is 4.37 and ζ is 0.08, which corresponds to stable behavior but a highly oscillatory underdamped response. The poles of this system are (-0.764; -0.0181 ± 0.228 i).

When a setpoint change is made to a process using the PID algorithm given by Equation 7.3.1, a spike in the calculated value of $de(t)/dt$ occurs, causing a spike in $c(t)$. The spike in the derivative of the error from setpoint results because the difference between the setpoint and the CV value changes instantaneously when a setpoint change is implemented. This behavior is called **derivative kick** and can be eliminated by replacing $de(t)/dt$ with $-dy_s(t)/dt$ yielding

$$c(t) = \bar{c} + K_c\left[e(t) + \frac{1}{\tau_I}\int_0^t e(t)\,dt - \tau_D\frac{dy_s(t)}{dt}\right] \qquad \textbf{7.3.3}$$

Note that because

$$e(t) = y_{sp}(t) - y_s(t)$$

if it is assumed that y_{sp} is constant,

$$\frac{de(t)}{dt} = -\frac{dy_s(t)}{dt}$$

Equation 7.3.3 and Equation 7.3.1 are equivalent as long as there are not changes in the value of the setpoint. In fact, there is generally no need to have derivative action on the error from setpoint because it just takes action against a setpoint change. The **derivative-on-measurement** form of the PID algorithm (Equation 7.3.3) is recommended because it provides derivative action, but it is not susceptible to derivative kick.

Equation 7.3.3 can be implemented in a digital form by using the following approximations

$$\int_0^t e(t)\,dt \approx \sum_{i=1}^{n} e(i\,\Delta t)\,\Delta t$$

where n is the number of control moves because the controller was turned on and is equal to $t/\Delta t$. In addition,

$$\frac{d\,y_s(t)}{dt} \approx \frac{y_s(t) - y_s(t-\Delta t)}{\Delta t}$$

where Δt is the time interval between applications of the PID algorithm, i.e., the **control interval**. These approximations result in the digital version of the position form of the PID controller.

$$c(t) = \bar{c} + K_c\left[e(t) + \frac{\Delta t}{\tau_I}\sum_{i=1}^{n} e(i\,\Delta t) - \tau_D\,\frac{y_s(t) - y_s(t-\Delta t)}{\Delta t}\right] \qquad \textbf{7.3.4}$$

Another way to represent the controller gain is the **proportional band** (***PB***), which is an approach that was in more frequent use 30 to 40 years ago. *PB* is given by

$$PB = \frac{100\%}{K_c^D} \qquad \textbf{7.3.5}$$

where K_c^D is given by $$K_c^D = K_c\,\frac{\Delta y_s}{\Delta c}$$

where Δc **is the range of the controller output** (e.g., 100%) and Δy_s **is the span of the sensor** that measures the CV. That is, K_c^D is a dimensionless form for the controller gain. The proportional band is small when the controller gain is large and *PB* is large when K_c is small. The position form of the PID algorithm using proportional band is given as

$$c(t) = \bar{c} + \frac{100\%}{PB}\,\frac{\Delta c}{\Delta y_s}\left[e(t) + \frac{1}{\tau_I}\int_0^t e(t)\,dt + \tau_D\,\frac{d\,e(t)}{dt}\right] \qquad \textbf{7.3.6}$$

When proportional band is used with a controller, the controller output is expressed as a percentage.

Example 7.6 Conversion from Proportional Band to K_c

Problem Statement. Determine the dimensional version of the controller gain corresponding to a proportional band of 200%. The span of the sensor is 200 psi and the controller output is 0% to 100%.

Solution. Applying Equation 7.3.5 yields

$$K_c^D = \frac{100\%}{PB} = \frac{100\%}{200\%} = 0.5$$

Then, using the span of the sensor and the controller output results in

$$K_c = 0.5\left(\frac{100\,\%}{200\,\text{psi}}\right) = 0.25\ \%/\text{psi}$$

Example 7.7 Conversion from K_c to Proportional Band

Problem Statement. Determine the proportional band corresponding to a controller gain of 15 %/ºF. The span of the sensor is 25ºF and the controller output is 0% to 100%.

Solution. First, the process gain must be converted into a dimensionless form. Using the span of the sensor and the controller output,

$$K_c^D = \left(\frac{15\%}{{}^\circ F}\right)\left(\frac{25\ {}^\circ F}{100\%}\right) = 3.75$$

Then, application of Equation 7.3.5 yields

$$PB = \frac{100\%}{3.75} = 26.7\%$$

Self-Assessment Questions

Q7.3.1 In relation to the units of the gain of a process, what are the units of the controller gain?
Q7.3.2 What is the difference between the controller gain and the proportional band?
Q7.3.3 In Equation 7.3.4, which term represents proportional action and which term represents integral action?

Self-Assessment Answers

Q7.3.1 The units for the controller gain are the inverse of the units for the process gain.

Q7.3.2 The proportional band is inversely related to the controller gain. Equation 7.3.5 gives the explicit relationship between the two.

Q7.3.3 In Equation 7.3.4, $K_c e(t)$ represents the proportional action and the integral action is given by $K_c \frac{\Delta t}{\tau_I}\sum_{i=1}^{n} e(i\,\Delta t)$.

Self-Assessment Problem

P7.3.1 Using the characteristic equation, determine the closed-loop dynamic behavior for a first order process (K_p=1; τ_p=2) with a PI controller (K_c=1; τ_I=2). Assume that $G_a(s)$=$G_s(s)$=1.

Self-Assessment Answers

P7.3.1 Applying the characteristic equation,

$$K_c\left(1+\frac{1}{\tau_I s}\right)\left(\frac{K_p}{\tau_p s+1}\right)+1=(1)\left(1+\frac{1}{2s}\right)\left(\frac{1}{2s+1}\right)+1=0$$

Rearranging into the standard form for a second-order system, $4s^2+4s+1=0$, which indicates that the natural period is 2 and the damping factor is one (critically damped response).

7.4 Velocity Forms of the PID Algorithm

The PID algorithm can also be applied in the **velocity form**, i.e., a form that determines changes to the current controller output. Consider Equation 7.3.4.

$$c(t) = \bar{c} + K_c \left[e(t) + \frac{\Delta t}{\tau_I} \sum_{i=1}^{n} e(i\Delta t) - \tau_D \frac{y_s(t) - y_s(t-\Delta t)}{\Delta t} \right] \tag{7.3.4}$$

Applying Equation 7.3.4 at $t - \Delta t$ results in the following equation

$$c(t-\Delta t) = \bar{c} + K_c \left[e(t-\Delta t) + \frac{\Delta t}{\tau_I} \sum_{i=1}^{n-1} e(i\Delta t) - \tau_D \frac{y_s(t-\Delta t) - y_s(t-2\Delta t)}{\Delta t} \right] \tag{7.4.1}$$

Subtracting Equation 7.4.1 from Equation 7.3.4 results in the velocity form of the PID algorithm.

$$\Delta c(t) = K_c \left[e(t) - e(t-\Delta t) + \frac{\Delta t}{\tau_I} e(t) - \tau_D \frac{y_s(t) - 2y_s(t-\Delta t) + y_s(t-2\Delta t)}{\Delta t} \right] \tag{7.4.2}$$

where

$$c(t) = c(t-\Delta t) + \Delta c(t) \tag{7.4.3}$$

The velocity form for the derivative on the error from setpoint is given by

$$\Delta c(t) = K_c \left[e(t) - e(t-\Delta t) + \frac{\Delta t}{\tau_I} e(t) + \tau_D \frac{e(t) - 2e(t-\Delta t) + e(t-2\Delta t)}{\Delta t} \right] \tag{7.4.4}$$

Another popular version of the velocity form of the PID algorithm can be developed by eliminating proportional action for setpoint changes. For certain control systems, implementing large setpoint changes unduly upsets the process. That is, a setpoint change abruptly affects the process operation. Because the error from setpoint changes as soon as the setpoint change is implemented, the response of a standard proportional controller to a setpoint change is virtually instantaneous. The sharp response of a proportional controller to a setpoint change is called **proportional kick.** Noticing that the proportional part of Equation 7.4.2 is simply the difference between $y(t-\Delta t)$ and $y(t)$ when the setpoint remains unchanged leads to replacing the difference between errors from setpoint with the difference between measured values of the CV in the velocity form of the PID controller, i.e.,

$$\Delta c(t) = K_c \left[y_s(t-\Delta t) - y_s(t) + \frac{\Delta t}{\tau_I} e(t) - \tau_D \frac{y_s(t) - 2y_s(t-\Delta t) + y_s(t-2\Delta t)}{\Delta t} \right] \tag{7.4.5}$$

The advantage of this form of the PID controller is that it does not act as abruptly to setpoint changes as Equation 7.4.2 or 7.4.4, i.e., this form eliminates proportional kick and makes it easier to recover from windup (Section 14.2). In fact, from Equation 7.4.5 only the integral action moves the process toward a new setpoint. This reduction in aggressive setpoint tracking has an effect that is similar to bumpless transfer, which is introduced in Chapter 14.

Note that the position form of the PID algorithm calculates the absolute value of the output of the controller while the velocity form calculates the change in the controller output, which is added to the current level of the controller output. The position and velocity modes are different forms of the same equation; therefore, they are generally equivalent while the velocity form is usually used industrially. In general, PLC and DCS vendors offer a variety of PID algorithms in their commercial process control system products, each with their pros and cons, that users can choose from when configuring applications. A caution for users (because of the different versions of PID algorithms available) is that PID control applications developed in one vendor's system may not necessarily be directly portable to another vendor's system. Nevertheless, a control computer should offer velocity forms that eliminate proportional and derivative kick.

Self-Assessment Questions

Q7.4.1 The term "$e(t)$-$e(t-\Delta t)$" in the velocity form of the PID algorithm represents what type of control action?

Q7.4.2 Is it possible to eliminate proportional kick while maintaining proportional action for disturbance upsets using the position form of a PID controller?

Q7.4.3 What is proportional kick and derivative kick?

Self-Assessment Answers

Q7.4.1 The term "$e(t)$-$e(t-\Delta t)$" in the velocity form of the PID algorithm represents proportional action.

Q7.4.2 It is not possible to eliminate proportional kick while maintaining proportional action for disturbance upsets using the position form of a PID controller because the proportional action from the position form of a PID controller is equal to $K_c e(t)$ and removing the setpoint from this expression would remove proportional action from the controller.

Q7.4.3 Proportional kick is the sharp change in the proportional action resulting from a change in the setpoint. Derivative kick is a sharp change in the derivative action resulting from a change in the setpoint.

7.5 Interactive Form of the PID Controller

An older version of the PID algorithm, originally applied using analog devices, is called an **interactive PID controller**. Figure 7.5.1 shows a block diagram of this controller, which is also referred to as "**rate before reset**" because the derivative action is in series with and precedes the integral action. Note that K_c, τ_I and τ_D are the controller gain, reset time and derivative time, respectively, for the interactive form of a PID controller, which can be different from the tuning parameters for the non-interactive form of a PID controller (Figure 7.3.1). Equating the transfer functions for the non-interactive (Equation 7.3.2) and interactive (Figure 7.5.1) forms of the PID algorithm yields

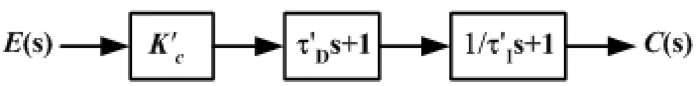


Figure 7.5.1 Block diagram for an interactive PID controller.

$$K_c\left(1+\frac{1}{\tau_I s}+\tau_D s\right) = K_c\left(1+\frac{1}{\tau_I s}\right)[\tau_D s+1] = K\left(\frac{\tau_D\tau_I s^2+\tau_I s+\tau_D s+1}{\tau_I s}\right)$$

Further, dividing the numerator of the term in the bracket on the right-hand side by the denominator yields

$$K_c\left(1+\frac{1}{\tau_I s}+\tau_D s\right) = K_c\left(1+\frac{\tau_D}{\tau_I}+\frac{1}{\tau_I s}+\tau_D s\right)$$

Factoring out $1+\frac{\tau_D}{\tau_I}$ from the terms inside the bracket on the right-hand side yields

$$K_c\left(1+\frac{1}{\tau_I s}+\tau_D s\right)=K'_c\left(1+\frac{\tau'_D}{\tau'_I}\right)\left[1+\frac{1}{\tau'_I\left(1+\frac{\tau'_D}{\tau'_I}\right)s}+\frac{\tau'_D}{1+\frac{\tau'_D}{\tau'_I}}s\right]$$

Equating like terms between the non-interactive form and the rearranged interactive form results in the following formulas for converting from interactive tuning parameters (i.e., the tuning parameters with primes) to tuning parameters for the conventional non-interactive PID form.

$$K_c = K'_c(1+\tau'_D/\tau'_I)$$
$$\tau_I = \tau'_I(1+\tau'_D/\tau'_I)$$
$$\tau_D = \tau'_D\frac{1}{1+\tau'_D/\tau'_I}$$

A PI or a P-only interactive controller is no different from the earlier form presented (i.e., non-interactive PID, Equation 7.3.1). This can be seen by comparing Figures 7.3.1 and 7.5.1 or the previous equations with τ_D equal to zero. The only difference between an interactive and a non-interactive controller occurs for the PID controller. While there are differences in the values of the tuning constants, both controllers apply the same PID algorithm. That is, for the same amount of proportional, integral and derivative action, the interactive and non-interactive controllers can have different values of the tuning constants (i.e., K_c, τ_I, and τ_D) as shown in the previous set of equations.

Even though **interactive controllers** are an option on most DCSs, it is not recommended to use this form because it can cause confusion concerning the tuning parameters and **offers no advantage over the non-interactive form of the PID controller**. As a result, only the formulas for converting the settings from the interactive form to the non-interactive form are presented here.

Self-Assessment Questions

Q7.5.1 Why are the interactive and non-interactive forms of a PI controller equivalent?

Self-Assessment Answer

Q7.5.1 The noninteractive form, i.e., the conventional form, applies proportional, integral and derivative action in parallel to each other (Figure 7.3.1) while the interactive form applies the proportional, integral and derivative terms in series (Figure 7.5.1). They are equivalent because both apply PID control according to Equation 7.3.1 although the block diagrams for each are different. The tuning constants for both controllers are different for PID controllers, but the tuning parameters are the same when a PI or P-only controller is used.

7.6 Direct- and Reverse-Acting Controllers

Depending on the process and the final control element, the change in the controller output, $\Delta c(t)$ can be either added to or subtracted from the control action previously implemented, $c(t-\Delta t)$. A PID controller that adds $\Delta c(t)$ to the previous control action is called a **reverse-acting controller** and a controller that subtracts $\Delta c(t)$ is referred to as a **direct-acting controller**, i.e., for a PID controller in the velocity form, i.e.,

Direct-acting controller $c(t) \quad c(t \quad t) \quad c(t)$ **7.6.1**

Reverse-acting controller $c(t) \quad c(t \quad t) \quad c(t)$ **7.6.2**

These definitions of direct-acting and reverse-acting controllers have evolved from industrial use even though their designation is counterintuitive based on Equations 7.6.1 and 7.6.2. The standard definition states that **the output of a direct-acting controller increases as the value of the process measurement y increases**. This definition agrees with Equation 7.6.1 because negative sign for y in the error from setpoint (i.e., y_{sp}-y) cancels the negative sign in Equation 7.6.1 preceding $c(t)$.

Figure 7.6.1 shows a general feedback loop considering only the gains of the controller, actuator, process and sensor. From the analysis in Section 7.2, the characteristic equation considering only these gains is

$$K_c K_a K_p K_s \quad 1 \quad 0 \tag{7.6.3}$$

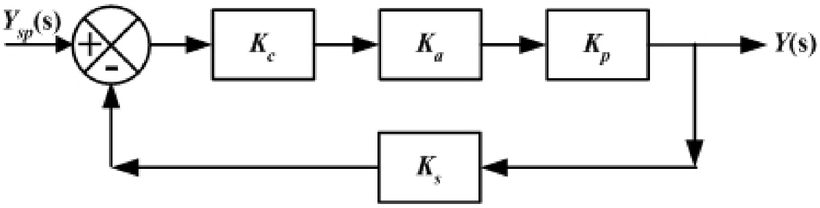


Figure 7.6.1 General feedback loop including only the gains of each element.

If $K_c K_a K_p K_s \quad 0$, positive feedback will result, causing unstable closed-loop behavior (See Example 7.2). Therefore, the product of $K_c K_a K_p K_s$, which is known as the **loop gain**, must be positive for closed-loop stability. K_p can be either positive or negative, depending on the process and K_a can be either positive (direct-acting actuator) or negative (reverse-acting actuator). K_s is usually positive. As a result, K_c must be chosen so that negative feedback is used. Therefore, from Equation 7.6.1 and 7.6.2 **when a negative value of K_c is required to maintain $K_c K_a K_p K_s$ greater than zero, a direct-acting controller is selected and when a positive value of K_c is required, a reverse-acting controller is used.**

Table 7.1 lists the guidelines for selecting a direct-acting or a reverse-acting controller based on the sign of the process gain and whether the actuator is direct or reverse acting. An air-to-open valve actuator for a control valve is referred to as a direct-acting actuator because when the pneumatic signal to the valve actuator is increased, the flow through the control valve increases, i.e., an increase in the input causes an increase in the output. Likewise, an air-to-close valve actuator is a reverse-acting actuator because an increase in the pneumatic signal to the valve causes a decrease in the flow rate through the valve. The previous analysis is based on assuming that the valve is not equipped with a valve positioner. **A valve equipped with a valve positioner will behave as a direct-acting actuator regardless of whether an air-to-open or an air-to-close valve actuator is used.** This results because an increase in the specified valve position for a valve with a positioner (actuator input) produces an increase in the flow rate through the valve (actuator output) and a decrease in the specified valve position produces a decrease in the flow rate. In addition, variable speed pumps behave as direct-acting actuators.

Table 7.1
Guidelines for Selection of Direct- and Reverse-Acting Controllers

Process Gain	Direct-Acting Actuator (Air-to-open valve actuator or a control valve with a valve positioner or a variable speed pump)	Reverse-Acting Actuator (Air-to-close valve actuator without a valve positioner)
Positive	Reverse-Acting PID	Direct-Acting PID
Negative	Direct-Acting PID	Reverse-Acting PID

To illustrate the application of Table 7.1, consider a heat exchanger (Figure 7.6.2) in which the steam flow rate to the heat exchanger is manipulated to control the temperature of the process stream leaving the heat exchanger. Because an increase in the steam flow rate to the heat exchanger results in an increase in the outlet temperature of the process stream, the process gain of this system is positive. Consider a direct-acting final control element on the steam. For this case, the loop gain is positive; therefore, a reverse-acting PID controller should be used. Table 7.1 also indicates that a reverse-acting controller (Equation 7.6.2) should be used for this case. On the other hand, if the direct-acting final control element is replaced with a reverse-acting final control element, the loop gain would be negative; therefore, a direct-acting controller (Equation 7.6.1) is required, which is also in agreement with Table 7.1. **Remember that the choice between an air-to-open or an air-to-close valve actuator usually depends upon whether the final control element should fail open or closed when instrument air pressure is lost.**

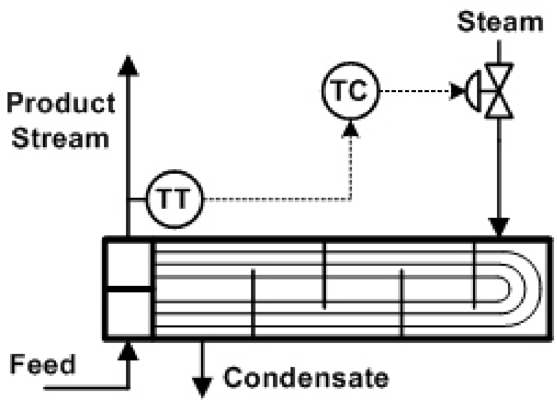


Figure 7.6.2 Control diagram for a temperature controller applied to a steam-heated heat exchanger.

Now consider a heat exchanger in which the flow rate of coolant to the heat exchanger is manipulated to control the temperature of the process stream leaving the heat exchanger (Figure 7.6.3). Because an increase in the flow rate of the coolant to the heat exchanger results in a decrease in the CV for this process, the process gain is negative. Also, consider a direct-acting final control element. Under these conditions, the loop gain is negative; therefore, a direct-acting controller should be used. Finally, if a reverse-acting final control element is substituted for the direct-acting final control element, a reverse-acting controller should be used. These examples represent all the possible combinations of process gains and reverse- and direct-acting actuators as shown in Table 7.1. Obviously, these different combinations of positive and negative process gains and reverse- and direct-acting final control elements can each occur in the implementation of process control in industry. As a result, the process control engineer needs a way to conveniently choose a

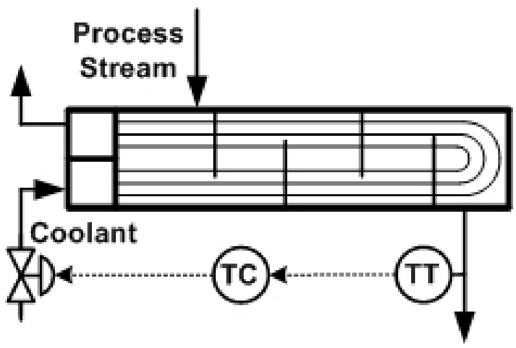


Figure 7.6.3 Control diagram for a temperature controller applied to a liquid/liquid heat exchanger.

direct-acting or a reverse-acting controller. On a DCS, when a control loop is set up, there is typically a box to check to select a direct- or reverse-acting controller. For analog controllers, there is a switch on the back of the controller that allows the user to select the proper form for the controller, direct or reverse acting.

If the position form of the PID controller is used, whether a direct-acting or reverse-acting controller is used determines whether the proportional, integral and derivative terms are added or subtracted, respectively, from $\bar{c}$. That is, Equation 7.3.1 represents a reverse-acting PID controller in the position form and the following equation represents a direct-acting position form PID controller.

$$c(t) = \bar{c} - K_c \left[e(t) + \frac{1}{\tau_I}\int_0^t e(t)\,dt - \tau_D \frac{d\,y_s(t)}{dt} \right] \qquad \textbf{7.6.4}$$

Example 7.8 Selecting the Proper Form of the PID Algorithm

Problem Statement. Write the position form of the PID algorithm similar to Equation 7.3.1 for Figure 7.6.4 from Example 3.3 and assume that the control valve on the feed has an air-to-close actuator without a valve positioner. Use the form that is not susceptible to derivative kick.

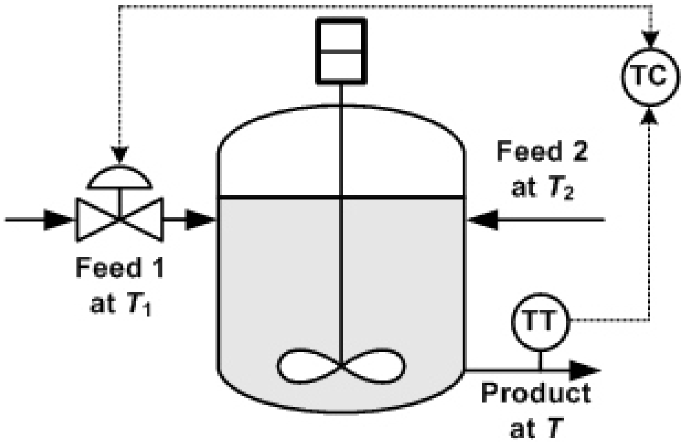


Figure 7.6.4 Control diagram of a CST thermal mixing process. Note that T_1<T_2.

Solution. From Figure 7.6.4, because T_1 is less than T_2, as F_1 is increased, the outlet temperature from the CST thermal mixer decreases; therefore, the process gain is negative. Because an air-to-close valve actuator, which is a reverse-acting actuator, is used, from Table 7.1 a reverse-acting controller should be used. To eliminate derivative kick, the derivative action should be based on the process measurement. The resulting position form of the PID algorithm is given by

$$c(t) = \bar{c} + K_c \left[e(t) + \frac{1}{\tau_I}\int_0^t e(t)\,dt - \tau_D \frac{d\,y_s(t)}{dt} \right]$$

Example 7.9 Selecting the Proper Form of the PID Algorithm

Problem Statement. Write the digital version of the position form of the PID algorithm similar to Equation 7.3.4 for Example 3.6 (Figure 7.6.5) and assume that the control valve on the steam line for the CSTR has an air-to-close actuator with a valve positioner. Use the form that is not susceptible to derivative kick.

Solution. From Figure 7.6.5, as the steam flow to the heat exchanger on the reactor is increased, the outlet temperature from the CSTR increases; therefore, the process gain is positive. Because an air-to-close valve actuator with a valve positioner, which is a direct-acting actuator, is used, Table 7.1 indicates that a reverse-acting controller should be used. To eliminate derivative kick, the derivative action should be based on the process measurement. The resulting digital form of the PID algorithm is given by

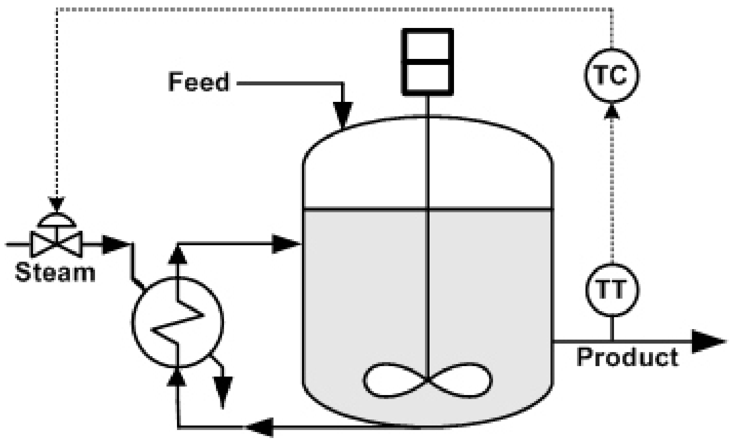


Figure 7.6.5 Control diagram for an endothermic CSTR.

$$c(t) = \bar{c} - K_c\left[e(t) + \frac{\Delta t}{\tau_I}\sum_{i=1}^{n} e(i\,\Delta t) - \tau_D \frac{y_s(t) - y_s(t-\Delta t)}{\Delta t}\right]$$

Example 7.10 Selecting the Proper Form of the PID Algorithm

Problem Statement. Write the velocity form of the PID algorithm similar to Equation 7.4.2 for Figure 7.6.6 from Example 3.9 (dissolved oxygen process). Use the form that is not susceptible to derivative kick or proportional kick.

Solution. From Figure 7.6.6, as the air flow to the bio-reactor is increased, the concentration of oxygen in the broth increases; therefore, the process gain is positive. Because a variable speed air compressor, which is a direct-acting actuator, is used, from Table 7.1 a reverse-acting controller should be used. To eliminate derivative kick, the derivative should be based on the measured value of the CV and to eliminate proportional kick, the proportional action should also be based on the measured value of the CV. The resulting PID algorithm is given by

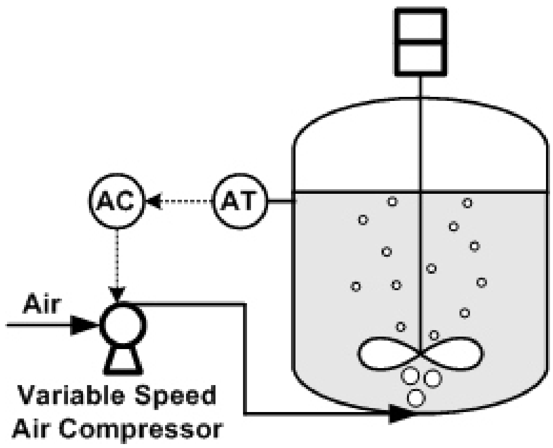


Figure 7.6.6 Schematic of a DO process.

$$\Delta c(t) = K_c\left[y_s(t-\Delta t) - y_s(t) + \frac{\Delta t}{\tau_I} e(t) - \tau_D \frac{y_s(t) - 2y_s(t-\Delta t) + y_s(t-2\Delta t)}{\Delta t}\right]$$

$$c(t) = c(t-\Delta t) + \Delta c(t)$$

Self-Assessment Questions

Q7.6.1 What is a reverse-acting final control element?

Q7.6.2 When a reverse-acting final control element is used, for what cases is the controller reverse acting?

Q7.6.3 What would happen if a direct-acting or reverse-acting controller is incorrectly specified?

Self-Assessment Answers

Q7.6.1 For a reverse-acting final control element, the MV decreases when the signal to the final control element is increased.

Q7.6.2 When a reverse-acting final control element is used, a reverse-acting controller results when the process gain is negative (see Table 7.1).

Q7.6.3 If a direct-acting or reverse-acting controller is incorrectly specified, a positive feedback controller will result and the control loop will go unstable when the loop is turned on.

Self-Assessment Problem

P7.6.1 Write the velocity form of the PID algorithm similar to Equation 7.4.1 for Example 3.5 and assume that the control valve on the exit line has an air-to-open actuator without a valve positioner. Use the form that is not susceptible to derivative kick, but is susceptible to proportional kick

Self-Assessment Answer

P7.6.1 Because Example 3.5 has a negative process gain and the actuator is direct-acting, a direct-acting PID controller is used:

$$c(t \quad t) \quad c_0 \quad K_c \quad e(t \quad t) \quad \frac{t}{_I} \; {}^{n\ 1}_{i\ 1} \, e(i \quad t) \quad {}_D \frac{y_s(t \quad t) \quad y_s(t \quad 2 \quad t)}{t}$$

$$c(t) \quad c(t \quad t) \quad c(t)$$

7.7 Analysis of P, I and D Action

The PID controller can be represented by

$$c(t) \quad \bar{c} \quad K_c \quad e(t) \quad \frac{1}{_I} \; {}^{t}_{0} \, e(t)\,dt \quad {}_D \frac{d\,e(t)}{dt} \qquad \textbf{7.7.1}$$

where $c(t)$ is the controller output at time t, $\bar{c}$ is the controller output when the controller is turned on at time t=0, $e(t)$ is the error from setpoint [$e(t) \quad y_{sp} \quad y_s(t)$], K_c is the controller gain, ${}_I$ is the controller reset time and ${}_D$ is the controller derivative time. The first term inside the bracket in Equation 7.7.1 provides proportional action while the second and third terms provide integral and derivative action, respectively. K_c, ${}_I$ and ${}_D$ are the PID tuning parameters. Figures 7.7.1 and 7.7.2 show the response of a PID controller to a setpoint change and to a disturbance upset, respectively. Figures 7.7.1a and 7.7.2a show the CV response in relation to the setpoint while Figures 7.7.1.b and 7.7.2b show the MV determined by the PID controller. Finally, Figures 7.7.1c and 7.7.2c show the proportional, integral and derivative components of the control signal. That is, if the components shown in Figures 7.7.1c and 7.7.2c were added together, the result would equal the MV result shown in Figure 7.7.1b and 7.7.2b, respectively. The analysis of proportional, integral and derivative action will refer to these results.

The transfer function for a PID controller (See Example 5.9) is given as

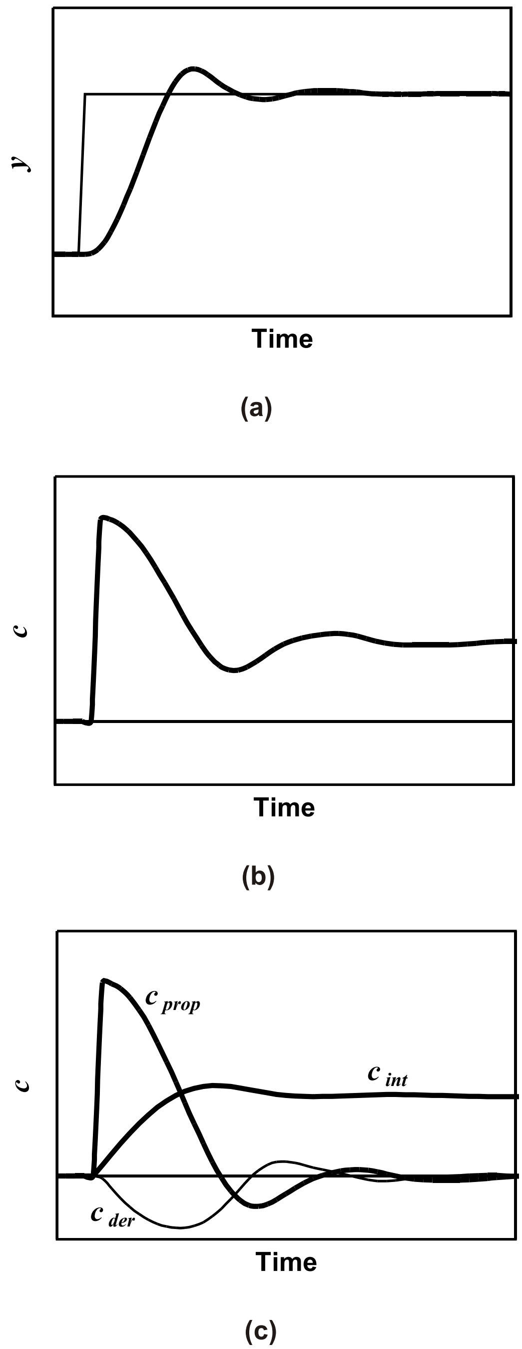


Figure 7.7.1 Response of a PID controller to a setpoint change. (a) y and y_{sp}, (b) c and (c) proportional action (c_{prop}), integral action (c_{int}) and derivative action (c_{der}).

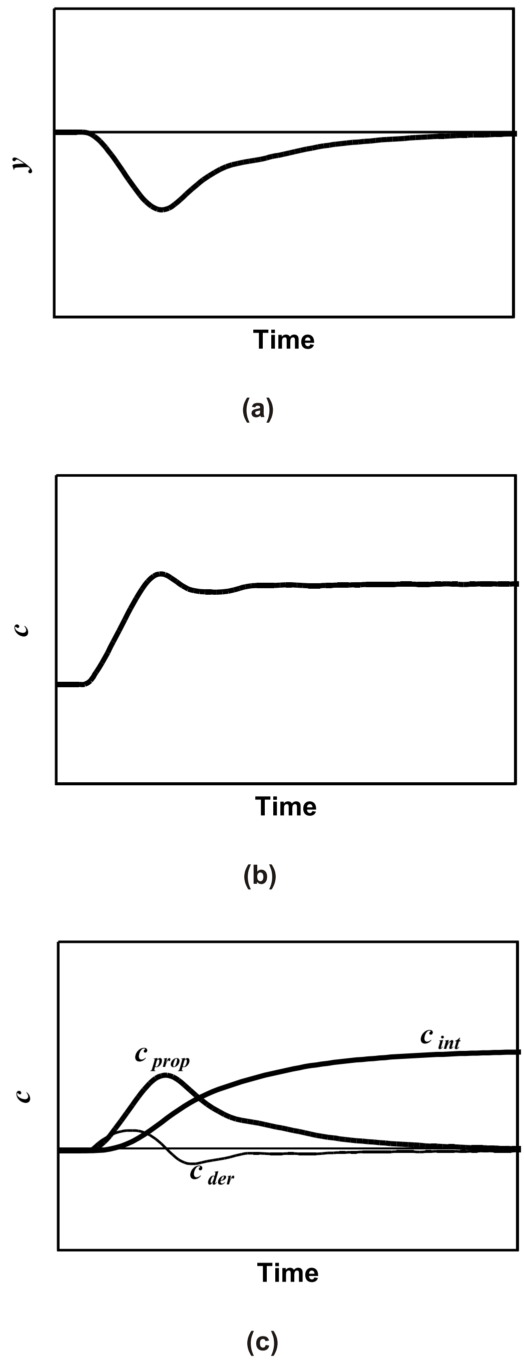


Figure 7.7.2 Response of a PID controller to a disturbance upset. (a) y and y_{sp}, (b) c and (c) proportional action (c_{prop}), integral action (c_{int}) and derivative action (c_{der}).

$$G_c(s) = \frac{C(s)}{E(s)} = K_c\left(1 + \frac{1}{\tau_I s} + \tau_D s\right) \qquad \textbf{7.7.2}$$

To investigate the fundamental characteristics of proportional, integral and derivative action, we assume that a process is first order and that the sensor and the actuator respond much faster than the process (i.e., $G_a(s) = G_s(s) = 1$). Therefore, the combination of the actuator, process and sensor is represented as a first-order process, i.e.,

$$G_a(s)G_p(s)G_s(s) = \frac{K_p}{\tau_p s + 1}$$

The closed-loop transfer function for a setpoint change (Equation 7.2.7) for this system will be used to analyze the fundamental characteristics of proportional, integral and derivative action.

Proportional action. From Equation 7.7.1, feedback control based on **proportional action**,

$$c(t) = \bar{c} + K_c e(t)$$

takes action based on the current error from setpoint where K_c is the **controller gain**. The larger the error from setpoint, the larger the control action. From Equation 7.7.2, the transfer function for a proportional controller is

$$G_c(s) = K_c$$

The closed-loop transfer function for a setpoint change (Equation 7.2.7) for the case of a first-order actuator/process/sensor becomes

$$\frac{Y(s)}{Y_{sp}(s)} = \frac{G_c(s)G_p(s)}{G_c(s)G_p(s) + 1} = \frac{\dfrac{K_c K_p}{\tau_p s + 1}}{\dfrac{K_c K_p}{\tau_p s + 1} + 1} = \frac{K_c K_p}{\tau_p s + 1 + K_c K_p}$$

Putting this result into the standard form for a first-order process results in

$$\frac{Y(s)}{Y_{sp}(s)} = \frac{\dfrac{K_c K_p}{K_c K_p + 1}}{\dfrac{\tau_p}{K_c K_p + 1} s + 1} \qquad \textbf{7.7.3}$$

This result can be used to identify several fundamental characteristics of proportional action:

1. The closed-loop time constant $[\ _p / (K_c K_p \quad 1)]$ is smaller than the open-loop time constant, $_p$. That is, proportional action makes the closed-loop process respond faster than the open-loop process. **Increasing the speed of the response of the process is the primary benefit of proportional control.**
2. The closed-loop response of a first-order process remains first-order. In general, proportional action does not change the order of the process.
3. The steady-state closed-loop gain is not equal to unity. Figure 7.7.3 shows setpoint changes for three different values of K_c. Note that the steady-state closed-loop response differs from the setpoint value, which indicates **offset**. Offset is the error between the setpoint and the steady-state CV value after an input change. Applying the final value theorem to Equation 7.7.3 yields the value of the CV after a setpoint change, y_{ss},

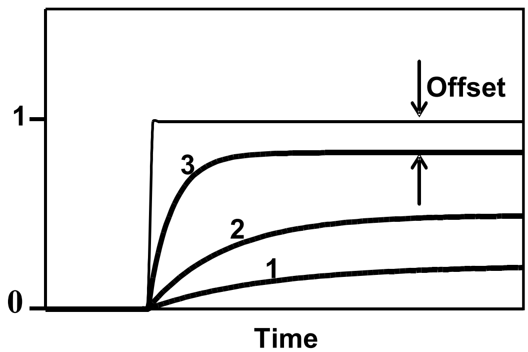


Figure 7.7.3 The effect of K_c on the response of a P-only controller for a setpoint change for a first-order process. K_c is increased from "1" to "3".

$$y_{ss} \quad \frac{K_c K_p y_{sp}}{K_c K_p \quad 1}$$

Then using the definition of offset,

$$\text{Offset} = y_{sp} \quad y_{ss} \quad y_{sp} \ 1 \quad \frac{K_c K_p}{K_c K_p \quad 1} \quad y_{sp} \ \frac{1}{K_c K_p \quad 1}$$

Note that the offset decreases as K_c is increased, which is consistent with Figure 7.7.3. The same three conclusions for proportional action can be reached if a disturbance upset were considered instead of a setpoint change.

Figures 7.7.1c and 7.7.2c show the portion of a controller output signal resulting from proportional action (u_{prop}) for a PID controller after a setpoint change and a disturbance upset, respectively. The proportional control action is positive when y is less than y_{sp} and negative when y is greater than y_{sp} and its magnitude is directly proportional to the error from setpoint. Initially, the setpoint change causes a spike in proportional action (Figure 7.7.1c) but, as y moves toward the setpoint, the proportional action is reduced and eventually diminishes as y settles at the setpoint. Note that there is no offset in these cases, which results from the integral action in the PID controller and integral action is considered next.

Integral action. From Equation 7.7.1, feedback control based on **integral action**, i.e.,

$$c(t) \quad \bar{c} \quad \frac{K_c}{_I} \quad {}_0^t \ e(t)\,dt$$

acts on the long-term error from setpoint. $_I$ is the **reset time** (**integral time**), which is the tuning parameter for integral action and has units of time. Consider Equation 7.7.1 after a setpoint change or a disturbance has affected the process. Further, assume that the process has reached steady-state conditions at the specified setpoint after the effects of the disturbances or the setpoint changes have been absorbed by the controller. Under these conditions, $e(t)$ and $de(t)/dt$ are both zero, which indicates that the proportional action and derivative action are also zero at steady-state conditions at the new setpoint. Because a setpoint change or a disturbance change has occurred, $c(t)$ must be significantly different from $\bar{c}$; therefore, the integral term is responsible for providing the incremental change at steady-state from $\bar{c}$ necessary to maintain operation at the new operating condition (See Figures 7.7.1c and 7.7.2c). As a result, integral action is a critically important feature of PID feedback control. From Equation 7.7.2, the transfer function for an integral-only controller is

$$G_c = \frac{K_c}{\tau_I s}$$

The closed-loop transfer function for an I-only controller for a setpoint change (Equation 7.2.7) applied to a first-order process becomes

$$\frac{Y(s)}{Y_{sp}(s)} = \frac{G_c(s)G_p(s)}{G_c(s)G_p(s)+1} = \frac{\left(\dfrac{K_c}{\tau_I s}\right)\left(\dfrac{K_p}{\tau_p s+1}\right)}{\left(\dfrac{K_c}{\tau_I s}\right)\left(\dfrac{K_p}{\tau_p s+1}\right)+1} = \frac{K_c K_p}{\tau_I \tau_p s^2+\tau_I s+K_c K_p}$$

Putting this result into the standard form for a second-order process (i.e., Equation 6.4.2) results in

$$\frac{Y(s)}{Y_{sp}(s)} = \frac{1}{\dfrac{\tau_I \tau_p}{K_c K_p}s^2+\dfrac{\tau_I}{K_c K_p}s+1} = \frac{1}{\tau_n^2 s^2+2\zeta\tau_n s+1}$$

By comparing this equation with the standard second-order form, the closed-loop natural period, τ_n, is given as

$$\tau_n = \sqrt{\frac{\tau_I \tau_p}{K_p K_c}}$$

and solving for the closed-loop damping factor, ζ, yields

$$\zeta = \frac{1}{2}\sqrt{\frac{\tau_I}{\tau_p K_c K_p}}$$

and the closed-loop gain is 1. These results, along with the previous analysis, provide several fundamental characteristics of integral action.

1. Because the gain of the closed-loop transfer function is 1, there is no offset at steady state. **Eliminating offset is the primary advantage provided by integral action.**
2. All the steady-state corrections for disturbances or setpoint changes must come from integral action for offset-free operation.
3. Integral action increases the order of the process dynamics by 1.

4. Based upon the equations for $_n$ and , as $_I$ is decreased, the process becomes faster but at the expense of larger overshoots and more sustained oscillations. For example, assume that $K_c\ K_p \quad 1$. Then

$$_n \quad _p\sqrt{\frac{_I}{_p}} \qquad\qquad 0.5\sqrt{\frac{_I}{_p}}$$

Figure 7.7.4 shows and $_n$ as a function of the ratio of $_I/\ _p$ for this case. Note that increasing the amount of integral action (decreasing $_I$) results in a faster-responding feedback process, but increases the degree of oscillatory behavior (i.e., reduces the value of).

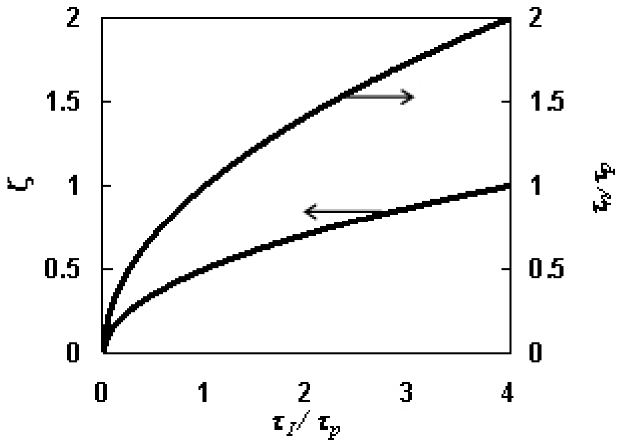


Figure 7.7.4 The effect of the ratio of the reset time to the process time constant on the closed-loop damping factor and $_n$. (K_cK_p=1).

Figures 7.7.1c and 7.7.2c show the portion of the controller output resulting from integral action for a PID controller for a setpoint change and a disturbance upset, respectively. Note that the peaks and valleys in c_{int} occur when y crosses y_{sp}. Also note that, as the process settles to the setpoint, all the control action comes from integral action because c_{int} is the only non-zero component of the control action when the system settles to setpoint.

Derivative Action. From Equation 7.7.1, feedback control based on **derivative action** is given by

$$c(t) \quad \bar{c} \quad K_c\ _D \frac{d\,e(t)}{dt}$$

where $_D$ is the **derivative time**, which is the tuning parameter for derivative action and has the units of time. From Equation 7.7.2, the transfer function for derivative action is

$$G_c(s) \quad K_c\ _D\, s$$

The closed-loop transfer function for a derivative-only controller for a setpoint change (Equation 7.2.7) applied to a first-order process becomes

$$\frac{Y(s)}{Y_{sp}(s)} \quad \frac{G_c(s)G_p(s)}{G_c(s)G_p(s) \quad 1} \quad \frac{K_c\ _D s \frac{K_p}{_p s \quad 1}}{K_c\ _D s \frac{K_p}{_p s \quad 1} \quad 1} \quad \frac{K_c K_p\ _D\, s}{(K_c K_p\ _D \quad _p)s \quad 1}$$

which indicates that derivative action does not change the order of the process or eliminate offset, but does increase the time constant for a first-order process.

It is also instructive to consider the effect of derivative action applied to a second-order process. The closed-loop transfer function for a setpoint change applied to this case yields

$$\frac{Y(s)}{Y_{sp}(s)} = \frac{G_c(s)G_p(s)}{G_c(s)G_p(s)+1} = \frac{K_c\tau_D s\dfrac{K_p}{\tau_n^2 s^2 + 2\zeta\tau_n s + 1}}{K_c\tau_D s\dfrac{K_p}{\tau_n^2 s^2 + 2\zeta\tau_n s + 1} + 1}$$

Rearranging

$$\frac{Y(s)}{Y_{sp}(s)} = \frac{K_c K_p \tau_D s}{\tau_n^2 s^2 + (2\zeta\tau_n + K_c K_p \tau_D)s + 1}$$

For a second-order process, τ_n remains unchanged while the closed-loop damping factor is larger than the open-loop damping factor, which indicates that the closed-loop process will have the same dynamic response, but with less oscillations. These results indicate several fundamental characteristics of derivative action:

1. Derivative action reduces the oscillatory nature of a feedback system. **Reducing the oscillatory nature of a feedback response is the primary advantage of derivative action.**
2. Derivative action does not perform well for noisy sensor readings
3. Derivative action does not change the order of the process.
4. Derivative action does not eliminate offset.

Figures 7.7.1c and 7.7.2c show the portion of the output from a PID controller resulting from derivative action (c_{der}) for a setpoint change and a disturbance upset, respectively. Note that c_{der} is zero at the peaks and valleys of y because it is directly related to the local slope of y. Derivative action acts to oppose the slope of the CV in an effort to stabilize the feedback process regardless of whether the CV is moving away from or toward the setpoint. As the slope of the CV increases, derivative control action also increases in an effort to reduce the slope to allow for more gradual changes in the CV.

Self-Assessment Questions

Q7.7.1 What are the primary benefit and shortcoming of proportional action?

Q7.7.2 What are the primary benefit and shortcoming of integral action?

Q7.7.3 What is the primary benefit of derivative action?

Self-Assessment Answers

Q7.7.1 The primary benefit of proportional action is that it increases the speed of the closed-loop response and the primary shortcoming is that proportional action results in offset for most systems.

Q7.7.2 The primary benefit of integral action is that it eliminates offset and the primary shortcoming is that it increases the oscillatory nature of the closed-loop response.

Q7.7.3 The primary benefit of derivative action is that it reduces the oscillatory response of the closed-loop response.

7.8 Choosing the Proper Mode of a PID Controller

Not all PID-type controllers use proportional, integral and derivative action together. For example, a P-only controller, which is a type of PID controller, uses only proportional action, i.e.,

$$c(t) \quad \bar{c} \quad K_c e(t)$$

On the other hand, a PI controller uses proportional and integral action, but does not use derivative action, i.e.,

$$c(t) \quad \bar{c} \quad K_c \quad e(t) \quad \frac{1}{{}_{I}} \; {}^{t}_{0} \, e(t)dt$$

Likewise, a PD controller uses proportional and derivative action, but does not use integral action:

$$c(t) \quad \bar{c} \quad K_c \quad e(t) \qquad {}_{D} \frac{de(t)}{dt}$$

And of course, a PID controller uses all three modes (Equation 7.4.1).

When choosing between P-only, PI, PD or PID controllers, you should consider the dynamics of the combined actuator/process/sensor system and the objectives of the control loop. For conventional control loops in the CPI, it has been estimated[1] that approximately 93% use PI controllers, 2% use P-only controllers and 5% use PID controllers. Probably less than one percent of industrial control loops use PD controllers. The following guidelines, based on process dynamics and control objectives, can be used to choose the proper mode for PID controllers in the CPI and bio-tech industries.

P-Only Control. P-only control is used for fast-responding processes (i.e., processes that are not sluggish) and for which some degree of offset is acceptable or for integrating processes (e.g., most level control processes). A sluggish process is characterized by the fact that the process does not respond quickly to changes in the MV (i.e., not a first-order-like response). Typical applications for a P-only controller are for level controllers and pressure controllers. At times certain control loops that should use P-only controllers, but instead use a typical PI or PI with a relatively small amount of integral action.

PI Control. PI controllers are used for processes that are not sluggish and for which it is necessary to have offset-free operation. Typical applications are flow control, pressure control, temperature control and composition control.

PID Control. PID controllers are useful for certain sluggish processes. Typical applications are temperature control and composition control. Because of the inertia of a sluggish process, the process exhibits a tendency to oscillate (i.e., a low damping factor) under PI control. Derivative action reduces the oscillatory nature of the response and allows more proportional action to be used, both of which contribute to improved control performance. A key issue here is to determine whether a process is sufficiently sluggish to warrant a PID controller. Assume that a FOPDT model has been fit to an open-loop step test for the process in question. If the resulting deadtime, $_p$, and time constant, $_p$, are such that

$$\frac{_p}{_p} \quad \frac{1}{2}$$

the process is not sufficiently sluggish to warrant a PID controller. **If**

$$\boldsymbol{\frac{_p}{_p} \quad 1}$$

the process is sufficiently sluggish that a PID controller should offer significant benefits over a PI controller. For

$$\frac{1}{2} \quad \frac{_p}{_p} \quad 1$$

either PI or PID can be preferred. In the event that a FOPDT model is not available, excessive oscillations of a PI controller or a sluggishly responding PI controller are indications that a PID controller may provide improved control performance. Measurements of the CV with significant noise levels can make the use of derivative action ineffective because of the sensitivity of the derivative to noise on the measurement. That is, because the measurements of the CV have so much noise, the lag added by a filter can negate any benefit produced by the derivative action.

PD Control. PD controllers are used for sluggish processes that behave as an integrator. If significant integral action is used on such a process, unstable behavior results. Exothermic CSTRs are an example of a process that benefits from PD control.

Example 7.11 Selection of the Proper Mode of a PID Controller

Problem Statement. Consider the wastewater neutralization process shown in Figure 7.8.1. The objective of this process is to maintain the pH of the effluent stream at a pH value of 7 (neutrality). Determine the proper mode of a PID controller for this process if the residence time for the reactor is 3 minutes and the time constant for the pH sensor/transmitter is 5 s.

Solution. Because the objective is to maintain neutrality, integral action should be used. Because primarily strong acid/strong base reactions are occurring in the reactor, the reactions can be assumed instantaneous. Therefore, the reactor should behave as a first-order process due to the mixing of the feed streams in the reactor volume. Because the sensor dynamics and the dynamics of the final control element are much faster than the process in this case, the overall dynamic behavior should be similar to a first-order process; therefore, the effective deadtime-to-time constant ratio of this

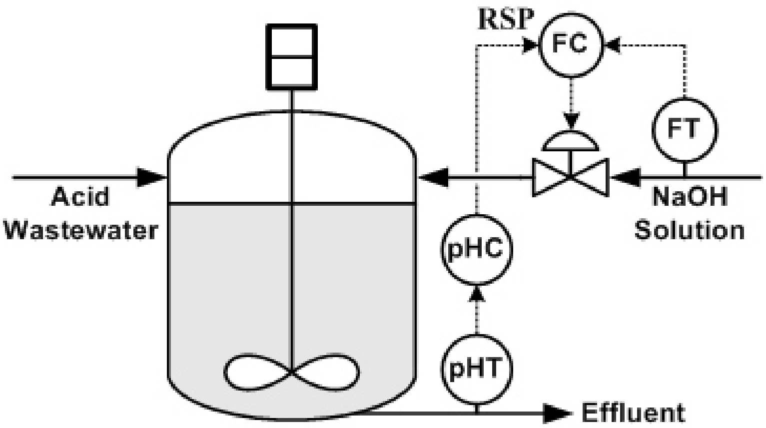


Figure 7.8.1 Control diagram for a pH control system applied to a wastewater neutralization process.

system should be relatively small. As a result, a PI controller should be selected for this application.

Example 7.12 Evaluation of a P-Only Controller Applied to an Integrating Process.

Problem Statement. Evaluate the performance of a P-only controller applied for level control of a tank (Figure 7.8.2). Assume that the transfer function for this process (see Equation 6.5.2) is given by

$$G_p(s) = \frac{L(s)}{F_{in}(s)} = \frac{1}{A_c s}$$

Also, assume $G_a(s) = G_s(s) = 1$.

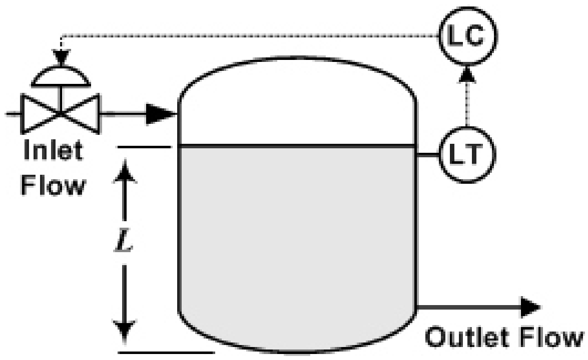


Figure 7.8.2 Control diagram for a tank level control system.

Solution. Determine the closed-loop transfer function for setpoint changes for the P-only controller (Equation 7.2.7).

$$\frac{L(s)}{L_{sp}(s)} = \frac{G_c(s)G_p(s)}{G_c(s)G_p + 1} = \frac{K_c \dfrac{1}{A_c s}}{K_c \dfrac{1}{A_c s} + 1} = \frac{1}{\dfrac{A_c}{K_c} s + 1}$$

Even though a P-only controller is applied, the gain of the closed-loop transfer function is unity, indicating offset-free operation. This results because the process is an integrator because the denominator of the transfer function contains a zero pole (i.e., s=0). As a result, the closed-loop response is first order. Also, note that the time constant for the closed-loop response varies inversely with the controller gain.

Self-Assessment Questions

Q7.8.1 What is the most commonly used mode of a PID controller?

Q7.8.2 How do you decide between using a P-only and a PI controller?

Q7.8.3 How do you decide between using a PI and a PID controller?

Self-Assessment Answers

Q7.8.1 Over 90% of the PID controllers in the CPI use the PI mode.

Q7.8.2 If a control loop is fast responding, a PI controller should be used if offset-free operation is required. Otherwise, a P-only controller should be used. In addition, a P-only controller should be used for integrating processes.

Q7.8.3 If a process is sufficiently sluggish, a PID controller should be applied instead of a PI controller. Quantitatively, if the deadtime to time constant ratio of the open-loop process is greater than 0.5, PID control may provide significant improvement in the closed-loop performance. If this ratio is greater than one, a PID controller should provide significant performance improvement.

7.9 Commonly Encountered Control Loops

In this section, several control loops commonly encountered in the CPI and the biotechnology industries are analyzed. The relevant control characteristics of each loop are discussed from an actuator/process/sensor point of view and the problem of selecting the proper mode for a PID controller is addressed for each case.

Flow Control Loop. A flow control loop is the most common control loop used in the CPI. Most of the control loops in the CPI, other than a flow control loop itself, use a setpoint to a flow controller as their MV in a cascade control arrangement (Chapter 12). A schematic of a flow control loop is shown in Figure 7.9.1. An orifice meter/differential pressure sensor is used to measure the flow rate and the actuator is composed of an I/P converter, instrument air and assembled control valve. The objective of this control loop is to maintain the flow rate at the setpoint for changes in the upstream and downstream pressures and for changes in the setpoint to the flow controller.

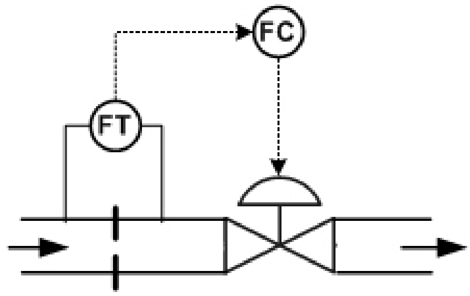


Figure 7.9.1 Control diagram for a flow control loop.

The dynamics of the process (i.e., flow rate change as the result of a change in the valve stem position) and the sensor (i.e., change in the measured pressure drop for a change in the flow rate) are quite fast compared with the dynamics of the control valve (i.e., change in valve stem position for a change in signal to the control valve). Because the overall process is relatively fast and accurate control to setpoint is required, a PI controller is the proper choice for most flow control applications.

The most interesting aspect of flow control loops is that, in spite of the fact that industrial control valves have a deadband of 10% to 25%, flow control loops are able to precisely meter the average flow rate typically to within a deadband of 0.5% to as low as 0.1%. Figure 7.9.2 shows a plot of the actual valve position and the instrument air pressure delivered to the valve in an open-loop case as a function of time. This behavior is caused by deadband in the valve primarily from friction between the valve stem and the valve packing. As a result of the drag of the packing on the valve stem, a minimum force is necessary to cause the valve to move and when it does move, it breaks loose and significant valve travel results, thus causing a significant deadband. From Figure 7.9.2, note that in the latter stages of the time plot the stem position remains constant even though the air signal to the valve actuator decreases significantly and increases significantly.

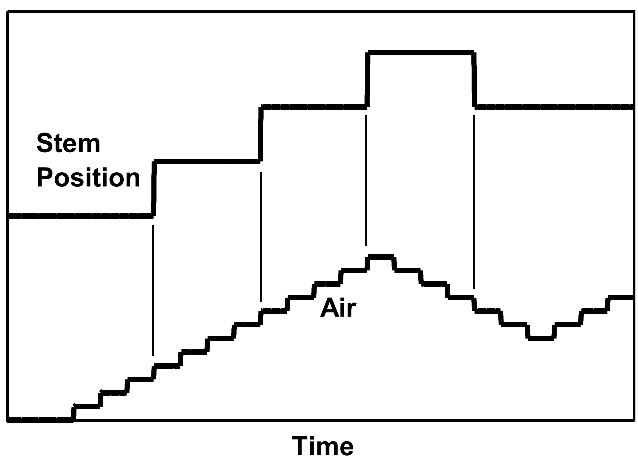


Figure 7.9.2 Valve stem position and air pressure to the valve actuator for a valve with significant deadband.

To understand how a flow control loop can very accurately control the flow rate using such an imprecise actuator, consider the measured flow rate and specified flow rate shown in Figure 7.9.3. The significant variation in the flow rate (i.e., sustained oscillations) is due to the combined effect of the deadband of the valve and feedback control. On the other hand, the average flow is precisely controlled due to the high frequency feedback control provided by the flow controller or a valve positioner. Because the period of the flow variations is in the range of a few seconds and most chemical processes have time constants several minutes or larger, the process is usually sensitive only to the average flow and not to flow fluctuations; therefore, flow control loops can provide very precise metering of the **average flow rate** in spite of the fact that a very imprecise actuator is used. Because the primary objective of a flow control loop is to control the average flow rate and short-term deviations from setpoint are unimportant, PI flow controllers are usually tuned with much more integral action than proportional action. If a valve positioner is used, the valve positioner provides the high-frequency feedback necessary to counteract the detrimental effects of the deadband of the control valve on the metering precision of the average flow rate. That is, the high-gain P-only controller applied by the valve positioner opens and closes the valve in a manner similar to the results shown in Figure 7.9.3. A PI flow controller applied to a control valve with a positioner eliminates the offset for which the positioner does not account and absorbs unmeasured disturbances such as changes in upstream and downstream pressures.

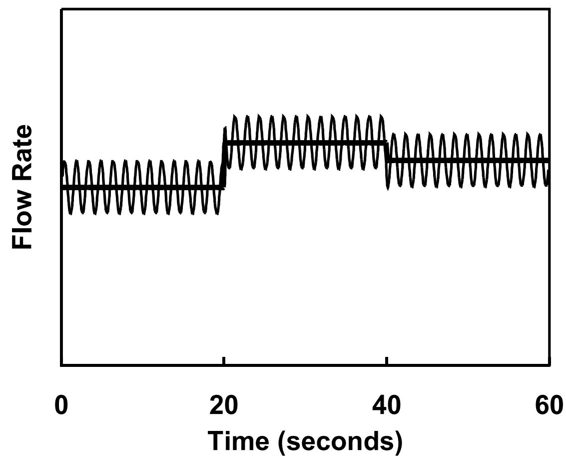


Figure 7.9.3 Measured flow rate and the specified average flow rate for a flow control loop or a valve with a positioner. Thin line - measured flow rate. Thick line - flow rate setpoint.

Level Control Loop. Level control loops are used to control the liquid levels in the accumulator and reboiler of distillation columns, steam boilers, reactors and intermediate storage tanks. A schematic for a level control loop used to control the level in a tank is shown in Figure 7.9.4. A differential pressure sensor is used to measure the liquid level and the actuator is a control valve on the line to the tank. The MV determined by the level controller is the valve position for the control valve on the line feeding the tank. The objective of this loop is to maintain the level within a certain range, for example, from 30 to 40% of full range (i.e., 0% when the tank is empty and 100% when the tank is full) for changes in the feed rate to the tank and changes in operating conditions as well as reducing the effect of the variations in the inlet flow on downstream operations (see Example 7.13).

The dynamics of the sensor are quite fast and the dynamics of the actuator are usually fast compared with the dynamics of the process (i.e., percentage level change for a change in flow entering the tank). Because level systems are usually integrating processes, the rate of change of the level depends upon the change in flow rate and the cross-sectional area of the vessel. For a typical system under open-loop conditions, a 5% level change can occur in about one minute for about a 10% change in feed rate to the tank. Thus, the response of the actuator/process/sensor system is typically controlled by the process dynamics. Because the overall process is not generally sluggish, a P-only controller is the proper choice if the level process is an integrating process (Example 7.12), which is usually the case. Using a PI controller for an integrating process is not necessary and will only contribute to the oscillatory nature of the closed-loop

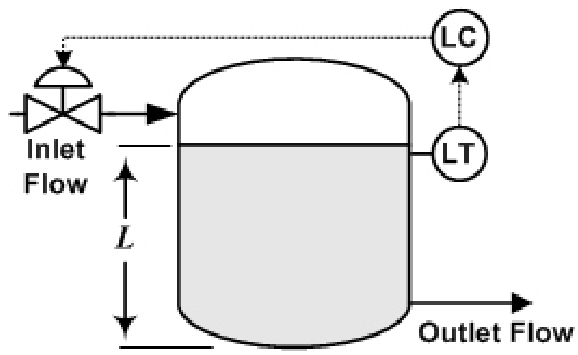


Figure 7.9.4 Control diagram for a level controller applied to a tank.

response. If the level process is self-regulating (i.e., not an integrating process), a PI controller is required for offset-free operation.

Example 7.13 Attenuation of Inlet Flow Variations by Level Control in a Tank.

Problem Statement. As pointed out, one of the functions of level control on tanks is to reduce the variations in the inlet flow rate that are passed on to the outlet flow of the tank and the downstream processing units. Consider the tank shown in Figure 7.9.4. Assume that the nominal level in the tank is L_0, the cross section area of the tank is A_r, the density of the liquid in the tank is , a P-only control is used with a proportional gain of K_c, the inlet flow to the tank is given by F_0+Asin(t), which is given in units of mass per time and the setpoint for the level controller is equal to L_0. Develop an expression for the variation in the outlet flow in terms of parameters of this system.

Solution. The outlet flow can be expressed as F_0+A'sin(t) and the objective of this example is to derive an expression for A'. Performing an unsteady-state mass balance on this system yields

$$A\frac{dL}{dt} \quad F_{in} \quad F_{out} \quad F_0 \quad A\sin(\;\; t) \quad F_0 \quad K_c(L \quad L_0) \quad A\sin(\;\; t) \quad K_c(L \quad L_0)$$

with the initial conditions: at t=0 L=L_0. Following is the MATLAB code and results for the application of `dsolve`, which was introduced in Chapter 4, for the analytical solution of this problem.

```
function OdeSoln
syms L(t) Ar w den Kc L0 A
S=dsolve(den*Ar*diff(L)+Kc*L==A*sin(w*t)+Kc*L0,L(0)==L0)
end
```

```
(Kc^2*L0 + A*Kc*sin(t*w) + Ar^2*L0*den^2*w^2 - A*Ar*den*w*cos(t*w))/(Ar^2*den^2*w^2 + Kc^2) +
exp(-(Kc*t)/(Ar*den))*(L0 - (L0*Ar^2*den^2*w^2 - A*Ar*den*w + L0*Kc^2)/(Ar^2*den^2*w^2 + Kc^2))
```

Note that there are two primary terms in the solution. The second term is multiplied by exp(-K_c*t/A_r) and therefore will go to zero at long times. The first term contains two sinusoidal components (i.e., sin t and cos t) and these terms represent the variation in the outlet flow from the tank. Considering only these two terms yields

$$\frac{A K_c \sin(\;\; t) \quad A A_r \quad\;\; \cos(\;\; t)}{K_c^2 \quad A_r^2 \;\; ^2 \;\; ^2}$$

Then using the trigometric identity: $A\sin \quad t \quad B\cos \quad t \quad \sqrt{A^2 \quad B^2}\sin[\;\; t \quad\;\;]$ where $\quad \tan^{\;1}(B/A)$, the above term becomes

$$\frac{A\sqrt{K_c^2 \quad A_r^2 \;\; ^2 \;\; ^2}}{K_c^2 \quad A_r^2 \;\; ^2 \;\; ^2}\sin(\;\; t \quad\;\;) \quad \frac{A}{\sqrt{K_c^2 \quad A_r^2 \;\; ^2 \;\; ^2}}\sin(\;\; t \quad\;\;)$$

Therefore, the amplitude of the sinusoidal input flow changes entering the tank is reduced by a factor of $\sqrt{K_c^2 \quad A_r^2 \;\; ^2 \;\; ^2}$. Tanks also attenuate variations in composition due to mixing in the tank.

As an example of the application of this relationship, consider a P-only level control system applied to a 6-ft diameter horizontal cylindrical tank that is 10 feet long. For this case for water as the liquid in the tank, the density is equal to 62.4 lb_m/ft^3 and the cross-sectional area at 20% full is 48 ft^2. Using the level tuning procedure for P-only control presented in Section 9.13, the controller gain is 0.463 lb_m/ft-s. Therefore, the attenuation factor for this case is equal to

$$A \;\; / A \quad 1/\sqrt{K_c^2 \quad A_r^2 \;\; ^2 \;\; ^2} \quad 1/\sqrt{0.463^2 \quad (48 \quad 62.4)^2 \;\; ^2}$$

For this case, by comparing the magnitude of the two terms under the square root sign, you can see that the attenuation factor for higher frequencies effective eliminates these variations (e.g., for a period of 5 minutes per cycle, the attenuation factor is 250) and the controller gain K_c has little effect. On the other hand, changes on the order of one cycle per 12 minutes, the controller gain begins to have an effect on the attenuation factor.

Pressure Control Loop. Pressure control loops are used to maintain system pressure for distillation columns, reactors and other process units. A pressure control loop for maintaining overhead pressure in a column is shown in Figure 7.9.5. The actuator is a control valve on the vent line and the sensor is a pressure sensor mounted on the top of the column. Note that the output from the pressure controller goes directly to the control valve on the vent line. The objective of this loop is to maintain the column overhead pressure at or near setpoint for changes in condenser duty and changes in vapor flow rate up the column.

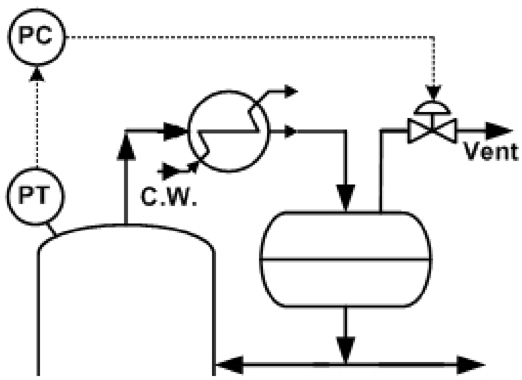


Figure 7.9.5 Control diagram for a pressure controller for the overhead of a distillation column.

The pressure sensor is quite fast while the process (change in pressure for a change in vent valve stem position) and the actuator are generally the slowest elements in the feedback system, making this a relatively fast-responding process. Pressure systems are many times integrating processes, and therefore, only requires a P-only controller for offset-free operation. Likewise, a self-regulating pressure system will required a PI controller for offset-free operation.

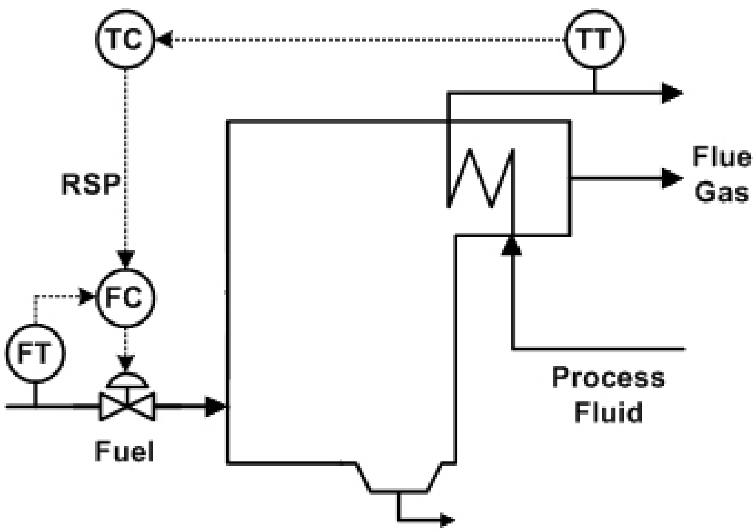


Figure 7.9.6 Control diagram for a temperature controller applied to a gas-fired heater.

Temperature Control Loop. Temperature control loops can be applied to control the temperature of a stream exiting a heat exchanger, the temperature of a tray in a distillation column and the temperature of a CSTR. Figure 7.9.6 shows a schematic of a temperature controller applied to control the temperature of a process stream leaving a gas-fired heater. The sensor is a RTD element or a thermistor placed in a thermowell located in the process line leaving the heater and the actuator is a flow control loop on the gas line to the heater. The objective of the temperature control loop is to maintain the temperature of the exiting process stream at setpoint in the face of changes in the temperature of the process stream entering the heater and changes in the heating value of the gas.

The dynamics of the actuator are generally much faster than the dynamics of the process (i.e., change in outlet process temperature for a change in gas flow rate to the heater) and the sensor, which typically has a dynamic time constant between 6 and 20 seconds for a properly installed RTD or a thermistor. The process fluid entering the gas-fired heater flows by plug flow through the furnace tubes that are exposed to high-temperature combustion gas. There is a thermal lag associated with changing the temperature of the metal of the furnace tubes as well as transport delay caused by plug flow through the heater tubes. The transport delay and resulting overall process deadtime increases as the feed rate of the process fluid is reduced. As a result, the process can be sluggish, particularly for low feed rate operations. Because the heater is likely to behave as a sluggish process, a PID controller should be selected in this example. Excessive sensor noise can make the use of derivative action ineffective, which is why an RTD or a thermistor would be preferred in this application over a TC. If this process were less sluggish, a PI controller can be preferable. The guidelines presented in the previous section should be used to determine if the process is sufficiently sluggish to warrant the use of PID control.

Composition Control Loop. Composition control loops are used to keep products produced by distillation columns on specification, to maintain constant conversion in a reactor and to maintain oxygen levels in the flue gas of a boiler to minimize carbon monoxide emissions. Figure 7.9.7 shows a control diagram for a composition loop that controls the impurity level in the overhead product of a distillation column. The sensor is a gas chromatograph that samples the distillation product and the output of the controller for this loop is the setpoint for the reflux flow controller. The objective of the composition control loop is to keep the impurity level in the overhead product at setpoint during changes in the feed flow rate and feed composition to the column.

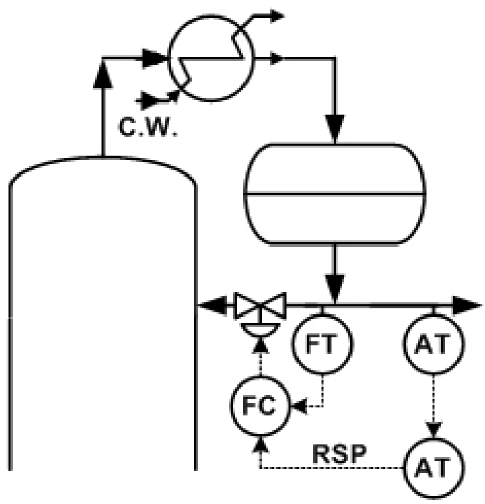


Figure 7.9.7 Control diagram for a composition controller for the overhead of a distillation column.

The actuator dynamics are relatively fast while the sensor typically can have three to ten minutes of analyzer delay. The process (i.e., change in impurity level in the overhead product for a change in the setpoint for the reflux flow controller) can be sluggish or relatively fast. For the relatively fast case, the sensor delay is usually significant; therefore, the dynamics of the process and sensor typically determine the dynamic response of this type of process. For this case, a PI controller should be used. On the other hand if the process is sluggish, the analyzer delay can be insignificant compared to the dynamics of the process. Therefore, the actuator/process/sensor for the sluggish case, the process dynamic is controlling and a PID controller may be preferred. Once again, the guidelines presented in the previous section should be used to determine if the process is sluggish enough to warrant the use of PID control.

Dissolved Oxygen Control Loop. Dissolved oxygen (DO) control loops are used to maintain uniform O_2 levels in aerobic bio-reactors. A control diagram for a DO control loop applied to a bio-reactor is shown in Figure 7.9.8. An ion-specific electrode is used to measure the dissolved O_2 in the reactor and the flow rate of air is used as the MV for the control loop. Note that the output of the DO controller goes directly to the variable speed air compressor, which is the actuator for this loop. Note that in industry, there are several ways of controlling DO, including use of variable speed agitation, oxygen gas feed, and/or combination of several MVs (see Chapter 18). Use of air flow rate is just one of several options. The objective of this controller is to maintain relatively constant DO levels in the bio-reactor in the face of changes in O_2 demand by the microorganisms in the bio-reactor.

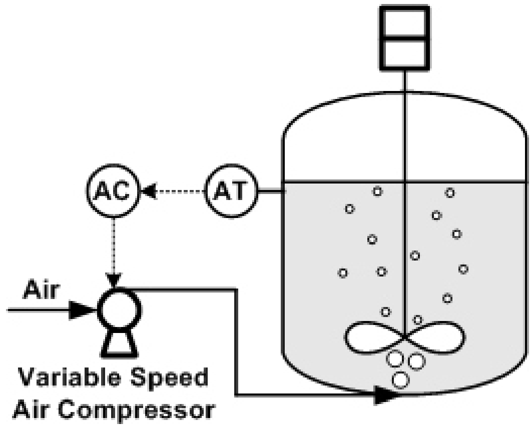


Figure 7.9.8 Control diagram for a DO controller applied to a bio-reactor.

The actuator is quite fast with a time constant less than 0.1 s while the sensor typically has a time constant between 2-5 s (Table 2.3). The process (i.e., change in DO concentration for a change in air flow rate) is the slowest element in the loop (i.e., the time constant for the process is typically 30-40 s), but overall the process is relatively fast responding. Therefore, the dynamics of the process and sensor determine the dynamic response of the actuator/process/sensor system. Therefore, because the overall feedback process is fast responding and control to setpoint is desirable, a PI controller is the PID mode of choice for this application.

Table 7.2 Summary of PID Controller Modes for Commonly Encountered Control Loops

	Are the dynamics of a system component important to the control loop performance?			Preferred Mode of a PID Controller
	Actuator	Process	Sensor	
Flow controllers				**PI**
Pressure controllers				**P-only or PI for non-integrating process**
Level controllers				**P-only or PI for a non-integrating process**
Temperature controllers				**PI or PID**
Composition Controllers				**PI or PID**
DO Controllers				**PI**

Self-Assessment Questions

Q7.9.1 What are the relative dynamics (i.e., fastest to slowest) of the actuator, process and sensor for a typical control valve-based flow control loop?

Q7.9.2 What are the relative dynamics (i.e., fastest to slowest) of the actuator, process and sensor for a typical temperature control loop?

Self-Assessment Answers

Q7.9.1 For a control valve-based flow control loop, the fastest to slowest elements are the process, sensor and actuator.

Q7.9.2 For a typical temperature control loop, the fastest to the slowest elements are the actuator, sensor and process.

7.10 Summary

- PID controllers are extremely flexible and are the most commonly used controllers in the CPI and bio-tech industries.
- The roots of the characteristic equation of the closed-loop transfer function define a major portion of the dynamic behavior of a PID feedback loop. The characteristic equation is given by

$$G_c(s)\,G_a(s)\,G_p(s)\,G_s(s) + 1 = 0$$

  where $G_c(s)$, $G_a(s)$, $G_p(s)$ and $G_s(s)$ are the transfer functions for the controller, actuator, process, and sensor, respectively.

- The PID algorithm can be implemented using the position form or the velocity form, but the velocity form is usually used in the process industries. The velocity form of the PID controller, which is not susceptible to derivative kick, but is susceptible to proportional kick, is given by

$$\Delta c(t) = K_c\left[e(t) - e(t-\Delta t) + \frac{\Delta t}{\tau_I}e(t) - \tau_D\,\frac{y_s(t) - 2y_s(t-\Delta t) + y_s(t-2\Delta t)}{\Delta t}\right]$$

$$c(t) = c(t-\Delta t) + \Delta c(t)$$

- The sign of the controller gain must be selected such that negative feedback is used (i.e., $K_c K_a K_p K_s > 0$). This is accomplish by selecting the controller as a reverse-acting controller [i.e., $c(t) = c(t-\Delta t) + \Delta c(t)$] for a positive controller gain or as a direct-acting controller [$c(t) = c(t-\Delta t) - \Delta c(t)$] for a negative controller gain assuming a direct-acting final control element.
- Proportional action makes the process respond faster but does not eliminate offset, integral action eliminates offset but increases the oscillatory nature of the feedback response and derivative action reduces the oscillatory nature of the closed-loop response but is susceptible to noisy sensor readings.
- When deciding among P-only, PI and PID controllers, you should consider the combined dynamic behavior of the actuator, process and sensor. Processes with non-sluggish dynamic behavior should use P-only control when offset is not important or for integrating processes and should use PI control when offset elimination is important. PID control should be used for sluggish processes, i.e., when the effective deadtime-to-time constant ratio is significant. Otherwise, a PI controller should be used.

7.11 References

1. Private communication, Jim Downs, Tennessee Eastman Company (Nov 1998).

2. Shuler, M.L. and F. Kargi, *Bioprocess Engineering*, Prentice-Hall, Englewood, NJ, p. 283 (1992).

7.12 Additional Terminology

$\bar{c}$ - the value of the controller signal to the actuator when a PID controller is turned on.

Characteristic equation - the equation formed by setting the denominator of the closed-loop transfer function equal to zero. The roots of the characteristic equation determine the dynamic behavior of the closed-loop system.

Control interval - the cycle time for applying control action, i.e., equal to the time interval between executions of the controller.

Controller gain (K_c) - tuning parameter for proportional action in a PID controller; it determines the aggressiveness of the controller.

Derivative action - control action that is proportional to the derivative of the CV.

Derivative kick - a spike in control action resulting from a setpoint change when the derivative is based on the error from setpoint.

Derivative-on-measurement - derivative action in a PID controller that is calculated based upon the slope of the measurement and, therefore, does not suffer from derivative kick.

Derivative time ($_D$) - the tuning parameter for derivative action in a PID controller.

Digital filter - a numerical running average that is used to reduce the effect of noisy sensor readings.

Direct-acting controller - a controller that subtracts the proportional, integral and derivative correction from the previous control action (velocity form) or from $\bar{c}$ (position form).

Disturbance rejection - controlling to setpoint in the face of disturbance upsets.

Integral action - control action that is proportional to the time integral of the error from setpoint.

Integral time ($_I$) - the tuning parameter for integral action in a PID controller.

Interactive PID controllers - an older form of PID control based upon the sequential application of derivative and integral action. Also known as “rate-before-reset”.

K_c - PID controller gain, tuning parameter for proportional action.

Loop gain - the product of the gain of the controller, the actuator, the process and the sensor.

Negative feedback - when the sensor reading is subtracted from the setpoint. Also, occurs when the loop gain is positive.

Offset - a sustained error from setpoint.

PB - the proportional band tuning constant, which is inversely related to K_c.

PID controller - a linear controller that applies proportional, integral and derivative action.

Position form - PID algorithm that calculates the total value of the controller output.

Positive feedback - when the sensor reading is added to the setpoint. Also, occurs when the loop gain is negative and indicates unstable closed-loop behavior.

Proportional action - control action that is proportional to the current error from setpoint.

Proportional band - a term that indicates the amount of proportional action used by a PID controller (inversely related to K_c).

Proportional kick - the sharp response of a PID-type controller to a setpoint change when the proportional term is based on the error from setpoint.

Rate-before-reset - an older form of PID control based on the sequential application of derivative and integral action.

Regulatory control - controlling to setpoint in the face of disturbance upsets.

Repeatability reduction ratio (R) - the repeatability of a sensor before filtering divided by the repeatability of the sensor after filtering.

Reset time ($_I$) - the tuning parameter for integral action in a PID controller.

Reverse-acting controller - a controller that adds the proportional, integral and derivative correction to the previous control action (velocity form) or to $\bar{c}$ (position form).

Servo control - controlling to setpoint in the face of setpoint changes.

Setpoint tracking - controlling to setpoint in the face of setpoint changes.

Velocity form of PID - PID algorithm that calculates the change in the controller output.

$_{D}$ - PID derivative time, the tuning parameter for derivative action.
$_{I}$ - PID reset time, the tuning parameter for integral action.

7.13 Preliminary Questions

7.2 Closed-Loop Transfer Functions

Q7.2.1 What are the poles of the closed-loop transfer function.

Q7.2.2 What is positive feedback and what is negative feedback and what effect does each have on control performance?

Q7.2.3 What is the effect of feedback control of an open-loop unstable processes?

7.3 Position Forms of the PID Algorithm

Q7.3.1 What is $\bar{c}$ in the position form of the PID algorithm?

Q7.3.2 How can derivative kick be eliminated?

Q7.3.3 How does the control interval affect the digital version of the position form of the PID algorithm?

Q7.3.4 How can you determine whether a process gain is positive or negative?

Q7.3.5 How are the units of the controller gain related to the units of the process gain?

7.4 Velocity Forms of the PID Algorithm

Q7.4.1 How can proportional kick be eliminated?

Q7.4.2 The term "$e(t)$" in the velocity form of the PID algorithm represents what type of control action?

7.6 Direct- and Reverse-Acting Controllers

Q7.6.1 For a reverse-acting controller, what type of controller output was used if the value of the CV increases?

Q7.6.2 If the sign of the loop gain is negative, what does this indicate?

7.7 Analysis of P, I and D Action

Q7.7.1 Compare and contrast the fundamental characteristics of proportional, integral and derivative action.

Q7.7.2 What effect does integral action have on the oscillatory nature of the closed-loop performance?

Q7.7.3 What effect does derivative action have on the oscillatory nature of the closed-loop performance?

7.8 Choosing the Proper Mode of a PID Controller

Q7.8.1 In general, what kind of processes use P-only controllers?

Q7.8.2 In general, what kind of processes use PI controllers?

Q7.8.3 In general, what kind of processes use PID controllers?

7.9 Commonly Encountered Control Loops

Q7.9.1 How can a control valve with a deadband of 10% be used to control the flow rate to a deadband of less than 0.5%?

Q7.9.2 What are the relative dynamics (i.e., fastest to slowest) of the actuator, process and sensor for a typical pressure control loop?

Q7.9.3 What are the relative dynamics (i.e., fastest to slowest) of the actuator, process and sensor for a typical dissolved oxygen control loop?

Q7.9.4 What are the relative dynamics (i.e., fastest to slowest) of the actuator, process and sensor for a typical composition control loop?

7.14 Analytical Questions and Exercises

7.2 Closed-Loop Transfer Functions

P7.2.1* Using the characteristic equation, determine the dynamic behavior of a PI controller with K_c equal to 0.1 and $_I$ equal to 10 applied to a first-order process in which the process gain is equal to 12 and the time constant is equal to 9. Assume that $G_s(s)$ and $G_a(s)$ are equal to unity.

P7.2.2* Using the characteristic equation, determine the dynamic behavior of a P-only controller with K_c equal to 3 applied to a second-order process (K_p=0.3, $_n$=5 and =2). Assume that $G_s(s)$ and $G_a(s)$ are equal to unity.

P7.2.3** Using the characteristic equation, determine the dynamic behavior of a PI controller with $_I$ equal to 4 applied to a second-order process (K_p=2, $_n$=5 and =1.5). Assume that $G_s(s)$ and $G_a(s)$ are equal to unity. Find the values of K_c that render this closed-loop process unstable.

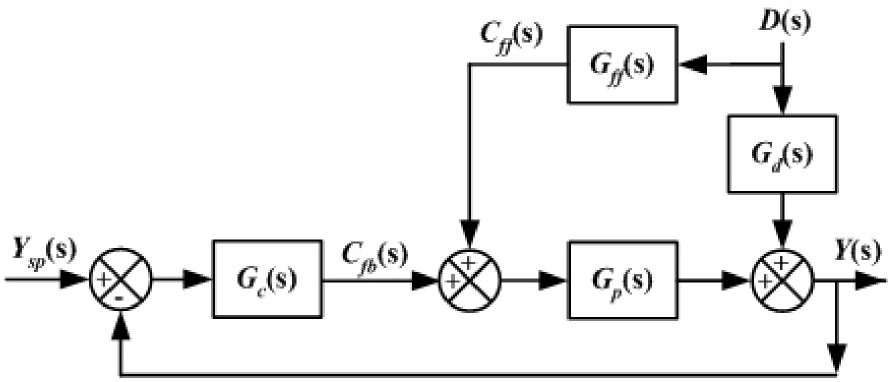


Figure P7.2.4 Block diagram of a combined feedforward and feedback controller.

P7.2.4** Figure P7.2.4 shows a block diagram for a combined feedforward and feedback controller. Develop the transfer functions for the effect of $D(s)$ and $Y_{sp}(s)$ on $Y(s)$.

P7.2.5** Consider the actuator, process and sensor models for the DO process (Figure P7.2.5) presented in Example 3.9:

Actuator:

$$F_{air} \quad (F_{air})_{spec}$$

Process:

$$k_L a \quad 0.25 \quad 0.001(F_{air} \quad 500)$$

$$\frac{dC_{O_2}}{dt} \quad k_L a \; C^*_{O_2} \quad C_{O_2} \quad K_{O_2} \quad {}_{\max} x$$

Sensor:

$$\frac{dC_{O_2,s}}{dt} \quad \frac{1}{\ _s} \quad C_{O_2} \quad C_{O_2,s}$$

- C_{O_2} - concentration of O_2 in the reaction broth (initially 1.1×10^{-4} g-moles/l)

- $C^*_{O_2}$ - saturated concentration of O_2 in the broth (2.20×10^{-4} g-moles/l)
- $C_{O_2,s}$ - the measurement of the O_2 concentration in the broth (initially 1.1×10^{-4} g-moles/l)
- F_{air} - the volumetric flow rate of air to the bio-reactor (500 cfm)
- K_{O_2} - cellular uptake of O_2 (1.98 g-moles O_2/g-cells)
- k_La - the overall liquid phase mass transfer coefficient for transport from the bubble surface to the bulk broth (initially 0.25 s^{-1})
- T - broth temperature (35ºC)
- t- time (s)
- V - the volume of broth in the bio-reactor (1000 l)
- x - constant cell concentration in the bio-reactor (0.25 g/l)
- $_{max}$ - maximum specific growth rate (5.56×10^{-5} s^{-1})
- $_s$ - the time constant of the DO sensor (30 s)

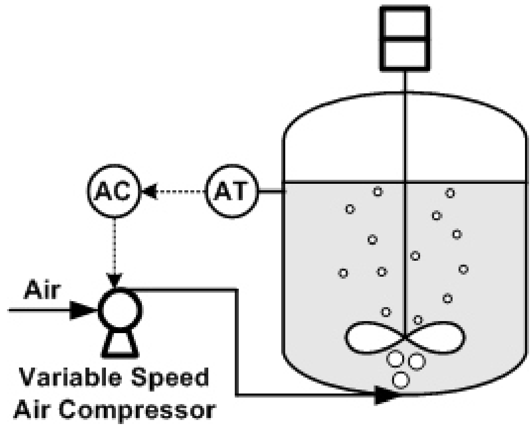


Figure P7.2.5 Schematic of a DO controller applied to a batch bio-reactor.

Using the characteristic equation and the Routh stability criterion, determine the stability limits for the controller gain for a PI controller with a reset time equal to 30 s.

P7.2.6S** Consider the actuator, process and sensor models for the level in a tank process (Figure P7.2.5) presented in Example 3.5:

Actuator
$$\frac{dF_{out}}{dt} \quad \frac{1}{_v} \quad F_{out,spec} \quad F_{out}$$

Process
$$A_c \frac{dL}{dt} \quad F_{in} \quad F_{out}$$

Sensor
$$L_s \quad L$$

- A_c- cross-sectional area of the tank (0.3 m^2)
- L - the level of liquid in the tank (initially 2 meters)
- F_{in} - the mass flow rate of liquid into the tank (1.0 kg/s)
- F_{out} - the mass flow rate of liquid leaving the tank (initially 1.0 kg/s)
- t - time (s)
- - the fluid density (1 kg/l)
- $_v$ - the time constant for the control valve on stream 1 (5 s).

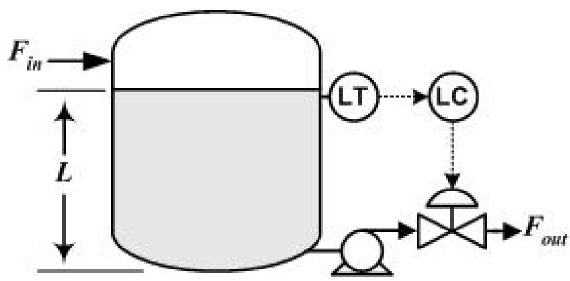


Figure P7.2.6 Control diagram of a level control process for a tank.

Using the characteristic equation and the Routh stability criterion, determine the stability limits for the controller gain for a PI controller with a reset time equal to 30 s. Apply these limits to the simulator that comes with this textbook and compare results.

P7.2.7[S]*** Consider the actuator, process and sensor models for the composition mixer process (Figure P7.2.7) presented in Example 3.4:

Actuator $$\frac{dF_1}{dt} \quad \frac{1}{_v}(F_{1,spec} \quad F_1)$$

Process $$V\frac{dC}{dt} \quad F_1 C_1 \quad F_2 C_2 \quad (F_1 \quad F_2)C$$

Sensor $$C_s(t) \quad C(t \quad _A)$$

- C - the concentration of the component in the mixed stream (g mol/l)
- C_1 - the concentration of the component in stream 1 (0.5 g mol/l)
- C_2 - the concentration of the component in stream 2 (1.0 g mol/l)
- F_1 - the mass flow rate of stream 1 (initially 500 kg/min)
- F_2 - the mass flow rate of stream 2 (500 kg/min)
- t - time (min)
- V - the volume of the mixer (1000 l)
- - the constant density of the feed and product streams (1 kg/l)
- $_v$ - the time constant for the flow controller on stream 1 (2 s).
- $_A$ - analyzer deadtime (5 min).

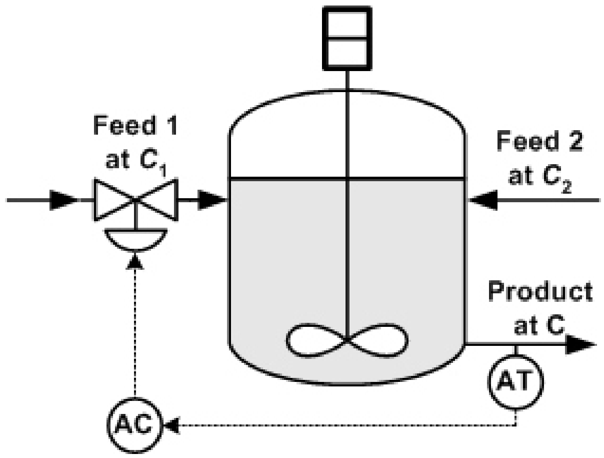


Figure P7.2.7 Control diagram for a controller applied to the composition mixing process.

Using the characteristic equation and the Routh stability criterion, determine the stability limits for the controller gain for a PI controller with a reset time equal to 3 min. Apply these limits to the simulator that comes with this textbook and compare results. [Hint: Use the first-order Padé approximation (see Example 8.2) to approximate the time delay in the model for the composition analyzer on the product stream.]

7.3 Position Forms of the PID Algorithm

P7.3.1** Write the digital version of the position form of the PID algorithm similar to Equation 7.3.4 for the process shown in Figure P7.3.1 and assume that the control valve on the steam line to the heat exchanger has an air-to-open actuator with a valve positioner. Use the form that is not susceptible to derivative kick.

P7.3.2** Write the position form of the PID algorithm similar to Equation 7.3.1 for Figure P7.3.2 and assume that the control valve on the steam line has an air-to-open actuator without a valve positioner. Use the form that is not susceptible to derivative kick and use proportional band instead of K_c.

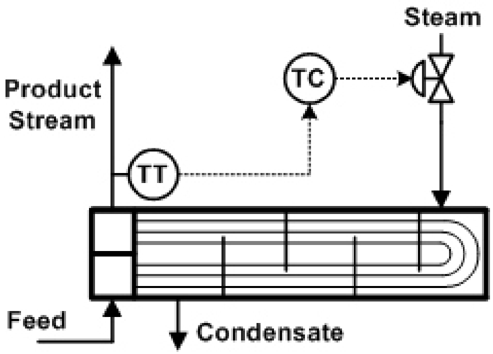


Figure P7.3.1 Control diagram for a temperature control loop applied to a steam-heated heat exchanger.

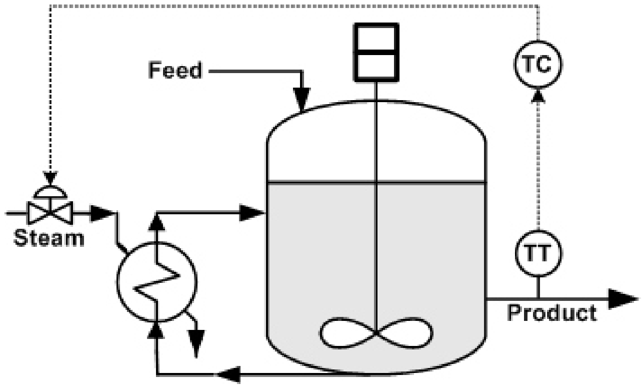


Figure P7.3.2 Control diagram for a temperature control loop applied to an endothermic CSTR

P7.3.3** Write the digital version of the position form of the PID algorithm similar to Equation 7.3.4 for Example 3.3 (see Figure P7.3.3) and assume that the control valve on the feed line to the mixer has an air-to-close actuator without a valve positioner. Use the form that is susceptible to derivative kick.

P7.3.4* Determine the dimensional version of the controller gain corresponding to a proportional band of 400% for a flow controller. The range of the error from setpoint is 1000 lb/h and the controller output is 0 to 100%.

P7.3.5* Determine the proportional band corresponding to a controller gain of 0.01 %-h/lb for a flow controller The range of the error from setpoint is 3000 lb/h and the controller output is 0 to 100%.

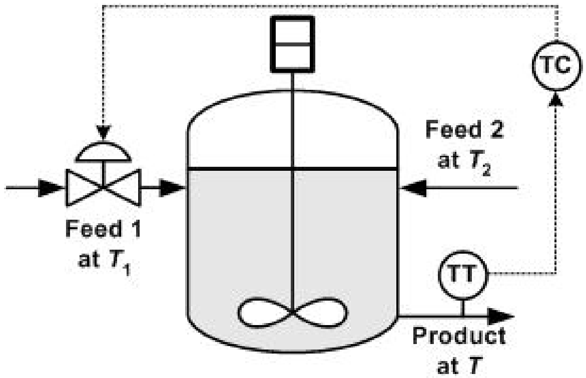


Figure P7.3.3 Control diagram for a temperature control loop applied to a thermal mixing process

P7.3.6** Write the position form of the PID algorithm similar to Equation 7.3.1 for the process shown in Figure P7.3.2) and assume that the control valve on the steam line has an air-to-close actuator with a valve positioner. Use the form that is susceptible to derivative kick.

7.4 Velocity Forms of the PID Algorithm

P7.4.1** Write the velocity form of the PID algorithm similar to Equation 7.4.4 for Figure P7.3.1and assume that the control valve on the steam line to the heat exchanger has an air-to-close actuator. Use the form that is susceptible to derivative kick and proportional kick.

P7.4.2** Write the velocity form of the PID algorithm similar to Equation 7.4.4 for Example 3.5 (see Figure P7.3.2) and assume that the control valve on the feed line to the mixer has an air-to-close actuator without a valve positioner. Use the form that is not susceptible to derivative kick or proportional kick.

P7.4.3** Write the velocity form of the PID algorithm similar to Equation 7.4.4 for Example 3.2 (see Figure P7.3.3) and assume that the control valve on the feed line to the mixer has an air-to-open actuator with a valve positioner. Use the form that is not susceptible to derivative kick or proportional kick. Use proportional band instead of K_c.

7.6 Direct- and Reverse-Acting Controllers

P7.6.1 Consider a process with a positive process gain that uses an air-to-open control valve. Should you use a direct-acting or a reverse-acting controller in this case?

P7.6.2 Consider a process with a negative process gain that uses an air-to-open control valve. Should you use a direct-acting or a reverse-acting controller in this case?

P7.6.3 Consider a process with a positive process gain that uses an air-to-close control valve. Should you use a direct-acting or a reverse-acting controller in this case?

P7.6.4 Consider a process with a negative process gain that uses an air-to-close control valve. Should you use a direct-acting or a reverse-acting controller in this case?

7.7 Analysis of P, I and D Action

P7.7.1** Derive each of the characteristics of proportional, integral and derivative action similar to Section 7.7 except using a disturbance upset instead of a setpoint change. Assume that the effect of $D(s)$ on $Y(s)$ [i.e., $G_d(s)$] is equal to K_d.

P7.7.2*** Consider a temperature control loop with the following characteristics:

$$K_a = 2 \text{ gpm}/\% \qquad K_p = \frac{0.5}{10s+1}\ {}^{\circ}\text{C}/\text{gpm} \qquad K_s = \frac{1}{0.5s+1}\ {}^{\circ}\text{C}/{}^{\circ}\text{C}$$

a. What are the units for gain of a controller for this process?

b. Develop the closed-loop transfer function for a P-only controller applied to this process.

c. What the range of stability for a P-only controller applied to this process.

d. For K_c=10 and a setpoint change of 5°C with a P-only controller, what would be the offset from setpoint?

7.8 Choosing the Proper Mode of a PID Controller

7.8.1** When are the dynamics of the actuator system important to the overall dynamic behavior of the feedback loop? Give an example.

7.8.2** When are the dynamics of the sensor system important to the overall dynamic behavior of the feedback loop? Give an example.

7.8.3** Consider the control of a DC motor, which is commonly encountered in the bio-tech industries (e.g., control of the speed of a mixer on a bio-reactor). The following transfer function represents the performance of a DC motor

$$\frac{\theta_m(s)}{V_A(s)} = \frac{K_p}{s(\tau_p s+1)}$$

where $_m$ is the rotational velocity of the shaft of the motor and V_A is the applied voltage. It is desired to design a control system for a DC motor that provides offset-free operation.

1. What type of controller (P, PI, PID or PD) would you select for this application and why?

2. Determine the closed-loop transfer function for setpoint changes and determine the conditions under which this system may become unstable if appropriate.

Chapter 8

Behavior and Tuning of PID Controllers

Chapter Objectives

- Demonstrate that as the controller aggressiveness is increased most closed-loop systems go from an overdamped response to critically damped to oscillatory to ringing to unstable behavior.
- Examine the effect of PID tuning parameters on closed-loop performance.
- Demonstrate how to determine if a control loop has excessive oscillations from too much proportional or too much integral action.
- Present classical and advanced PID tuning methods

8.1 Introduction

P-only, PI and PID controllers are the most commonly used modes of a PID-type controller. Of these three modes, PI is by far the most commonly used in industry. In Chapter 7, the fundamental characteristics of proportional, integral and derivative action were identified. In this chapter, we will consider the effect of the amount of proportional, integral and derivative action on the closed-loop performance of P-only, PI and PID controllers, which should allow you to better understand the effect of controller tuning (i.e., selecting the values of K_c, $_I$ and $_D$). The latter portion of the chapter presents classical and advanced tuning methods. Recommended tuning procedures are presented in Chapter 9.

8.2 Effect of Tuning Parameters for P-only Control

P-only controllers are used industrially for some pressure and level control loops. It was shown in Section 7.7 that a P-only controller increases the speed of the dynamic response of a process but does not change the order of the process. Because P-only control uses only one tuning parameter, K_c, it is the simplest form of PID control.

Example 8.1 Dynamics of a P-only Controller applied to a First-Order Process

Problem Statement. Analyze the dynamic behavior of a first-order process with a P-only controller.

Solution. Applying the characteristic equation (Equation 7.2.9) using the transfer function for a first-order process and $G_c(s)=K_c$ results in the following equation

$$G_c(s)G_p(s) + 1 = K_c \frac{K_p}{\tau_p s + 1} + 1 = 0$$

Rearranging,
$$K_c K_p + \tau_p s + 1 = \frac{\tau_p}{K_p K_c + 1} s + 1 = 0$$

This indicates that the closed-loop response remains first-order regardless of the value of K_c, and becomes faster (i.e., the closed-loop time constant becomes smaller) as K_c is increased. This result indicates that the best approach is to operate at very large controller gains because the response becomes virtually instantaneous and the offset becomes negligibly small (see Figure 7.7.3). If this situation existed for real processes, process control would be extremely simple. Unfortunately, this analysis does not apply to real processes because it is based on the assumption that the process is described as a first-order system without deadtime. Even if a process, by itself, behaves as a first-order process, as K_c is increased and the response of the closed-loop system becomes faster, the dynamics of the actuator and sensor will at some point become significant, and the actuator/process/sensor system becomes at least third-order. Example 7.4 showed that an open-loop overdamped third-order system can become unstable at large values of the controller gain. In addition, the first-order models for the actuator and sensors neglect deadtime because it is usually small compared to the time constant of the response for these systems. As K_c is increased and the process response becomes faster, even a small amount of deadtime can also affect the stability of the system. **The analysis of a P-only controller applied to a first-order process is physically unrealistic in the extreme because no real process behaves as a first-order system without deadtime at large controller gains.** The analysis of a FOPDT model (Example 8.3) is much more realistic than even a high-order linear model without deadtime because of the sensitivity of feedback control to deadtime.

Example 8.2 Evaluation of the Effect of K_c on the Closed-Loop Dynamics of a P-Only Controller Applied to a Second-Order Process

Problem Statement. Determine the effect of controller gain, K_c, for a P-only controller applied to a second-order process, assuming that the second-order process represents the combined effect of the actuator, process, and sensor.

Solution. Applying the characteristic equation (Equation 7.2.9) using a transfer function for a second-order process (Equation 6.4.2) and a P-only controller [$G_c(s)=K_c$] results in the following equation.

$$G_c(s)G_p(s) + 1 = \frac{K_c K_p}{\tau_n^2 s^2 + 2\zeta\tau_n s + 1} + 1 = 0$$

Rearranging and transforming this equation into the standard form for a second-order process yields

$$\frac{\tau_n^2}{K_cK_p+1}s^2+\frac{2\zeta\tau_n}{K_cK_p+1}s+1=0$$

Then, the closed-loop natural period (τ_n) and the closed-loop damping factor (ζ) are given as

$$\tau_n = \tau_n\sqrt{\frac{1}{K_cK_p+1}}$$

$$\zeta = \zeta\sqrt{\frac{1}{K_cK_p+1}}$$

As K_c is increased, both the closed-loop natural period and the closed-loop damping factor decrease. As K_c is increased, the response of the closed-loop process becomes faster and, eventually, more oscillatory. But no matter how large the controller gain becomes, the response of the system remains stable (i.e., $\zeta > 0$). Real processes become unstable if the controller gain is increased sufficiently due to nonlinearity and deadtime. Therefore, a second-order model without deadtime does not represent a real process for large values of the controller gain.

Example 8.3 Effect of K_c on Closed-Loop Dynamics for a FOPDT Process

Problem Statement. Determine the effect of controller gain, K_c, for a P-only controller applied to a FOPDT process. The transfer function for a FOPDT process is

$$G_p(s)=\frac{K_p e^{-\theta_p s}}{\tau_p s+1} \qquad \textbf{8.2.1}$$

Solution. Using the first-order Padé approximation[1] for the deadtime term, $e^{-\theta_p s}$,

$$e^{-\theta_p s}\approx\frac{1-½\,\theta_p s}{1+½\,\theta_p s}$$

[The Padé approximation is used here to approximate the effect of deadtime. Using frequency response analysis (Chapter 11), the effect of deadtime can be evaluated without approximation.] Substituting the Padé approximation into Equation 8.2.1 results in

$$G_p(s)=\frac{K_p(1-½\,\theta_p s)}{(\tau_p s+1)(1+½\,\theta_p s)}$$

Then, using the closed-loop transfer function for a setpoint change (Equation 7.2.7) for P-only control [$G_c(s)=K_c$] and assuming $G_a(s) = G_s(s) = 1$ yields

$$\frac{Y(s)}{Y_{sp}(s)} \quad \frac{G_c(s)G_p(s)}{G_c(s)G_p(s) \quad 1} \quad \frac{\dfrac{K_cK_p(1 \quad ½ \quad {}_p s)}{{}_p s \quad 1)(1 \quad ½ \quad {}_p s)}}{\dfrac{K_cK_p(1 \quad ½ \quad {}_p s)}{(\quad {}_p s \quad 1)(1 \quad ½ \quad {}_p s)} \quad 1}$$

Rearranging results in

$$\frac{Y(s)}{Y_{sp}(s)} \quad \frac{K_c K_p(1 \quad ½ \quad {}_p s)}{½ \quad {}_p \quad {}_p \, s^2 \quad [\, {}_p \quad ½ \quad {}_p(1 \quad K_c K_p)]s \quad 1 \quad K_c K_p}$$

The natural period for the closed-loop system is given by

$${}_n \quad \sqrt{\frac{½ \quad {}_p \quad {}_p}{1 \quad K_c K_p}}$$

The coefficient of the linear term in s is equal to $2 \quad {}_n$; therefore,

$$\frac{{}_p \quad ½ \quad {}_p(1 \quad K_c K_p)}{2 \quad {}_n(1 \quad K_c K_p)}$$

For stability, ; therefore, $\quad {}_p \quad ½ \quad {}_p(1 \quad K_c K_p) \quad 0$

which reduces to $\quad K_c \quad \dfrac{2 \quad {}_p \quad {}_p}{K_p \quad {}_p}$

Note that this system has a right-half plane zero at $2/ \quad {}_p$. The poles of the transfer function determine the majority of the dynamic behavior of the closed-loop system. The poles of the previous closed-loop transfer function are the roots of the denominator and can be evaluated analytically using the quadratic formula, i.e.,

$$p_1 \quad \frac{(\quad {}_p \quad ½ \quad {}_p[1 \quad K_c K_p]) \quad \sqrt{(\quad {}_p \quad ½ \quad {}_p[1 \quad K_c K_p])^2 \quad 2 \quad {}_p \quad {}_p(1 \quad K_c K_p)}}{{}_p \quad {}_p} \qquad \textbf{8.2.2}$$

$$p_2 \quad \frac{(\quad {}_p \quad ½ \quad {}_p[1 \quad K_c K_p]) \quad \sqrt{(\quad {}_p \quad ½ \quad {}_p[1 \quad K_c K_p])^2 \quad 2 \quad {}_p \quad {}_p(1 \quad K_c K_p)}}{{}_p \quad {}_p} \qquad \textbf{8.2.3}$$

Figure 8.2.1 shows plots of the two poles, p_1 and p_2, on a complex plane (Section 5.3) as a function of K_c assuming that K_p = 1, $\quad {}_p$ = 1, and $\quad {}_p$ = 0.5, where the real portion of the pole is plotted on the x-axis and the imaginary portion on the y-axis. This plot is called a **root locus diagram** and graphically shows the effect of K_c on the dynamic characteristics of the feedback system. The arrows on the lines in Figure 8.2.1 indicate the direction of increasing values of K_c.

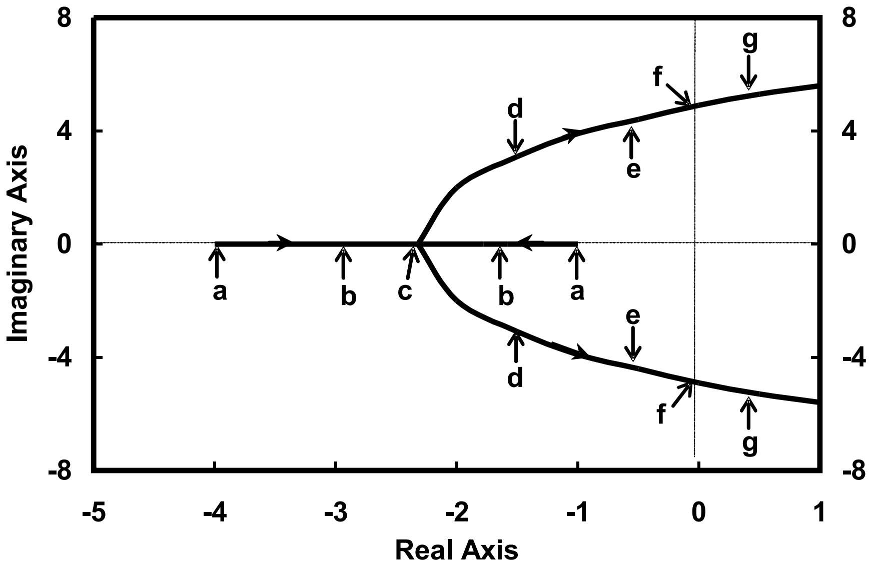


Figure 8.2.1 Root locus diagram for the results of Example 8.3. A P-only controller applied to a FOPDT process (K_p=1; $_p$=1; $_p$=0.5) with a controller gain, K_c, that increases from points a to points g: (a) K_c=0, (b) K_c=0.30, (c) K_c=0.35, (d) K_c=1.0, (e) K_c =3.0, (f) K_c =5.0, (g) K_c=6.0. The zero for this system is at +4.

A root locus diagram is an analysis tool based on a specific theoretical framework with significant analytical capabilities. The theoretical development of the root locus diagram is beyond the scope of this text. On the other hand, root locus diagrams are used here as a convenient visual representation of the range of dynamic behavior resulting from feedback control. The root locus diagram is a plot in the complex plane of the poles of the characteristic equation of the closed-loop transfer function in which a parameter, usually a controller tuning parameter (e.g., K_c or $_I$), is varied over a range. By knowing the dynamic character indicated by the poles of a transfer function (Section 5.3), you can identify the transition from one type of dynamic behavior to another as the parameter of interest is varied. For example, point c in Figure 8.2.1 (K_c=0.35) indicates the boundary between overdamped and underdamped behavior, and point f (K_c=5.0) represents the boundary between stable and unstable behavior. The root locus diagram presented here is qualitatively representative of a major portion of industrial process. While root locus diagrams are a convenient means to represent the dynamics of feedback control and assist in the fundamental understanding of feedback dynamics, they are not generally used industrially because their use requires process models, which are not usually available.

Between points a and c in Figure 8.2.1 ($0 < K_c < 0.35$), the poles are real and negative, indicating overdamped dynamic behavior. Remember from Section 5.3 that real negative poles yield time-domain solutions that involve exponential decay with time (e^{-at}). At point c (K_c = 0.35), the closed-loop system is critically damped because both poles are -2.32 and any increase in K_c results in oscillatory behavior. For the poles between points c and f ($0.35 < K_c < 5.0$), the system is underdamped. The poles in this region are complex conjugate pairs ($p_1 = a + i$ and $p_2 = a - i$), which in the time-domain behavior result in terms of the form e^{-at} sin t. Because the real part of these complex conjugate poles is negative, damped oscillatory behavior is indicated. Moreover, the magnitude of the real portion of the complex conjugate (a) decreases as K_c is increased from 0.35 to 5, which indicates that the rate of damping of the oscillations is also decreasing. For the poles between points c and f, the magnitude of the

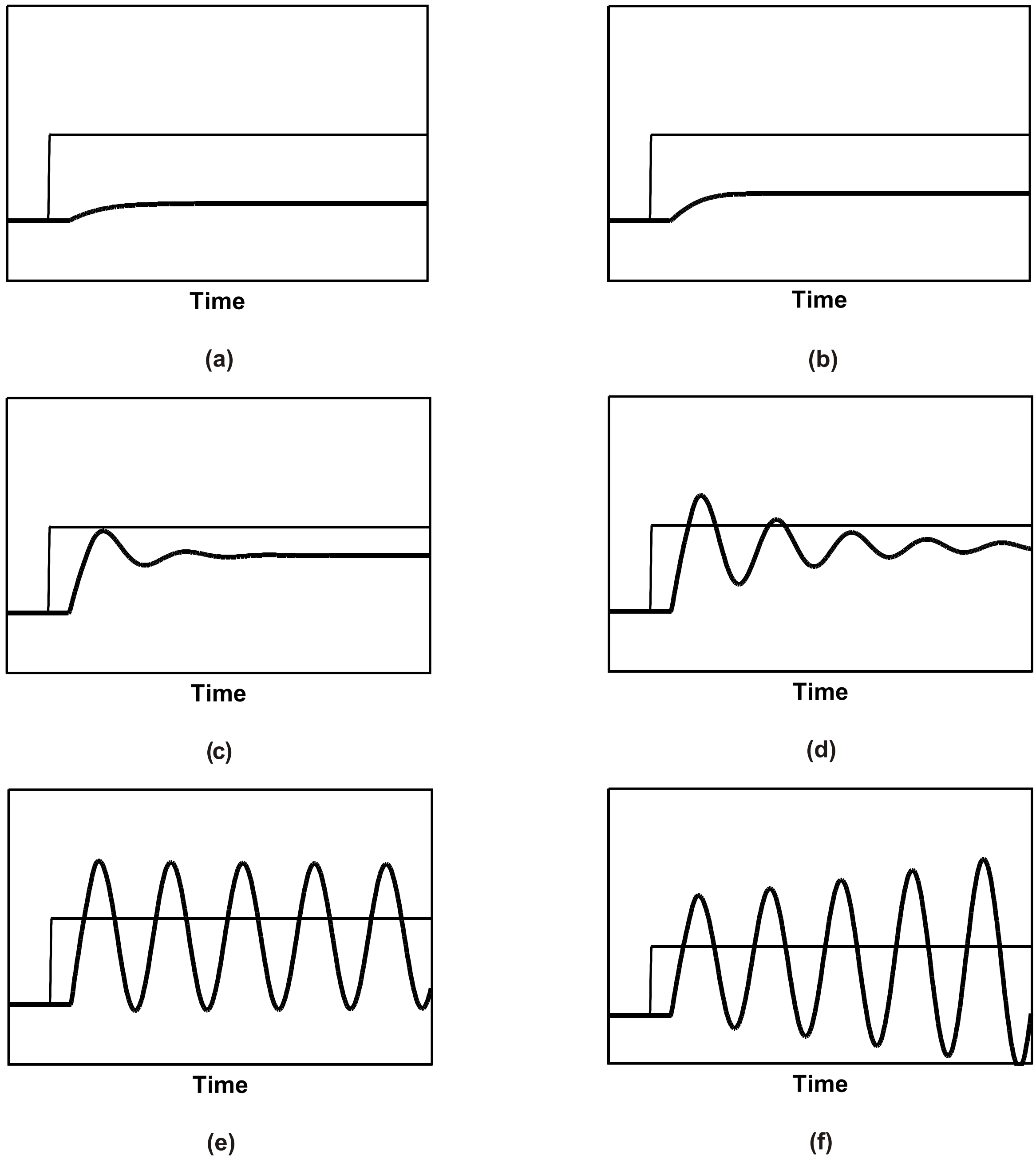


Figure 8.2.2 The response of a FOPDT process with a P-only controller to a setpoint change. (a) K_c=0.25; (b) K_c=0.35; (c) K_c=1.0; (d) K_c=3.0; (e) K_c=5.0; (f) K_c=5.2.

imaginary part of the pole increases, indicating an increase in the oscillatory nature of the response. At point f (K_c = 5), sustained oscillations result. This marks the boundary between stable operation (K_c < 5) and unstable operation (K_c > 5) and is indicated as a vertical line in Figure 8.2.1. Poles to the right of this vertical line are said to lie in the right-half plane and represent unstable behavior. As K_c is increased above a value of 5, the rate of exponential growth increases because the magnitude of the real part of the pole increases.

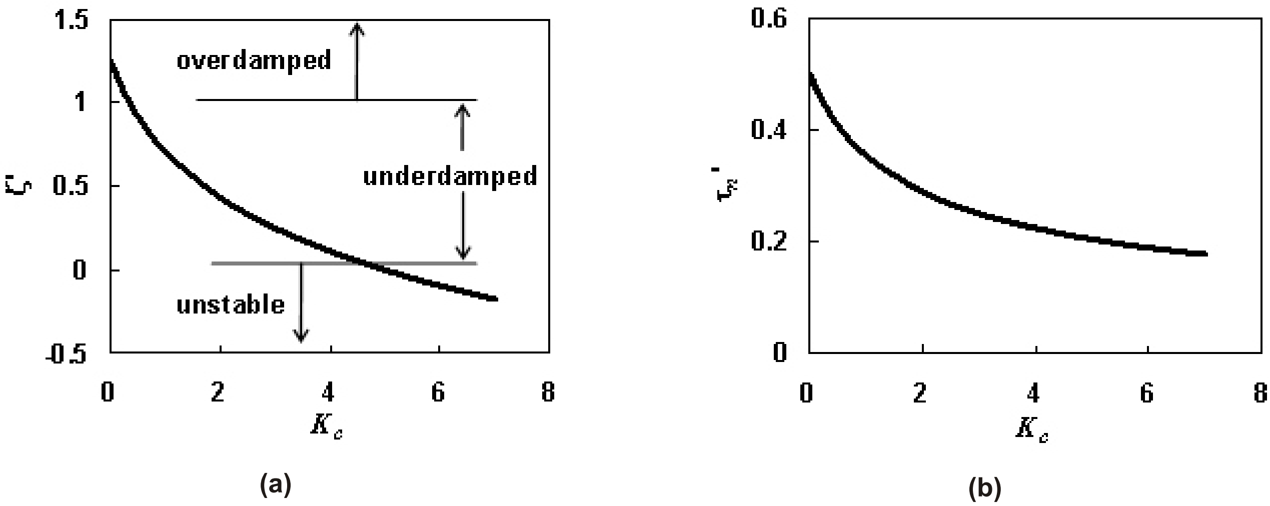


Figure 8.2.3 The effect of the controller gain on the dynamic behavior of a FOPDT process (K_p=1; $_p$=1; $_p$=0.5). (a) shows how the closed-loop damping factor changes with K_c and (b) shows how the closed-loop natural period varies with K_c.

Figure 8.2.2 shows the time-domain response of the FOPDT model (K_p=1; $_p$=1; $_p$=0.5) for several values of K_c (0.25, 0.35, 1.0, 3.0, 5.0, 5.2) for a setpoint change in y. The dynamic response of the closed-loop system corresponds to (a) sluggish behavior, (b) critically damped, (c) oscillatory performance, (d) ringing, (e) sustained oscillations, and (f) unstable oscillations. Note that as K_c increases, the system response becomes faster. Although this is a simple case, industrial control loops show the same general behavior; that is, **as the controller gain is increased, an open-loop overdamped process moves from overdamped behavior to critically damped to underdamped to ringing to sustained oscillations to unstable oscillations.** Therefore, for a P-only controller, determining whether a controller is tuned too aggressively (i.e., K_c is too large) or too sluggishly (i.e., K_c is too small) is relatively simple by comparing the dynamic response of the process to the sequence of dynamic behavior shown in Figure 8.2.2. If the dynamic response of the process is similar to Figure 8.2.2a, the value of K_c used in the P-only controller is too small. On the other hand, if the dynamic response is similar to one of the responses shown in Figures 8.2.2d-f, the value of K_c is too large and should be reduced.

Because the application of a P-only controller to the FOPDT model considered here results in a second-order closed-loop response, the closed-loop damping factor and natural period can be used to characterize the dynamic behavior of this system. Figure 8.2.3 shows the closed-loop damping factor and natural period for this system as functions of the controller gain, K_c. Note that the same modes of dynamic behavior that were observed in Figures 8.2.2a-f are also shown in Figure 8.2.3a. In Figure 8.2.3b, the closed-loop second-order natural period is shown to decrease monotonically as the controller gain is increased.

Self-Assessment Questions

Q8.2.1 Define a root locus diagram and explain what it shows.

Q8.2.2 When a P-only controller is applied to a FOPDT process, list the different types of dynamic behavior that result as the controller gain is increased from zero.

Q8.2.3 Consider a closed-loop response with sustained oscillations. Where is it located on a root locus diagram?

Q8.2.4 For a plot of the closed-loop damping factor versus the controller gain, what point corresponds to critically damped behavior?

Self-Assessment Answers

Q8.2.1 A root locus plot is a plot of the real and imaginary components of the poles of a closed-loop transfer function plotted as a function of controller gain. As a result, you can observe the range of dynamic behavior as the controller gain is changed. For example, when the poles all lie on the negative real axis, this indicates overdamped behavior and when the poles separate from the real axis, this indicates the onset of underdamped behavior and when one of the poles crosses the imaginary axis and has a positive real component, this indicates the onset of instability.

Q8.2.2 As the controller gain is increased, the closed-loop response goes from overdamped to critically damped to oscillatory to ringing to sustained oscillations to unstable oscillations.

Q8.2.3 Sustained oscillations correspond to a complex conjugate pair with positive and negative imaginary components without a real component. Therefore, sustained oscillations are represented by a complex conjugate pair located on the imaginary axis of the complex plane.

Q8.2.4 A critically damped response corresponds to a closed-loop damping factor of 1.0.

8.3 Effect of Tuning Parameters for PI Control

PI controllers are the most commonly used form of PID controllers, accounting for over 90% of industrial PID applications. Tuning a PI controller involves setting the controller gain, K_c, and the reset time, τ_I. Proportional action increases the speed of the closed-loop response, and integral action ensures offset-free operation. As a result, combining proportional and integral action achieves the advantages of both approaches, but tuning a PI controller is more complicated than a P-only controller because two tuning parameters must be specified. To better understand the effect of the PI tuning parameters, the following examples are presented to illustrate the individual effects of K_c and τ_I on a PI controller applied to a process represented by a FOPDT model.

Example 8.4 Evaluation of the Effect of K_c on the Closed-Loop Dynamics for a PI Controller applied to a FOPDT Model of a Process

Problem Statement. Analyze the effect of the controller gain on the dynamic behavior of a PI controller applied to a FOPDT model of a process. Assume that the reset time ($\tau_I = 1$) and the FOPDT parameters ($K_p = 1$; $\tau_p = 1$; $\theta_p = 0.5$) are constant. Assume that the FOPDT model represents the combined effect of the actuator, process and sensor.

Solution. Substituting the transfer function of a general FOPDT model as the process and the transfer function of the controller,

$$G_c(s) = K_c\left(1 + \frac{1}{\tau_I s}\right)$$

into the characteristic equation for this system assuming $G_a(s) = G_s(s) = 1$ results in

$$G_c(s)G_p(s) + 1 = \frac{K_c\left(1 + \frac{1}{\tau_I s}\right)K_p(1 - \tfrac{1}{2}\theta_p s)}{(\tau_p s + 1)(1 + \tfrac{1}{2}\theta_p s)} + 1 = 0 \qquad \textbf{8.3.1}$$

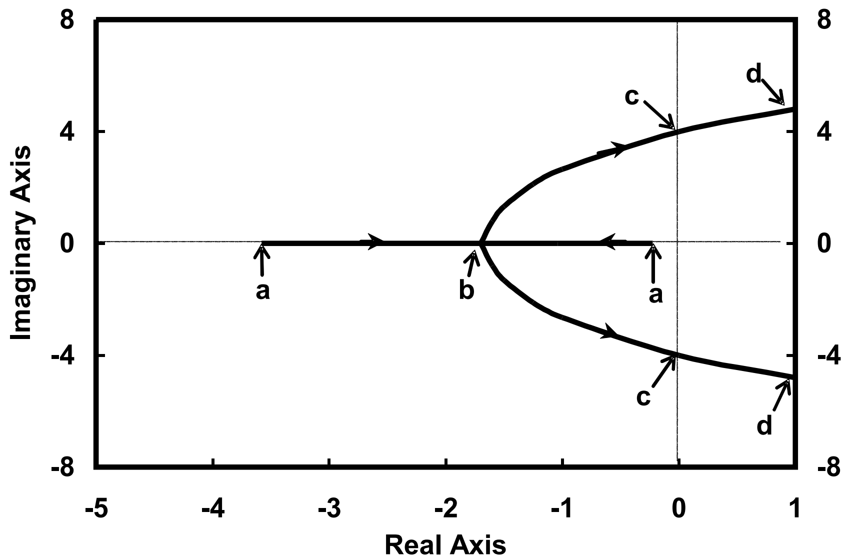


Figure 8.3.1 Root locus diagram for the results of Example 8.4. A PI controller applied to a FOPDT process (K_p=1; $_p$=1; $_p$=0.5; $_I$=1) with a controller gain, K_c, that increases from points a to points d: (a) K_c=0.2, (b) K_c=0.75, (c) K_c=4.0, (d) K_c=6.0. The zero for this system is at +4.

using the Padé approximation for deadtime.

Figure 8.3.1 shows the root locus diagram of the characteristic equation for a range of K_cs for $_I$ 1.0, K_p 1.0, $_p$ 1.0, and $_p$ 0.5. Because the rearrangement of Equation 8.3.1 results in a cubic equation in s, there are three poles for this closed-loop system. Using the parameter values, the characteristic equation for this system reduces to

$$\tfrac{1}{4}s^3 \quad (1.25 \quad \tfrac{1}{4}K_c)s^2 \quad (1 \quad \tfrac{3}{4}K_c)s \quad K_c \quad 0 \qquad \textbf{8.3.2}$$

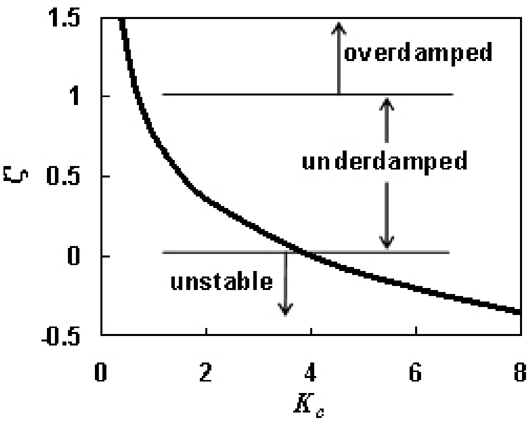


Figure 8.3.2 The effect of the controller gain on the damping factor for a PI controller applied to a FOPDT process (K_p=1; $_p$=1; $_p$=0.5, $_I$=1).

As K_c is increased, the dynamic behavior goes from overdamped to critically damped to underdamped to sustained oscillations to unstable behavior, which is consistent with the results obtained in Example 8.3. One pole (p_3 = -1.0) remains invariant for the full range of controller gains and is denoted in Figure 8.3.1 by a diamond symbol. Figure 8.3.2 plots the damping factor of the complex conjugate pair for this system as a function of K_c, which shows the transition from overdamped to unstable. The complex conjugate determines the oscillatory component of the dynamic response. By comparing Figures 8.3.2 and 8.2.3a, you can see that integral action causes the system to become unstable at a lower value of K_c than a P-only controller. The addition of integral action provides offset-free operation but produces a more sensitive

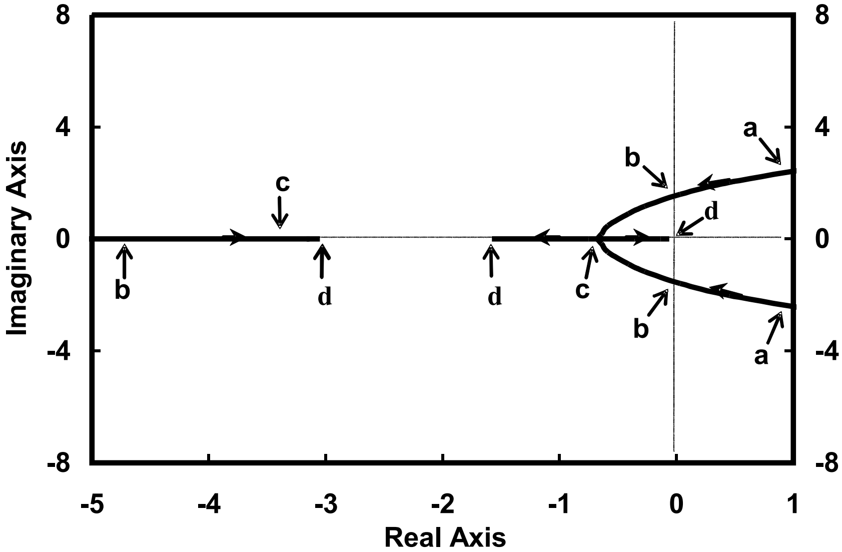


Figure 8.3.3 Root locus diagram for the results of Example 8.5. A PI controller applied to a FOPDT process (K_p=1; $_p$=1; $_p$=0.5, K_c=0.3) with a controller reset time, $_I$, that increases from points a to points d: (a) $_I$ =0.04, (b) $_I$= 0.19, (c) $_I$= 0.9, (d) $_I$ =4.0. The zero for this system is at +4.

closed-loop system. The speed of the closed-loop dynamics for this system increases with K_c in a similar fashion to the results shown in Figure 8.2.3b for this case.

Example 8.5 Evaluation of the Effect of $_I$ on the Closed-Loop Dynamics for a PI Controller applied to a FOPDT Model of a Process

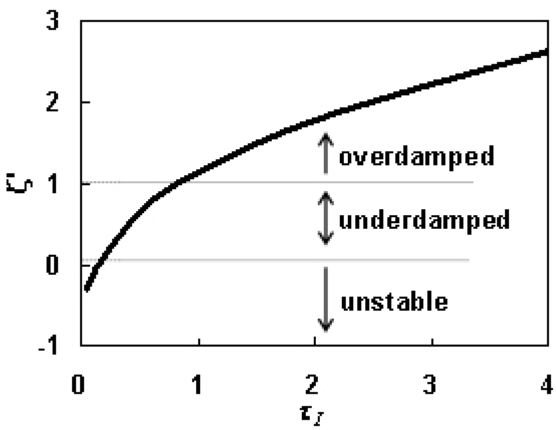


Figure 8.3.4 The effect of the controller reset time on the damping factor for a PI controller applied to a FOPDT process (K_p=1; $_p$=1; $_p$=0.5, K_c=0.3).

Problem Statement. Analyze the effect of the reset time on the dynamic behavior of a PI controller applied to a FOPDT model of a process. Assume that the controller gain (K_c = 0.3) and the FOPDT parameters (K_p = 1; $_p$ = 1; $_p$ = 0.5) are constant, and the combined effect of the actuator, process, and sensor is represented by a FOPDT model.

Solution. The closed-loop transfer function developed in Example 8.4 can be applied to this case. As in the previous example, solving for the roots of the characteristic equation requires the solution of a cubic equation. Figure 8.3.3 shows the poles for the closed-loop transfer function for a range of values of reset times from 0.04 to 4.0. In this case, as $_I$ is increased, the dynamic behavior goes from unstable oscillations to sustained oscillations to critically

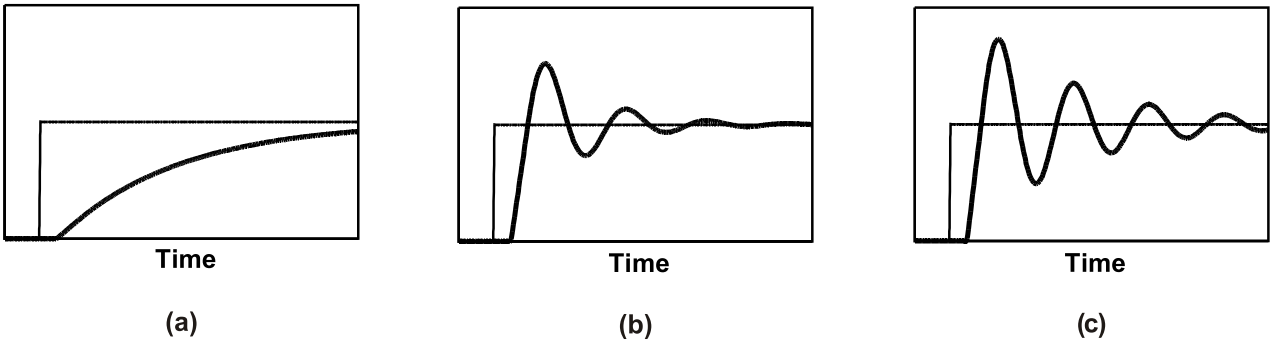


Figure 8.3.5 PI controller response for a FOPDT process with varying amounts of proportional action. (a) K_c is too low. (b) well-tuned controller. (c) K_c is too large.

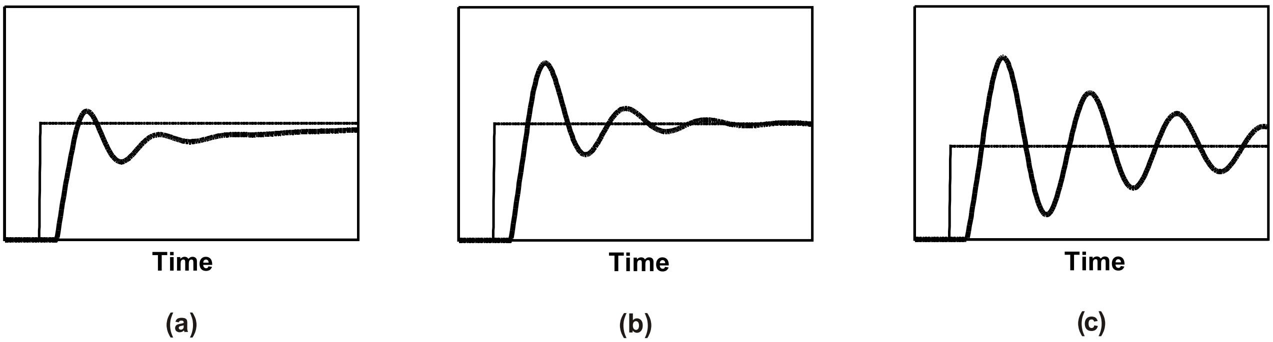


Figure 8.3.6 PI controller response for a FOPDT process with varying levels of integral action. (a) $_I$ is too large. (b) well-tuned controller. (c) $_I$ is too small.

damped to overdamped. As the amount of integral action is increased (i.e., as $_I$ is decreased), the dynamic behavior goes through the same sequence of phases as an increase in proportional action produced. Figure 8.3.4 shows the effect of $_I$ on the damping factor of the complex conjugate pair for this system.

Figure 8.3.5 shows the dynamic behavior of the same FOPDT process considered in Examples 8.4 and 8.5 for a PI controller with different amounts of proportional action. Figure 8.3.5b shows the results for a PI controller for which both K_c and $_I$ were adjusted to provide a well-tuned response. In addition, this tuning was modified by increasing K_c (Figure 8.3.5c) while keeping $_I$ constant and by decreasing K_c (Figure 8.3.5a) while keeping $_I$ constant. The increase in K_c results in ringing while the decrease in K_c results in sluggish behavior. Note that too much or too little proportional action results in longer settling times than the well-tuned controller.

Figure 8.3.6 shows similar results for the effect of variations in $_I$. Figure 8.3.6b shows the results for a well-tuned controller and is the same result as shown in Figure 8.3.5b. A decrease in $_I$ from these settings results in ringing (Figure 8.3.6c) and an increase results in a slow removal of offset (Figure 8.3.6a). By comparing Figures 8.3.5a and 8.3.6a, it can be seen that when K_c is too low, long response times and sluggish behavior result, and when integral action is too low (i.e., $_I$ is too large), offset elimination is slow. Therefore, determining that a PI controller has too little proportional or too little integral action is relatively straightforward. On the other hand, ringing from too much proportional action (Figure 8.3.5c) and ringing from too much integral action (Figure 8.3.6c) are quite similar; therefore, when controller ringing results, it is difficult to tell whether it is caused by excessive proportional action, excessive integral action, or both.

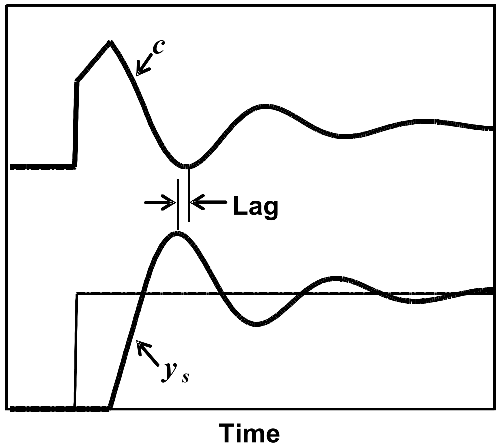


Figure 8.3.7 The lag between the controller output and the CV for a well-tuned PI controller. c is the controller output and y_s is the CV.

Figure 8.3.7 shows the controller output and CV for a well-tuned controller. Note that the controller output "lags" behind the CV for this case. The lag between the CV and the controller output is discussed in more detail in Chapter 11. Figure 8.3.8 shows the same system with K_c increased by 50% (i.e., the ringing is caused by too much proportional action). Note that the lag between the CV and the controller output is significantly reduced. Excessive controller gain reduces the lag between the CV and the controller output. Figure 8.3.9 shows the case in which $_I$ is reduced by a factor of 2 compared with the settings for the well-tuned controller. In this case, the lag increases significantly. Excessive integral action results in an increase in the lag of the system. **Therefore, the lag between the controller output and the CV can be used to determine if a controlled process is ringing from too much proportional action or too much integral action.** If a controller is ringing and there is a small amount of lag between the controller output and the CV, the ringing is caused by an excess amount of proportional action and can be corrected by reducing the controller gain. On the other hand, if a controller is ringing and there is an appreciable amount of lag between the controller output and the CV, the ringing is caused by an excessive amount of integral action and can be corrected by increasing the reset time. In the case in which there is an excessive amount of proportional and integral action, there should be some degree of lag between the controller output and the CV, but when the integral is reduced, the response usually still exhibits some degree of ringing.

With proportional-only control action, the maximum c (controller output) occurs at the maximum deviation from setpoint, which corresponds to zero lag (i.e., **in-phase**). For integral-only control action, the maximum c occurs when the error from setpoint changes sign, which corresponds to a large lag.

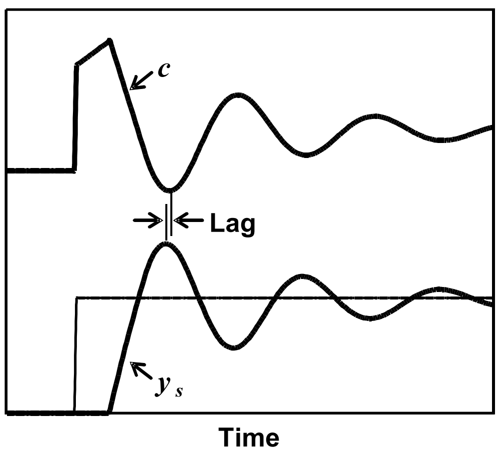


Figure 8.3.8 The lag between the controller output and the CV for a controller with too much proportional action. c is the controller output and y_s is the CV.

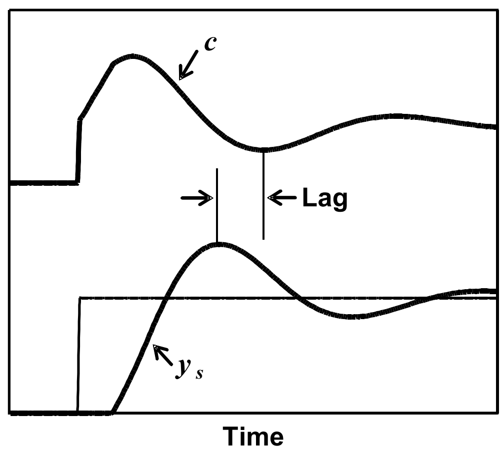


Figure 8.3.9 The lag between the controller output and the CV for a controller with too much integral action. c is the controller output and y_s is the CV.

Example 8.6 Stability Analysis of a PI Controller Applied to a FOPDT Process

Problem Statement. Using the Routh stability criterion, determine the minimum value of $_I$ for stability for a PI controller applied to a FOPDT process (Equation 8.3.1) for a specified controller gain (K_c = 0.3) and specified FOPDT parameters model (K_p = 1; $_p$ = 1; $_p$ = 0.5).

Solution. From Equation 8.3.1 after some rearranging, the characteristic equation for a PI controller applied to a FOPDT process using the first-order Padé approximation for the deadtime is given by

$$\tfrac{1}{2} \; _p \; _I \; _p s^3 \quad _I \; \tfrac{1}{2} \; _p(1 \quad K_c K_p) \quad _p \; s^2 \quad K_c K_p(\; _I \quad \tfrac{1}{2} \; _p) \quad _I \; s \quad K_c K_p \quad 0$$

Substituting the numerical values for the known parameters yields

$$\frac{_I}{4} s^3 \quad 1.175 s^2 \quad 1.3 \; _I \quad 0.25 \; s \quad 0.3 \quad 0$$

First, each coefficient of s in this polynomial must be positive for stability. Only the coefficient of s can be negative; therefore, for stability

$$_I \quad \frac{0.25}{1.3} \quad 0.192$$

Next, the Routh stability criterion is applied. The Routh array is given by

$0.25 \; _I$	$1.3 \; _I \quad 0.25$
1.175	0.3
$1.453 \; _I \quad 0.294$	0
0.3	0

From the Routh stability criterion, each of the elements in the first column must be positive. Only the third element is in question. Based on the third element in the first column, $_I$ >0.202 for stability. The latter restriction on the reset time is more restrictive; therefore, based on this analysis $_I$ >0.202 for stability. Note that this agrees with the reset time corresponding to sustained oscillations shown in Figures 8.3.3 and 8.3.4.

Self-Assessment Question

Q8.3.1 How are the dynamic responses of a PI and a P-only controller alike? How are they different?

Q8.3.2 Consider a PI controller. How can you tell if there is too little proportional action? How can you tell if there is too little integral action?

Q8.3.3 Consider a PI controller that results in ringing. How can you tell if there is too much proportional action or too much integral action?

Self-Assessment Answers

Q8.3.1 A PI and a P-only controller are alike in that as the controller gain is increased, the dynamics of the process goes from overdamped to critically damped to oscillatory to sustained oscillations to unstable. They are different in that the P-only controller exhibits offset while the PI does not. Also, with the addition of integral action, the closed-loop damping factor changes more sharply with gain for the PI controller.

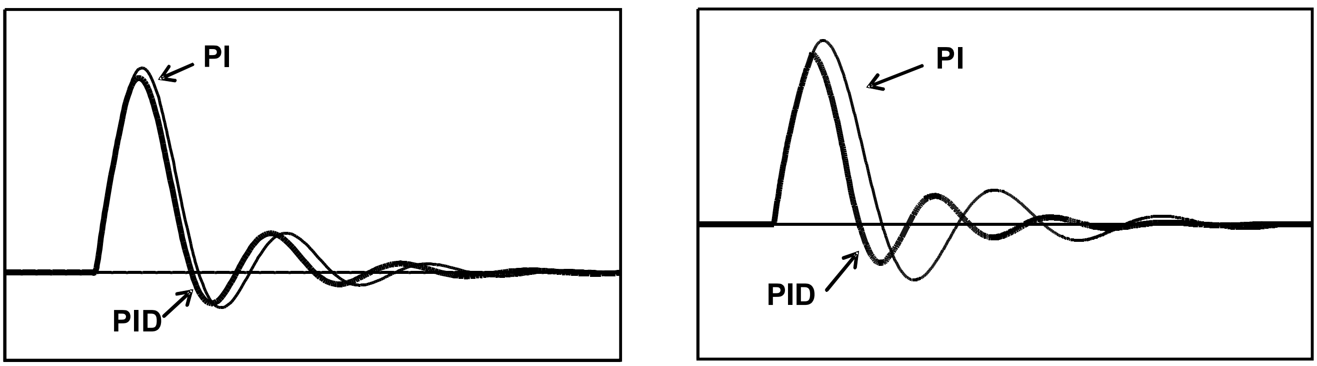
Q8.3.2 For a PI controller with too little proportional action, a sluggish response results (i.e., the proper decay ratio is not obtained and the response can be overdamped in the extreme), but offset is eventually eliminated (see Figure 8.3.5a). For a PI controller with too little integral action, offset removal is slow (see Figure 8.3.6a). For this latter case, if the level of proportional action is correct, the response will exhibit the correct underdamped behavior except that the offset removal is slow.

Q8.3.3 If a controller exhibits ringing, whether there is too much proportional action or too much integral action can be assessed by plotting the CV and the output of the controller (or the measured value of the MV for a fast responding actuator) on the same graph and evaluating the lag between the two. When y and c are in phase (i.e., the peaks or peaks and valleys occur very nearly at the same point in time), this indicates that the ringing is being caused by excessive proportional action. When u significantly lags y, this indicates that there is excessive integral action.

8.4 Effect of Tuning Parameters for PID Control

PID controllers are used for sluggish processes, e.g., certain temperature and composition control loops. Derivative action is the least understood and the least used mode in a PID controller, but can provide significant benefits in specific cases. Section 7.8 indicates that derivative action improves control performance for processes that have deadtime-to-time constant ratios greater than one. Because derivative action opposes the slope of the CV, it reduces the oscillatory nature of the feedback response (Section 7.7). The effect of proportional and integral action on the feedback behavior of a PID controller is similar to that observed for the PI controller studied in Examples 8.4 and 8.5.

Figure 8.4.1a shows the results of PID and PI control applied to a FOPDT process: (K_p 1, $_p$ 1, $_p$ 0.1). Figure 8.4.1b shows the results of PID control and PI control applied to another FOPDT process with more deadtime: (K_p 1, $_p$ 1, $_p$ 2). These results support the conclusion that derivative action is useful for processes that have large deadtime-to-time constant ratios. Figure 8.4.2 shows a case that has too much derivative action in the PID controller. Note that a "stair-step" feedback response indicates that too much derivative action is being used. The stair-step behavior is caused because, as the process moves toward the setpoint, excessive derivative action causes the process to stall or level out. After the process stalls, the



Time Time

(a) (b)

Figure 8.4.1 Comparison between a PI and PID controllers for a process with (a) a small deadtime and (b) a large deadtime.

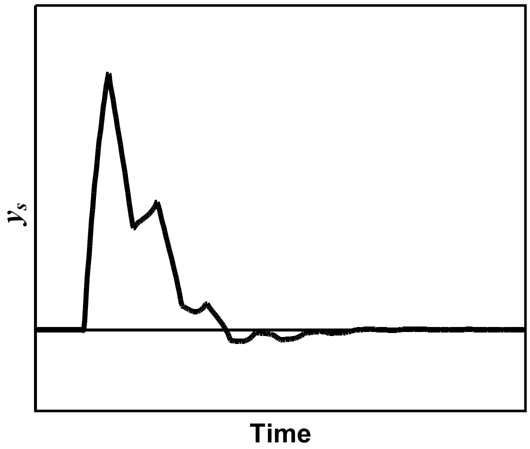


Figure 8.4.2 The control performance of a PID controller with too much derivative action.

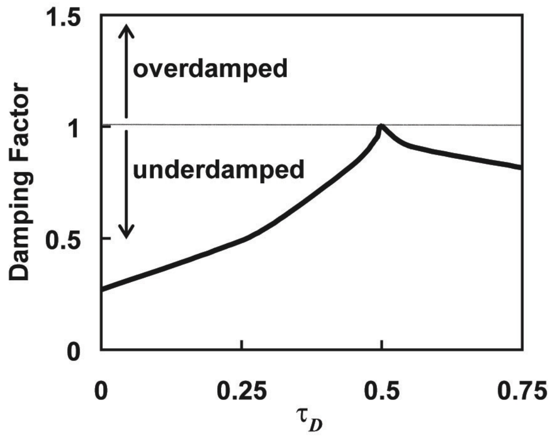


Figure 8.4.3 The effect of the controller derivative time on the damping factor for a PID controller applied to a FOPDT process (K_p=1; $_p$=1; $_p$=2, K_c=1; $_I$=2).

proportional and integral action act on the process to move the CV toward the setpoint. When this occurs, the derivative of the CV builds up and the derivative action acts against it, causing the response to stall again, resulting the stair-step effect.

Example 8.7 Evaluation of the Effect of $_D$ on the Closed-Loop Dynamics for a PID controller applied to a FOPDT Process

Problem Statement. Evaluate the effect of $_D$ on the closed-loop dynamics of a PID controller (K_c=1; $_I$=2) applied to a FOPDT process (K_p = 1; $_p$ = 1; $_p$ = 2).

Solution. Substituting the transfer function for a PID controller, i.e.,

$$G_c(s) \quad K_c \; 1 \quad \frac{1}{_I s} \quad _D s$$

and the transfer function for a FOPDT process model with the Padé approximation for deadtime into the characteristic equation (Equation 6.6.8) result in

$$G_c(s)G_p(s) \quad 1 \quad K_c \; 1 \quad \frac{1}{_I s} \quad _D s \; \frac{K_p(1 \quad \tfrac{1}{2} \quad _p s)}{(\; _p s \quad 1)(1 \quad \tfrac{1}{2} \quad _p s)} \quad 1 \quad 0$$

Substituting the controller and process parameters results in the following cubic equation in s after rearranging.

$$2(1 \quad _D)s^3 \quad 2(\; _D \quad 1)s^2 \quad 3s \quad 1 \quad 0$$

Because the coefficients must all be positive for stability, the coefficient for s^3 indicates that for stability $_D$ 1.

As $_D$ is increased from 0 to 0.75, one of the poles of the characteristic equation, which is a real pole, changes from -0.4 to -6.0. The other two poles form a complex conjugate pair that determine the oscillatory nature of the closed-loop response. Figure 8.4.3 plots the damping factor for the complex conjugate pair as a function of the derivative time. As $_D$ is increased from 0 to 0.5, the damping factor increases, indicating that the oscillatory nature of the system is reduced by the derivative action. On the other hand, as the derivative time is increased above 0.5, the oscillatory nature of the response increases. Derivative action reduces the oscillatory nature of the response up to a point, but, above that point, a further increase in $_D$ increases oscillatory behavior, which corresponds to the results shown in Figure 8.4.2.

Self-Assessment Questions.

Q8.4.1 How can you tell when too much derivative action is being used for a PID controller?

Q8.4.2 How does the derivative action affect the closed-loop damping factor?

Self-Assessment Answers

Q8.4.1 If the closed-loop response of a process exhibits a stair-step behavior, it has too much derivative action.

Q8.4.2 An increase in the derivative time increases the closed-loop damping factor up to a point. Above that point the closed-loop damping factor decreases with increasing values of $_D$.

8.5 Tuning Criteria and Performance Assessment

Tuning Criteria. There are a number of mathematical expressions that can be used as a basis for PID tuning:

Integral Absolute Error (***IAE***)

$$IAE \quad {}_0 \left| y_{sp}(t) \quad y_s(t) \right| dt$$

Integral Time Absolute Error (***ITAE***)

$$ITAE \quad {}_0\, t \left| y_{sp}(t) \quad y_s(t) \right| dt$$

Integral Square Error (***ISE***)

$$ISE \quad {}_0 [y_{sp}(t) \quad y_s(t)]^2 \, dt$$

Integral Time Square Error (***ITSE***)

$$ITSE \quad {}_0\, t [y_{sp}(t) \quad y_s(t)]^2 \, dt$$

Each of these statistical measures values the error from setpoint differently. *ITAE* and *ITSE* penalize later deviations more severely than *IAE* and *ISE*. *ISE* and *ITSE* penalize larger deviations more severely than *IAE* and *ITAE*.

Table 8.1
Several Performance Statistics as a Function of Decay Ratio for the Endothermic CSTR (Example 3.5)

Decay Ratio	IAE	ITAE	ISE	ITSE
1/1.5	39.6	1244	31.1	470
1/2	28.3	628	22.8	231
1/3	20.9	347	17.8	117
1/4	19.8	387	16.8	92.8
1/5	20.7	503	16.8	91.2
1/6	22.0	635	17.1	97.4
1/8	24.9	903	17.9	119
1/10	27.4	1141	18.8	145

Self-Assessment Question

Q**8.5.1** How is using a minimum IAE tuning criterion different from using a minimum ITAE tuning criterion?

Self-Assessment Answer

Q8.5.1 Using a minimum ITAE penalizes long time deviations from setpoint more than a minimum IAE tuning criterion, resulting in a faster settling time for the minimum ITAE tuned controller.

8.6 Classical Tuning Methods

A wide range of PID tuning methods has been proposed. This section considers two of the earliest methods: Cohen and Coon method[2] and the Ziegler-Nichols method[3]. In addition, a more recent technique (the Ciancone and Marlin method[5]) is also considered.

Each of these methods is based upon a preset tuning criterion. As a result, even when they are used industrially, they are usually used as initial controller settings. Then control engineers use their knowledge of the nonlinearity and the severity of disturbances for the particular control loop in question to adjust the controller tuning to meet the proper compromise between reliability and performance for that particular loop. The methods presented here can be used for initial controller settings (Example 8.2) using process knowledge. These methods also provide insight into the relative tuning of P-only, PI and PID controllers. Sections 9.8-9.11 present the recommended tuning procedures for industrial control loops.

Cohen and Coon Method. The **Cohen and Coon** approach[2] assumes that a FOPDT model of the process (i.e., the combined effect of the actuator, process and sensor) is available. The Cohen and Coon parameters for P-only, PI and PID controllers are listed in Table 8.2. These results are based on a combination of QAD (i.e., a decay ratio of ¼), minimum ISE and minimum offset tuning for a FOPDT process model.

Table 8.2
Cohen and Coon PID Settings Based on a FOPDT Model[2]

Controller	K_c	τ_I	τ_D
P-only	$\frac{1}{K_p}\frac{\tau_p}{\theta_p}\left(1+\frac{\theta_p}{3\tau_p}\right)$		
PI	$\frac{1}{K_p}\frac{\tau_p}{\theta_p}\left(0.9+\frac{\theta_p}{12\tau_p}\right)$	$\frac{\theta_p\left(30+3\frac{\theta_p}{\tau_p}\right)}{9+20\frac{\theta_p}{\tau_p}}$	
PID	$\frac{1}{K_p}\frac{\tau_p}{\theta_p}\left(\frac{16+3\frac{\theta_p}{\tau_p}}{12}\right)$	$\frac{\theta_p\left(32+6\frac{\theta_p}{\tau_p}\right)}{13+8\frac{\theta_p}{\tau_p}}$	$\frac{4\theta_p}{11+2\frac{\theta_p}{\tau_p}}$

Ziegler-Nichols Tuning. Ziegler-Nichols (ZN) tuning[3] uses experimental measurements of the **ultimate gain**, K_u and the **ultimate period**, P_u to calculate the controller settings. The ultimate parameters are obtained by applying a P-only controller to attain sustained oscillations and measuring the resulting period of the oscillations and noting the gain of the P-only controller. The procedure is as follows

1. Turn off integral and derivative action to give a P-only controller.
2. Increase K_c until oscillations are sustained for a relatively small setpoint change (Figure 8.6.1).
3. K_u is the P-only controller gain that results in the sustained oscillations.
4. P_u is the period of the sustained oscillations (Figure 8.6.1).
5. Calculate the controller settings using Table 8.3.

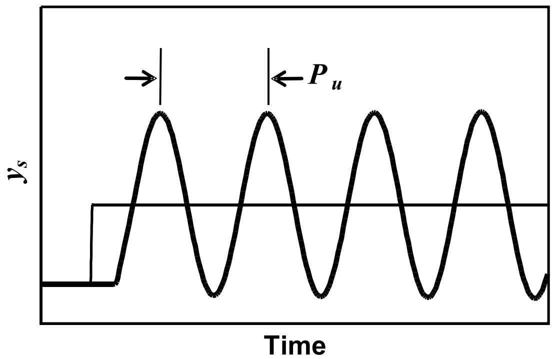


Figure 8.6.1 The CV for a ZN ultimate test, i.e., operation at sustained oscillations with a P-only controller.

The Ziegler-Nichols settings are based on a QAD tuned response. A QAD response for a second-order process corresponds to a damping factor (ζ) of 0.22 with a 50% overshoot; therefore, ZN setting are relatively aggressive. According to these settings, a PID controller uses a 33% larger K_c than a corresponding PI controller (i.e., K_c is equal to $0.6K_u$ for PID compared to $0.45K_u$ for PI) and a P-only controller uses a K_c that is 10% larger than a corresponding PI controller (i.e., K_c is equal to $0.45K_u$ for PI compared to $0.5K_u$ for P-only) .

Table 8.3

Ziegler-Nichols PID Settings[3]

Controller	K_c	$_I$	$_D$
P	$0.5K_u$	—	—
PI	$0.45K_u$	$P_u/1.2$	—
PID	$0.6K_u$	$P_u/2$	$P_u/8$

Ciancone and Marlin Tuning. The Ciancone and Marlin controller tuning approach[4,5] uses FOPDT parameters along with dimensionless tuning parameter plots to determine the controller settings. This approach was derived using the closed-loop transfer function to develop dimensionless relationships that can be used to select tuning parameters. Consider a FOPDT model of the process

$$G_a(s)\,G_p(s)\,G_s(s) \qquad G_p\ (s) \qquad \frac{K_p e^{\quad _p s}}{_p s \quad 1}$$

Then, the closed-loop transfer function for disturbance rejection is given by

$$\frac{Y(s)}{D(s)} \qquad \frac{G_d(s)}{G_c(s)\,G_p(s) \quad 1} \qquad \frac{G_d(s)}{K_c \quad 1 \quad \frac{1}{_I\, s} \qquad _D\, s \qquad \frac{K_p e^{\quad _p s}}{_p s \quad 1} \qquad 1} \tag{8.6.1}$$

The modified Laplace transform variable ($\bar{s}$) is defined as

$$\bar{s} \qquad s(\ _p \qquad _p)$$

Rearranging this equation to solve for the Laplace transform variable (s) in terms of $\bar{s}$ yields

$$s \qquad \frac{\bar{s}}{_p \qquad _p}$$

Using this equation to eliminate s from Equation 8.3.1 yields

$$\frac{Y(\bar{s})}{D(\bar{s})} \qquad \frac{G_d(\bar{s})}{K_c K_p \quad 1 \quad \frac{1}{_I\, \bar{s}\,/(\ _p \qquad _p)} \qquad \frac{_D\, \bar{s}}{_p \qquad _p} \qquad \frac{e^{\quad _p \bar{s}/(\ _p \quad _p)}}{1 \quad \frac{_p \bar{s}}{_p \qquad _p}} \qquad 1}$$

Note that

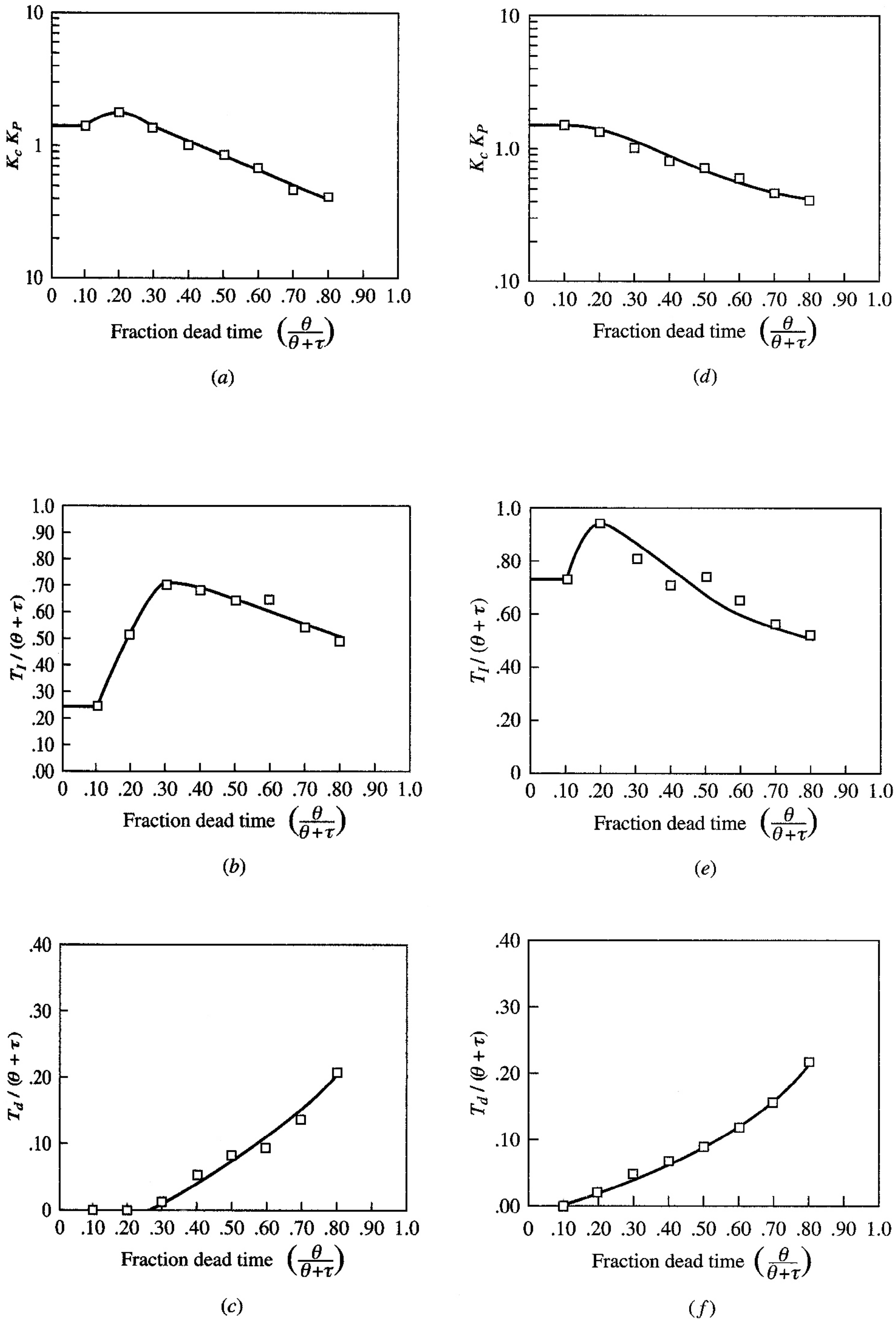


Figure 8.6.2 Ciancone and Marlin's correlation functions for dimensionless tuning constants. For disturbance rejection (a) K_c (b) $_I$ (c) $_D$. For setpoint changes (d) K_c (e) $_I$ (f) $_D$. Note that T_d and T_I in this figure correspond to $_D$ and $_I$. This figure is reprinted with permission from McGraw-Hill Publishing Co.

$$\frac{\theta_p}{(\theta_p+\tau_p)} = 1 - \frac{\tau_p}{\tau_p+\theta_p}$$

and [$\theta_p/(\theta_p+\tau_p)$] is referred to as the **fractional deadtime**[4]. Therefore, the process model is converted from six parameters (K_c, τ_I, τ_D, K_p, τ_p, θ_p) into four parameters [K_pK_c, $\theta_p/(\theta_p+\tau_p)$, $\tau_I/(\theta_p+\tau_p)$, $\tau_D/(\theta_p+\tau_p)$]. As a result, dimensionless forms of the tuning parameters can be developed, i.e.,

$$\text{Dimensionless Gain} = K_cK_p$$

$$\text{Dimensionless Reset Time} = \frac{\tau_I}{\theta_p+\tau_p}$$

$$\text{Dimensionless Derivative Time} = \frac{\tau_D}{\theta_p+\tau_p}$$

Ciancone and Marlin[5] correlations for the dimensionless gain, reset time and derivative time as a function of the fractional deadtime are shown in Figure 8.6.2. These correlations are based on tuning for minimum IAE performance considering ±25% error in the model parameters (i.e., variations in K_p, τ_p and θ_p). These results show some of the differences between tuning for setpoint changes and disturbance rejection. The correlations for disturbances and setpoint changes are similar, but the dimensionless reset time for disturbances is quite different at low values of fractional deadtime [$\theta_p/(\theta_p+\tau_p) < 0.3$]. Note that more integral action is used for disturbances than for setpoint changes. There is also some difference between the dimensionless derivative time at low fractional deadtime for setpoint tracking and disturbance rejection.

While the FOPDT model is flexible enough to reasonably model a wide range of real processes, developing accurate FOPDT models for industrial processes can be a difficult and time-consuming process. Also, the assumption of ±25% parameter uncertainty is arbitrary. Relatively linear processes with low disturbance levels can be expected to result in FOPDT parameters that have much less parameter variation than ±25%. On the other hand, highly nonlinear processes are expected to result in FOPDT parameter variations well in excess of ±25% for major disturbance upsets. While this approach is certainly interesting and provides insight into the PID tuning process, it should be viewed as providing only initial estimates of tuning parameters.

Overview. The Cohen and Coon method and the Ciancone and Marlin method require a FOPDT process model, which is difficult and time consuming to develop. The settings for these tuning methods are based on a tuning criterion that may not be consistent with the requirements of the control loop under consideration.

The Ziegler-Nichols method requires an ultimate test that can unnecessarily upset the process. It does have the advantage that it is a direct measurement on the process. But once again, the controller settings are based on QAD tuning, and nonlinear processes using these settings can lead to ringing or unstable behavior. A very large number of tuning methods have been proposed. Most of these methods suffers from one or more of the same limitations as the three tuning methods presented here.

Example 8.8 Comparison of Classical Tuning Methods

Problem Statement. Determine the PID controller setting for the Cohen and Coon and the Ciancone and Marlin methods using the FOPDT model determined in Example 6.10 (K_p = -1.5; $_p$ = 1.43; $_p$ = 2.43). Use the Ciancone and Marlin settings for disturbance rejection.

Solution. For the Ciancone and Marlin method, the fractional deadtime is the independent variable for the dimensionless correlations and has a value of 0.63 for this case. The values of the dimensionless tuning parameters can be read directly from Figure 8.6.2 (i.e., the dimensionless gain is 0.60, the dimensionless reset time is 0.60, and the dimensionless derivative time is 0.12). The controller settings for the Ciancone and Marlin method result from using the definitions of the dimensionless tuning parameters to solve for the PID tuning parameters. Using the FOPDT model parameters from Example 6.10 for the Cohen and Coon method (Table 8.2) results in the following PID settings.

	K_c	0.690		K_c	0.90
Cohen and Coon Method	$_I$	3.86	**Ciancone and Marlin:**	$_I$	2.32
	$_D$	0.675		$_D$	0.463

Note that the Cohen and Coon settings are less aggressive than the Ciancone and Marlin settings in this case. Also, note that even though the process gain is negative, the controller gain is specified as a positive number because the effect of the negative gain is addressed by whether a direct-acting or a reverse-acting controller is used (Section 7.6).

Example 8.9 Initial Controller Tuning Estimates

Problem Statement. Using the Ciancone and Marlin method, develop initial controller settings for a PI controller applied to a temperature control loop on a steam-heated heat exchanger, which operates under regulatory control. By observing the process operation, it has been estimated that the process gain is approximately 0.01 ºF-hr/lb. The response time of the process has been estimated equal to approximately 10 minutes. Further, by observing the process, it has been determined that the amount of deadtime in the system is relatively small, i.e., the process responds quickly to changes in the flow rate of steam to the heat exchanger.

Solution. Figure 8.6.2d, assuming a fractional deadtime equal to approximately zero, yields a value of K_cK_p of 1.3. This results in a controller gain of 130 lb/hr-ºF. From Figure 8.6.2e, the dimensionless reset time is 0.73 and the dimensionless derivative time is zero. From the response time, the process time constant is estimated as 2.5 minutes. Neglecting the deadtime, the reset time is calculated as 1.83 minutes. Therefore, K_c=130 lb-ºF/h, $_I$=1.83 min and $_D$=0.

Self-Assessment Questions

Q8.6.1 To apply the Cohen and Coon method or the Ciancone and Marlin method, what information about the process is required?

Q8.6.2 How do you determine the ultimate period and ultimate gain used in the Ziegler-Nichols tuning?

Q8.6.3 Which of the three tuning methods considered in this section provides the most conservative tuning settings?

Self-Assessment Answers

Q8.6.1 Both the Cohen and Coon method and the Ciancone and Marlin method require a FOPDT model to determine the controller settings.

Q8.6.2 The ultimate gain and the ultimate period used by the ZN method are determined by performing an ultimate test on the process. For an ultimate test, apply a P-only controller and increase K_c until sustained oscillations result. K_c is the ultimate gain and the period of the oscillations is the ultimate period.

Q8.6.3 The Cioanione and Marlin method provide the more conservative setting compared to the Cohen and Coon method and the ZN method because the Ciancone and Marlin method considers ±25% parameter uncertainty.

8.7 Controller Tuning by Pole Placement[AT]

One way to tune a PID controller using transfer function process models is to use a pole-placement approach (also known as **direct synthesis**). For **pole placement** applied to PID tuning, the poles of the closed-loop response (i.e., the desired dynamic response of the closed-loop system) are specified, and the PID tuning parameters are then calculated. Tuning the control loop using pole placement requires the user to specify the desired closed-loop dynamics (e.g., by specifying the closed-loop natural period, $_n$, and the damping factor,). Consider the characteristic equation for a first-order process with a PI controller, i.e.,

$$K_c \; 1 \; \frac{1}{\; s} \; \frac{K_p}{\;_p s \;\; 1} \;\; 1 \;\; 0$$

Rearranging into the standard form for a second-order equation results in

$$\frac{\;_I \;\;_p}{K_c K_p} s^2 \qquad _I \; 1 \; \frac{1}{K_c K_p} \;\; s \;\; 1 \;\; 0$$

Then, the closed-loop time constant and damping factor are given by

$$_n \qquad \sqrt{\frac{\;_I \;\;_p}{K_c K_p}} \tag{8.7.1}$$

$$½\sqrt{\frac{\;_I}{\;_p}} \; \sqrt{K_c K_p} \qquad \frac{1}{\sqrt{K_c K_p}} \tag{8.7.2}$$

Assume that we have specified the values of $_p$ and (i.e., the dynamic response of the closed-loop system). Specifying $_p$ and is, in effect, specifying the poles of the closed-loop response, which can be calculated directly by applying the quadratic formula to the following equation.

$$(\;_n)^2 s^2 \quad 2 \;_n \quad s \;\; 1 \;\; 0$$

Solving Equations 8.7.1 and 8.7.2 simultaneously for K_c and $_I$ yields

$$K_c \quad \frac{1}{K_p} F \qquad \textbf{8.7.3}$$

$$ {}_I \quad \frac{(\;{}_p)^2}{{}_p} F \qquad \textbf{8.7.4}$$

where

$$F \quad 2 \; \frac{{}_p}{{}_p} \quad 1 \qquad \textbf{8.7.5}$$

F must be greater than zero to maintain proper values of K_c and $_I$. For a conservatively tuned controller ($_n$= $_p$ and =1), F is equal to 1. For an aggressively tuned controller ($_n$= $_p$/4 and =0.5), F is equal to 3. Therefore, F is an indication of the aggressiveness of the controller.

Example 8.10 Application of Pole Placement for Tuning a PI Controller

Problem Statement. Determine the PI controller settings for a first-order process (K_p = 3; $_p$ = 10) if it is desired to obtain a closed-loop damping factor of 0.4 and a closed-loop time constant of 3.

Solution. From Equation 8.7.5, F is equal to 1.67 for this case. Then, using the value of K_p, K_c is calculated equal to 0.556. Likewise, $_I$ is calculated directly as 1.50. The two closed-loop poles that correspond to a damping factor of 0.4 and a time constant of 3 are (-.133 0.306 *i*); therefore, specifying the closed-loop time constant and damping factor is equivalent to selecting the poles of the closed-loop response.

Example 8.11 Application of Pole Placement for Tuning a PI Controller

Problem Statement. Determine the PI controller settings for a first-order process (K_p = 3; $_p$ = 10) if it is desired to obtain a closed-loop damping factor of 0.4 and a closed-loop time constant of 10.

Solution. From Equation 8.7.5, F is equal to -0.2 for this case. The negative sign for F indicates that the specifications for $_p$ and were not consistent. For this case, to obtain a closed-loop damping factor of 0.4, the closed-loop natural period would be significantly smaller than $_p$. Therefore, when using this approach, consistent performance specification must be used because the values of $_p$ and are not completely independent.

The concept of pole placement leads to the derivation of a general transfer function representation of a controller based on a specified closed-loop response and a transfer function for the process. Let us now assume that we want a specific second-order response for a first-order process. Then the closed-loop transfer function for setpoint changes (Equation 7.2.7) is set equal to a specified second-order transfer function, i.e.,

$$\frac{G_c(s)\dfrac{K_p}{{}_p s \quad 1}}{G_c(s)\dfrac{K_p}{{}_p s \quad 1} \quad 1} \qquad \frac{1}{{}_n)^2 s^2 \quad 2 \; {}_n \quad s \quad 1}$$

Here ${}_n$ and are specified by the user to set the desired closed-loop second-order response. Because an offset-free response is desired, the gain of the specified second-order response is one. Rearranging and solving for $G_c(s)$ yields

$$G_c(s) \qquad \frac{\dfrac{1}{(\;{}_n)^2 s^2 \quad 2 \; {}_n \quad s \quad 1}}{\dfrac{K_p}{{}_p s \quad 1} \quad 1 \quad \dfrac{1}{(\;{}_n)^2 s^2 \quad 2 \; {}_n \quad s \quad 1}}$$

Combining the two terms in the denominator and simplifying yields

$$G_c(s) \qquad \frac{{}_p s \quad 1}{K_p \; {}_n s \quad {}_n s \quad 2}$$

Because the formulation of the problem was based on a general unknown controller, the functional form of this controller does not correspond to a PID controller. Earlier in this section, it was demonstrated that a PI controller could, under certain circumstances, be tuned for a specified second-order closed-loop response. As a result, this general controller reduces to a PI controller under certain conditions. There are limits on the closed-loop performance specifications that you can choose when using a PID-type controller, and these limits correspond to the conditions necessary to transform this general controller into a PI controller.

Consider the problem of deriving a general controller, $G_c(s)$, for a general closed-loop performance specification, $S_{cl}(s)$. Once again, we can equate the closed-loop transfer function for a setpoint change (Equation 7.2.7) to the user-selected performance specification:

$$\frac{G_c(s)G_p(s)}{G_c(s)G_p(s) \quad 1} \qquad S_{cl}(s)$$

Rearranging and solving for $G_c(s)$ yields

$$G_c(s) \qquad \frac{S_{cl}(s)}{G_p(s)[1 \quad S_{cl}(s)]} \qquad \textbf{8.7.6}$$

Example 8.12 Derivation of a General Controller for a First-Order Process

Problem statement. Derive the transfer function for the controller for a first-order process that provides a specified first-order response.

Solution. In this case,

$$S_{cl}(s) \quad \frac{1}{{}_{p}s \quad 1}$$

Applying Equation 8.7.6 results in

$$G_c(s) \quad \frac{\dfrac{1}{{}_{p}s \quad 1}}{\dfrac{K_p}{{}_{p}s \quad 1} \quad 1 \quad \dfrac{1}{{}_{p}s \quad 1}}$$

Simplifying,

$$G_c(s) \quad \frac{{}_{p}s \quad 1}{K_p \; {}_{p}s} \quad \frac{{}_{p}}{K_p \; {}_{p}} \quad 1 \quad \frac{1}{{}_{p}s}$$

which is equivalent to a PI controller with the following settings.

$$K_c \quad \frac{{}_{p}}{K_p \; {}_{p}} \qquad\qquad {}_{I} \qquad {}_{p}$$

As the specifications for the closed-loop response become more aggressive (i.e., the ratio of $_p$ to $_p$ becomes larger), K_c increases proportionally. If the reset time is less than the process time constant, an underdamped response results. In addition, the general controller can also be reduced to a P-only controller by placing restrictions on the values of $_p$.

Overview. Pole placement provides an important perspective for PID controller tuning and provides an important insight: closed-loop performance specifications must be consistent. Moreover, the concept of pole placement can be extended to a methodology for developing a general controller (i.e., not limited to a PID-type controller) that is designed to meet preset performance specifications.

While pole placement provides useful insights and fundamental understanding, it is not usually used to tune PID controllers in the process industries or to develop general controllers that are applied in the process industries. This results primarily because process models are not generally available, and the time required to develop a process model far exceeds the time required to tune a loop using other methods. In addition, PID controllers are a standard function on control computers whereas implementation of a general controller requires the development and implementation of custom software. Also, pole placement cannot be used for process models with time delay ($e^{\quad {}_{p}s}$) or right half-plane zeros as this results in unstable controllers. Sections 9.8-9.11 present the recommended procedures for tuning PID controllers for the process industries.

Self-Assessment Questions

Q8.7.1 For which specific cases does the pole placement method result in a PID controller?

Q8.7.2 When using the pole placement method, can you independently specify the closed-loop natural period and damping factor?

Q8.7.3 What restrictions are associated with using the pole placement method for industrial control loops?

Self-Assessment Answers

Q8.7.1 Pole placement results in a PID-type controller when the specified response is in the form of a first-order response and the process model is also first order.

Q8.7.2 From Example 8.11, $_n$ and must be consistent, otherwise F<0.

Q8.7.3 To use pole placement for an industrial control loop, you must have a process model and usually the controller determined will not be a PID-type controller. In addition, pole placement cannot be used for process models containing deadtime and right-half plane zeros.

8.8 PID Tuning Based on Internal Model Control (IMC)^{AT}

The **internal model control** (**IMC**) algorithm[6,7] is a convenient means to tune a PID controller if a process model is available. The IMC algorithm can be derived by using a logic flow diagram that is equivalent to a conventional feedback control loop. Figure 8.8.1 shows a traditional feedback control loop in which the disturbance effect, $D(s)G_d(s)$, is added to the process effect [i.e., the output of $G_p(s)$]. Note that $G_p(s)$ is equal to the product of $G_p(s)$ and $G_a(s)$ from Figure 7.2.1. Figure 8.8.2 shows a logic flow diagram for an IMC controller in which $\underline{G}_p(s)$ is a model of the process, i.e., an approximation of the process.

To establish the conditions for equivalence between an IMC controller (Figure 8.8.2) and a conventional feedback controller (Figure 8.8.1), we will equate the closed-loop transfer functions. The closed-loop transfer function for a conventional controller is given by Equation 7.2.7, which yields the following results when applied to the case represented by Figure 8.8.1 (i.e., $G_s(s)$=$G_a(s)$=1).

$$Y(s) \quad \frac{G_c(s)G_p(s)}{G_c(s)G_p(s) \quad 1} Y_{sp}(s) \quad \frac{G_d(s)}{G_c(s)G_p(s) \quad 1} D(s) \qquad \textbf{8.8.1}$$

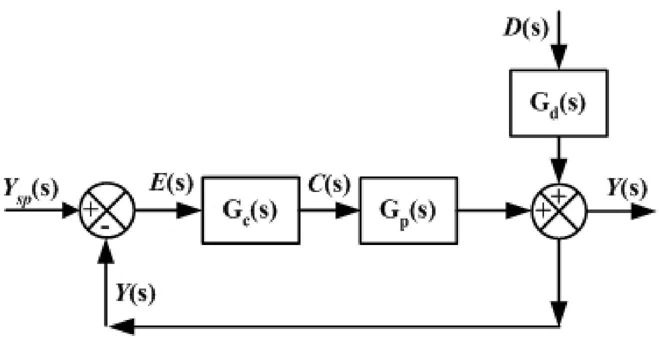


Figure 8.8.1 Block diagram of a traditional feedback control loop.

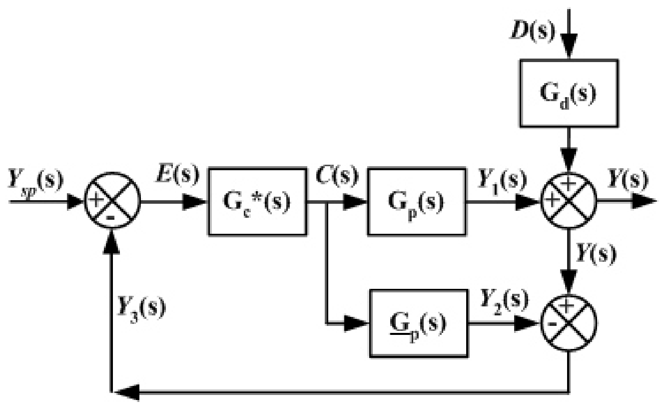


Figure 8.8.2 Block diagram for the IMC algorithm.

Now, let's derive the closed-loop transfer functions for setpoint changes and disturbance rejection for the IMC controller (Figure 8.8.2). First, the value of the error from setpoint, $E(s)$, in Figure 8.8.2 is expressed as

$$E(s) \quad Y_{sp}(s) \quad Y_3(s) \tag{8.8.2}$$

Then $Y_3(s)$ from Figure 8.8.2 can be written as a function of $E(s)$, i.e.,

$$Y_3(s) \quad Y(s) \quad Y_2(s) \quad Y(s) \quad G_c^*(s)\underline{G}_p(s)E(s)$$

Substituting this equation into Equation 8.8.2 and solving for $E(s)$ yields

$$E(s) \quad \frac{Y_{sp}(s) \quad Y(s)}{1 \quad G_c^*(s)\underline{G}_p(s)} \tag{8.8.3}$$

Next, $Y(s)$ is expressed as the sum of $G_d(s)D(s)$ and $Y_1(s)$. Then an expression for $Y_1(s)$ in terms of $E(s)$ is substituted yielding

$$Y(s) \quad G_d(s)D(s) \quad Y_1(s) \quad G_d(s)D(s) \quad G_c^*(s)G_p(s)E(s)$$

Substituting for $E(s)$ using Equation 8.8.3 yields the following equation after collecting terms and rearranging

$$Y(s) \quad \frac{G_c^*(s)G_p(s)}{1 \quad G_c^*(s)[G_p(s) \quad \underline{G}_p(s)]}Y_{sp}(s) \quad \frac{[1 \quad G_c^*(s)\underline{G}_p(s)]G_d(s)}{1 \quad G_c^*(s)[G_p(s) \quad \underline{G}_p(s)]}D(s) \tag{8.8.4}$$

Equating the coefficients of $Y_{sp}(s)$ for Equations 8.8.1 and 8.8.4 yields

$$\frac{G_c(s)G_p(s)}{G_c(s)G_p(s) \quad 1} \quad \frac{G_c^*(s)G_p(s)}{1 \quad G_c^*(s)[G_p(s) \quad \underline{G}_p(s)]}$$

Solving for $G_c(s)$ yields

$$G_c(s) \quad \frac{G_c^*(s)}{1 \quad G_c^*(s)\underline{G}_p(s)} \tag{8.8.5}$$

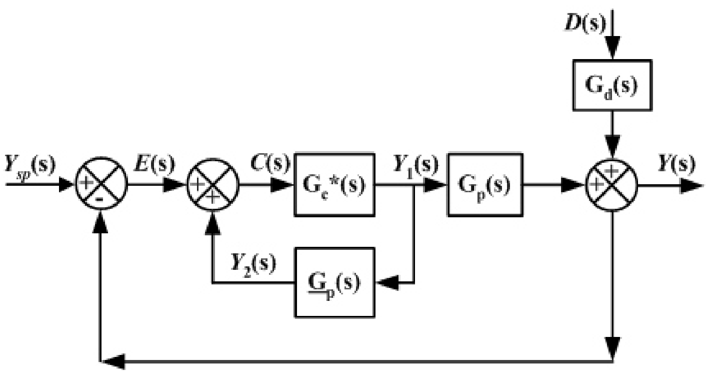


Figure 8.8.3 Alternative block diagram for the IMC algorithm.

This relationship between $G_c(s)$ and $G_c^*(s)$ indicates that the controller block in Figure 8.8.1 can be replaced by a positive feedback loop containing the IMC controller, $G_c^*(s)$, as shown in Figure 8.8.3. Therefore, Figure 8.8.3 is equivalent to Figure 8.8.2. Note that in Figure 8.8.3, the internal feedback is positive, but this does not cause instability because stabilization is provided by the outer feedback loop. Remember that based on the previous derivation, Figure 8.8.2 is equivalent to a general feedback loop (Figure 8.8.1) where $G_c(s)$ in Figure 8.8.1 is given by Equation 8.8.5.

Ideal Case. Let's assume that the process model is known exactly; therefore,

$$\underline{G}_p(s) \quad G_p(s)$$

Then Equation 8.8.4 becomes

$$Y(s) \quad G_c^*(s) G_p(s) Y_{sp}(s) \quad [1 \quad G_c^*(s) G_p(s)] D(s) \qquad \mathbf{8.8.6}$$

If the IMC controller $[G_c^*(s)]$ is chosen such that

$$G_c^*(s) \quad \frac{1}{G_p(s)}$$

Then $$Y(s) \quad Y_{sp}(s)$$

which indicates perfect setpoint tracking and perfect disturbance rejection.

Non-Ideal Case. Unfortunately, it is impossible to specify an approximate model $[\underline{G}_p(s)]$ that perfectly matches the process behavior $[G_p(s)]$. Moreover, even if the process model were known exactly, it is not possible to form the inverse of a process model if the model contains a time delay or a right half-plane zero. Morari and coworkers[6] have suggested that the process model, in such cases, is factored into "invertible" and "non-invertible" parts, i.e.,

$$\underline{G}_p(s) \quad \underline{G}_p \quad \underline{G}_p$$

where $\underline{G}_p(s)$ only contains invertible parts of $\underline{G}_p(s)$ while $\underline{G}_p(s)$ contains the non-invertible parts of $\underline{G}_p(s)$. The non-invertible parts of a transfer function correspond to the parts of a transfer function, which when inverted, result in unstable or unrealizable controllers (e.g., deadtime and right-half plane zeros). For example, consider the following model

$$\underline{G}_p(s) \quad \frac{K_p(1 \quad {}_1 s) e^{\quad s}}{(\quad {}_2 s \quad 1)}$$

Hence,

$$\underline{G}_p(s) \quad \frac{K_p}{{}_2 s \quad 1} \qquad \underline{G}_p(s) \quad e^{\quad s}(1 \quad {}_1 s)$$

Morari and co-workers[7] further suggest that the controller can be designed using only the invertible part of the transfer function of the process model and that a "filter" is added, e.g.,

$$G_c^*(s) = \frac{1}{\underline{G}_p(s)} f(s)$$

where $f(s)$ has the following form

$$f(s) = \frac{1}{(\tau_f s + 1)^p}$$

Here τ_f is chosen such that it is equivalent to the desired closed-loop time constant while p is chosen so that the relative order of the controller $[G_c^*(s)]$ is less than or equal to one. The relative order of a controller is based on the order of the numerator and the order of the denominator, i.e.,

$$G_c^*(s) = \frac{N(s)}{D(s)}$$

The relative order of a controller is the order of the numerator $[N(s)]$ minus the order of the denominator $[D(s)]$. Normally, p is set equal to one.

Thus, an IMC controller can be designed by using the invertible portion of the process transfer function $[\underline{G}_p(s)]$, which is assumed to be known, and a filter time constant (τ_f), which represents the desired time constant for the closed-loop response.

Assuming that the transfer function of the process model is known perfectly [i.e., $\underline{G}_p(s) = G_p(s)$] and considering the invertible and non-invertible portions of the process model, Equation 8.8.4 becomes

$$Y(s) = \frac{f(s)}{\underline{G}_p(s)} \underline{G}_p(s) \underline{G}_p(s) Y_{sp}(s) + \left[1 - \frac{f(s)\, \underline{G}_p(s) \underline{G}_p(s)}{\underline{G}_p(s)}\right] D(s)$$

Simplifying, yields

$$Y(s) = f(s)\, \underline{G}_p(s)\, Y_{sp}(s) + [1 - f(s)\underline{G}_p(s)]\, D(s)$$

This results indicates that perfect control is not possible even if the process model is perfectly known when there are non-invertible portions of the process model.

Example 8.13 IMC Applied to a FOPDT Process

Problem Statement. Develop an IMC controller for a general FOPDT process (K_p, τ_p, θ_p).

Solution. The process transfer function is given by

$$\underline{G}_p(s) \quad \frac{K_p\, e^{\quad {}_p s}}{{}_p s \quad 1}$$

Therefore,
$$\underline{G}_p(s) \quad \frac{K_p}{{}_p s \quad 1}; \qquad \underline{G}_p(s) \quad e^{\quad {}_p s}$$

Using a first-order tuning filter, the IMC controller for this case becomes

$$G_c^*(s) \quad \frac{f(s)}{\underline{G}_p(s)} \quad \frac{{}_p s \quad 1}{K_p(\ {}_f s \quad 1)}$$

The equivalent conventional feedback controller is given by

$$G_c(s) \quad \frac{G_c^*(s)}{1 \quad G_c^*(s)\underline{G}_p(s)} \quad \frac{\dfrac{{}_p s \quad 1}{K_p(\ {}_f s \quad 1)}}{1 \quad \dfrac{{}_p s \quad 1}{K_p(\ {}_f s \quad 1)} \quad \dfrac{K_p e^{\quad {}_p s}}{{}_p s \quad 1}}$$

Simplifying,

$$G_c(s) \quad \frac{{}_p s \quad 1}{K_p(\ {}_f s \quad 1 \quad e^{\quad {}_p s})}$$

Note that $e^{\quad {}_p s}$ is not realizable, i.e., because it is located in the denominator of the transfer function, its evaluation requires process measurements before they are measured. Using a first-order Taylor series expansion,

$$e^{\quad {}_p s} \quad 1 \quad s \quad {}_p$$

Substituting this approximation into the last equation for $G_c(s)$ yields

$$G_c(s) \quad \frac{{}_p s \quad 1}{K_p(\ {}_f \qquad {}_p)s}$$

Rearranging this equation into the form of a PI controller (Equation 7.3.2 with ${}_D$=0) yields

$$G_c(s) \quad \frac{{}_p}{K_p(\ {}_f \qquad {}_p)} \quad 1 \quad \frac{1}{{}_p s}$$

Therefore, equating coefficients between this equation and Equation 7.3.2 yields the PI tuning parameters in terms of the FOPDT model parameters and the IMC filter time constant, i.e.,

$$K_c \quad \frac{_p}{K_p(\ _f \quad _p)} \qquad _I \quad _p$$

Thus, an IMC controller design method can be used to set PID tuning parameters based on process model parameters and the desired closed-loop response. IMC design methods can be used to derive the corresponding PI/PID controller settings for a variety of assumed approximate model forms. Table 8.4 provides PI/PID settings based on several commonly used model forms based on IMC design methods. Note that as $_f$ is increased, the aggressiveness of the controller is reduced. Details of this approach are provided elsewhere[6].

Table 8.4 PI/PID Tuning Setting for Various Model Types based on an IMC Design

$G_p(s)$	K_cK_p	$_I$	$_D$
$\frac{K_p}{_p s \quad 1}$	$\frac{_p}{_f}$	$_p$	0
$\frac{K_p}{(\ _1 s \quad 1)(\ _2 s \quad 1)}$	$\frac{_1 \quad _2}{_f}$	$_1 \quad _2$	$\frac{_1\ _2}{_1 \quad _2}$
$\frac{K_p}{^2_n s^2 \quad 2 \quad _n s \quad 1}$	$\frac{2 \quad _n}{_f}$	$2 \quad _n$	$\frac{_n}{2}$
$\frac{K_p(1 \quad s)}{^2_n s^2 \quad 2 \quad _n s \quad 1};$	$\frac{2 \quad _n}{_f}$	$2 \quad _n$	$\frac{_n}{2}$
$\frac{K_p}{s(\ _p s \quad 1)}$	$\frac{2 \quad _f \quad _p}{^2_f}$	$2 \quad _f \quad _p$	$\frac{2 \quad _f \quad _p}{2 \quad _f \quad _p}$
$\frac{K_p e^{\ _p s}}{_p s \quad 1}$ *	$\frac{_p \quad _p/2}{_f \quad _p/2}$	$_p \quad _p/2$	$\frac{_p\ _p}{2 \quad _p \quad _p}$

* Based on the first-order Padé approximation

Overview. While IMC provides a number of fundamental insights and has several convenient features, it assumes that a model of the process is available. Unfortunately, accurate process models are rarely available in an industrial setting. As a result, IMC is rarely used for tuning most industrial control loops because the time required to identify a process model greatly exceeds the time to implement other PID tuning approaches which are presented in Sections 9.8-9.11. Tuning a PID controller using the Cohen and Coon method, the Ciancone and

Marlin method or IMC design approach each requires a process model. An open-loop identification method (e.g., FOPDT identification presented in Section 6.8) is usually required to develop a process model for an industrial process. On the other hand, closed-loop methods, which are presented later in Sections 9.7 and 9.8, require much less time to implement.

Example 8.14 Tuning a pH Control Loop for a Fed-Batch Bio-Process

Problem Statement. Consider the pH control loop for a fermentor shown in Figure 8.8.4. This process is represented by the following transfer functions

$$G_a(s) = 10\ \frac{\text{ml / min}}{\text{mA}}$$

$$G_p(s) = \frac{0.5\ \dfrac{\text{pH units}}{\text{ml / min}}}{(5s+1)(50s+1)}$$

$$G_s(s) = 1$$

where the time constants have units in minutes. The desired closed-loop time constant is 2 min. Determine the PID controller settings based on (a) IMC tuning, (b) the pole placement method and (c) Cohen and Coon method or the Ciancone and Marlin method.

Solution. The transfer function for the actuator/process/sensor is given by

$$\underline{G}_p(s) = \frac{5}{(5s+1)(50s+1)}$$

(a) From Table 8.4 using the settings for the second model from the top, K_c=0.2(55)/τ_f=11/τ_f ml/min-pH units; τ_I=5+55=55 min; and τ_D=(5)(50)/(5+50)=4.55 min.

(b) Applying Equation 8.8.6,

$$G_c(s) = \frac{S_{cl}(s)}{G_p(s)[1 - S_{cl}(s)]} = \frac{\dfrac{1}{2s+1}}{\dfrac{5}{(5s+1)(50s+1)}\left[1 - \dfrac{1}{2s+1}\right]} = \frac{(5s+1)(50s+1)}{10s} = 5.5\left(1 + \frac{1}{55s} + 4.55s\right)$$

Therefore, for pole placement for this case, $K_c = 5.5$; $\tau_I = 55$; and $\tau_D = 4.55$, which corresponds exactly to the IMC tuning with τ_f=2.

(c) It is not possible to apply the Cohen and Coon method to this problem because the process model does not have any deadtime in the process model. Note that the controller gain for the Cohen and Coon method contains the process deadtime in the denominator. The Ciancone and Marlin method is also based on a FOPDT model, but

Figure 8.6.2 can be used for zero deadtime. The process model in this case is second order, but the larger time constant (50 min) is an order of magnitude larger than the smaller one; therefore, it is reasonable to assume that

$$G_p(s) = \frac{5}{50s + 1}$$

Using Figure 8.6.2 for disturbance rejection, $K_c K_p = 2$; $\tau_I/(\tau_p + \theta_p) = 0.25$; and $\tau_D/(\tau_p + \theta_p) = 0$. Therefore, $K_c = 0.4$; $\tau_I = 12.5\,\text{min}$; and $\tau_D = 0$. Based on the guidelines for choosing between PI and PID controllers presented in Section 7.9, this process should not use derivative action because the deadtime to time constant ratio is zero. This result is consistent with the tuning from the Ciancone and Marlin method, but is not consistent with the results from IMC tuning and the pole placement approach.

Example 8.15 Temperature Controller Tuning Using IMC Design for a Bio-Reactor

Problem Statement. Certain therapeutic proteins are produced industrially in recombinant *E Coli*. The cells are first grown at optimal growth conditions at a broth temperature of 35ºC. The metabolic heat generated by the *E Coli* during the growth phase is removed by cooling water that is circulated through the bio-reactor cooling jacket (Figure 8.8.5). When the cell concentration for the *E. Coli* reaches the desired concentration, the setpoint for the temperature controller is changed from 35ºC to 25ºC, which initiates the formation of the therapeutic protein by the *E. Coli*. During the production formation mode, the glucose feed and other nutrients are converted into the desired protein product. It is important to maintain accurate temperature control during both the cell growth phase and the product production phase.

The transfer function for the effect of cooling water flow rate (F_{cw}) on the broth temperature (T) during the growth stage is given as

$$\underline{G}_p(s) = \frac{T(s)}{F_{cw}(s)} = \frac{(0.1\ {}^\circ\text{C} \cdot \text{h/kg})\, e^{-5s}}{20s + 1}$$

where time is in h.

1. Design an IMC controller (i.e. select the tuning parameters) for controlling the temperature during the growth phase.

2. Assuming that the controller is not retuned, evaluate the performance of the controller during the product production phase if the process transfer function during the production phase is given as

$$\underline{G}_p(s) = \frac{T(s)}{F_{cw}(s)} = \frac{(0.01\ {}^\circ\text{C} \cdot \text{h/kg})\, e^{-5s}}{20s + 1}$$

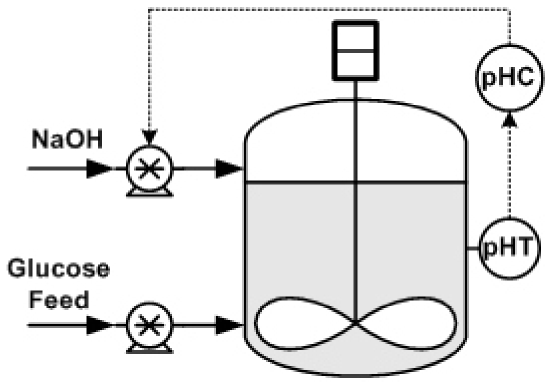


Figure 8.8.4 Schematic of a pH controller applied to a fed-batch reactor.

Solution. For $K_p = 0.1$; $\tau_p = 20$; $\theta_p = 5$, using Table 8.4, $K_c = 22.5/(\tau_f + 2.5)$; $\tau_I = 22.5$; $\tau_D = 1.11$. Applying the characteristic equation and using the Padé approximation for the deadtime yields

$$K_c\left(1+\frac{1}{\tau_I s}+\tau_D s\right)\frac{K_p}{\tau_p s+1}\,\frac{1-\theta_p s/2}{1+\theta_p s/2}+1=\frac{22.5}{\tau_f+2.5}\,\frac{22.5s+1+1.11s^2}{22.5s}\,\frac{0.1}{20s+1}\,\frac{1-2.5s}{1+2.5s}+1=0$$

Rearranging yields $([\tau_f+2.5]1125-0.624)s^3+([\tau_f+2.5]506-12.4)s^2+([\tau_f+2.5]22.5+4.5)s+0.225=0$

First, a real root is found by numerically solving this equation. Then, this root is factored out by long division, leaving a quadratic equation. The natural period and the damping factor are determined from the quadratic equation in s. For τ_f equal to 1, the real root was -0.393 and the remaining second-order equation has a damping factor equal to 4.45. Even when τ_f is set equal to zero, the real root is -0.391 and the damping factor drops only to 1.92. Obviously, the gain of the controller can be increased further to reach critically damped behavior or even underdamped behavior, but this cannot be attained using the IMC design approach for this case.

Assuming that τ_f is maintained equal to zero, the characteristic equation when the process is operated at 25ºC for the production mode becomes

$$(2.5\cdot 1125-0.0624)s^3+(2.5\cdot 506-1.24)s^2+(2.5\cdot 22.5+0.45)s+0.0225=0$$

$$2812.4s^3+1263.8s^2+56.7s+0.0225=0$$

Once again a root is found numerically equal to -0.399 and then the damping factor for the remaining quadratic term is equal to 5.81. Due to the ten-fold decrease in the process gain when the operating temperature is reduced from 35ºC to 25ºC without changing the controller tuning, the closed-loop response becomes highly overdamped.

Lambda Tuning. Lambda tuning is a special form of IMC tuning that is designed to provide an overdamped closed-loop response for a PI controller. Similar to IMC tuning, Lambda tuning requires a model of the process, which is normally a FOPDT model obtained from a single step test. Because Lambda tuning uses an overdamped tuning criterion, an accurate model is not required. Therefore, Lambda tuning is normally applied to less-important control loops, such as, most level and pressure loops.

The controller settings for a PI controller[8] are given by

$$K_c=\frac{\tau_p}{K_p(\lambda+\theta_p)}\qquad\qquad \tau_I=\tau_p$$

where K_p, τ_p and θ_p are the FOPDT model parameters and λ is the tuning factor. Thus, λ is the only tuning parameter and it normally has a value $\tau_p<\lambda<3\tau_p$. Moreover, VanDoven[8] indicates that λ is related to the closed-loop time constant.

Self-Assessment Questions

Q8.8.1 How were the PID tuning settings shown in Table 8.4 developed?

Q8.8.2 What is the "invertible" part of a transfer function in the context of IMC?

Q8.8.3 What is the "filter" in an IMC design?

Self-Assessment Answers

Q8.8.1 Equation 8.8.7 is applied to produce $G_c^*(s)$ using the realizable portion of the process model. Then, Equation 8.8.6 is used to derive the feedback controller, $G_c(s)$. The $G_c(s)$ is rearranged into the form of a PID controller yielding the relationships between the PID setting and the model equations.

Q8.8.2 The non-invertible parts of a transfer function correspond to the parts of a transfer function, which when inverted result in unstable or unrealizable controllers (e.g., deadtime and right-half plane zeros). The remaining portion is invertible.

Q8.8.3 The filter in the IMC design is used to reduce the controller aggressiveness to compensate for the non-invertible portion of the process model and modeling error.

8.9 Summary

- For an open-loop overdamped process, as the controller aggressiveness is increased (K_c increased or $_I$ decreased) the closed-loop dynamics go from overdamped to critically damped to oscillatory to ringing to sustained oscillations to unstable oscillations.
- Root locus diagrams are a concise means to represent the range of closed-loop dynamics as a function of a tuning parameter.
- A controller that does not have a large enough controller gain will exhibit a closed-loop damping factor that is larger than the desired response.
- A controller that does not have enough integral action will exhibit slow offset elimination.
- By comparing the lag between the measured value of the CV and the controller output, you can determine whether a controller is ringing from too much proportional action or too much integral action.
- Too much derivative action results in a closed-loop response that has a stair-step shape.

8.10 References

1. Stephanopoulos, G., *Chemical Process Control*, Prentice-Hall, Inc., p. 215 (1984).

2. Cohen, G.H. and G.A. Coon, “Theoretical Considerations of Retarded Control”, *Trans ASME*, Vol. 75, p. 827 (1953).

3. Ziegler, J.G. and N.B. Nichols, “Optimum Settings for Automatic Controllers”, *Trans ASME*, Vol. 64, p. 759 (1942).

4. Marlin, T.E., *Process Control*, McGraw-Hill, p. 300 (1995).

5. Cianione, R. and T.E. Marlin, “Tune Controllers to Meet Plant Objectives”, *Control*, Vol. 5, p. 50 (1992).

6. Garcia, C.E. and M. Morari, "Internal Model Control 1. A Unifying Review and Some New Results, *Ind. Eng. Chem. Process Des. Dev*., Vol. 21, pp. 308-320 (1982).

7. Rivera, D.E., M. Morari and S. Skogestad, "Internal Model Control 4. PID Controller Design", Ind. Eng. Chem. Process Des. Dev., Vol. 25, pp. 252-261 (1986).

8. VanDoren, V., "Fundamentals of Lambda Tuning", Control, April 13, 2013.

8.11 Additional Terminology

Cohen and Coon tuning - control settings for P-only, PI and PID controllers based on a minimum IAE tuning criterion using FOPDT process models.

Direct synthesis - a method for specifying a controller based on closed-loop performance specifications.

Fractional deadtime - $_p$ / ($_p$ $_p$).

IAE - integral absolute error.

ISE - integral squared error.

ITAE - integral time absolute error.

ITSE - integral time squared error.

In-phase - y_s is in-phase with c when the peaks or valleys for y_s and c both occur at the same point in time.

Internal model control (IMC) - a model based control approach that can be used to set PID tuning parameters if a model of the process is available.

K_u - the ultimate controller gain of a loop, i.e., the controller gain for a P-only controller that causes sustained oscillations.

P_u - the period of sustained oscillations using a P-only control.

Pole placement - a method for specifying a controller based on closed-loop performance specifications.

Quarter amplitude damping (QAD) - a decay ratio of ¼.

Ringing - an industrial term referring to a response that exhibits slow damping of the oscillations.

Root locus diagram - a plot of the real and imaginary components of the poles of a closed-loop transfer function as a function of the controller tuning parameters.

Ultimate controller gain - the gain of a P-only controller that corresponds to sustained oscillations.

Ultimate period - the period of sustained oscillations for a P-only controller.

Ziegler-Nichols (ZN) tuning - controller settings for P-only, PI or PID controllers using K_u and P_u which is based upon a QAD tuning criterion.

8.12 Preliminary Questions

8.2 Effect of Tuning Parameters for P-only Control

8.2.1 For a process that is open-loop overdamped, how do the poles of the closed-loop response change as the controller gain is increased for a P-only controller?

8.2.2 Where is a critically damped response located on a root locus diagram?

8.2.3 For a plot of the closed-loop damping factor versus the controller gain, what point corresponds to sustained oscillations?

8.2.4 Why does a first-order process model without deadtime not adequately represent the closed-loop behavior of a process with an aggressively tuned controller?

8.3 Effect of Tuning Parameters for PI Control

8.3.1 Consider the application of P-only and PI control to a FOPDT process. How are they alike and how are they different?

8.3.2 For a PI controller, how does the reset time affect the aggressiveness of the controller?

8.4 Effect of Tuning Parameters for PID Control

8.4.1 Compare the response of a PI and a PID controller applied to a sluggish responding process.

8.4.2 Explain why too much derivative action causes stair step behavior (e.g., Figure 8.4.2).

8.4.3 What is the connection between Figures 8.4.2 and 8.4.3?

8.6 Classical Tuning Methods

Q8.8.1 What knowledge about the process is necessary to apply the Cohen and Coon tuning method?

Q8.6.2 What knowledge about the process is necessary to apply the Ziegler-Nichols tuning method?

Q8.6.3 How are the dynamics of the sensor and actuator related to the process models used in the Cohen and Coon method?

Q8.6.4 What information about the process is required to use the Ciancone and Marlin tuning method?

Q8.7.5 Why are the classical tuning methods presented in Section 9.3 not recommended for tuning industrial control loops?

8.7 Controller Tuning by Pole Placement^{AT}

8.7.1 How is pole placement used to tune a PID controller**?**

8.7.2 How can you specify a general feedback controller using pole placement, which is not necessarily a PID type controller, to attain a prescribed feedback behavior.

8.8 PID Tuning Based on Internal Model Control (IMC)^{AT}

Q8.8.1 What is the relationship between the controller in Figure 8.8.2 [i.e., $G_c^*(s)$] and the controller in Figure 8.8.1 [i.e., $G_c(s)$]?

Q8.8.2 For an IMC controller, what is the difference between the ideal case and the non-ideal case?

8.13 Analytical Questions and Exercises

8.2 Effect of Tuning Parameters for P-only Control

8.2.1** Consider a first-order process with a P-only controller. Assuming that the controller gain is increased without limit, will the following systems combined with the first-order process result in unstable behavior at some value of K_c?

- **a.** Assuming that $G_a(s)$=$G_s(s)$=1.
- **b.** Assuming that $G_a(s)$=1 and that $G_s(s)$ is a first-order system.
- **c.** Assuming that $G_a(s)$ and $G_s(s)$ are each first-order systems.
- **d.** Assuming that $G_a(s)$=1 and that $G_s(s)$ has analyzer delay.

8.2.2** Develop a root locus plot for a P-only controller applied to a second-order process (K_p=10, $_n$=6, =10). Can this process be made to become unstable by increasing the controller gain?

8.2.3*** Develop a root locus plot by varying K_c for a P-only controller applied to the thermal mixing process from Example 3.3 (Figure P8.2.3). The model for this process is given by

Actuator
$$\frac{dF_1}{dt} \quad \frac{1}{_v}(F_{1,spec} \quad F_1)$$

Process $$M\frac{dT}{dt} \quad F_1T_1 \quad F_2T_2 \quad (F_1 \quad F_2)\,T$$

Sensor $$\frac{dT_s}{dt} \quad \frac{1}{_{Ts}}(T \quad T_s)$$

where

- F_1 - mass flow rate of stream 1 (initially 5 kg/s)
- F_2 - mass flow rate of stream 2 (5 kg/s)
- M - mass of liquid in the mixer (100 kg)
- T - temperature of the mixed liquid (initially 50°C)
- T_1 - temperature of stream 1 (25 C)
- T_2 - temperature of stream 2 (75 C
- t - time (s)
- $_v$ - the time constant for the flow controller on stream 1 (2 s).
- $_{Ts}$ - the time constant for the temperature sensor on the product stream (6 s).

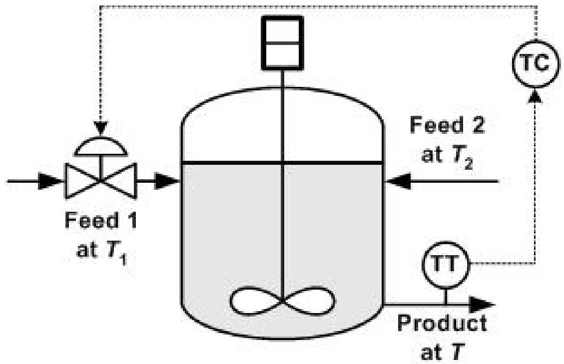


Figure P8.2.3 Control diagram of a CST thermal mixer with temperature controller.

8.3 Effect of Tuning Parameters for PI Control

8.3.1** Develop a root locus plot for a PI controller (K_c=1) applied to a first-order process (K_p=2, $_p$=3) for which the reset time is varied. Why does this root locus diagram not represent a real process?

8.3.2*** Develop a root locus plot by varying K_c for a PI controller with $_I$=30 s applied to the DO process from Example 3.9 (Figure P8.3.2). The model for this process is given by

Actuator: $$F_{air} \quad (F_{air})_{spec}$$

Process: $$k_La \quad 0.25 \quad 0.001(F_{air} \quad 500)$$

$$\frac{dC_{O_2}}{dt} \quad k_La\; C^*_{O_2} \quad C_{O_2} \qquad K_{O_2} \quad {}_{\max} x$$

Sensor: $$\frac{dC_{O_2,s}}{dt} \quad \frac{1}{_s} \quad C_{O_2} \quad C_{O_2,s}$$

where

- C_{O_2} - concentration of O_2 in the reaction broth (initially 1.1×10^{-4} g-moles/l)
- $C^*_{O_2}$ - saturated concentration of O_2 in the broth (2.20×10^{-4} g-moles/l)
- $C_{O_2,s}$ - the measurement of the O_2 concentration in the broth (initially 1.1×10^{-4} g-moles/l)

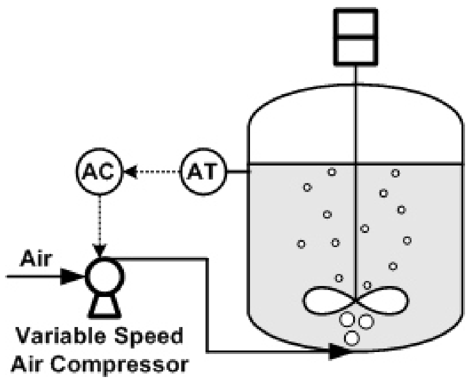


Figure P8.3.2 Control diagram of a bio-reactor with a dissolved oxygen controller.

- F_{air} - the volumetric flow rate of air to the bio-reactor (initially 500 cfm)
- K_{O_2} - cellular uptake of O_2 (1.98 g-moles O_2/g-cells)
- k_La - the overall liquid phase mass transfer coefficient for transport from the bubble surface to the bulk broth (initially 0.25 s^{-1})
- T - broth temperature (35ºC)
- t- time (s)
- V - the volume of broth in the bio-reactor (1000 l)
- x - cell concentration in the bio-reactor (0.25 g/l)
- $_{max}$ - maximum specific growth rate (5.56×10^{-5} s^{-1})
- $_s$ - the time constant of the DO sensor (10 s)

8.3.3^S** The following PI controller settings applied to the heat exchanger simulation for the simulation software that comes with this text results in a closed-loop performance that is ringing (K_c=15; $_I$=10). Using the approach shown in Section 8.3, determine whether the ringing is the result of too much proportional or too much integral action. Develop a plot of y_s and c versus time to perform your analysis. Then check your conclusion by reducing either the gain or increasing the reset time and perform a simulation of a setpoint change and compare to the response of the original controller settings.

8.3.4^S** The following PI controller settings applied to the thermal mixer simulation for the simulation software that comes with this text results in a closed-loop performance that is ringing (K_c=0.04; $_I$=4). Using the approach shown in Section 8.3, determine whether the ringing is the result of too much proportional or too much integral action. Develop a plot of y_s and c versus time to substantiate your conclusion. Then check your conclusion by reducing either the gain or increasing the reset time and perform a simulation of a setpoint change and compare to the response of the original controller settings.

8.4 Effect of Tuning Parameters for PID Control

8.4.1** Develop a root locus plot for a PID controller (K_c=1, $_I$=3) applied to a first-order process (K_p=2, $_p$=4) for which the derivative time is varied.

8.6 Classical Tuning Methods

P8.6.1* Determine the PID controller setting for the Cohen and Coon and Ciancone and Marlin methods using the following FOPDT model (i.e., K_p=10, $_p$=3, $_p$ 1). Use the Ciancone and Marlin settings for disturbance rejection.

P8.6.2* Using the Cohen and Coon method, develop initial controller settings for a PI controller applied to a temperature control loop on a water-cooled CSTR that operates under regulatory control. By observing the process operation, it has been estimated that the process gain is approximately -0.03 ºF-hr/lb. The response time of the process has been estimated to be approximately 6 minutes. Further, by observing the process, it has been determined that the amount of deadtime in the system is approximately 1 minute.

P8.6.3^S** Using the visual basic simulators that come with this textbook, apply open-loop tests for positive and negative 10% changes in the MV for the CSTR simulator. From these open-loop tests, develop FOPDT models for the positive and negative MV changes. Average the FOPDT model parameters from these two tests and with the average FOPDT parameters determine PI tuning settings using the Cohen and Coon and Ciancone and Marlin methods. Apply these tuning settings to the simulator and determine the decay ratio for positive and negative 1.428% setpoint changes from the base case conditions.

P8.6.4^S** Using the simulators that come with this textbook, apply open-loop tests for positive and negative 10% changes in the MV for the thermal mixing tank simulator. From these open-loop tests, develop FOPDT models for the positive and negative MV changes. Average the FOPDT model parameters from these two tests and with the average FOPDT parameters determine PI tuning settings using the Cohen and Coon and Ciancone and Marlin methods. Apply these tuning settings to the simulator and determine the decay ratio for positive and negative 10% setpoint changes from the base case conditions.

8.7 Controller Tuning by Pole Placement^{AT}

P8.7.1** Determine the PI controller settings for a first-order process (K_p = 6; $_p$ = 5) using pole placement if it is desired to obtain a closed-loop damping factor of 0.3 and a closed-loop natural period of 3.

P8.7.2** Determine the PI controller settings for a first-order process (K_p = 13; $_p$ = 15) using pole placement if it is desired to obtain a closed-loop damping factor of 0.7 and a closed-loop natural period of 3.

P8.7.3*** For the following process transfer function design a pole-placement controller so that the closed-loop pole is located at -5/4 and the closed-loop response should be offset free.

$$G_p(s) \quad \frac{5(\; s \quad 1)}{(10s \quad 1)(1s \quad 1)}$$

Discuss the effect of for 1 10 on the performance of the pole-placement controller.

P8.7.4** Consider the model for a non-self-regulating level in a tank (Example 3.5).

$$A_c \frac{dL}{dt} \quad F_{in} \quad F_{out}$$

where L is the tank level, is the fluid density, A_c is the cross-sectional area of the tank, F_{in} is the inlet mass flow rate and F_{out} is the outlet mass flow rate. Design a pole-placement controller so that the closed-loop system has not offset, and the closed-loop time constant is 25 s considering the outlet flow rate as the MV.

8.8 PID Tuning Based on Internal Model Control (IMC)^{AT}

P8.8.1* For the second-order process with a right-half plane zero in Table 8.4 (i.e., the fourth model from the top), determine the PID tuning settings based on IMC tuning if the process model is given by K_p 1; 0.5; $_n$ 1; and 2. Can you use any other tuning method for this model?

P8.8.2*** A process transfer function for the response of the bottom temperature, T_B, of a distillation column to changes in the reboiler duty (Q) is given by

$$G_p(s) \quad \frac{T_B(s)}{Q(s)} \quad \frac{10(^{\circ}\mathrm{C} \quad \mathrm{min/kg})\,(1 \quad 10s)}{(10s \quad 1)(s \quad 1)}$$

The transfer function for the effect on feed composition disturbances (x_F) on the bottom temperature is given as

$$G_d(s) \quad \frac{T_B(s)}{X_F(s)} \quad \frac{5\;(^{\circ}\mathrm{C/\,mole\;\;fraction})}{(50s \quad 1)(10s \quad 1)}$$

Find an appropriate IMC controller based on $G_p(s)$ and evaluate the effect of disturbances in x_F on the regulatory behavior of the closed-loop system as a function of $_f$.

Chapter 9

PID Tuning: Industrially Relevant Approaches

Chapter Objectives

- Appreciate the effect of controller reliability on controller tuning.
- Understand how to select the proper tuning criterion for a control loop.
- Understand how to choose the proper filter factor for a sensor.
- Present an industrially relevant method for tuning fast- and slow-responding PI control loops.
- Present an approach for tuning a PID controller.
- Illustrate a direct means to set the tuning parameters for slow-responding level control loops.
- Perform controller tuning exercises.

9.1 Introduction

Tuning PID loops is one of the major responsibilities of a process control engineer and the resulting controller settings have a dominant effect on the performance of a PID control loop. Tuning a PID controller requires selecting values for K_c, τ_I and τ_D that meet the operational objectives of the control loop, which usually requires making the proper compromise between performance (minimizing deviations from setpoint) and reliability (the controller's ability to remain in service while handling major disturbances). At other times, controller tuning is determined by the overall objectives of the process.

The classical methods presented in Chapter 8 will work for many processes, but the time required to implement them will, in general, be significantly more than the methods presented later in this chapter. That is, certain classical methods (e.g., Cohen and Coon, Ciancone and Marlin and IMC) require a model of the process (e.g., a FOPDT model) and the time required to develop the model is usually significant. Many times in industry time is not available to develop a process model and then tune the controller, especially when less time-consuming tuning approaches can be used. For example, for batch processes, which are about half of all processes in industry, the processes are typically noncontinuous, non steady-state, and often nonlinear. Sometimes the process characteristics and loads change significantly as a batch proceeds. Other times, different products (with different properties) are made from time to time using the same equipment. In addition, for new plants or major retrofits to existing plants, resources are usually stretched thin, with tight project time-lines in force, and so not much time is available to study, test, and tune each loop when a plant consisting of thousands of loops is being commissioned. So, it is common to, as a first pass, configure and tune control loops as quickly, practically, and safely as possible using simple procedures, rules of thumb, equipment manufacturer's recommendations, and experience with similar processes. Additional time and effort can be pursued later using more traditional testing and analysis if such loops prove to be troublesome. Commercial auto-tuning products are also available to assist with loop tuning.

9.2 Summary of Theoretical Results

Following is a listing of some of the key results relating to controller tuning and controller mode selection that were developed from theoretical considerations in Chapters 7 and 8:

- Proportional action (K_c) increases the speed of the closed-loop response, but exhibits offset from setpoint (Section 7.7)
- Integral action ($K_c/\ _I$) removes offset, but increases the oscillatory nature of the closed-loop response (Section 7.7).
- Derivative action ($K_c\ _D$) decreases the oscillatory nature of the closed-loop response for a sluggish process (Section 7.7)
- As proportional action or integral action is increased for a typical process, the closed-loop response will go through the following range of dynamic responses: overdamped, critically damped, oscillatory, ringing and instability in that order (Sections 8.3 and 8.4).
- When too much proportional action for a control loop is causing ringing, the CV and the MV are in-phase (Section 8.4).
- When too much integral action for a control loop is causing ringing, the CV and the MV are out of phase (Section 8.4).
- Use P-only or PI control for a non-sluggish process (Section 7.8). Use a PI controller if offset-free operation is desired. An integrating process can be offset-free for P-only control (Example 7.4).
- Use PID control for a sluggish process (i.e., $_p/\ _p$>1). Too much derivative action will result in a stair step closed-loop response (Section 8.5).

9.3 Tuning Criteria and Performance Assessment

Tuning Criteria. Following are some common objectives for controller tuning.

- Minimize deviations from setpoint
- Attain good setpoint tracking performance
- Avoid excessive variation of the MVs
- Maintain process stability for major disturbance upsets
- Eliminate offset

It is not possible to simultaneously satisfy each of these objectives; therefore, tuning is usually a compromise among these objectives. For example, tuning for minimum deviation from setpoint for normal disturbances is

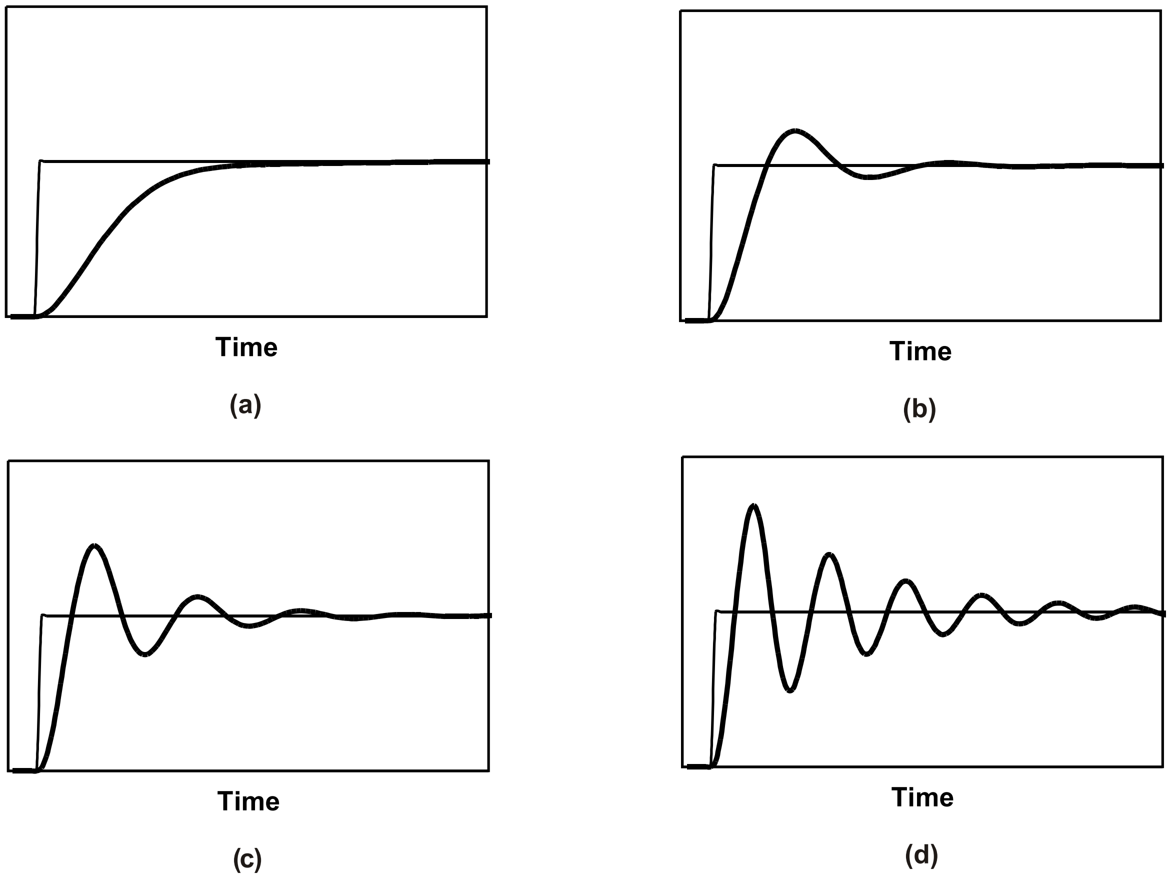


Figure 9.3.1 Control response for a setpoint change. (a) critically damped response. (b) 1/10 decay ratio. (c) QAD. (d) Ringing response.

contrary to tuning the controller to remain stable for major disturbances. If the controller is tuned for normal disturbances, the closed-loop system may go unstable when a major disturbance enters the process. On the other hand, if the controller is tuned for the largest possible disturbance, control performance is likely to be excessively sluggish for normal operation.

Performance Assessment. Figure 9.3.1 shows the setpoint tracking response of an endothermic CSTR (Example 3.4) tuned for a critically damped response (Figure 9.2.1a), a decay ratio of 1/10 (Figure 9.3.1b), a decay ratio of ¼ (**quarter-amplitude damping, QAD**, Figure 9.3.1c) and a ringing response (e.g., a decay ratio of 1/1.5; Figure 9.3.1d). The decay ratio is a convenient measurement of the closed-loop dynamic response because it can be easily estimated from setpoint changes and some disturbance upsets. The 1/1.5 decay ratio results in excessive cycling, which in industry is called **ringing**. The aggressiveness of the controller is increased from Figure 9.3.1a to Figure 9.3.1d. Table 9.1 lists the IAE, ITAE, ISE and ITSE for a range of decay ratios for this process. Note that QAD tuning results in the best overall performance without regard to **reliability** (i.e., the ability to maintain stable operation in the face of significant disturbances). Also, each of these statistics goes through a minimum as the decay ratio is increased. Note that from Table 9.1, the optimum ITSE level occurred at a decay ratio of 1/5 because of the emphasis on faster damping. It should be emphasized that these statistics are

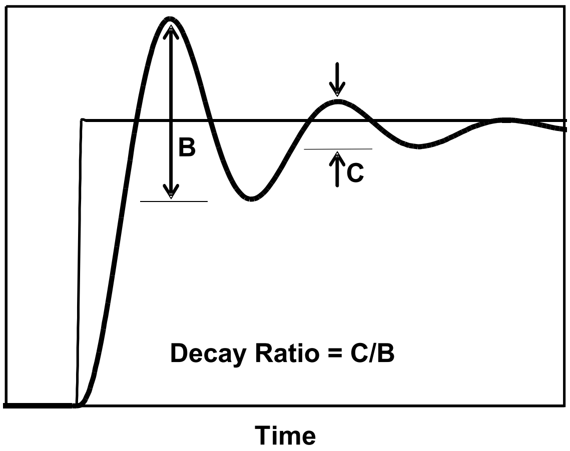


Figure 9.3.2 A method for estimating the decay ratio using peak-to-valley measurements.

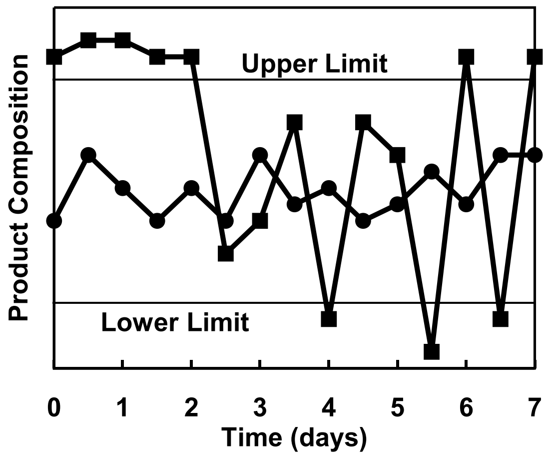


Figure 9.3.3 SPC chart for two different controllers on the same process during different time periods.

used in academic research to compare the control performance for different controllers using dynamic process simulations, but are not usually convenient to use in industry.

The decay ratio is an excellent measure of the aggressiveness of a controller with an underdamped response. The classical definition of the decay ratio (Figure 6.4.4), which is based on peak heights, can be difficult to apply in certain cases because the oscillations may not be symmetric about the setpoint. The decay ratio can also be estimated using the difference between a peak and a valley (Figure 9.3.2). That is, the decay ratio can be calculated by the ratio of the peak-to-valley difference for two adjacent cycles, i.e., C/B in Figure 9.3.2. As a result, symmetric oscillations about the setpoint are not required to estimate the decay ratio from the closed-loop response to a setpoint change or during the return to setpoint for a disturbance upset.

Industrial control performance is often assessed by the variability in the final products. The deviation from setpoint can be used as a measure of product variability and thus as a measure of control performance. The variation from setpoint (s) is defined as

$$s = \sqrt{\frac{\sum_{i=1}^{N} [y_s(t_i) - y_{sp}]^2}{N}}$$

where $y_s(t_i)$ is the sampled CV value at time equal to t_i and N is the number of samples. s is an estimate of the standard deviation of the error from setpoint. The smaller the deviation from setpoint, the better the control performance. Remember that this statistic is based on the error from setpoint while the standard deviation is based on the error from the average value of a set of data.

Most companies keep statistical process control (SPC) charts that track the laboratory analysis of final products, which are typically sampled one to three times daily. Figure 9.3.3 is an example of an industrial SPC chart. Note the upper and lower limits on the product composition. This chart shows control results for two different controllers for two different seven day periods. It is easy to see which controller performed better.

Table 9.1
Several Performance Statistics as a Function of Decay Ratio for the Endothermic CSTR (Example 3.5)

Decay Ratio	IAE	ITAE	ISE	ITSE
1/1.5	39.6	1244	31.1	470
1/2	28.3	628	22.8	231
1/3	20.9	347	17.8	117
1/4	19.8	387	16.8	92.8
1/5	20.7	503	16.8	91.2
1/6	22.0	635	17.1	97.4
1/8	24.9	903	17.9	119
1/10	27.4	1141	18.8	145

Self-Assessment Questions

Q9.3.1 List several commonly used tuning objectives and explain why they cannot be simultaneously met.

Q9.3.2 Why is using a decay ratio a convenient tuning criterion for tuning PID control loops?

Self-Assessment Answers

Q9.3.1 The following tuning criterion can be considered:

- Deviations from setpoint should be minimized.
- Good setpoint tracking performance should be attained.
- Excessive variation of the MV levels should be avoided.
- The controlled process should remain stable for major disturbance upsets.
- Offset elimination may or may not be important.

Note that minimizing deviations from setpoint and maintaining stable operation during severe upsets cannot generally be satisfied simultaneously; therefore, a compromise between these two objectives has to be made. In addition, minimizing deviations from setpoint can be contrary to avoiding excessive variations in the MVs.

Q9.3.2 A decay ratio is a convenient tuning criterion because it can be estimated easily from setpoint changes or from a response to a sharp disturbance. It explicitly indicates the degree of aggressiveness that is to be implemented by a controller when an underdamped response is used.

9.4 Controller Reliability

Controller reliability is concerned with whether a controller remains stable during severe upsets. Consider the closed-loop transfer function for disturbance rejection (Equation 7.2.8):

$$\frac{Y(s)}{D(s)} \qquad \frac{G_d(s)}{G_c(s)\,G_a(s)\,G_p(s)\,G_s(s) \quad 1}$$

For a particular set of controller tuning parameters and fixed process models (i.e., $G_a(s)$, $G_p(s)$, and $G_s(s)$ remain unchanged), this equation indicates that the dynamic behavior of the process, i.e., the roots of the characteristic equation, are fixed and remain unchanged regardless of the size of the disturbance. Figure 9.4.1 shows the dynamic behavior of a linear process subjected to several different levels of a disturbance (i.e., $d_3 > d_2 > d_1$). Note that the larger disturbances produce larger deviations, but the dynamic behavior (decay ratio and response time) remains unchanged.

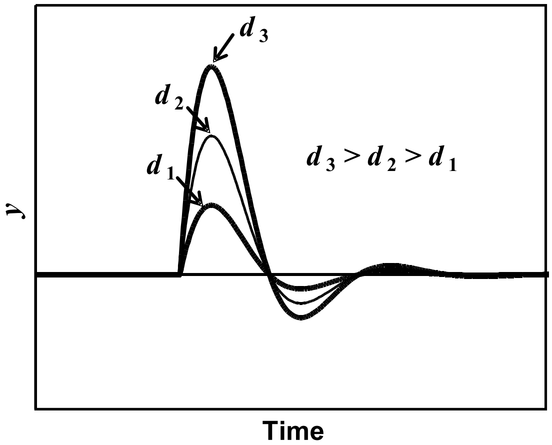


Figure 9.4.1 The effect of disturbance magnitude on the dynamic response of a linear process.

It is well known from industrial experience that certain control loops, which are stable under normal conditions, can become unstable when subjected to severe upsets, thus indicating that the dynamic characteristics of feedback processes are **not** always constant. Let's examine this apparent discrepancy.

Up until this point we have assumed that the dynamic characteristics of the process, sensor and actuator all remain constant as the operating conditions change. In fact, the characteristics of these systems can change with operating conditions. For example, all real processes exhibit some degree of nonlinearity. Figure 9.4.2 shows the process gain, K_p, as a function of the CV, y, for (a) a linear process, (b) a moderately nonlinear process and (c) a severely nonlinear process. For a very narrow region near y_0, even the severely nonlinear process behaves relatively linearly (i.e., exhibits a relatively constant process gain). For case (c), the farther removed from y_0, the larger the resulting gain change.

Feed flow rate changes can cause process time constant and deadtime changes, which represent another type of process nonlinearity. For the CST thermal mixer (Example 3.3), the process time constant, $_p$, is given by

$$_p \qquad \frac{V_r}{F_1 \quad F_2}$$

This equation shows that the process time constant varies inversely with the total feed rate to the process. Also, Section 6.7 shows that transport deadtime varies inversely with flow rate.

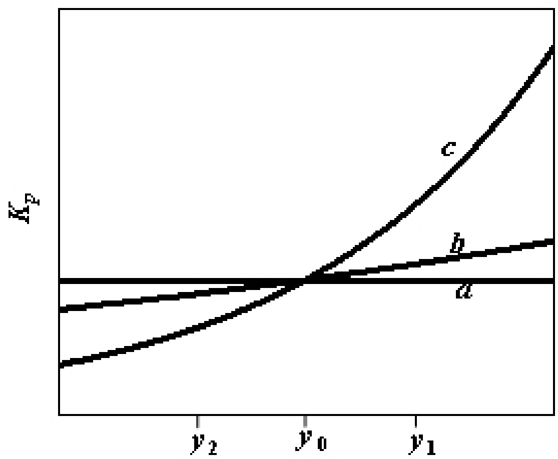


Figure 9.4.2 Process gain as a function of the CV for (a) a linear process, (b) a moderately nonlinear process and (c) a severely nonlinear process.

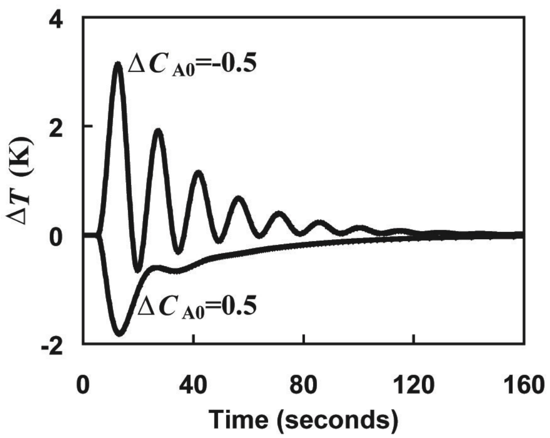


Figure 9.4.3 The effect of the disturbance direction on the dynamic behavior of the endothermic CSTR.

To demonstrate the effect of process nonlinearity combined with disturbances, consider the application of a PI controller to the endothermic CSTR presented in Example 3.6. This process exhibits changes in both process gain and time constant as the operating conditions change. A PI controller was tuned for a setpoint change using the QAD tuning criterion. Table 9.2 shows the results of this system for different magnitude upsets in the inlet feed composition, C_{A_o}.

The decay ratio changes sharply with an increase in the disturbance magnitude, which leads to ringing for a large reduction in the feed composition. For the largest feed composition upset, unstable closed-loop dynamics result. Therefore, for this nonlinear process, as the disturbance magnitude increases, the dynamic behavior of the process goes through the same sequence of phases that was observed for the controller tuning in previous sections. Remember from Figure 9.4.1, decay ratio remains constant irrespective of the disturbance size for a linear system.

Table 9.2
Effect of Disturbance Magnitude on Closed-Loop Dynamic Behavior

C_{A_0} $(gmole / L)$	Decay Ratio
-0.1	1/2.5
-0.2	1/2.35
-0.3	1/2.18
-0.4	1/1.94
-0.5	1/1.64
-0.6	1/1.34
-0.7	1/0.9 (unstable)

Figure 9.4.3 shows the closed-loop results for the endothermic CSTR for a positive and negative 0.5 g mol/l feed composition upset. The negative feed composition change results in ringing while the positive change results in sluggish behavior. The effect of process nonlinearity and disturbance magnitude on dynamic behavior can be understood by recognizing that a disturbance moves the CV from its normal operating range, resulting in process gain and time constant changes. These changes to the process parameters cause changes in the dynamic behavior of the closed-loop process. Consider the severely nonlinear process gain depicted in Figure 9.4.2. A disturbance

enters this process and results in an increase in y from y_0 to y_1, where the process gain is over twice the gain at y_0. Likewise, if a different disturbance moves the process to y_2, the process gain would be less than half the original process gain at y_0. Increasing the process gain by a factor of two can cause a properly tuned controller to exhibit ringing while a reduction in the process gain by a factor of two can cause a properly tuned controller to become sluggish. As a result, a nonlinear process can exhibit severe ringing or instability and, at other times, behave in a very sluggish manner depending upon the type, magnitude and direction of the disturbance. **An industrial control loop can be identified as highly nonlinear when the loop exhibits ringing and, at other times, exhibits sluggish behavior with the same controller settings.**

This analysis shows that **the combined effect of the nonlinearity of a process and the type and severity of disturbances directly affects the reliability of a controller**. If a controller proves unreliable, reducing the aggressiveness of the controller usually improves its reliability but at the expense of control performance. Many times, selecting the proper tuning parameters for a controller is a compromise between reliability and performance.

Self-Assessment Questions

Q9.4.1 How does a linear system under feedback control react to disturbances?

Q9.4.2 How does the magnitude and direction of a disturbance affect the dynamic response of a feedback controller?

Q9.4.3 If a feedback system with the same controller tuning exhibits ringing and at other times sluggish behavior, what does this indicate?

Self-Assessment Answers

Q9.4.1 A linear system reacts with the same dynamic behavior regardless of the size of the disturbance upset.

Q9.4.2 Process nonlinearity causes the dynamic characteristics of the closed-loop system to change with the magnitude and direction of a disturbance. For example, a disturbance increase can result in overdamped behavior while a decrease can cause ringing. Disturbances can have significant effects on the closed-loop behavior by altering the nature of the process, i.e., changing the process gain, time constant, and dead time, thereby changing the closed-loop response.

Q9.4.3 This would indicate that the process has significant nonlinearity.

9.5 Selection of Tuning Criterion

The first step in tuning a controller is selecting the tuning criterion (Section 9.3). You cannot arbitrarily select the desired dynamic response for a PID controller. For example, choosing the closed-loop damping factor limits the closed-loop time constant that is attainable. In fact, Figure 8.3.3 shows that as the controller gain of the P-only controller is increased, the closed-loop time constant and the damping factor decrease in a consistent manner. The same general behavior occurs for PI and PID controllers. Therefore, selecting the damping factor, to a large degree, determines the closed-loop time constant for a particular control loop. As a result, the aggressiveness of a controller can be specified in terms of the damping factor of the closed-loop system. Because the decay ratio, which is directly related to the damping factor (see Equation 6.4.4), is relatively easy to estimate from an underdamped response (see Figure 9.3.2), it is the preferred form for the tuning criterion.

Selecting the tuning criterion for a control loop is equivalent to determining the aggressiveness of the feedback controller. If you were to choose a critically damped response as the tuning criterion, a controller with a low aggressiveness would result. On the other hand, QAD tuning criterion corresponds to a very aggressively-tuned controller.

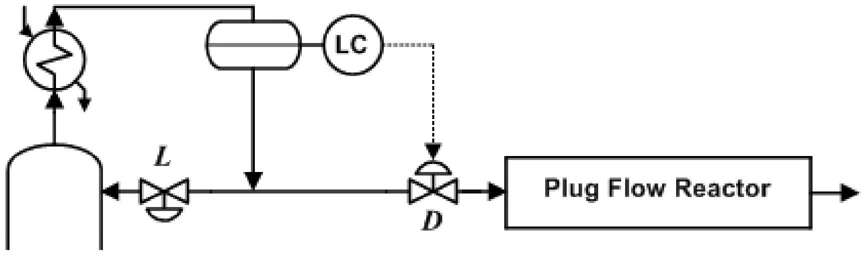


Figure 9.5.1 Schematic of an accumulator of a distillation column feeding a plug flow reactor.

The first factor that should be considered when choosing the tuning criterion for a control loop is how the performance of the control loop in question affects the overall objectives of the process. Consider an accumulator for a distillation column that feeds a plug flow reactor as shown in Figure 9.5.1. Tight level control for the accumulator would result in sharp changes in the feed rate to the reactor. Sharp and frequent changes of the feed rate to the reactor can significantly upset the operation of the reactor because this would result in changes in the residence time of the reactor, which would cause changes in the conversion in the reactor. On the other hand, if the level controller for the accumulator were tuned for sluggish performance, the change in the feed rate to the reactor would be much more gradual and smooth, allowing for better operation of the reactor. From an overall process point of view, maintaining smooth operation of the reactor is much more economically important than maintaining tight level control for the accumulator. Therefore, a critically damped or overdamped tuning criterion is the proper choice for the tuning criterion for the level controller in Figure 9.5.1.

As another example of how the overall process objective can affect the selection of the tuning criterion, consider the CSTR and separation train shown in Figure 9.5.2. If the level controller for the CSTR is tuned loosely, the CSTR level can change significantly. Because the production rates of the various products are related to the residence time in the reactor, large variations in the CSTR level directly affect the product distribution produced by the reactor. The resulting composition changes in the stream leaving the CSTR represent major upsets for the composition controllers for the distillation columns that comprise the separation train. Even though tight level control for the CSTR causes short term variations in the feed to the first column, these upsets are much easier to handle than composition changes; therefore, considering the overall process objectives, a tuning criterion corresponding to tight level control should be chosen for the level controller on the CSTR. From these two level control examples, it should be clear that it is important to consider how a control loop affects the overall process when selecting its tuning criterion.

The following procedure is recommended for selecting the tuning criterion for a control loop.

1. Evaluate the effect of tight and loose control for the control loop in question on the overall objectives of the process.

2. If based on the overall process objectives loose control is selected, the tuning criterion should be between highly overdamped to critically damped behavior. Consideration of the overall objectives of the plant will determine where in this range the tuning criterion should be selected.

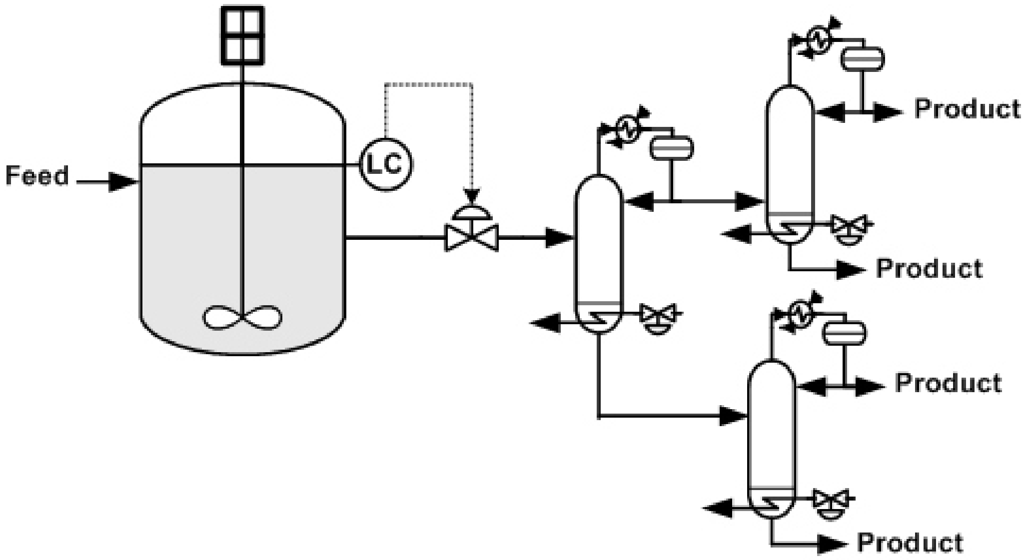


Figure 9.5.2 Schematic of a level in a CSTR that feeds a separation train.

3. If based on the overall process objectives tight control is selected, the tuning criterion should be between critically damped behavior to a 1/6 decay ratio. Determining the tuning criterion from this range will be based on a compromise between performance and reliability.

From the previous section, it was shown that process nonlinearity and disturbances determine the reliability of a controller. If a process is highly nonlinear and subject to large disturbances, controller reliability will likely be a problem, and a more conservative tuning criterion should be selected (e.g., a critically damped response). On the other hand, if the process is relatively linear and the disturbances are relatively mild, a more aggressive tuning criterion should be selected (e.g., a 1/6 decay ratio). Therefore, **when control engineers are faced with choosing a tuning criterion that is a compromise between performance and reliability, they must use their knowledge of the process to evaluate the relative nonlinearity of the process and the relative degree of severity of the disturbances.**

Even though Table 9.1 shows that QAD provides the best overall performance in terms of errors from setpoint, many companies are reluctant to have their control engineers tune even well-behaved control loops for QAD because of the 50% overshoot associated with QAD and because QAD is relatively close to the onset of instability. In addition, because QAD causes significant variation in the MV levels, QAD can result in unduly upsetting other parts of the process. For these reasons, it is probably better to tune well-behaved loops for decay ratios of 1/6 to 1/8. For a process that is more nonlinear with more severe disturbances 1/10 amplitude damping or a critically damped response is more appropriate. In extreme cases, an overdamped tuning criterion may be the proper choice. From Figure 8.2.1, the controller gain for critically damped performance is 0.35, and it is approximately 1.0 for a controller tuned for a 1/6 decay ratio; therefore, the range in the tuning parameters from a 1/6 decay ratio to critically damped is, in general, relatively large (e.g., in this case K_c changes by a factor of 3). **No single tuning criterion works effectively for all control loops because the process nonlinearity, disturbance type and magnitude, and operational objectives must all be considered when choosing the proper tuning criterion, and these factors change from application to application.**

Self-Assessment Questions

Q9.5.1 Explain why different control loop applications can require a different level of controller aggressiveness.

Q9.5.2 Explain in your own words how controllers are tuned when you must consider a tradeoff between performance and reliability.

Self-Assessment Answers

Q9.5.1 Each control loop can have a different effect on the overall objectives of a process. In certain cases, from the overall objectives of a process, the control loop should be detuned, otherwise the operation of the overall process will be adversely affected. At other times, from the overall objectives of a process, the control loop should be operated with as tight control as possible.

Q9.5.2 To make the proper tradeoff between performance and reliability, the control engineer must select the most aggressive controller that provides the required reliability for the system. As the aggressiveness of the controller is increased, the immediate performance of the control loop usually improves but the reliability of the control loop decreases. In general, the range of tuning criterion for this case is between critically damped (conservative) and a decay ratio of 1/6 (most aggressive). As the nonlinearity of the process and the magnitude of the disturbances increases, the tuning criterion should be made more conservative.

9.6 Filtering Sensor Readings

Sensor measurements are well known to contain some degree of noise (Section 3.8). Sensor noise is the variation in the sensor reading that does not correspond to changes in the process and can be caused by background electrical interference, mechanical vibrations and process fluctuations. In feedback control loops, sensor measurements containing significant levels of noise result in noisy errors from setpoint, noisy controller output and noisy MV behavior. Consider the proportional term in a PID controller, i.e., $K_c(y_{sp} - y_s)$. Clearly, if the magnitude of the noise is large compared to the difference between the setpoint and the measured value of the CV, a major portion of the proportional control action will result directly from the noise.

Figure 9.6.1 shows the effect of excessive sensor noise on the feedback control performance of the CST thermal mixer (Example 3.3) for a setpoint increase of 5ºC. The effect of the noise (σ=1.5ºC) on the temperature sensor reading can be seen in the strong variations in the sensor readings. The feedback controller reacts to these variations as evidenced by the variation in the MV. The thick line in Figure 9.6.1 represents the temperature sensor reading without noise (i.e., the noise-free reading). The variability in the noise-free value of the temperature is caused by the response of the feedback controller to relatively large variations in the sensor reading caused by the sensor noise. The variations in the temperature measurement due to noise do not reflect real process changes. On the other hand, because of the action of the feedback

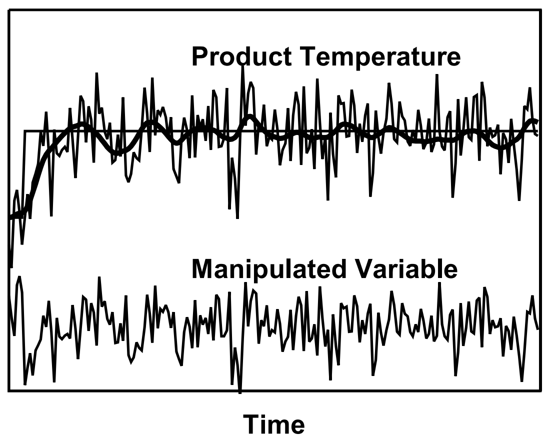


Figure 9.6.1 Feedback control results for the CST thermal mixer for a noisy sensor.

controller on the noisy sensor measurement, the noise does affect the process as shown by the variations in the noise-free sensor reading.

Filters are used to reduce the effect of noise on control loops. Filters, in effect, average a number of sensor readings to reduce the amount of noise on the CV value used by the feedback controller. A first-order **digital filter** is given by

$$y_f(t) = f\,y_s(t) + (1-f)\,y_f(t-\Delta t_f) \qquad \textbf{9.6.1}$$

where $y_f(t)$ is the sensor reading after a digital filter has been applied, $y_f(t-\Delta t_f)$ is the previous value of the filtered sensor reading, $y_s(t)$ is the current sensor reading and f is the filter factor, which is limited to $0 \le f \le 1$. A value of 0.1 for the filter factor is roughly equivalent to a running average of the last 10 sensor measurements while a filter factor of 0.01 is equivalent to an average of the last 100 sensor measurements. The digital filter provides a running average and absorbs some of the short term variations caused by the noise. In general, the value of f used in Equation 9.6.1 for sensor filtering sets the **repeatability reduction ratio**, R (the repeatability of the sensor reading before filtering divided by the repeatability after filtering). Box and Jenkins[2] have shown that the relationship between R and f for a first-order filter (Equation 9.6.1) is given by

$$f = \frac{2}{R^2+1} \quad \text{or} \quad R = \sqrt{\frac{2-f}{f}} \qquad \textbf{9.6.2}$$

Therefore, based on Equation 9.6.2 to attain an R of 5 requires that f=0.077 and a repeatability reduction of 10 requires that f=0.02. Figure 9.6.2 shows the dependence of R on f. A more complete discussion of signal filtering is presented in Appendix C. From Appendix C, the transfer function for a first-order filter is given by

$$G_f(s) = \frac{y_f(s)}{y_s(s)} = \frac{1}{\tau_f s + 1}$$

$$\tau_f = \Delta t_f \left(\frac{1}{f} - 1\right) \qquad \textbf{9.6.3}$$

where Δt_f is the time between applications of the filter. The latter equation shows that the effective time constant associated with the filtering (τ_f) varies directly with the time interval with which the filter is applied (Δt_f).

Figure 9.6.3 shows the results of the application of sensor filtering to the CST thermal mixer for the sensor noise shown in Figure 9.6.1. A digital filter (Equation 9.6.1) was applied to the noisy sensor reading to produce a filtered temperature reading (thin line in Figure 9.6.3), which is used as the measurement of the CV by the feedback temperature controller. A filter factor of 0.05 (R=6.24) was used in this case. Note that the noise-free value of the product temperature (thick line in Figure 9.6.3) shows significantly lower variability about the setpoint than for the case without sensor filtering (Figure 9.6.1). Also, the filter was able to significantly reduce the variation in the MV. To produce the desired dynamic response, the aggressiveness of the controller was reduced when filtering was applied. That is, filtering of the CV adds additional lag to the overall process, requiring less aggressive tuning. The use of less aggressive tuning for the controller and the delay caused by the

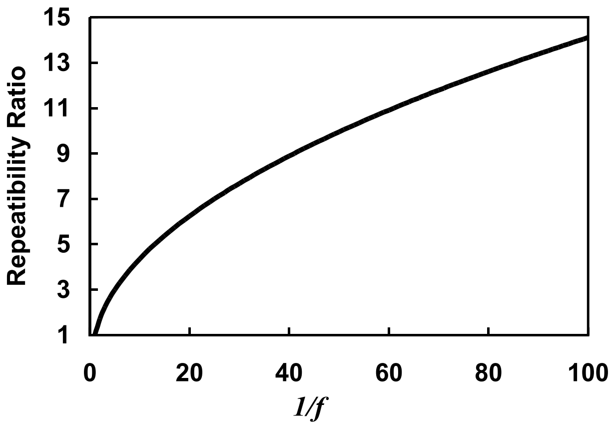


Figure 9.6.2 The repeatability reduction ratio versus the filter factor for a first-order digital

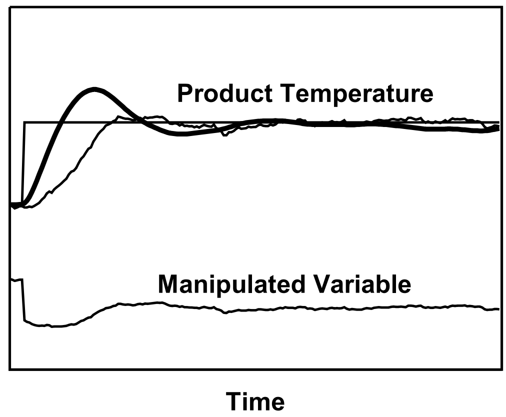


Figure 9.6.3 Feedback control results for the CST thermal mixer with sensor filtering.

filtering process causes the filtered case to respond more slowly than in the case without filtering, i.e., compare Figures 9.6.1 and 9.6.3.

Derivative action is much more sensitive than proportional action to sensor noise because the value of the derivative is proportional to the change in the CV value. Consider the noisy temperature measurement shown in Figure 9.6.1. Note that the instantaneous slope changes sharply from a large positive value to a large negative value due to the sensor noise. If control action is taken based on the instantaneous value of the derivative, the sensor noise can cause large positive and negative swings in the MV that are the result of the noise and not real changes the CV of the process.

In general, derivative action should not be used for a control loop with a noisy sensor. For cases in which there is a relatively large amount of noise, the degree of filtering necessary to effectively apply derivative action usually adds so much extra lag to the loop that it negates the benefit of derivative action.

Example 9.1 Analysis of the Dynamic Behavior of a Feedback System with a Filter on the Sensor Reading

Problem Statement. Analyze the effect of a filter for the sensor reading on the dynamic behavior of a first-order process with a P-only controller, i.e., $G_c(s) = K_c$ (Figure 9.6.4). Assume that the first-order process represents the combined effect of the actuator, process and sensor.

Solution. Figure 9.6.4 shows a block diagram of a feedback controller for which filtering is used on the measurement of the CV. The input to the filter is the sensor reading and the output is the filtered value of the sensor reading, which is compared with the setpoint value.

In a manner similar to the derivation presented in Section 7.2 or by applying Equation 7.2.10, the characteristic equation for the system given in Figure 9.6.4 is

$$G_p(s)G_a(s)G_c(s)G_s(s)G_f(s) \quad 1 \quad 0$$

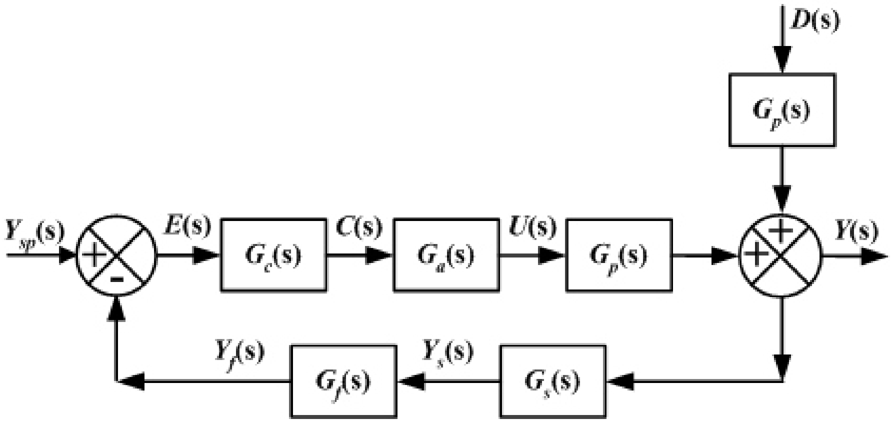


Figure 9.6.4 Block diagram of a feedback control loop with sensor filtering.

From Equation 9.6.3, the transfer function of a first-order filter is given by

$$G_f(s) \quad \frac{1}{{}_f s \quad 1}$$

Note that as ${}_f$ is increased, more filtering is applied (i.e., *f* is decreased). The characteristic equation for this case assuming that $G_a(s)$=$G_s(s)$=1 is given by

$$G_c(s)G_p(s)G_f(s) \quad 1 \quad K_c \frac{K_p}{{}_p s \quad 1} \quad \frac{1}{{}_f s \quad 1} \quad 1 \quad 0$$

Rearranging this equation and putting it into the standard form for a second-order system (Equation 6.4.1) yields

$$\frac{{}_p \quad {}_f}{K_c K_p \quad 1} s^2 \quad \frac{{}_p \quad {}_f}{K_c K_p \quad 1} s \quad 1 \quad 0$$

which based on the standard form for a second-order system [${}_n^2 s^2 \quad 2 \quad {}_n s \quad 1 \quad 0$] indicates that

$${}_n \quad \sqrt{\frac{{}_p \quad {}_f}{K_c K_p \quad 1}}$$

$$\frac{{}_p \quad {}_f}{2\sqrt{{}_p \quad {}_f (K_c K_p \quad 1)}}$$

Therefore, as ${}_f$ is increased (i.e., an increased amount of filtering), the closed-loop response becomes slower and the closed-loop damping factor increases, i.e., it becomes more overdamped. increases because as ${}_f$ is increased, the linear term in the numerator increases at a faster rate than the square root term in the denominator. The comparison of Figures 9.6.1 and 9.6.3 showed that filtering the sensor reading slowed the closed-loop response of the CST thermal mixer, which is consistent with the results of this example.

Example 9.2 The Effect of Filtering a DO Measurement

Problem Statement. Determine the effect on the closed-loop response time for filtering a DO measurement from a bio-reactor. Assume that the unfiltered measurement has a repeatability equal to ±5% and that it is desired to reduce the repeatability of this measurement to ±0.5% by applying a first-order digital filter. Assume that the process time constant is equal to 30 s, the DO sensor has a time constant equal to 5 s and that the cycle time, Δt, is equal to 1 s. Also, assume that a P-only controller is used and that K_cK_p is equal to 1.

Solution. Based on the problem statement, the repeatability reduction ratio, R, required is equal to 10. From Equation 9.6.2, f is equal to 0.02. From Equation 9.6.3, τ_f is equal to 49 s. Applying the characteristic equation for this system yields

$$K_c\left(\frac{K_p}{\tau_p s+1}\right)\left(\frac{1}{\tau_s s+1}\right)\left(\frac{1}{\tau_f s+1}\right)+1=\frac{1}{(30s+1)(5s+1)(49s+1)}+1=0$$

Note that the filtering in this case adds an additional lag that is the largest lag in the closed-loop response. Using simulations without sensor noise, the effective closed-loop time constant of the response for this process increased by 270% with sensor filtering compared to the case with no sensor filtering (see Figure 9.6.5). In this case, the effective closed-loop time constants were estimated by the time required to reach 63.2% of the final response. As a point of reference, assume that filtering corresponding to a filter time constant of 3 s is used instead of a 49 s filter time constant in this example. Then the effective time constant for the closed-loop response would have increased by only 7%. Note that by comparing the relative magnitudes of the time constants of the process, sensor and filter, you can usually determine whether filtering will significantly reduce the speed of the closed-loop response.

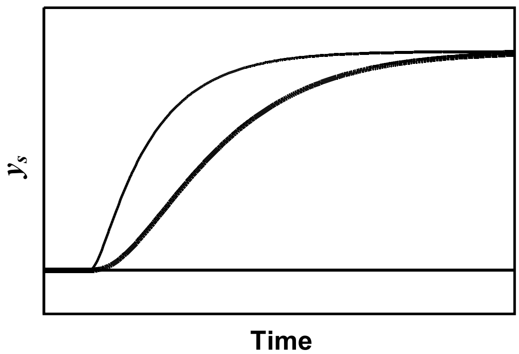


Figure 9.8.5 Comparison between the response of the DO process with filtering (thick line) and without filtering (thin line).

Example 9.3 Comparison Between RTDs and TCs.

Problem Statement. Consider the application of a TC and an RTD for the measurement of the CV in a temperature control loop. Assuming that the sensor filter is applied every 0.2 s, determine the filter time constants for the TC and the RTD if it is desired to attain a filtered value of the CV with a repeatability equal to ±0.02ºC. Assume the repeatability for the unfiltered TC reading is ±1.0ºC and is ±0.1ºC for the unfiltered RTD measurement, which is consistent with Table 2.4.

Solution. Consider the RTD first. The repeatability reduction, R, is equal to 0.1/0.02=5. Using Equation 9.6.2, f=0.0770. Then from Equation 9.6.3, τ_f=0.2s(1/0.077-1)=**2.4 s**.

Now consider the TC. The repeatability reduction, R, is equal to 1/0.02=50. Using Equation 9.6.2, f=8×10^{-4}. Then from Equation 9.6.3, $_f$=0.2s(1/8×10^{-4}-1)=**250 s**. That is, the filter time constant for the TC is larger than the one for the RTD by a factor of 100! This is consistent with the guideline given in Chapter 2 that RTDs should be used over TCs for important temperature control points.

Filter Tuning Guidelines. Sensor readings should usually be filtered to reduce the influence of sensor noise on feedback control performance, as demonstrated earlier in this section. Filtering, however, adds lag to the closed-loop response (Example 9.1). In certain cases, tuning a filter on a sensor can involve balancing the benefits of reducing the noise against the detrimental effects of adding lag to the overall process. In other cases, if possible, the time constant of the filter should be significantly smaller than the other dominant time constants in the actuator/process/sensor system so that filtering does not significantly slow down the response of the system.

It is usually more convenient to use the filter time constant (i.e., the time constant of the first-order filter), $_f$, to specify the amount of filtering applied to a sensor reading. In this manner, the time constant of the filter can be directly compared to the time constants of the actuator, process and sensor to determine whether it will significantly affect the closed-loop dynamics. The filter factor, f, and the cycle time for applying the filter, t_f, can be used to calculate the filter time constant by Equation 9.6.3:

$$_f \qquad t_f \; \frac{1}{f} \quad 1$$

For most DCSs, sensor readings can be updated 5 or 6 times per second (i.e., t_f is 0.16 to 0.2 s) by specification instead of the standard one time per second. For a filter time constant of 5 s, the filter factor is equal to 0.04, which from Figure 9.6.2 indicates almost a factor of 7 reduction in the repeatability of the sensor reading; therefore, relatively extensive sensor filtering can result for most sensors using high-frequency updating by a DCS.

The amount of sensor filtering required is dependent on the amount of noise on the reading. For example, a reading from a thermocouple is expected to require much more filtering than a reading from an RTD, which has an order of magnitude smaller repeatability (Table 2.4), for the same application as shown in Example 9.3. From an examination of Table 2.4, you can see that most properly functioning sensors usually have a relatively small amount of noise. As a result, most flow, level, pressure and temperature sensors can be filtered effectively using a filter time constant of 3 to 5 s, which does not normally affect the closed-loop dynamics. Composition analyzer readings from GCs are updated so infrequently that filtering is usually not used for them. On the other hand, if composition measurements are available at a sufficiently high frequency, filtering can be used effectively.

In certain cases, the filtering of a noisy sensor is required. Noisy sensors present a challenge. A nuclear-based level sensor, a pressure sensor located too close to a 90º elbow in a line or an orifice flow meter located immediately downstream of a control valve are examples of noisy sensors. For these cases, tuning the filter is a compromise between removing the noise from the sensor reading and adding lag to the closed-loop response when you are forced to use a noisy sensor reading.

While digital first-order filters are commonly used in industry for noisy signals, there are also other forms of filters utilized[1], such as moving average filters where the output of a filter represents the average of the past "n" measurements, with "n" specified by the user. A key difference between a first-order digital filter and a moving average filter is that a moving average filter weighs the current and recent past measurements the same while a first-order digital filter puts a different emphasis on the current measurement than a moving average filter depending on the value of the filter constant. Another technique sometimes used in industry is to subject incoming measurements to if-then-else rules to look for "data outliers" (signal noise from sources other than the process itself), flag them as invalid, and prevent them from being passed on to any PID controllers. This technique does not filter out high frequency process noise, but does prevent invalid measurements from reaching the PID controller.

Self-Assessment Questions

Q9.6.1 How does filtering a sensor measurement reduce the variation in the MV value for a control loop?

Q9.6.2 What price, in terms of controller performance, is paid for using a filter on sensor readings?

Q9.6.3 When should you not use derivative control?

Q9.6.4 Explain why a filter time constant between 3-5 s provides adequate filtering for properly functioning sensors.

Q**9.6.5** Explain why a filter time constant between 3-5 s does not usually significantly slows the closed-loop response.

Self-Assessment Answers

Q9.6.1 A filter reduces the magnitude of sensor noise by keeping a running average of past measurements. That is, the previous filtered value of the sensor reading is averaged with the current reading based on the value of the filter factor.

Q9.6.2 Filtering reduces the variation in the a sensor reading, but filtering adds an additional lag to the feedback system. If the filter time constant is small compared to the lags of the actuator, process and sensor, the extra lag added by the filter will be relatively small. On the other hand, if the filter time constant is comparable to the largest lags of the actuator, process and sensor, sensor filtering will significantly slow the closed-loop response.

Q9.6.3 You should not use derivative action if the process sensor has a significant noise level.

Q9.6.4 With normal updating of the sensor reading, a filter time constant of 3-5 s does not normally add enough additional lag to affect the dynamics of the closed-loop system.

Q9.6.5 A filter time constant of 3-5 s adds additional lag that is equal to 3-5 s. For most control loops, the dominant lags in the system are usually several minutes which makes the effect of a 3-5 s lag insignificant.

9.7 Recommended Approach for Controller Tuning

The following procedure is recommended for tuning important PID control loops, i.e., control loops for which it is important that the control loop perform according to the selected tuning criterion.

1. Select the tuning criterion for the control loop. The tuning criterion depends on how the control loops affect the overall process objectives and can involve applying a compromise between performance and reliability (Sections 9.4 and 9.5).

2. Apply filtering to the sensor reading. Sensor filtering (Section 9.6) reduces the effect of sensor noise on the variability in the CV but can introduce lag to the feedback system, which is detrimental to control performance. Therefore, filtering should be applied carefully.

3. Determine if the control loop is a fast- or slow-responding control loop. The distinction between fast- and slow-responding control loops is concerned with the closed-loop response time of the system. If a set of controller settings can be tested using setpoint changes in a reasonable period of time (e.g., less than 10 minutes), the process is a fast-responding control loop. If not, it is a slow-responding control loop.

4. **For fast-responding control loops, apply field tuning** (Section 9.8).

5. **For slow-responding control loops, apply the ATV-based tuning procedure** (Section 9.9).

6. **For process with significant deadtime-to-time constant ratios, apply PID tuning** (Section 9.10).

Table 9.3
Typical Tuning Parameters for Common Loops in the CPI

Loop Type	PB	$_I(s)$	$_D(s)$
Flow Controller	100 - 500%	0.2 - 2.0	0
Gas Pressure Controller	1 - 15%	5-100	0
Liquid Pressure controller	100 - 500%	0.2 - 2.0	0
Level Controller	5 - 50%	5 - 60	0
Temperature Controller	10 - 50%	40 - 4000	30 - 2000*
Composition Controller	100 - 1000%	100 - 5000	30 - 4000*

* $_D$ should always be smaller than $_I$.

Table 9.3 lists ranges of PID tuning parameters for flow controllers, gas pressure controllers, liquid pressure controllers, level controllers, temperature controllers and composition controllers. Note that the gain is expressed in proportional band (Section 7.3) and the larger the *PB*, the lower the value of K_c. Flow control loops are actually a special case for tuning. Because of the sustained oscillations that result about the setpoint (Figure 7.9.3), flow control loops are usually tuned with more integral action than proportional action compared with most other control loops, which is consistent with Table 9.3.

9.8 Tuning Fast-Responding Control Loops

Fast-Responding Loops. For fast-responding loops, such as flow control and certain pressure control loops, the simplest and quickest tuning method available is field tuning, which is based upon a trial-and-error selection of tuning parameters. Some level and temperature loops also behave as fast-responding control loops. A fast-responding control loop is defined here as a control loop that has a closed-loop response time of 10 minutes or less. Because these processes respond quickly, trial-and-error tuning is effective. It is easier to field tune a

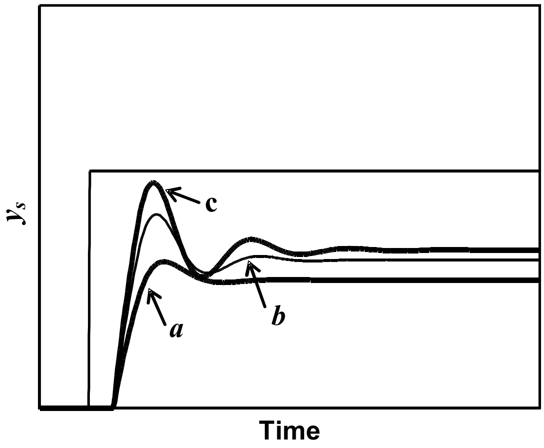


Figure 9.8.1 Selection of K_c during field tuning (a) Results for initial value of K_c (b) Results for an increase in K_c (c) Results for final value of K_c (1/6 decay ratio).

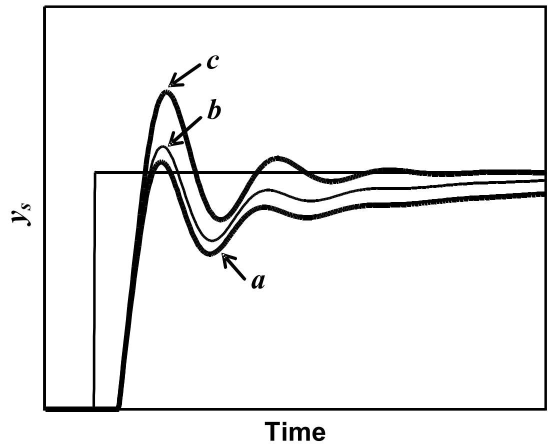


Figure 9.8.2 Selection of $_I$ during field tuning (a) Results for initial value of $_I$ (b) Results for a decrease in $_I$ (c) Results for the final value of $_I$ (1/6 decay ratio).

fast-responding loop rather than identify FOPDT parameters, use initial tuning parameters from a chosen tuning method and adjust the tuning to meet the selected tuning criterion. The recommended procedure for field tuning, assuming that the tuning criterion has been selected and the sensor reading filtered, follows.

1. Turn off the derivative action ($_D$ 0) and the integral action ($_I$).
2. Use an initial estimate of K_c , e.g., $K_c = \frac{1}{2K_p}$. Estimate K_p from process knowledge.
3. Using setpoint changes, increase K_c in small increments until the response meets the tuning criterion. (See Figure 9.10.1, which is based upon a 1/6 decay ratio). For tuning a P-only controller, the tuning procedure is completed.
4. For tuning a PI controller, decrease K_c by 10%.
5. Use an initial value of $_I$, i.e., $_I$ 5 $_p$. Estimate $_p$ from process knowledge.
6. Decrease $_I$ until offset is eliminated and the tuning criterion is met for setpoint changes. (See Figure 9.10.2, which is also based upon a 1/6 decay ratio).
7. Check to ensure that adequate levels of proportional and integral action are being used.

PID controllers, which are not usually applied to fast-responding processes, are discussed in Section 9.10.

Self-Assessment Questions

Q9.8.1 How do you determine if field tuning is the proper approach for a particular control loop?

Q9.8.2 After a P-only controller is tuned in the recommended field tuning approach, why is K_c reduced by 10% for a PI controller?

Self-Assessment Answers

Q9.8.1 When the process is relatively fast responding (e.g., a flow controller, a level controller, and certain pressure or temperature controllers), field tuning should be used. When the process responds more slowly, ATV-based tuning is preferred because field tuning of slow-responding processes can take an excessive amount of time and is susceptible to corruption due to unmeasured disturbances.

Q9.8.2 After tuning a P-only controller, K_c is reduced by 10% because adding integral action increases the oscillatory nature of the closed-loop response (See Section 7.7). In addition, based on ZN tuning the PI controller uses a controller gain that is 10% smaller than the P-only controller (Table **9.3**).

9.9 Tuning Slow-Responding Control Loops

Slow-Responding Processes. For slow-responding loops (e.g., cell concentration and certain temperature and composition control loops), field tuning can be a time-consuming procedure that leads to less than satisfactory results. Step test results can be used to generate FOPDT models and tuning parameters can be calculated from a variety of techniques. This approach suffers from the fact that it takes approximately the open-loop response time of the process to implement a step test and during that time, measured and unmeasured disturbances can affect the process, thus corrupting the results from the step test. In addition, it is unlikely that the tuning approach selected is based on the proper tuning criterion. Because of model mismatch and the likely mismatch of the tuning criterion, significant adjustments to the tuning parameters are usually required.

The **ATV**[3] (autotune variation) method determines the ultimate gain and period in a manner similar to that of the ultimate method, but ATV tests can be implemented without unduly upsetting the process. Controller settings are calculated, and the controller is then tuned on-line to meet the selected tuning criterion.

Figure 9.9.1 graphically demonstrates the ATV method. The user selects h, the relay height used or the change in the MV applied. The chosen value of h should be small enough that the process is not unnecessarily upset, yet large enough that the resulting amplitude, a, can be accurately measured.

To initiate an ATV test, the process should be at or near steady-state conditions, c_0 and y_0. Next, the controller output is set to $c_0 + h$ (or $c_0 - h$) until y deviates noticeably from y_0. At that point, the controller output is set to $c_0 - h$ (or $c_0 + h$), which turns the process back toward y_0. Then, each time y crosses y_0, the controller output is switched from $c_0 + h$ to $c_0 - h$ or vice versa. This process is also referred to as a relay feedback experiment. A standing wave is established after 3 to 4 cycles; therefore, the values of a and the ultimate period, P_u, can be measured directly and the ATV test is concluded. The **ultimate controller gain**, K_u , is calculated by

$$K_u \quad \frac{4h}{a} \qquad \textbf{9.9.1}$$

The **ultimate period** is obtained directly from the ATV test (see Figure 9.9.1). K_u and P_u can be used in one of several tuning schemes. One tuning approach is the Ziegler-Nichols (ZN) ultimate settings (Table 9.3). Consider the ZN settings for a PI controller:

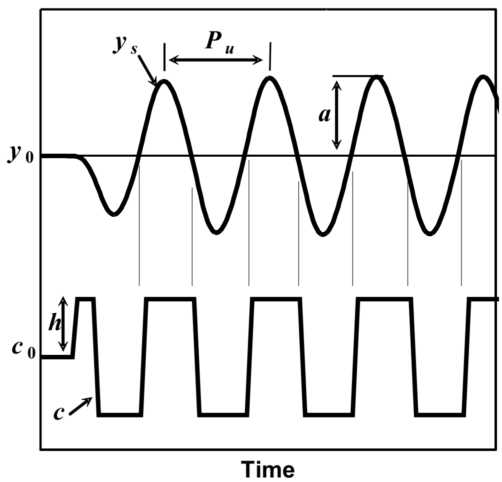


Figure 9.9.1 Graphical representation of an ATV test.

$$K_c^{ZN} \quad 0.45\,K_u$$
$$_I^{ZN} \quad P_u / 1.2$$

ZN settings are fairly aggressive and can lead to ringing behavior for nonlinear processes due to the relatively small value of $_I$ (i.e., large integral action).

Another tuning approach that was developed for processes that behave like an integrator plus deadtime system is the Tyreus and Luyben (TL) settings[4]:

$$K_c^{TL} \quad 0.31K_u$$
$$_I^{TL} \quad P_u / 0.45$$

The TL settings are less aggressive with considerably less integral action than the ZN settings. The TL settings are recommended for more sluggish processes that are well represented as integrator plus deadtime for a good portion of its step test (e.g., a sluggish distillation column). After the ZN or TL settings are calculated, they may require on-line tuning, particularly for the ZN settings to meet the desired tuning criterion (e.g., 1/6 decay ratio or critical damping). For example, the ZN settings are tuned on-line as follows:

$$K_c \quad K_u^{ZN} / F_T$$
$$_I \quad _I^{ZN} \quad F_T \qquad \textbf{9.9.2}$$

by adjusting $\boldsymbol{F_T}$ on-line. F_T, which is the on-line tuning factor, is adjusted to select the desired controller aggressiveness. As F_T is increased, K_c decreases while $_I$ increases by the same proportion, resulting in detuning of the controller. As F_T is decreased, K_c increases and $_I$ deceases, both of with increase the aggressiveness of the controller tuning. Therefore, F_T can be adjusted to meet the performance requirements for each individual application. As a result, on-line tuning has been reduced to a one-dimensional search for the proper level of controller aggressiveness for a PI controller. If the controller is too aggressive, F_T is increased; if the controller is too sluggish, F_T is decreased.

The procedure based on ATV identification with on-line tuning is applicable to PI controllers. For certain cases, after this procedure has been applied, it will be evident that the proper balance between proportional and integral action has not been found, e.g., if offset elimination is slow. In these cases, adjustments in the relative amount of proportional or integral action may be required. For example, if the TL settings were used and not enough integral action resulted, the 0.45 factor in the TL settings for integral action (i.e., P_u /0.45) could be increased to speed up offset elimination. Figures 8.3.5 and 8.3.6 can be helpful in determining if not enough proportional action or not enough integral action is being used.

Consider the application of an ATV test to a dynamic simulator of a C_3 (propylene/propane) splitter. Figure 9.9.2 shows an ATV test and an open-loop step test on the same time scale for the bottom product composition control loop. Note that the four cycles of the ATV test require 6-8 hours while the open-loop test requires in excess of 60 hours. The ATV results were used with TL settings and the results for three different tuning factors are shown in Figure 9.9.3.

Summarizing, identifying the ultimate controller gain and ultimate period of a slow-responding loop using the ATV method is relatively fast, providing a "snap shot" of the process without unduly upsetting the system. In

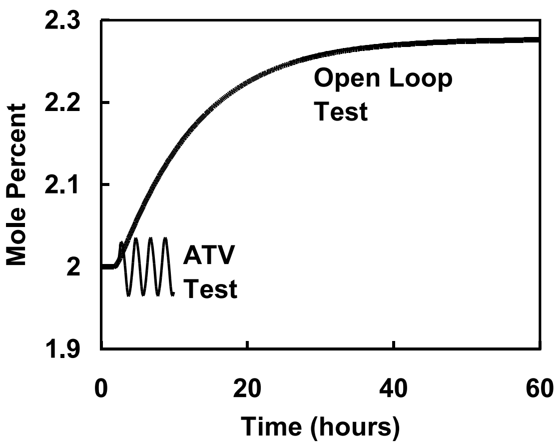


Figure 9.9.2 Comparison of an ATV and an open-loop step test.

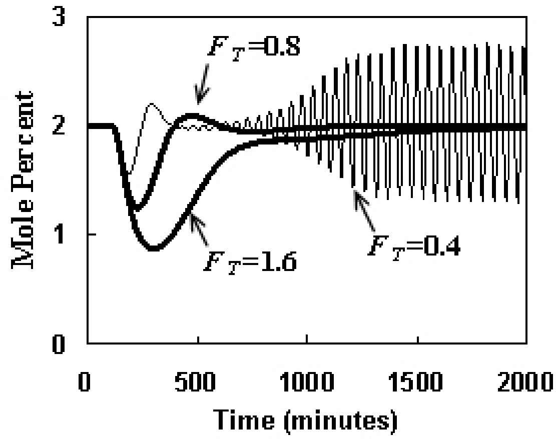


Figure 9.9.3 Effect of FT on dynamic response for a disturbance upset.

addition, the on-line tuning procedure provides a systematic method of selecting the proper degree of controller aggressiveness. Therefore, the ATV test with on-line tuning represents an industrially relevant means of attaining high quality controller tuning for loops with large response times.

Example 9.4 Example of ATV Identification and On-line Tuning

Problem Statement. Apply ATV identification with on-line tuning to a PI controller applied to the CST composition mixer process given in the visual basic simulator that comes with this text. Tune for a 1/6 decay ratio.

Solution. Figure 9.9.4 shows an ATV test using a value of h equal to 0.5 kg/s applied to the specified value of F_1. From Figure 9.9.4, the values of a and P_u are

$$a = 0.107 \text{ g mol/l}$$

$$P_u = 30 \text{ min}$$

Because y_s does not oscillate symmetrically about y_0, an average value of a is computed, i.e., a is calculated as one-half of the peak to valley difference. Then, using Equation 9.9.1 results in

$$K_u = 5.96 \text{ (kg/s) /(g mol/l)}$$

Then, the ZN settings are

$$K_c^{ZN} = 2.68 \text{ (kg/s) / (g mol/l)}$$

$$_{I}^{ZN} = 25 \text{ min}$$

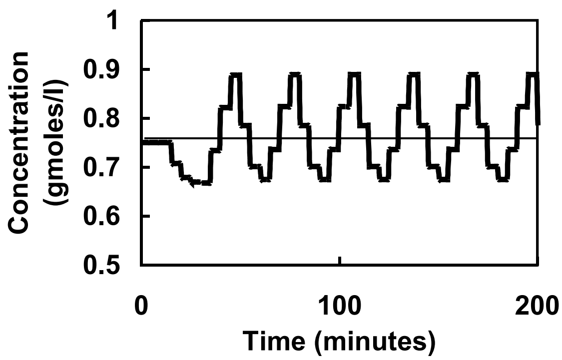
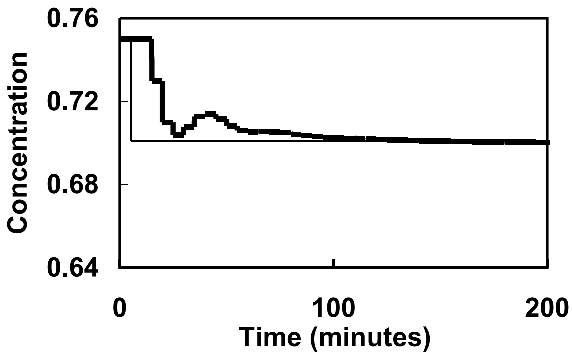
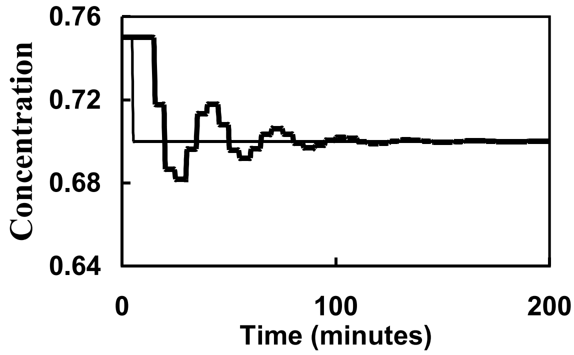


Figure 9.9.4 ATV test on CST composition mixer.



(a)



(b)

Figure 9.9.5 Dynamic response for the CST composition mixer for different tuning factors. (a) F_T=0.75 (b) F_T=0.5.

Therefore,

$$K_c \quad \frac{2.68}{F_T} \quad (\mathrm{kg/s})/(\mathrm{g\ \ mol/l})$$

$$_I \quad 25 \quad F_T \quad (\mathrm{min})$$

Figures 9.9.5a and b show the closed-loop results for a setpoint change in the product composition for F_T equal to 0.75 and 0.5, respectively. For F_T equal to 0.75 (Figure 9.9.5a), the decay ratio is 1/12 and for F_T equal to 0.5 (Figure 9.9.5b) the decay ratio is 1/2.6; therefore, the tuning factor for a decay ratio of 1/6 is between 0.5 and 0.75. In fact, for a decay ratio of 1/6, F_T is equal to 0.61. In this case, because the oscillations are not always symmetric about the setpoint, the decay ratio can be determined based upon adjacent peak to valley heights. Figure 9.9.6 shows the control performance for the tuned controller for a step change in feed composition for stream 1 from 0.5 to 0.55 g mol/l. The results shown in Figure 9.9.6 indicate that, because the oscillations are not symmetric about the setpoint and the offset is removed slowly, not enough integral action was applied. The results shown in Figure 9.9.7 correspond to a 30% reduction in the reset time. Note that the oscillations are much more symmetric

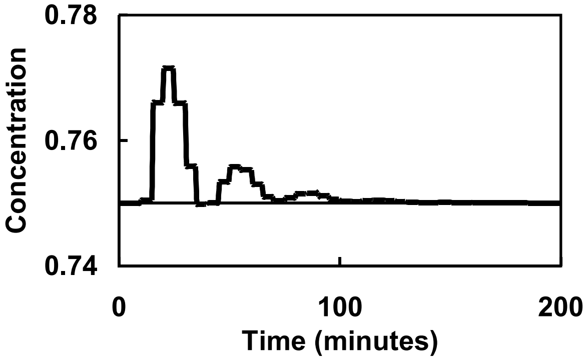


Figure 9.9.6 Control performance for the CST composition mixer for a feed composition upset.

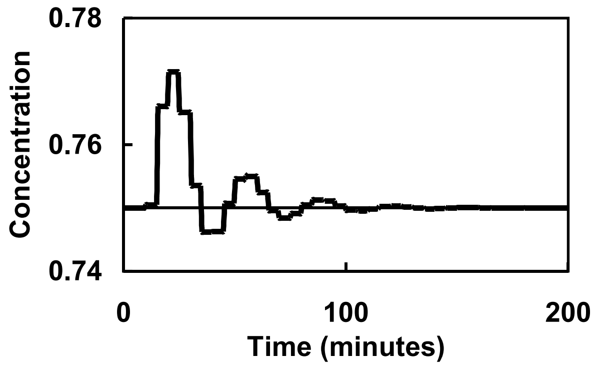


Figure 9.9.7 Control performance for the CST composition mixer for a feed composition upset with 30% additional integral action.

about the setpoint.

Example 9.5 PI Tuning for a Sensor with Noise

Problem Statement. Using the simulator that comes with this text, tune a PI controller for the thermal mixing tank (Example 3.2). Assume that the sensor has a repeatability of ±0.2ºC. Tune this controller for a 1/6 decay ratio and filter the sensor reading so that the MV is not noisy.

Solution. Based on the repeatability, (Section 3.6) is equal to 0.1ºC. Using open-loop step changes to the MV, the effect of a filter factor equal to 0.05 and 0.01 are shown in Figure 9.9.8. While a filter factor of 0.01 provides improved smoothing of the filtered value of the sensor reading, the results for a filter factor of 0.05 smooth the reading enough without excessively slowing the process response. Therefore, the filter factor for the sensor filter is set equal to 0.05. With sensor filtering applied, an ATV test is conducted, yielding a=0.383 ºC and P_u=25.5 s. Applying Equation 9.9.1,

$$K_u \quad \frac{4h}{a} \quad \frac{4 \quad 0.05}{3.14 \quad 0.383} \quad 0.166$$

Using the ZN settings with the tuning factor yields

$$K_c \quad \frac{0.45K_u}{F_T} \quad \frac{0.45 \quad 0.166}{F_T} \quad \frac{0.0747}{F_T}$$

$$_I \quad \frac{P_u F_T}{1.2} \quad \frac{25.5F_T}{1.2} \quad 21.25F_T$$

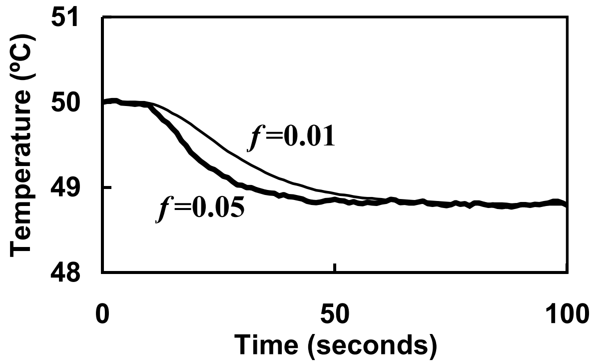


Figure 9.9.8 Effect of filter factor on open-loop response.

Figure 9.9.9 shows the closed-loop response for a tuning factor (F_T) of 2 and 1.1. The response for F_T=1.1 meets the specified decay ratio, but the oscillation about the setpoint are below the setpoint, indicating that more integral action is required. Figure 9.9.10 shows the closed-loop response after the reset time was reduced by 14% from the previous case with F_T=1.1.

The effect of the controller aggressiveness on the decay ratio for the thermal mixing process (TMIXER) and the heat exchanger (HEATEX) used in the visual basic simulator that accompanies this text is shown in Figure 9.9.11. For both simulators, the controller tuning was adjusted to attain a critically damped response using a PI controller. Then, the PI settings for a critically damped response were adjusted for increasing controller aggressiveness using the following formulas:

$$K_c \quad K_c^{CD} \quad F_A \qquad\qquad _I \quad {}_I^{CD} \,/\, F_A$$

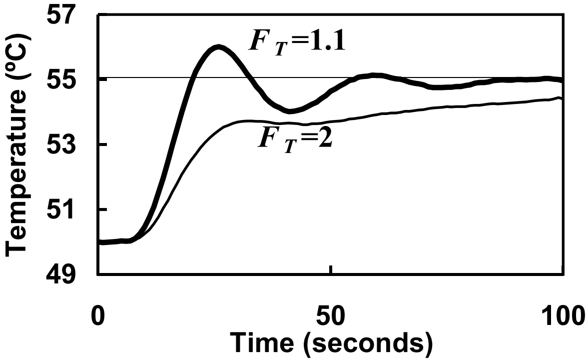


Figure 9.9.9 Effect of F_T on closed-loop response.

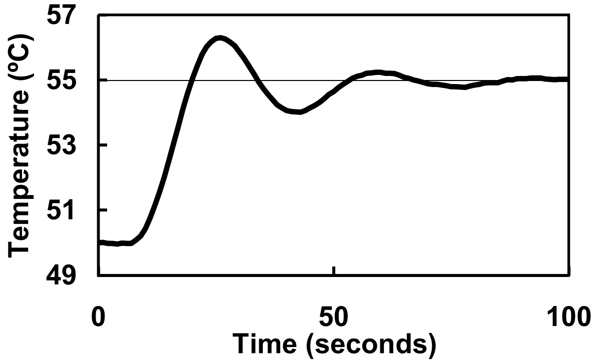


Figure 9.9.10 Desired tuning results for Example 9.9 (K_c=0.0679; τ_I=20 s)

Therefore, as F_A is increased, the aggressiveness of the controller increases. SO in Figure 9.9.11 indicates the decay ratio (*DR*) for sustained oscillation (*DR*=1), which also indicates the onset of instability. From this figure, the thermal mixer tuned for a critically damped response remains stable if the critically damped tuning factors are increased in aggressiveness by a factor of 3 while the heat exchanger can handle an increase in aggressiveness by a factor of 2.4. Therefore, if the thermal mixer was tuned for a critically damped response, the control loop would be expect to remain stable as long as the process gain did not change by more than a factor greater than 3 and the heat exchanger would remain stable as long as the gain did not increase by a factor of greater than 2.4. Using this same reasoning, if the controller is tuned for a 1/6 decay ratio, the gain of the thermal mixer and the heat exchanger could increase by about 50% and remain stable. Therefore, if the process gain is not expected to change by more than 50%, tuning for a 1/6 decay ratio should remain stable.

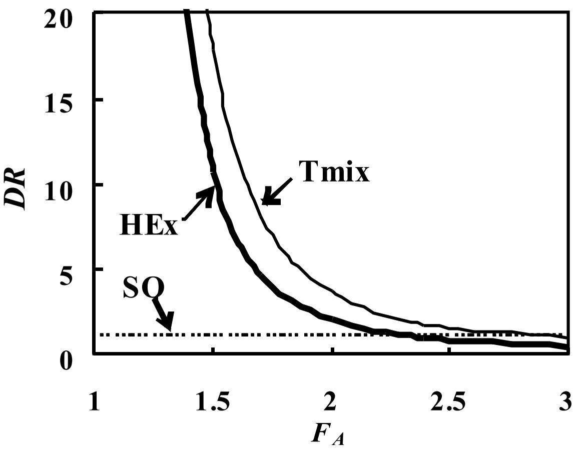


Figure 9.9.11 Decay ratio (DR) as a function of the controller aggressiveness (F_A).

Self-Assessment Questions

Q9.9.1 Why is ATV-based tuning recommended for tuning slow-responding processes?

Q9.9.2 What is the difference between ZN and TL settings?

Self-Assessment Answers

Q9.9.1 A slow-responding process will have a large time constant. An open-loop step test takes a very long time for a slow process. In this case, ATV will be advantageous in identifying the ultimate gain and period relatively quickly without unduly upsetting the process.

Q9.9.2 ZN settings result in much more aggressively tuned controllers than when the TL settings are used. For example, ZN settings use controller gains that are 45% larger than TL settings and ZN settings use reset times that are a factor of 2.6 smaller than TL settings.

9.10 PID Tuning

The tuning of PID controllers applied to slow-responding processes is less systematic than tuning PI controllers because the on-line tuning procedure (Equation 9.9.2) is not generally effective for setting the amount of derivative action. That is, applying a tuning factor, F_T, to the derivative time and tuning a PID controller by adjusting only F_T does not, in general, lead to a well-tuned PID controller. The recommended procedure for tuning PID controllers is as follows

1. Tune a PI controller using ATV identification with on-line tuning. Make sure that the proper balance between proportional and integral action is used. It may be necessary to reduce τ_I to produce symmetric oscillations about the new setpoint.
2. Add derivative action and tune τ_D for minimum response time for a setpoint change keeping K_c and τ_I fixed. Initially set τ_D equal to $P_u/8$ where P_u comes from the ATV test.
3. Because the application of step 2 moves the dynamic response toward critically damped behavior, increase K_c and τ_D by the same factor until the desired dynamic response is obtained.
4. Check the response to ensure that the correct level of integral action is being used.

Example 9.6 PID Tuning Example

Problem Statement. Apply the proposed PID tuning procedure to tune a PID controller for a 1/6 decay ratio for the heat exchanger used in the visual basic simulator provided with this text.

Solution. An ATV test was applied to the simulation of the heat exchanger and the amplitude of the resulting standing wave was 0.228ºF with an ultimate period of 42 s for a relay height (h) of 2.4ºF for the steam temperature. The TL settings were applied because of the sluggish nature of the process and the on-line tuning factor was adjusted until a 1/6 decay ratio was obtained for a setpoint change from 104ºF to 114.4ºF, resulting in F_T equal to 0.3. Additional integral action was added to produce symmetric oscillation about the setpoint. The control results for the PI controller (K_c=13.8 and τ_I=15 s) are shown in Figure 9.10.1. Next, derivative action was adjusted until the minimum settling time (Figure 9.10.2) was obtained (i.e., τ_D was set equal to 2.0 s) with K_c and τ_I fixed. As expected, the response became less oscillatory as shown in Figure 9.10.2. Therefore, K_c and τ_D were increased together to meet the specified decay ratio (i.e., 1/6). During these tests, it was determined that too much integral action was being applied; therefore, τ_I was increased from 15 s to 18 s. Then, K_c and τ_D were both increased by 33% yielding a 1/6 decay ratio. From Figure 9.10.3, which compares the PID and PI controllers for this problem, the PID controller reduced the settling time (based on ±5% variation about setpoint) by 55% and reduced the overshoot by 47% compared to the corresponding PI controller tuned for a 1/6 decay ratio.

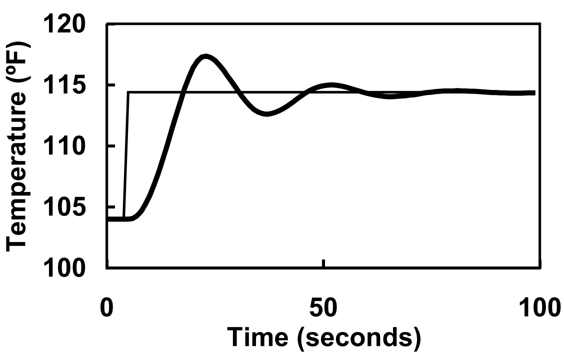


Figure 9.10.1 Temperature control performance for PI controller (1/6 decay ratio).

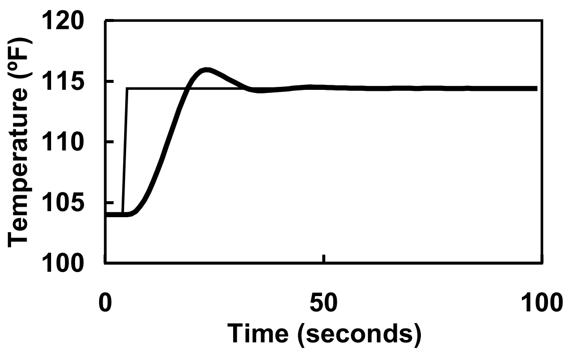


Figure 9.10.2 Temperature control performance after $_D$ adjusted for minimum response time.

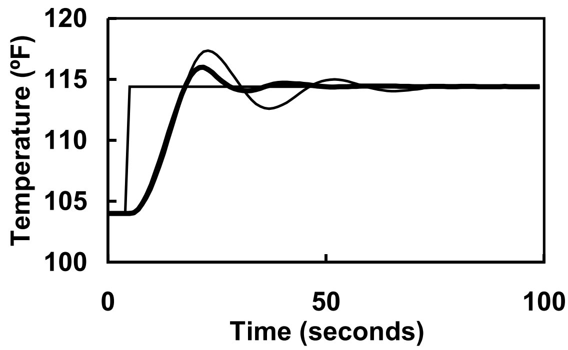


Figure 9.10.3 Temperature control performance for the PI (thin line) and the PID (thick line) controllers.

Self-Assessment Questions

Q9.10.1 Consider the addition of derivative action to the PI tuned controller in the PID tuning procedure. Why does the closed-loop damping factor decrease when derivation action is added?

Q9.10.2 Why does the use of derivation action allow for an increase in the controller gain for a sluggish process?

Self-Assessment Answers

Q9.10.1 When derivative action is added to a PI controller, the closed-loop damping factor decreases because derivative action acts against the local slope, reducing the oscillatory nature of the response.

Q9.10.2 When derivative action is added, it reduces the closed-loop damping factor, which allows for an increase in the controller gain. That is, when the closed-loop damping factor is reduced, the closed-loop response is not at the specified tuning criterion, allowing additional controller aggressiveness to be added (i.e., an increase in the controller gain).

9.11 Simplified Field Tuning

When commissioning a new plant or completing a major retrofit to an existing plant, up to thousands of control loops may require tuning with the majority of these loops considered as non-critical (i.e., not directly affecting product quality or safety). A challenge for batch processes (most of which are non-linear and time varying) is establishing representative conditions in the process in which to pursue loop tuning. As a result, the control engineer needs a simplified means to tune these non-critical loops quickly and easily. Most level and pressure loops and many flow control loops fall into the category of non-critical loops. Field tuning (Section 9.8) and ATV-based tuning (Section 9.9) are efficient ways to tune loops but still take more time and effort than these low-priority loop require.

The field tuning method for a PI controller can be simplified by simply setting the reset time ($_I$) equal to the estimated process open-loop response time and then adjusting the controller gain (K_c) by trial-and-error to obtain an overdamped response. The estimate of the open-loop response time can be made based upon experience with similar loops, i.e., based on the approximate time required a change to occur in a CV after an input change. An overdamped response is selected in order to guarantee reliability. Therefore, tuning most PI controller can be reduced to a simple adjustment of the controller gain and many times the control engineer will have a rough idea of what controller gain to use. As a result, these control loops can be reliably tuned in a short period of time.

In certain sectors of industry, this simplified tuning procedure is known as one form of Lambda tuning. This is due, in part, to Lambda tuning being a relatively conservative tuning procedure. When used as a simplified field tuning procedure, the implementation is normally to use a PI controller, set the integral time equal to the process time constant and set K_c by trial and error, making use of whatever control engineer experience and/or tuning settings on similar equipment that may exist. Note, however, that a more formal version of Lambda tuning exists, where K_c is calculated using process parameters determined by an open loop test and choosing lambda to achieve a desired closed loop response (i.e., lambda is defined as the closed loop time constant of the process and so has actual physical meaning). Since the more formal version of Lambda tuning is a form of IMC tuning, which uses a step test to identify FOPDT properties of the process, it suffers from the same limitations as IMC tuning.

9.12 Tuning Level Controllers

Level Controller Tuning. If a level control process is fast responding, then field tuning is effective. If the level control process is relatively slow responding, it can be helpful to use the following approach to select the initial settings for the level controller. Marlin[5] developed closed-form solutions for the dynamic behavior of PI and P-only control of the level in a constant cross-section tank. He used these expressions to derive analytical expressions for the tuning parameters that result in a **critically damped response** for the closed-loop level control process:

$$K_c \quad \frac{F_{MAX}}{L_{\max}} \quad \text{P} \quad \text{only control} \tag{9.12.1}$$

$$\begin{aligned} K_c &\quad \frac{0.736\, F_{MAX}}{L_{MAX}} \\ {}_I &\quad \frac{4A_c}{K_c} \end{aligned} \quad \text{PI control} \tag{9.12.2}$$

where A_c is the cross-sectional area of the tank,   is the density of the liquid, F_{MAX} is the maximum expected step change in the feed rate to the tank and L_{MAX} is the desired level change that F_{MAX} should cause under feedback conditions. Remember from Example 7.12 that P-only control provides offset-free operation for integrating systems. Therefore, if a level system is non-self-regulating with respect to the MV, P-only control (Equation 9.12.1) should be used and if a level system is self-regulating, a PI controller (Equation 9.12.2) is required for offset-free operation.

These tuning relations can be used for both tight level control and loose level control depending upon the selection of L_{MAX}. If L_{MAX} is selected to correspond to about a 2% level change, it represents tight level control and K_c has a correspondingly high value. On the other hand, if L_{MAX} is selected to correspond to a 20% level change, it represents quite loose level control and K_c is lower by a factor of 10 in this case.

This analysis is based on an idealized model of the level of a tank and does not consider sensor or actuator dynamics and does not consider that horizontal cylindrical tanks do not have a constant cross-section. For these reasons, it is recommended that Equation 9.12.1 and 9.12.2 be used as initial estimates of the tuning parameters and that an on-line tuning factor, F_T, be used to tune for the desired level control performance:

$$K_c \qquad K_c / F_T$$
$$_I \qquad _I \; F_T$$

Example 9.7 Calculation of Initial Tuning Parameters for a Level Controller

Problem Statement. Consider level control in a horizontal cylindrical tank 6 feet in diameter and 20 feet long. Normally, the feed rate to the tank is 10,000 pounds per hour of a dilute aqueous solution. Feed rate step changes are normally within the range of ± 10% of the normal feed rate. The setpoint for the level is usually 20%. The pressure taps for the level indicator are located at the top and bottom of the tank. Determine the tuning parameters for a PI controller that maintains the level within 5% of setpoint based upon Equation 9.12.2 for 10% feed rate changes. Assume that this level process is self-regulating.

Solution. By geometric analysis, the width of the liquid level in the tank at 20% full is 4.8 feet; therefore, the cross-sectional area is 96 ft^2. Using the density of water,

$$F_{MAX} \quad (0.1)(10{,}000\ \text{lb/h}) \frac{\text{h}}{60\,\text{min}} \quad 16.67\ \ \text{lb/min}$$

$$K_c \quad \frac{0.736\ \ (1000\ \ \text{lb/h})}{5\%} \quad 147.4\ \frac{\text{lb/h}}{\%}$$

$$_I \quad \frac{(4)\ (96\ \ \text{ft}^2)\ (62.4\ \ \text{lb/ft}^3)\ (6\ \text{ft}/100\%)}{(147.4\ \frac{\text{lb/h}}{\%})\ (\frac{\text{h}}{60\ \text{min}})} \quad 585\ \text{min}$$

9.13 Control Interval

The PID control results presented so far have been based on a continuous application of the controller. The digital application of feedback control is applied at discrete points in time, i.e., the controller is periodically called and the resulting control action implemented. Control computers use sequential microprocessors that perform control calculations for a large number of control loops. In the CPI, typical control loops are executed every 0.2 to 0.5 seconds for regulatory loops and 10 to 120 seconds for supervisory loops. In the bio-process industry, most loops are operated with a one second cycle time, unless circumstance require a faster cycle time. The time between control applications is the control interval, t. PID control is applied industrially on control computers using digital formulas, which are applied at discrete control intervals (Equations 7.5.2 to 7.5.5).

As a general rule[6], **the control interval should be selected such that**

$$\boldsymbol{t \quad 0.05\,(\ _p \qquad _p)} \qquad \textbf{9.13.1}$$

to obtain control performance approaching that of continuous control for which $_p$ and $_p$ are the FOPDT model parameters of the process. For t above this limit, the control performance deteriorates as t is increased. For t below this limit, no significant improvement in control performance results as t is decreased. Equation 9.13.1 as written indicates if control performance will not be improved by reducing t.

For feedback control using an on-line GC, the control interval is set by the cycle time for the analyzer updates (typically 3-10 minutes). No advantage is gained by applying control action more frequently than the GC updates because new information on the process response is available only when the GC updates. Based on Equation 9.13.1, for a 5 minute analyzer delay for an analyzer applied to a column with an open-loop time constant of 5 hours, the control performance is not improved by reducing the analyzer delay. On the other hand, for a 5 minute analyzer delay for an analyzer applied to a column with an open-loop time constant of 30 minutes, the control performance should improve if the analyzer delay were reduced. For sensors that provide continuous readings (e.g., temperature sensor), the maximum recommended control interval is typically equal to one sensor time constant. For level, temperature, pressure and flow loops, sensor dynamics do not usually present a significant constraint for the choice of the control interval.

Example 9.8 Estimation of the Control Interval

Problem Statement. Estimate the maximum control interval that can be used on the process studied in Example 6.10 without affecting the control performance.

Solution. From the results of Example 6.10, $_p$=2.43 s and $_p$=1.43 s. Applying Equation 9.13.1 yields t<0.193 s; therefore, if the control interval is less than 0.193 s, the control performance should be equivalent to the continuous application of feedback control.

9.14 Tuning Exercises

Because controller tuning requires lots of experience, tuning exercises based on simulations are an excellent method to develop your controller tuning skills. Several simulation methods are presented in this section for tuning PID loops and each method has its own unique advantages. The most versatile method is to simply add a controller to a MATLAB or Python simulation of a process (e.g., a model developed in Chapter 3). In this manner, you can make your model as detailed as necessary (e.g., including nonlinearity, sensor noise and nonstationary changes to the process). The downside to this approach is that it often requires a significant effort to develop the model.

When you have a transfer function representation of a process (e.g., a FOPDT model), you can use a Simulink model to implement feedback control using a convenient graphical user interface. As a result, it is much easier to choose a transfer function model and apply PI control with specific settings than to develop a full MATLAB model. Simulink can also be used with a nonlinear MATLAB model, but again the development time for the model can be significant. Therefore, Simulink is quite convenient as long as you are have a transfer function model of the process that you want to control. As a downside, Simulink does not conveniently handle derivative action so it is generally only used to implement P-only and PI control. Simulink is a fully developed modeling environment and as such has capabilities far beyond those demonstrated in this text. As a result, you can use Simulink to perform many additional tasks, but it will require special effort on your part.

From a user's standpoint, the visual basic simulator (VBS) that accompanies this text is the easiest simulation tool to use to develop PID controller tuning experience. The VBS uses several of the actuator, process and sensor models developed in Chapter 3 [i.e., the thermal mixer (Example 3.3), the composition mixer (Example 3.4), the level in a tank (Example 3.5) and the endothermic CSTR (Example 3.6)] as well as a FOPDT model and a model of a heat exchanger[7]. The VBS offers a point-and-click user interface; therefore, once you learn how to navigate using the visual basic interface, which is presented below, you are ready to use it without any programming requirements. The VBS offers open loops tests for changes in the MV or DV. Closed-loop tests for the three modes of the PID controller are available as well as feedforward control (Chapter 12). Sensor noise and sensor filtering can also be added to a simulation. After the features of a simulation and any tuning parameters are selected, the simulation can be executed producing MS Excel plots of each of the key variables. Of course, the VBS is limited to the models just mentioned although this collection of models does offer a range of process behavior.

The following subsections will show how each of these approaches can be used for tuning exercises along with examples of their application.

MATLAB Simulation. The challenge with implementing PID control using a MATLAB model is to calculate the integral of the error from setpoint and the derivative of the CV so that the PID control law can be used to determine the manipulated variable value. If the control law is evaluated in the main program, you would have to go into and come out of the ODE integrator at a high frequency in order to approximate continuous control and this formulation can be a bit tedious. On the other hand, if an ODE is added to the function that calculates the time derivatives of the dependent variables to calculate the integral of the error from setpoint, the PID control law can be applied directly in the same function. The derivative of the CV is also calculated in the ODE function and can be used by the PID control law. Following is an example of how a PID-type controller can be added to the model of a thermal mixing process (Example 3.3) and used to tune this controller.

Example 9.9 Tuning Example Using a MATLAB Model

Problem Statement. Tune a PI controller for a 1/6 decay ratio applied to the MATLAB model of the thermal mixing process (Example 3.3) using a setpoint change from 50°C to 55°C.

Solution. Listed below is the MATLAB code used to solve this example. Note that only changes made to the MATLAB program for Example 3.3 are denoted by bold and underlined text. In the main program, only the initial conditions for integral of the error from setpoint is added. In the ODE function (`dydt=f(t,y)`), the equation for the control is specified as well as the controller gain, reset time and setpoint. Following the controller data input, the position form of a PI controller is applied for times greater than or equal to 10 s. Finally, an ODE for the error from setpoint is added. Figure 9.14.1 shows the results for K_c=-1 and τ_I=30.

```
function Ex9_9Tmixer
clear; clc;
t0=0; tf=150; y0=[5;50;50;0];  % Input specs (y(1)-F1spec; y(2)-T; y(3)-Ts)
soln=ode45(@f,[t0 tf],y0);     % Call ode45 and store solution in soln
```

```
x=linspace(0,tf,100);         % Generate the values of t using linspace
y1=deval(soln,x,3);              % Retrive value of y from soln

plot(x,y1,'k-','LineWidth',2); % Plot t/y data and specify labels
xlabel('t'); ylabel('Outlet Temp (deg C)');
end
% User specified function for dydx for each dependent variable
function dydt=f(t,y)
tauv=2; F1spec=5; row=1; F2=5; T1=25; T2=75; tauTs=6; M=100;
F10=5; Kc=-1; tauI=30; Tsp=55; % Initial MV and controller settings
if t>=10; F1spec=F10+Kc*(Tsp-y(3))+Kc/tauI*y(4); end
dydt(1)=(F1spec-y(1))/tauv;
```

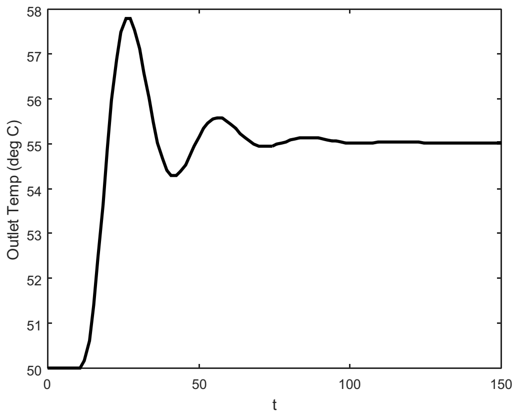


Figure 9.14.1 Results for Example 9.9.

```
dydt(2)=(y(1)*T1+F2*T2-(y(1)+F2)*y(2))/M;
dydt(3)=(y(2)-y(3))/tauTs;
dydt(4)=Tsp-y(3);
dydt=dydt'                       % Return dydt as column vector

end
```

Python Simulation. The challenge with implementing PID control using a Python model is to calculate the integral of the error from setpoint and the derivative of the CV so that the PID control law can be used to determine the manipulated variable value. If the control law is evaluated in the main program, you would have to go into and come out of the ODE integrator at a high frequency in order to approximate continuous control and this formulation can be a bit tedious. On the other hand, if an ODE is added to the function that calculates the time derivatives of the dependent variables to calculate the integral of the error from setpoint, the PID control law can be applied directly in the same function. The derivative of the CV is also calculated in the ODE function and can be used by the PID control law. Following is an example of how a PID controller can be added to the model of a thermal mixing process (Example 3.3) and used to tune this controller.

Example 9.10 Tuning Example Using a Python Model

Problem Statement. Tune a PI controller for a 1/6 decay ratio applied to the Python model of the thermal mixing process (Example 3.3) using a setpoint change from 50°C to 55°C.

Solution. Listed below is the Python code used to solve this example. Note that only changes made to the Python

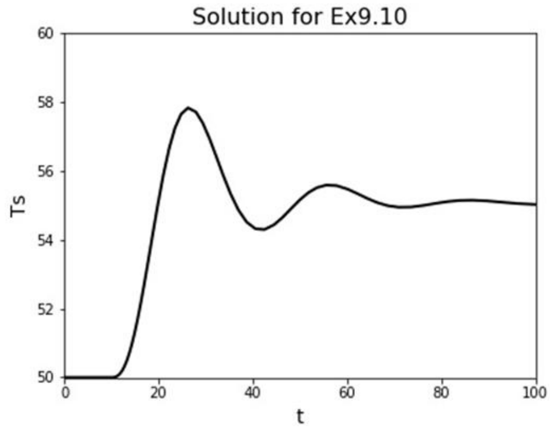


Figure 9.14.2 Results for Example 9.10.

program for Example 3.3 are denoted by bold and underlined text. In the main program, only the initial conditions for integral of the error from setpoint is added. In the user-defined function fun, the equation for the control is specified as well as the controller gain, reset time and setpoint. Following the controller data input, the position form of a PI controller is applied for times greater than or equal to 10 s. Finally, an ODE for the error from setpoint is added. Figure 9.14.2 shows the results for K_c=-1 and $_I$=30.

```
import scipy.integrate
import numpy as np
import matplotlib.pyplot as plt
def fun(t, y):
  tauv, F1spec, F2, T1, T2, tauTs, M = 2, 5, 5, 25, 75, 6, 100
  F10, Kc, tauI, Tsp = 5, -1, 30, 55  # Specify controller parameters
  if t>= 10:
     F1spec=F10+Kc*(Tsp-y[2])+Kc/tauI*y[3]
  dy1=(F1spec-y[0])/tauv
   dy2=(y[0]*T1+F2*T2-(y[0]+F2)*y[1])/M
   dy3=(y[1]-y[2])/tauTs
   dy4=Tsp-y[2]
   return [dy1, dy2, dy3, dy4]
tspan = [0, 100]
y0=[5, 50, 50, 0]
tvalues=np.linspace(0, 1, 100)
soln=scipy.integrate.solve_ivp(fun, tspan, y0, method='RK45', t_value=tvalues, rtol=1.E-6, atol=1.E-6)
time=soln.t
y3=soln.y[2]
plt.figure(figsize=(6,4.5))
```

```
plt.plot(time, y3, 'k-', linewidth=2)
plt.axis([0, 100, 50, 60])
plt.title('Solution for Ex9.3', fontsize=16)
plt.xlabel('t', fontsize=14)
plt.ylabel('Ts', fontsize=14)
plt.show
```

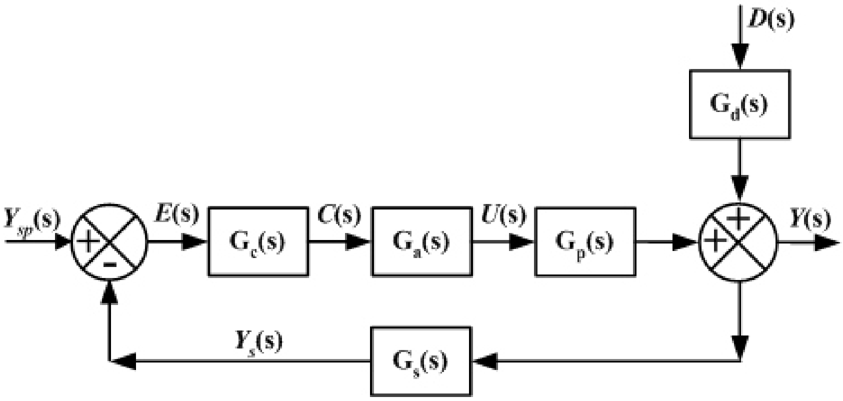


Figure 9.14.3 Block diagram of a general feedback control loop.

Simulink. Simulink can be used to represent a feedback control system by using the general feedback control loop block diagram shown in Figure 9.14.3, which was introduce earlier in Chapter 7 (Figure 7.2.1):

That is, by using input blocks to represent a step change in the setpoint [$Y_{sp}(s)$] and the disturbance [$D(s)$]; transfer function blocks for $G_c(s)$,$G_a(s)$,$G_p(s)$,$G_s(s)$ and $G_d(s)$; and Sum blocks for the commutator, which compares the measured value of the CV with the setpoint, and the addition of the effect of the disturbance with the output of the process, Simulink can be used to model a feedback control loop for setpoint changes and/or disturbance upsets.

Example 9.11 Tuning Example Using Simulink

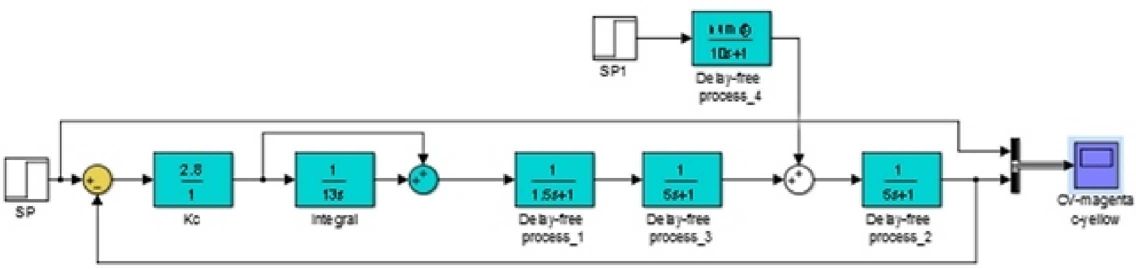


Figure 9.14.4 Simulink model for Example 9.11

Problem Statement. Tune a PI controller using a Simulink model of a process with an actuator time constant of 1.5 s, a first-order process with a time constant of 5s and a gain of 1.0 and a sensor with a time constant of 5 s. After the PI controller is tuned for a decay ratio of 1/7, test it for a unit disturbance test with a disturbance model with a gain of 0.5 and a time constant of 5 s.

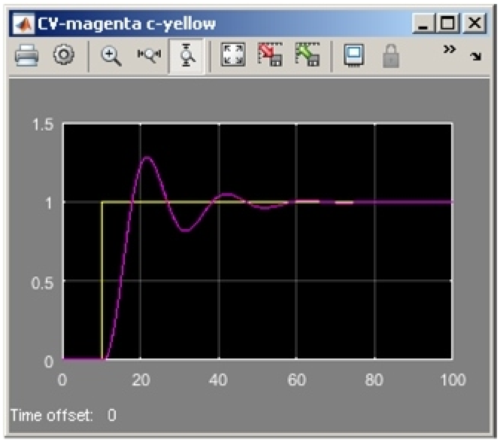


Figure 9.14.5 Setpoint tracking results.

Solution. In order to apply PI control, we will use the block diagram for a PI controller shown in Figure 5.6.5 with $_D$ set equal to zero. A Simulink model can be formed directly from Figure 9.14.3 using the block diagram for a PI controller. The Simulink model for this case is shown in Figure 9.14.4. Note that the effect of the MV is combined with the effect of the disturbance before implementing the sensor lag. The controller setpoint c_{sp} is plotted by the Scope block along with the CV value. For the first part of the problem, the final value for the input block for the setpoint is set equal to unity at t=10 and the final value for the input block for the disturbance is set equal to zero. The values of K_c and $_I$ are were adjusted until the desired response to the setpoint change was attained, which resulted in using K_c=2.8 and t_I =13. The results of the tuned PI controller for the setpoint change are shown in Figure 9.14.4. Note that the CV starts at the point (0,0).

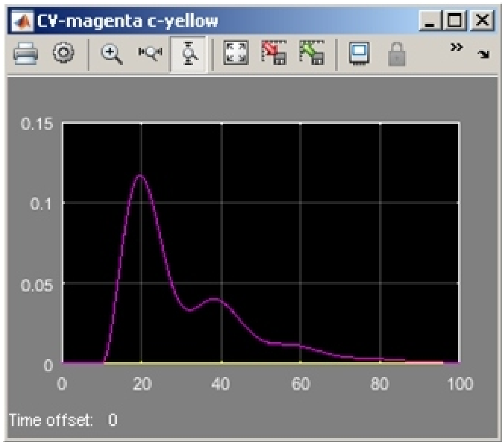


Figure 9.14.6 Disturbance rejection results.

Next, the final value for the setpoint input block is set equal to zero and the final value for the disturbance is set equal to unity at t=10. The results for the disturbance test are shown in Figure 9.14.6. Note that the CV settles to zero.

Example 9.12 Tuning Example Using Simulink

Problem Statement. Tune a PI controller for a decay ratio of 1/6 applied to a FOPDT process (K_p=0.5; $_p$=6; $_p$=2).

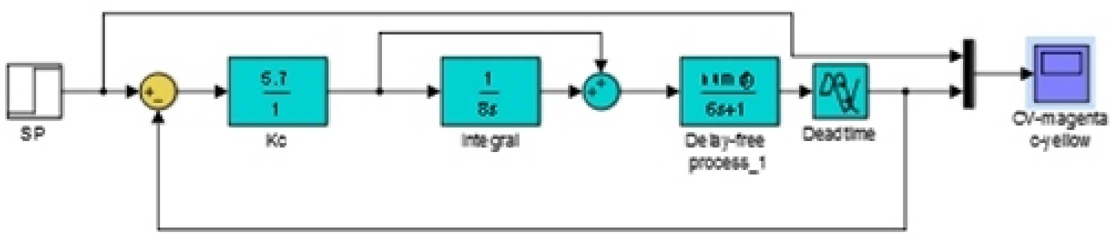


Figure 9.14.7 The Simulink model for Example 9.16.

Solution. Figure 9.14.7 shows the Simulink model for this case. The block diagram for a PI controller shown in Figure 5.6.5 with $_D$ set equal to zero was used to implement the PI controller and a first-order model plus a dead-time element was used to represent the FOPDT process model. The CV and the controller output are shown in Figure 9.14.8 for the PI tuning parameters (K_c=5.7; $_I$=8) that resulted in 1/6 decay ratio for the setpoint change implemented at t=10.

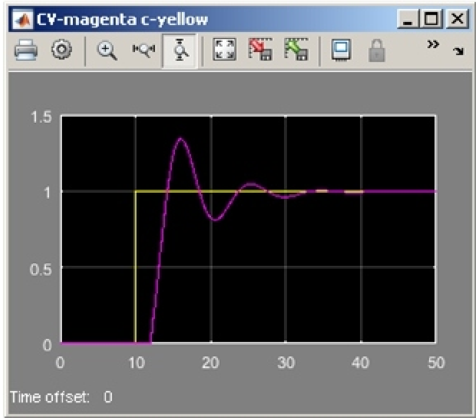


Figure 9.14.8 Setpoint change results for Example 9.16.

Example 9.13 Tuning the CST Thermal Mixing Process

Problem Statement. Use Simulink to develop tuning parameters for the CST thermal mixing process (Example 3.3) using a 10% setpoint change in the outlet temperature.

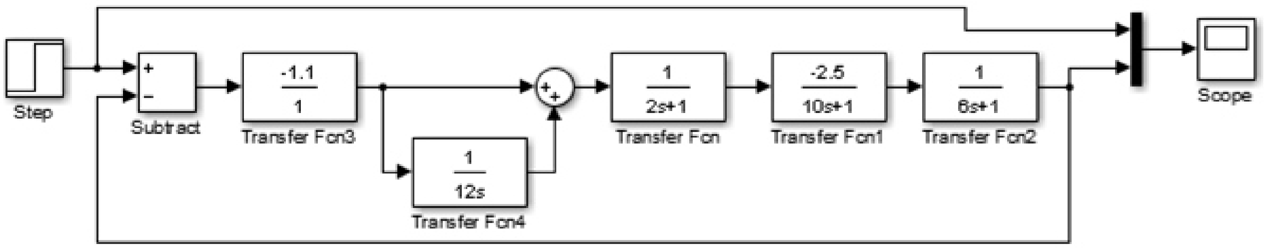


Figure 9.14.9 The Simulink model for Example 9.13.

Solution. The Simulink model for the open loop behavior for the CST thermal mixer (Example 6.15) can be modified for this example. That is, by adding the block diagram components for a PI controller and adding a recycle of the sensor measurement, the open-loop Simulink model can be converted into a closed loop model as shown in Figure 9.14.9. By applying the field tuning procedure, this Simulink model provided the results shown in Figure 9.14.10 using K_c=-1.1 and $_I$=12 s .

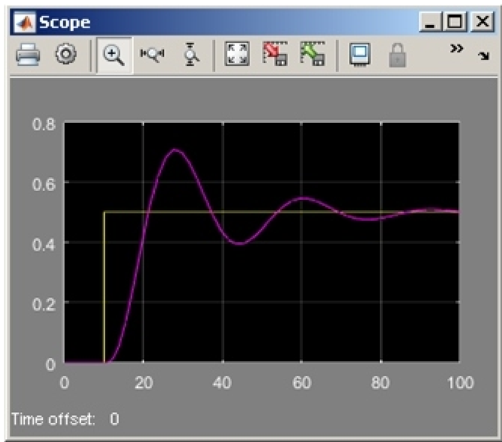


Figure 9.14.10 Results for Ex9.14.

Visual Basic Simulator. In order to use the VBS for tuning exercises, first select the model that you want to use: CST thermal mixer, CST composition mixer, level in a tank, CSTR reactor, the heat exchanger or the FOPDT process. Then, select a setpoint change, choose the type of controller (P-only, PI or PID) and set the controller tuning parameters. Finally, click the "Run Simulation" button and the results are provided as Excel plots. The additional details for using the VBS are provided in Section 3.9.

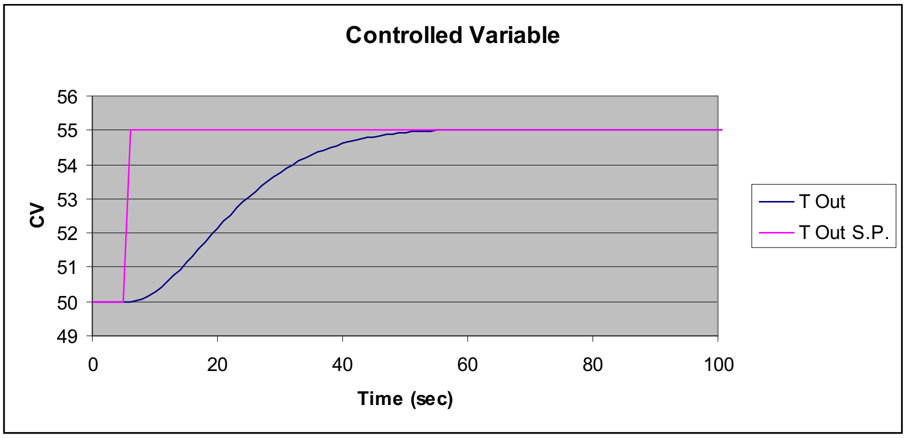


Figure 9.14.11 The setpoint and CV for Example 9.17.

Example 9.14 Tuning Example Using the VBS

Problem Statement. Tune a PI controller using the VBS for the TMIXER simulation for a critically damped response for a 10% increase in the CV setpoint using the field tuning procedure.

Solution. First, a P-only controller is tuned for non-oscillatory behavior resulting in a controller gain of 0.02. Then a PI controller is applied with this controller gain minus 10% (i.e., K_c=0.018) and the integral time is decreased until overshoot is observed using $_I$=12 s. Then the integral time is decreased slightly to 11 and critically damped offset-free response results (Figure 9.14.11).

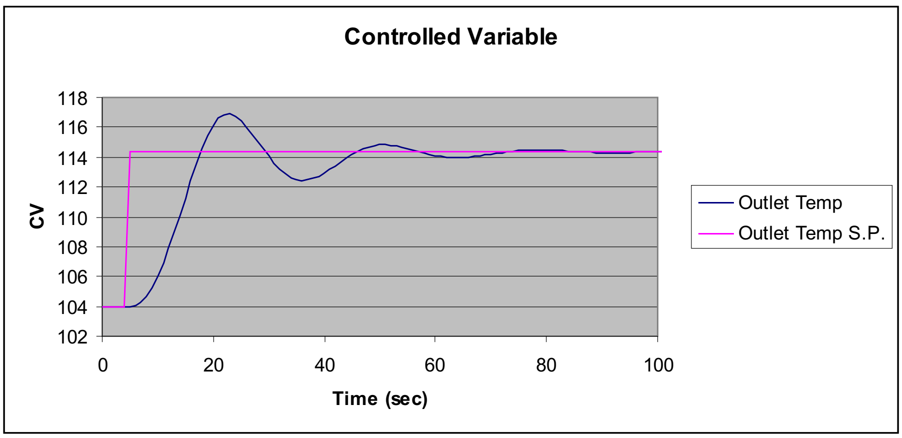


Figure 9.14.12 Setpoint and CV for Example 9.18.

Example 9.15 Tuning Example Using the VBS

Problem Statement. Tune a PI controller using the VBS for the HEATEX simulation for a 1/6 decay ratio for a 10% increase in the CV setpoint using the ATV-based tuning procedure.

Solution. First, an ATV test is run using a 10% relay height. By measuring the peak to valley separation and dividing the result by 2, the amplitude of the oscillation was determined to be 0.96^{o}C. The time between successive peaks or valley was 22 s, which is the period of oscillations. The relay height for the MV for the ATV test was 24^{o}C. From these results, K_u=31.8. Using the Ziegler-Nichols settings, K_c=14.3/F_T and τ_I=18.3F_T. For F_T=1, the specified decay ratio was obtained and the results are shown in Figure 9.14.12.

9.15 Summary

- One of the major responsibilities of a process control engineer is PID controller tuning, which is generally a compromise between controller performance and controller reliability with consideration given to the operational objectives of the process. Control performance is measured by the error from setpoint while reliability is determined by the severity of the disturbances that a controller is able to absorb effectively without becoming unstable.
- Classical tuning methods require a FOPDT process model or an ultimate test and are based on preset tuning criterion.
- Pole placement methods allow you to tune a controller for preset performance specifications, but require a process model and can result in a non-standard controller (i.e., something other than a PID controller).
- IMC-based PID tuning is convenient when a process model is available.
- The control engineer should consider the combined effect of process nonlinearity, the severity of disturbance and the operational objectives of the process when selecting the tuning criterion (e.g. 1/6 decay ratio to overdamped in the extremes) for a particular control loop.
- Sensor filtering reduces the effects of senor noise on the feedback control action, but it can increase the overall lag of the system. For properly functioning, temperature, flow, level and pressure sensors, a filter time constant of 3-5 seconds is usually adequate to remove sensor noise and does not significantly add lag to the system.
- The recommended approach for PID tuning for important loops involves the following steps. 1. Select the tuning criterion. 2. Apply filtering to the sensor reading. 3. Determine if the control loop is fast responding or slow responding. 4. For fast-responding loops, apply field tuning. 5. For slow-responding loops, apply ATV-based tuning. 6. Apply PID control for sluggish processes by adding derivative action to a tuned PI controller.
- The recommended approach for tuning less-important loops (e.g., most level and pressure loops and many flow control loops) is to use the simplified field tuning method and tune for an overdamped response.
- For a P-only controller applied to a slow responding level for critically damped performance,

$$K_c = \frac{F_{MAX}}{L_{\max}} \quad \text{P only control}$$

where F_{MAX} is the maximum expected step change in the feed rate to the tank, and L_{MAX} is the desired level change that F_{MAX} should cause under feedback conditions.

- If the control interval, t, satisfies $t \quad 0.05\,(\; _p \quad \; _p)$, effectively continuous control performance results.
- Tuning exercises can be performed using MATLAB models, Simulink models and the visual basic simulator that goes with this text.

9.16 References

1. Alford, J., Rhinehart, R, and Hrankowski, B., "Process Data Filtering", Intech, July- Aug., 2018

2. Box, G.E.P and G.M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, 1970.

3. Astrom, K.J. and T. Hagglund, *Automatic Tuning of PID Controllers,* Instrument Society of America, p. 233 (1998).

4. Tyreus, B.D. and W.L. Luyben, “Tuning PI Controllers for Integrator/Deadtime Processes”, *Ind. Eng. Chem. Res.*, Vol. 31, p. 2625 (1992).

5. Marlin, T.E., *Process Control*, McGraw-Hill, p. 588-590 (1995).

6. Marlin, T.E., *Process Control*, McGraw-Hill, p. 393 (1995).

7. Riggs, J.B. and M.N. Karim, *Chemical and Bio-Process Control*, 3rd Ed., Ferret Publishing, Austin, TX (2006).

9.17 Additional Terminology

ATV - autotune variation; a relay feedback experiment designed to measure the ultimate controller gain and ultimate period of a control loop.

Controller performance - a measure related to the error from setpoint.

Controller reliability - a measure of how well a controller stays in service. It can be quantified by the maximum severity of a disturbance that a control loop can handle and remain in service.

F_T - the on-line tuning factor for PI controller tuning.

Reliability - the ability of a controlled process to remain stable when subjected to severe disturbances.

9.18 Preliminary Questions

9.3 Tuning Criteria and Performance Assessment

Q9.3.1 How is using a minimum IAE tuning criterion different from using a minimum ISE tuning criterion?

Q9.3.2 Explain what a statistical process control chart is and how it can be used to assess control performance.

Q9.3.3 How do many companies track the variability of the products that they produce?

Q9.3.4 From a process control point of view, why is it better to use the standard deviation from the setpoint instead of the standard deviation as a measure of variability for a controller?

Q9.3.5 What factors determine the tuning criterion for a control loop?

Q9.3.6 From Table 9.1, it is clear that QAD tuning provides the best performance. Why are industrial control loops not normally tuned for QAD behavior?

9.4 Controller Reliability

9.4.1 Why does a linear process model not explain the controller reliability limitations of industrial processes?

9.4.2 How does process nonlinearity affect the dynamic character of the closed-loop response?

9.4.3 How do disturbances affect the dynamic character of the closed-loop response?

9.4.4 If a control loop exhibits ringing and, at other times, sluggish behavior with the same controller settings, what process characteristic does this indicate?

9.6 Filtering Sensor Readings

9.6.1 Explain how an increase in the noise level for a sensor can cause an increase in the variability produced by the control loop using the sensor.

9.6.2 What is the purpose of a digital filter? How do you decide how much filtering is appropriate?

9.6.3 Explain why you cannot simply add enough filtering to remove all the noise from a noisy measurement of the CV before applying derivative action.

9.6.4 Why is filtering of the sensor reading applied before tuning of a PID controller is started?

9.6.5 How is the filter on a sensor reading tuned?

9.8 Tuning Fast-Responding Control Loops

9.10.1 What determines whether a process is fast responding or slow responding?

9.10.2 What tuning procedure should be applied to fast-responding processes?

9.9 Tuning Slow-Responding Control Loops

9.11.1 What is an ATV test and how is it used to tune a controller?

9.11.2 Why should you use an ATV test for tuning a slow-responding process?

9.10 PID Tuning

9.12.1 How does the PI-control tuning factor, F_T, affect the closed-loop dynamics?

9.12 Tuning Level Controllers

9.13.1 What information is required to choose the P-only controller setting for a level controller using Equation 9.13.1?

9.13.2 What information is required to choose the PI controller setting for a level controller using Equation 9.13.2?

9.13 Control Interval

9.14.1 What happens if Equation 9.14.1 is not satisfied?

9.19 Analytical Questions and Exercises

9.3 Tuning Criteria and Performance Assessment

P9.3.1^{S}** Using the simulator for the heat exchanger that comes with this textbook, test a PI controller for a +20% and -20% disturbance change and characterize the dynamic behavior using the decay ratio for each if possible. Use the following controller settings: K_c 10; $_I$ 14. As a point of reference, determine the decay ratio for a +10% setpoint change and a -10% setpoint change. Compare these results. What can you conclude?

9.6 Filtering Sensor Readings

P9.6.1[S]* For the thermal mixing tank simulator provided with this text, assume that the repeatability of the temperature sensor on the product stream is equal to 0.4ºC (i.e., =0.2ºC). Using a 10% increase in the MV, plot the filtered value of the sensor reading versus time for f=0.5, f=0.1 and f=0.01.

P9.6.2* Consider the application of an orifice meter and a magnetic flow meter for the measurement of the flow rate in a flow control loop. Assuming that the sensor filter is applied every 0.1 s, determine the filter time constant for these two flow sensors if it is desired to attain a filtered value of the CV with a repeatability equal to ±0.05%. Assume that the repeatability of the unfiltered orifice meter reading is ±0.8% and the repeatability of the unfiltered magnetic flow meter reading is ±0.1%.

P9.6.3*** Consider a first-order process with a process gain of 0.5 and a time constant of 15 and a sensor with a time constant equal to 6. If the cycle time for the sensor filter is equal to 0.1, determine how much the repeatability of the sensor can be reduced without increasing the effective time constant of the closed-loop system by more than 30%. Neglect the dynamics of the actuator.

P9.6.4** Determine the dynamic behavior of a P-only controller with K_c equal to 0.1 applied to a first-order process with a process gain of 3 and a time constant of 200. Assume that a first-order filter is applied to the sensor reading and uses a control interval of 1. Assume that $G_s(s)$ and $G_a(s)$ are equal to unity. The sensor reading is quite noisy and it is desired to reduce the repeatability of the noisy sensor by a factor of 10.

P9.6.5*** Consider the control of the outlet temperature of the cold stream from a double-pipe heat exchanger shown in Figure P7.51. The flow rate of the hot stream (F_h) is manipulated to control the outlet temperature of the cold stream ($T_{c,out}$). The overall process model for this system is given by

$$\frac{T_{c,out}(s)}{F_h(s)} = \frac{0.5e^{-5s}}{(20s+1)(5s+1)}$$

Assume the first-order Padé approximation to represent the deadtime term, i.e.,

$$e^{-\theta_p s} \approx \frac{1 - \frac{1}{2}\theta_p s}{1 + \frac{1}{2}\theta_p s}$$

1. For a P-only controller with a gain K_c, will this process become unstable at certain value of K_c? Show the details of your analysis.

2. What happens to the stability limits for this system if a sensor filter with a filter time constant equal to 5 time units is added?

9.8 Tuning Fast-Responding Control Loops

P9.8.1[S]** For the thermal mixing simulator provided with this text, tune the temperature controller for this process for a decay ratio of 1/8 using the field tuning approach using a +10% setpoint change. Assume that there is no noise on the sensor reading.

P9.8.2[S]** For the CSTR simulator provided with this text, tune the temperature controller for this process for a decay ratio of 1/6 using ATV-based tuning approach for a +1.428% setpoint change and test the controller for a +10% disturbance upset. Assume that there is no noise on the sensor reading.

P9.8.3[S]** For the level process simulator provided with the text, tune a P-only controller. With the same controller gain, add integral action in increments. What can you conclude?

9.9 Tuning Slow-Responding Control Loops

P9.9.1^S** For the heat exchanger simulation provided with this text, tune the temperature controller for this process for a critically damped response, a decay ratio of 1/10 and a decay ratio of 1/6 using ATV-based tuning for a 10% setpoint increase. For each tuning case, test with the same disturbance upset (i.e., a +10% increase in the disturbance). Assume that there is no noise on the sensor.

P9.9.2^S*** For the CSTR simulation provided with this text, tune the temperature controller for this process for a critically damped response, a decay ratio of 1/10 and a decay ratio of 1/6 using ATV-based tuning using a 1.5% increase in the setpoint. For each tuning case, test with the same disturbance upset (i.e., a +10% increase in the disturbance). Assume that the sensor has a repeatability of ±0.5ºC. Use enough sensor filtering to reduce the repeatability of the sensor by a factor of 10.

9.10 PID Tuning

P9.10.1^S** For the heat exchanger process simulator provided with the text, tune a PI and a PID controller for a 1/8 decay ratio* using small amplitude setpoint changes. Then test both controllers for a step disturbance upset and compare the results.

9.12 Tuning Level Controllers

P9.12.1** Calculate the initial PI controller settings for a level controller with a critically damped response for a 10 ft diameter tank (i.e., a cylinder placed on its end) with a measured height of 10 ft that normally handles a feed rate of 1000 lb/h. Assume that it is desired to have a maximum level change of 5% for a 20% feed rate change and that the liquid has a density corresponding to that of water.

9.13 Control Interval

P9.13.1* Consider a FOPDT process (i.e., K_p=4, $_p$=10, $_p$ 4.5). What is the maximum size of the control interval that should be used for this case without the control performance deteriorating? What happens if a larger or a smaller control interval is used?

9.14 Tuning Exercises

MATLAB or Python

P9.14.1 Tune a PI controller applied to the MATLAB or Python model of the CSTR (Example 3.6) using a setpoint change from 350K to 360 for a critically damped response using the field tuning procedure. Once tuned, test the controller for an increase in the feed composition from 1 to 1.1 gmol/l.

P9.14.2 Tune a PI controller applied to the MATLAB or Python model of the CSTR (Example 3.6) using a setpoint change from 350K to 360 for a 1/14 decay ratio using the field tuning procedure. Once tuned, test the controller for an increase in the feed composition from 1 to 1.1 gmol/l.

Simulink

P9.14.3 Using Simulink, tune a PI controller for a unit setpoint change applied to a FOPDT process (K_p=3; $_p$=10; $_p$=1) for a 1/6 decay ratio using the field tuning procedure.

P9.14.4 Using Simulink, tune a PI controller for a unit setpoint change applied to a FOPDT process (K_p=0.2; $_p$=1; $_p$=0.1) for a critically damped response using the field tuning procedure.

P9.14.5 Using Simulink, tune a PI controller for a unit setpoint change applied to a FOPDT process (K_p=12; $_p$=3; $_p$=0.5) for a 1/6 decay ratio using the field tuning procedure.

P9.14.6 Using Simulink, tune a PI controller for a unit setpoint change applied to a FOPDT process (K_p=0.3; $_p$=10; $_p$=6) for a critically damped response using the field tuning procedure.

P9.14.7 Using Simulink, tune a PI controller for a unit setpoint change applied to a system with an actuator with a 5s time constant, a sensor with a 10 s time constant and a first-order process with a gain of 2.5 and a time constant of 12 s for a 1/6 decay ratio using the field tuning procedure.

P9.14.8 Using Simulink, tune a PI controller for a unit setpoint change applied to a system with an actuator with a 3s time constant, a sensor with a 15 s time constant, a sensor filter with a 5 s time constant and a first-order process with a gain of 12.5 and a time constant of 20 s for a 1/6 decay ratio using the field tuning procedure.

P9.14.9 Using Simulink, tune a PI controller for a unit setpoint change applied to a system with an actuator with a 2s time constant, a sensor with a 1 s time constant and a first-order process with a gain of 0.4 and a time constant of 8 s for a critically damped response using the field tuning procedure.

P9.14.10 Using Simulink, tune a PI controller for a unit setpoint change applied to a system with an actuator with a 3s time constant, a sensor with a 6 s time constant and a first-order process with a gain of 6 and a time constant of 6 s for a 1/8 decay ratio using the field tuning procedure.

A problem number with a **superscript S** denotes that the problem requires the use of the visual basic simulators.

Chapter 10

Troubleshooting Control Loops

Chapter Objectives

- Present a systematic approach to troubleshooting control loops in the CPI.
- Demonstrate how a block sine wave test can be used to evaluate the performance of an actuator.
- Lists the most common failure modes for the control computer, actuator and sensor systems.
- Present an approach for analyzing the performance of the entire control loop.
- Discuss troubleshooting the control systems for bio-reactors.

10.1 Introduction

Chemical and bio-process control engineers spend a major portion of their time troubleshooting control loops. An operator may point out that a particular loop has been behaving erratically and ask the control engineer to improve the performance of the control loop. The control engineer may discover that an important control loop is under manual operation (i.e., open-loop operation). A final product may have excessive variability in its impurity levels and the control engineer's job is to reduce the variability to an acceptable level. In this latter example, a number of control loops may require scrutiny. For bio-processes, certain batches can perform poorly due to poorly performing control loops. When one or more loops are not performing properly, troubleshooting is required to return them to the expected performance levels or at least identify the source of the problem. To effectively troubleshoot control loops, the control engineer must understand the proper design and expected performance of the hardware that comprise a control loop, which are addressed in Chapter 2.

Troubleshooting control loops involves identifying the source of the problem with a control loop from an overwhelming number of possible causes. The size of this problem requires a systematic approach when troubleshooting. Control loop troubleshooting is too often treated as an afterthought and performed haphazardly. This chapter presents a general troubleshooting procedure as well as a detailed analysis of fault detection for the final control element, the sensor system, the control computer or DCS and the process.

10.2 Overall Approach to Troubleshooting Control Loops

The key to effective troubleshooting is expressed in the old adage, “divide and conquer”. It is important to locate the portion of the control-loop hardware that is causing the poor performance: the final control element, the sensor system, the controller or the process. The place to start is to test each system separately to determine whether that portion of the control loop is operating properly. The final control element can be evaluated by applying a series of input step tests. That is, the input to the final control element, which is normally set by the controller, can be manually adjusted. This test allows the determination of the dynamic response and deadband of the actuator system. If the performance in these two areas is satisfactory, there is no need to evaluate the actuator system further. Section 10.3 addresses troubleshooting actuator systems and contains a listing of common failure modes.

The sensor system is more difficult to check than the actuator system. Using the DCS or control computer to track the measured value of the CV can be helpful to point out certain abnormalities. The repeatability of the sensor can be estimated during a steady-state period. Determining the time constant for the sensor dynamics is usually much more difficult. A time history of the actual control variable, which may not be available, must be compared to the time history of the sensor reading to estimate the sensor time constant. The sensor may also require calibration. Section 10.3 addresses troubleshooting sensor systems and contains a listing of common failure modes for a number of sensor types.

Controllers are implemented by the DCS or control computers and affect the control performance through the controller tuning, the control interval and sensor filtering. The controller tuning is very often the source of erratic control performance and can be corrected by simply retuning the control loop. Controller tuning is the easiest thing to change, but retuning the controller may mask the real problem with the control loop. When a component of the feedback system is not functioning properly and the controller is retuned, erratic behavior may not be present, but a sacrifice in control performance can result. Controller design and implementation issues can be the source of poor control loop performance, e.g., using the wrong mode of the PID controller, using a reverse-acting controller for a system that requires a direct-acting controller and having the controller output go to the wrong MV. Section 10.3 addresses troubleshooting the controller and contains a listing of common problems.

The process should also be evaluated with regard to its effect on controller performance. Excessive disturbances entering the process can be reduced by modifying upstream operations. Fouling or mechanical failure can create a situation in which it is not possible for the controller to maintain the CV at its setpoint. The control of the process may not be satisfactory because an unidentified constraint prevents normal control of the process.

Sudden changes from satisfactory to unacceptable performance for a control loop may be caused by recent changes in the process. Considering what changes have been made to the process can expedite the troubleshooting process. For example, the controller could have been retuned. A new analyzer could have been installed. A new instrument technician could be responsible for calibrating and maintaining an analyzer. The feed to the unit could have significantly changed. These examples and many more can be directly related to the source of poor performance of a control loop. When something significant has changed, it can provide a valuable clue that allows quicker determination of the source of the problem with a poorly performing control loop. Section 10.3 addresses troubleshooting the process to identify process problems or limitations affecting control loop performance.

Finally, the entire loop should be tested under closed-loop conditions. First, the **closed-loop deadband** should be determined, i.e., the largest positive and negative setpoint changes that can be implemented without causing measurable changes in the CV. The dynamic response of the closed-loop process can also be assessed from the results of the closed-loop deadband test. Section 10.3 addresses evaluating the closed-loop performance of a control loop.

When troubleshooting a more complex control system, it is advisable to start by comparing the existing control loops with those from the P&ID. The current problem may be due to inappropriate modifications of the original control configuration.

Bio-systems require a somewhat different approach for troubleshooting than control loops in the CPI. Bio-reactors are the primary operation concern because if the bio-reactor is not operated properly there will not be any product to purify. While troubleshooting in the CPI largely involves checking individual control loops, for bio-reactors troubleshooting requires a holistic approach, i.e., the entire system must be considered. Section 10.4 discusses troubleshooting control systems on bio-processes.

Self-Assessment Questions.

Q10.2.1 What is troubleshooting control loops?

Q10.2.2 Because there are a very large number of possible faults that can undermine the performance of a control loop, how does the control engineer check each of these potential faults?

Self-Assessment Answers

Q10.2.1 Troubleshooting control loops involves identifying the source of the problem with a control loop from an overwhelming number of possible causes.

Q10.2.2 Because there are a very large number of possible faults that can undermine the performance of a control loop, it is important for a control engineer to locate the portion of the control-loop hardware that is causing the poor performance: the final control element, the sensor system, the controller or the process. Hence, the control engineer tests each system separately to determine whether that portion of the control loop is operating properly. That is, divide and conquer.

10.3 Troubleshooting Control Loops in the CPI

Actuator Systems. The final control element consists of the instrument air system, the I/P converter and the control valve (the valve body and the valve actuator). The fastest way to identify gross problems with the final control element is to plot both the MV and the controller output. If the MV does not follow the controller output, there is probably a problem with the final control element.

Even if the MV seems to follow the controller output, there could be a problem with the actuator. Estimates of the actuator deadband and dynamic response are required to determine if the actuator system is performing properly (both of which can be determined by a **block sine wave** test). A block sine wave test, which is a series of step changes applied to approximate one cycle of a sin wave, is shown in Figure 10.3.1. For the test shown in Figure 10.3.2, initially the amplitude of the step change used in the block sine wave is small enough that consistent positive and negative changes in the measured value of the flow rate of the MV are not observed. The next block sine wave uses a larger amplitude step change, and for

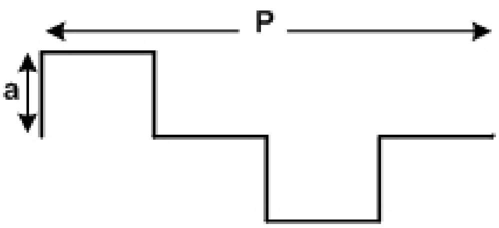


Figure 10.3.1 A block sine wave with an amplitude, *a* and a period, P.

this case, the measured MV can be seen to make both positive and negative changes corresponding to the positive and negative changes in the controller output. Therefore, the deadband of the actuator in this case is larger than the step size used in the first block sine wave and smaller than the one used in the second. The same block sine wave test can be used to estimate the time constant of the final control element. If the time between step changes used in the block sine wave is large enough, the settling time of the actuator can be estimated. Then, the time constant of the actuator is estimated as one-quarter of the settling time.

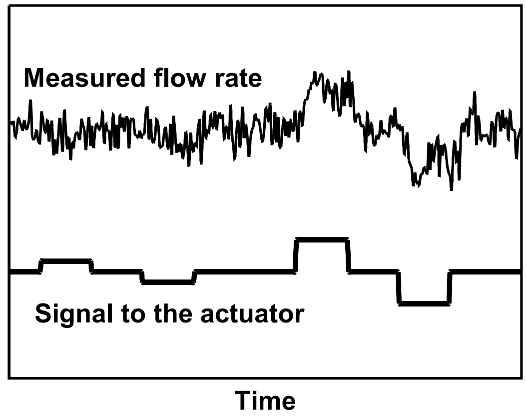


Figure 10.3.2 Graphical representation of a block sine wave test applied to an actuator system

Once the deadband and time constant of the actuator have been determined, the performance of the actuator can be assessed. The deadband for valves with positioners typically ranges from 0.1 to 0.5 % of the flow rate for properly implemented systems and depends on the size of the valves, the pressure drop across the valve, the fluid properties, etc. Deadband for an industrial valve without a positioner typically ranges from 10 to 25% and even higher for older valves that have not been maintained. The time constant of a properly functioning final control element is less than 2 s for a valve with a positioner or a control valve in a flow control loop. Otherwise, the actuator time constant is usually between 3 and 15 s.

The easiest way to check an actuator system is to use a block sine wave test with an amplitude equal to 0.005% and a hold time sufficient to observe the full change. Then, if the measured flow rate is observed to have positive and negative changes about the nominal value and the response time for these changes is less than 8s, the actuator is performing to specification. If not, further investigation is required to identify the problem.

If you determine that the actuator system is not functioning properly, you should first determine if the instrument air pressure system is operating properly. This can be done by observing the instrument air pressure at the control valve after a step change in the signal to the final control element has been implemented in the DCS or in the control computer. If the instrument air pressure at the control valve increases sharply after the step test has been implemented, then the control valve is the source of the slow or erratic response. Another common problem is valve packing that is over tightened, which primarily increases the valve deadband. A valve that is operating below 10% or above 90% opening typically performs below standards and indicates improper valve sizing. Another problem is an improperly tuned valve positioner. If the valve positioner is tuned too aggressively, oscillatory control performance results for the control valve, resulting in an increase in the actuator deadband. If the valve positioner is not tuned aggressively enough, the dynamic response of the actuator is slower than it should be. Table 10.1 lists a number of common problems with the components of the actuator system.

Table 10.1 Common Problems with the Final Control Element
• Excessive lag in the instrument air system. • Wrong type of instrument air connected to control valve. Some plants have high and low pressure instrument air. • Low instrument air pressure. • Wet or dirty instrument air. • Excessive deadband due to stiction or hysteresis.* • Improperly sized control valve*. • Excessive resistance to movement of valve stem*. • Leak in diaphragm of control valve. • Debris stuck in opening to control valve. • Plugged or obstructed instrument air line. • Plug/seat erosion in the control valve. • By-pass line open or leaking. • Flashing and cavitation. • Improperly tuned valve positioner*. * More frequently observed problems

Self-Assessment Questions.

Q10.3.1 How is the deadband of a final control element determined from a block sine wave test?

Q10.3.2 How is a block sine wave test used to determine the time constant of the actuator system?

Q10.3.3 What are the most frequently observed problems with the actuator systems?

Self-Assessment Answers

Q10.3.1 A block wave signal is applied to actuator system such that initially the amplitude of the step change used is small enough that consistent positive and negative changes in the measured value of MV are not observed. The next block sine wave uses a larger amplitude change, and the measured MV can be seen to make both positive and negative changes corresponding to the positive and negative changes in the controller output. From this observation the deadband is determined.

Q10.3.2 If the step changes used for the block sine wave test are held long enough for a complete process response, the time constant for the actuator system can be estimated.

Q10.3.3 The most frequently observed problems with actuator systems: (1) Excessive valve deadband which can result from excessively tightened packing [excessive resistance to movement of the stem] or a worn valve. (2) An improperly sized control valve. (3) An improperly tuned valve positioner.

Sensor Systems. The sensor system is composed of the sensor, the transmitter and the sampling system that allows the sensor to make its measurement. The performance of a sensor can be assessed by determining its repeatability and time constant and sometimes its accuracy. Accuracy is important for a composition analyzer on a final product to ensure that the product meets its specifications. In certain cases, the accuracy of a flow transmitter is usually not important because the flow rate is adjusted incrementally by a supervisory controller so that its actual flow rate is unimportant. On the other hand, when a flow rate is used to calculate a characteristic of the process (e.g., computed MV control (Section 13.5) and the calculation of oxygen demand for a fermentation reactor), an accurate flow measurement is important. The dynamics of the sensor can affect the feedback control performance if it is too slow, and a poor repeatability can increase the variability in the CV.

The accuracy of the sensor can be checked by comparing the sensor reading to a standard or known condition. For example, a composition standard can be processed to verify the accuracy of the GC or a thermocouple can be placed in boiling water to check its accuracy. The repeatability of a sensor can be estimated by observing sensor readings during a period of relatively steady-state operation. The repeatability is the variation in the sensor reading caused by noise. You can assume that during steady operation, the process is not, in fact, changing.

Determining the time constant of the sensor is usually more difficult than estimating the repeatability. To determine the time constant of the sensor system, you need to know the actual process measurement. Consider a temperature measurement. A measurement of the actual process temperature is required to estimate the time constant of a sensor. Instead, the thermal resistance, which causes excessive thermal lag of the temperature sensor, can be evaluated. The location of the thermowell should be checked to ensure that it extends far enough into the line so that the fluid velocity past the thermowell is sufficient; the possibility of buildup of insulating material on the outside of the thermowell should be assessed; and the thermal contact between the end of the temperature probe and the walls of the thermowell should be evaluated. In this manner, an indirect estimate of the responsiveness of the temperature sensor can be developed. The velocity of a sample in the line, which delivers a sample from a process line to a gas chromatograph, can indicate the transport delay associated with the sample system. A low velocity in the sample line from the process stream to a GC can result in excessive transport delay, which can greatly reduce controller effectiveness.

You should be careful to determine if the sensor used is really measuring the CV of interest. Differential pressure sensors are used for pressure, flow and level measurements. They are particularly susceptible to plugging of the sensing lines that connect the differential pressure sensor to the process itself. Plugging of the sensing lines can result from the buildup of coatings or solids or from freezing of the fluid in the pressure taps (Figure 2.4.9). The calibration of a differential pressure sensor is quite sensitive to the conditions of the fluid in the sensing lines. Condensate buildup in lines that should be dry can lead to large calibration errors. Table 10.2 lists some commonly encountered sources of problems for sensor systems. For a complete analysis of the sensor system, an instrument engineer or other expert familiar with that particular sensor may be required. Table 2.3 lists the expected ranges for the repeatability and the time constant for several commonly used sensors in the CPI. Deviation of the apparent repeatability or time constant from these expected values identifies a poorly performing sensor.

Table 10.2 Commonly Encountered Problems with Components of a Sensor System	
Sensor	• **Common Problems**
Transmitter	• Not calibrated correctly.* • Low resolution. • Excessive signal filtering.* • Slow sampling.
Thermocouple/ RTD	• Off-calibration.* • Short in the electrical circuit. • Improperly located thermowell or build up material on outside surface.* • Thermowell with excessive thermal resistance (e.g. stainless steel thermowells). • Partially burned out thermocouple. • Interference from heat tracing.
Pressure Indicators	• Plugged line to pressure indicator.* • Confusion about absolute pressure readings, gauge pressure readings and vacuum pressure readings. • Condensation in lines to pressure indicator.*
Sampling System For GC	• Excessive transport delay for an analyzer. • Sample drawn from wrong process point. • Plugged sample system.* • Sample system closed off.
GC	• Out of calibration. • Plugging in the GC column. • Failure of electrical components in GC. • Excessive noise on measurement.
Flow Indicator	• Square root compensation applied for non-DP-type flow indicator not applied properly. • Orifice plate installed backwards. • Damaged orifice plate. • Plugged line to DP sensor.* • Flashing of liquids as they flow through an orifice meter.
Level Indicator	• Plugged line from process to DP cell.* • Leak in line to DP cell or in DP cell itself. • Failure of stream tracing line resulting in (1) boiling of liquid in line to or from DP cell or solidification of liquid in line to or from DP cell. • Formation of emulsions, which can confound interface level measurements. • Leak in float type level indicators. • Formation of foams which can interfere with level measurements.
* Indicates a more frequently observed problem	

Self-Assessment Questions.

Q10.3.4 How can you estimate the repeatability of a sensor?

Q10.3.5 Explain how you determine whether a sensor system is functioning properly.

Q10.3.6 What is the most common problem with sensors based on a differential pressure measurement?

Self-Assessment Answers

Q10.3.4 The repeatability of a sensor can be estimated by observing sensor reading during a period of relatively steady-state operation. The repeatability would be the variation of the sensor reading caused by noise.

Q10.3.5 First check the repeatability of the sensor to ensure that it is in the proper range. Next, check the dynamic response time of the sensor system. This is much more difficult and may require an independent measurement of the CV if possible. Otherwise, an engineering analysis of the dynamic resistance of the sensor system is required.

Q10.3.6 Plugged line from DP cell to the process is the most common problem with a DP measurement.

Control Computer/DCS System. The control computer/DCS system consists of controllers, A/D and D/A converters except for digital field communication (e.g., Fieldbus) and the signal conditioning hardware and software, i.e., filtering and validation. Each of these components requires separate evaluation. Table 10.3 lists possible problems with the controller/DCS system. One way to initially check controller tuning is to place the control loop in manual (open the control loop) and observe whether the CV lines out to a steady-state or near steady-state value. Comparing the open-loop and closed-loop performance indicates if the controller is upsetting the process. If the controller is not upsetting the process, disturbances to the control loop in question are the primary source of the upsets. When evaluating whether a control loop is properly tuned, it is important to use a fundamental understanding of proportional, integral and derivative control (Section 7.7) and to understand the recommended approaches to tuning PID controllers (Sections 9.8-9.11).

Table 10.3
Possible Problems with the Controller/DCS System

- Improperly tuned controller.*
- Use of derivative action on a noisy CV.
- Wrong scaling for A/D and D/A converter.
- Improper or lack of pressure/temperature compensation for flow measurement.
- Improper selection of reverse-acting or direct-acting controller (Section 7.7).
- Too much or not enough filtering of the measured CV (Section 7.8).*
- Signal aliasing due to excessive control interval (see Appendix C).
- Poor resolution on A/D or D/A converters.
- Derivative action based on error from the setpoint instead of the measurement.
- Control loop is not in the proper mode (e.g., manual or auto).

* More frequently observed problems

Process Effects. The effect of the process on the closed-loop behavior can be examined directly by opening the control loop in question and observing the process behavior. Open-loop oscillatory behavior indicates a problem internal to the process or an external control loop. The noise level on the analyzer reading can also be assessed under open-loop conditions.

Fluctuating disturbances and process gain changes due to nonlinearity are a natural part of process control. Extraordinary disturbances combined with nonlinearity can cause an otherwise properly tuned controller to oscillate or go unstable during upset periods. Severe process nonlinearity can be identified if a closed-loop process exhibits ringing and sluggish behavior during different periods with the same controller tuning. Scheduling the controller tuning (Chapter 13) is one way to compensate for process nonlinearity. It may be possible to reduce the magnitude of the disturbances to acceptable levels by modifying the upstream operations (e.g., tuning upstream controllers). Excessive disturbances, when not measurable, can be inferred by observing the range of the average MV levels. If large magnitude disturbances are affecting the process, large changes in the average MV level are required to maintain the process near its desired operating point. Excessive fouling of heat exchangers or deactivation of catalyst can result in process gain changes that result in sluggish or unstable behavior.

Process changes that require MV levels in excess of what is physically available can also occur. After feed rate increases to a distillation column, the reboiler might be unable to provide enough heat transfer to maintain the purity of the bottom product. When this occurs, it is a physical limitation of the process and not the fault of the controller. Loss of steam pressure can also cause a constraint that can affect control loop performance. Downstream pressure changes can cause a constraint on the maximum flow rate due to an inadequate pressure driving force. Constraint control techniques (Chapter 13) should be used when MVs saturate. It should be clear that a thorough understanding of the process is a prerequisite for control loop troubleshooting.

Testing the Entire Control Loop. After each of the components has been evaluated and corrected, wherever possible, the closed-loop system should be checked. From an overall point of view, there are three general factors that affect the closed-loop performance of a control loop: (1) the type and magnitude of disturbances, (2) the lag associated with the components that comprise the control loop and (3) the precision to which each component of the control loop performs. For example, actuator deadband affects the variability in the CV, and the addition of lag to a control loop (e.g., sensor filtering) results in slower disturbance rejection, which can increase the variability in the CV. Disturbance magnitude directly affects CV variability.

The performance of a closed-loop system can be assessed by the settling time, closed-loop deadband and the variability of the CV evaluated over an extended period of time. The closed-loop settling time and the closed-loop deadband can be determined using a closed-loop block sine wave test (Figure 10.3.3). For a closed-loop block sine wave test, the setpoint for the control loop is applied in the form of a block sine wave and the amplitude of the block sine wave is varied until the deadband is determined. During these tests, the settling time of the controller can also be estimated by holding the setpoints long enough to determine the settling time. An accurate determination of the variability of a CV generally requires an extended period of operation (e.g., an SPC chart, Section 9.2). An evaluation of the variability based on a short period of time

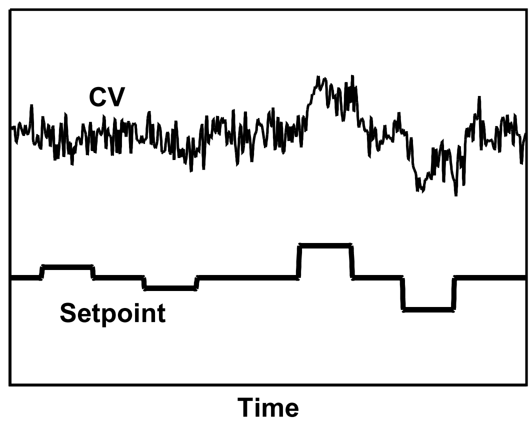


Figure 10.3.3 An example of a closed-loop block sine wave test.

may not be representative of the true performance of the system.

The closed-loop settling time is an indication of the combined lags of the control loop components. The closed-loop deadband is an indication of the variability in the CV that results from the combined effects of actuator deadband, sensor noise and resolution of the A/D and D/A converters and the combined lags of the control loop components. The closed-loop performance assessment is a means of determining if all the major problems within a control loop have been rectified. As a preventive measure, it is advisable to perform closed-loop block sine wave tests on the key loops in a process to document normal operating conditions. In this manner, when the operation of the process is less than satisfactory, you can perform another closed-loop block sine wave test and compare results to the earlier test to determine if the important control loop is performing as before.

Consider a control system with an excessive lag (e.g., buildup of scale on the exterior of a thermowell) added to a control loop. The controller can be tuned for any tuning criterion (e.g., a decay ratio of 1/6 to critically damped); therefore, tuning a control loop to the desired tuning criterion is not, in itself, an indication of the performance of the control system. The closed-loop settling time and the closed-loop dead-band provide a measure of the performance of the control system. A process with the additional lag exhibits a longer settling time than a process without the additional lag. The closed-loop dead-band also shows that the system without the additional lag exhibits superior control performance. When troubleshooting a control loop, you can determine the relative change in controller performance by comparing the closed-loop settling time and closed-loop dead-band before and after control loop troubleshooting was undertaken.

Example 10.1 Troubleshooting Example

Problem Statement. Present a step-by-step troubleshooting process, along with intermediate results, for a temperature controller that results in sluggish closed-loop performance.

Solution. Step 1: Determine the deadband of the final control element using a series of block sine wave tests. **Result:** The deadband of the final control element is less than 0.4% and the dynamic response time of the final control element is 2 seconds; therefore, the final control element is functioning properly.

Step 2: Retune the temperature controller. **Result:** The controller settings do not change significantly; therefore, the controller tuning does not appear to be the problem.

Step 3: Evaluate the sensor. Check the repeatability of the sensor by observing the temperature measurements during a steady-state or near steady-state period. **Result:** The repeatability is less than 0.1 ºC, which is good for an RTD. An independent measurement of the temperature is made and compared with the sensor reading. **Result:** The sensor reading is observed to have excessive lag, i.e., a dynamic response time for the sensor is estimated equal to about 5 minutes. Based on Table 2.3, the maximum response time for a properly functioning temperature sensor is less than 20 s. It is determined upon further examination that there was an excessive air space between the RTD element and the surface of the thermowell. The proper installation of the RTD in the thermowell is made and the dynamic lag of the sensor is found in the proper range. The controller is retuned and control performance is significantly improved.

Example 10.2 Troubleshooting Example

Problem Statement. Present a step-by-step troubleshooting process, along with intermediate results, for a reactor temperature controller that exhibits excessive reactor temperature excursions from setpoint. Reactor temperature control is achieved by manipulating the heat added to the reactor feed. The reactor temperature controller is supervisory and selects the setpoint for the flow controller on the steam to the feed preheater, which is the regulatory controller.

Solution. Step 1: First, the performance of the flow controller on the steam line is evaluated**.** Perform a series of closed-loop block sine wave tests on the steam flow controller. **Result:** The repeatability of the flow measurement is observed to be larger than it should be. It is determined that the differential pressure sensor has a partially plugged pressure tap. This problem is corrected and the flow controller is tested over the expected flow rate range. **Result:** It is observed that in the low flow rate range the flow control performance is poor. Upon examination of the control valve it is determined that an equal percentage valve is used while a linear valve should have been used because the pressure drop across the valve remains relatively constant. The valve plug and cage are replaced with ones that result in linear inherent valve characteristics and good controllability of the steam flow over the entire flow rate range is observed after the flow controller is retuned.

Step 2: Evaluate the temperature sensor on the product stream from the reactor. **Result:** It is determined that the repeatability and dynamic response of the temperature sensor are good.

Step 3: Retune the temperature controller on the reactor exit temperature. **Result:** After the temperature control loop is retuned, the variability in the reactor temperature from setpoint is observed to be a factor of three lower than was previously observed, thus meeting the operational objectives of this control loop.

Example 10.3 Troubleshooting Example

Problem Statement. Present a step-by-step troubleshooting procedure for a composition control loop on the overhead of a distillation column for which the variability of the impurity level in the overhead product is in excess of the product specifications. The output of the composition controller goes to a flow control loop on the reflux flow.

Solution. Step 1: Evaluate the deadband of the reflux flow controller. **Result:** The deadband of the flow control loop is found equal to ±0.3% with a time constant of approximately 1.5 seconds; therefore, the flow control loop is functioning properly.

Step 2: Check the tuning for the composition controller. **Result:** The controller is properly tuned.

Step 3: Evaluate the on-line GC. **Result:** The repeatability of the analyzer is found equal to ±2% by observing GC readings during a steady-state period. This repeatability is well within the product variability limits and is consistent with the analysis of this type of mixture. Upon further examination it is determined that there is excessive sample transport delay; therefore, the sample pump is replaced and the sample transport delay is reduced to an acceptable level because the velocity through the sample line is increased to a proper level. After this change and the retuning of the composition controller, the variability of the overhead product is reduced, but it is found to periodically exceed the product variability specifications.

Step 4: Evaluate the closed-loop deadband for the composition control loop. **Result**: It is found that the deadband is acceptable, but the dynamic response is slower than expected. Upon further evaluation it is determined that excessive filtering of the analyzer reading is being used. The proper level of filtering is applied and the composition controller is retuned. The resulting product variability is found to be well within the product specifications.

Self-Assessment Questions.

Q10.3.7 How is a closed-loop block sine wave test different from a block sine wave test applied to an actuator system?

Q10.3.8 What information is contained in the deadband obtained from a closed-loop block sine wave test?

Q10.3.9 What information does the closed-loop settling time of a control loop contain?

Self-Assessment Answers

Q10.3.7 For a closed-loop block sine wave test, the setpoint for the control loop is applied in the form of a block sine wave. For the block sine wave test applied to actuator system, the controller output is applied in the form of a block sine wave.

Q10.3.8 The closed-loop deadband is an indication of the variability in the CV that results from the combined effects of actuator deadband, sensor noise and resolution of the A/D and D/A converters.

Q10.3.9 The closed-loop settling time is an indication of the combined lags of the control loop components.

Self-Assessment Question

P10.3.1 In a manner similar to Examples 10.1, 10.2 and 10.3, present a troubleshooting study and results for a level controller malfunctioning because of a plugged line to the DP sensor in the level indicator.

Self-Assessment Answer

P10.3.1 Following is a step-by-step troubleshooting process along with intermediate results for a level controller that was observed to be performing poorly.

Step 1: Determine the deadband of the final control element using a series of block sine wave tests. **Result:** The deadband of the final control element was less than 0.4% and the dynamic response time of the final control element was 2 seconds; therefore, the final control element was found to be functioning properly.

Step 2: Retune the level controller. **Result:** The controller settings did not change significantly; therefore, the controller tuning does not appear to be the problem.

Step 3: Evaluate the sensor. Check the repeatability of the sensor by observing the level measurements during a steady-state or near steady-state period. **Result:** The repeatability was found to be ±5% which is excessively high. Upon examination and testing of the DP cell used to measure the level of the tank, it was determined that the lines between the process and the DP cell were partially plugged. After the DP cell taps were cleaned and the DP cell was calibrated and put back into service, the controller was retuned and control performance was found to be significantly improved.

10.4 Troubleshooting Control Systems for Bio-Processes

Troubleshooting control systems in the CPI usually reduces to analyzing the performance of a single control loop. On the other hand in the bio-tech industries, troubleshooting usually involves the operation of a bio-reactor, involving several control loops and a number of process sensors. A bio-reactor can produce an off-specification product (i.e., bad batch) due to improperly functioning control loops (e.g., DO, pH or temperature control loops) or due to malfunctioning sensors that are used by the operator to manage the batch (e.g., HPLC used to measure the substrate or product level in a bio-reactor). As a result, troubleshooting bio-reactors is a more global problem than troubleshooting most control loops in the CPI.

From another perspective, the standard control loops for most bio-reactors are for temperature, agitation speed, back pressure, inlet air flow and pH. Perfect control of all of these parameters provides very little information or clues as to the health and viability of the cell culture in the bio-reactor that is growing and producing product. That is why one or more additional sensors are typically installed on bio-reactors [e.g., dissolved oxygen probes, turbidity measurement and interface to a mass spectrometer (gas analyzer)] which give a better indication of how the cell culture is progressing.

When an "expert" analyzes the operation of a bio-reactor, he/she simultaneously considers all the available process data for the reactor. In particular, the expert will use a variety of information related to the metabolic rate of the microorganism in the bio-reactor (e.g., mass spectrometer for the CO_2 production rates, turbidity measures of the cell concentration, DO level, etc.) to assess the general health of that batch. Some pharmaceutical companies have compiled the experience of their process experts (operators, technical service, and the scientists who developed the process) into “real-time” expert systems, which continuously diagnose the operation of each batch using an extensive set of “if-then-else” rules. The expert system not only provides an indication of an actual or impending abnormal situation but provides a priority listing of potential root causes, which can be sent directly to the operators and/or support technical staff. Often there is ample time to correct the faults and return the bio-reactor to the desired operating trajectory. Since a particular abnormal situation might have several different causes, having a prioritized list greatly helps in efficient troubleshooting because the cause at the top of the list (highest probability of being the correct cause) will be the first one investigated.

Interviews with the experts are used to compile the “if-then-else” rules that the expert system uses for fault detection, determination of impending faults, root cause lists and assistance with troubleshooting. Note that the rules look not only at all the current values associated with the process, but also with several attributes of the trend graphs (e.g., extrapolating filtered slopes). Therefore, the expert system can not only identify current issues, but trends toward abnormal situations that will not actually occur until later (unless addressed immediately). While expert systems can be time consuming to develop, implement, and validate, they have been shown to be quite effective in representing human expertise about the process and making that expertise continuously available on-line as the process proceeds. This helps preserve the expertise of key process experts (who will eventually retire or change jobs) and makes that expertise immediately available to the process, rather than needing to call out an expert.

While it is useful to consider the bio-reactor as a whole, it can also be beneficial to evaluate each element of the control and instrumentation system of a bio-reactor. As demonstrated in the previous sections, each actuator and sensor system should be evaluated separately. Following is a summary of the key failure modes and problem areas for typical actuators and sensor systems used in the bio-tech industries.

Actuator Systems. Actuator systems used by the biotechnology industries include variable speed positive displacement pumps, variable speed air compressors (i.e., air blowers) and control valves. The evaluation approaches and failure modes for control valves listed in the Section 10.3 for the CPI also apply for actuators used in the bio-tech industries. In most cases, a flow sensor reading is available for each actuator system used. Therefore, the block sine wave approach shown Section 10.3 can be used to determine the deadband and the time constant for each actuator. Then these values can be compared with the expected performance specifications listed in Table 2.3 to determine if the actuator is functioning properly. That is, for positive displacement variable speed pumps, if the measured flow rate does not correspond to the control signal to the pump (i.e., a 4-20 mA signal), either the pump is malfunctioning or the flow sensor is malfunctioning.

Sometimes, the reactants for a bio-reactor during the initial make-up of the reactor (when tank gauge pressure is zero) are fed to the reactor by gravity flow. In these cases, the flow control is not crucial and ball or diaphragm valves are used as the control valves instead of globe valves. For globe valves, the valve stem moves up and down which can draw fluid into and out of the packing. Therefore, globe valves are much more difficult to clean and to maintain sterile conditions and are rarely used for bio-processes. On the other hand, the stem movement for ball and diaphragm valves does not tend to draw the fluid into the packing. As a result, ball and diaphragm valves are much easier to clean and maintain sterile conditions, and therefore, are used for certain bio-process applications.

Sensor Systems. In the bio-tech industries, preventive maintenance, which involves testing and recalibrating most sensors and actuators, is conducted on a periodic basis (e.g., 6 months). Certain high maintenance sensors, e.g, DO and pH sensors, are replaced at more frequent intervals based on the observed reliability of these sensors for a particular application. The same applies to separation columns in HPLCs. Some sensors (e.g., mass spectrometers) are self calibrating and self checking and so preventive maintenance checks can be less frequent, if needed at all.

Coriolis flow meters. Coriolis flow meters are very easy to clean and have proven to be highly reliable. The primary concern with Coriolis flow meters is to maintain their calibration. The electronics associated with a Coriolis meter tend to cause drift in the readings. As a result, Coriolis meters should be regularly calibrated.

DO sensors. Most ion-specific electrodes are galvanic cells. While an electrode is operating, a current is passed through the system, causing the dissolution of the silver chloride reference electrode (Figure 2.4.11). Eventually, the silver reference electrode is consumed, resulting in a poorly performing electrode. It is routine before inoculation to start the air compressor and allow the sterile broth to become saturated with O_2. The voltage generated by the DO electrode is measured and if this value is in the normal operating range, the DO sensor should be operational and the sensor is calibrated to match 100% saturation. On the other hand, if the measured voltage is low, the sensor should be replaced.

DO sensors typically use a Teflon membrane and this membrane is usually replaced after every 2-3 batches and in certain longer batch runs (e.g., 10-30 days for most mammalians cell cultures), it is replaced after each batch. The Teflon membranes are much more fragile than the glass ones used for pH sensors. As a result, DO electrodes, due to their Teflon membranes, are much more susceptible to damage, which can affect sensor reliability. If the membrane is not properly installed, the electrolyte in the electrode will leak into the bio-reactor, thus affecting the performance of the DO sensor. If a DO sensor is improperly mounted, its useful life and performance can be significantly affected. For example, if a DO sensor is mounted where it is in direct contact with air bubbles, the sensor reading will exhibit excessive sensor noise. In addition, if it is located too closed to where anti-foaming agents (e.g., oils) are injected, excessive fouling of the sensor can result, causing a significant increase in the time constant for the sensor. In addition, failure of a DO sensor can occur if the electrolyte in the electrode is heated excessively during the sterilization process.

pH sensors. Most bio-reactors are operating at a pH between 5-8 pH units. The reliability of pH sensors is generally good, but the most significant problem with pH sensors is calibration drift. In certain cases, broth samples are withdrawn from a bio-reactor and bench-top pH readings for the both samples are taken to calibrate the pH sensors for the bio-reactor. Many times the bio-reactors are operated under significant pressure so that when broth samples are analyzed at atmospheric pressure, CO_2 degassing results, which in turn affects the bench-top pH reading. Therefore, there is generally an unknown bias between the bench-top reading and the pH in the broth, which complicates the calibration process.

Similar to the DO sensor, pH electrodes have a limited operating life and require replacement periodically. In addition, care must be taken to ensure that a pH sensor is mounted in the proper location. Otherwise, a noisy pH reading can result from air bubbles contacting the probe or excessive fouling from locating it too close to the injection point for anti-foaming agents.

Redox sensors. Redox sensors can be used, if needed, to monitor the operation of a bio-reactor when the DO level is less than 5 % of saturation. Redox sensors, similar to pH sensors, are generally quite reliable, but are subject to sensor drift requiring regular calibration. Also, redox sensors should be replaced at regular intervals due to the consumption of the AgCl reference electrode and fouling of the membrane.

Turbidity sensors. Turbidity sensors are used to measure cell concentrations for systems that do not contain suspended solids in the broth, i.e., the cells represent the only suspended material in the broth. Turbidity sensors are susceptible to accumulation of cells in the optical chamber, where the light scattering measurements are made. Accumulation of cells in the optical chamber can result is errors in the measured cell concentration. In addition, it should be remembered that a turbidity meter cannot distinguish between live and dead cells, which can be a problem in certain applications.

Mass Spectrometers. Mass spectrometers are widely used in the bio-tech industries for bio-reactor off-gas analysis primarily because the culture respiration rate calculations based on mass spectrometer near real-time data results is the single best indicator of the health and viability of the cell culture. Use of mass spectrometers also benefits in that one instrument can provide on-line measurements for up to 32 bio-reactors which greatly reduces the capital cost on a per reactor basis. In addition, mass spectrometers are very fast and highly accurate and reliable. In fact, other than a sampling multiplexing valve and a vacuum pump, there are almost no moving parts in a mass spectrometer. Usually, during each cycle of the sampling valve, ambient air is also automatically tested as a calibration check. That is, because the concentrations of N_2, O_2, Argon, and CO_2 in air are known, periodically testing air provides an immediate indication of a poorly performing instrument.

HPLC. HPLCs require regular calibration, and their performance can degrade with use due to fouling of the HPLC separation column. In fact, HPLCs are usually installed with "guard" columns, which are designed to remove fouling agents from the test stream before it enters the HPLC unit. The separation of the species provided by the HPLC column is a function of operating temperature. Therefore, the operating temperature is adjusted to maintain optimum separation of the species for the column.

FIA. One of the major failure modes for a FIA analyzer is malfunctioning valves, which control the mixing of the test stream with the colorimetric solutions. Gas in the sample will introduce noise on the composition analysis from a FIA unit. It may be necessary to provide continuous degassing of the sample to reduce this effect.

Process Effects. The performance of a bio-process is dependent on many factors. It is truly a multivariant environment, and as such, monitoring and controlling single variables at a given value may not necessarily result in the production of products with the expected product quality. Multivariable statistical analysis[1] (e.g., principal component analysis, PCA) can be suitable for detecting and analyzing faults in bio-processes.

Self-Assessment Questions

Q10.4.1 What are the primary difference between control loop troubleshooting in the CPI and the bio-tech industries?

Q10.4.2 How are expert systems used to troubleshoot the operation of bio-reactors?

Q10.4.3 What are the similarities and difference between pH electrodes and DO electrodes?

Q10.4.4 Why are mass spectrometers widely used by large-scale pharmaceutical companies for off-gas analysis?

Self-Assessment Answers

Q10.4.1 In the CPI, troubleshooting usually reduces to testing an individual control loop while for the bio-tech industries, troubleshooting involves the entire process, which involves several control loops and a number of sensors used for monitoring.

Q10.4.2 Expert systems are a compilation of experience of process experts that consider all the available process measurements as a whole and use this information to identify operational problems and when this occurs, direct the operations personnel to the most likely causes of the problems.

Q10.4.3 pH electrodes and DO electrodes both base their measurement on voltage generated by the electrode. Because they are galvanic cells, they need to be replace periodically because the reference electrode become consumed as these sensors remain in service. Both electrodes are susceptible to fouling of their membranes and sensor noise from contact with air bubbles. pH sensors tend to last longer than DO sensors because pH sensors have a glass electrode while DO electrodes use Teflon membranes, which are usually replaced every 1-3 batches.

Q10.4.4 Mass spectrometers provide fast, accurate measurements of the full range of components in the off-gas from a bio-reactor. In addition, they are extremely reliable. And most important, one mass spectrometer instrument can provide on-line measurements for as many as 32 bio-reactor, greatly reducing the cost on a per bio-reactor basis.

10.5 Summary

- The key to troubleshooting control loops in the CPI is to independently check the actuator system, the sensor system, the process and the control computer to isolate the source of the problem.
- The actuator system is the easiest to check using a block sine wave test because the flow measurement is generally available.
- The sensor system may require an instrument technician to determine whether it is functioning properly.
- Checking the control computer generally involves evaluating the controller tuning but can also involve A/D and D/A converters, sensor signal conditioning or simple oversights, such as specifying a direct-acting controller for a case that requires a reverse-acting controller.
- Changes to the process (e.g., changes in the type or magnitude of disturbance) can also be the source of a poorly performing control loop. Finally, the closed-loop performance of the control loop should be tested.
- Troubleshooting bio-reactors in the bio-tech industries involves considering the process as a whole and many times expert systems are used to aid in this activity.

10.6 References

1. Karim, M.N., D Lodge and L. Simon, "Data Driven Approaches to Modeling and Analysis of Bioprocesses: Some Industrial Examples", *Biotechnol. Progress*, 19(5) pp. 1591-1605 (2003).

10.7 Additional Terminology

Block sine wave - a series of step changes that approximate a sine wave.

Closed-loop deadband - the maximum positive and negative setpoint change to a control loop that can be implemented without a noticeable change in the measured value of the CV.

10.8 Preliminary Questions

10.2 Overall Approach to Troubleshooting Control Loops

Q10.2.1 Summarize the recommended approach to troubleshooting a process control loop.

10.3 Troubleshooting Control Loops in the CPI

Q10.3.1 How can a control engineer determine if the final control element is functioning properly?

Q10.3.2 What is a block sine wave?

Q10.3.3 How is a block sine wave used to evaluate the performance of an actuator system?

Q10.3.4 What is the usual cause of excessive valve deadband?

Q10.3.5 Explain how you determine whether a sensor system is functioning properly.

Q10.3.6 Is the accuracy of a sensor always important to the performance of a control loop?

Q10.3.7 Is the repeatability of a sensor always important to the performance of a control loop?

Q10.3.8 Is the time constant of a sensor always important to the performance of a control loop?

Q10.3.9 Why do some engineers prefer a vortex shedding meter or a magnetic flow meter to an orifice flow meter?

Q10.3.10 How can condensate in a pressure tap affect a level reading?

Q10.3.11 What is the most common problem with the DCS/controller system?

Q10.3.12 Explain how controller tuning can mask problems with the hardware in a control loop. Give an example.

Q10.2.13 How can the wrong level of sensor filtering affect control performance?

Q10.3.14 What is the resolution of an A/D converter?

Q10.3.15 How does a control engineer determine if a process has excessive nonlinearity?

Q10.3.16 How can you determine if large disturbances are entering a process?

Q10.3.17 How do control engineers test the closed-loop performance of a control loop? What metrics do they use?

Q10.3.18 How is the variability of a control loop measured?

10.4 Troubleshooting Control Systems for Bio-Processes

Q10.4.1 What are the expert systems used to troubleshoot bio-reactors?

Q10.4.2 How are expert systems constructed?

Q10.4.3 How would you determine if a variable speed pump or a variable speed air compressor is functioning properly?

Q10.4.4 What is the primary maintenance problem associated with Coriolis meters?

Q10.4.5 What factors contribute to limiting the life of an ion-specific electrode, such as DO, pH and Redox sensors?

Q10.4.6 What is the primary sources of noise for DO, pH and Redox sensors?

Q10.4.7 How can fouling of HPLC analyzers be reduced?

10.9 Analytical Questions and Exercises

10.3 Troubleshooting Control Loops in the CPI

P10.3.1 In a manner similar to Examples 10.1, 10.2 and 10.3, present a troubleshooting study and results for

a. A pressure controller malfunctioning because of an improperly tuned valve positioner on the control valve.

b. A flow control loop malfunctioning because of low resolution of the A/D and D/A converters.

P10.3.2 What would happen if the output of a controller were sent to the valve on the MV and to a valve that affects a disturbance of the process? How difficult would it be to identify this fault during a troubleshooting effort?

Chapter 11

Frequency Response Analysis

Chapter Objectives

- Introduce the Bode plot and define the variables associated with it.
- Present several ways to develop a Bode Plot for a process.
- Present the Bode stability criterion.
- Demonstrate how to tune a controller using gain margins and phase margins.
- Introduce the Nyquist plot.
- Discuss closed-loop Bode plots and show how they determine the most sensitive disturbance frequencies for a control loop.

11.1 Introduction

To this point, we have used open-loop and closed-loop time domain behavior as well as Laplace domain characteristics (e.g., the characteristic equation) to characterize the dynamic behavior of a process. In this chapter, we will use the frequency domain to characterize the dynamic of a process by considering the open-loop effect of sinusoidal inputs (Figure 11.1.1) over a range of frequencies, , where

$$c(t) \quad a \sin \quad t$$

This procedure is called **frequency response analysis**. The frequency of an input can have a very significant effect upon the resulting behavior of the process. Consider a mixing tank. If sinusoidal variations in the inlet concentration are applied at a high frequency, no measurable sinusoidal variation in the outlet concentration results. The peaks and the valleys of the sinusoidal variation of the input average out because the input changes are occurring faster than the process can respond. On the other hand, if sinusoidal variations in the inlet concentration are applied at a low frequency, significant variation in the output would result because the process has ample time to respond to its input changes. Frequency response analysis can allow you to better understand the effect of the time scale of inputs on the behavior of a process.

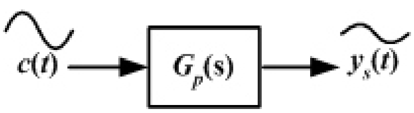


Figure 11.1.1 Schematic of a process exposed to a sinusoidal input.

Frequency response analysis is important to the understanding of the feedback control behavior of industrial processes and provides important process control terminology.

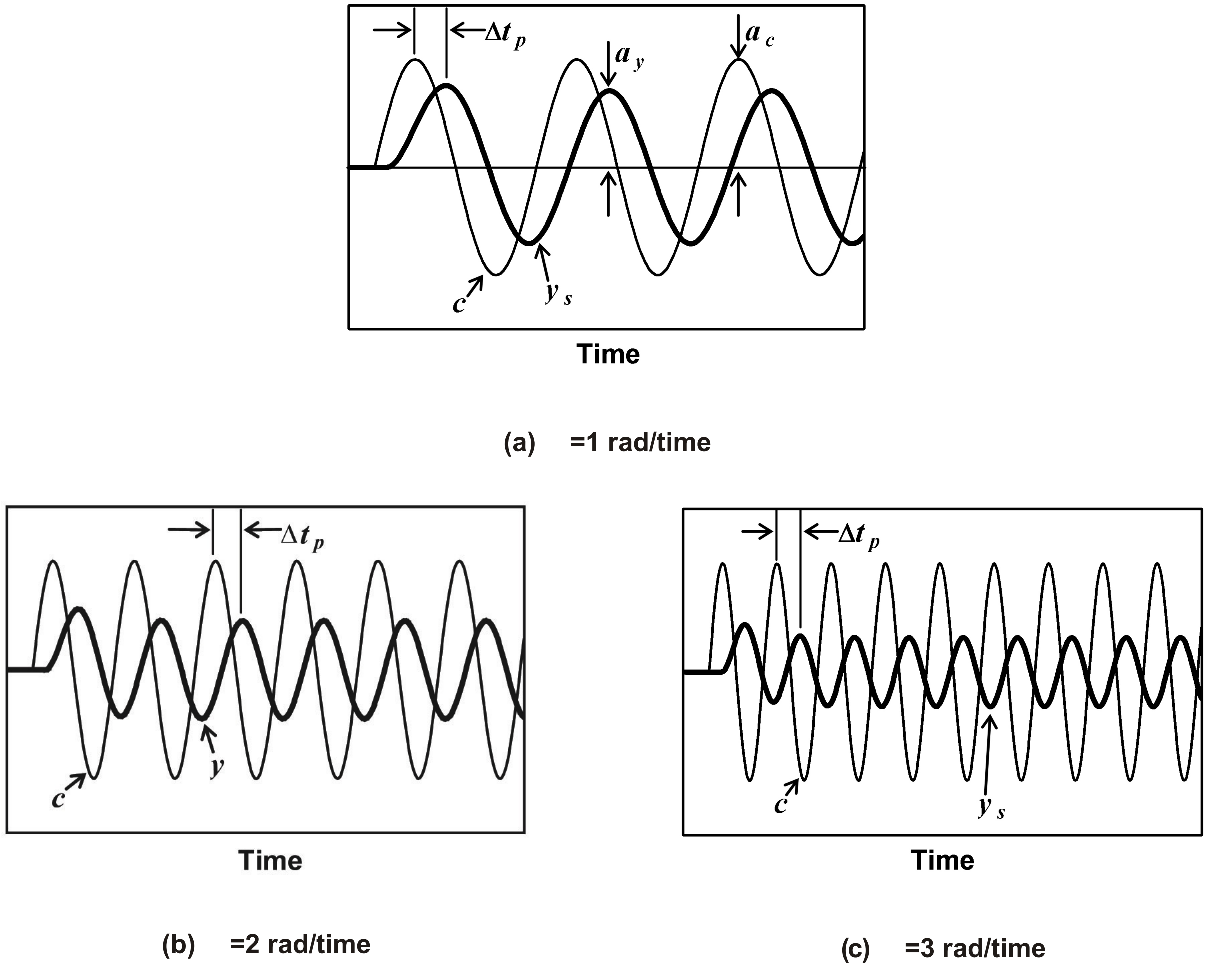


Figure 11.2.1 The effect of frequency on the amplitude and phase lag of a FOPDT process.

11.2 Bode Plots

Figure 11.2.1 shows the open-loop dynamic behavior of a FOPDT process model ($K_p = 1$, $\tau_p = 1$, $\theta_p = 0.5$) that represents the combined effect of the actuator, process and sensor subjected to a sinusoidal input with three different frequencies. Sinusoidal changes in c are the input to the process and the output is y_s. The amplitude of the variations in y_s and the difference in the timing of the peaks in y_s and c at each frequency characterize the frequency response behavior of this process. To normalize the results, the **amplitude ratio** is used, i.e.,

$$A_r = \frac{a_y}{a_c}$$

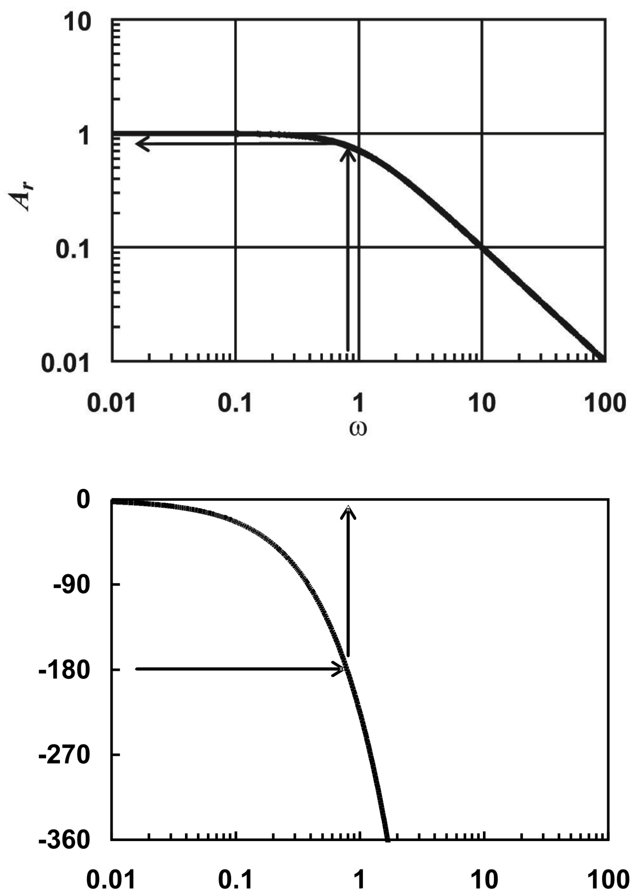


Figure 11.2.2 Bode plot for a FOPDT process (a) The upper plot is the amplitude ratio (b) The lower plot is the phase angle.

where a_y (Figure 11.2.1a) is the amplitude of the sinusoidal variation in the CV, y_s and a_c (Figure 11.2.1a) is the amplitude of the sinusoidal variations in the system input, c. The time difference between peaks, t_p (Figure 11.2.1) can be converted into an angle and is referred to as the **phase angle** of the response, ,

$$\frac{t_p}{2} \quad 360$$

where is the frequency of the input variations in radians per second. ***t_p* in this equation is assumed positive when *c* lags *y_s*.** For Figure 11.2.1, $_p$ is negative because y_s lags c. A **Bode plot** of a process is a plot of and the logarithm of A_r versus the logarithm of . Note that these are common logarithms with base 10. Figure 11.2.2 shows the Bode plot for the previous FOPDT process model. From Figure 11.2.1, as the frequency increases, the amplitude ratio decreases, which is consistent with Figure 11.2.2. Because y_s lags c in Figure 11.2.1, the phase angle is negative as is shown in Figure 11.2.2. Therefore, the direct way to generate a Bode plot is to excite the process with sinusoidal inputs of varying frequencies, wait until standing waves have been established and measure the amplitude ratio and phase angle.

The Bode plot of a process can also be generated by using the transfer function of the process if it is known. Consider a first-order process without deadtime

$$G_p(s) \quad \frac{Y(s)}{C(s)} \quad \frac{K_p}{{}_p s \quad 1}$$

with a sinusoidal input [$c(t)$=a_c sin t] $\qquad C(s) \quad \frac{a_c}{s^2 \quad {}^2}$

Then, combining the transfer function with the input and applying partial fraction expansions yield

$$Y(s) \quad \frac{K_p \, a_c}{({}_p s \quad 1)(s^2 \quad {}^2)} \quad \frac{K_p \, a_c}{{}^2_p \; {}^2 \quad 1} \quad \frac{{}^2_p}{{}_p s \quad 1} \quad \frac{s \quad {}_p}{s^2 \quad {}^2} \quad \frac{}{s^2 \quad {}^2}$$

Taking the inverse Laplace transform yields

$$y(t) \quad \frac{K_p a_c}{{}^2 \ {}^2_p \quad 1} [\quad {}_p e^{\ t/\ p} \qquad {}_p \cos \quad t \quad \sin \quad t]$$

and applying trigonometric relations results in

$$y(t) \quad \frac{K_p a_c \quad p}{{}^2 \ {}^2_p \quad 1} e^{\ t/\ p} \quad \frac{K_p a_c}{\sqrt{{}^2 \ {}^2_p \quad 1}} \sin(\quad t \quad)$$

where $\tan^{\ 1}(\quad {}_p)$. The first term on the right-hand-side of the equation is the transient response, and it decays to zero at large time. The second term is the forced response, and it describes the standing wave in y_s that results from the sinusoidal input. Therefore, the amplitude ratio is determined by the forced response and is given by

$$A_r \quad \frac{a_y}{a_c} \quad \frac{\frac{K_p a_c}{\sqrt{{}^2 \ {}^2_p \quad 1}}}{a_c} \quad \frac{K_p}{\sqrt{{}^2 \ {}^2_p \quad 1}}$$

and the phase angle is simply i.e., $\tan^{\ 1}(\quad {}_p)$.

Another way to use a transfer function to generate a Bode plot is to substitute $s = i$ into $G_p(s)$ and factor the result into the real and imaginary components, i.e.,

$$G_p(i \ ) \quad R(\ ) \quad iI(\ ) \qquad \textbf{11.2.1}$$

then

$$A_r(\ ) \quad \left|G_p(i \ )\right| \quad \sqrt{R^2(\ ) \quad I^2(\ )} \qquad \textbf{11.2.2}$$

$$G_p(i \ ) \quad \tan^{\ 1}[I(\ )/R(\ )] \qquad \textbf{11.2.3}$$

For example, consider a first-order process without deadtime. Substituting $s=i$,

$$G_p(i \ ) \quad \frac{K_p}{i \quad {}_p \quad 1}$$

To remove *i* from the denominator, multiply both the numerator and denominator by the complex conjugate of the denominator , i.e., $(\ i \quad {}_p \quad 1)$

$$G_p(i\quad) \qquad \frac{K_p}{i\quad _p\quad 1}\ \frac{i\quad _p\quad 1}{i\quad _p\quad 1}\ \frac{iK_p\quad _p\quad K_p}{^{2}\ {}^{2}_{p}\quad 1}\ \frac{K_p}{^{2}_{p}\quad 1}\ i\frac{K_p\quad _p}{^{2}\ {}^{2}_{p}\quad 1}$$

Then, using Equations 11.2.2 and 11.2.3,

$$A_r(\ \ ) \quad \sqrt{R^2(\ \ )\quad I^2(\ \ )} \quad \sqrt{\left(\frac{K_p}{\sqrt{^{2}\ {}^{2}_{p}\quad 1}}\right)^2 \quad \left(\frac{K_p\quad _p}{\sqrt{^{2}\ {}^{2}_{p}\quad 1}}\right)^2} \quad \frac{\sqrt{K_p^2\quad K_p^2\ {}^{2}\ {}^{2}_{p}}}{^{2}\ {}^{2}_{p}\quad 1} \quad \frac{K_p}{\sqrt{^{2}\ {}^{2}_{p}\quad 1}}$$

and
$$(\ \ ) \quad \tan^{1}\frac{R(\ \ )}{I(\ \ )} \quad \tan^{1}\frac{\dfrac{K_p\quad _p}{^{2}\ {}^{2}_{p}\quad 1}}{\dfrac{K_p}{^{2}\ {}^{2}_{p}\quad 1}} \quad \tan^{1}(\quad _p)$$

This method agrees with the results of the earlier approach and is generally easier to apply than developing a time-domain solution using Laplace transforms.

Consider a transfer function given by

$$G_p(s) \qquad \frac{G_a(s)\,G_b(s)}{G_c(s)\,G_d(s)}$$

Substituting $s{=}i$ yields

$$G_p(i\quad) \qquad \frac{G_a(i\quad)\,G_b(i\quad)}{G_c(i\quad)\,G_d(i\quad)}$$

The overall amplitude ratio using the rules of complex algebra is given by

$$(A_r)_{overall} \qquad \left|G_p(i\quad)\right| \qquad \frac{\left|G_a(i\quad)\right|\left|G_b(i\quad)\right|}{\left|G_c(i\quad)\right|\left|G_d(i\quad)\right|}$$

Because $\left|G_j(i\quad)\right|$ is equal to the amplitude ratio of the jth transfer function,

$$(A_r)_{overall} \qquad \frac{(A_r)_a\ (A_r)_b}{(A_r)_c(A_r)_d} \tag{11.2.4}$$

and for the phase angle

$$G_p(i\quad) \qquad G_a(i\quad) \qquad G_b(i\quad) \qquad G_c(i\quad) \qquad G_d(i\quad) \tag{11.2.5}$$

Table 11.1
Amplitude Ratios and Phase Angles for Several Transfer Functions

Transfer Function	A_r	( )
K_p	K_p	0
$\frac{K_p}{{}_p s \quad 1}$	$\frac{K_p}{\sqrt{\;^2\;{}_p^2 \quad 1}}$	$\tan^{1}(\quad {}_p)$
$\frac{K_p}{{}_n^2 s^2 \quad 2 \quad {}_n \quad s}$	$\frac{K_p}{\sqrt{(1 \quad ^2\;{}_n^2)^2 \quad (2 \quad {}_n \quad)^2}}$	$\tan^{1} \frac{2 \quad {}_n}{1 \quad ^2\;{}_n^2}$
$e^{\;{}_p s}$	1	${}_p \; \frac{360}{2}$
$\frac{K_p}{s}$	$\frac{K_p}{\quad}$	- 90
s		+90
${}_{ld} s \quad 1$	$\sqrt{\;^2\;{}_{ld}^2 \quad 1}$	$\tan^{1}(\quad {}_{ld})$
$1 \quad s$	$\sqrt{\;^2\;^2 \quad 1}$	$\tan^{1}(\quad)$
$K_c \; 1 \quad \frac{1}{{}_I s} \quad {}_D s$	$K_c \sqrt{\quad {}_D \quad \frac{1}{\quad {}_I}^{2} \quad 1}$	$\tan^{1} \quad {}_D \quad \frac{1}{\quad {}_I}$

Bode plots of individual transfer functions can be graphically combined to yield the Bode plot of the product of the transfer functions. Taking the logarithm of Equation 11.2.4 results in

$$\log(A_r)_{overall} \qquad \log(A_r)_a \qquad \log(A_r)_b \qquad \log(A_r)_c \qquad \log(A_r)_d$$

Because a Bode plot contains the logarithm of A_r, the above equation shows that the logarithm of A_r of a process that is composed of the product of several transfer function is the sum of the logarithms of the A_r's of the individual transfer functions. Likewise, Equation 11.2.5 indicates that the phase angle of a process that is composed of the product of several transfer functions is simply the sum of the phase angles of the individual transfer functions. In this way, the Bode plots of complex transfer functions can easily be constructed using the known Bode plots of the individual components.

Table 11.1 lists A_r and as a function of for several commonly encountered transfer functions. The transfer functions considered are for a constant, a first-order process, a second-order process, a deadtime element, an integrator, a derivative element, a lead element, a right-half plane zero and a PID controller, respectively.

Example 11.1 Developing the Equations for a Bode Plot for a FOPDT Process

Problem Statement. Using the functions listed in Table 11.1, develop the equations that describe the dependence of the amplitude ratio and phase angle of a FOPDT process ($K_p = 2$, $\tau_p = 3$, $\theta_p = 1.5$) on the frequency of the input.

Solution. Using Table 11.1, the amplitude ratio and phase angle for a first-order transfer function and deadtime with the given parameter values are

$$(A_r)_{first\ order} = \frac{2}{\sqrt{9\omega^2 + 1}} \qquad \phi_{first\ order} = -\tan^{-1}(3\omega)$$

$$(A_r)_{deadtime} = 1 \qquad \phi_{deadtime} = -85.9\omega$$

Using Equation 11.2.4,

$$(A_r)_{FOPDT} = (A_r)_{first\ order}(A_r)_{deadtime} = \frac{2}{\sqrt{9\omega^2 + 1}}$$

$$\phi_{FOPDT} = \phi_{first\ order} + \phi_{deadtime} = -\tan^{-1}(3\omega) - 85.9\omega$$

These equations allow you to directly generate a Bode plot for the FOPDT system by specifying a range of frequencies and calculating the corresponding values of the amplitude ratio and the phase angle at each frequency.

Example 11.2 Developing the Equations for a Bode Plot for a PI Controller Applied to a First-Order Process

Problem Statement. Develop the equations that can be used to construct a Bode plot for a PI controller (K_c=2; τ_I=3) applied to a first-order process ($K_p = 0.7$, $\tau_p = 5$).

Solution. Using Table 11.1 for the amplitude ratio and phase angle for each component in this system,

$$(A_r)_{PI} = 2\sqrt{1 + \frac{1}{9\omega^2}} \qquad \phi_{PI} = -\tan^{-1}\left(\frac{1}{3\omega}\right)$$

$$(A_r)_{first\ order} = \frac{0.7}{\sqrt{25\omega^2 + 1}} \qquad \phi_{first\ order} = -\tan^{-1}(5\omega)$$

Then

$$(A_r)_{overall} = \frac{0.467}{\omega}\sqrt{\frac{9\omega^2 + 1}{25\omega^2 + 1}} \qquad \phi_{overall} = -\tan^{-1}\left(\frac{1}{3\omega}\right) - \tan^{-1}(5\omega)$$

Self-Assessment Questions

Q11.2.1 What is the amplitude ratio used in a Bode plot?

Q11.2.2 What is the phase angle used in a Bode plot?

Q11.2.3 Explain what a Bode plot is. Indicate how you would directly generate a Bode plot using the simulator of the CST thermal mixer (Example 3.3).

Self-Assessment Answers

Q11.2.1 The amplitude ratio used in the Bode plot is A_r where

$$A_r = \frac{a_y}{a_c}$$

and a_y is the amplitude of the sinusoidal variation in the CV, y, and a_c is the sinusoidal variation in the controller output, c.

Q11.2.2 The phase angle used in a Bode plot is ϕ in degrees where

$$\phi = -\frac{\omega \Delta t_p}{2\pi} 360$$

and ω is the frequency of the input variations in radians per second and Δt_p is the time difference between peaks of the controller output and the CV.

Q11.2.3 A Bode plot is a plot of the logarithm of the amplitude ratio versus the logarithm of the frequency of the input and a plot of the phase angle versus the logarithm of the frequency of the input. For the CST thermal mixer, a Bode plot could be generated by sinusoidally varying the specified flow rate for stream 1 under open-loop conditions and measuring the resulting amplitude of the measured temperature of the product after it settles to a fixed standing wave for a range of different frequencies. The amplitude ratio is the ratio of the amplitude in the measured temperature divided by the amplitude of the sinusoidal variation in the specification of the flow rate of stream 1. The phase angle is given by

$$\phi = -\frac{\omega \Delta t_p}{2\pi} 360$$

where ω is the frequency of the input variations in radians per second and Δt_p is the amount of time that c lags y_s.

Self-Assessment Problem

P11.2.1 Develop the set of equations that could be used to develop a Bode plot for a second order process (K_p=1, τ_n=2 and ζ=1) with deadtime (θ_p=1).

Self-Assessment Answer

P11.2.1 Using Table 11.1 for a second-order process with deadtime and applying the specified process parameters yields

$$A_r = \frac{K_p}{\sqrt{(1-\tau_p^2\omega^2)^2 + (2\zeta\tau_p\omega)^2}} = \frac{1}{\sqrt{(1-4\omega^2)^2 + 16\omega^2}}$$

$$\phi = \tan^{-1}\left(\frac{-2\zeta\tau_p\omega}{1-\tau_p^2\omega^2}\right) - \theta_p\omega\frac{360}{2\pi} = \tan^{-1}\left(\frac{-4\omega}{1-4\omega^2}\right) - \omega\frac{360}{2\pi}$$

11.3 Bode Stability Criterion, Gain Margin and Phase Margin

Consider the frequency in Figure 11.2.2 that corresponds to a phase angle of -180 , i.e., from Figure 11.2.2, = 0.79 radians per unit time. Figure 11.3.1a is a schematic of a loop with a sinusoidally varying setpoint with an amplitude of one and a frequency of 0.79 radians per second. A P-only controller with a gain of 1.0 is used, but the feedback of the measurement is broken before it is compared with the setpoint value (i.e., an open-loop process). Because the amplitude ratio at = 0.79 radians per second is 0.78, the amplitude of the variation of the measured value of *y* is 0.78. Note that, because the phase angle is -180 , the sinusoidal variation in the measured value of *y* is the negative of the setpoint variation, i.e.,

$$y_{sp} \quad 1.0\sin(0.79t)$$

$$y_s \quad 0.78\sin(0.79t)$$

because

$$\sin(0.79t \quad 180 \ ) \qquad \sin(0.79t)$$

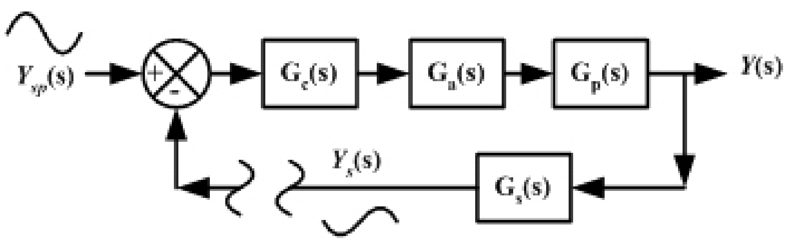


Figure 11.3.1a Block diagram of a feedback loop with the feedback broken and a sinusoidal variation in the setpoint.

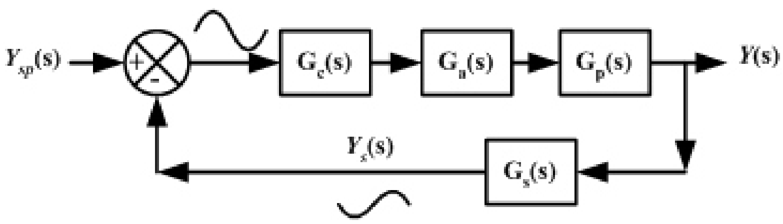


Figure 11.3.1b Block diagram of a feedback loop with simultaneous termination of the setpoint variation and closing of the loop.

Figure 11.3.1b represents simultaneously closing the feedback loop and replacing the sinusoidal variation in the setpoint with a constant setpoint. Note that because the measured value of the CV is subtracted from the setpoint, the variations in y_s are in phase with the original variations in y_{sp}. Once the loop is closed, the sinusoidal variations are fed back around the loop, but because the amplitude ratio is 0.78, the variation in y_s damps out with each subsequent cycle of the loop. The second time around the loop y_s has a sinusoidal amplitude of 0.61 (i.e., 0.78×0.78), the third time has an amplitude of 0.48 and so on. If the amplitude ratio were exactly 1.0, the oscillations would be sustained. If the amplitude ratio were greater than 1.0, y_s would grow without bound (i.e., unstable operation). Therefore, the amplitude ratio at a phase angle of -180 indicates the stability of the system. Therefore, the **Bode stability criterion states that if the open-loop amplitude ratio is greater than unity at a phase angle of -180°, the system is unstable under closed-loop conditions**. The frequency at which the phase angle is equal to -180 is referred to as the **critical frequency**, $_c$.

Figure 11.3.2 shows the Bode plot of the FOPDT process (K_p 1, $_p$ 1, $_p$ 0.5) for a P-only controller in an open-loop configuration. It shows the results for K_c = 1.0 and K_c = 1.28, which corresponds to underdamped stable operation and sustained oscillations at the critical frequency, respectively. The difference between the amplitude ratio for K_c = 1.0 and K_c = 1.28, denoted by *M* in Figure 11.3.2, is a measure of how close the controller with K_c =1.0 is to the onset of instability. The **gain margin** (**GM**) is defined by

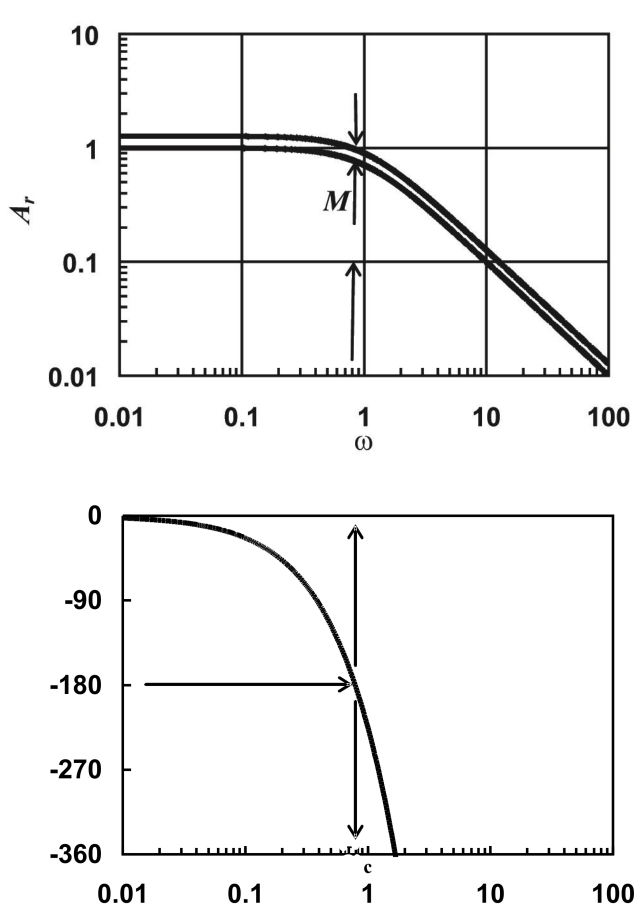


Figure 11.3.2 Bode plot of a FOPDT process that indicates the gain margin.

$$GM = \frac{1}{A_r^*} \qquad \textbf{11.3.1}$$

where A_r^* is the amplitude ratio at the critical frequency, or 0.78 for Figure 11.2.2. In this latter case, the *GM* is equal to 1/0.78 or 1.28. When the *GM* >1, the system is stable. *GM* equal to one corresponds to sustained oscillations. Figure 11.3.3 shows a Bode plot for a second-order process with a P-only controller with K_c = 1.0. The **phase margin** (**PM**), is the difference between a phase angle corresponding to an amplitude ratio of 1.0 and a phase angle of -180°, i.e.,

$$PM = \phi^* - (-180^\circ) = \phi^* + 180^\circ \qquad \textbf{11.3.2}$$

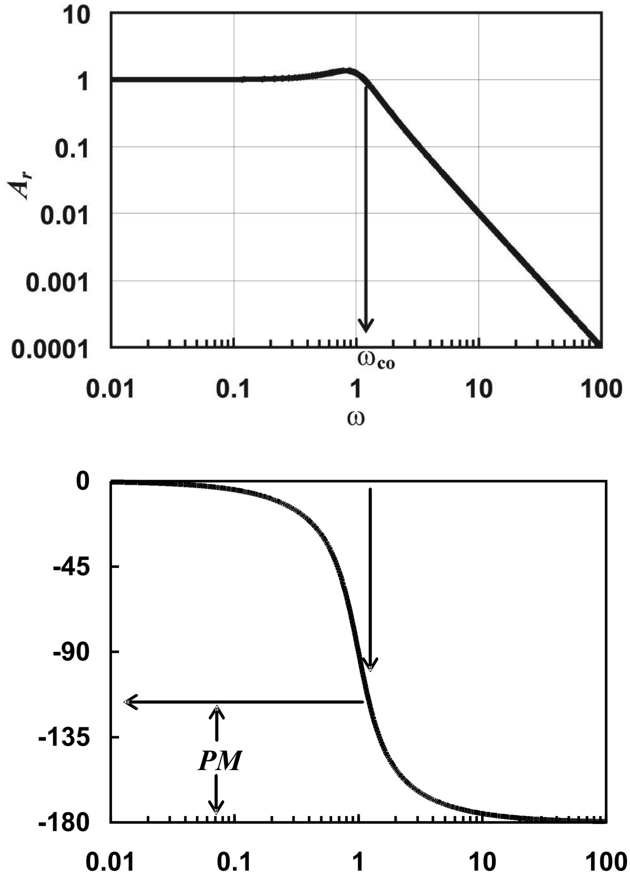
where ϕ^* is the phase angle that corresponds to A_r=1. For Figure 11.3.3, the *PM* is 67.5º. When *PM*>0, the system is stable, when *PM*=0, the system operates under sustained oscillations, and when *PM*<0, the system is unstable. The **crossover frequency**, ω_{co}, is the frequency that corresponds to A_r=1. A gain margin or a phase margin can be used to a tune controller. The larger the *GM* or *PM* values used for tuning, the more conservative the controller tuning. Typical *GM* values used for tuning range from 1.4 to 1.8 while *PM* values typically range from 30° to 45°. *GM* and *PM* can be used similar to how the decay ratio was used in Chapter 9 to select the controller tuning criterion, i.e., the more nonlinear a process and the larger the magnitude of disturbances, the larger the values of *PM* or *GM* that should be used.

Example 11.3 Tuning a Controller for a Specified Gain Margin

Problem Statement. Determine the controller gain, K_c, for a P-only controller applied to a FOPDT process ($K_p = 2$, $\tau_p = 3$, $\theta_p = 1.5$) for a gain margin equal to 1.7.

Solution. From Table 11.1, the amplitude ratio and the phase angle are

$$(A_r)_{overall} = \frac{2K_c}{\sqrt{9\omega^2 + 1}} \qquad \textbf{11.3.3}$$



11.3.3 Bode plot a second-order process that indicates the phase margin.

$$_{overall} \quad \tan^{1}(\;3\;)\quad 85.9$$

First, the critical frequency, $_c$, is determined by setting the overall phase angle equal to (-180º) and solving for the corresponding frequency, i.e.,

$$180 \quad \tan^{1}(\;3\;_c)\quad 85.9\;_c$$

By trial-and-error, the critical frequency is calculated equal to 1.218 radians per unit time.

The gain margin allows for the calculation of the amplitude ratio at the critical frequency (A_r^*), i.e.,

$$A_r^* \quad \frac{1}{1.7} \quad 0.588$$

Finally, the controller gain can be determined by applying Equation 11.3.3 at the critical frequency, i.e.,

$$A_r^* \quad \frac{2K_c}{\sqrt{9\;_c^2\quad 1}} \quad 0.588$$

Rearranging and solving for K_c, results in the controller gain equal to 1.12.

Example 11.4 Tuning a Controller for a Specified Phase Margin

Problem Statement. Determine the controller gain, K_c, for a P-only controller applied to a second-order process (K_p 2, $_p$ 3, 1.5) for a phase margin equal to 40º.

Solution. From Table 11.1 in a manner similar to Examples 11.1 and 11.2, the amplitude ratio and the phase angle can be expressed by

$$(A_r)_{overall} \quad \frac{2K_c}{\sqrt{(1\quad 9\;^2)^2\quad (9\;)^2}}$$

$$_{overall} \quad \tan^{1}\frac{9}{1\quad 9\;^2}$$

The 40º phase margin determines the crossover frequency by applying the equation for the phase angle, Equation 11.3.2

$$-40 - 180 = -\tan^{-1}\left(\frac{9\,\omega_{co}}{1-9\,\omega_{co}^2}\right)$$

$$\tan(-140) = \frac{9\,\omega_{co}}{1-9\,\omega_{co}^2}$$

Rearranging yields $$7.55\,\omega_{co}^2 - 9\,\omega_{co} - 0.839 = 0$$

Solving for the positive root of this equation determines that the crossover frequency is equal to 1.28 radians per unit time. At the crossover frequency, the amplitude ratio is equal to unity; therefore, applying the amplitude ratio equation results in

$$1 = \frac{2K_c}{\sqrt{(1-9\,\omega_{co}^2)^2 + (9\,\omega_{co})^2}}$$

Using the value of the crossover frequency in this equation and solving for the controller gain yields the controller gain equal to 8.95.

Self-Assessment Question

Q11.3.1 Explain the Bode stability criterion in your own words.

Q11.3.2 How are the gain margin and the phase margin alike and how are they different?

Q11.3.3 Explain why the larger the *GM* or the *PM* the more conservative a loop is tuned.

Self-Assessment Answers

Q11.3.1 When the amplitude ratio of the open-loop transfer function including the controller exceeds unity at the critical frequency, then the closed-loop transfer function will be unstable. At the critical frequency (θ = -180º), the feedback of the measurement back into the controller becomes unstable when the amplitude ratio is greater than unity because each time it passes through the feedback loop it will get larger without bound.

Q11.3.2 GM and PM are alike because both measures indicate how close a closed-loop process is to instability. In fact, the smaller that either GM or PM are, the closer the control loop is to the onset of instability. Sustained oscillations (i.e., the onset of instability) occur when the GM is one while a PM of zero corresponds the onset of instability. Typical GM's used for tuning range from 1.4 to 1.8 while typical PM's used for tuning range from 30º to 45º. Certain processes (e.g., a FOPDT model) has a GM but does not have a PM because it does not have an amplitude ratio of unity. Likewise, certain processes (e.g., a second order process) has a PM but do not have a GM because there is not a phase angle of -180º.

Q11.3.3 When the GM is 1.0 and the PM is 0º, the closed-loop system is at sustained oscillation. As GM or PM is increased, the closed-loop system becomes less aggressively tuned.

Self-Assessment Problem

P11.3.1 Determine the controller gain for a P-only controller applied to a gain plus deadtime model (K_p=1 and θ_p=1) if the *GM* is 1.5.

Self-Assessment Answer

P11.3.1 First, determine the critical frequency

$$180^{o} \qquad \frac{360}{2} ; \qquad {}_{c}$$

$$A_r \qquad \frac{K_c K_p}{} \quad \frac{K_c}{}$$

A_r^* 1.0/ 1.5=0.667; applying the amplitude equation at the critical frequency yields

$$0.667 \quad A_r^* \quad \frac{K_c}{} \quad \frac{K_c}{} ; \; K_c \quad 2.09$$

11.4 Pulse Tests

Earlier in this chapter, it was shown that Bode plots can be developed by testing the process directly and using the transfer function of the process. A **pulse test** is an experimental approach that can also be used to generate a Bode plot of an industrial process without directly using transfer functions. The process considered here is the combined system of the actuator, process and sensor, i.e.,

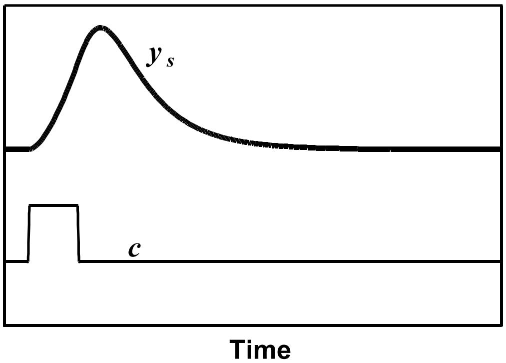


Figure 11.4.1 Example of a pulse test.

$$\overline{G}_p(s) \qquad G_a(s)\, G_p(s)\, G_s(s)$$

The input to this process is the output of the controller. For a pulse test, a rectangular pulse (Chapter 6) is used and the resulting measured values of the CV are recorded (e.g., Figure 11.4.1). This is an open-loop test and y_s should return to or near its starting point in the response time of the process if no significant disturbance occurs during the test. The transfer function for this process is given by

$$\overline{G}_p(s) \qquad \frac{Y_s(s)}{C(s)} \qquad \frac{{}_0 \; y_s(t)\, e^{\;st}\, dt}{{}_0 \; c\,(t)\, e^{\;st}\, dt} \tag{11.4.1}$$

which is based upon the definition of the Laplace transform (Chapter 4) and where $y_s(t)$ and $c\,(t)$ are deviation variables. The transfer function can be converted into a Bode plot by substituting $s = i$. Thus

$$\overline{G}_p(i \;\;) \qquad \frac{{}_0 \; y_s(t)\, e^{\;i\;\;t}\, dt}{{}_0 \; c\,(t)\, e^{\;i\;\;t}\, dt} \tag{11.4.2}$$

Using the Euler identity

$$e^{\;i\;\;t} \qquad \cos \;\; t \quad i \sin \;\; t$$

results in

$$\overline{G}_p(i\omega) = \frac{A(\omega) - iB(\omega)}{C(\omega) - iD(\omega)} \tag{11.4.3}$$

where

$$\begin{aligned} A(\omega) &= \int_0^\infty y_s(t)\cos\omega t\,dt \\ B(\omega) &= \int_0^\infty y_s(t)\sin\omega t\,dt \\ C(\omega) &= \int_0^\infty c(t)\cos\omega t\,dt \\ D(\omega) &= \int_0^\infty c(t)\sin\omega t\,dt \end{aligned} \tag{11.4.4}$$

After multiplying by the complex conjugate of the denominator,

$$R(\omega) = \frac{A(\omega)C(\omega) + B(\omega)D(\omega)}{C^2(\omega) + D^2(\omega)} \tag{11.4.5}$$

$$I(\omega) = \frac{A(\omega)D(\omega) - B(\omega)C(\omega)}{C^2(\omega) + D^2(\omega)} \tag{11.4.6}$$

where $R(\omega)$ and $I(\omega)$ are the real and imaginary components of $\overline{G}_p(i\omega)$, respectively. Then, finally, the amplitude ratio, A_r, and the phase angle, ϕ, can be calculated directly

$$A_r(\omega) = \left|\overline{G}_p(i\omega)\right| = \sqrt{R^2(\omega) + I^2(\omega)} \tag{11.4.7}$$

$$\phi(\omega) = \angle\overline{G}_p(i\omega) = \tan^{-1}\frac{I(\omega)}{R(\omega)} \tag{11.4.8}$$

After the experimental pulse test is generated, Equations 11.4.3 to 11.4.6 are applied at each value of ω to generate the Bode plot. A value ω_1 is selected and the values of $A(\omega_1)$, $B(\omega_1)$, $C(\omega_1)$ and $D(\omega_1)$ [Equation 11.4.4] are calculated using the pulse test results and a numerical integration method (e.g., the trapezoidal method[1]). Then, $R(\omega_1)$ and $I(\omega_1)$ are calculated using Equations 11.4.5 and 11.4.6. Finally, $A_r(\omega_1)$ and $\phi(\omega_1)$ are determined from Equations 11.4.7 and 11.4.8. Another frequency is selected and the procedure is repeated until the Bode plot is complete.

Once the Bode plot is generated, the Bode plot of a P-only, PI, or PID controller can be combined with it and used to tune a controller to meet gain margin or phase margin specifications. The Bode plot of the controller can be plotted on the Bode diagram developed from the experimental pulse test. For a set of tuning parameters, the A_r*'s* and ϕ*'s* for the controller and the process can be added together to yield the overall Bode plot. A_r*'s* are added because they are plotted as logarithms and adding logarithms is equivalent to multiplying the A_r*'s*. In this manner, controller tuning parameters can be adjusted until the desired *GM* or *PM* is obtained.

This approach[2] was first used in the 1960's because it allowed for a systematic procedure to tune PID controllers applied to complex industrial processes. This approach to tuning suffers from the following limitations.

1. It requires an open-loop response time to complete the pulse test.
2. Disturbances during the test can corrupt the results.
3. Bode plots developed by this approach can be noisy, particularly around the crossover frequency, affecting the accuracy of the resulting *PM* or *GM* used for tuning.

The ATV method of identification with on-line tuning (Section 9.9) can be applied with much less time and effort and yields more accurate results. As a result, ATV-based tuning is recommended over pulse-based tuning for slow-responding loops.

Example 11.5 Pulse Test for the CST Thermal Mixer

Problem Statement. Using the model for the CST thermal mixer (Example 3.3), apply a pulse test.

Solution. Using the MatLab or the Python version of the simulator, a pulse test can be applied to the CST thermal mixer. The pulse test for a +0.25 kg/s change in F_1 at t=5 s followed by a -0.25 kg/s change at t=35s is shown in Figure 11.4.2.

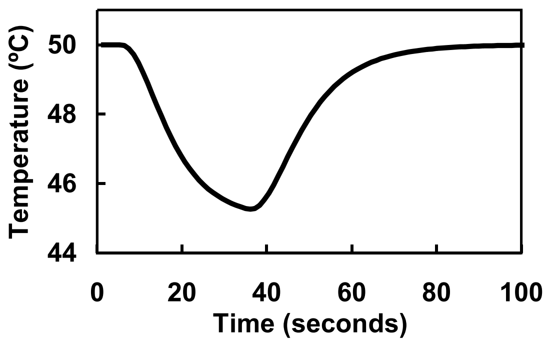


Figure 11.4.2 Pulse test for CST thermal mixer.

Self-Assessment Question

Q11.4.1 Describe how a pulse test can be used to tune a controller on an industrial process. Why would you not use this approach on an industrial process?

Self-Assessment Answer

Q11.4.1 To apply a pulse test, the process should be at or very near to steady-state conditions. A rectangular pulse in applied to the input to the process (i.e., the controller output) and the resulting measured values of the output variable are recorded until the process returns to its original condition. Note that if there is significant offset between where the process variable started and finished the test, an unmeasured disturbance most likely entered the process during the pulse test; therefore, the pulse test should be repeated. After the experimental pulse test is generated, Equations 11.4.4 to 11.4.8 are applied at each value of to generate the Bode plot. That is, a value of is selected and the values of *A*($_1$), *B*($_1$), *C*($_1$), and *D*($_1$) are calculated using the pulse test results and a numerical integration method (e.g., the trapezoidal method). Then *R*($_1$) and *I*($_1$) are calculated using Equations 11.4.5 and 11.4.6. Finally, A_r($_1$) and (are determined from Equations 11.4.7 and 11.4.8. Another frequency is selected and the procedure is repeated until the Bode plot is completed. Then the Bode plot can be used to tune the controller using the GM or PM. This approach to controller tuning is not recommended for tuning industrial control loops because there are faster and more accurate ways to tune industrial control loops (e.g., Section 9.9).

11.5 Nyquist Diagrams

The **Nyquist diagram** is an alternate method for presenting frequency response behavior. Bode plots present separate curves for the amplitude ratio and phase angle as a function of frequency. The Nyquist diagram presents the frequency response behavior in a more compact form, i.e., with a single curve. At a specific frequency, the real and imaginary portion of $G_p(i\ \ )$ [Equation 11.1] defines a point on the complex plane. When these points are plotted on the complex plane for a range of frequencies, the result is a Nyquist diagram. Because the Nyquist diagram plots the real and imaginary components of $G_p(i\ \ )$, the points on a Nyquist diagram can also be defined in terms of the amplitude ratio, A_r, and the phase angle,

$$R(\ \ )\quad A_r \cos \qquad I(\ \ )\quad A_r \sin \tag{11.5.1}$$

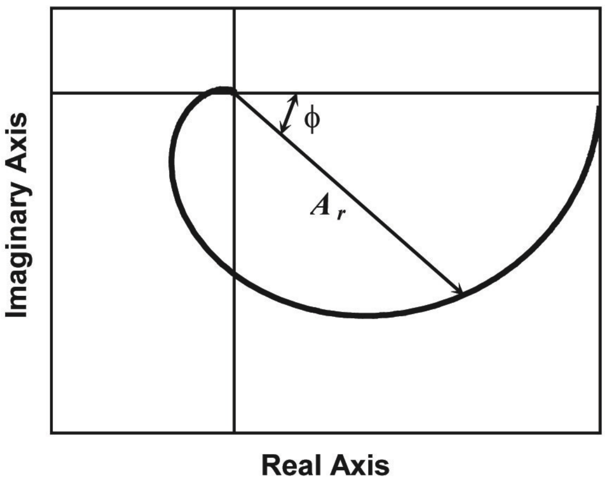


Figure 11.5.1 A Nyquist diagram showing the relationship between A_r and and a point on the diagram.

Figure 11.5.1 shows the relationship between a point on a Nyquist diagram and the amplitude ratio and phase angle. The Nyquist diagram is generated by varying the frequency over a given range. The methods previously discussed for generating Bode plots can be used to generate a Nyquist diagram by applying the previous equations using A_r, and over a range of frequencies and using Equation 11.5.1. Nyquist diagrams, like Bode plots, can be used to tune controllers and analyze closed-loop stability.

Figure 11.5.2 shows a Nyquist diagram for a FOPDT process model (K_p 0.45, $_p$ 1, $_p$ 0.5). The magnitudes of the real and imaginary components of $G_p(i\ \ )$ are plotted in a complex plane for a range of values of frequency, . Each point on the Nyquist plot corresponds to a different frequency. Note that the Nyquist plot of a

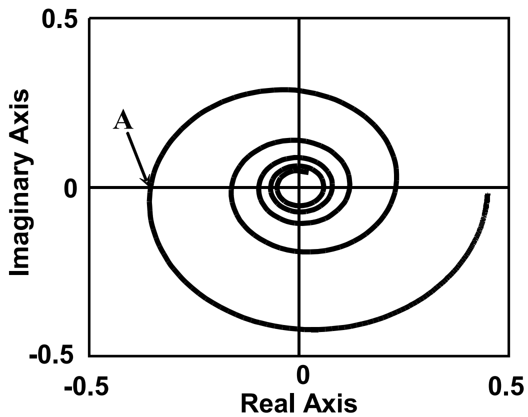


Figure 11.5.2 Nyquist diagram of a FOPDT process.

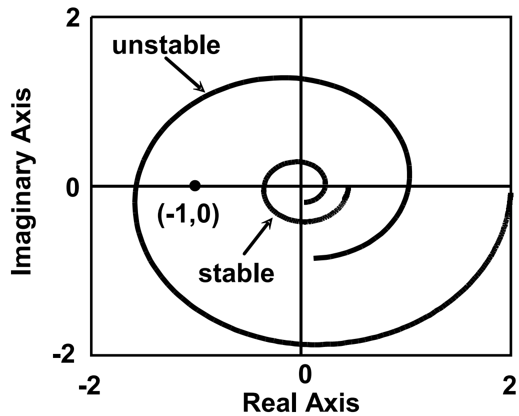


Figure 11.5.3 Nyquist plot that demonstrates the Nyquist stability criterion.

FOPDT process model spirals inward towards the origin of the complex plane. Even though it is not shown in Figure 11.5.2, the curve asymptotically approaches the origin of the complex plane as the frequency is increased.

The **Nyguist stability criterion: If the Nyquist plot of the overall open-loop transfer function encircles the (-1,0) point on the complex plane, as frequency goes from zero to , the system will be unstable under closed-loop conditions.** Figure 11.5.3 shows the Nyquist plot for a stable and an unstable system. Note that the Nyquist plot of the stable system spirals inward toward the origin of the complex plane without encircling the point (-1,0). On the other hand, for the unstable system, the Nyquist plot encircles the point (-1,0). The point where the Nyquist plot first crosses the *x*-axis (e.g., point A in Figure 11.5.2) indicates how close the system is to instability. In addition, the gain margin (*GM*) is equal to the inverse of the absolute value of the point where the Nyquist plot first crosses the *x*-axis. For the unstable case in Figure 11.5.3, the *GM* is less than unity, indicating that the system will be unstable under closed-loop conditions. In addition, if the Nyquist plot were to pass through the point (-1,0), the gain margin would be equal to unity, indicating the system would produce sustained oscillations under closed-loop conditions. Also, for the stable case shown in Figure 11.5.2, the gain margin is in excess of 2. The PM can also be determined from a Nyquist plot but is not shown here.

Self-Assessment Questions

Q11.5.1 Explain how you can generate a Nyquist diagram for a P-only controller applied to a second-order plus deadtime process using Table 11.1.

Q11.5.2 How can you determine the GM from a Nyquist diagram of the process and the controller?

Self-Assessment Answers

Q11.5.1 First, using Table 1.1 develop equations for the amplitude ratio and phase angle as functions of frequency for the controller and process lumped together. Then for each frequency over the desired range of frequencies, Equation 11.5.1 can be used to generate a series of points on the complex plane, which is the Nyquist diagram.

Q11.5.2 The inverse of the absolute value of the point where the Nyquist diagram first crosses the x-axis is equal to the GM.

11.6 Closed-Loop Frequency Response

A Bode plot provides frequency response information about a process or a process and a controller in an open-loop form. It is also informative to consider the frequency dependence of a closed-loop system for disturbance upsets.

Figure 11.6.1 is a schematic of a closed-loop feedback system subjected to a sinusoidally-varying disturbance. Figure 11.6.2 shows the **closed-loop frequency response** for a FOPDT process model (K_p 1, $_p$ 1, $_p$ 0.5) and a FOPDT disturbance model (K_d 1, $_d$ 1, $_d$ 0.5) for sinusoidal disturbance changes with a P-only controller (K_c = 1). Note that A_r is plotted on a linear scale in Figure 11.6.2. **The closed-loop amplitude ratio, A_r, is defined here as the ratio of the amplitude of the variations in the CV divided by the amplitude of the variations in the disturbance.** The phase angle is the phase lag between the disturbance and the CV.

At high frequencies, A_r drops off sharply because the process is not fast enough to respond to high-frequency variations in the disturbance level, and the variations become filtered out (i.e., averaged out by the process). At

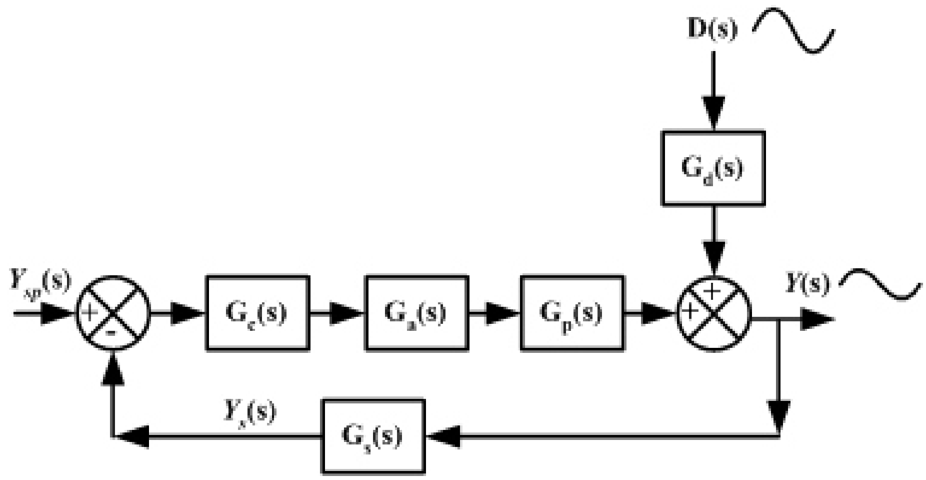


Figure 11.6.1 Block diagram of a feedback loop that is excited by a sinusoidal disturbance.

low frequencies, A_r drops off as well because, for very slow-varying disturbances, the feedback controller has time to absorb the disturbance and maintain operation at the setpoint. At frequencies between these extremes, is large enough that the feedback controller is unable to remove all the variations and is small enough that the process does not filter out the variations in the disturbance. The peak in the closed-loop frequency response is called the **closed-loop peak amplitude**. The **peak frequency** is the disturbance frequency that corresponds to the closed-loop peak amplitude and represents the frequency at which the maximum sensitivity to the disturbance occurs.

Industrial feedback controllers exhibit the same general behavior as shown in Figure 11.6.2. There is a range of disturbance frequencies for which a controller is most sensitive. Analyzing the peak frequencies of the individual loops of a number of processing units in a series can provide insight into the disturbance rejection performance of the overall system.

Consider two distillation columns in series (Figure 11.6.3) for which the bottoms of the first column is the feed to the second column. First, consider the case in which the peak frequency for the bottom loop of the first column is equal to the peak frequency for both the top and bottom composition loops of the second column. In this case, the disturbance frequencies for which the two loops on the second column are most sensitive are the same as the frequencies for which the bottom loop on the first column is most sensitive. Therefore, the largest variations in the bottom product are expected to significantly affect both loops on the second column.

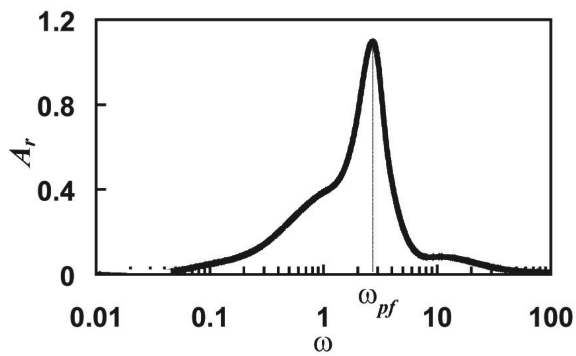


Figure 11.6.2 The closed-loop amplitude ratio for a FOPDT process. Note that A_r is plotted on a linear scale instead of the log scale used for a Bode plot. $_{pf}$ is the peak frequency.

Next, consider the case in which the peak frequency of the bottom loop on the first column is significantly lower than the peak frequency for the two loops on the second column (Figure 11.6.4). For this case, the control loops on the second column should be able to handle the largest

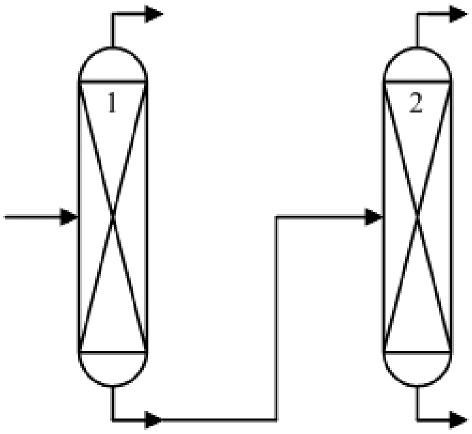


Figure 11.6.3 Two distillation columns in series.

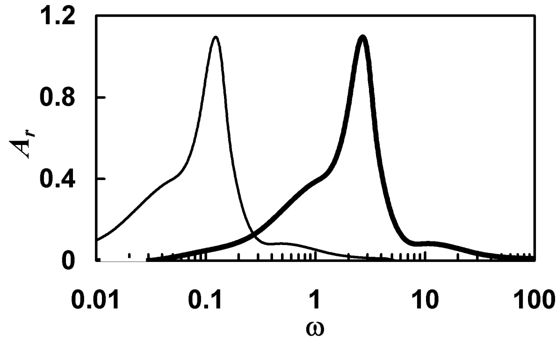


Figure 11.6.4 Closed-loop amplitude ratios for two control loops with very different peak frequencies values.

variations coming from the first column. From Figure 11.6.4, you can see that the disturbances that pass through the bottoms controller on the first column do not significantly affect the control loops for the second column. The controllers on the second column have sufficient time to handle these frequencies.

Finally, assume that the bottom loop of the first column is much faster than the control loops on the second column. For this case, the largest variations coming from the first column are filtered out by the slower responding second column. Qualitatively using closed-loop frequency analysis can be helpful in analyzing the propagation of disturbances through a sequence of processing units.

Self-Assessment Question

Q11.6.1 What is the peak frequency of a control loop and what is its significance?

Self-Assessment Answer

Q11.6.1 The peak frequency for a closed-loop system is the frequency corresponding to the maximum value of the amplitude ratio. The peak frequency indicates the frequency of disturbances for which a control loop is most sensitive.

11.7 Summary

- The frequency or time scale of inputs to a process can have a significant effect on the resulting closed-loop behavior.
- Bode plots, which are plots of log(A_r) versus log( ) and versus log( ), are a convenient means to present the frequency-dependent characteristics of a process. Bode plots can be generated by (1) directly exciting the process with sinusoidal inputs, (2) applying a sinusoidal input to the transfer function of the process, (3) substituting $s = i$ into the transfer function and (4) applying a pulse test.
- The Bode plot of the combined process and controller can be used to determine the closed-loop stability of the system using the Bode stability criterion. The gain margin and the phase margin can be used as tuning specifications for PID controllers.

- Nyquist diagrams are another way to represent the frequency dependent behavior of a process. The Nyquist diagram is a plot of the real and imaginary components of $G_p(i$) on a complex plane for a range of frequencies.
- The closed-loop frequency response of a process indicates the disturbance frequencies for which the controller is most sensitive. This information can be used to analyze how disturbances damp out or propagate from one control loop to another through a sequence of processes.

11.8 References

1. Riggs, J.B., *Computational Methods for Engineers with MATLAB Applications*, Ferret Publishing, Austin, Texas, pp. 240-244 (2013).
2. Hougen, J.O., *Methods for Solving Process Plant Problems*, Instrument Society of America (1996).

11.9 Additional Terminology

Amplitude ratio - (A_r) the ratio of the amplitude of the variations in y_s divided by the amplitude of the sinusoidal variations in an input to the process.
Bode plot - a plot of log(A_r) versus log( ) and a plot of versus log( ).
Bode stability criterion - states that a process is stable if $A_r < 1$ at the critical frequency.
Closed-loop frequency response - a Bode plot for a closed-loop process subjected to sinusoidally varying disturbance inputs.
Closed-loop peak amplitude - the maximum A_r for a closed-loop frequency response.
Critical frequency - the frequency that corresponds to = -180
Crossover frequency - the frequency that corresponds to A_r=1.
Frequency response analysis - the study of the effect of varying input frequencies on process behavior.
Gain margin (GM) - defined by Equation 11.5 and indicates how aggressively a controller is tuned.
Nyquist diagram - a polar-coordinate plot of $G_p(i$ for a range of frequencies.
Peak frequency - the frequency that corresponds to the closed-loop peak amplitude.
Phase angle - ( ) an indication of how much the controller output lags behind the CV.
Phase margin (*PM*) - defined by Equation 11.6 and indicates how aggressively a controller is tuned.
Pulse test - the response of a process to a rectangular pulse input that can be used to develop a Bode plot for an experimental system.

11.10 Preliminary Questions

11.2 Bode Plots

Q11.2.1 Even though frequency response analysis is not generally used by industrial process control engineers, why is frequency response analysis important to the understanding of feedback systems?

Q11.2.2 Explain, using equations, how you would generate a Bode plot from the transfer function of a process.

Q11.2.3 What information does a Bode plot contain?

Q11.2.4 How do you calculate the amplitude ratio of a sequence of transfer functions?

Q11.2.5 How do you calculate the phase angle of a sequence of transfer functions?

Q11.2.6 What five methods shown in this chapter can be used to generate a Bode plot?

11.3 Bode Stability Criterion, Gain Margin and Phase Margin

Q11.3.1 Using the results from Table 11.1, describe how the *PM* could be used to tune a PI controller applied to a second-order process.

11.4 Pulse Tests

Q11.4.1 Explain how a pulse test can be used to identify the Bode plot of an experimental process.

Q11.4.2 Describe how a pulse test can be used to tune a controller on an industrial process. Why would you not use this approach on an industrial process?

11.5 Nyquist Diagrams

Q11.5.1 Define what a Nyquist diagram is and how it can be generated.

11.6 Closed-Loop Frequency Response

Q11.6.1 Define what a closed-loop Bode plot is and explain how it can be used to determine the propagation of variability from one processing unit to another.

Q11.6.2 What does the peak frequency of a controller indicate?

Q11.6.3 What input is varied to generate a closed-loop Bode plot?

Q11.6.4 What information does a closed-loop Bode plot provide?

11.11 Analytical Questions and Exercises

11.2 Bode Plots

P11.2.1** Using Table 11.1, develop equations, in terms of the frequency of the input, for the amplitude ratio and phase angle:

a. For a PI controller (K_c=4; $_I$=5) applied to a FOPDT process (K_p 0.4; $_p$ 6; $_p$ 1).

b. For a PI controller (K_c=1; $_I$=5) applied to an integrating process (K_p=-3).

c. For a PI controller (K_c=13; $_I$=10) applied to a second-order plus deadtime process (K_p 0.1; $_p$ 30; 4; $_p$ 15).

d. For a PI controller (K_c=3; $_I$=1) applied to a FOPDT process (K_p 2; $_p$ 4; $_p$ 2).

P11.2.2** Derive the amplitude and phase angle as a function of frequency for a PID controller using Equation 11.2.1. Compare your answer to the results listed in Table 11.1.

11.3 Bode Stability Criterion, Gain Margin and Phase Margin

P11.3.1** Determine the controller gain for the following process models and gain margin specifications.

a. For a PI controller ($_I$=5) applied to a FOPDT process (K_p 0.4; $_p$ 6; $_p$ 1) with a gain margin equal to 1.6.

b. For a P-only controller applied to a second-order plus deadtime process (K_p 0.25; $_p$ 30; 2; $_p$ 10) with a gain margin equal to 1.8.

c. For a P-only controller applied to a second-order plus deadtime process (K_p 25; $_p$ 3; 1.5; $_p$ 2) with a gain margin equal to 1.5.

P11.3.2** Determine the controller gain for the following process models and phase margin specifications.

a. For a PI controller ($_I$=5) applied to a FOPDT process (K_p 0.4; $_p$ 6; $_p$ 1) with a phase margin equal to 30º.

b. For a P-only controller applied to a second-order plus deadtime process (K_p 0.25; $_p$ 30; 2; $_p$ 10) with a phase margin equal to 35º.

c. For a P-only controller applied to a second-order plus deadtime process (K_p 25; $_p$ 3; 1.5; $_p$ 2) with a phase margin equal to 40º.

P11.3.3** Determine the change in the controller gain corresponding to a change in the gain margin from 1.6 to 1.4 for Problem 11.16a.

P11.3.4** Determine the change in the controller gain corresponding to a change in the phase margin from 30º to 45º for Problem 11.17a.

P11.3.5** Consider a P-only controller (K_c=2) applied to a FOPDT process (K_p 2; $_p$ 3; $_p$ 1). Determine the gain margin for this feedback system.

P11.3.6** Consider a P-only controller (K_c=0.04) applied to a second-order process (K_p 3; $_p$ 30; 2). Determine the phase margin for this feedback system.

P11.3.7*** Explain how the reset time for a PI controller can be calculated to meet a specific gain margin if the controller gain is known.

P11.3.8*** Explain how the reset time for a PI controller can be calculated to meet a specific phase margin if the controller gain is known.

P11.3.9** Consider a second-order model of a process without deadtime. Is there a gain margin for a controller applied to this model? If not, explain.

P11.3.10** Consider a first-order model of a process without deadtime. Is there a phase margin for a controller applied to this model? If not, explain.

P11.3.11** A process is described by a FOPDT model, with K_p = 10, = 100 min, = 3 min.

a) Find the maximum value of gain (for a P-only controller) that can be used for this process, before it is unstable. What are the critical frequency and the amplitude ratio at this gain?

b) Use the Zeigler Nichols tuning parameters to find the PID controller settings for this process, noting that ultimate period, P_u = 2 $_c$; where $_c$ is the critical frequency.

11.5 Nyquist Diagrams

P11.5.1** Using Table 11.1, develop a Nyquist plot for a first-order process (i.e., K_p 2; $_p$ 3).

P11.5.2** Using Table 11.1, develop a Nyquist plot for a third-order process, i.e.,

$$G_p(s) \quad \frac{2}{(0.5s \quad 1)(s \quad 1)(3s \quad 1)}$$

P11.5.3** For Problem 11.28, if a P-only controller is used with K_c=1.0, find the GM and PM. Show these on a Nyquist plot.

Part IV

Advanced PID Control

Chapter 12

Cascade, Ratio and Feedforward Control

Chapter Objectives

- Present the advantages of cascade control along its requirements.
- Introduce the benefits of ratio control and illustrate when dynamic compensation is required.
- Present a general equation for a feedforward controller and show its relation to a lead-lag element.
- Compare the advantages and disadvantages of feedback and feedforward control.
- Demonstrate how feedback and feedforward control compliment each other.

12.1 Introduction

This chapter considers cascade, ratio and feedforward control. The primary advantages of cascade, ratio and feedforward control are related to their ability to reject disturbances more effectively than conventional PID controllers. Cascade control rejects specific types of disturbance upsets, which can either be measured or unmeasured. Ratio control effectively handles measured feed flow rate disturbances for a wide range of processes. Feedforward control is a general methodology for directly compensating for measured disturbances.

When a disturbance upsets a conventional PID control loop, all the correction must come from feedback action. No corrective action is taken until the disturbance has affected the process. Return to the everyday control example of driving a car, which was presented in Chapter 1. When driving a car, if the driver looks only at the car's position on the road when negotiating a turn (e.g., looking right in front of the car), the safe car speed through a turn is greatly reduced compared with the feedforward approach wherein the driver anticipates an upcoming curve. Cascade, ratio and feedforward control provide performance enhancement for process control systems because, in each case, corrective action is taken before the disturbance has significantly affected the process. As a result, the amount of corrective action required from the PID controller, the resulting maximum deviation from setpoint and response time of the feedback system can each be significantly reduced.

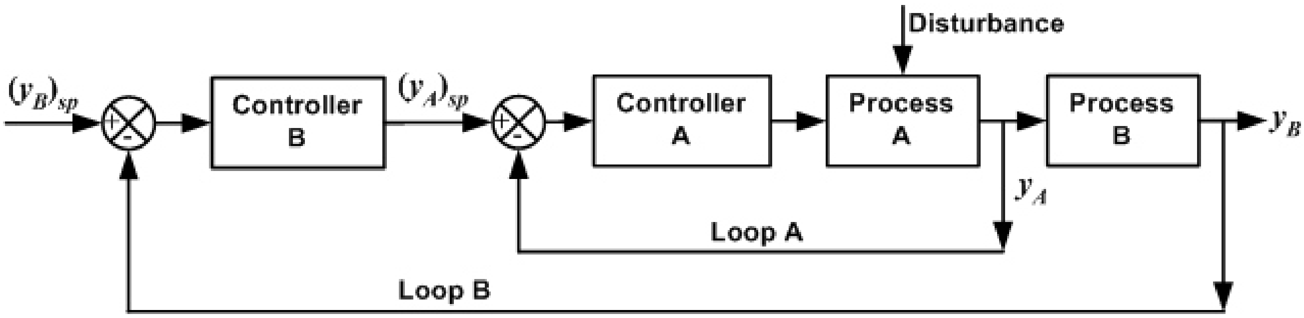


Figure 12.2.1 Block diagram of a cascade control loop. Controller A is the secondary controller and controller B is the primary controller.

12.2 Cascade Control

Cascade control offers a means of reducing the effect of certain disturbances on the primary control objective of a control loop. Cascade control uses two control loops in tandem (Figure 12.2.1). The inner loop (A) receives its setpoint from the outer loop (B). The inner loop is used to react to certain disturbances on a high-frequency basis before these disturbances can significantly upset the process by maintaining y_A at its setpoint. The outer loop is applied to maintain the primary control objective (y_B) on setpoint. The inner loop is called the **secondary loop**, and the outer loop is called the **primary loop**. For an effective cascade arrangement, y_A in Figure 12.2.1 must have a strong and immediate effect on y_B, and the closed-loop dynamics of the secondary loop must be much faster than those of the primary loop. In most cases, maintaining y_A at its setpoint should reject specific disturbances, but in certain cases (e.g., Example 12.3), cascade control is used because the CV (i.e., y_B in Figure 12.2.1) for the primary loop responds best to changes in the CV for the secondary loop (i.e., y_A in Figure 12.2.1).

Example 12.1 Cascade Control to a Flow Control Loop

Problem Statement. The most common application of cascade control in the CPI involves using a flow control loop as the secondary loop. That is, the majority of control valves in the CPI are applied in flow control loops where the primary loop (e.g., level control loops, temperature control loops or composition control loops)

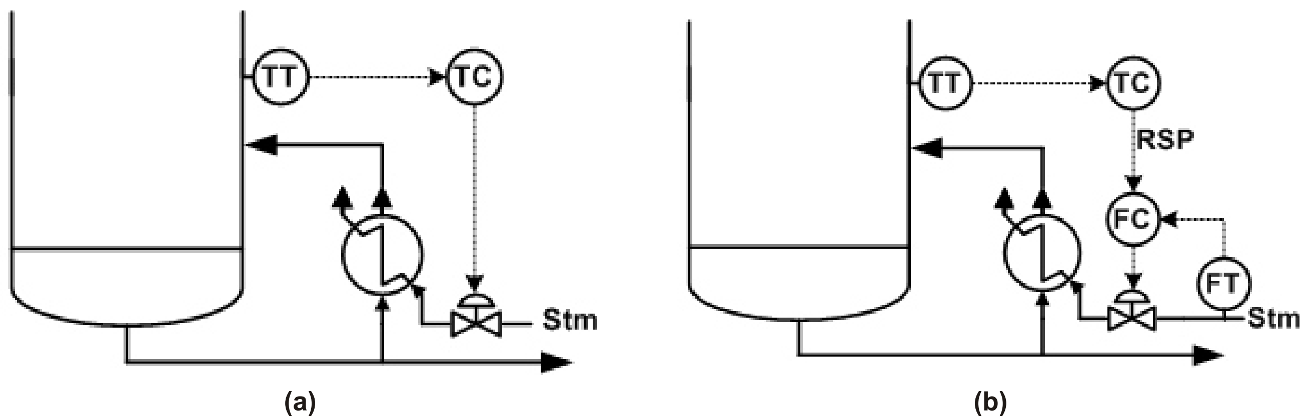


Figure 12.2.2 Control of a tray temperature in the stripping section of a column (a) without cascade control and (b) with cascade control.

provides the setpoint for the flow control loop. Compare the control of a tray temperature in the stripping section of a distillation column with and without cascade control (Figure 12.2.2).

Solution. First consider the control loop without cascade control (Figure 12.2.2a). In this case, the output of the temperature controller goes directly to the valve on the steam line. The control valve will likely have a valve positioner, but this arrangement is susceptible to changes in the steam supply pressure. That is, as the steam pressure changes, the flow rate of steam to the reboiler will change. These changes in the steam flow rate to the reboiler will affect the separation produced by the column, and therefore, the tray temperature of interest. The temperature controller will respond to this disturbance after it has affected the separation produced by the column and the bottom product composition.

Now, consider the cascade approach for this system (Figure 12.2.2b). For a steam supply pressure upset, the cascade arrangement will immediately recognize the change in the steam flow rate caused by the steam pressure change and adjust the valve position for the control valve to return the steam flow to the flow rate specified by the temperature controller. In this manner, the flow control loop is able to eliminate a common process disturbance (i.e., steam pressure upsets) in a relatively short time before this disturbance can significantly affect the process. Note that the cascade arrangement for this process (Figure 12.2.2b) will not offer any advantage over the case without cascade (Figure 12.2.2a) for other types of disturbances (e.g., column feed flow rate changes or feed composition changes)

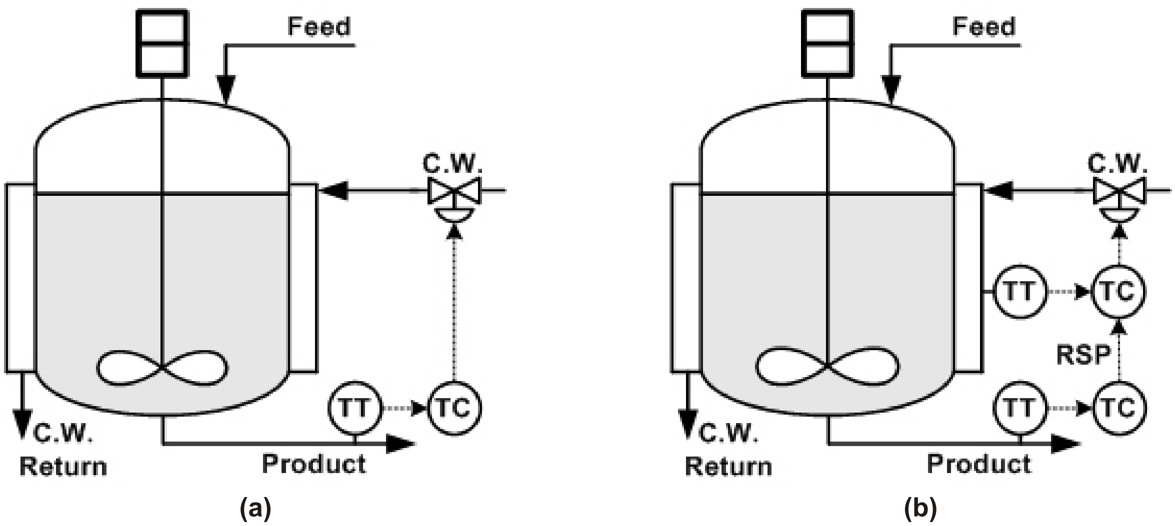


Figure 12.2.3 Control diagrams for a temperature control loop for an exothermic CSTR. (a) Without cascade control. (b) With cascade control.

Example 12.2 An Exothermic CSTR with a Water-Jacketed Heat Exchanger

Problem Statement. Consider the exothermic CSTR with a water-jacketed heat exchanger without cascade control shown in Figure 12.2.3a. Apply cascade control and qualitatively analyze the control performance with and without cascade control.

Solution. Considering the control configuration without cascade (Figure 12.2.3a), assume that a significant change in the temperature of the cooling water occurs. A change in the inlet cooling water temperature will cause a change in the reactor temperature before the reactor temperature controller can compensate for this disturbance. Figure 12.2.3b shows the CSTR with a cascade configuration. The control loop on the jacket cooling water temperature is the secondary control loop while the temperature control loop on the product is the primary loop. When an inlet cooling water temperature change occurs, a change in the jacket water temperature occurs before the reactor temperature starts to change. The secondary loop (the water-jacket temperature loop) can react quickly to inlet cooling water temperature changes, thus significantly reducing the effect of cooling water temperature changes on the reactor temperature.

Example 12.3 Bio-Cascade Example

Problem Statement. pH has a direct effect on the transport of substrates (e.g., glucose) through the cell membranes of the microorganisms in a bio-reactor. The transport rate of substrates affects the consumption rate of the substrate by the cell, and as a result, the cellular metabolic rate. Based on this description, develop a cascade controller for controlling the substrate concentration in a fed-batch bio-reactor and analyze its performance.

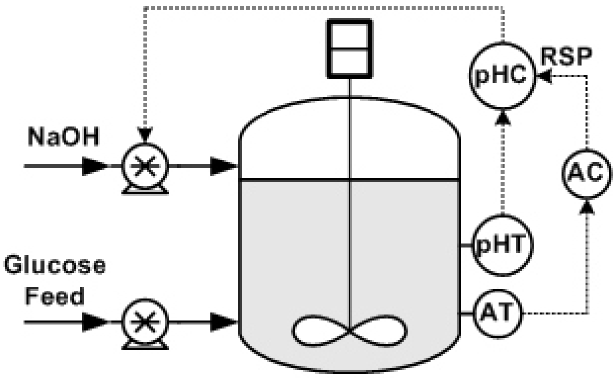


Figure 12.2.4 Control diagram for a cascade controller applied for glucose composition control for a fed-batch bio-reactor.

Solution. Because the broth pH directly affects the glucose consumption rate, the broth pH level can be adjusted by a glucose concentration controller to maintain a specified substrate concentration as measured by an on-line glucose analyzer (AT). And the pH controller, which is a much faster responding loop than the glucose concentration loop, receives its setpoint from the glucose composition controller. Therefore, the pH loop is the secondary control loop and the glucose composition loop is the primary control loop in this case. The control diagram for this control system is shown in Figure 12.2.4. It is important for this application to ensure that the broth pH remains within specific pH limits or microorganisms may go into shock and become dormant or die. Note that in this case, the application of cascade control is not to reject specific disturbances, but to use a MV for which the process is responsive.

Example 12.4 Multiple Cascade Arrangement

Problem Statement. Consider the stripping section of a distillation column with multiple cascade control loops applied as shown in Figure 12.2.5. Analyze the control performance differences for this process with and without cascade control.

Solution. First, recognize that Figure 12.2.5 is an extension of Figure 12.2.2b, i.e., an additional composition control loop has been added on top of the tray temperature control loop. Figure 12.2.5 shows a multiple cascade configuration that is designed to maintain the impurity level in the bottoms product of a distillation column at its setpoint. The innermost loop is flow control on the steam to the reboiler. The flow controller provides fast

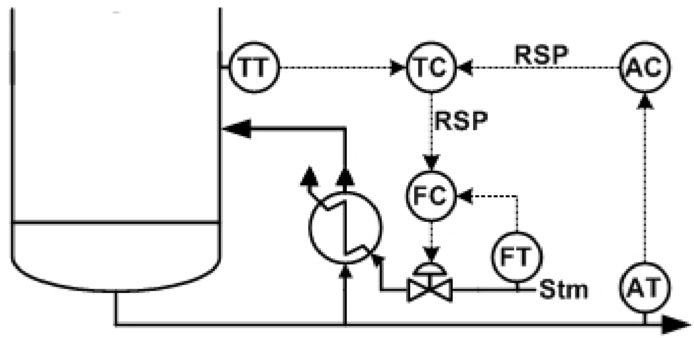


Figure 12.2.5 Control diagram for a multiple cascade configuration applied for bottoms composition control of a distillation column.

response to steam pressure changes. The setpoint for the flow control loop is the MV of the intermediate loop (i.e., tray temperature control loop.

Tray temperature correlates strongly with product composition for a large class of industrial columns. Utilizing such correlations is the basis of inferential control (Section 13.2). The advantage of controlling tray temperatures on distillation columns is that composition changes resulting from process disturbances (e.g., column feed flow and composition upsets) can be identified much more quickly using tray temperatures than using on-line product analyzers. For fast-acting columns (i.e., columns with a relatively low reflux ratio), feedback control using the GC can result in poor control performance because the deadtime-to-time constant ratio of the process is too large. The relatively fast response of temperature sensors gives tray temperature control loops a much smaller deadtime-to-time constant ratio and, therefore, tray temperature control loops exhibit better control performance with shorter closed-loop response times than control directly from a GC.

For example, as a disturbance enters the column, the tray temperature changes long before the online analyzer begins to change; therefore, adjustments to the setpoint for the steam flow controller are made by the tray temperature control loop to absorb the effect of the disturbance. The analyzer controller, which is applied only when a new composition measurement is available, acts slowly and removes any offset from the composition setpoint. This multiple cascade arrangement works effectively because the flow control loop is much faster than the temperature control loop, which is much faster than the composition control loop.

From an analysis of these examples of cascade control, the following conclusions can be drawn:

1. The inner loop (secondary loop) must be considerably faster than the outer loop (primary loop) for effective cascade control.
2. The CV for the secondary loop must have a direct effect on the CV for the primary loop.
3. High-frequency feedback action of the secondary loop many times eliminates specific disturbances before they can significantly affect the primary loop. Because the process responds much faster to the secondary loop than to the primary loop, the secondary loop can absorb these specific disturbances much more quickly.

Theoretical Analysis. Figure 12.2.6 shows a block diagram of a generalized cascade loop. For the flow control cascade shown in Example 12.1, the inner loop is the flow control loop wherein the output of the flow controller goes directly to the control valve. $G_{ps}(s)$ represents the effect of valve stem position on steam flow to the reboiler. $G_{pp}(s)$ represents the effect of steam flow on the column tray temperature. The output of the temperature controller is the setpoint for the flow controller. Note that the disturbance, $D(s)$, enters the secondary loop and, therefore, can be effectively absorbed by the high frequency feedback action of the secondary loop. The transfer function for the effect of the disturbance $D(s)$ on the primary control loop $Y(s)$ can be derived using the properties of a block diagram or by applying Equation 7.2.10 yielding

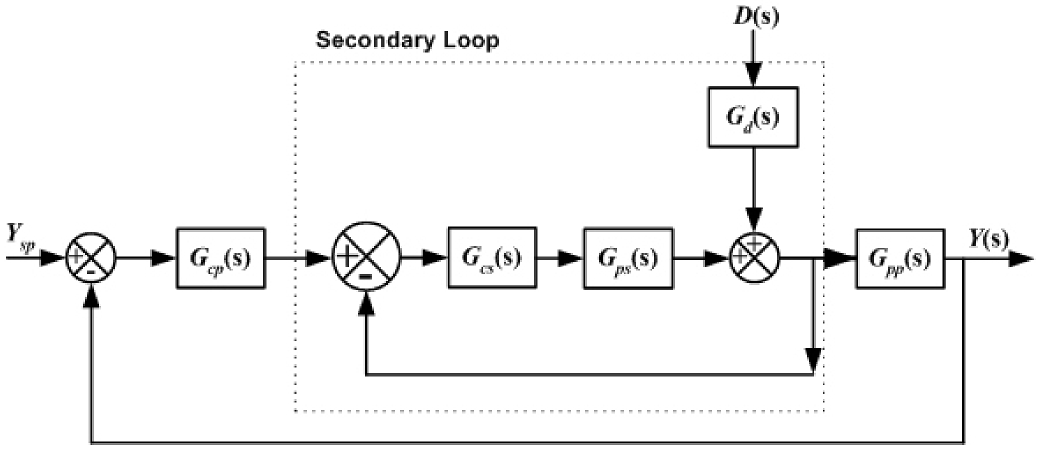


Figure 12.2.6 Block diagram of a generalized cascade control loop. Note that the actuators and sensors are omitted from this diagram.

$$\frac{Y(s)}{D(s)} \qquad \frac{G_d(s)\,G_{pp}(s)}{1 \quad G_{cs}(s)\,G_{ps}(s) \quad G_{cp}(s)\,G_{pp}(s)\,G_{cs}(s)\,G_{ps}(s)} \tag{12.2.1}$$

Assuming first-order processes for $G_{pp}(s)$ and $G_{ps}(s)$ and P-only controllers for the primary and secondary loops, a second-order response results. The closed-loop second-order process natural period, $_n$, is given by

$$_n \qquad \sqrt{\frac{_{pp} \quad _{ps}}{1 \quad K_{cs}K_{ps}(1 \quad K_{cp}K_{pp})}} \tag{12.2.2}$$

where $_{pp}$ is the time constant and K_{pp} is the gain of the primary process [$G_{pp}(s)$], $_{ps}$ is the time constant and K_{ps} is the gain of the secondary process [$G_{ps}(s)$], K_{cp} is the gain of the controller on the primary loop and K_{cs} is the gain for the secondary loop. From Equation 12.2.2, $_n$ can be significantly smaller than $_{pp}$ when $_{ps} \quad _{pp}$.

Example 12.5 The Effect of the Relative Dynamics of the Primary and Secondary Loops

Problem Statement. Evaluate the control performance of a cascade control loop using a FOPDT process model for the primary and secondary loops:

$$G_{pp}(s) \qquad \frac{1.0\,e^{\;0.5\,s}}{s \quad 1}$$

$$G_{ps}(s) \qquad \frac{1.0\,e^{\;(0.5/r)\,s}}{s/r \quad 1}$$

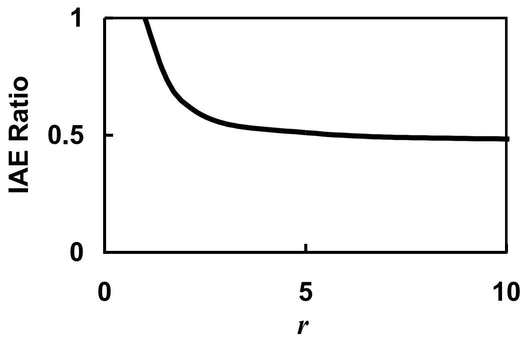


Figure 12.2.7 The effect of the relative speed of the secondary loop compared to the primary loop on the relative control performance of cascade control.

$$r \quad \frac{pp}{ps}$$

where r is the relative speed of the secondary loop compared to the primary loop.

Solution. Figure 12.2.7 shows that the relative control performance (ratio of IAE for the cascade loop divided by IAE for the case without a cascade) improves as r increases. Note that these results are based on the disturbance, $D(s)$, affecting the primary loop only through the secondary loop (see Figure 12.2.6). As a general rule, **the secondary loop should be at least three times as fast as the primary loop** to justify the use of cascade control. This point corresponds to the "knee" of the curve shown in Figure 12.2.7. To obtain this advantage for cascade control, **the secondary loop must be tuned tightly**, i.e., allowing only short term deviations from the setpoint of the secondary loop.

Example 12.6 Schematic Development for a Cascade Control System

Problem Statement. Consider the steam-heated heat exchanger shown in Figure 12.2.8. This process has a temperature measurement for the outlet temperature of the process fluid, a pressure measurement on the steam on the shell side of the exchanger and a control valve on the condensate leaving the exchanger. The valve on the condensate affects the level of condensate in the exchanger which, in turn, affects the area available for condensation of the steam (heat transfer). The heat-transfer rate is also directly related to the steam pressure. The heat-transfer rate determines the outlet temperature of the process stream. Draw the schematic of a control system for which a temperature controller on the outlet of the process stream is cascaded to a pressure controller on the steam, the output of which is sent to the control valve on the condensate.

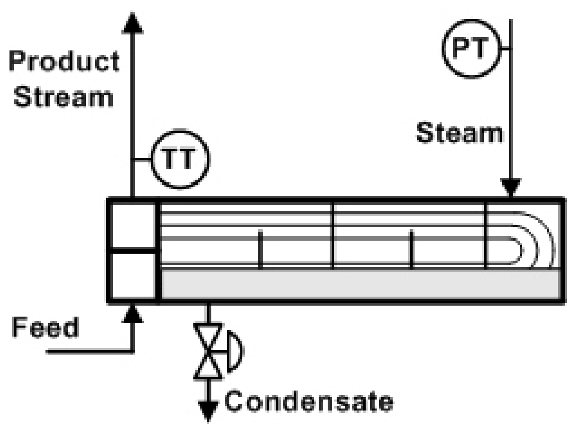


Figure 12.2.8 Schematic of a steam-heated heat exchanger without controls.

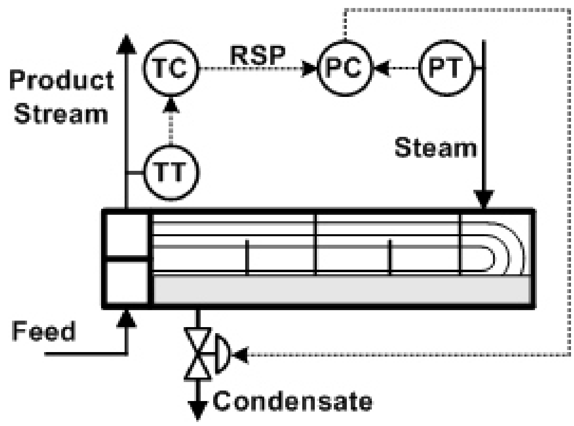


Figure 12.2.9 Schematic of the heat exchanger with multiple cascade controls.

Solution. The cascade system for the steam-heated heat exchanger is shown in Figure 12.2.9. The output of the temperature controller on the outlet of the process fluid becomes the setpoint for the pressure controller on the steam. The output of the pressure controller goes directly to the valve on the condensate. This configuration has the advantage that it can absorb changes in the steam pressure, and it uses a relatively small control valve (i.e., less expensive) because the condensate line is a smaller diameter line than the steam line.

Self-Assessment Questions

Q12.2.1 What relationships between the primary and secondary loops are required for effective cascade control?

Q12.2.2 For Figure 12.2.3b, what does the primary loop adjust to control the temperature of the product stream?

Q12.2.3 What disturbances does a flow control cascade reject?

Q12.2.4 What disturbances does the cascade controller shown in Figure 12.2.3b reject that the controller without cascade control (Figure 12.2.3a) can less effectively reject?

Self-Assessment Answers

Q12.2.1 In order for cascade control to be effective, the secondary loop should be at least three times faster than the primary loop. Once the ratio of time constants of the secondary and primary loops reaches a ratio of 1/10, little incremental benefit is obtained by increasing the relative speed of the secondary loop compared with the primary loop. Note that in order to maximize the speed difference between the primary and secondary loops, the secondary loop should be tightly tuned. In addition, the secondary loop must be able to reject key disturbances that affect the primary loop.

Q12.2.2 The primary loop adjusts the setpoint of secondary loop (water-jacket temperature loop) to control the product temperature.

Q12.2.3 A flow control cascade rejects the effect of changes in upstream and downstream pressure.

Q12.2.4 The cascade controller effectively rejects inlet cooling water temperature changes. The case without cascade control must wait until the reactor temperature changes before taking action.

12.3 Ratio Control

The flow rates associated with many processes scale directly with the feed rate to the process, e.g., distillation, wastewater neutralization and reactors with multiple feed streams. For distillation, all the liquid and vapor flow rates within the column are directly proportional to the feed rate if the product purities are maintained and the tray efficiency is constant. For wastewater neutralization, the amount of reagent necessary to maintain a neutral pH for the effluent varies directly with the flow rate of the wastewater feed, as long as the titration curve of the wastewater remains constant. A reactor may have to be fed at specified stoichiometric ratios to attain a high conversion to the desired product.

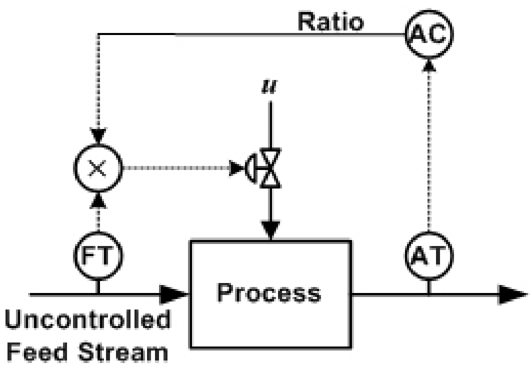


Figure 12.3.1 Schematic of a general ratio controller applied for composition control.

When the MV of a process is directly proportional to the feed rate, ratio control can significantly reduce the effect of feed rate disturbances on the process. Figure 12.3.1 shows how a ratio controller can be applied for a general case. When a feed flow rate

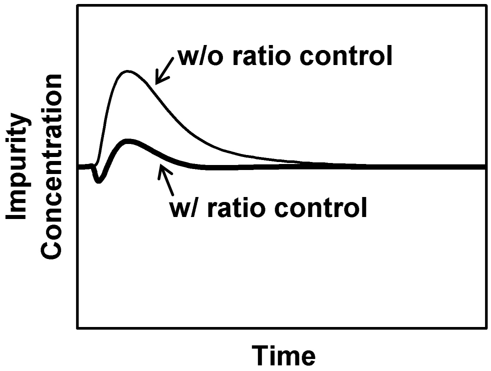


Figure 12.3.2 Distillation overhead composition response with and without ratio control.

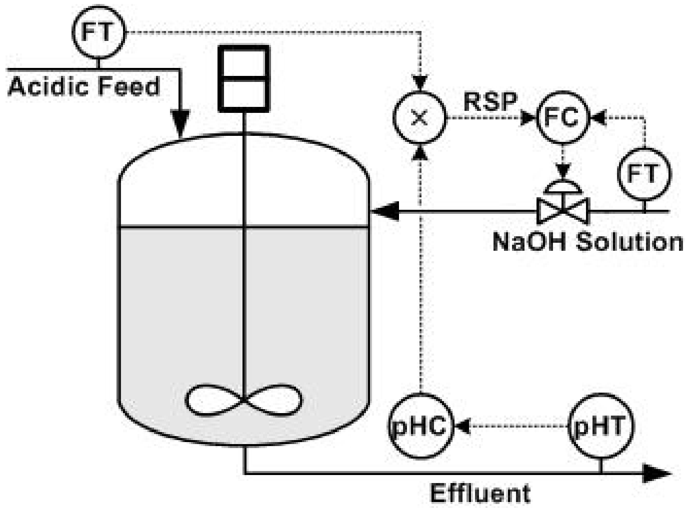


Figure 12.3.3 Schematic of ratio control applied for pH control of an acid wastewater neutralization process.

change, which is referred to as the "wild" stream, is measured, the MV is proportionally adjusted immediately, i.e., the measured flow rate of the feed is multiplied by the MV-to-feed rate ratio (output of the composition controller). Feedback corrections (i.e., changes in the ratio specified by the controller) are made to the ratio based on analyzer readings. That is, **the controller output is a ratio**.

Figure 12.3.2 shows a comparison in control performance between a conventional feedback controller and a ratio controller for a distillation column. The CV is the overhead composition and the disturbance is a change in the feed rate. Note that the maximum deviation from setpoint and the settling time for this disturbance are significantly reduced by ratio control. The MV for a ratio controller is a ratio; therefore, during an ATV test to tune a ratio controller, the ratio is adjusted to create the standing wave in the CV.

Example 12.7 Ratio Control applied to a pH Neutralization Process

Problem Statement. Consider the pH neutralization process based on a mixing tank shown in Figure 12.3.3. Analyze the control performance of this process with and without ratio control.

Solution. Without ratio control, acid wastewater flow rate changes cause significant pH excursions for the effluent. The wastewater is typically discharged to a biological treatment pond. If the pH of the pond is more than 9 or less than 6, the biological agents go into shock and are unable to effectively remove the organic matter from the wastewater. Sharp changes in pH within the 6-9 pH unit range can also reduce the effectiveness of the treatment pond.

When the chemical makeup of the wastewater remains relatively constant, the ratio controller shown in Figure 12.3.3 can effectively handle wastewater feed flow rate changes and prevent feed flow rate changes from upsetting the treatment pond. Small changes in the chemical makeup of the wastewater can usually be handled by the feedback controller, which adjusts the reagent-to-wastewater ratio to maintain the specified effluent pH.

Example 12.8 Ratio Control for a Bio-Reactor

Problem Statement. Mammalian cell cultures (e.g., Chinese Hamster Ovary cells, which are known as CHO cells) require glutamine as a nutrient in proportion to the glucose feed to maintain a healthy cell culture. Glutamine is an amino acid that is a building block for the formation of mammalian cells. CHO cells are used by the pharmaceutical industries to manufacture human therapeutics based on recombinant DNA. In the bio-tech industries, glutamine is typically mixed into the glucose feed tank. Because it is important to maintain a relatively fixed ratio of glutamine to glucose, ratio control may be an effective means to add glutamine to a bio-reactor based on mammalian cells. Develop a control diagram for the addition of glutamine to a bio-reactor based on ratio control.

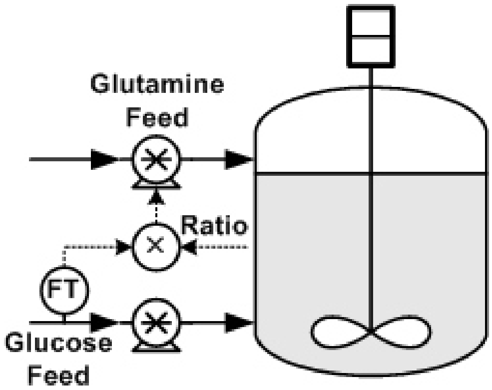


Figure 12.3.4 Ratio control for glutamine addition for a bio-reactor.

Solution. Figure 12.3.4 shows the application of glutamine to glucose ratio control for a bio-reactor. Note that this application is an open-loop version of ratio control because the glutamine to glucose ratio was set at a fixed level. In the previous ratio control cases, the ratio is set by a feedback controller. In this case, the glucose feed rate is the wild stream and the glutamine feed rate is controlled by the ratio controller. Because the ratio of glutamine to glucose is quite small, the flow rate of glutamine is relatively small. The advantage of this arrangement is that it removes a batch element (i.e., manually adding the glutamine to the glucose feed tank) from the operation of this fed-batch process.

Example 12.9 The Application of Ratio Control for a Distillation Column

Problem Statement. Consider the stripping section of a distillation column shown in Figure 12.3.5. Analyze the control performance differences for this process with and without ratio control.

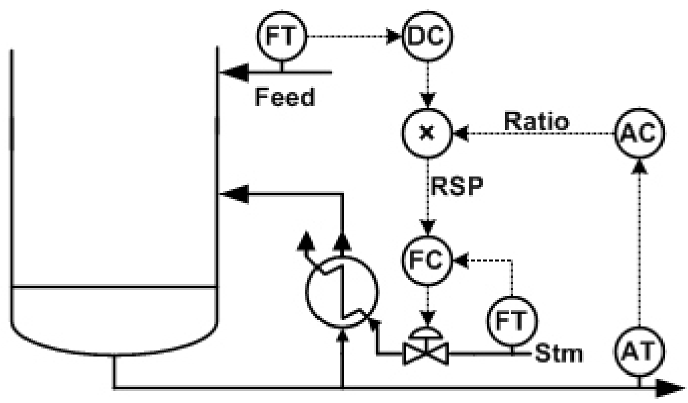


Figure 12.3.5 Control diagram for ratio control for feed rate changes applied to the stripping section of a distillation column.

Solution. Consider the system shown in Figure 12.3.5 without ratio control. In this case, as feed flow rate changes to the column occur, the feedback controller from the bottoms product analyzer to the flow controller on the steam must absorb the entire upset, which can be significant. On the other hand, the application of ratio control can significantly reduce the size of the upset that the feedback controller must handle. This results because, at steady-state conditions, the steam rate necessary to maintain the product purities scales directly with the feed rate to the column.

This application of ratio control is similar to the previous example except that dynamic compensation is added to the measured column feed rate. If the steam flow to the reboiler is increased immediately after an increase in column feed rate, the corrective action is initially an over-correction. This results because, when a feed rate change occurs, it takes some time for the bottom product composition to be affected. The purpose of the dynamic compensation (DC) element is to allow for the correct timing for the compensation for feed rate changes. The dynamic element for this case could be simply a first-order lag element, e.g., a digital filter described by Equation 9.8.1. The wastewater neutralization case (Figure 12.3.3) does not require dynamic compensation because the process pH responses to feed rate and NaOH flow rate changes have similar dynamic behavior. In this case, the bottom composition responds differently to feed rate changes than to reboiler steam changes; therefore, dynamic compensation is required for ratio control in this case.

Self-Assessment Questions

Q12.3.1 What kinds of processes benefit from ratio control?

Q12.3.2 When is dynamic compensation required and when is it not required for ratio control?

Q12.3.3 Explain how the ratio controller shown in Figure 12.3.3 reduces the effect of feed flow rate disturbances for the acidic feed on the effluent pH.

Self-Assessment Answers

Q12.3.1 Processes whose flow rates scale directly or fairly directly with feed rate to the process are likely to benefit from ratio control.

Q12.3.2 Dynamic compensation is not required when the dynamic effect of the feed rate on the CV is similar to the dynamic effect of the MV on the CV. When these two dynamic responses are significantly different, dynamic compensation is required.

Q12.3.3 The output of pH controller is the ratio of flow of NaOH-to-flow of acid wastewater. The measured flow rate of acid wastewater is multiplied by the ratio to give the setpoint to NaOH flow controller. Therefore, as the feed rate changes, the flow rate of NaOH automatically adjusts.

12.4 Feedforward Control

Feedforward control can be applied to process control loops that are significantly affected by disturbances that are measurable (or can be estimated) on-line. A feedback controller (Figure 12.4.1a) reacts to deviations from setpoint caused by a disturbance until the process is returned to setpoint. As was pointed out in Chapter 7, because the proportional and derivative terms are zero during steady-state operation at the setpoint, the integral term in the PID controller is responsible for the long term compensation for disturbances. A feedforward controller (Figure 12.4.1b) anticipates the effects of a measured change in a disturbance (i.e., a **load change**) and takes corrective action before the disturbance affects the process. In effect, the feedforward controller applies corrective MV changes corresponding to the integral action that a feedback controller generates; therefore, when a feedback controller and feedforward controller are used together, the feedback controller has much less integral "work" to do to compensate for a measured disturbance.

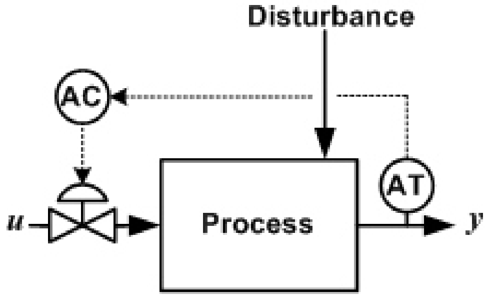


Figure 12.4.1a Schematic of a general feedback controller.

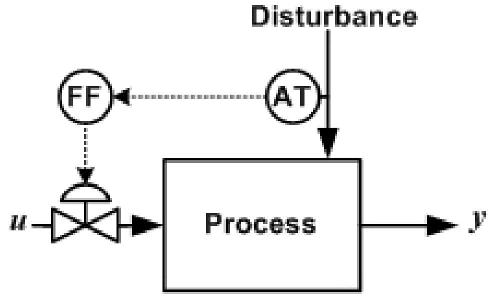


Figure 12.4.1b Schematic of a general feedforward controller.

Example 12.10 Feedforward and Feedback Control of a Boiler Drum Level

Problem Statement. Evaluate feedforward, feedback and combined feedforward and feedback control applied for level control of a boiler drum.

Solution. Figure 12.4.2a shows a feedback controller applied for the level control of a boiler drum. The feedback controller compares the measured value of the level with the setpoint and adjusts the flow rate of the feedwater to the drum. Therefore, when changes in the demand for steam occur, changes in the drum level result. If large changes in steam demand occur, a large gain is required for the feedback controller to maintain the level near its setpoint. But, for large controller gains, the process is more susceptible to oscillatory or unstable behavior. Also, high-gain controllers are sensitive to noisy measurements of the CV and, in this case, level indicators can have significant noise levels.

Figure 12.4.2b is a schematic of a feedforward controller applied for steam drum level control. The idea is quite simple: if the flow rate of the makeup feedwater is equal to the steam usage, the drum level remains constant. You are tempted to conclude that the feedforward controller is all that is needed for this application. Unfortunately,

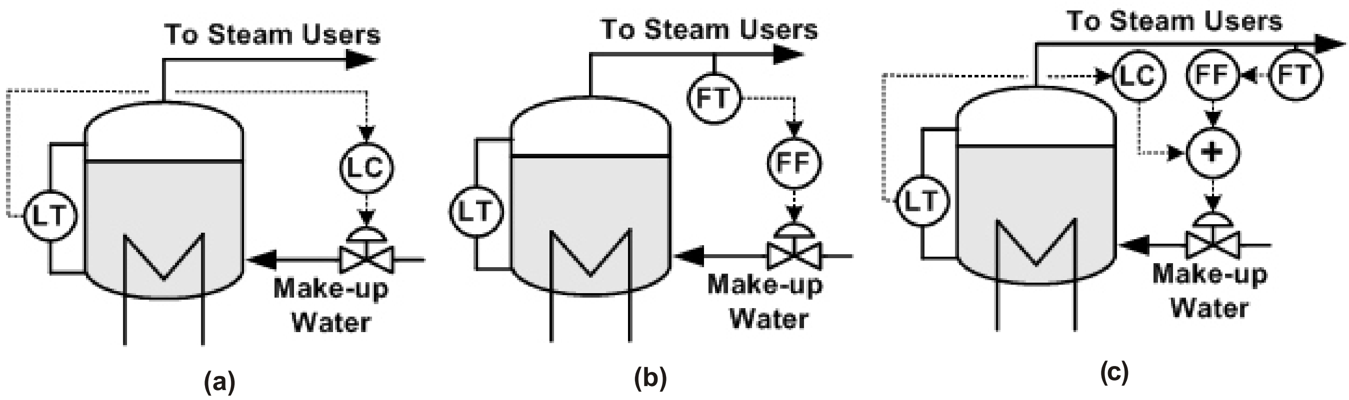


Figure 12.4.2 Boiler drum level control (a) feedback (b) feedforward (c) feedback and feedforward combined.

the measurements of the steam usage and the feedwater flow rate are not perfectly accurate. Even small errors in measured flow rates add up over time, leading to one of two undesirable extremes. The drum can fill with water and put water into the steam system or the liquid level can drop, exposing the boiler tubes, which can damage the boiler tubes. As a result, neither feedback nor feedforward are effective by themselves for this case. In general, feedforward-only controllers are susceptible to measurement errors and unmeasured disturbances and, as a result, some type of feedback correction is typically required.

Figure 12.4.2c shows a combined feedforward and feedback controller for the control of the level in the steam drum. The feedforward controller provides most of the control action required by responding to the measured steam usage. The feedback controller can be a relatively low-gain controller because it is required to compensate only for measurement errors and unmeasured disturbances.

General feedforward controller. The previous example demonstrates how a feedforward controller can be developed from an analysis of the process. A more generalized feedforward controller design procedure can also be used. Consider a block diagram for a generalized feedforward controller shown in Figure 12.4.3. The disturbance is measured by a sensor with a dynamic response given by $G_{ds}(s)$. The feedforward controller [$G_{ff}(s)$] uses the measured value of the disturbance to calculate its feedforward correction [$C_{ff}(s)$]. The feedforward correction affects the actuator, which, in turn, changes the MV level, which affects the CV. Notice that the CV [$Y(s)$] changes are due to changes in the feedforward controller output, $C_{ff}(s)$, and in the disturbance, $D(s)$, i.e.,

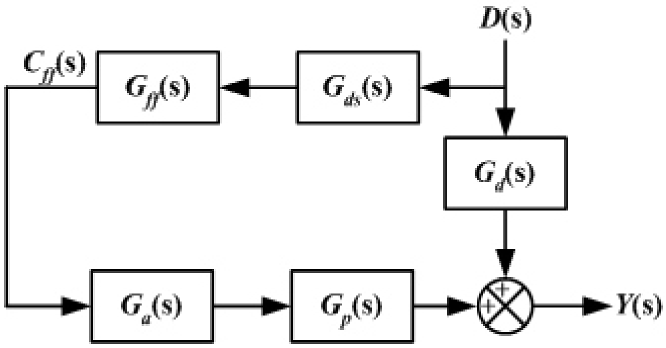


Figure 12.4.3 A block diagram of a general feedforward controller.

$$Y(s) \quad C_{ff}(s)G_a(s)G_p(s) \quad D(s)G_d(s)$$

The feedforward controller determines the control action $C_{ff}(s)$, i.e.,

$$C_{ff}(s) \quad G_{ff}(s)G_{ds}(s)D(s)$$

where $G_{ds}(s)$ represents the sensor response to the measured disturbance, then

$$Y(s) \quad D(s)G_{ds}(s)G_{ff}(s)G_a(s)G_p(s) \quad D(s)G_d(s)$$

Because we want to design a feedforward controller that keeps the process at setpoint in spite of disturbances, we set $Y(s)$ equal to zero and solve for $G_{ff}(s)$ yielding

$$G_{ff}(s) \quad \frac{G_d(s)}{G_{ds}(s)G_a(s)G_p(s)} \qquad \textbf{12.4.1}$$

This equation gives us a means of directly determining the feedforward controller using a model of the effect of the disturbance on the process and a model of the effect of the MV on the process. Let's assume that we have FOPDT models for $G_d(s)$ and the product $G_{ds}(s)\ G_a(s)\ G_p(s)$, i.e.,

$$G_{ds}(s)G_a(s)G_p(s) = \frac{K_p\, e^{-\theta_p s}}{\tau_p s + 1}$$

$$G_d(s) = \frac{K_d\, e^{-\theta_d s}}{\tau_d s + 1}$$

Then the application of Equation 12.4.1 using the FOPDT models yields

$$G_{ff}(s) = -\frac{K_d(\tau_p s + 1)\, e^{-\theta_d s}}{K_p(\tau_d s + 1)\, e^{-\theta_p s}} = \frac{K_{ff}(\tau_{ld}\, s + 1)\, e^{-\theta_{ff} s}}{(\tau_{lg}\, s + 1)} \qquad \textbf{12.4.2}$$

The feedforward controller gain is given by

$$K_{ff} = -\frac{K_d}{K_p} \qquad \textbf{12.4.3}$$

The lead of the feedforward controller, τ_{ld}, is

$$\tau_{ld} = \tau_p \qquad \textbf{12.4.4}$$

The lag of the feedforward controller, τ_{lg}, is

$$\tau_{lg} = \tau_d \qquad \textbf{12.4.5}$$

The deadtime of the feedforward controller is

$$\theta_{ff} = \theta_d - \theta_p \qquad \textbf{12.4.6}$$

Equation 12.4.2 represents a lead-lag element (Section 6.10) with deadtime, which is a standard feature on a DCS or control computer. A feedforward controller can be implemented on a DCS or control computer by applying a lead-lag element with deadtime with the proper values of K_{ff}, τ_{ld}, τ_{lg} and θ_{ff} to measured changes in the disturbance. Figure 12.4.4 shows the effect of the ratio of τ_{ld}/τ_{lg} on the dynamic response of a lead-lag element. When τ_{ld}/τ_{lg} is greater than one, an initial overcompensation is used. That is, when the process responds faster to the disturbance than to the MV (i.e., controller output), larger than steady-state changes in the controller output are required to initially compensate for dynamic mismatch. On the other hand, when τ_{ld}/τ_{lg} is less than one, the controller output results in a step change in c_{ff} , which is less than the steady-state change, followed by a monotonic approach to its steady-state level. That is, when the process responds faster to the controller output than to the disturbance, a more gradual increase in the controller output level is used to compensate for a

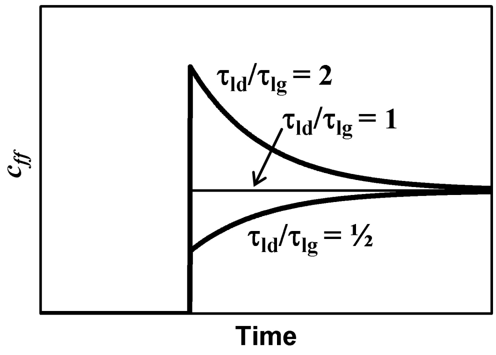


Figure 12.4.4 The effect of the ratio of $_{ld}$ to $_{lg}$ on the dynamic response of a lead-lag element. c_{ff} is the output from the feedforward controller for a step input.

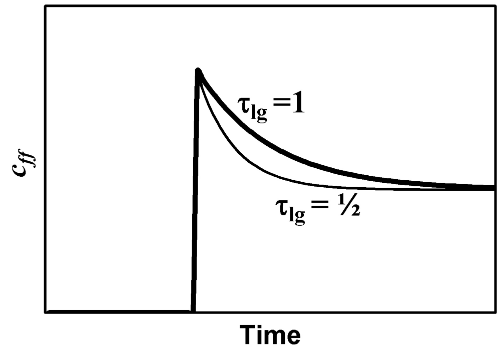


Figure 12.4.5 The effect of $_{lg}$ on the dynamic response of a lead-lag element. c_{ff} is the output of the feedforward controller for a step input.

disturbance. Figure 12.4.5 shows the effect of $_{lg}$ on the dynamic response of a lead-lag element for which the ratio of lead to lag is maintained constant at a value of 2. As you would expect, the case with the larger value of the lag requires a longer time to settle to the steady-state value. It should be pointed out that you cannot use a negative value for $_{ff}$ (i.e., $_{d}$ $_{p}$), which is called an unrealizable controller. In such cases, set $_{ff}$ equal to zero.

Example 12.11 Static Feedforward Controller

Problem Statement. Develop a feedforward controller for inlet feed temperature changes for a stirred-tank heater (Figure 12.4.6). The stirred-tank heater is equivalent to an endothermic CSTR (Example 3.5) without reactions occurring.

Solution. The dynamic energy balance for a stirred-tank heater is given by

$$V \quad C_v \frac{dT_{out}}{dt} \quad FC_p(T_{in} \quad T_{out}) \quad Q$$

where V is the volume of fluid in the stirred-tank heater, is the density of the fluid, C_v and C_p are heat capacities of the fluid, T_{out} is the temperature of the product leaving the stirred-tank heater, T_{in} is the temperature of the feed entering the stirred-tank heater and Q is the rate of heat addition. Applying Equation 5.7.3 to this equation and rearranging results in the process transfer function,

$$G_p(s) \quad \frac{T_{out}(s)}{Q(s)} \quad \frac{\frac{1}{VC_p}}{s \quad \frac{F}{V}} \quad \frac{1}{FC_p} \quad \frac{1}{\frac{V}{F}s \quad 1}$$

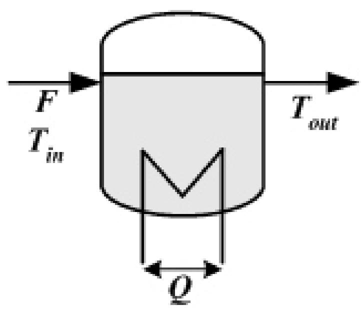


Figure 12.4.6 PFD for a stirred-tank heater.

and the transfer function for the effect of changes in the feed temperature, i.e.,

$$G_d(s) = \frac{T_{out}(s)}{T_{in}(s)} = \frac{\frac{F}{V}}{s + \frac{F}{V}} = \frac{1}{\frac{V}{F}s + 1}$$

assuming that the heat capacity at constant volume (C_v) is equal to the heat capacity at constant pressure (C_p). Applying Equation 12.4.1, assuming that $G_{ds}(s)$ and $G_a(s)$ are unity (i.e., the dynamics of $G_{ds}(s)$ and $G_a(s)$ are relatively fast), the feedforward controller is given as

$$G_{ff}(s) = -\frac{G_d(s)}{G_p(s)} = -F\,C_p$$

Because the feedforward controller does not provide dynamic compensation, it is called a **static feedforward controller**. The effect of the MV and the effect of the disturbance on the process have the same dynamic behavior (i.e., first-order dynamics); therefore, the dynamic terms for the application of Equation 12.4.2 cancel, and a static feedforward controller results. The response of a static feedforward controller to a unit step input is shown in Figure 12.4.4 for τ_{ld}/τ_{lg} equal to one. In general, **a static feedforward controller can be used when the process has the same dynamic response to changes in the MV and the disturbance.** The feedforward controller for the boiler level in Example 12.11 is another example of a static feedforward controller.

Example 12.12 Feedforward Controller When τ_p is Small

Problem Statement. Evaluate feedforward control for changes in the feed flow rate for the liquid/liquid heat exchanger shown in Figure 12.4.7.

Solution. A qualitative examination of the liquid/liquid heat exchanger shown in Figure 12.4.7 yields the following conclusions: (1) the product temperature is expected to respond relatively quickly to changes in the MV (i.e., the bypass flow rate) and (2) the product temperature should respond much more slowly to changes in the feed rate of the process stream because its response depends on the dynamic response of the heat exchanger to a feed rate change. Therefore, application of Equation 12.4.2 for this problem results in a feedforward controller with the following form:

$$G_{ff}(s) = \frac{K_{ff}\,e^{-\theta_{ff} s}}{(\tau_{lg} s + 1)}$$

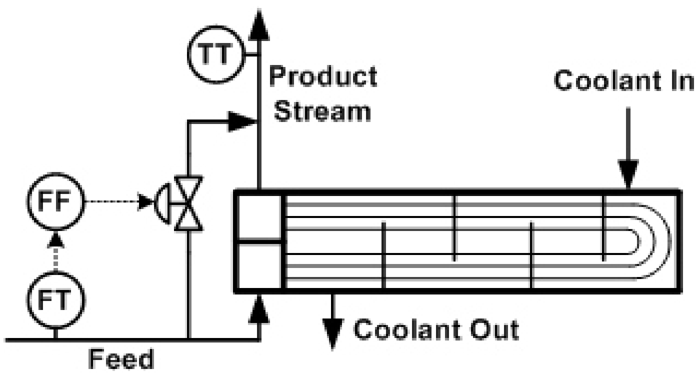


Figure 12.4.7 Schematic of a feedforward-only controller for a liquid/liquid heat exchanger.

because τ_{ld} can be assumed to equal to zero in this case (i.e., τ_p is small). As a result, the feedforward controller provides a first-order plus deadtime compensation for changes in the feed rate to the system. From Equation 6.10.1, you can observe that in the limit of τ_{ld} approaching zero, the feedforward response becomes FOPDT model.

Example 12.13 Feedforward Controller When the Process Responds Faster to the MV

Problem Statement. Evaluate feedforward control for changes in the feed composition for the stripping section of a distillation column (Figure 12.4.8).

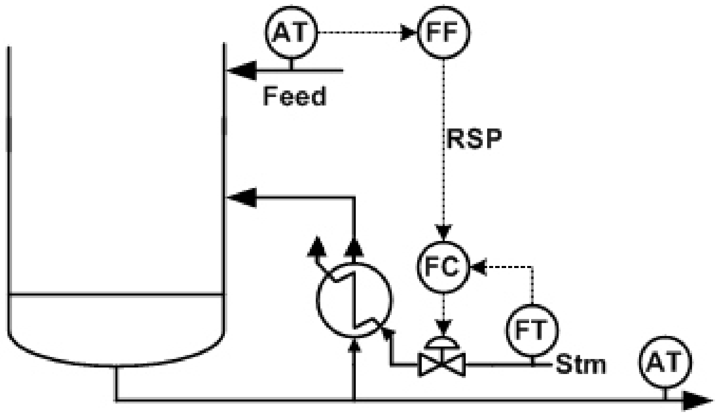


Figure 12.4.8 Control diagram for feedforward control applied to the stripping section of a column.

Solution. For this case, the process should respond more quickly to changes in the MV than to changes in the disturbance. That is, once a change in the setpoint to the flow controller on the steam is implemented, the bottoms product composition changes more quickly than for a change in the feed composition. This results because the feed composition change must work its way down the column tray-by-tray before it affects the bottom product composition. As a result, if compensation for feed composition changes is implemented immediately using a static feedforward controller, overcompensation results. Unlike Example 12.12, the dynamics of $G_p(s)$ are not fast enough to be neglected. As a result, the complete version of Equation 12.4.2 is required in this case, i.e.,

$$G_{ff}(s) = \frac{K_{ff}(\tau_{ld}\, s + 1)\, e^{-\theta_{ff} s}}{(\tau_{lg}\, s + 1)}$$

and the response of the feedforward controller for this case should be similar to the response shown in Figure 12.4.4 in which τ_{ld} is less than τ_{lg}. Because MVs that have the most immediate effect on the process are usually selected, this case (i.e., the process responds more quickly to the MV than to the disturbance) represents the most common form for feedforward controllers.

Tuning. Tuning a feedforward controller involves selecting the values of K_{ff}, τ_{ld}, τ_{lg} and θ_{ff}. Equations 12.4.3 to 12.4.6 can be used to estimate these tuning parameters if FOPDT models are available. Because identifying $G_d(s)$ can be difficult, it is usually advisable to field tune feedforward controllers. The following field tuning procedure is recommended.

1. Make initial estimates of K_{ff}, τ_{ld}, τ_{lg}, and θ_{ff} based on process knowledge.

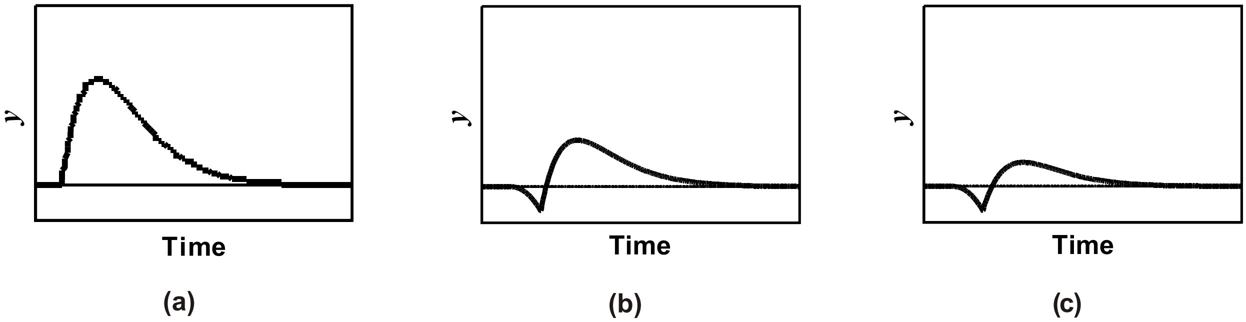


Figure 12.4.9 Tuning results for a feedforward controller for a step change in the disturbance. (a) Results for initial settings with correct feedforward gain. (b) Results after deadtime tuned. (c) Final tuning results.

2. Under open-loop conditions, adjust K_{ff} while maintaining the rest of the tuning parameters at their initial levels to minimize the steady-state deviation from setpoint. Figure 12.4.9a shows the dynamic response of a feedforward controller for a step change in the disturbance after K_{ff} has been adjusted to eliminate offset.
3. By analyzing the dynamic mismatch, adjust $_{ff}$. The direction of the deviation should indicate whether the feedforward correction is applied too soon or too late, causing dynamic mismatch. Figure 12.4.9b shows the feedforward control performance after $_{ff}$ is tuned.
4. Finally, adjust either $_{ld}$ or $_{lg}$ until approximately equal areas above and below the setpoint result. Figure 12.4.9c shows the results after $_{ld}$ and $_{lg}$ are adjusted.

Combined Feedforward and Feedback. Figure 12.4.10 is a general schematic of a combined feedforward controller and a feedback controller. The input to the feedforward controller is the measure value of the disturbance and the output is the feedforward correction for the disturbance. Likewise, the

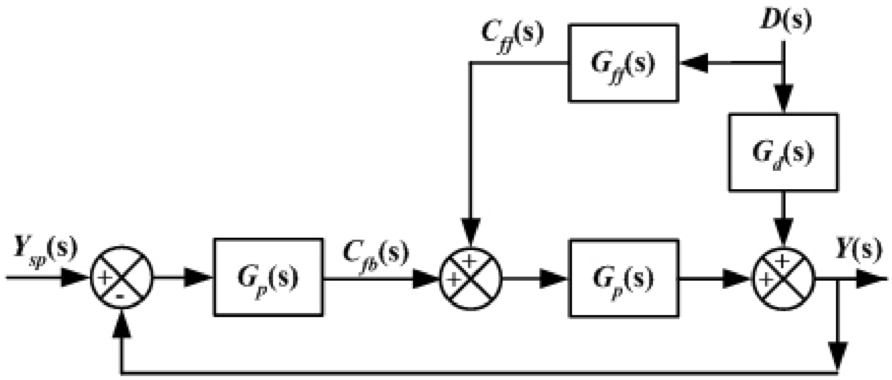


Figure 12.4.10 Block diagram of a combined feedforward and feedback controller. Note that the actuator and sensors are omitted.

input to the feedback controller is the error from setpoint and the output is the feedback controller correction. Note that the feedforward and the feedback correction are simply added and applied to the process.

Example 12.14 Feedforward Control Applied to the Endothermic CSTR

Problem Statement. Field tune a feedforward controller for feed temperature changes to the endothermic CSTR (Example 3.6) using the visual basic simulator that accompanies this text (Section 3.9). In addition, test the feedforward-only controller, the feedback-only controller and the combined feedforward and feedback controller for the same disburbance upset.

Solution. Using open-loop step tests, the following FOPDT models were developed for $G_d(s)$ and $G_p(s)$

$$G_d(s) \quad \frac{T(s)}{T_0(s)} \quad \frac{0.381\ e^{\ 5.62s}}{6.57\,s \quad 1} \ (^{o}\mathrm{C}/^{o}\mathrm{C})$$

$$G_p(s) \quad \frac{T(s)}{Q(s)} \quad \frac{1.50 \quad 10^{\ 4}\ e^{\ 4.31s}}{10.7\,s \quad 1} \ \frac{^{o}\mathrm{C}}{\mathrm{cal/s}}$$

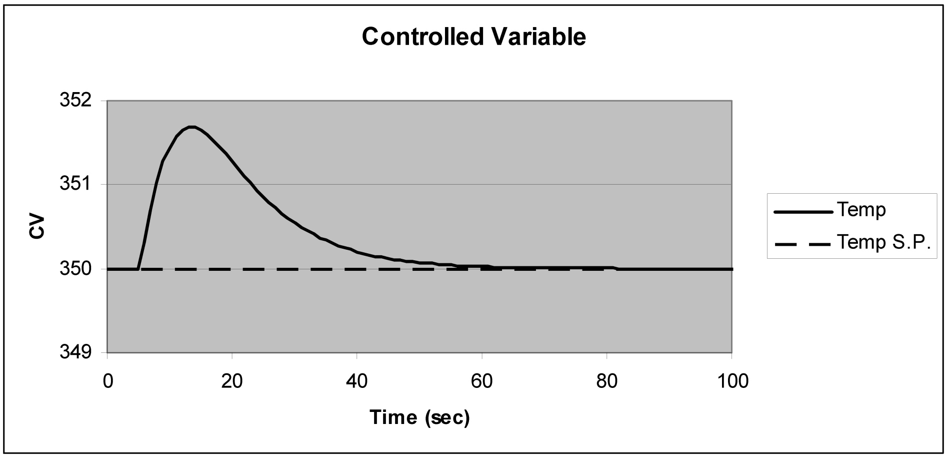


Figure 12.4.11 Results for the static feedforward controller.

where time constants and deadtimes are reported in seconds. The process responds more quickly to the MV than to the disturbance, but not so fast that the dynamics of $G_p(s)$ can be neglected. The feedforward controller was tuned for a 10% increase in feed temperature (T_0). Using a static feedforward controller (i.e., no dynamic compensation), K_{ff} was set at -1.0 10^4 cal/sec/ C and the results for the CV are shown in Figure 12.4.11. Next, $_{ff}$ was evaluated and it was found that $_{ff}$ should be set to zero. Due to positive deviations of the reactor temperature from setpoint, $_{ld}$ was increased from 10.7 s (i.e., $_p$) to 12 s, which reduced the maximum deviation from setpoint without resulting in excessive negative deviations. $_{lg}$ was maintained equal to $_d$ or 6 seconds and the results for this feedforward controller are shown in Figure 12.4.12. Figure 12.4.13 shows the results for a step increase in T_0 from 400 K to 500K for the feedforward controller, a feedback controller and the combined feedforward and feedback controller. The open-loop effect of this disturbance resulted in a 38 K increase in reactor temperature while the combined feedforward and feedback controller had a maximum deviation of only 1.7 K. The results in Figure 12.4.13 show that combined feedforward and feedback out perform either

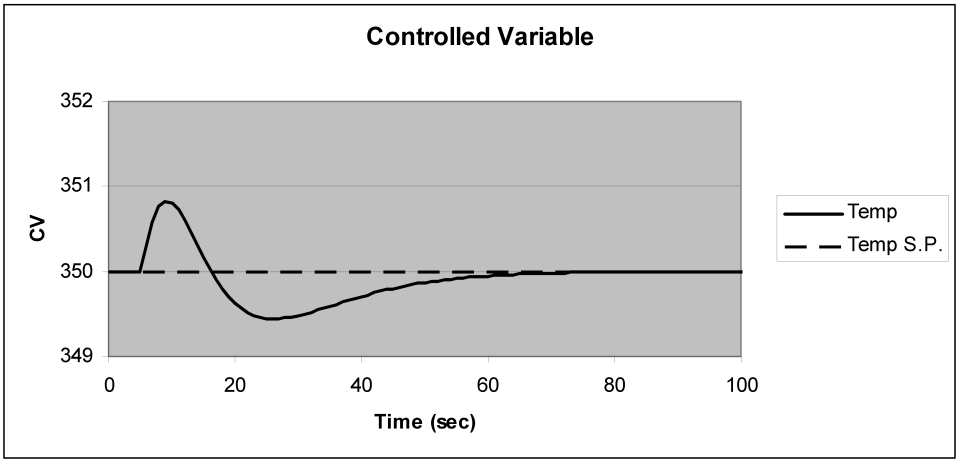


Figure 12.4.12 Final results for tuned feedforward controller.

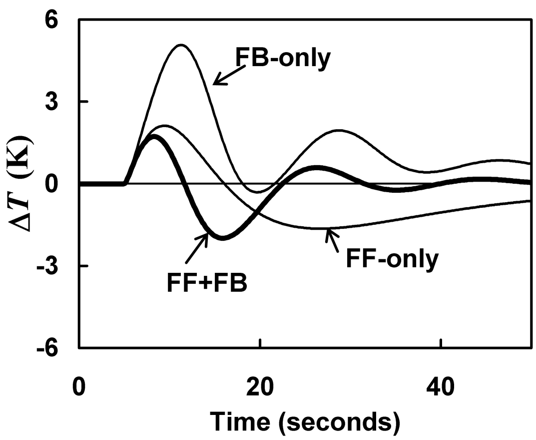


Figure 12.4.13 Comparison among feedforward (FF-only), feedback (FB-only) and combined feedforward and feedback (FF+FB) for a disturbance upset (Example 12.15).

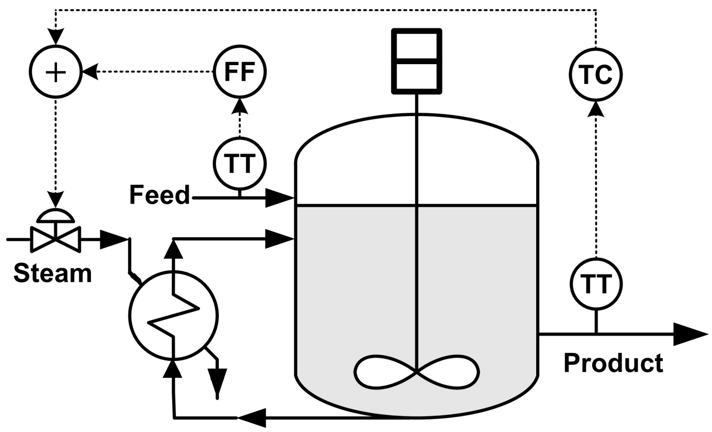


Figure 12.4.14 Schematic of an endothermic CSTR with combined feedforward and feedback control.

feedforward or feedback alone. The feedforward-only controller significantly reduces the initial deviation from setpoint compared to the feedback-only controller, but is sluggish in returning to the setpoint. The combined feedforward and feedback controller uses the feedforward action to reduce the initial deviation from setpoint and uses the feedback action to quickly settle at the setpoint. Figure 12.4.14 shows a schematic representing a combined feedforward and feedback controller applied to the endothermic CSTR.

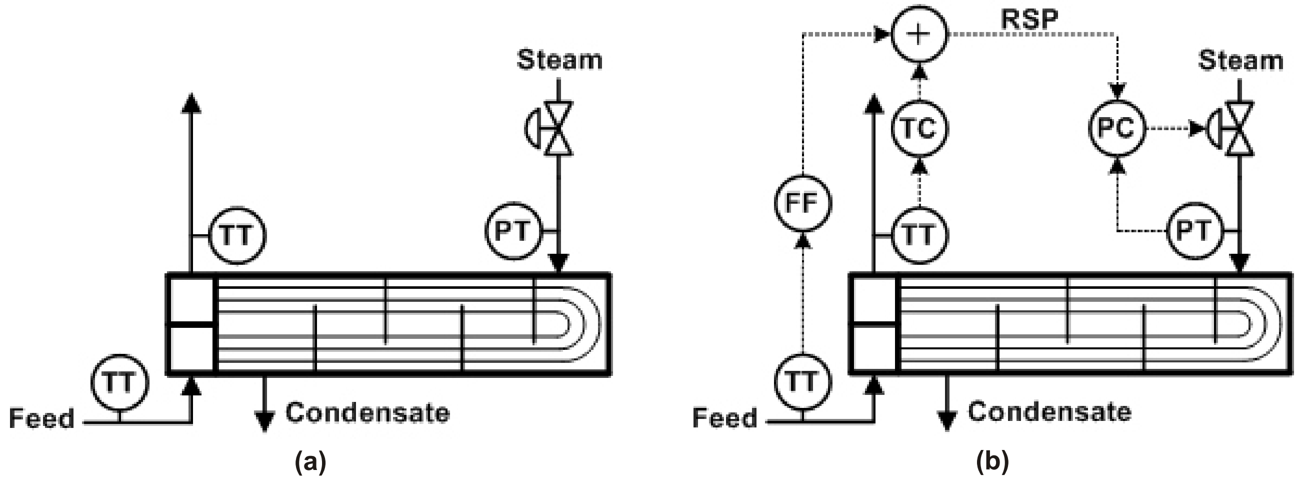


Figure 12.4.15 Schematic of a steam-heated heat exchanger. (a) without controls (b) with feedforward and feedback controllers.

Example 12.15 Schematic of a Combined Feedforward and Feedback Control System for a Heat Exchanger

Problem Statement. Consider the steam-heated heat exchanger shown in Figure 12.4.15a for which there is a temperature measurement on the inlet and outlet of the feed to the exchanger, a pressure measurement on the steam to the exchanger and a control valve on the steam to the exchanger. Draw a schematic of a combined feedforward and feedback control system for which the inlet temperature is the feedforward variable and the outlet temperature is the CV for the feedback loop. The combined feedforward and feedback control action is the setpoint for a pressure controller on the steam.

Solution. The combined feedforward and feedback controller is shown in Figure 12.4.15b. The inlet temperature of the feed is the input to the feedforward controller while the outlet temperature is the CV for the feedback controller. The outputs from the feedback and feedforward controllers are added together, and the resultant is the setpoint for the pressure controller on the steam.

Overview. Table 12.1 summarizes the advantages and disadvantages of feedforward and feedback control. Feedforward and feedback control are complementary, i.e., they each can overcome the disadvantages of the other so that together they are superior to either method alone. Feedforward control does not offer a significant advantage for fast-responding processes because a feedback-only controller can usually absorb disturbances efficiently for those cases. But for slow-responding processes or processes with significant deadtime, by the time a feedback-only controller starts to respond to the effects of a disturbance, the process can already be severely upset. For these cases, the effect of the disturbance can cause the CV to change significantly from its setpoint, resulting in relatively large process parameter changes (e.g., changes in K_p, τ_p and θ_p). In some cases this can lead to closed-loop instability. When feedforward control is added to a slow process or a process with significant deadtime, the deviation of the CV from setpoint can be significantly reduced, resulting in smaller process parameter changes. Therefore, feedforward control can provide significantly more reliable feedback control performance when the feedforward control performance compensates for a major disturbance to the process. In general, feedforward control is useful when (1) feedback control by itself is not satisfactory, i.e., for slow-responding processes or processes with significant deadtime and (2) the major disturbance to a process is measured on-line.

Table 12.1 Comparison of Feedback and Feedforward Control

Feedback	
Advantages	**Disadvantages**
1. Does not require a measurement of the disturbance.	1. Waits until the disturbance has affected the process before taking action.
2. Can effectively reject disturbances for fast-responding process.	2. Susceptible to disturbances when the process is slow or when significant deadtime is present.
3. Simple to implement.	3. Can lead to instability of the closed-loop system due to nonlinearity.

Feedforward	
1. Compensates for disturbances before they affect the process.	1. Requires measurement or estimation of the disturbance and does not compensate for unmeasured disturbances.
2. Can improve the reliability of the feedback controller by reducing the deviation from setpoint.	2. Requires additional tuning parameters for implementation.
3. Offers noticeable advantages for slow processes or processes with significant deadtime.	3. Because it is a linear based correction, its performance deteriorates with nonlinearity.

Feedforward control provides a linear correction and, therefore, can provide only partial compensation for a nonlinear process. Nevertheless, feedforward control can be effective when properly implemented because it can reduce the amount of feedback correction required. When tuning a feedforward controller for a nonlinear process, care should be taken to ensure that the feedforward controller is tuned considering both increases and decreases in the disturbance level. For example, if the feedforward controller is tuned for a certain size increase in the disturbance, it may work quite well for that case but actually contribute to poorer performance when a different size disturbance decrease is encountered. The changes in the MV calculated by the feedforward and feedback controllers are simply added.

Self-Assessment Questions

Q12.4.1 How are ratio control and feedforward control alike and how are they different?

Q12.4.2 When is it advisable to use a static feedforward controller?

Q12.4.3 How are feedback and feedforward controllers combined for the same control loop?

Q12.4.4 How can a feedforward controller improve the reliability of a feedback controller?

Self-Assessment Answers

Q12.4.1 Ratio and feedforward control are alike because they both are used to reduce the effect of measured disturbances on a control loop. They are different in that ratio control handles only feed flow rate disturbances while feedforward can be applied to any measured disturbance. In addition, for ratio control the output of the controller is a ratio that is multiplied by the measured feed rate while for feedforward the measured disturbance is input to a feedforward control element that calculates an incremental change in u that is added to the output of the feedback controller.

Q12.4.2 A static feedforward controller can be used when the process has the same dynamic response to changes in the MV and the disturbance.

Q12.4.3 The changes in the MV calculated by the feedforward and feedback controllers are simply added.

Q12.4.4 Feedforward controller can improve the reliability of the feedback controller by reducing the deviation from setpoint, i.e., reducing the amount of "work" that the feedback controller has to do. In addition, because the feedforward controller reduces the deviations from setpoint due to disturbances, there are less nonlinear changes in the process behavior, resulting in more consistent process parameters (e.g., FOPDT parameters) and thus more reliable feedback control performance.

12.5 Summary

- Cascade, ratio and feedforward control are each designed to reduce the effect of disturbances on feedback control performance. Because disturbances tend to undermine controller reliability, cascade, ratio and feedforward control can contribute to controller performance and reliability when they are properly implemented.
- Cascade control involves applying two controllers in tandem instead of using a single control loop. The control loops are arranged such that the primary loop provides the setpoint for the secondary loop. In this manner, the secondary loop acts as the MV for the primary loop. Because the secondary loop responds more quickly than the primary loop, the secondary loop is able to more effectively reject certain disturbances than if the primary loop were applied by itself.
- Ratio controllers can effectively handle process feed rate changes for processes for which the process flow rates generally scale with feed rate. For ratio controllers, the controller selects the manipulated stream flow rate-to-feed flow rate ratio. Then, this ratio is multiplied by the measured feed rate to give the setpoint for the flow controller on the manipulated stream. In certain cases, dynamic compensation is required for ratio control applications.
- Feedforward control offers significant benefits to a feedback controller when (1) the process is slow responding or when significant deadtime is present and (2) a major disturbance to a process is measurable. A feedforward controller is typically implemented using a lead-lag element with deadtime.
- The equation for a general feedforward controller is given by

$$G_{ff}(s) \qquad \frac{G_d(s)}{G_{ds}(s)G_a(s)G_p(s)}$$

where $G_{ff}(s)$, $G_d(s)$, $G_{ds}(s)$, $G_a(s)$ and $G_p(s)$ are the transfer function for the feedforward controller, the effect of the disturbance on the CV, the measurement of the DV, the actuator and the effect of the MV on the CV, respectively. Note that $G_{ff}(s)$ is generally well represented using a lead-lag element.

12.6 Additional Terminology

Lead-lag element - the ratio of two FOPDT transfer function models; defined by Equation 12.6.
Load change - a change in the disturbance level to a process.
Primary loop - the outer loop for a cascade controller. Also, the slowest loop in a cascade arrangement.
Secondary loop - the inner loop for a cascade controller. Also, the fastest loop in a cascade arrangement.
Static feedforward controller - a feedforward controller which contains only a gain, i.e., it contains no dynamic compensation.

12.7 Preliminary Questions

12.1 Introduction

Q12.1.1 What do cascade, ratio and feedforward control have in common?

12.2 Cascade Control

Q12.2.1 What factors are necessary for an effective cascade arrangement?

Q12.2.2 Is a typical flow control loop on a DCS an example of cascade control? Explain your reasoning.

Q12.2.3 What kinds of disturbances does the cascade arrangement shown in Figure 12.2.3b reject effectively?

Q12.2.4 What advantages does the cascade arrangement shown in Figure 12.2.4 offer?

Q12.2.5 For Figure 12.2.2b, what does the primary loop adjust to control the tray temperature in the stripping section of the column?

Q12.2.6 For Figure 12.2.4, what does the primary loop adjust to control the glucose composition?

12.3 Ratio Control

Q12.3.1 Explain how the ratio controller shown in Figure 12.3.5 reduces the effect of feed flow rate disturbances on the bottoms product composition.

Q12.3.2 How do you determine when the application of ratio control should be beneficial?

12.4 Feedforward Control

Q12.4.1 For the combined feedforward-feedback controller shown in Figure 12.4.2c, indicate how the control structure is able to overcome the limitations of the feedback-only and feedforward-only controller while retaining their advantages.

Q12.4.2 For a feedforward controller using a lead-lag element, what kinds of systems result in a lead that is larger than its lag?

Q12.4.3 Why is a lead-lag element useful for applying feedforward control?

Q12.4.4 When a process responds more quickly to the MV than to the disturbance, what form of the dynamic response does a feedforward controller provide for a step change in the measured disturbance?

Q12.4.5 When a process responds more slowly to the MV than to the disturbance, what form of the dynamic response does a feedforward controller provide for a step change in the measured disturbance?

Q12.4.6 When tuning a lead-lag element with deadtime for a feedforward controller, in what order are the parameters tuned?

Q12.4.7 How does process nonlinearity affect feedforward controller tuning?

Q12.4.8 To what types of processes does feedforward control provide the most significant benefits?

Q12.4.9 How are a feedforward and a feedback controller combined?

12.8 Analytical Question and Exercises

12.2 Cascade Control

P12.2.1** Derive Equation 12.2.1 using Figure 12.2.6 and the properties of block diagrams.

P12.2.2* Consider the CST thermal mixer presented in Figure P12.2.2 (Example 3.3). Draw a schematic for this process demonstrating the application of cascade control.

P12.2.3*** Consider the fixed-bed reactor in which the feed is preheated by a gas-fired heater shown in Figure P12.2.3. The current control configuration has a temperature controller on the outlet from the fixed-bed reactor setting the setpoint for the flow controller on the gas to the furnace. Draw a schematic showing a further application of cascade control (i.e., beyond flow control on the gas flow rate) that provides improved control for this process for changes in the heating value of the gas fired to the heater.

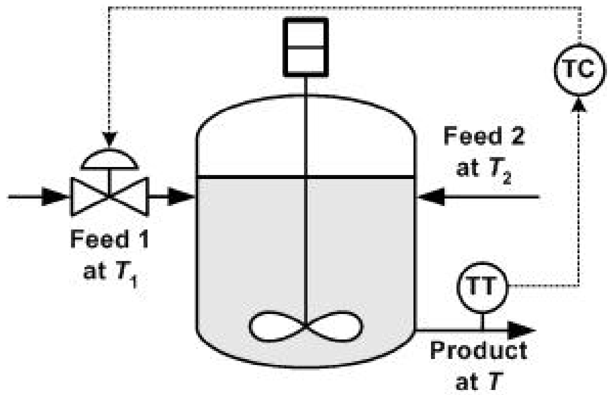


Figure P12.2.2 Control diagram for a CST thermal mixing process.

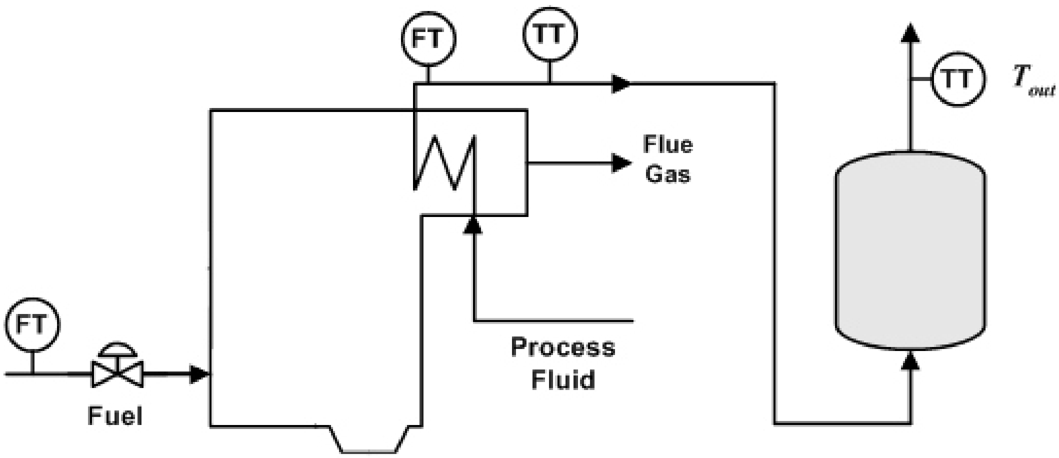


Figure P12.2.3 Schematic of a furnace that preheats the feed to a fixed-bed reactor.

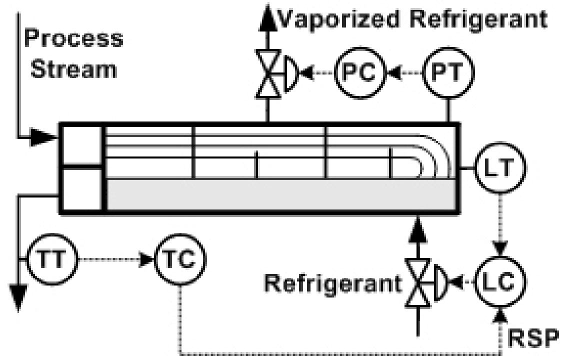


Figure P12.2.4 Control diagram for a cascade loop applied to heat exchanger.

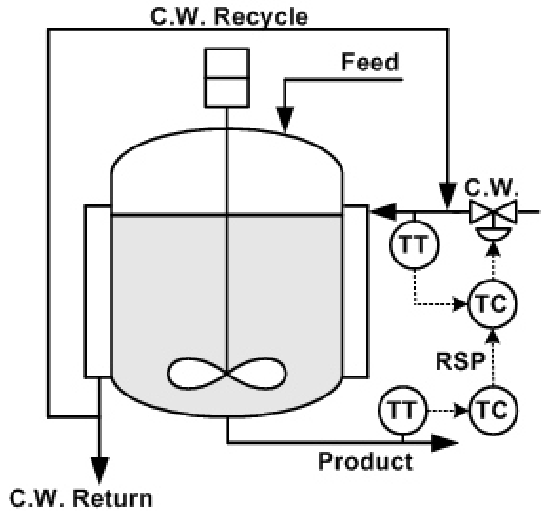


Figure P12.2.5 Schematic of a temperature controller for an exothermic CSTR in cascade with the inlet temperature of the

P12.2.4** Consider a heat exchanger that cools a process stream using a liquid refrigerant. The process stream is on the tube-side of the exchanger and the liquid refrigerant is maintained as a level on the shell-side. The vast majority of the heat transfer from the refrigerant to the process stream occurs from the liquid refrigerant due to the larger heat-transfer coefficient for the liquid than for the vapor refrigerant. As heat is transferred from the process stream to the liquid refrigerant, the refrigerant boils and leaves the heat exchanger as a vapor. A cascade control configuration has a temperature controller on the exit of the process stream setting the setpoint for the level controller on the liquid level of refrigerant in the exchanger. The level controller sets the valve position for a valve on the inlet flow of liquid refrigerant to the heat exchanger. What advantage does this cascade control arrangement have compared to having the output of the temperature controller go directly to the valve on the liquid refrigerant?

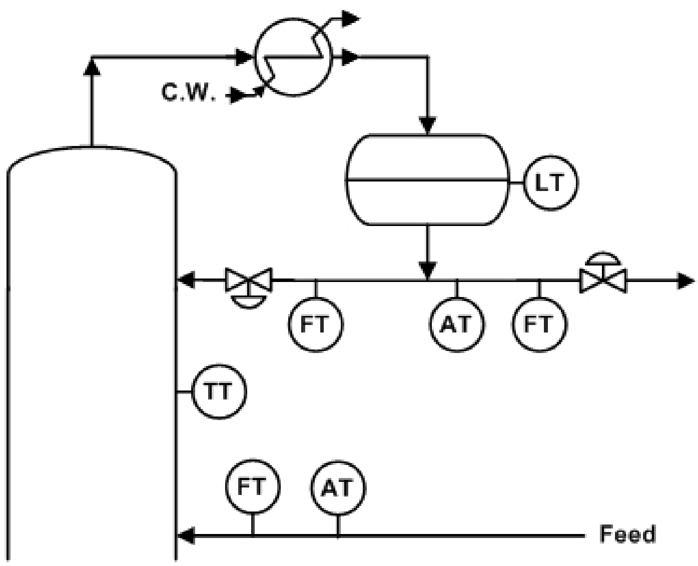


Figure P12.2.6 Schematic of the rectifying section of a column .

P12.2.5*** Figure P12.2.5 shows a control diagram of an exothermic CSTR that uses the cooling water flow to control the temperature of the CSTR. The output of the temperature controller on the product stream is cascaded to another temperature controller that adjusts the valve position for the control valve for the cooling water to control the inlet temperature of the cooling water to the reactor jacket. Identify the CVs, MVs, DVs, processes, sensors and actuator for this process. Also, make a logic flow diagram similar to Figure 1.5.1 for this cascade control system.

P12.2.6** Draw a schematic of a rectifying section of a distillation column (Figure P12.2.6) that has a composition controller for the overhead product cascaded to a tray temperature controller, which is cascaded to a flow controller for the reflux. In addition, the level controller for the accumulator is cascaded to a flow controller on the distillate product.

P12.2.7** Draw a schematic of a rectifying section of a distillation column (Figure P12.2.6) that has a composition controller for the overhead product cascaded to a tray temperature controller, which is cascaded to a flow controller for the distillate product. In addition, the level controller for the accumulator is cascaded to a flow controller on the reflux.

12.3 Ratio Control

P12.3.1** Figure P12.3.1 shows a CSTR equipped with a vent stream so that light impurities do not accumulate in the vapor space above the reaction mixture in the reactor. Add a ratio controller that adjusts the ratio of the flow rate of the vent stream to the reactor feed flow rate to maintain the composition of the light component (AT) in the reactor vapor space at a prescribed setpoint.

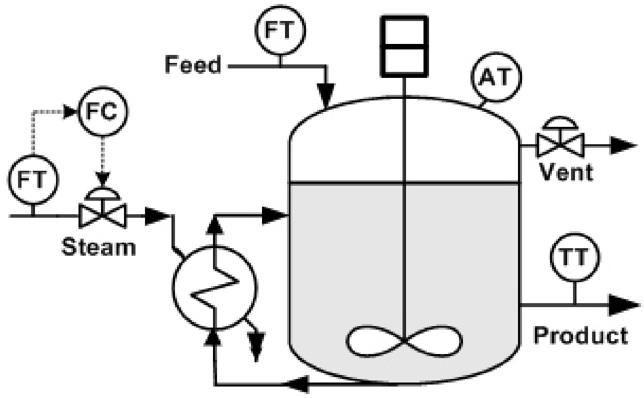


Figure P12.3.1 Schematic of a CSTR with a gas vent.

P12.3.2** Draw a schematic for the rectifying section of a distillation column (Figure P12.2.6) in which the ratio of the distillate product rate to column feed rate is set by a composition controller on the overhead product. Remember to set up a level control scheme for the accumulator.

P12.3.3** Draw a schematic for the rectifying section of a distillation column (Figure P12.2.6) in which the ratio of reflux flow rate to distillate flow rate is set by a composition controller on the overhead product. (Hint: Use a level controller to set either the reflux flow rate or the distillate flow rate and determine the other flow rate using the ratio).

P12.3.4*** Draw a schematic for the rectifying section of a distillation column (Figure P12.2.6) in which the ratio of reflux flow rate to distillate flow rate is set by a composition controller on the overhead product, which is cascaded to a tray temperature controller. The tray temperature controller is cascaded to a flow controller on the reflux. In addition, the level controller for the accumulator is cascaded to a flow controller on the distillate product.

P12.3.5*** Draw a schematic for the rectifying section of a distillation column (Figure P12.2.6) in which a composition controller on the overhead product is cascaded to a tray temperature controller. The tray temperature controller is a ratio controller (i.e., the controller output is the reflux-to-feed ratio) and its output is combined with the measured feed rate to form the setpoint for the flow controller on the reflux. In addition, the level controller for the accumulator is cascaded to a flow controller on the distillate product.

P12.3.6*** Draw a schematic for the rectifying section of a distillation column (Figure P12.2.6) in which a composition controller on the overhead product is cascaded to a tray temperature controller. The tray temperature controller is a ratio controller (i.e., the controller output is the reflux-to-feed ratio) and its output is multiplied by the measured feed rate to form the feedback portion of the setpoint for the flow controller on the reflux. In addition, the output of a feedforward controller for feed composition upsets is combined with the output of the feedback controller and the resultant is the setpoint for the flow controller for the reflux The level controller for the accumulator is cascaded to a flow controller on the distillate product.

12.4 Feedforward Control

P12.4.1** Draw a schematic of a control system that provides the same function as the ratio control system shown in Figure 12.3.5 using a combined feedforward and feedback controller.

P12.4.2** Consider the steam-heated heat exchanger (Figure P12.4.2). Draw a schematic for this process for combined feedforward and feedback control where the feedforward disturbance is the temperature of the process stream entering the heat exchanger.

P12.4.3** Consider the level process shown in Figure P12.3.3. Draw a schematic for this process for combined feedforward and feedback control where the feedforward disturbance is the inlet flow to the tank.

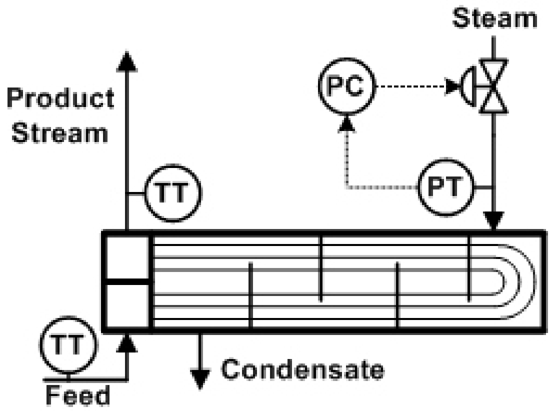


Figure P12.4.2 Schematic of a steam heated heat exchanger.

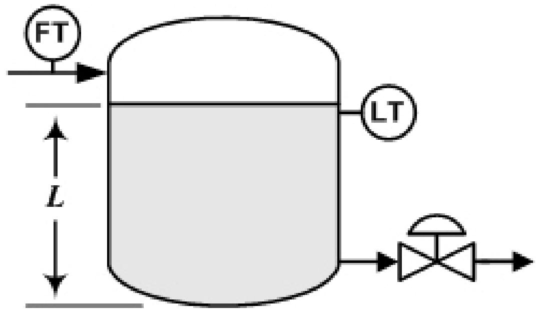


Figure P12.4.3 Schematic of a tank level process.

P12.4.4*** Draw a schematic for the rectifying section of a distillation column (Figure P12.3.6) in which a composition controller on the overhead product is cascaded to a tray temperature controller. The tray temperature controller is cascaded to the flow controller on the reflux. In addition, the feed composition is fed forward to the reflux flow controller. The level controller for the accumulator is cascaded to a flow controller on the distillate product.

P12.4.5^{S}** Apply and tune a feedforward controller to each of the following process simulators that are provided with this text. Develop FOPDT models of $G_p(s)$ and $G_d(s)$ using step test applied to the simulators. From these models, make initial guesses of the tuning parameters for the feedforward controller. Fine tune the feedforward controller and show results for a series of step changes in the measured disturbance.

a. CST thermal mixer
b. CST composition mixer
c. Endothermic CSTR
d. Heat exchanger

A superscript "**S**" indicates that the problem requires the use of one of process simulators.

Chapter 13

PID Enhancements

Chapter Objectives

- Show how inferential control can dramatically reduce analyzer deadtime using several different examples.
- Demonstrate that scheduling of the controller tuning can improve the reliability of a PID controller applied to certain nonlinear processes.
- Illustrate how override and select controls are used to satisfy process constraints.
- Demonstrate the advantages of computed MV control for specific types of disturbances.

13.1 Introduction

This chapter is concerned with enhancements for PID controllers that are designed to overcome the effects of measurement deadtime, process nonlinearity, process constraints and specific disturbances. Inferential control can greatly reduce the effect of measurement deadtime, scheduling of controller tuning can compensate for process nonlinearity, override/select control provides a direct means to use PID controls on systems that encounter process constraints and computed MV control can be used to effectively reduce the effect of certain types of disturbances.

13.2 Inferential Control

To this point, it has been assumed that the sensor in a control loop provides a direct measurement of the CV. In fact, the output of the sensor only correlates with the value of the measured variable. For example, from Chapter 2, a thermocouple exposed to a process stream at a specific temperature generates a millivolt signal that correlates strongly with the temperature of the process stream. Likewise, the level in a tank can be inferred from the pressure difference between the top and the bottom of the tank and a flow rate can be estimated from the pressure drop across an orifice plate. In this section, it is shown that **easily measured quantities, such as pressures, temperatures and flow rates, can be effectively used to infer quantities which are more difficult to measure, such as composition, extent of reaction and total cell mass in a bio-reactor.** The inferred value of the CV can be used as the value of the CV in a feedback control loop, greatly reducing the associated measurement delay, or to monitor the performance of a process.

There are three main reasons for using inferential measurement of a CV:

1. Excessive analyzer deadtime undermines the performance of a feedback loop. In Chapter 9, it was shown that a deadtime-to-time constant ratio in excess of 0.5 requires reduction to the aggressiveness of the controller and, therefore, the performance of the control loop suffers. Figure 13.2.1 shows the setpoint tracking performance for two FOPDT processes (K_p=1, $_p$=1, $_p$=0.5; K_p=1, $_p$=1, $_p$=1.5). Both systems are tuned for a 1/6 decay ratio, but the process with the larger deadtime results in a much longer response time than the smaller deadtime process. Certain techniques, such as **Smith Predictors**, have been developed to directly compensate for the deadtime of a process using process models. Smith Predictors have been studied extensively in academia, but have sparingly been applied industrially. Smith Predictors are typically difficult to implement and their effectiveness is sensitive to modeling errors. Therefore, inferential measurements are the industrial method of choice for counteracting large measurement delays for CVs. Inferential measurements can greatly reduce the measurement deadtime because they are based on measurements (e.g., temperatures, pressures and flows) that have relatively low levels of measurement deadtime.

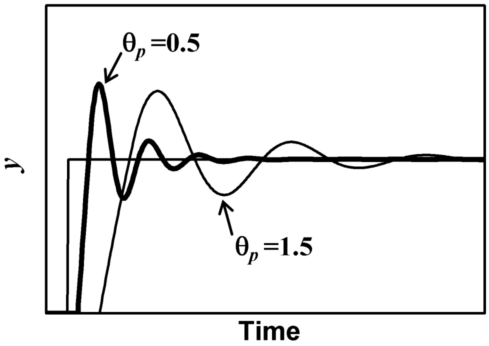


Figure 13.2.1 Setpoint tracking performance showing the effect of deadtime on closed-loop performance.

2. The total cost (i.e., the purchase price and maintenance cost) of an on-line analyzer can be excessive. Because inferential measurements are typically based on temperature, pressure and flow measurements, they are much less expensive to install and maintain.

3. An on-line analyzer may not be available and making inferential measurement is the only option for feedback control or process monitoring.

For inferential control to be effective, the inferential measurement must correlate strongly with the CV of interest and this correlation should be relatively insensitive to unmeasured disturbances. Following are several examples that illustrate how inferential measurements can be effectively applied in the CPI and bio-tech industries.

Example 13.1 Inferential Temperature Control for Distillation

Problem Statement. Evaluate the use of tray temperature measurements to infer product compositions for distillation columns.

Solution. Tray temperatures correlate very well with product compositions for many distillation columns; therefore, inferential control of distillation product composition is a widely used form of inferential control. Figure 13.2.2 shows the correlation between propane content in the bottoms product and the tray temperature for two trays in the stripping section of a propane/butane binary column. Because this is a binary separation, the temperature and pressure of a tray define the composition on that tray. The largest temperature change for a fixed change in the bottom product composition occurs for tray number 10; therefore, the temperature of tray 10 can be

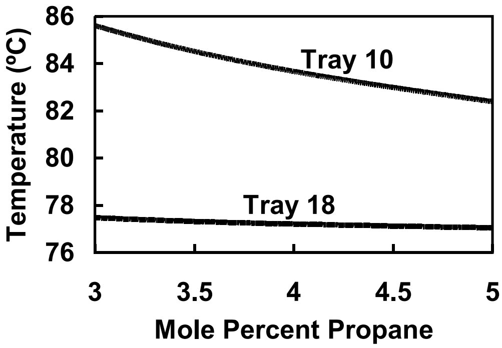


Figure 13.2.2 The effect of product impurity level on two different tray temperatures.

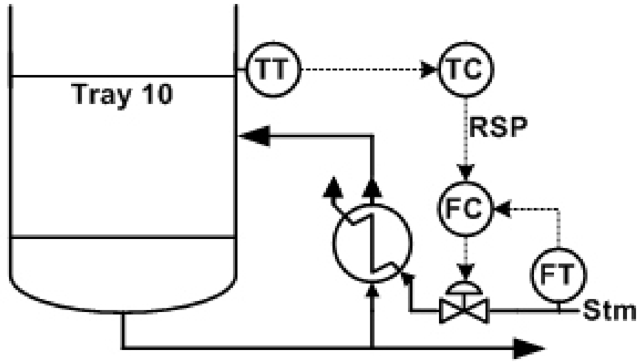


Figure 13.2.3 Control diagram for inferential control of the bottoms product composition of a distillation column.

used to infer the bottom product purity of this column. Figure 13.2.3 shows the control diagram for inferential temperature control of the bottom product composition for this column. The tray temperature controller is cascaded to a flow controller on the steam. This case is equivalent to the cascade control example shown in Figure 12.2.2b. Nevertheless, this is an application of inferential control because the setpoint for the temperature controller in Figure 13.2.3 is chosen so that the desired bottom product composition is attained.

For a multi-component distillation column, the tray temperature does not define the product composition. The liquid on a tray can have the same concentration of light and heavy keys with different relative amounts of heavy non-key and light non-key, and the resulting tray temperature changes significantly. As the amount of heavy non-key is increased and the amount of light non-key is decreased, the tray temperature increases even though the proportions of light and heavy keys remain unchanged. Figure 13.2.4 shows a tray temperature in the stripping section of a multi-component distillation column as a function of the light key in the bottoms product for different ratios of heavy non-key to light non-key for the column feed (i.e., a light and a heavy feed). Note that the two curves are parallel with a difference of about 2 C. As a result, controlling a tray temperature to a fixed temperature results in offset as the feed composition changes. To remove this offset, a composition controller

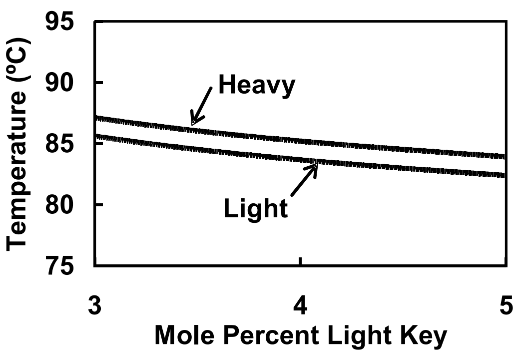


Figure 13.2.4 The effect of feed composition on the correlation between tray temperature and impurity level in the bottoms product.

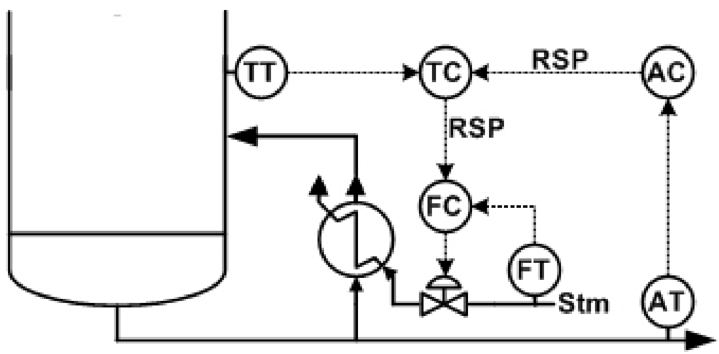


Figure 13.2.5 Control diagram of an analyzer controller cascaded to a temperature controller for bottoms composition control of a distillation column.

that uses an on-line composition analyzer can be cascaded to the tray temperature control loop (Figure 13.2.5). In certain cases, laboratory analysis results taken once per shift or once per day are used by the operator to select the setpoint for the temperature controller in an effort to remove this offset.

Example 13.2 Inferential Reaction Conversion Control

Problem Statement. Consider an adiabatic fixed-bed reactor (Figure 13.2.6). Using macroscopic mole and energy balances for this system, develop an inferential estimate of the conversion in this reactor. Assume an irreversible first-order reaction.

Solution. The total rate of consumption of component A (R_A) can be expressed in terms of the fractional conversion of A [X_A, where $X_A = (C_{A_0} - C_A)/C_{A_0}$].

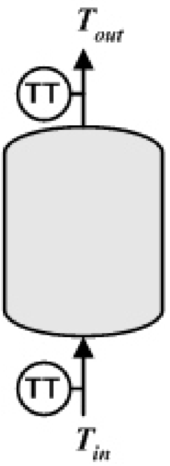


Figure 13.2.6 Schematic of a fixed-bed reactor.

$$R_A = F_V C_{A_0} X_A \qquad \textbf{13.2.1}$$

Assuming no phase changes occur in this system, the macroscopic energy balance (Equation 3.4.3) for this process is given as

$$0 = F_V \rho C_p (T_{in} - T_{out}) - \Delta H_{rxn} R_A$$

where C_{A_0} is the inlet concentration of A to the reactor, ΔH_{rxn} is the heat of reaction, F_V is the volumetric feed rate to the reactor, ρ is the average density of the process stream, C_P is the average heat capacity of the process stream, T_{out} is the temperature of the outlet stream from the reactor and T_{in} is the temperature of the inlet stream to the reactor. Substituting Equation 13.2.1 into this equation and rearranging yields

$$X_A = \frac{\rho C_P}{C_{A_0}(-\Delta H_{rxn})}(T_{out} - T_{in}) \qquad \textbf{13.2.2}$$

Note that this relationship is not affected by changes in the feed rate although the feed rate will affect T_{out}, and thus, affects X_A. In an industrial reactor, there are heat losses, side reactions and variations in the physical parameters; therefore based on the form of Equation 13.2.2, the assumed inferential relationship is

$$X_A = a(T_{out} - T_{in}) + b \qquad \textbf{13.2.3}$$

A plot of the experimental data for a reactor (X_A, T_{out} and T_{in}) can be used to determine a and b as well as check the validity of this functional form (i.e., a plot of X_A versus $(T_{out} - T_{in})$ should be linear). Note that the temperature difference across the reactor needs to be large enough that noise on the temperature measurement does not significantly affect the measured temperature change across the reactor. Once a and b are identified, the inlet temperature (T_{in}) can be adjusted to maintain a fixed reaction conversion, X_A. Periodically, composition measurements for the product leaving the reactor can be made and the results used to update the value of b in the previous equation. The value of b, instead of the value of a, should be updated with plant data because a is less likely to change significantly with changes in the operating conditions compared with b.

Example 13.3 Inferential Estimate of Total Cell Mass for a Bio-Reactor

Problem Statement. Bio-reactors start with the contents of the seed reactor (Figure 1.1.2) and must grow the microorganisms to a level where the final products are produced. As a result, when operating a bio-reactor, it is important to monitor the total cell mass. The total cell mass can be measured on-line by drawing a sample of the reaction broth, removing the water from the sample and weighing the remaining cell mass. Thus, the mass of cells measured from the sample, the sample volume and the total volume of the bio-reactor are used to determine the total cell mass. This analysis is slow, and it is desirable to have more frequent cell mass estimates than are feasible to test by sampling the broth. Therefore, an inferential measurement of the total cell mass in a bio-reactor is useful. During the cell growth stages of a bio-reactor, CO_2 is evolved from the consumption of the substrate (e.g., glucose) for cell maintenance and cell growth. Develop an inferential measurement of the total cell mass in a bio-reactor (Figure 13.2.7). Assume that periodically broth samples are measured off-line for the mass of cells in the broth sample and at the same time the off-gas from the bio-reactor is measured for the CO_2 concentration and flow rate.

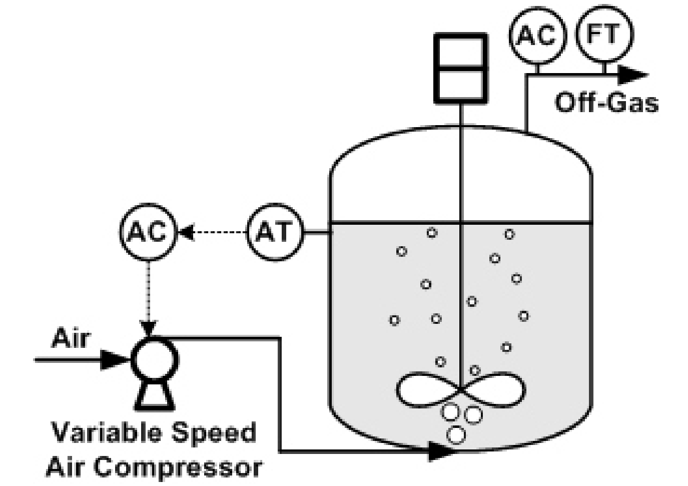


Figure 13.2.7 Schematic of a bio-reactor with process measurements for an inferential total cell mass sensor.

Solution. The CO_2 evolution rate resulting from maintaining the existing cells ($F_{CO_2}^{Main}$) is directly proportional to the current cell concentration, x, i.e.,

$$F_{CO_2}^{Main} \quad k_1 V x$$

where k_1 is a proportionality constant, V is the volume of broth in the bio-reactor and x is the cell concentration. The CO_2 evolution rate resulting from cell growth ($F_{CO_2}^{Growth}$) is directly proportional to the rate of cell growth, i.e.,

$$F_{CO_2}^{Growth} \quad k_2 V \frac{dx}{dt}$$

where k_2 is a proportionality constant. During the cell growth stage, the total CO_2 production rate is given as

$$F_{CO_2}^{Tot} \quad F_{CO_2}^{Main} \quad F_{CO_2}^{Growth} \quad k_1 V x \quad k_2 V \frac{dx}{dt} \qquad \textbf{13.2.4}$$

During this stage, it is important that the cell growth is not inhibited by a large substrate concentration or limited by a substrate concentration that is too low. Therefore, during this phase it is reasonable to model cell growth as an exponential function of time, i.e.,

$$x \quad x_0 e^{\ t}$$

where   is an assumed constant and x_0 is the initial concentration of cells. Substituting this equation into Equation 13.2.4 and rearranging results in

$$F_{CO_2}^{Tot} \quad Vx_0 e^{\ t}(k_1 \quad k_2 \ ) \qquad \textbf{13.2.5}$$

Now consider how to match this model to process data. Assume that cell mass measurements are made periodically during the growth stage. Considering Figure 13.2.7, the total production rate of CO_2 can be determined using the measured concentration of CO_2 and the flow rate of the off gas. By applying Equation 13.2.5 to the cell mass measurement and its corresponding CO_2 generation rate at time t_1, the term $(k_1 \quad k_2)$ can be determined

$$(k_1 \quad k_2 \ ) \quad \frac{F_{CO_2}^{Tot}(t_1)}{Vx(t_1)}$$

Once $(k_1 \quad k_2)$ is determined, Equation 13.2.5 can be used to estimate the current value of the total cell mass $(Vx_0 e^{\ t})$ based on the CO_2 generation rate using the following equation

$$\text{Total Cell Mass}\,(t) \quad Vx_0 e^{\ t} \quad \frac{F_{CO_2}^{Tot}(t)}{(k_1 \quad k_2)} \qquad \textbf{13.2.6}$$

Because changes in the CO_2 generation rate are the most immediate indication of a change in the growth rate of the cells during the growth stage in a bio-reactor, Equation 13.2.6 provides a direct measurement of the total cell mass in the bio-reactor.

Soft Sensors Based on Neural Networks. In electric power generating stations, restricting the NO_x (nitrogen oxide compounds) emissions in the flue gas to acceptable levels is important because NO_x compounds contribute to air pollution. Typically, on-line analyzers are used to measure the NO_x in the flue gas from the boilers. Occasionally, the NO_x analyzers on a boiler fail. If the NO_x level is not measured, the power companies must pay a fine for emissions. Instead of installing additional on-line NO_x analyzers, which are quite expensive, a number of power companies have applied a type of inferential estimator to predict the NO_x level in their flue gas.

Instead of using one or two process measurements, all the measured process conditions (e.g., fuel feed rate, oxygen in the flue gas, heating value of the fuel, ambient air temperature, etc.) can be empirically correlated to predict the NO_x concentration in the flue gas. The empirical correlation is based on training an **artificial neural network** (ANN) to predict the flue gas NO_x concentration from all the available data. A network with three input nodes, four nodes in the hidden layer and one output node is shown schematically in Figure 13.2.8. Inputs to each node are summed and the resultant is transformed (to a -1 to +1 node output) by a nonlinear function (usually a sigmoidal function such as the hyperbolic tangent function. Weights (Figure 13.2.8) are multiplied by the values that pass through them and are automatically selected by the neural net tool during a several iteration process in order to minimize the error between the

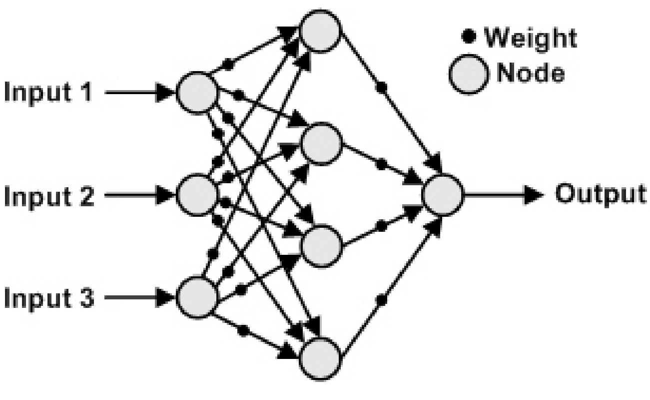


Figure 13.2.8 Diagram of a network with three input nodes, four nodes in the hidden layer and one output node.

neural net output prediction and a portion of the actual data of the output parameter. For Figure 13.2.8, there are a total of 16 adjustable weights for the neural network that are determined such that the neural network predicts the subset of process output data used in its "training" (i.e., learning) phase. That is, developing a neural net model involves separating the available process data into two parts (often a 50:50 split), with one part used to train the neural net (which sets the weights) and the second part used to test the now trained model to see how well it predicts data that was not used in its development . As a result, neural networks can be used as empirical nonlinear input/output models. A neural network for a NO_x analyzer could have over 500 weights (adjustable parameters). Note that: it is important that the amount of data used in training the neural net be substantially more than the number of weights to prevent the neural net from memorizing the data presented to it. This inferential NO_x analyzer is also referred to as a **soft sensor** because the neural network software, along with the process measurements, is used to provide the on-line measurement.

Note that neural nets are generally very good at modeling non-linear processes and interpolating incoming data that is within the boundaries of its training data set. However, such models can be poor at extrapolating outside the boundaries of its training data set. Also, neural nets are created without an engineer having to create a structure of the "soft sensor" algorithm such as would be needed for an algorithm based on deterministic/mechanistic first principles or a statistical stochastic model. This is a very powerful feature, as often, in desiring to develop a soft sensor, it is not known if a good correlation exists between a parameter to be predicted/estimated and the wealth of process measurements available. A neural net can establish efficiently and quickly whether a good correlation exists. Commercial neural network tools which operate on a PC are available for developing neural networks.

A common practice in developing a neural net is to begin by assigning an input node to each of the input measurements available, then having the neural net tool train the neural network application. Once a good correlation is shown to exist, the user can then delete one or more of the input nodes that are thought as possibly unimportant and retrain the remaining network. A resulting good correlation then confirms that the deleting of certain inputs is okay, which then, of course, results in a simpler model with fewer nodes and weights. This process can be repeated until just the important input data remains.

One of the challenges in developing neural networks for a batch process can be aligning the input data such that all slices of data represent measurements in approximately the same points in time. Sometimes, time adjustments to certain data (e.g., ones with significant dead time) are needed so that the input data slices represent measurements at the same points in time. In the final analysis, it is up to the engineer whether to keep the correlation as a neural network in a process monitoring or control application or whether to then investigate other structural forms for the model (e.g., first principles equations).

ANN-based soft sensors are used for a variety of applications in the bio-tech industries. For example, during the product production phase of a bio-reactor (i.e., after the growth phase when the operating conditions are such that the microorganisms produce the desired product), the carbon from the substrate is converted by the cells into CO_2, aldehydes, the product and compounds to build cells. Due to the complex nature of this process, ANNs are used to correlate the total cell mass to several operating parameters of the process. Operating data and cell mass measurements are used to train these ANNs. Once trained the ANN is used to predict the total cell mass in the bio-reactor during the product phase of the operation. As another example, ANNs are used to estimate the concentration of product proteins inside microorganisms (e.g., *E. Coli*) based on the process operating conditions and certain measurements.

Self-Assessment Questions

Q13.2.1 Why are inferential measurements used industrially?

Q13.2.2 Indicate how an energy balance on a CSTR can be used to estimate the amount of conversion occurring in the reactor. What assumptions and limitations does this inferential estimator have?

Q13.2.3 What is a soft sensor?

Self-Assessment Answers

Q13.2.1 Inferential measurements are used industrially in certain cases because they offer on-line measurements that have less analyzer delay and/or are less expensive to install and maintain.

Q13.2.2 An energy balance on a CSTR can be used to determine the amount of heat given off by reactions in the reactor. If there is one dominant reaction or if several reactions occur in some known proportions, the heat of reaction can be used directly to determine the degree of reaction. If there are more than one main reaction and the relative proportions of the main reactions are unknown and variable, using the energy balance to estimate the degree of reaction can be quite inaccurate. In addition, this analysis assumes that the process is at or near steady state.

Q13.2.3 A soft sensor is a software-based model of a property of a process that uses process measurements, such as temperatures, pressures and flow rates, as inputs to calculate the property of interest.

13.3 Scheduling Controller Tuning

In Chapter 9, it was demonstrated that a controller on a nonlinear process can become highly oscillatory in certain situations and can become extremely sluggish (e.g., Figure 13.3.1) at other times. If the process gain increases by over 50%, the controller is likely to ring or become unstable and if the process gain decreases by 50% or more, the process can be expected to behave sluggishly. Tuning PID controllers for the case with the largest process gain can eliminate unstable operation, but at the expense of largely sluggish performance. The combination of the magnitude of the disturbances and the inherent process nonlinearity determine the degree of variation in the FOPDT model parameters of the process. For a number of processes, certain measurements (e.g., the CV or a measured disturbance) directly indicate whether the process parameters have increased or decreased and by how much; therefore, **scheduling of the controller tuning based on process measurements can be an effective means of compensating for process nonlinearity in certain cases**. The CV and the feed rate are examples of such key process measurements that can typically be used to schedule the controller tuning.

Example 13.4 The Application of Scheduling of the Tuning Parameters for a CSTR

Problem Statement. Develop a scheduling procedure for a PI controller for the CSTR process presented in Example 3.5.

Solution. For the CSTR in Example 3.4, the reaction rate constant is represented by an Arrhenius rate expression, i.e.,

$$k \qquad k_o\, e^{\,E/RT}$$

which indicates strong nonlinearity for this process with regard to changes in the reactor temperature. Figure 13.3.1 shows the response of the temperature to a feed composition disturbances under PI control. The controller is tuned for a region near the setpoint. The feed composition decrease causes the reactor temperature to increase and a feed composition increase causes the reactor temperature to decrease. Note that when the reaction

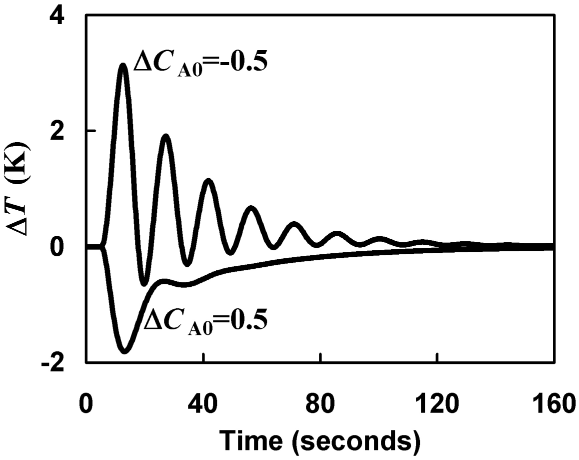


Figure 13.3.1 The effect of feed composition upsets on the PI feedback behavior for the endothermic CSTR.

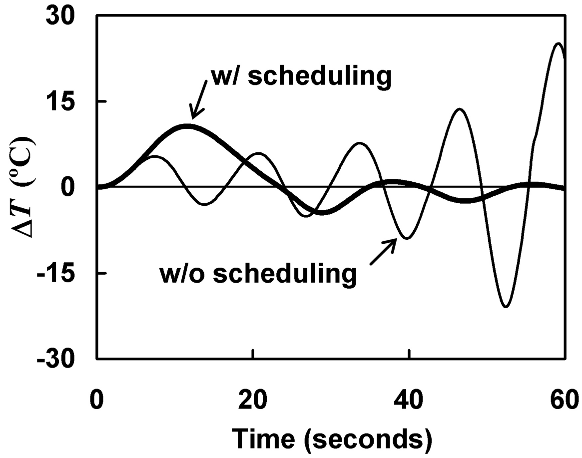


Figure 13.3.2 Comparison between a conventional PI controller and a PI controller with scheduling of the controller tuning for a severe feed composition.

temperature increases, the closed-loop response begins to ring and when the temperature decreases, the closed-loop response becomes sluggish.

Now consider an approach for which the controller aggressiveness is adjusted based on the reactor temperature. The PI tuning factor, F_T (Equation 9.9.2), is scheduled as a function of the reactor temperature (Table 13.1). Figure 13.3.2 shows a comparison between a conventional PI controller and a PI controller with scheduling for a severe feed composition upset for the endothermic CSTR. For the scheduled controller, F_T is adjusted based on the error from setpoint according to Table 13.1. Note that when the temperature is less than the setpoint, the controller is tuned more aggressively, and when the temperature is larger than the setpoint, it is detuned. In fact, F_T changes by a factor for the gain scheduling shown in Table 13.1. For this case, scheduling of the controller tuning was able to maintain stability while a conventional PI controller was not. The slow movement towards setpoint indicates that not enough integral action is used when the reactor temperature is below setpoint. Scheduling the entire set of controller parameters can provide the correct integral action for each temperature region. This can be accomplished by tuning the controller at several different reactor temperatures and using the results to schedule each of the tuning parameters.

Table 13.1 Scheduling of the Controller Tuning for the CSTR.

$(T\text{-}T_{sp})$	-4	-2	0	2	4	6	8	10
F_T	1.4	1.6	1.8	2.8	3.8	4.8	5.8	6.8

Example 13.5 The Application of Scheduling of the Tuning Parameters for a Heat Exchanger

Problem Statement. Develop a scheduling procedure for a PI controller for the heat exchanger process presented in an earlier version of this text[1].

Solution. Consider the heat exchanger shown in Figure 13.3.3. As the feed to the heat exchanger flows through the tube bundle, it is heated by steam condensing on the shell side. As the feed rate changes, the residence time of the feed in the tubes changes. Figure 13.3.4 shows the open-loop responses for three different feed rates. The feed rate is represented by the average fluid velocity (v) in the tubes. Both the gain and the dynamic response change as the feed rate is changed. Table 13.2 lists the FOPDT model parameters for each flow rate. Note that the gain and the deadtime of each change by about 250% from the largest to the smallest flow rates. The PI Cohen and Coon settings (Table 9.2) at each flow rate are also listed in Table 13.2. The controller gain changes by a factor of 5 while the reset time changes are more gradual. It is clear from these results that it is not reasonable to expect one set of PI controller settings to work effectively for significant changes in the feed rate to this heat exchanger. The temperature controller for the outlet of the heat exchanger tuned for v=7 ft/s becomes unstable when the feed rate is reduced to v=4 ft/s. Conversely, if the controller is tuned for the low flow rate condition, it performs sluggishly for the high flow rate conditions. Figure 13.3.5 shows results with and without scheduling of the controller tuning based on feed rate for a change in the velocity through the tubes from 7 to 4 ft/s. The controller without scheduling was tuned for a feed rate corresponding to v= 7 ft/s.

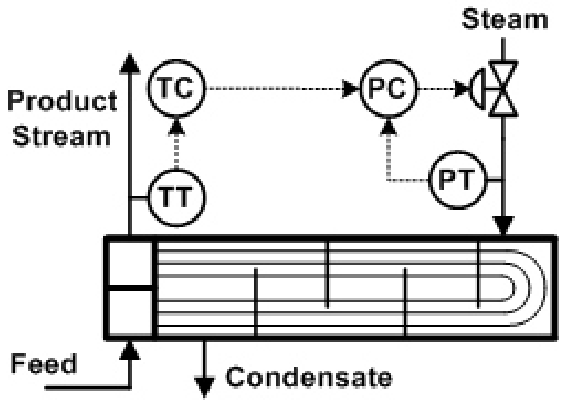


Figure 13.3.3 Control diagram of a heat exchanger for which the outlet temperature is controlled by the steam pressure.

Table 13.2
FOPDT and PI Tuning Parameters for the Heat Exchanger Case as a Function of Feed Rate

	v=4 ft/s	v=7 ft/s	v=10 ft/s
K_p	0.25	0.15	0.11
$_p$	10.7	9.9	10.0
$_p$	10.2	5.8	4.0
K_c	4.2	10.8	21.3
$_I$	12.0	8.8	7.3

Non-stationary Behavior. Consider a wastewater neutralization process (Figure 12.3.3). If the titration curve of the wastewater and the other process parameters remain fixed, the process is referred to as **stationary** or **time-invariant**. If the titration curve changes with respect to time, the process is **non-stationary** or **time-varying**. In the case of this pH control example, changes in the titration curve can have an overwhelming effect on the process gains. There are many more examples of non-stationary behavior that result in much more gradual process gain changes. The following are several examples of non-stationary behavior in the CPI and bio-tech industries:

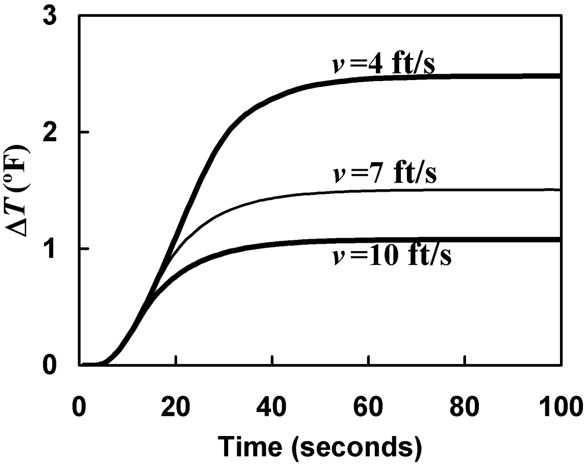


Figure 13.3.4 Open-loop response for a heat exchanger for different feed rates.

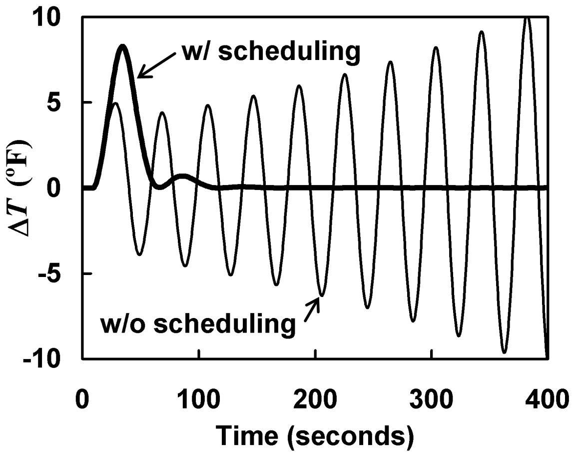


Figure 13.3.5 Closed-loop results for a step change in feed rate with and without scheduling of the controller tuning for the heat exchanger case.

1. Catalyst deactivation
2. Heat exchanger fouling
3. Fouling of trays in a distillation column
4. Feed composition changes that affect the process parameters (K_p, $\ _p$ and $\ _p$)
5. Changes in the production rate of bio-reactors due to an increase in the product concentration
6. A dramatic change to the airflow and/or agitation needed in a bio-reactor due to exponentially increasing cell mass and its associated demand for oxygen

These effects can be large enough that controller retuning is required. If an overall tuning factor, F_T (Equation 9.9.2) has been used, you can adjust F_T in a straightforward manner to compensate for the non-stationary behavior. Control methods that adjust controller tuning to adapt to non-stationary behavior are referred to as **adaptive control** techniques. Adaptive control techniques can be effectively applied for processes that vary slowly. An adaptive controller is expected to handle gradual catalyst deactivation that occurs over several days, but is not expected to handle sharp changes in catalyst activity that occur within one hour. A number of commercially available adaptive controllers are referred to as **self-tuning controllers** and can usually be installed on a DCS or control computer. While there is a range of approaches used for self-tuning controllers, they are generally limited to processes that vary in a gradual, consistent manner.

Self-Assessment Questions

Q13.3.1 How does scheduling controller tuning prevent a nonlinear process from going unstable or behaving sluggishly?

Q13.3.2 How do you determine whether scheduling of the tuning parameters of a controller will be effective?

Q13.3.3 Why is the controller scheduling based on the CV for Example 13.4 and based on a measured DV for Example 13.5?

Self-Assessment Answers

Q13.3.1 When a process measurement (e.g., the CV or a disturbance) correlates directly with the nonlinearity of the process, scheduling the tuning parameters of the controller can eliminate oscillatory or unstable behavior and sluggish behavior by adjusting the controller tuning parameters to compensate for the nonlinear changes in the process.

Q13.3.2 To determine if scheduling of the controller parameters will be effective, first determine if the process is sufficiently nonlinear to warrant this approach. A process is usually nonlinear enough to warrant scheduling of the controller parameters when both sluggish and oscillatory or unstable behavior is observed for a constant set of controller tuning parameters. If it is determined that the process is sufficiently nonlinear, the next step is to determine if a process measurement correlates strongly with the observed nonlinear behavior of the process. For example, when oscillatory behavior is observed, is the CV always in its high range and is the CV in its low range when the process is sluggish? A phenomenological understanding of the process is useful in answering the latter question.

Q13.3.3 Example 13.4 schedules the controller tuning based on the CV value because the nonlinearity of this process varies with the value of the CV. For Example 13.5, the disturbance (flow rate of the process fluid) causes significant changes in the process gain and process deadtime; therefore, the controller tuning in this case is based on the flow rate.

13.4 Override/Select Control

Constraints are a natural part of industrial process control. As processes are pushed to produce as much product as possible and thus increase profits, process limits are inevitably encountered. When an upper or lower limit on a MV is encountered or when an upper or lower value of a controlled or output variable from the process is reached, it is necessary to alter the control configuration to prevent the violation of a constraint. **Effective industrial controller implementation requires that safeguards be installed to prevent the process from violating safety, environmental or economic constraints.** These constraints can be met using override/select controls.

Example 13.6 Override/Select Control of a Gas-Fired Heater

Problem Statement. Analyze the operation of the override/select controls for the gas-fired heater shown in Figure 13.4.1.

Solution. Under normal operating conditions, the fuel flow rate is adjusted to control the exit temperature of the process fluid. As the feed rate of the process fluid increases, the furnace tube temperature increases. At some point, the upper limit on furnace tube temperature (an operational constraint) is encountered. The fuel flow rate to the furnace must be adjusted to keep the furnace tube temperature from exceeding its upper limit. If the tube temperature constraint is exceeded, damage to the furnace tubes results, significantly reducing their useful life. Figure 13.4.1 shows that the output of both control loops (the temperature controller on the process fluid and the temperature controller on the furnace tubes) are combined and the lower fuel feed rate is actually applied. The "LS" symbol in Figure 13.4.1 is called a **low select (LS)** and indicates that the lower fuel feed rate is chosen. There are two separate loops that use fuel flow rate as a MV and the LS controller switches between them as the flow rate of the process fluid changes. When the feed rate is sufficiently low that the temperature of the process fluid can be controlled to setpoint, the output of the process fluid temperature controller is selected because it is lower than the output of the tube temperature controller. That is, when the tube temperature is below the maximum tube temperature constraint, the tube temperature controller will call for an increase in the fuel firing rate, which is rejected by the LS. Likewise, when the tube temperature exceeds its upper limit, the output of the tube temperature controller is reduced by the tube temperature controller and it is selected by the

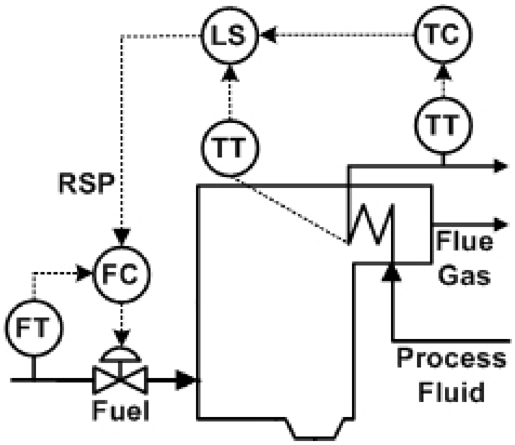


Figure 13.4.1 A control diagram of a furnace fired heater with low select firing controls.

LS. When the tube temperature controller is controlling the furnace firing rate, the outlet temperature is not controlled to setpoint because the output of the product temperature controller is rejected by the LS.

Example 13.7 Override/Select Control of a Distillation Column

Problem Statement. Analyze the operation of the override/select controls for the stripping section of the distillation column shown in Figure 13.4.2.

Solution. Flooding in a distillation column can result as the feed to a column is increased. The onset of flooding is usually identified when the pressure drop across the column or a portion of the column increases sharply. When the pressure drop across the column reaches an upper operational limit (usually identified by experience), the reboiler duty is switched from controlling the bottom product composition to maintaining operation at the maximum pressure drop across the column (Figure 13.4.2). A LS controller is used for this application. Two separate control loops use reboiler duty as a MV, and the LS controller switches between them as the feed rate to the process changes. When the column feed rate is reduced while operating at the maximum differential pressure across the column, the composition of the impurity in the bottoms product becomes less than its setpoint and the reboiler duty called for by the composition control loop will be less than that called for by the differential pressure controller. At this point, the LS controller uses the output from the composition controller. On the other hand, when the column differential pressure reaches its upper limit, the control scheme switches to using the output of the differential pressure controller.

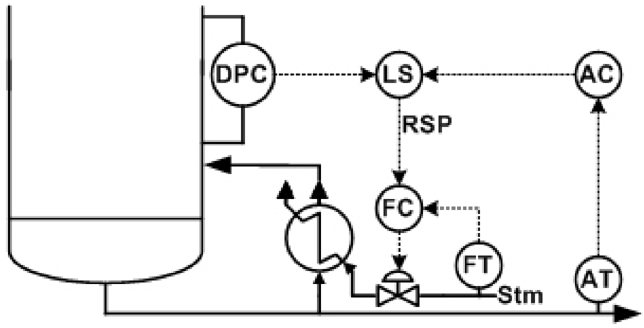


Figure 13.4.2 Control diagram of the stripping section of a distillation column with low select controls applied to prevent flooding of the column.

Example 13.8 High Select Control of a Fixed-Bed Reactor

Problem Statement. Analyze the operation of the high select controls for the fixed-bed reactor shown in Figure 13.4.3.

Solution. A **high select (HS)** controller (Figure 13.4.3) can be used to control the maximum temperature in a fixed-bed reactor even when the maximum reactor temperature occurs at different locations in the reactor. For certain reactors, if the catalyst in the reactor exceeds an upper temperature limit, damage to the catalyst occurs. The HS controller chooses the largest temperature measurement from a number of temperature measurements and the largest reading is sent to the temperature controller. In this manner, the highest reactor temperature can be maintained below a preset upper limit. Low select (LS) controllers can also be used where the lowest reading is selected from several readings.

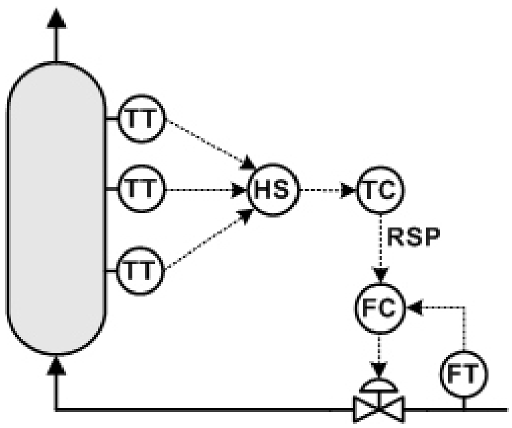


Figure 13.4.3 Control diagram of a fixed-bed reactor with a high select.

Example 13.9 Override/Select Control of a Reboiler

Problem Statement. Analyze the operation of the override/select controls for the stripping section of the distillation column shown in Figure 13.4.4, which has an upper limit on its reboiler duty.

Solution. When the remote setpoint for the steam flow rate to the reboiler is consistently greater than the measured steam flow, a select controller switches to using the column feed rate as a MV to keep the bottom product purity on specification. When the column feed rate is adjusted back to its normal level and the control valve on the steam to the reboiler is no longer saturated (i.e., fully open), the control configuration is changed so that the reboiler duty is manipulated to control the bottom product purity. This is an example of the selection of a secondary MV when the primary MV reaches a limit (becomes saturated).

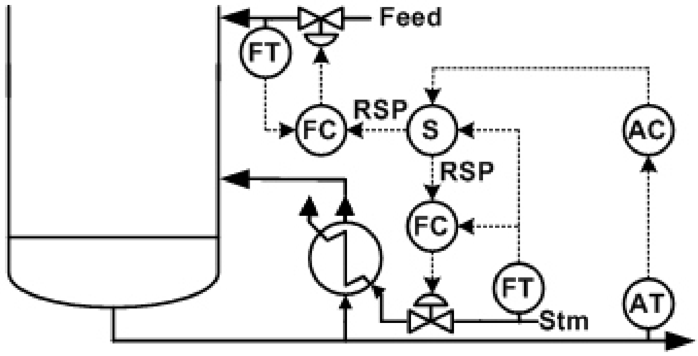


Figure 13.4.4 Control diagram of the stripping section of a distillation column with select control to maintain bottom product purity when a maximum reboiler constraint is encountered.

Example 13.10 Cross-Limiting Firing Controls for a Boiler

Problem Statement. Analyze the operation of the cross-limiting firing controls for a boiler shown in Figure 13.4.5.

Solution. For furnaces, it is important to ensure that excess air is always supplied with the fuel to prevent the formation of carbon monoxide (CO), a serious safety hazard. That is, when there is insufficient oxygen for complete combustion of the fuel to CO_2 and H_2O, CO will form. Furnaces are normally equipped with CO sensors that shut down the furnace if CO levels exceed specified limits. **Cross-limiting firing controls** (Figure 13.4.5) are designed to reduce the likelihood that CO is formed during changes in the firing rate to the furnace. To understand the schematic of cross-limiting firing controls (Figure 13.4.5), you must recognize that all of the signals in Figure 13.4.5 are based on fuel flow rate, except for the signal from the air flow measurement. The setpoint to the flow controller for the air is the fuel equivalent for the desired air flow rate. The measured air flow rate is multiplied by the fuel-to-air ratio and the resultant is compared with the setpoint by the flow controller. Therefore, both FCs, the LS and the HS are based on the fuel firing rate.

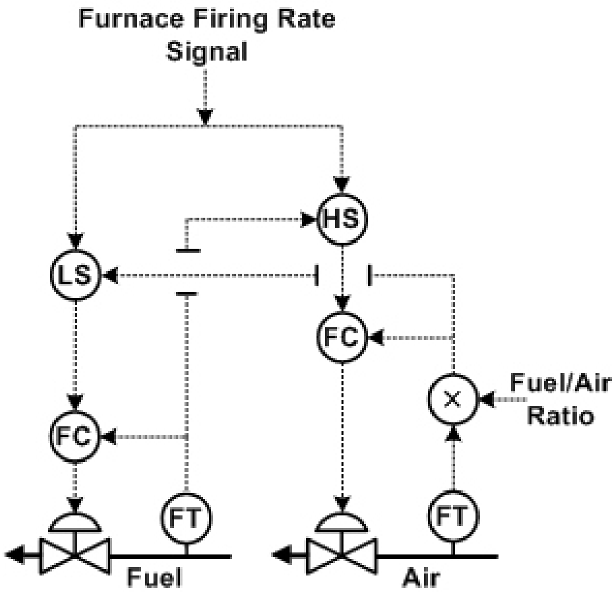


Figure 13.4.5 Control diagram of cross-limiting firing controls.

Consider the firing rate increase shown in Figure 13.4.6a. The increased firing rate signal is rejected by the LS

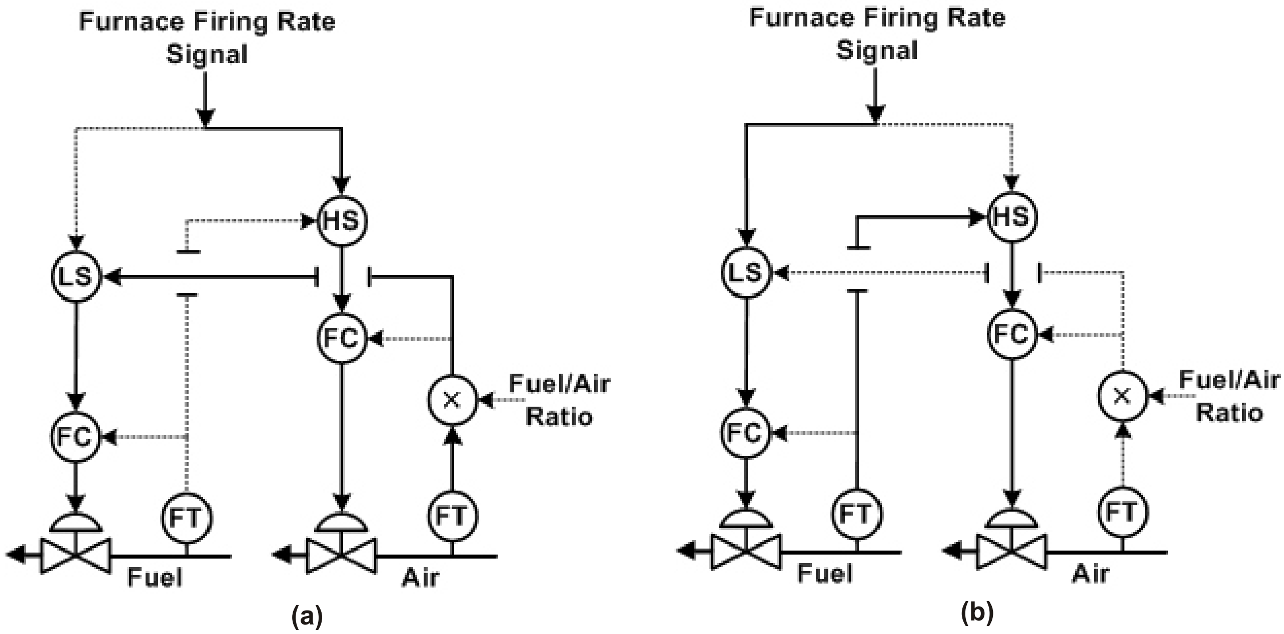


Figure 13.4.6 Control diagram of cross-limiting firing controls showing the signal route for (a) an increase in firing rate and (b) a decrease in firing rate.

because the fuel flow rate corresponding to the measured air flow rate is smaller. On the other hand, the HS selects the increased firing rate signal because it is larger than the current measured fuel flow rate. As a result, the setpoint for the FC on the air flow rate receives the firing rate increase, which in turn increases the air flow rate. As the air flow increases, the fuel flow rate corresponding to the air flow rate increases, thus increasing the setpoint for the FC on the fuel. Therefore, for a firing rate increase, the air flow rate increases first followed by the fuel flow rate, thus maintaining an excess of O_2 preventing CO formation during firing rate increases. Similarly, for firing rate decreases (Figure 13.5.6b), cross-limiting firing controls cuts the fuel flow rate first and the air flow rate follows, thus maintaining an excess of O_2.

There are also several other applications where on-line programmable/configurable logic (often implemented as "if-then-else" rules) is involved in making selections and/or dealing with constraints. These include, e.g., 1) enabling/disabling alarms and changing an alarm's priority attribute as a function of batch steps and 2) selecting controller setpoints (or putting controllers in manual and selecting valve positions and other programmable actions when an operator (or computer) puts a process in an ABORT state (e.g., due to a runaway exothermic reactor or an accidental environmental release of a hazardous material). Also, various "interlocks" associated with a process may need to be selected for certain process operations and released for others.

Self-Assessment Questions

Q13.4.1 How is a low select element (LS) used to satisfy certain process constraints?

Q13.4.2 When is a select element (S) required instead of a low select?

Q13.4.3 Why is the fuel/air ratio used for cross-limiting firing controls?

Self-Assessment Answers

Q13.4.1 Many times several controllers use the same MV and a LS is a convenient means to choose the proper controller. For example, consider the reboiler duty on a distillation column as the MV. This MV can be set by a bottoms composition controller, a column pressure constraint controller or a column flooding constraint controller. As a result, the outputs of each of these controllers are sent to a LS element to choose smallest MV value. In this manner, each constraint will be satisfied and when the constraints do not affect the operation of the column, the bottom product composition can be controlled to setpoint.

Q13.4.2 A select element is normally used when there are multiple MVs that can be used by the controller to maintain the CV at its setpoint.

Q13.4.3 Most of the signals used in the cross-limiting firing controls are based on the fuel flow rate. As a result, for the air flow rate controller, the air flow rate measurement is converted to an equivalent fuel flow rate by the fuel to air ratio.

13.5 Computed Manipulated Variable Control

In certain cases, it is not possible to directly adjust the desired MV for a particular process. In these cases, by indirectly adjusting the desired MV, specific disturbances are effectively rejected. For example, consider the reboiler on a distillation column that uses waste heat in the form of quench water (water used to cool hot gases) to provide reboiler duty (Figure 13.5.1). The inlet temperature of the quench water can vary over a wide range, which is a significant disturbance for the distillation column. When the inlet temperature increases, extra boilup results for the column and the bottoms product becomes over purified. The composition controller on the bottoms product can eventually compensate for this disturbance, but this affects the variability in the products produced by the distillation column. The desired operation of the reboiler has the composition controller setting the reboiler duty directly. For steam-heated reboilers with constant enthalpy steam, the reboiler duty is directly related to the steam flow rate, but, for the case under consideration, the reboiler duty changes with inlet temperature as well as the flow rate of the quench water. The solution is to use a steady-state energy balance on the quench water to calculate the flow rate of the quench water that provides the desired reboiler duty. The macroscopic energy balance on the quench water is given by

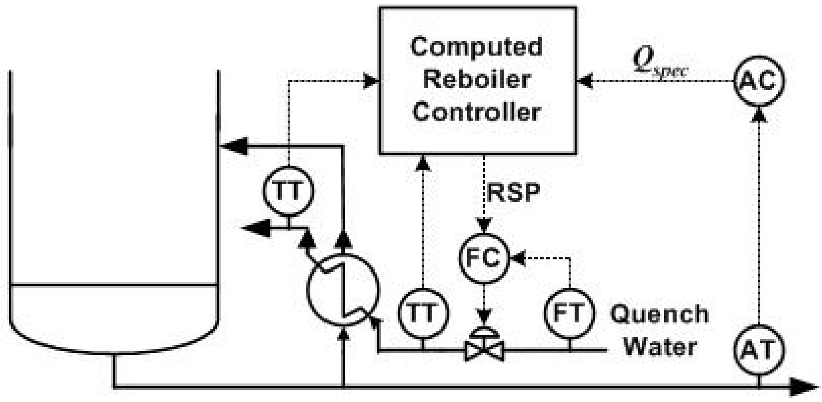


Figure 13.5.1 Control diagram for applying computed reboiler duty control for a distillation column.

$$0 \quad FC_p(T_{in} \quad T_{out}) \quad Q$$

where F is the quench water flow rate, C_p is the heat capacity of the quench water, Q is the rate at which heat is removed from the quench water and T_{in} and T_{out} are the inlet and outlet quench water temperatures. Rearranging to solve for the quench water flow rate yields

$$F_{sp} = \frac{Q_{spec}}{C_p (T_{in} - T_{out})} \qquad \textbf{13.5.1}$$

where F_{sp} is the setpoint for the flow controller on the quench water to the reboiler, Q_{spec} is the reboiler heat duty specified by the bottom product composition controller, C_p is the heat capacity of the quench water, T_{in} and T_{out} are the measured inlet and outlet temperatures of the quench water, respectively. In this manner, as the inlet and outlet temperatures for the quench water change, the quench water flow rate can be adjusted accordingly before upsetting the product compositions of the distillation column. Figure 13.5.1 shows how computed MV control can be applied to this case. The inlet and outlet temperatures, along with the specified reboiler duty, are input to the computation block where Equation 13.5.1 calculates the required quench water flow rate which, in turn, is passed on as the setpoint for the flow controller on the quench water.

Example 13.11 Computed MV Control for a Furnace

Problem Statement. Analyze the operation of the computed MV control for a furnace that is fired using two different types of gas with different heats of combustion.

Solution. Computed MV control can also be used to control a furnace that uses two different grades of fuel. Consider a process that produces a low-heating-value gas as a byproduct. It is desirable to burn all the low-heating-value gas in a furnace that is used to heat a process stream. Unfortunately, the production rate of the low-heating-value gas is not sufficient to provide all the heat duty for the furnace; therefore, natural gas is also fed to the furnace and the flow rate of natural gas is adjusted to control the temperature of the process stream leaving the furnace. Because the flow rate of the low-heating-value gas varies over a wide range, it represents a major disturbance for the temperature controller on the process stream leaving the furnace. A computed MV controller (Figure 13.5.2) can be used to calculate the flow rate of natural gas necessary to meet the heat duty requirements specified by the temperature controller using an energy balance for the heat duty of the furnace along with the heats of combustion for the low-heating-value gas and the natural gas. The macroscopic energy balance for the furnace is given by

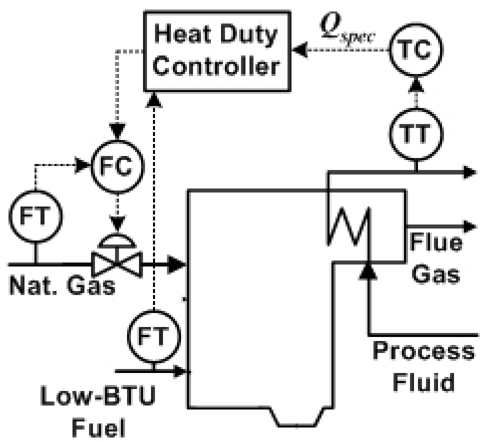


Figure 13.5.2 Control diagram for computed MV control of a furnace with two grades of fuel.

$$0 = F_{LHV}\,\Delta H_{c,LHV} + F_{NG}\,\Delta H_{c,NG} - Q$$

where F_{LHV} is the flow rate of the low-heating value gas, $\Delta H_{c,LHV}$ is the heat of combustion of the low-heating-value gas, F_{NG} is the flow rate of natural gas to the furnace, $\Delta H_{c,NG}$ is the heat of combustion of the natural gas and Q is the heat released in the furnace by the combustion of the low-Btu gas and the natural gas. Solving for the flow rate of natural gas yields.

$$F_{sp,NG} = \frac{Q_{spec} - F_{LHV}\,\Delta H_{c,LHV}}{\Delta H_{c,NG}}$$

where $F_{sp,NG}$ is the computed setpoint for the natural gas firing rate, Q_{spec} is the heat duty specified by the temperature controller, F_{LHV} is the measured flow rate of the low-heating value gas, $H_{c,LHV}$ is the heat of combustion of the low-heating-value gas and $H_{c,NG}$ is the heat of combustion of the natural gas. In this manner, flow rate changes in the low-heating-value gas can be readily compensated for by the computed MV controller. Computed MV control is an effective means of providing heat duty control in certain cases.

Example 13.12 Internal Reflux Control

Problem Statement. Analyze the operation of internal reflux control applied to the column shown in Figure 13.5.3.

Solution. Distillation columns are particularly sensitive to sudden changes in ambient conditions, which usually accompany weather fronts and thundershowers. When the ambient air temperature drops sharply, the temperature of the reflux can also decrease sharply because of the increased cooling provided by the condenser. This subcooled reflux causes added condensation from the vapor in the top of the column, increasing the internal reflux ratio and improving the separation in the top portion of the column. On the other hand, the increase reflux can cause the bottom product to become less pure. In this case, it is much more desirable to control the internal reflux flow rate than the external reflux flow rate because ambient changes can effect the internal reflux of a column and upset the composition control for the column.

An energy balance based on equating the heat lost by the condensing vapor to the heat required to heat the subcooled reflux to the temperature of the top tray results in the following equation

$$C_p F_{ex} (T_{oh} \quad T_r) \qquad F_{\text{int}} \quad H_{vap}$$

where C_p is the heat capacity of the reflux, T_{oh} is the overhead temperature, T_r is the subcooled reflux temperature, F_{ex} is the external reflux (the setpoint for the flow controller on the reflux), F_{int} is the change in the reflux caused by the condensing vapor and H_{vap} is the heat of vaporization of the vapor. F_{int} combines with the external reflux to form the internal reflux. Noting that the internal reflux is the external reflux plus the amount condensed (F_{int}), the equation for the internal reflux flow (F_{int}) is given by

$$F_{int} \quad F_{ex} (1 \quad C_p [T_{oh} \quad T_r] / \ H_{vap})$$

This equation can be rearranged to calculate the external reflux that maintains a specified internal reflux (F_{int}^{spec}), i.e.,

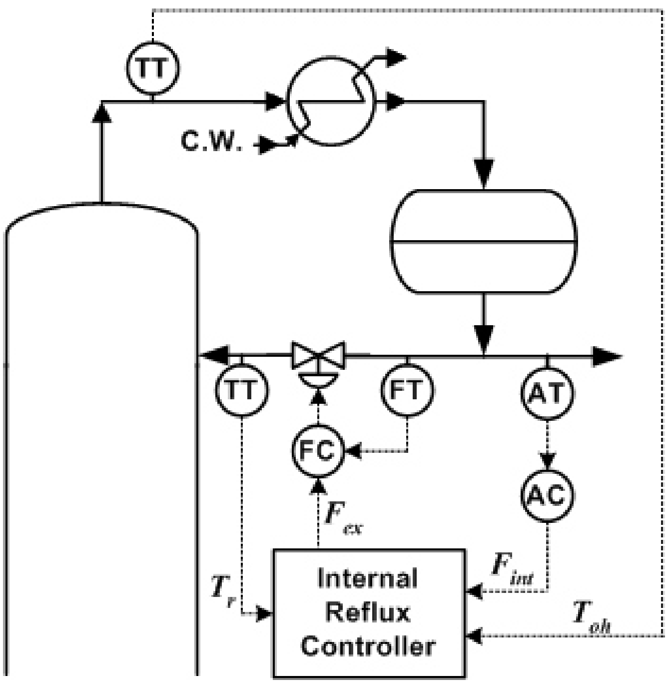


Figure 13.5.3 Control diagram of an internal reflux controller applied for composition control of the overhead of a column.

$$F_{ex} = \frac{F_{int}^{spec}}{1 + C_p(T_{oh} - T_r)/\Delta H_{vap}} \qquad \textbf{13.5.2}$$

This approach is called internal reflux control and is shown schematically in Figure 13.5.3. The composition controller sets the specified value of the internal reflux flow rate and the internal reflux controller calculates the corresponding external reflux flow rate, which is used as the setpoint for the flow controller on the reflux. In this manner, when changes in the reflux temperature occur, the internal reflux controller will make adjustments to the external reflux flow to maintain the specified internal reflux flow rate.

Self-Assessment Questions

Q13.5.1 What kind of disturbance is the computed reboiler duty controller designed to compensate for?

Q13.5.2 How does an internal reflux controller reduce the effect of changes in the cooling water temperature?

Self-Assessment Answers

Q13.5.1 The computed reboiler duty controller is designed to absorb changes in the inlet and outlet quench water temperature.

Q13.5.2 An internal reflux controller uses an energy balance to compensate for changes in internal reflux flow rate caused by changes in the cooling water temperature due to a thunderstorm or a cold front.

13.6 Summary

- Inferential control uses fast-responding process measurements, such as pressures, temperatures, and flow rates, and sometimes also include near-real time measurements (e.g., a mass spectrometer), to estimate the value of a CV.
- When the characteristics of a process (K_p, τ_p and θ_p) change significantly with the value of a measured process variable (e.g., a CV or a DV), scheduling of the controller tuning parameters can result in improved control performance and reliability. For these cases, scheduling of the tuning parameters allows for the variation in the controller tuning as the process conditions change. Scheduling of the tuning parameters can provide stable controller performance without sluggish behavior.
- Override/select control switches between control loops when process constraints are encountered. High and low select controllers are applied to cases where the same MV is used by two different control loops to maintain process constraints. Override/select controls switch between MVs and possible control loops to meet the operational objectives of the process as process conditions change.
- Computed MV control is used when direct manipulation of the desired MV is not possible. For these cases, process measurements are used to calculate the flow rate that is adjusted to maintain the desired MV at its prescribed level.

13.7 Additional Terminology

Adaptive controller - a controller that adjusts its tuning parameters on-line in response to changes in the process.
Artificial neural networks - (ANN) a special class of nonlinear empirical models.

Cross-limiting firing controls - firing controls based on low and high selects that maintain excess air during changes in the firing rate to a furnace.

HS - high select controller.

Inferential control - the use of readily measured quantities, such as pressures, flows and temperatures, to estimate values of the CVs for control purposes.

LS - low select controller.

Non-stationary process - a process whose characteristics (K_p, $_p$, $_p$) change in response to disturbances entering the process.

Self-tuning control - a controller that adjusts its tuning parameters on-line in response to changes in the process.

Smith Predictor - an approach that uses a process model to reduce the effects of deadtime.

Soft sensor - an algorithm that estimates the value of difficult-to-measure process variables using correlation functions based on available process measurements.

Stationary process - a process whose process characteristics (K_p, $_p$, $_p$) remain constant with time.

Time-invariant process - a process whose process characteristics (K_p, $_p$, $_p$) remain constant with time.

Time-varying process - a process whose characteristics (K_p, $_p$, $_p$) change in response to disturbances entering the process.

13.8 References

1. Riggs, J.B. and M.N. Karim, Chemical and Bio-Process Control, 3rd Ed., Ferret Publishing, Austin, Texas, 2006, pp. 104-106.

13.9 Preliminary Questions

13.2 Inferential Control

Q13.2.1 How do you determine which tray temperature should be used to infer the product composition of a distillation column?

Q13.2.2 Explain why inferential temperature control is extensively used industrially.

Q13.2.3 How do changes in the feed composition affect the use of inferential temperature control? How does you compensate for feed composition changes?

Q13.2.4 What assumptions were used in the derivations of Equation 13.2.2?

Q13.2.5 How would you evaluate parameters *a* and *b* in Equation 13.2.3 using plant data?

Q13.2.6 Explain how a neural network can be trained and then used as a soft sensor.

13.3 Scheduling Controller Tuning

Q13.3.1 Identify a process for which scheduling of the controller tuning parameters is likely to be beneficial. Outline how the scheduling of the controller tuning parameters could be accomplished. Choose a system not described in the text.

13.4 Override/Select Control

Q13.4.1 Explain how the control configuration shown in Figure 13.4.1 can prevent the furnace tubes from overheating.

Q13.4.2 Explain how the control configuration shown in Figure 13.4.2 can prevent the distillation column from flooding.

Q13.4.3 Explain how the control configuration shown in Figure 13.4.4 can maintain the bottom product composition at setpoint for a full range of operation.

Q13.4.4 Explain how the control configuration shown in Figure 13.4.5 can prevent CO formation for a decrease in the firing rate.

13.5 Computed Manipulated Variable Control

Q13.5.1 Explain how the control configuration shown in Figure 13.5.1 can reduce the effect of changes in the quench water temperature on the bottoms product composition.

Q13.5.2 Explain how the control configuration shown in Figure 13.5.2 can reduce the effect of changes in the production rate of the low-BTU gas on the temperature of the stream heated by the furnace.

Q13.5.34 Explain how the control configuration shown in Figure 13.5.3 can reduce the effect of changes in the subcooling of the reflux on the overhead product composition.

13.10 Analytical Questions and Exercises

13.2 Inferential Control

13.2.1** The operating pressure of a distillation column has a significant effect on the temperatures of the trays of the column. Indicate how a tray temperature used to infer the product composition can be compensated for pressure changes. Assume that the tray temperature varies linearly with column pressure. Indicate how you would determine all unknown parameters.

13.2.2** Construct an inferential estimate of the fouling of the heat exchanger shown in Figure P13.2.2. Indicate how this estimator could be used to schedule cleaning of the tube bundle.

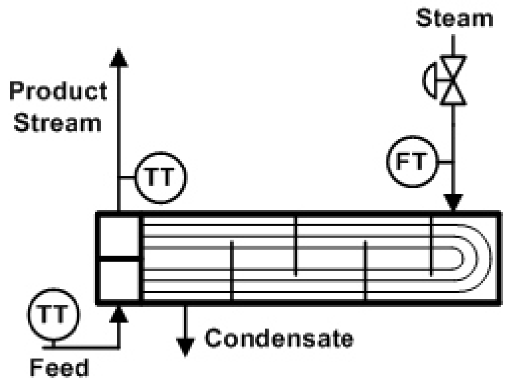


Figure P13.2.2 Schematic of a heat exchanger along with instrumentation.

13.4 Override/Select Control

13.4.1*** Consider the accumulator for a distillation column for which the distillate product flow rate is used to control the accumulator level and the reflux flow rate is used to control the composition of the overhead product (Figure P13.4.1). Draw a schematic showing select controls that will prevent the level from exceeding 95% or becoming less than 5% by overriding the composition controller on the overhead when the level is too high or too low.

13.4.2** Consider the stripping section of the distillation column shown in Figure P13.4.2. Modify this schematic by adding override controls that prevent the column pressure from exceeding its upper limit by overriding the composition controller when the pressure reaches its upper limit.

13.4.3*** Consider the stripping section of the distillation column shown in Figure P13.4.3. Under certain conditions, the column floods if the steam addition rate is not restricted and, under other conditions, excess steam flow to the reboiler causes the maximum temperature limit on the reboiler to be exceeded, resulting in severe fouling of the

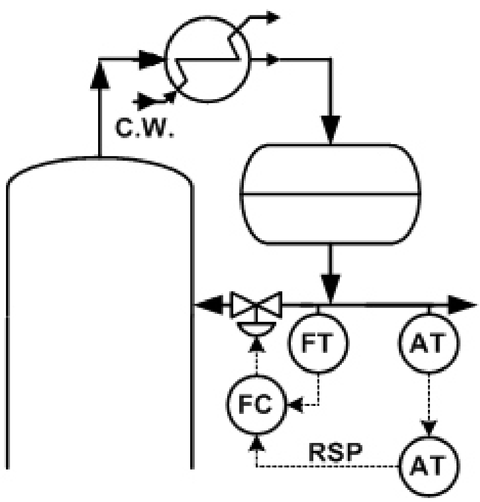


Figure P13.4.1 Schematic of the rectifying section of a column.

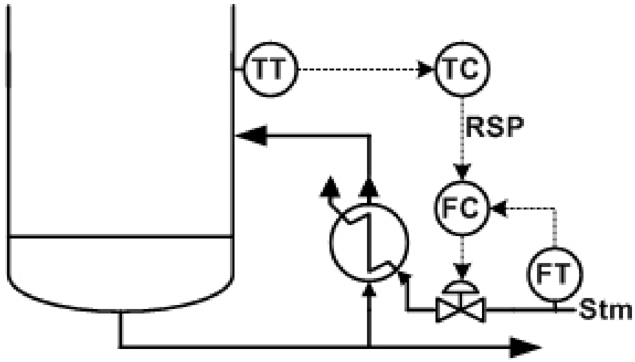


Figure P13.4.2 Schematic of the stripping section of a column.

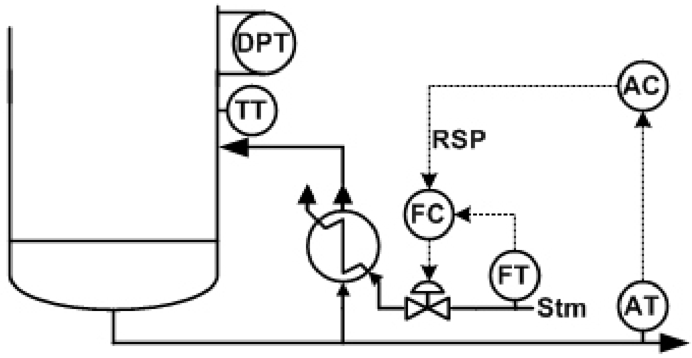


Figure P13.4.3 Control diagram for bottoms composition control of a distillation column.

reboiler. Draw a schematic showing the override/select controls that simultaneously prevents the column from flooding and from exceeding the upper limit on the reboiler temperature.

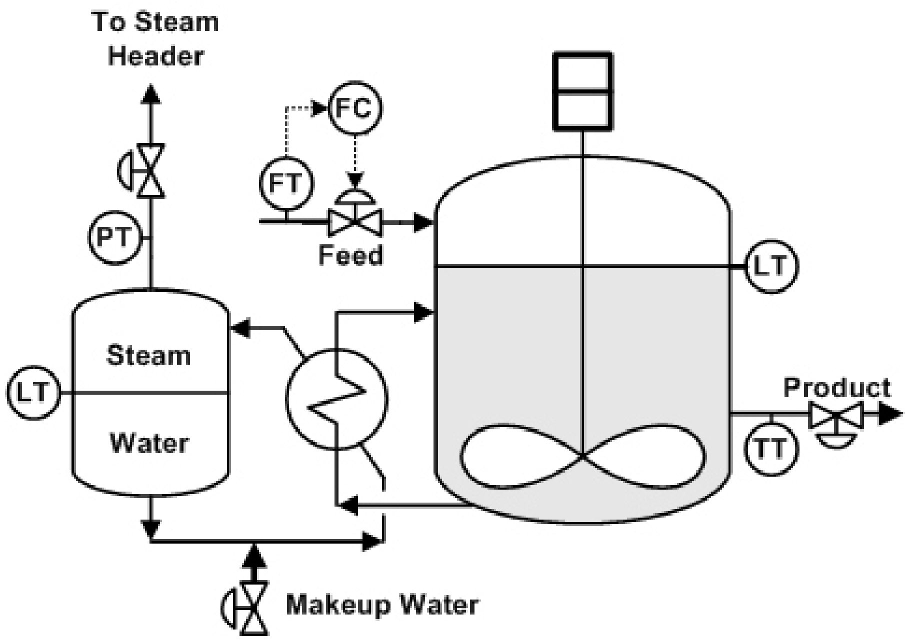
13.4.4*** Consider the schematic for an exothermic CSTR shown in the Figure P13.4.4 in which the heat produced by the reactor is used to generate steam [after W.L. Luyben, *Process Modeling, Simulation and Control for Chemical Engineers*, Second Edition, McGraw-Hill, p. 292 (1990)]. Draw a schematic for this process including each of the following control features:

a. The level in the steam drum is controlled by the make-up water.

b. The pressure of the steam drum is controlled by the valve on the steam line to the steam header.

c. The temperature controller for the reactor is cascaded to the steam pressure control loop.

d. The level in the reactor is controlled by the product flow rate.



Schematic for Problem 13.4.4

e. A low level in the steam drum overrides the setpoint for the flow controller on the feed to the reactor and cuts back on the feed to the reactor.

f. A high reactor temperature overrides the setpoint for the flow controller on the feed to the reactor and cuts back on the feed to the reactor.

13.5 Computed Manipulated Variable Control

13.5.1** Develop a computed MV controller for the reboiler shown in Figure 13.5.1, assuming that steam is used for reboiler duty instead of quench water. The computed MV controller should be able to make adjustments in the steam flow rate to account for changes in the steam enthalpy; therefore, the steam temperature and pressure upstream of the control valve should be measured and used by the computed MV. Assume that there is no condensed water in the steam.

13.5.2*** From your fluids course, you know that the mass flow rate of a gas through an orifice meter is dependent on the pressure drop across the orifice plate and the temperature and pressure of the gas. Therefore, if the temperature and pressure of a gas change significantly, using the pressure drop across an orifice meter as a measurement of flow rate can result in significant error. Devise a computed flow sensor for the mass flow rate of a gas for which the temperature and pressure of the gas change significantly. List all the necessary equations and draw a schematic showing the computed mass flow rate controller.

Chapter 14

PID Implementation Issues

Chapter Objectives

- Define windup and present several approaches to correct for it.
- Present bumpless transfer.
- Introduce split range flow and temperature control.

14.1 Introduction

This chapter is concerned with several techniques that have been developed to solve PID implementation problems. Anti-reset windup procedures protect against integral windup when a MV reaches an upper or lower limit or when two or more control loops are used with the same MV (e.g., override/select control). Bumpless transfer is a strategy for bringing a controller on-line in a manner that does not unduly upset the process. For a wide range of operation, split-range control uses two separate actuators to provide better performance than a single actuator.

14.2 Anti-Windup Strategies

Figure 14.2.1a shows the MV and CV for a standard PI controller for which the MV reaches a limit, i.e., the control valve is fully open or fully closed (i.e., saturated). Valve saturation can occur when a large disturbance enters the process or when the process is operated over a wide region (e.g., large setpoint changes). Because the MV cannot be increased further, the PI controller is unable to return the CV to its setpoint. As long as there is an error between the CV and its setpoint, the integral term in the PI controller (Equation 7.3.1) continues to accumulate, which is referred to as **reset windup** or **integral windup**. After some time, the disturbance level returns to its original value or the operation returns to its original operating region. At this point, integral windup in the PI controller keeps the MV at its maximum level even though the value of the CV is now above its setpoint. In effect, before the process can return to steady-state at the setpoint for the CV, an area "B" above the setpoint must be generated to compensate for area "A" shown in Figure 14.2.1a (i.e., area B must equal area A).

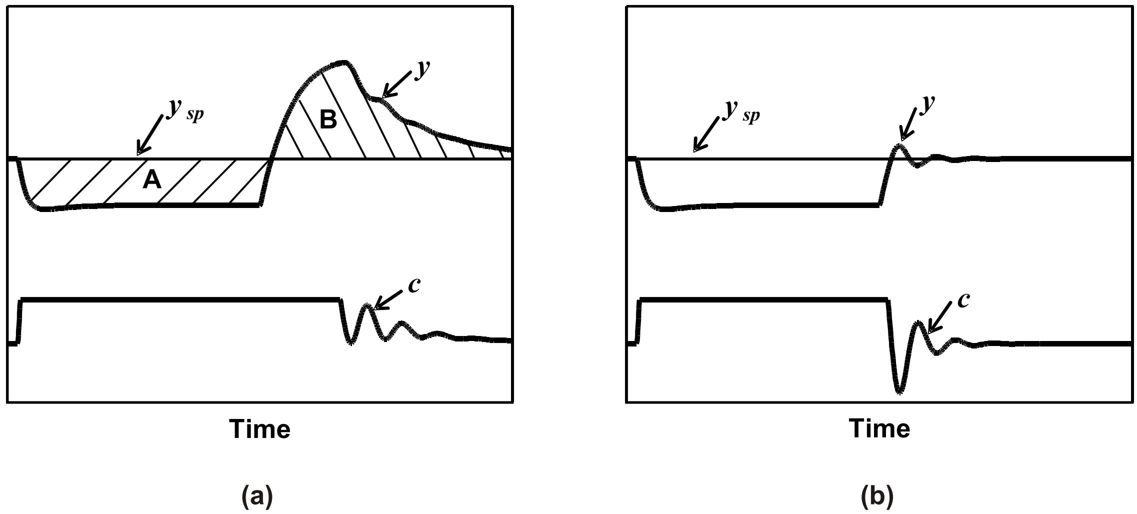


Figure 14.2.1 Response of a feedback system to a saturated MV. (a) Conventional PI controller. (b) PI controller with anti-reset windup.

This behavior results because the integral action is allowed to continue accumulating after control of the process has been lost (**MV saturation**). Figure 14.2.1b shows the same case as Figure 14.2.1a except that, when the MV saturates, the integral action is not allowed to accumulate (**windup**). Note that when control returns to the process (i.e., when the MV is no longer saturated), the CV moves directly back to its setpoint and does not exhibit prolonged deviations from setpoint as before. Because the integral action was turned off when the MV became saturated, the PI controller does not have to generate an area equivalent to area "A" above the setpoint.

Anti-reset windup (Figure 14.2.1b) can be implemented by simply not allowing the integral to accumulate when the MV is saturated. The MV is saturated when the control valve on the line supplying the MV is either closed or fully open. A saturated control valve can be identified when there is sustained offset between the MV level requested by the flow controller and the actual flow rate of the MV.

Clamping the Controller Output. Because DCSs and control computers use the velocity form of the PID controller, the output from the controller can be restricted or "clamped" so that it does not become less than 0% or more than 100%. In most cases, clamping the controller output prevents severe reaction to reset windup, but clamping the controller output still allows some degree of windup to occur.

Internal Reset Feedback. Figure 14.2.2a shows a block diagram for a conventional PI controller. Figure 14.2.2b shows a block diagram for **internal reset feedback**. Applying a balance around the summation block in Figure 14.2.2b for the internal reset feedback case yields

$$K_c E(s) \quad F(s) \quad C(s)$$

where

$$F(s) \quad \frac{C(s)}{{}_I s \quad 1}$$

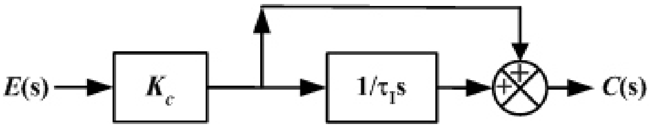


Figure 14.2.2a Block diagram of a conventional PI controller.

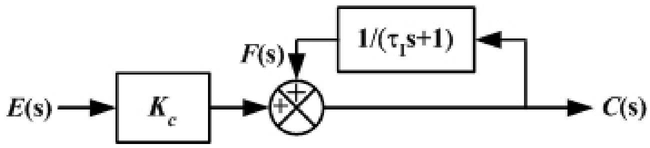


Figure 14.2.2b Block diagram of a PI controller with internal reset feedback.

Substituting and collecting terms results in

$$C(s)\left[1-\frac{1}{\tau_I s+1}\right]=K_c E(s)$$

Solving for $C(s)$ yields

$$C(s)=K_c\left[1+\frac{1}{\tau_I s}\right]E(s)$$

which is the transfer function for a PI controller; therefore, internal reset feedback (Figure 14.2.2b) is equivalent to PI control (Figure 14.2.2a). Because it is equivalent to a conventional PI controller, internal reset feedback will also windup, but because $C(s)$ can be clamped, it does not windup above 100% or below 0%. Therefore, internal reset feedback does not offer any advantage over using the velocity form of the PI controller and clamping the controller output. It is, however, a natural step from internal reset feedback to external reset feedback, which is superior to controller output clamping.

External Reset Feedback. Figure 14.2.3 shows a block diagram for **external reset feedback**. For this case the measured value of the MV, instead of the output from the controller is fed back through the filter to the summation block. The advantage of external reset feedback is that shortly after the MV saturates, i.e., $U_{meas}(s)$ becomes constant, reset windup is turned off and $C(s)$ becomes constant. For internal reset feedback, reset windup continues until the controller output reaches 0% or 100%. External reset feedback turns off the integral action much sooner than internal reset feedback. The disadvantage of external reset feedback is that a measurement of the MV, which is not available in all cases, is required. It should be pointed out that the measured value of the MV, U_{meas}(s), must also be scaled so that it has the same units as the controller output (i.e., %).

It is standard control practice to apply some type of anti-windup strategy (e.g., external reset feedback or turning off the integral action when the MV saturates) to all control loops that use integral action to prevent reset windup. This is essential for override/select loops because, when one loop is controlling the process, the other is not in service; therefore, the inactive loop can experience severe windup if anti-reset windup measures are not taken. Figure 14.2.4 demonstrates how external reset feedback is applied to a tray temperature controller on the stripping section of a distillation column. As shown in Figures 14.2.3 and 14.2.4, external reset feedback requires that the measurement of the steam flow, which

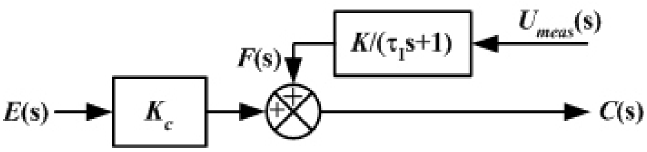


Figure 14.2.3 Block diagram of a PI controller with external reset feedback.

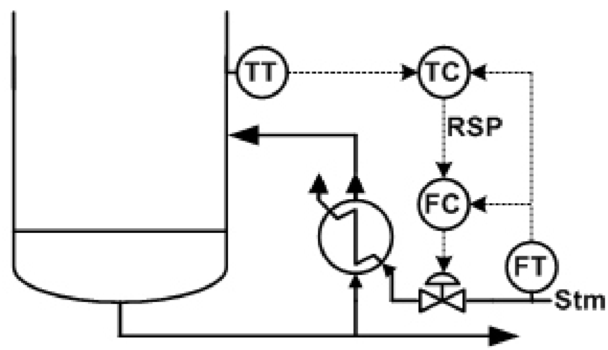


Figure 14.2.4 A control diagram for the implementation of external reset feedback for a tray temperature controller on the stripping section of a distillation column.

corresponds to $U_{meas}(s)$, be input to the temperature controller. From Figure 14.2.3, both the error in the measurement of the CV and the measured value of the MV are used to calculate control action and from Figure 14.2.4, the temperature measurement and the measured steam flow are inputs to the temperature controller.

Self-Assessment Questions

Q14.2.1 If a conventional PI (Figure 14.2.2a) and internal reset feedback (Figure 14.2.2b) are equivalent, why do they respond differently to windup?

Q14.2.2 Why is external reset feedback superior to internal reset feedback?

Q14.2.3 What extra requirement does external reset feedback have compared to internal reset feedback?

Self-Assessment Answers

Q14.2.1 Even though Figure 14.2.2a and 14.2.2b are equivalent from an overall point of view, a conventional PID controller (Figure 14.2.2a) is susceptible to windup while internal reset feedback (Figure 14.2.2b) is not because integral action operates off the controller output [$C(s)$] as part of a lag element. As a result, when the error from setpoint stops changing, the contribution from the integral term will also stop.

Q14.2.2 Internal reset feedback can be used to prevent the windup of the controller output below 0% or above 100% while external reset windup prevents windup shortly after the MV becomes constant. As a result, external reset feedback allows less windup than internal reset feedback.

Q14.2.3 External reset feedback requires the measurement of the MV while internal reset feedback does not.

14.3 Bumpless Transfer

Figure 14.3.1 shows the process behavior with and without **bumpless transfer**. Without bumpless transfer, if the controller is turned on (i.e., the controller is changed from manual to automatic or computer/remote control) when the CV is far removed from setpoint, the controller takes immediate action and drives the process to setpoint in an

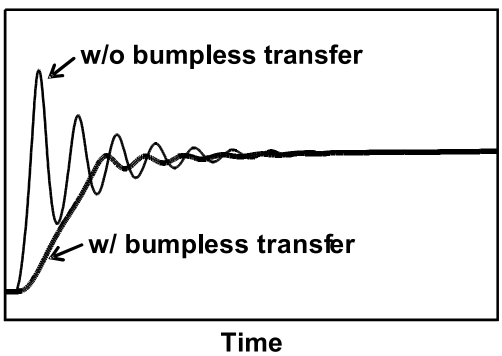


Figure 14.3.1 The startup response of a feedback system without bumpless transfer and with bumpless transfer.

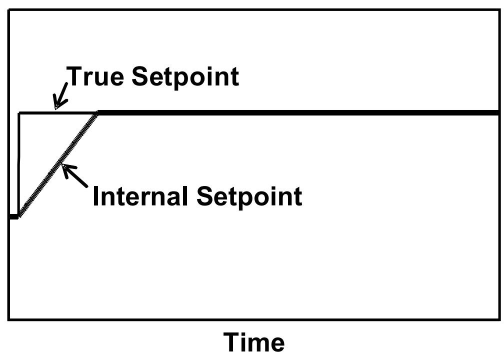


Figure 14.3.2 Comparison between the true setpoint and the internal setpoint for the bumpless transfer example.

underdamped fashion. In certain cases, the CV can be far enough away from setpoint and the process can be sufficiently nonlinear that the control loop becomes unstable. Even if the control loop does not become unstable, the abrupt action of the feedback controller can significantly upset other control loops on the process. As a result, operators find that the behavior of a controller without bumpless transfer is generally unacceptable, particularly for key loops such as composition and temperature control loops.

For bumpless transfer, there are two types of setpoints: the true setpoint, which corresponds to the desired operating point and the internal setpoint that is used for bumpless transfer (Figure 14.3.2). When a control loop is turned on, the setpoint used by the controller is actually different than the true setpoint when applying bumpless transfer. When the controller is turned on, the internal setpoint is set equal to the current CV value; therefore, there is no change in the MV level. After this, the internal setpoint is ramped toward the true setpoint and the process begins moving toward the true setpoint in a gradual fashion. After the internal setpoint reaches the true setpoint value, it remains constant. Note that the internal setpoint ramp (for those controllers that have such a separate internal setpoint) is normally set by the controller algorithm and so is not user configurable. By virtue of bumpless transfer, smooth and consistent startups for control loops result. Prevention of overly aggressive startup of a control loop or response to setpoint changes can also be attained by using the form of the PID algorithm that is not susceptible to proportional kick (Equation 7.5.5).

Self-Assessment Question

Q14.3.1 Why do operators prefer bumpless transfer?

Self-Assessment Answer

Q14.3.1 Operators prefer bumpless transfer because by using it they are able to startup control loops without upsetting the control loop in question and the rest of the process that is affected by changes in the manipulated or CVs of the control loop.

14.4 Split-Range Control

The basic idea behind split-range control is to use two or more actuators or MVs so that the flexibility of a number of actuators or MVs can be used by the controller. Split-range flow control and split-range temperature control are considered here.

Split-Range Flow Control. In a typical split-range control loop, the controller output is split and sent to two or more control valves. The splitter in the controller defines how each valve is sequenced as the controller output changes from 0 to 100%. In most split-range applications, the controller adjusts the opening of one of the valves when its output is in the range of 0 to 50% (i.e., 4-12 ma for an analog controller) and the other valve when its output is in the range of 50% to 100% (i.e., 12-20 ma for an analog controller). Consider a wastewater neutralization process in which an acidic wastewater is neutralized by adding NaOH. The titration curve for the wastewater is shown in Figure 14.4.1. To control the pH to pH unit at a setpoint of pH

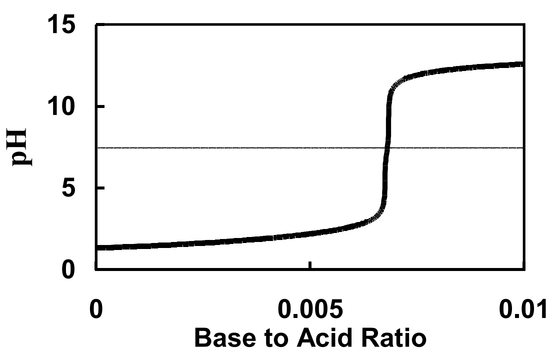


Figure 14.4.1 Titration curve for a strong acid/strong base system.

7, the base flow rate must be metered accurately to within 0.5%. A single flow-control loop with a control valve with a positioner can meet this metering precision. If the total flow rate of base were to range from 0.1 to 10 gallons per minute, one flow control loop could not simultaneously meter the base flow rate to within ±0.5% at both extremes.

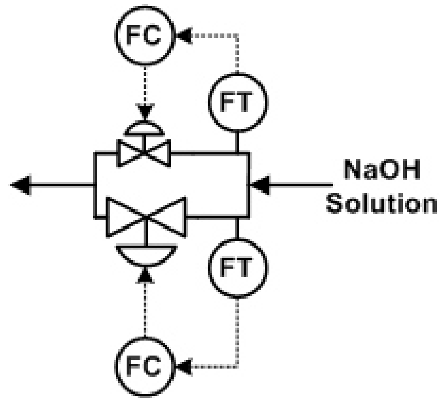


Figure 14.4.2 Schematic of a split range flow controller.

Two flow control loops that work together, as shown in Figure 14.4.2, can meet this requirement. Figure 14.4.2 shows a schematic of the hardware and the flow control loops necessary to apply split range flow control. At low flow rates, the large control valve is closed and the flow control loop with the smaller control valve can accurately meter the low-flow operation (Figure 14.4.3). As the total flow increases, the smaller control valve begins to approach saturation. Before this happens, the flow control loop with the larger control valve comes into service. At large flow rates, the small control valve is completely open and the flow control loop with the larger valve is accurately metering the base flow rate. Note that a single PID controller is used for pH control, but the output of the controller is split and sent to two different MVs. This is an example of **split-range flow control**, which is used when accurate flow control is required over a wider operating range than one control valve can provide.

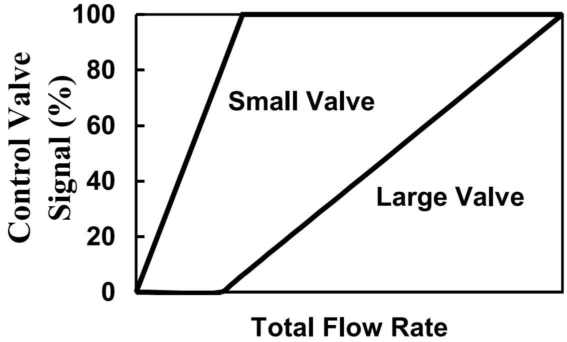


Figure 14.4.3 Controller output to the large and small valves in the split-range flow control arrangement shown in Figure 14.4.2.

Split-Range Reactor Temperature Control. Figure 14.4.4 shows a schematic of an exothermic CSTR for which steam is used to bring the reaction mixture up to the normal operating temperature and cooling water is used to maintain the reactor at the operating temperature. Initially, the reactor is near ambient temperature and the reactions are essentially extinguished. Steam is added to the reactor jacket to raise the temperature of the reaction mixture. As the reactor temperature increases, the exothermic reactions add heat to the system, further increasing the reactor temperature. As the reactor temperature begins to increase, the amount of steam added to the reactor is decreased and cooling water is added to the reactor jacket. Eventually, the steam is cut off and the cooling water removes the heat generated by the exothermic reactions. Figure 14.4.5 shows the control signals that are sent to the valves on the steam and the cooling water by the split-range temperature controller. Note that as the error between the remote setpoint for the product

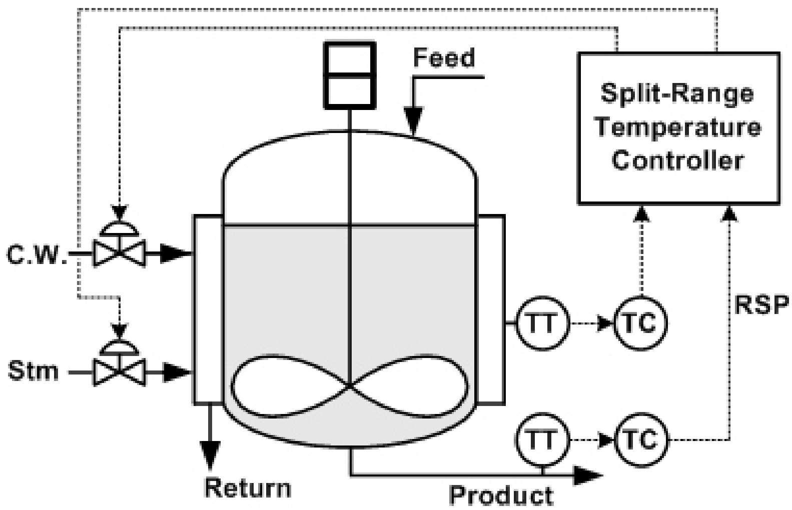


Figure 14.4.4 Schematic of a split-range temperature controller applied to a jacketed batch reactor.

temperature and the measured product temperature decreases, the control action switches from heating to cooling.

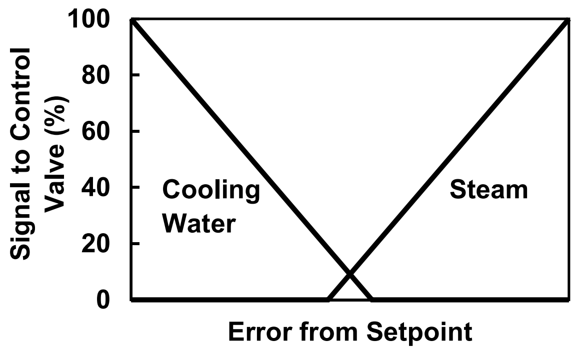


Figure 14.4.5 The signals to the control valves from the split-range temperature controller as a function of the error from setpoint.

Self-Assessment Questions

14.4.1 When is split-range flow control required?

14.4.2 When is split-range temperature control required?

Self-Assessment Answers

Q14.4.1 Split range flow control should be used when accurate flow control is required over a wider range than can be provided by a single flow control loop.

Q14.4.2 Split range temperature control can be used when heating and cooling are required by the same controller.

14.5 Summary

- Reset windup can occur when a control valve saturates. External reset feedback or turning off the integral action when a control valve saturates eliminates reset windup.
- Bumpless transfer provides a smooth startup procedure for a control loop by ramping the setpoint from the initial value of the CV to its desired final value.
- Split-range control uses two or more actuators or MVs in parallel so that the control system operates efficiently over a wide operating range.

14.6 Additional Terminology

Anti-reset windup - approaches that prevent reset windup, e.g., external reset feedback.

Bumpless transfer - a startup procedure used to gradually bring a control loop into service.

Clamping - restricting the output of a controller to less than a maximum amount (e.g., 100%) and greater than a minimum amount (e.g., 0%).

External reset feedback - an anti-windup approach that uses the measured value of the MV.

Integral windup - the accumulation of the integral of the error from setpoint caused by an uncontrollable error from setpoint.

Internal reset feedback - an anti-windup approach that does not use the measured value of the MV.

MV saturation - a MV that is either at its maximum or minimum value.

Reset windup - the accumulation of the integral of the error from setpoint caused by an uncontrollable error from setpoint.

Split-range flow control - using two flow control loops in parallel, one with a smaller valve than the other, to provide precise flow metering over a wide range of flow rates.

Windup - the accumulation of the integral of the error from setpoint caused by an uncontrollable error from setpoint.

14.7 Preliminary Questions

14.2 Anti-Windup Strategies

14.2.1 In your own words, explain how windup occurs and what problems it causes.

14.2.2 Show that internal reset feedback is equivalent to conventional control for a PI controller.

14.2.3 Explain, in your own words, how external reset feedback eliminates integral windup.

14.2.4 What advantages does internal reset feedback have over using the velocity form of the PID controller and clamping the controller output?

14.2.5 What is the advantage of external over internal reset feedback?

14.2.6 If you were applying an anti-windup strategy that turned off the integral when a control valve saturated, how would you determine whether or not a control valve was saturated?

14.2.7 What is the primary requirement for applying external reset feedback?

14.3 Bumpless Transfer

14.3.1 How can large setpoint changes or starting a control loop far from its setpoint result in unstable operation of the control loop?

14.3.2 Explain how bumpless transfer is applied and indicate what its advantages are.

14.3.3 Why do operators prefer bumpless transfer?

14.4 Split-Range Control

14.4.1 When should you use split-range flow control?

14.4.2 Explain how the controller outputs shown in Figure 14.4.2 are responsible for implementing split-range flow control.

14.4.3 Why is split-range temperature control necessary for the reactor shown in Figure 14.4.4?

14.4.4 Explain how the controller outputs shown in Figure 14.4.5 are responsible for implementing split-range temperature control.

Part V

Control of MIMO Processes

Chapter 15

PID Controllers Applied to MIMO Processes

Chapter Objectives

- Illustrate the problems with coupling and configuration selection for MIMO control systems.
- Present the three factors that affect configuration selection: steady-state coupling, dynamic factors and sensitivity to disturbances.
- Introduce the concepts of controllability and observability.

15.1 Introduction

A multiple-input/multiple-output (MIMO) process has two or more inputs and two or more outputs. A two-input/two-output system is shown schematically in Figure 15.1.1. Figure 15.1.1 can be represented mathematically as

$$y_1(s) \quad G_{11}(s)c_1(s) \quad G_{12}(s)c_2(s)$$
$$y_2(s) \quad G_{22}(s)c_2(s) \quad G_{21}(s)c_1(s)$$

Note that c_1 affects both y_1 and y_2 and c_2 affects both y_1 and y_2. When both inputs affect both outputs the process is **coupled**. Coupled MIMO processes are frequently encountered in the CPI and bio-tech industries.

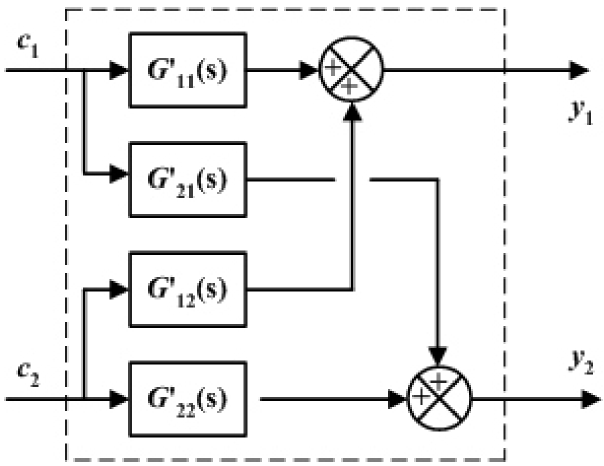


Figure 15.1.1 Block diagram of a two-input/two-output process.

This chapter considers the application of PID controllers to coupled MIMO processes. A key issue when applying PID controllers to MIMO systems is deciding which MV should be used to control which CV. This is referred to as choosing the **MV/CV pairings** [(c,y) pairings] or selecting the **control configuration**. The factors that affect the choice of (c, y) pairings are analyzed in this chapter. A strategy for tuning PID controllers that are applied to MIMO processes is presented as well as an introduction to decoupling. In this chapter for simplicity, the transfer function $G\ (s)$ is used to represent the combined effect of the actuator and process $[G_a(s)\ G_p(s)\ ]$. For example, $G_{11}(s)$ in Figure 15.1.1 represents the combined effect of c_1 on y_1 due to the actuator and process while the sensor dynamics are neglected. In addition, $G_{12}(s)$ in Figure 15.1.1 represents the combined effect of c_1 on y_2 due to the actuator and process.

15.2 SISO Controllers and (*c*, *y*) Pairings

Figure 15.2.1 shows two **single-loop PID controllers** applied to a two input/two output process (2 2 system). Applying single-loop PID controllers to a MIMO process is called **decentralized control**. Centralized control is addressed in the next chapter. The coupling in this 2 2 system causes the two control loops to interact. That is, while control loop 1 manipulates c_1 to keep y_1 at its setpoint, it affects control loop 2. Likewise, the operation of control loop 2 acts as an upset for control loop 1. Figure 15.2.2 shows schematically the coupling effect of control loop 2 (indicted by heavy lines) as an additive disturbance to control loop 1. As shown in Figure 15.2.2, when loop 1 changes c_1 in an effort to keep y_1 at its setpoint, changes in c_1 also affect y_2. Then loop 2 responds to maintain y_2 at setpoint, using changes in c_2 which, in turn, affect y_1. Therefore, when a decentralized controller is applied to a MIMO process, each controller output affects all the CVs and the controllers on these CVs respond to these changes and so on. For this 2 2 example, tuning control loop 1 must include the effects of control loop 2 and vice versa.

The selection of pairings for decentralized control can have a dramatic effect on the resulting overall control performance. Consider the 2 2 system represented by transfer function models shown in Table 15.1. Using c_1 to control y_1 and c_2 to control y_2 has the advantage that the magnitude of coupling is relatively small. The steady-state process gain for the effect of c_1 on y_1 is 1.0 and for the effect of c_2 on y_2 is 2.0 while the gain for the

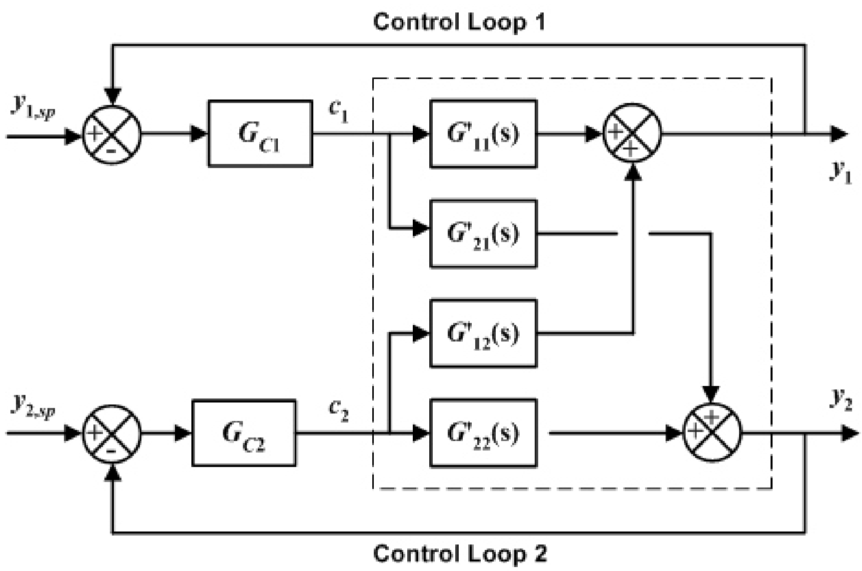


Figure 15.2.1 Block diagram of a 2 × 2 process with single-loop controllers (decentralized control).

effect of c_1 on y_2 is 0.05 and for the effect of c_2 on y_1 is 0.1. As a result, relatively small changes in y_1 and y_2 result

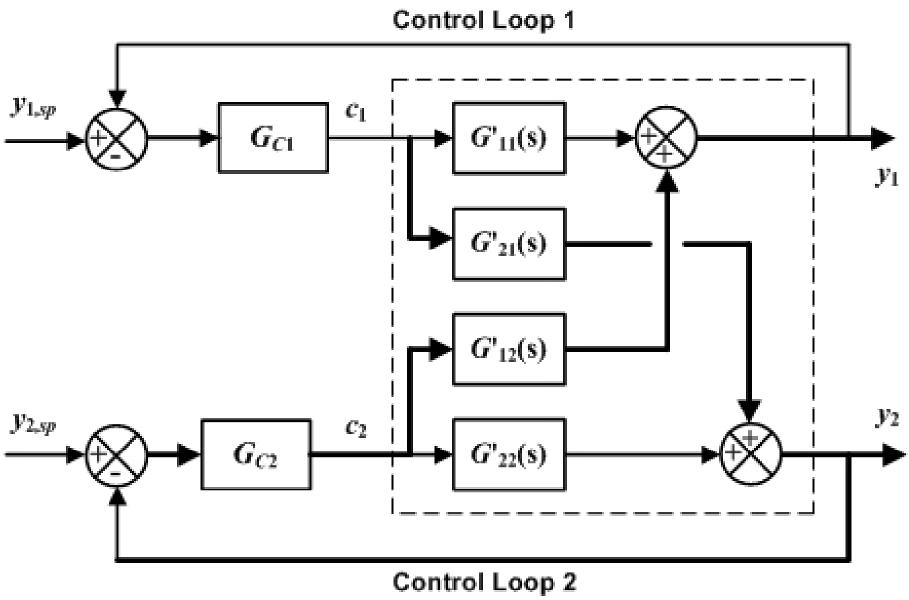


Figure 15.2.2 A block diagram of a 2 × 2 process with single-loop controllers showing the coupling effect of loop 2 on y_1 for changes in

due to coupling because the off-diagonal process gains (i.e., K_{21} and K_{12}) are relatively small. Moreover, the controller that uses c_1 to control y_2 and c_2 to control y_1 would have to use relatively large changes in c_1 and c_2 for the feedback controllers due to the small diagonal process gains, which result in even larger coupling effects. Thus, the resulting coupling would be severe.

Three factors determine the best pairings for a MIMO process: coupling, dynamic response and the sensitivity to disturbances. Each of these factors is considered separately in the next three sections.

Table 15.1
Transfer Function Models for a Two Input/Two Output Process.

	c_1	c_2
y_1	$\dfrac{1.0}{10s \quad 1}$	$\dfrac{0.1}{10s \quad 1}$
y_2	$\dfrac{0.05}{10s \quad 1}$	$\dfrac{2.0}{10s \quad 1}$

Self-Assessment Questions

Q15.2.1 For a MIMO process, explain why different control configurations provide different control performance.

Q15.2.2 For a MIMO process, how do the control loops affect each other?

Self-Assessment Answers

Q15.2.1 Each different control configuration for a MIMO process will have a different control performance because each control configuration will have different degrees of coupling, dynamic response and sensitivity to disturbances.

Q15.2.2 As each control loop applied to a MIMO process makes changes to its MV, all the other control loops will be affected as a result of coupling in the MIMO process.

15.3 Steady-State Coupling

Bristol[1] developed the **Relative Gain Array (*RGA*)**, which is a measure of steady-state coupling. The *RGA* for a 2 2 system is given by

$$RGA \qquad \begin{matrix} {}_{11} & {}_{12} \\ {}_{21} & {}_{22} \end{matrix}$$

where

$$_{11} \quad \frac{\left.\frac{y_1}{c_1}\right|_{c_2}}{\left.\frac{y_1}{c_1}\right|_{y_2}} \qquad\qquad _{12} \quad \frac{\left.\frac{y_1}{c_2}\right|_{c_1}}{\left.\frac{y_1}{c_2}\right|_{y_2}}$$

$$_{21} \quad \frac{\left.\frac{y_2}{c_1}\right|_{c_2}}{\left.\frac{y_2}{c_1}\right|_{y_1}} \qquad\qquad _{22} \quad \frac{\left.\frac{y_2}{c_2}\right|_{c_1}}{\left.\frac{y_2}{c_2}\right|_{y_1}} \qquad \textbf{15.3.1}$$

where

$$\left.\frac{y_i}{c_j}\right|_{c_k}$$

represents the steady-state change in y_i resulting from a change in c_j while keeping c_k constant and

$$\left.\frac{y_i}{c_j}\right|_{y_k}$$

represents the steady-state change in y_i for a change in c_j while keeping y_k constant. As a result, **the *RGA* is the ratio of the process gain without coupling to the process gain with coupling**. The numerator of $_{11}$ in Equation 15.3.1 is simply the open-loop gain for the effect of c_1 on y_1, i.e., the steady-state gain in $G_{11}(s)$ and this

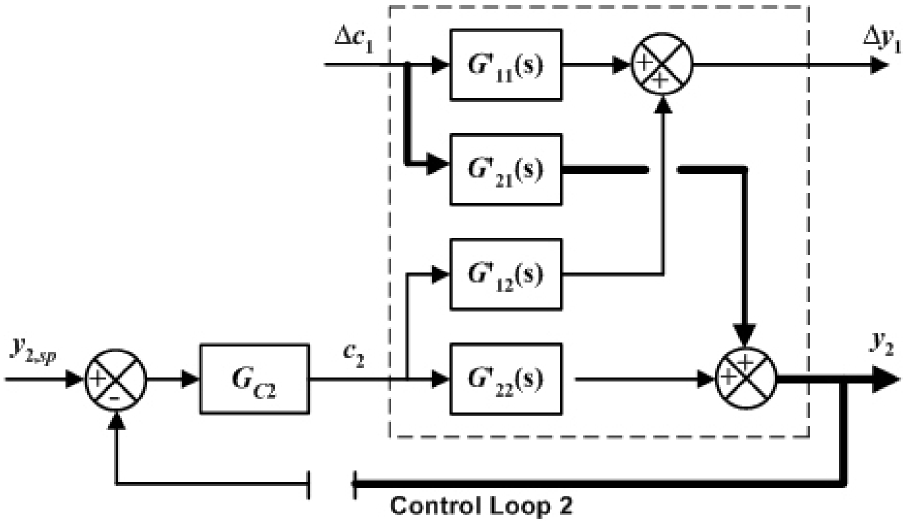


Figure 15.3.1 Block diagram for the determination of the numerator of $_{11}$.

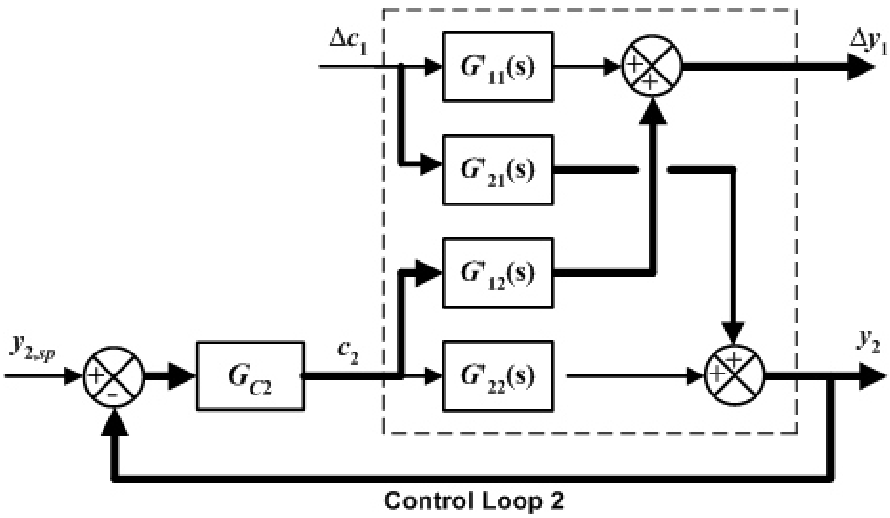


Figure 15.3.2 Block diagram for the determination of the denominator of $_{11}$.

can be determined by implementing a change in c_1 and measuring the resulting steady-state change in y_1 while control loop 2 is open (Figure 15.3.1). The denominator of $_{11}$ is the gain between c_1 and y_1 while keeping y_2 at its setpoint, which requires that the second loop is closed (Figure 15.3.2). If no coupling is present, the numerator of $_{11}$ equals the denominator. As a result, the closer $_{11}$ is to 1.0, the less steady-state coupling a configuration has.

Consider the system represented by Table 15.1. The steady-state gain matrix for this process is given by

$$\underline{\underline{K}} = \begin{matrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{matrix} \qquad \begin{matrix} 1.0 & 0.1 \\ 0.05 & 2.0 \end{matrix}$$

Consider $_{11}$. The numerator is the open-loop gain of the first transfer function, 1.0. The evaluation of the numerator of $_{11}$ is shown schematically in Figure 15.3.1. To calculate the value of the denominator of $_{11}$, i.e.,

$$\frac{y_1}{c_1}\ {}_{y_2}$$

consider a change in c_1 of 1.0. From Figure 15.3.2, the net effect on y_1 is the combined effect of the primary response and the result of coupling. The primary response is the product of c_1 and K_{11}, 1.0. The result of coupling is calculated by using the various steady-state gains. The effect of c_1 on y_2 is given by

$$y_2 \qquad c_1\ K_{21} \qquad 0.05$$

The (c_2, y_2) control loop must compensate for this change in y_2; therefore, the required change in c_2 is given by

$$c_2 \qquad \frac{y_2}{K_{22}} \qquad 0.025$$

Finally, the effect of c_2 on y_1 is given by

$$y_1 \qquad c_2\ K_{12} \qquad 0.0025$$

which represents the coupling result for a c_1 change.

Therefore, the total effect is the sum of the primary effect and the coupling effect, i.e.,

$$\frac{y_1}{c_1}\ {}_{y_2} \qquad \frac{1\quad 0.0025}{1.0} \qquad 0.9975$$

Then, $_{11}$ is given by

$$_{11} \qquad \frac{1.0}{0\ .9975} \qquad 1.0025$$

which indicates that this pairing is highly decoupled.

For a 2 2 system, the value of $_{11}$ is expressed as a function of the steady-state gains, i.e.,

$$_{11} \qquad \frac{K_{11}}{K_{11} \quad \dfrac{K_{12}\, K_{21}}{K_{22}}} \qquad \frac{1}{1 \quad \dfrac{K_{12}\, K_{21}}{K_{11}\, K_{22}}} \tag{15.3.2}$$

In this example, the effect of coupling works in the opposite direction of the primary action, where the primary action in this case is given by K_{11}. As a result, $_{11}$ is greater than unity. When the primary action and the coupling effect act in the same direction, $_{11}$ is less than one. This can be understood by recognizing that when both effects act in the same direction, the denominator of Equation 15.3.2 is greater than the numerator, which occurs when

$$\frac{K_{12}K_{21}}{K_{11}K_{22}} \quad 0$$

It can be shown that

$$_{11} \qquad _{12} \qquad 1$$

Likewise,

$$_{21} \qquad _{22} \qquad 1$$

It can also be shown that

$$_{11} \qquad _{21} \qquad 1$$

and

$$_{12} \qquad _{22} \qquad 1$$

In general, the sum of any row or any column of the RGA matrix is equal to unity. As a result,

$$_{11} \qquad _{22}$$

and

$$_{12} \qquad _{21} \qquad 1 \qquad _{11}$$

As a result, determining that

$$_{11} \qquad 1.0025$$

sets

$$_{22} \qquad 1.0025$$
$$_{12} \qquad 0.0025$$
$$_{21} \qquad 0.0025$$

Therefore, once $_{11}$ is determined, all the other ’*s* are specified for a 2 2 system. As a result, the *RGA* for a 2 2 system is typically reported as a single number (i.e., $_{11}$). In addition, if the pairing of c_1 and c_2 are switched, the *RGA* value is simply one minus the *RGA* ($_{11}$) value for the original pairing.

For the system represented in Table 15.1, if c_1 were used to control y_2 and c_2 were used to control y_1, the steady-state gain matrix would be

$$\underline{\underline{K}} \quad \begin{matrix} 0.1 & 1.0 \\ 2.0 & 0.05 \end{matrix}$$

Using Equation 15.3.2 yields

$$_{11} \quad \frac{1}{1 \quad \frac{2}{0.005}} \quad 0.0025$$

which is $_{12}$ for the original pairing scheme.

Consider the *RGA* for a 3 3 system

$$RGA \quad \begin{matrix} _{11} & _{12} & _{13} \\ _{21} & _{22} & _{23} \\ _{31} & _{32} & _{33} \end{matrix}$$

where

$$_{12} \quad \frac{\frac{y_1}{c_2} \; c_1, c_3}{\frac{y_1}{c_2} \; y_2, y_3}$$

Note that the gain in the numerator is based upon keeping c_1 and c_3 constant while the gain in the denominator is based upon keeping y_2 and y_3 constant. The closer the diagonal elements of the *RGA* are to unity, the more decoupled the process is. The sum of 's in any row or any column is equal to one. Therefore, for a 3 3 system, the *RGA* requires the determination of four of the nine possible 's

15.4 Dynamic Factors in Configuration Selection

For the 2 2 system represented in Table 15.1, all the input/output relationships have the same dynamic behavior; therefore, a steady-state analysis is sufficient. Consider the transfer function representation of a 2 2 system shown in Table 15.2. From Equation 15.3.2, the steady-state *RGA* ($_{11}$) for this system is 0.94, indicating that the control loop pairings listed in Table 15.2 are proper.

Table 15.2

Transfer Function Representation of a 2 2 System with Dynamic Coupling

	c_1	c_2
y_1	$\frac{1.0}{100s \quad 1}$	$\frac{0.3}{10s \quad 1}$
y_2	$\frac{0.4}{10s \quad 1}$	$\frac{2.0}{100s \quad 1}$

However, notice that the effect of c_1 on y_1 and the effect of c_2 on y_2 have much slower dynamics than the effect of c_1 on y_2 and the effect of c_2 on y_1 (i.e., the coupling). The time constants for the diagonal responses are ten times the time constants for the off-diagonal terms. When changes in c_1 are made to correct for deviations in y_1 from its setpoint, changes in y_2 result long before y_1 can be corrected. Then the (c_2, y_2) control loop makes changes in c_2 to correct for the coupling. Once again, because of the dynamic differences, y_1 responds to the coupling much more quickly than y_2 can be corrected. The (c_1, y_1) control loop responds to these additional changes in y_1 and the coupling process continues. This is an example of **dynamic coupling**. Figure 15.4.1 shows the dynamic response of y_1 and y_2 for a setpoint change in y_1 for the original (c, y) pairings and for the reverse pairings, i.e., (c_1, y_2) and (c_2, y_1). Even though the steady-state *RGA* of the reverse pairing is 0.06, the control performance of the reverse pairings is far superior because of its superior dynamic response.

The dynamic *RGA*[2] can be used to assess the effect of dynamics on coupling. The dynamic *RGA* is calculated by substituting $s = i$ into each transfer function comprising the input/output model. For a specific frequency, the magnitude of each transfer function (Equation 11.2.2) and the corresponding *RGA* at that frequency is calculated from Equation 15.3.2 for a 2×2 system with the transfer function magnitudes used in place of the static gains. In this manner, the *RGA* can be plotted as a function of frequency. That is,

$$ {}_{11}(\ ) \qquad \frac{1}{1 \qquad \frac{|G_{12}(i\ )||G_{21}(i\ )|}{|G_{11}(i\ )||G_{22}(i\ )|}} $$

At very low frequencies, the dynamic *RGA* approaches the steady-state *RGA* value. Therefore, a comparison of dynamic *RGA*s at intermediate frequencies distinguishes dynamic coupling effects.

Consider the calculation of the dynamic *RGA* for the 2 2 process represented in Table 15.2. Because each of the transfer functions in Table 15.2 is for a first-order process, the magnitude of G_p(i for a general first-order process is simply A_r and is given in Table 11.1

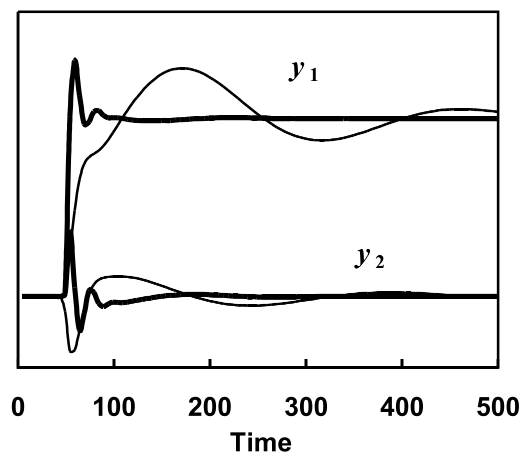


Figure 15.4.1 Response to a setpoint change in y_1 for the original pairings (thin line) and the reverse pairings (thick line) for the process represented by Table 15.2.

$$\left|G_p(i\ \ )\right| \qquad \frac{K_p}{\sqrt{\ {}^{2}_{p} \qquad 1}}$$

Then, using Equation 15.3.2, the dynamic *RGA* for the (c_1, y_1) pairing listed in Table 15.2 is given by

$$_{11}(\ ) \qquad \frac{1}{1 \quad \dfrac{|G_{12}(i\ )||G_{21}(i\ )|}{|G_{11}(i\ )||G_{22}(i\ )|}} \qquad \frac{1}{1 \quad \dfrac{\dfrac{0.3}{\sqrt{10^2\ {}^2\quad 1}}\ \dfrac{0.4}{\sqrt{10^2\ {}^2\quad 1}}}{\dfrac{1}{\sqrt{100^2\ {}^2\quad 1}}\ \dfrac{2}{\sqrt{100^2\ {}^2\quad 1}}}} \qquad \frac{1}{1 \quad \dfrac{0.06(100^2\ {}^2\quad 1)}{(10^2\ {}^2\quad 1)}}$$

Likewise, the dynamic *RGA* for the opposite pairing is given by

$$_{12}(\ ) \qquad \frac{1}{1 \quad \dfrac{10^2\ {}^2\quad 1}{0.06(100^2\ {}^2\quad 1)}}$$

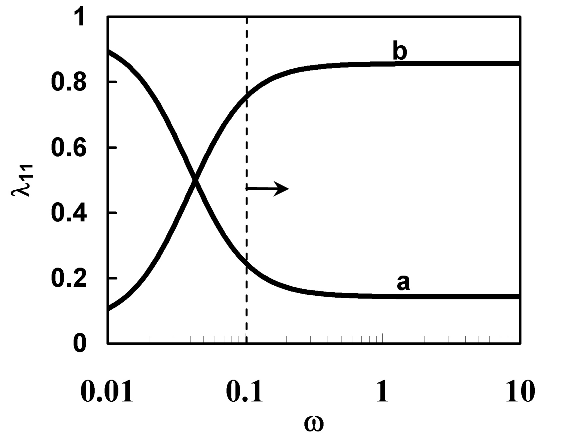


Figure 15.4.2 Dynamic RGA for (a) the original pairings and (b) the reverse pairings for the process represented in Table 15.2.

Figure 15.4.2 shows the dynamic *RGA*s for these two cases. At very low frequencies, which corresponds to the steady-state *RGA*, the original pairing provides superior decoupling, while the reverse pairing is preferred at high frequencies. On the other hand, as shown in Figure 15.4.1, the configuration with the reverse pairing responds much more quickly. In fact, the frequency corresponding to the time constant of the off-diagonal interaction (i.e., $_p$ equal to 10) corresponds to a frequency of 0.1 radians per unit time. At this frequency and higher, the reverse pairing is clearly preferred, as shown in Figure 15.4.2.

Because transfer function models are not usually available for industrial processes, it is recommended to qualitatively use the results of this section when choosing (c, y) pairings. That is, **when selecting a MV for a particular CV, choose a MV that causes the CV to exhibit a relatively fast dynamic response, i.e., choose pairings that have a relatively low effective time constant and a relatively low effective process deadtime**. It is important to remember to use your understanding of the process to guide your analysis.

15.5 Sensitivity to Disturbances

In general, each possible configuration has a different sensitivity to a particular disturbance. Consider the distillation column shown in Figure 15.5.1. The reflux flow rate, L, is used to control the overhead composition and the boilup rate, V, which is set by the reboiler duty, is used to control the bottom composition. Because L and

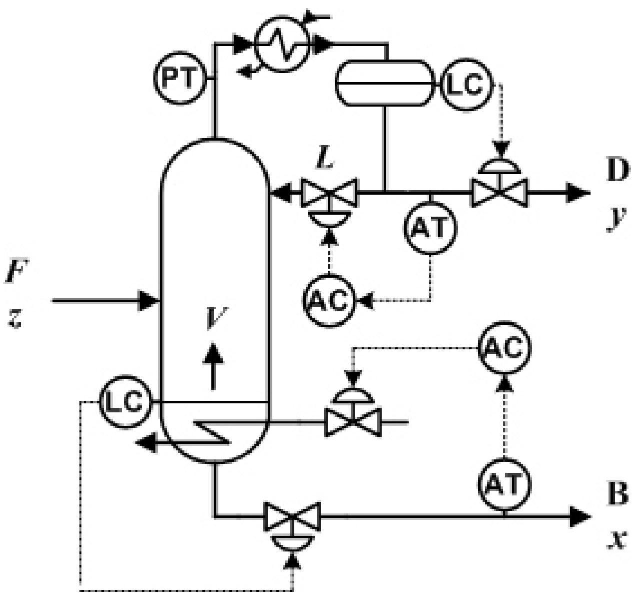


Figure 15.5.1 Schematic of a distillation column with the (*L*,*V*) configuration. Note that control valves represent flow control loops in this figure.

V are used for composition control, the distillate flow rate, *D*, is used to control the accumulator level and the bottoms flow rate, *B*, is used to control the reboiler level. This configuration is referred to as the (*L*,*V*) configuration because *L* is used to control the overhead product composition and *V* is used to control the bottoms product composition. Consider the (*L/D*, *V/B*) configuration. Figure 15.5.2 shows how the reflux ratio, *L/D*, can be applied to control the overhead product composition. Note that a ratio controller is used for the composition controller and *D* is used to control the level in the accumulator. In a similar manner, the boilup ratio *V/B* can be used to control the bottoms product composition. That is, a *V/B* ratio controller can be used to control the bottoms product composition while the bottoms product flow rate, *B*, is used to control the level in the reboiler. Figure 15.5.3 shows the "open-loop" response of a distillation column to a step increase in the light component composition in the feed for the (*L*,*V*) and (*L/D*,*V/B*) configurations. In this case, open-loop response refers to opening the composition control loops while maintaining the level controllers in closed-loop operation. The (*L*,*V*) configuration is less sensitive to feed composition changes than the (*L/D*,*V/B*) configuration. As a result, the (*L*,*V*) configuration is less affected by feed composition changes than the (*L/D*,*V/B*) configuration.

From an analysis of the previous sections it can be concluded that **the combined effect of coupling, dynamic behavior and sensitivity to disturbances determine the control performance for a particular control**

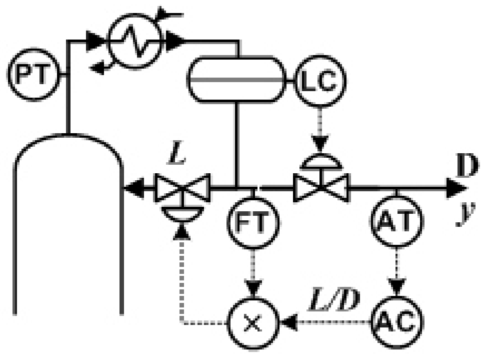


Figure 15.5.2 Schematic of a reflux ratio controller applied for the control of the overhead composition. Note that control valves represent flow control loops in this figure.

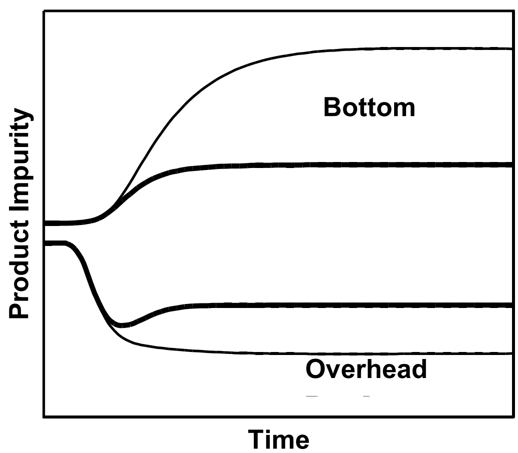


Figure 15.5.3 The open-loop response of a distillation column with the (*L*,*V*) configuration (thick line) and the (*L/D*,*V/B*) configuration (thin line) to a step change in feed composition.

configuration for a MIMO process. Process control engineers typically rely upon their understanding of the process and their experience when selecting a control configuration for a MIMO process.

Example 15.1 Configuration Selection for a C_3 Splitter

Problem Statement. Evaluate the configuration selection problem for a C_3 splitter. A C_3 splitter separates a feed mixture primarily composed of propane and propylene into polymer-grade propylene (< 0.5% propane) and fuel-grade propane (approximately 2% propylene).

Solution. The nomenclature used here refers to a particular configuration as (c_1, c_2) in which c_1 is the MV that is used to control the overhead composition and c_2 is the MV that is used to control the bottoms composition. If we limit ourselves to controlling the overhead composition with *L*, *D* or *L*/*D* (the reflux ratio) and the bottoms composition with *V*, *B* or *V*/*B* (the boilup ratio), there are a total of nine possible configurations. Here we limit the discussion to the following configurations for purposes of illustration: (*L, B*), (*L, V*), (*L/D, V/B*) and (*D, V*).

The steady-state *RGA*s for each of the configurations considered here are listed below:

Table 15.3
RGA Values for Example 15.1

Configuration	*RGA* ($_{11}$)
(*L, B*)	0.94
(*L, V*)	25.3
(*L/D, V/B*)	1.70
(*D, V*)	0.06

Shinskey[3] recommends *RGA* values between 0.9 and 3.0 for distillation columns, indicating that *RGA* values greater than one are preferred over *RGA* values less than one. Based upon Shinskey's guidelines and the results in Table 15.3, the (*L, B*) and (*L/D, V/B*) configurations appear the most promising.

The dynamics of distillation columns can be understood by recognizing that product composition changes result from changes in the vapor/liquid traffic in the column. That is, changing *L* or *V* directly affects the column vapor/liquid traffic and have the most immediate effect on the product composition. On the other hand, changes in *B* and *D* depend on the level controllers to change the vapor/liquid traffic of the column; therefore, the dynamic response of the product compositions is significantly slower when *B* and *D* are changed compared with changing *L* and *V*. The dynamic response to changes in *L/D* and *V/B* are intermediate between *L* and *V* on the fast side and *B* and *D* on the slow side. For example, the dynamics of changes in *L/D* are faster than changes in *D* but slower than changes in *L*. Based upon this analysis, (*L, B*) is expected to perform better for the overhead composition control than for the bottoms, but there is no clear winner between the (*L, B*) and the (*L/D, V/B*) configurations with regard to the overall dynamic response.

Table 15.4 shows the relative changes in each MV for a change in feed composition. This table is based on steady-state results in which the product compositions are maintained at a constant level. A lower relative change for a MV indicates a reduced sensitivity to feed composition changes for that MV. *L*, *L/D* and *V* show the least sensitivity to feed composition changes.

Table 15.4

Relative Changes in the MVs to Maintain the Product Purities for a 5 mole % Increase in Feed Composition.

MV	Percentage Change
L	4.2
D	7.4
L/D	-3.0
V	4.4
B	-16.8
V/B	25.5

Table 15.5 lists the integral absolute error (IAE) of each product for each configuration for a feed composition upset based on tray-to-tray dynamic distillation simulation results. A lower IAE value indicates closer control to setpoint. The (*L, B*) configuration provided the best overall control performance, especially for the overhead product. This is consistent with the observations that *L* is dynamically fast and relatively insensitive to feed composition changes coupled with the moderate steady-state coupling as indicated by the *RGA*.

The (*L,V*) configuration has the advantages of fast overall dynamics and insensitivity to feed composition upsets. These advantages are negated by the extreme degree of steady-state coupling as indicated by its steady-state RGA value. The control performance of the (*L,V*) configuration is the poorest of the four configurations listed in Table 15.5. The (*L/D, V/B*) configuration has a good steady-state *RGA* and dynamic characteristics, but is particularly sensitive to feed composition upsets for the bottom composition control loop. As a result, its performance is inferior to the (*L, B*) configuration. The steady-state *RGA* value of the (*D, V*) configuration indicates that this configuration will not function properly. Actually, the control performance of the (*D*,*V*) configuration is quite reasonable, i.e., the IAEs for the (*D*,*V*) configuration are only about 30% larger than those for the (*L*,*B*) configuration.

Table 15.5

Control Performance (IAE) for a Step Change in Feed Composition

Configuration	IAE for Overhead	IAE for Bottoms
(*L*,*B*)	0.067	1.49
(*L*,*V*)	0.250	15.3
(*L/D*,*V/B*)	0.095	2.00
(*D*,*V*)	0.098	1.91

For complex configuration selection problems, such as distillation columns, the previous analysis is helpful but does not always guarantee that the best configuration is identified. The performance differences between reasonable configuration choices and the best configuration can be substantial.

15.6 Tuning Decentralized Controllers

The recommended tuning procedure for a single PID loop can be extended to tuning single-loop PID controllers applied for decentralized control of a MIMO process. The first step in tuning a decentralized controller is to apply ATV tests to each MV/CV pair. While an ATV test is being applied to one loop, the other loops should be maintained in an open-loop condition.

Next, determine if any of the loops are significantly faster-responding than the other loops. This can be done by comparing the values of the ultimate periods, P_u, obtained in the ATV tests. If the smallest value of P_u is smaller by a factor of 3 or more than the next smallest P_u value, that loop should be implemented by itself before tuning the other loops. It can be tuned as a single PID loop as discussed in Chapter 9. Then, ATV tests on the remaining loops should be rerun with the tuned fast loop in service (closed-loop operation). Then, the remaining control loops can be tuned using the following procedure.

Consider the tuning of PI controllers for a 2 2 MIMO process. The ATV results are used to select the controller gain and reset time based on an appropriate tuning criterion, e.g., Ziegler-Nichols method. Then, a single tuning factor, F_T, is applied to the tuning parameters for **both control loops**.

$$
\begin{array}{lll}
K_c \quad K_c^{ZN} / F_T & & \text{First control loop} \\
{}_I \quad {}_I^{ZN} \quad F_T & & \\
 & & \\
K_c \quad K_c^{ZN} / F_T & & \text{Second control loop} \\
{}_I \quad {}_I^{ZN} \quad F_T & &
\end{array}
\qquad \textbf{15.6.1}
$$

F_T is adjusted until the dynamic response satisfies the control loop tuning criterion. For example, setpoint changes in y_1 and/or y_2 can be used to select the proper value of F_T. Alternatively, the value of F_T can be adjusted to provide reliable performance of the controllers based on day-to-day controller operating performance. While tuning, if the closed-loop response is sluggish, the value of F_T is decreased. If the controller exhibits periods of ringing, the value of F_T is increased.

After F_T has been adjusted to tune the set of decentralized PI controllers, fine tuning of the controller settings should be used. For example, if you observe that one of the control loops is slow to remove offset, an increase in integral action for that loop should be implemented. If one of the loops exhibits ringing, derivative action should be tested to determine if it improves the feedback control performance of that loop. In the latter case, derivative action should be tuned in the manner described in Chapter 9.

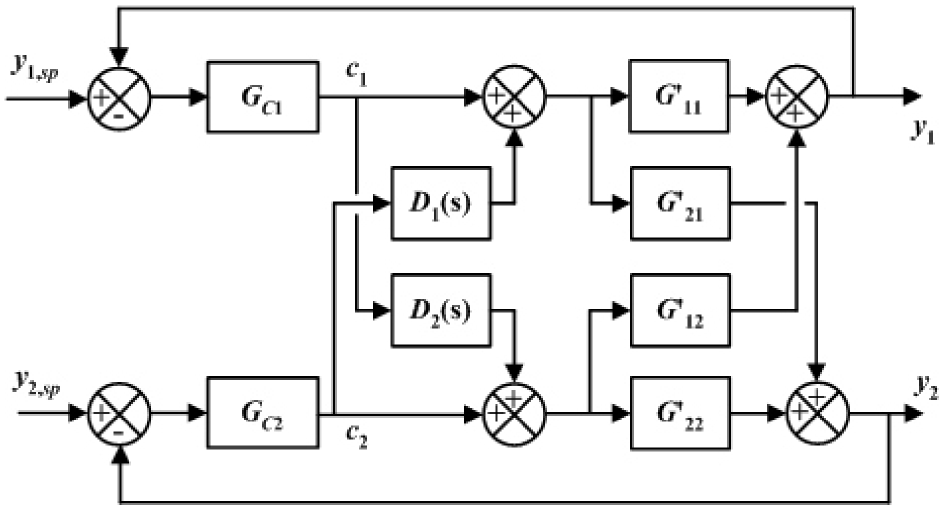


Figure 15.7.1 Block diagram of a two-input/two-output process with two-way decoupling.

15.7 Decouplers

Decouplers are designed to reduce the detrimental effects of coupling. Figure 15.7.1 shows two decouplers [$D_1(s)$ and $D_2(s)$] applied to a two-input/two-output process. D_1(s) is designed to reduce the effects of changes in c_2 on y_1 while D_2(s) is designed to reduce the effects of changes in c_1 on y_2.

Consider the design of D_1(s). The effect of changes in c_2 on y_1 is given by

$$G_{12}(s)C_2(s)$$

The corrective action on y_1 from D_1 is given by

$$D_1(s)G_{11}(s)C_2(s)$$

The objective of the decoupler is to eliminate the effect of coupling; therefore, the sum of the previous two terms is set to zero, i.e.,

$$G_{12}(s)C_2(s) \quad D_1(s)G_{11}(s)C_2(s) \quad 0$$

Solving for $D_1(s)$ yields

$$D_1(s) \quad \frac{G_{12}(s)}{G_{11}(s)} \qquad \textbf{15.7.1}$$

Using a similar analysis

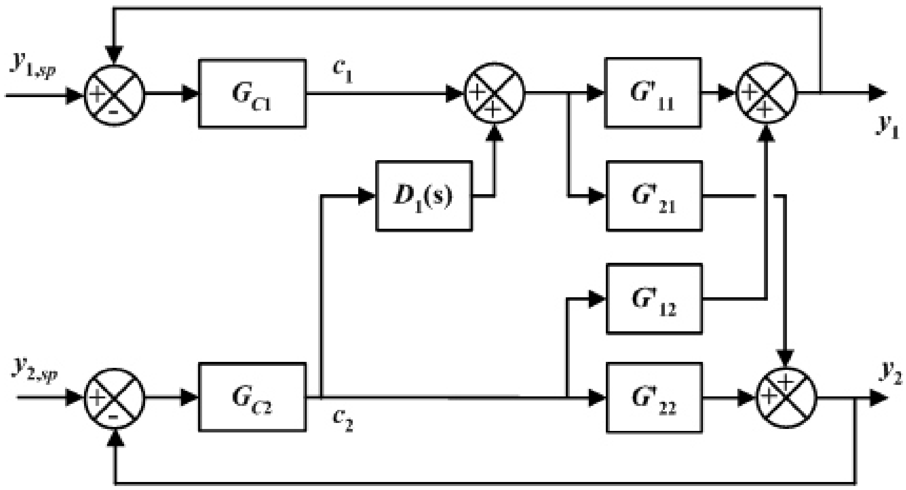


Figure 15.7.2 Block diagram of a two-input/two-output process with one-way decoupling to reduce the effect of c_2 on y_1.

$$D_2(s) \quad \frac{G_{21}(s)}{G_{22}(s)} \tag{15.7.2}$$

The decoupler $D_1(s)$ can be viewed as a "feedforward" correction to y_1 using c_1 for disturbances caused by changes in c_2. Note that Equations 15.7.1 and 15.7.2 are similar to the general equation for a feedforward controller (Equation 12.4.1) where $G'_{12}(s)$ corresponds to $G_d(s)$ and $G'_{11}(s)$ corresponds to $G_p(s)$. A multivariable decoupler design can be used, but is not presented here.

Figure 15.7.1 shows a 2×2 system with two decouplers, which is referred as a **two-way** or **complete decoupler**. Two-way decouplers are rarely used in industry because, many times, they result in poorer control performance than conventional control without decouplers. This results because two-way decouplers can be sensitive to nonlinearity and modeling errors in the decouplers. On the other hand, **one-way** or **partial decouplers** (Figure 15.7.2) are much more reliable and are more frequently used industrially. One-way decouplers are particularly useful when the key CV of a MIMO process suffers from significant coupling.

15.8 Controllability and Observability [AT]

Controllability. For the state space representation of a system, i.e.,

$$\begin{aligned} \frac{d\underline{x}}{dt} \quad & \underline{\underline{A}}\,\underline{x} + \underline{\underline{B}}\,\underline{u} \\ \underline{y} \quad & \underline{\underline{C}}\,\underline{x} \end{aligned} \tag{5.8.6}$$

it is important to know if you can control all the states by manipulating the inputs. For the state **controllability of a process**, is defined as, **given the initial state $\underline{x}_0 \quad \underline{x}(0)$ and any other arbitrary state $\underline{x}_i \quad \underline{x}(t_i)$, there exists a time, t_i >0, and a manipulative variable, $\underline{u}(t)$ defined on the interval (0 and t_i) which takes the state $\underline{x}_0$ to $\underline{x}_i$. If this condition is not satisfied, the process is uncontrollable.**

A mathematical test for a system given by the state space equation above is the following:

A system is controllable if the rank of the controllability matrix, $\underline{\underline{\ }}_c \quad \underline{\underline{B}} \quad \underline{\underline{AB}} \quad \underline{\underline{A}}^2\underline{\underline{B}} \quad \mathrm{K} \quad \underline{\underline{A}}^{n\ 1}\underline{\underline{B}}$, has full rank, n, where n is the dimension of matrix, $\underline{\underline{A}}$ $(n\times n)$. The rank of a matrix is equal to the number of independent rows or columns in the matrix.

Example 15.2 State Space Controllability Test

Problem Statement. Determine the controllability of a process defined by the following state space matrices.

$$\underline{\underline{A}} \quad \begin{matrix} 0 & 1 \\ 3 & 4 \end{matrix} \qquad \underline{\underline{B}} \quad \begin{matrix} 1 & 2 \\ 3 & 4 \end{matrix}$$

Solution. For the second term in the controllability matrix,

$$\underline{\underline{AB}} \quad \begin{matrix} 0 & 1 \\ 3 & 4 \end{matrix} \quad \begin{matrix} 1 & 2 \\ 3 & 4 \end{matrix} \quad \begin{matrix} 3 & 4 \\ 9 & 10 \end{matrix}$$

Forming the controllability matrix,

$$\underline{\underline{\ }}_c \quad \underline{\underline{B}} \quad \underline{\underline{AB}} \quad \begin{matrix} 1 & 2 & -3 & -4 \\ -3 & -4 & 9 & 10 \end{matrix}$$

which has full rank (2). That is, because the second row is not a scaled version of the first row, the rank of the controllability matrix is 2. Therefore, this system defined by the state space model is controllable.

Controllability can also be defined from the input-output representation (transfer function matrix, Equation 5.9.1) as well. **A system is called input/output controllable if the determinant of the transfer function matrix, $\underline{\underline{G}}_p(s)$, is nonzero.**

Example 15.3 Transfer Function Matrix Controllability Test

Problem Statement. For the following transfer function matrix, determine if the system is controllable.

$$\underline{\underline{G}}_p(s) \quad \begin{matrix} \dfrac{2}{s \quad 3} & \dfrac{2}{s \quad 3} \\ \dfrac{1}{s \quad 3} & \dfrac{1}{s \quad 3} \end{matrix}$$

Solution. The determinant of the transfer function matrix is given by

$$\det \underline{\underline{G}}_p(s) \quad a_{11}a_{22} \quad a_{12}a_{21} \quad \frac{2}{(s \quad 3)^2} \quad \frac{2}{(s \quad 3)^2} \quad 0$$

Therefore, this system is not controllable.

Observability. It is important to know if the output we are measuring represents states of the process. A mathematical representation, which addresses this issue, is called the observability. **A state space system is observable, if there exists a time, t_i>0, in which given the vectors $\underline{u}$ and $\underline{y}$, we can calculate the initial state vector, $\underline{x}_0$.** The mathematical criterion is as follows:

If the observability matrix, $\underline{\underline{\ }}_0$, is full rank, the system is observable where the observability matrix is given by

$$\underline{\underline{\ }}_0 \quad \underline{\underline{C}}^T \quad \underline{\underline{A}}^T\underline{\underline{C}}^T \quad \underline{\underline{A}}^T \,{}^2\, \underline{\underline{C}}^T \quad \mathrm{K} \quad \underline{\underline{A}}^T \,{}^{n \;\; 1}\, \underline{\underline{C}}^T$$

Example 15.4 State Space Observability Test

Problem Statement. Determine the observability of the following system.

$$\underline{\underline{A}} \quad \begin{matrix} 0 & 1 \\ 3 & 4 \end{matrix} \qquad \underline{\underline{C}} \quad \begin{matrix} 1 & 1 \\ 2 & 2 \end{matrix}$$

Solution. First, form the first two matrices in the expression for the observability matrix.

$$\underline{\underline{C}}^T \quad \begin{matrix} 1 & 2 \\ 1 & 2 \end{matrix} \qquad \underline{\underline{A}}^T\underline{\underline{C}}^T \quad \begin{matrix} 3 & 6 \\ 3 & 6 \end{matrix}$$

Then,

$$\underline{\underline{\ }}_0 \quad \begin{matrix} 1 & 2 & 3 & 6 \\ 1 & 2 & 3 & 6 \end{matrix}$$

Because the rows in the observability matrix are not independent, the rank is one. Therefore, this system is not observable. Observability is a precondition if we want to estimate the states of the process from the measurements of inputs and outputs.

15.9 Summary

- The control performance for a particular control configuration applied to a MIMO process depends on three factors:
  1. Coupling
  2. CV/MV dynamics
  3. Sensitivity to disturbances
- The steady-state RGA and the dynamic RGA can be used to assess the steady-state and dynamic coupling, respectively, for a particular control configuration.
- It is desirable to choose (c,y) pairings such that each y responds quickly to changes in the c with which it is paired.
- Each control configuration for a MIMO process has its own specific sensitivity to disturbances.
- Once a control configuration is selected, the decentralized controllers can be tuned using an extension of the ATV tuning procedure recommended for a single PID control loop.
- One-way decouplers can be beneficial when the most important CV in a MIMO process suffers from significant coupling from one or more of the other loops.

15.10 Additional Terminology

Complete decoupling - a decoupler for each CV.
Control configuration - the particular pairing of MVs and CVs for a MIMO process.
Coupling - the effect of one control loop on another in a MIMO process.
Decentralized control - applying single loop PID controllers to a MIMO process.
Dynamic coupling - coupling that includes dynamic differences between various input/output pairs. It can be evaluated using the dynamic RGA.
MV/CV pairings - the choice of which MV is used to control which CV.
One-way decoupler - a single decoupler applied to a MIMO process.
Partial decoupler - fewer decouplers applied than the number of CVs.
***RGA* -** the relative gain array, which indicates the degree of steady-state coupling.
Single-loop PID controllers - PID controllers that are applied to a MIMO process.
Two-way decoupler - two decouplers applied to a 2×2 process.

15.11 References

1. Bristol, E.H., *IEEE Trans Auto. Con.*, AC-11, p. 133 (1966).
2. McAvoy, T.J., *Interaction Analysis*, Instrument Society of America, pp. 190-192 (1983).
3. Shinskey, F.G., *Distillation Control*, 2nd Edition, McGraw-Hill, pp. 154-165 (1984).

15.12 Preliminary Questions

15.2 SISO Controllers and (c, y) Pairings

Q15.2.1 Explain how coupling occurs in a MIMO process.

Q15.2.2 Why is it important to choose a good pairing of Ms and CVs when applying control to a MIMO process?

15.3 Steady-State Coupling

Q15.3.1 Explain how the steady-state *RGA* is a measure of steady-state coupling.

Q15.3.2 When is the steady-state *RGA* the best criterion for selecting (c,y) pairings?

Q15.3.3 For a 2×2 system, why is the *RG*A given by a single number when the RGA is, in fact, a matrix of numbers?

Q15.3.4 What does a negative *RGA* indicate?

Q15.3.5 What does an RGA less than unity indicate?

Q15.3.6 What does an RGA greater than unity indicate?

15.4 Dynamic Factors in Configuration Selection

Q15.4.1 How does you take dynamic coupling into account when comparing control configurations?

Q15.4.2 Why is dynamic coupling important to the performance of a control system on a MIMO process?

15.5 Sensitivity to Disturbances

Q15.5.1 Why is the sensitivity of a configuration to disturbances important in the selection of a control configuration for a MIMO process?

Q15.5.2 What advantage results from using a control configuration that has a low sensitivity to a disturbance?

Q15.5.3 For Example 15.1, what factors contribute to the (L,B) configuration being the best performing configuration considered?

15.6 Tuning Decentralized Controllers

Q15.6.1 When tuning decentralized controllers applied to a MIMO process, how do you tune loops that are much faster than the other loops?

Q15.6.2 After applying Equation 15.6.1, what tuning adjustments should be made to a decentralized controller if the response of the process is too sluggish?

Q15.6.3 After applying Equation 15.6.1, what tuning adjustments should be made to a decentralized controller if the response of the process is ringing?

Q15.6.4 Assume that after applying Equation 15.6.1 to tune a decentralized controller on a MIMO process, one of the loop exhibits slow offset removal. What action would you take?

Q15.6.5 Assume that after applying Equation 15.6.1 to tune a decentralized controller on a MIMO process, one of the loops exhibits sluggish response to a setpoint change. What action should be taken?

15.7 Decouplers

Q15.19 When should a one-way decoupler be used?

Q15.20 Why are two-way decouplers rarely used in industry?

Chapter 16

Model Predictive Control

Chapter Objectives

- Introduce the multivariable approach to process control.
- Present step response modeling.
- Present the derivation of the DMC control law.
- Outline the organization of an industrial project for the application of an MPC controller.

16.1 Introduction

In Chapter 15, the application of conventional PID controllers to a multivariable (MIMO) process was presented. In this chapter, the application of **model predictive control (MPC)** for the control of MIMO systems is considered. MPC is a type of **multivariable control**, also known as **centralized control**, which can use all available process measurements (i.e., measured DVs and CVs) along with process models to simultaneously determine the values of all the MVs for control of a MIMO process. MPC can be defined as a multivariable controller that uses dynamic models of the MIMO process to calculate a control response into the future. Because MPC uses MIMO process models, it is also known as a **model-based controller**.

Diagrams comparing a conventional decentralized PID controller and a MPC controller are shown in Figures 16.1.1 and 16.1.2, respectively. The MPC controller in this case uses y_1 and y_2 and their setpoints as well as the disturbance levels (d_1 and d_2) to simultaneously calculate u_1 and u_2. The MPC model used in the controller considers the effect of u_1 on y_1 and y_2 and the effect of u_2 on y_1 and y_2 when determining the control action; therefore, the MPC controller provides decoupling. In addition, if the MPC model considers the effect of measured **disturbance variables** (**DVs**) on the process, it will also provide feedforward compensation for the DVs. In Figure 16.1.2, the DVs are d_1 and d_2, the MVs are u_1 and u_2 and the CVs are y_1 and y_2. In this case, the MVs and DVs are inputs to the process while the CVs are the outputs. On the other hand, MVs are set by the MPC controller while the DVs are measured on the process and are usually independent of the controller. For simplicity, these diagrams do not show the actuators, i.e., the outputs of the controllers are shown as MVs, or sensors.

Linear MPC is the most widely used form of multivariable control. MPC has been widely accepted by the process industries because it provides a convenient means to combine control and optimization by implementing constraint control. That is, a linearized optimization algorithm, which is discussed later in this chapter, is typically used to identify the economically optimum set of process constraints. In this manner, the optimizer

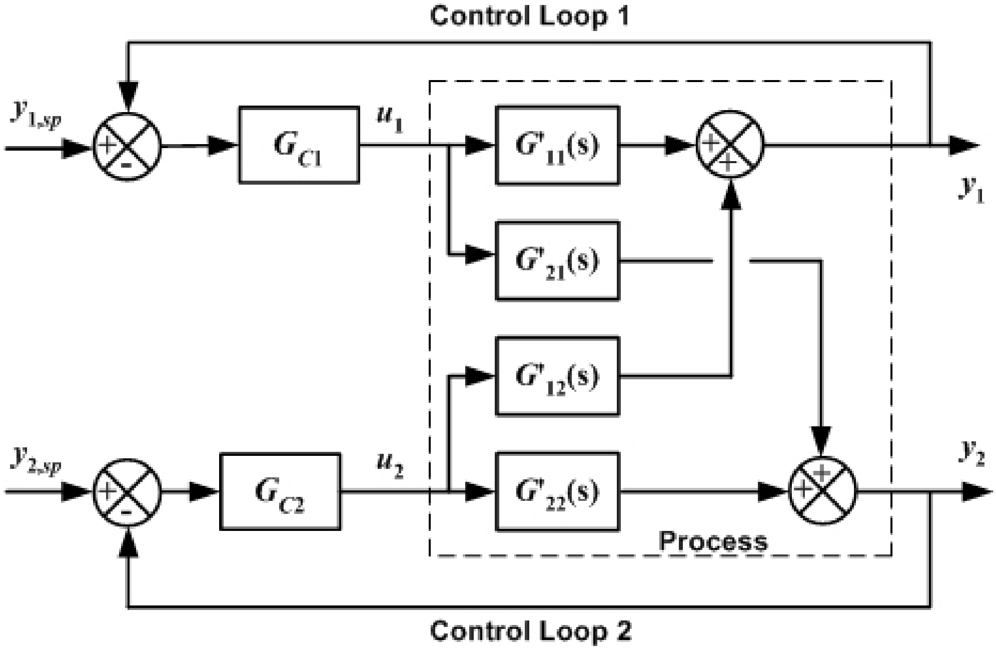


Figure 16.1.1 Schematic for single loop PID controllers applied to a 2×2 MIMO process.

determines the setpoints for the MPC controller that maximize the profit of the operation. It is unlikely that MPC would have been so widely used in the CPI without its ability to effectively coordinate control with a process optimization algorithm.

Single-loop PID controllers can also be applied to maintain a process at an optimum set of constraints, but as the sets of operative constraints change, the control configuration and the controller tuning usually must be changed. That is, controlling to each combination of constraints generally requires a separate control configuration and a separate set of controller tuning parameters. As a result, for industrial scale MIMO processes, e.g., a separation train in an ethylene plant or a fluidized catalytic cracking (FCC) unit in a refinery, the number of possible combinations of constraints is well over a million. Therefore, it is not normally practical to implement constraint control with single-loop PID controllers for industrial-scale processes because of the large number of constraints sets that are typically encountered. On the other hand, MPC is able to handle constraint control for a large set of constraints using a manageable set of tuning parameters because MPC uses models of the effect of each MV and DV on each constraint.

With the availability of computers in the 1960s, oil refiners began to attempt process optimization applications for key processing units, such as FCC units and crude units. Even though they were able to identify the optimum operating conditions for these units, their single-loop PID controllers were unable to perform the constraint control necessary to maintain operation at the desired operating conditions. As additional constraints combinations were added to the constraint controllers, it soon became apparent that it was not possible to anticipate all the possible combinations of constraint sets. This realization lead to the development and industrial implementation of linear MPC.

The first known industrial application of linear MPC was implemented by Charles R. Cutler at the Shell Refinery in New Orleans, LA in 1973 and was known as **Dynamic Matrix Control** (**DMC**). In the mid 1970s, Richalet

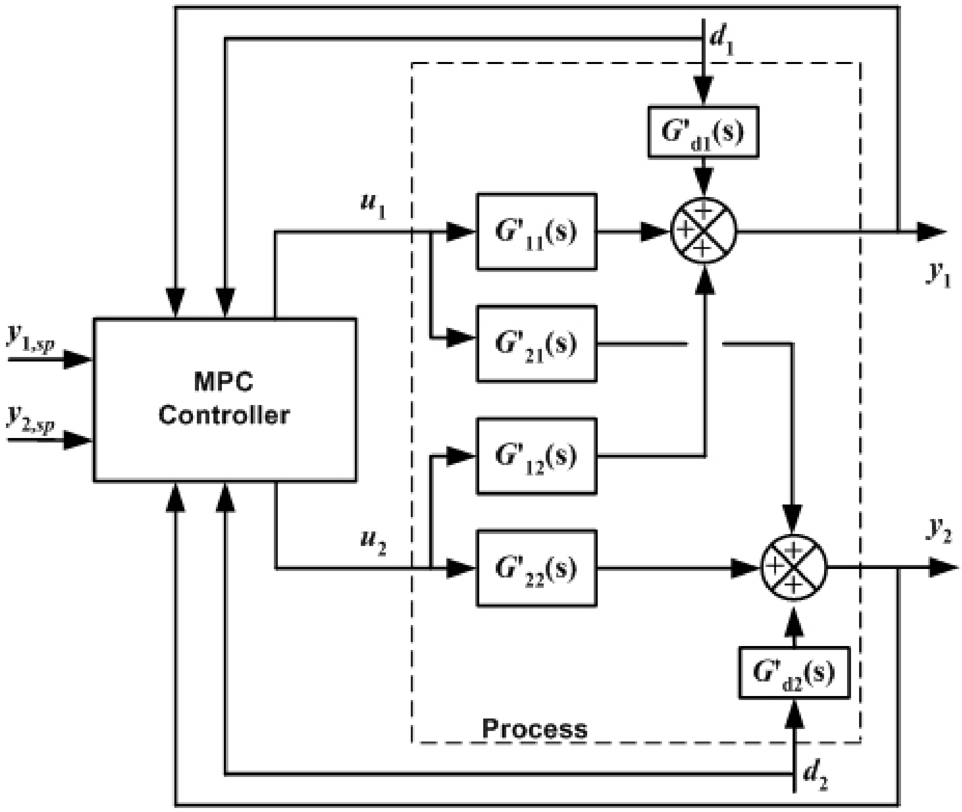


Figure 16.1.2 Schematic of a multivariable controller applied to a 2×2 MIMO process.

et.al[1] developed a model-based controller (IDCOM) which was originally designed to implement a fixed trajectory for the flight of a missile. Setpoint Inc. licensed the IDCOM controller and applied it to MIMO control problems in the CPI. Although IDCOM was advertized as a multivariable controller, it paired each CV with a single MV and used switching logic to change to a new MV when the original MV became saturated. During this period most of the development work for MPC was performed by Shell Oil Company and later by Setpoint, Inc. During the 1980s, a number of companies in the CPI began to implement MPC through collaboration with Setpoint and DMC Corporation, which was founded by Charles R. Cutler in 1984. During the 1990s, widespread adoption of linear MPC occurred. Today, there are a number of commercial vendors that offer MPC controllers and almost all companies in the CPI rely on MPC for their advanced control applications. It was estimated[2] that there were almost 5,000 industrial linear MPC applications worldwide in 1999. DMC[3] is the most popular form of MPC. The next several sections consider DMC applied to a SISO process and show how it can be extended to MIMO processes.

Initially, DMC is introduced for the SISO case. The modeling approach for DMC is described and used to develop the dynamic matrix. Then the effect of previous inputs on the future behavior is introduced as the prediction vector. The DMC control law is derived and the model identification process is considered. Next, the extension of DMC to MIMO cases is considered as well as how DMC is applied to constraint control and process optimization. Finally, the organization of a typical MPC industrial application project is presented.

16.2 Step Response Models (SRMs)

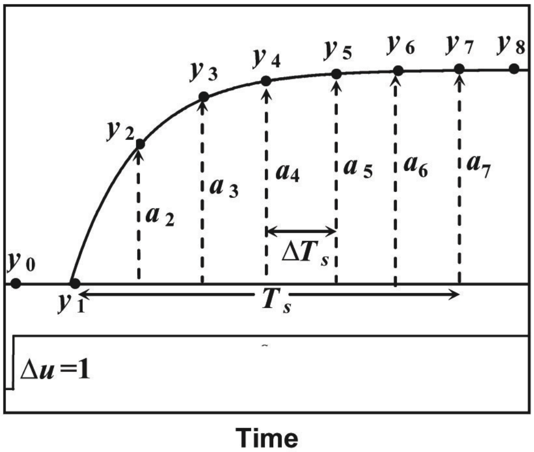


Figure 16.2.1 Open-loop response of a FOPDT process and the corresponding SRM coefficients.

DMC uses **step response models** (**SRM**s) of the process, which are linear input/output models, to calculate control action. Previously, we used transfer functions to represent the effect of the MV on the CV, i.e., $G_p(s)$. The same information contained in $G_p(s)$ can be represented using a SRM. Moreover, SRMs offer unique flexibility with regard to representing complex dynamic behavior, e.g., the dynamics of processes that involve recycle.

Consider a FOPDT process (K_p 1, $_p$ 1, $_p$ 1). The step response for u 1 applied at t =0 is shown in Table 16.1 and in Figure 16.2.1 for discrete points in time. **The coefficients of a SRM are defined as the open-loop response of a process at discrete points in time for a unit step input change.** Therefore, because the results shown in Table 16.1 and Figure 16.2.1 are based on a unit step input, the SRM coefficients are simply the noted process response at each sampling interval, T_s. Note that these results are based on a fixed sampling time, T_s, of 1.0. Because the step response is effectively complete after 7 sampling time intervals, a_i is constant for i greater than or equal to 7. A SRM of a process can be obtained from a step test based on an input other than a unit step using the following equation

$$a_i \qquad \frac{y(t_i)}{u(t_0)} \tag{16.2.1}$$

assuming that the process is at steady state at $t = t_0$ and a single step input change is made at $t \quad t_0$. a_i is called a step response coefficient. The values of the coefficients of the **SRM**, a_i, for this case are also listed in Table 16.1. The response of a process output variable [$y(t_i)$] can be expressed in terms of the SRM coefficients by rearranging Equation 16.2.1.

$$y(t_i) \quad y(t_i) \quad y(t_0) \quad a_i \quad u(t_0) \tag{16.2.2}$$

The **time to steady state** (***T***$_{ss}$) is defined as the time interval after an input change for the output variable to reach steady state. For the input/output relationship presented in Table 16.1 and in Figure 16.2.1, the T_{ss} is 7 time units. Note that T_{ss} can be easily identified from a SRM by determining when a_i reaches its steady-state value and noting the value of T_s.

Table 16.1

Step Response and Step Response Coefficients for a FOPDT Process (K_p 1, $_p$ 1, $_p$ 1) for a Step Change in u.

Time	Sample Number, i	$u(t)$	$y(t)$	a_i
0	0	1	0	0
1	1	0	0	0
2	2	0	0.63	0.63
3	3	0	0.87	0.87
4	4	0	0.95	0.95
5	5	0	0.98	0.98
6	6	0	0.99	0.99
7	7	0	1.00	1.00
8	8	0	1.00	1.00

The flexible form of a SRM, i.e., the actual response of the process determines the model coefficients (a_i), **accommodates a full range of dynamic behavior** (e.g., inverse action, overshoot, etc.). On the other hand, the price that is paid for this flexibility is a relatively large number of model parameters that are required. Figure 16.2.2 shows a step response for a more complicated process (i.e., an inverse response). Table 16.2 lists the coefficients of the SRM for this process based on t_0=100 (i.e., the input change was applied at t=100). While this modeling approach can be effectively applied to a large number of industrial processes, the models determined from process measurements are not perfect because they have varying amounts of **process/model mismatch**. Process/model mismatch is the error between the model prediction and the actual process response and results from process nonlinearity, process noise and unmeasured disturbances. All industrial processes have some degree of nonlinearity which causes mismatch for SRMs because they are linear models. Sensor noise and process variability due to physical behavior and coupled regulatory controllers are sources of process noise. In addition, unmeasured disturbances that occur during an MPC test can affect the model parameters determined by process measurements because they cause changes in the CV that are not caused by the MVs and the measured DVs. These issues will be addressed in more detail in Section 16.10, which considers identification of SRM from plant tests. SRMs are used to model the effect of MVs on CVs and the effect of DVs on CVs. Note that CVs include process constraints.

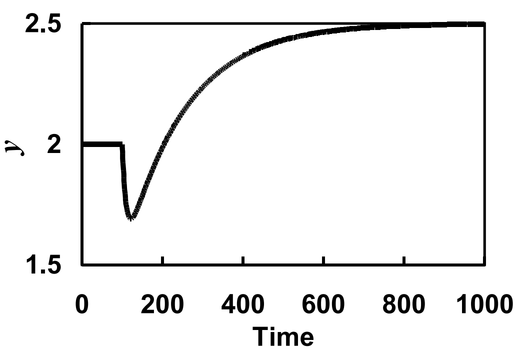


Figure 16.2.2 Dynamic response of a complex process.

Table 16.2
Coefficients of a SRM for a Process with a Complex Response (T_s 20)

i	a_i	i	a_i	i	a_i
1	-0.31	11	0.27	21	0.44
2	-0.26	12	0.30	22	0.45
3	-0.17	13	0.32	23	0.45
4	-0.09	14	0.34	24	0.46
5	-0.02	15	0.36	25	0.46
6	0.05	16	0.38	26	0.47
7	0.10	17	0.40	27	0.47
8	0.15	18	0.41	28	0.48
9	0.20	19	0.42	29	0.48
10	0.23	20	0.43	30	0.48

Example 16.1 SRM for the Thermal Mixer Process

Problem Statement. Using the VBS (Section 3.9) for the thermal mixing process (Example 3.3) for a 10% increase in the MV, develop a SRM that has 10 model coefficients.

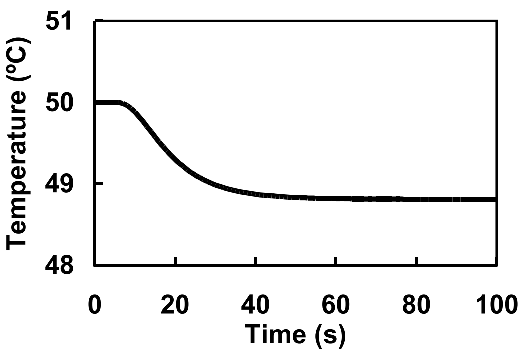


Figure 16.2.3 Open-loop response of the thermal mixing process to an MV change.

Solution. Figure 16.2.3 shows the open loop response of the thermal mixing process for 10% increase in the MV. From this figure, the time to steady state (T_{ss}) is equal to approximately (55-5) s because the process seems to settle after 55 s and the input change occurred at t= 5 s. Then, the sampling period is given by

$$T_s \quad \frac{50}{10} \quad 5\text{ s}$$

In this case the MV change is equal to 0.05 kg/s. Table 16.3 lists the step response coefficients for this system along with the values used to calculate them. Note that the units for a_i are ºC-s/kg.

Table 16.3

SRM Coefficients for Example 16.1 (u=0.05)

i	t_i	y_i	y_i	a_i
0	5	50.00	0.0	0.0
1	10	49.77	-0.23	-4.68
2	15	49.35	-0.65	-13.0
3	20	49.08	-0.92	-18.4
4	25	48.94	-1.06	-21.2
5	30	48.87	-1.13	-22.6
6	35	48.84	-1.16	-23.2
7	40	48.83	-1.17	-23.5
8	45	48.82	-1.18	-23.7
9	50	48.82	-1.18	-23.7
10	55	48.81	-1.19	-23.8

Example 16.2 An Average SRM for the Thermal Mixing Process

Problem Statement. Using the VBS (Section 3.9) for the thermal mixer (Example 3.3), develop SRMs for this system based on +10%, +5%, -5% and -10% step MV changes. Average the resulting SRM coefficients and compare the predicted responses with +10% and -10% step MV changes.

Solution. Using the simulator provided with this text, open-loop step tests were conducted and the step response coefficients for each step test were generated (Table 16.4). Finally, the SRM coefficients were averaged (Table 16.4). Note that due to the nonlinearity of the process, negative MV changes have a larger gain (i.e., a 10% difference in steady-state gain between the -10% and +10% MV change). Figure 16.2.4 compares the open-loop response of the process simulator with the SRM prediction based on the average model coefficients (Table 16.4) for a 10% increase in the MV while Figure 16.2.5 compares the two for a 10% decrease in the MV. If the SRM had been based on based only on the 10% increase, the SRM would have very closely matched the results for the +10% MV change (Figure 16.2.4), but would have had a much larger error for the -10% MV change (Figure 16.2.5). By averaging the coefficients of the SRMs, the maximum error between the SRM prediction and the process is reduced. For the industrial application of MPC, the SRM are identified for a number of step input changes based on an approach that, in effect, averages all the step test results.

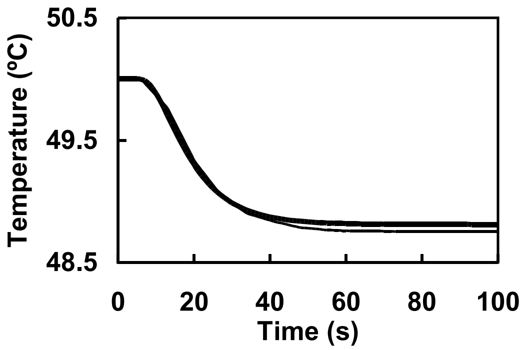


Figure 16.2.4 Comparison between thermal mixer simulator (thick line) and the SRM (thin line) for a 10% increase in the MV.

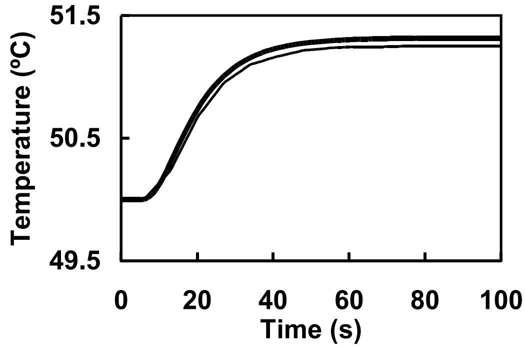


Figure 16.2.5 Comparison between thermal mixer simulator (thick line) and the SRM (thin line) for a 10% decrease in the MV.

Table 16.4
SRM Coefficients for the Step Tests Applied to the Thermal Mixing Process.

	+10%	+5%	-5%	-10%	Average
***u* (kg/s)**	0.05	0.025	-0.025	-0.05	
a_1	-4.68	-4.70	-4.73	-4.75	-4.72
a_2	-13.0	-13.2	-13.4	-13.5	-13.3
a_3	-18.4	-18.7	-19.2	-19.5	-19.0
a_4	-21.2	-21.6	-22.4	-22.8	-22.0
a_5	-22.6	-23.0	-24.0	-24.5	-23.3
a_6	-23.2	-23.7	-24.8	-25.8	-24.3
a_7	-23.5	-24.1	-25.2	-25.8	-24.7
a_8	-23.7	-24.2	-25.4	-26.1	-24.9
a_9	-23.7	-24.3	-25.5	-26.2	-24.9
a_{10}	-23.8	-24.4	-25.6	-26.3	-25.0

16.3 The Dynamic Matrix

The definition of a SRM is based on a single input change, but feedback control involves a series of input changes. This section will demonstrate that SRM models can be used to predict the response of a process to a series of input changes using the **Principle of Superposition, which assumes that the process is linear**. Figure 16.3.1 shows the combined effect of two input changes (u_1 or u_2). First, these figures show the response of the process to each input change as if only one input change were implemented on the process [i.e., $y($ $u_1)$ and $y($ $u_2)$] assuming that the system is at steady state before the input changes are applied. To obtain the combined effect of

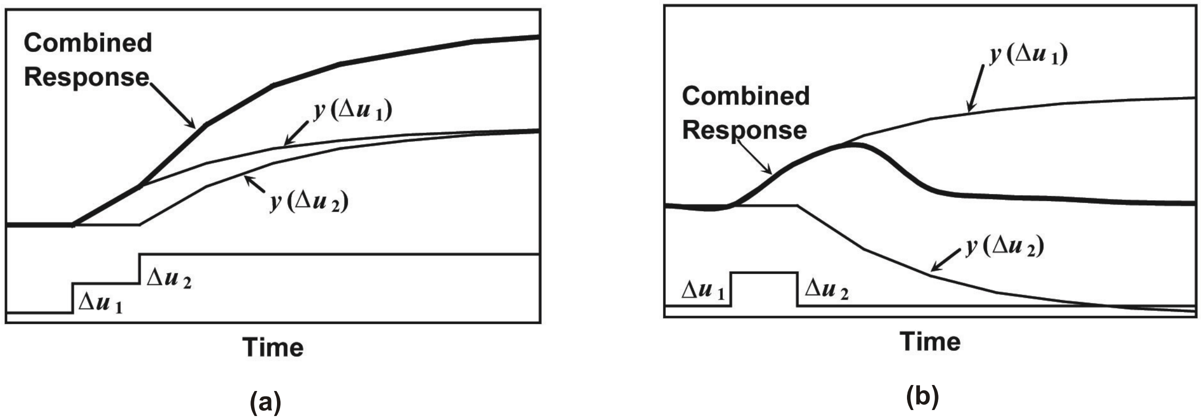


Figure 16.3.1 Application of superposition for two separate step changes in the MV. (a) additive step changes (b) step changes in the opposite direction.

both input changes, the responses are simply added together. That is, adding each response to get the combined response is an application of the Principle of Superposition, which assumes that the process responds to inputs in the same fashion regardless of the previous inputs. Because each of the individual responses [i.e., $y(\Delta u_1)$ and $y(\Delta u_2)$] can be determined directly from a SRM and the magnitude of the input changes (Δu), the combined response can also be determined from the SRM of the process if the timing of the input changes is taken into account. Applying Equation 16.2.2 for the effect of Δu_1 up to $t = t_{n_p}$, assuming $n_p > m$ where n_p is the **prediction horizon**, which is equal to the number of time steps into the future that predictions for $y(t_i)$ are made, yields

$$
\begin{aligned}
y(t_1) &= y(t_0) + a_1 \Delta u_1 \\
y(t_2) &= y(t_0) + a_2 \Delta u_1 \\
y(t_3) &= y(t_0) + a_3 \Delta u_1 \\
&\;\;\vdots \\
y(t_{n_p}) &= y(t_0) + a_m \Delta u_1
\end{aligned}
$$

which represents values of $y(t_i)$ at discrete times into the future as a result of the input change, Δu_1. Likewise, applying Equation 16.2.2 for the effect of Δu_2 yields

$$
\begin{aligned}
y(t_1) &= y(t_0) + 0 \\
y(t_2) &= y(t_0) + a_1 \Delta u_2 \\
y(t_3) &= y(t_0) + a_2 \Delta u_2 \\
&\;\;\vdots \\
y(t_{n_p}) &= y(t_0) + a_m \Delta u_2
\end{aligned}
$$

recognizing that u_2 is first applied one time interval after u_1 was applied. To obtain the combined effect of a series of two input changes, the last two equations are added together.

$$
\begin{array}{llllll}
y(t_1) & y(t_0) & a_1 & u_1 & & \\
y(t_2) & y(t_0) & a_2 & u_1 & a_1 & u_2 \\
y(t_3) & y(t_0) & a_3 & u_1 & a_2 & u_2 \\
 & & \cdot & & & \\
 & & \cdot & & & \\
 & & \cdot & & & \\
y(t_{n_p}) & y(t_0) & a_n & u_1 & a_{n\ 1} & u_2
\end{array}
$$

On the right-hand side of this equation, the first column represents the effect of the u_1 on $y(t)$ and the second column represents the effect of u_2.

Now consider the effect of a series of four input changes $u(t_i)$, where u changes each sampling time (T_s) and where i varies from 0 to 3. In addition assume that the SRM for this case has m coefficients. By assuming that the system is initially at steady-state and adding the effect of each input change [$u(t_i)$] on $y(t)$, the following set of equations result for the prediction for n_p time steps into the future.

$$
\begin{array}{lllllll}
y(t_1) & y(t_0) & y(t_1) & a_1\ u(t_0) & & & \\
y(t_2) & y(t_0) & y(t_2) & a_2\ u(t_0) & a_1\ u(t_1) & & \\
y(t_3) & y(t_0) & y(t_3) & a_3\ u(t_0) & a_2\ u(t_1) & a_1\ u(t_2) & \\
y(t_4) & y(t_0) & y(t_4) & a_4\ u(t_0) & a_3\ u(t_1) & a_2\ u(t_2) & a_1\ u(t_3) \\
y(t_5) & y(t_0) & y(t_5) & a_5\ u(t_0) & a_4\ u(t_1) & a_3\ u(t_2) & a_2\ u(t_3) \\
y(t_6) & y(t_0) & y(t_6) & a_6\ u(t_0) & a_5\ u(t_1) & a_4\ u(t_2) & a_3\ u(t_3) \\
y(t_7) & y(t_0) & y(t_7) & a_7\ u(t_0) & a_6\ u(t_1) & a_5\ u(t_2) & a_4\ u(t_3) \\
 & & \cdot & & & & \\
 & & \cdot & & & & \\
 & & \cdot & & & & \\
y(t_{n_p}) & y(t_0) & y(t_{n_p}) & a_m\ u(t_0) & a_m\ u(t_1) & a_m\ u(t_2) & a_m\ u(t_3)
\end{array}
\qquad \textbf{16.3.1}
$$

Note that for the right-hand side of Equation 16.3.1, the first column represents the response of the process to $u(t_0)$, the second column represents the response of the process to $u(t_1)$, the third column represents the response to $u(t_2)$ and so on. Also, note that because $u(t_1)$ is not applied until t=t_1, it does not start to affect the response of the process until t=t_2. Likewise, because $u(t_2)$ is not applied until t=t_2, it does not start to affect the response of the process until t=t_3. For this reason, the columns are offset by one sample interval as each column is added to the right-hand side of Equation 16.3.1. Putting Equation 16.3.1 into matrix form yields

$$
\begin{array}{c}
y(t_1) \\ y(t_2) \\ y(t_3) \\ y(t_4) \\ y(t_5) \\ y(t_6) \\ y(t_7) \\ \vdots \\ y(t_{n_p})
\end{array}
\qquad
\begin{array}{cccc}
a_1 & 0 & 0 & 0 \\
a_2 & a_1 & 0 & 0 \\
a_3 & a_2 & a_1 & 0 \\
a_4 & a_3 & a_2 & a_1 \\
a_5 & a_4 & a_3 & a_2 \\
a_6 & a_5 & a_4 & a_3 \\
a_7 & a_6 & a_5 & a_4 \\
 & & \vdots & \\
a_m & a_m & a_m & a_m
\end{array}
\qquad
\begin{array}{c}
u(t_o) \\ u(t_1) \\ u(t_2) \\ u(t_3)
\end{array}
\qquad \mathbf{16.3.2}
$$

For this case there are four input changes and the prediction is n_p time steps ahead. Therefore, the coefficient matrix is dimensioned $n_p \times 4$.

Now consider the case in which there is a general number of input changes. Once again assume that the system is initially at steady state and that the SRM for this system has m coefficients. Also, assume that the number of input changes is equal to n_c where $n_c<m$ and that there are n_p prediction steps into the future where $n_p>m$. n_c, which is the **control horizon**, represents the number of control moves into the future that are considered. The application of the Principal of Superposition to this case results in the following set of equations.

$$
\begin{array}{llllllll}
y(t_1) & a_1\ u(t_0) & & & & & \\
y(t_2) & a_2\ u(t_0) & a_1\ u(t_1) & & & & \\
y(t_3) & a_3\ u(t_0) & a_2\ u(t_1) & a_1\ u(t_2) & & & \\
y(t_4) & a_4\ u(t_0) & a_3\ u(t_1) & a_2\ u(t_2) & a_1\ u(t_3) & & \\
y(t_5) & a_5\ u(t_0) & a_4\ u(t_1) & a_3\ u(t_2) & a_2\ u(t_3) & a_1\ u(t_4) & \\
y(t_6) & a_6\ u(t_0) & a_5\ u(t_1) & a_4\ u(t_2) & a_3\ u(t_3) & a_2\ u(t_4) & \ldots \\
 & & & \vdots & & & \\
y(t_{n_p}) & a_m\ u(t_0) & a_m\ u(t_1) & a_m\ u(t_2) & a_m\ u(t_3) & a_m\ u(t_4) & \ldots
\end{array}
$$

Putting this equation into matrix form similar to Equation 16.3.2 yields

$$
\begin{array}{c}
y(t_1) \\ y(t_2) \\ y(t_3) \\ y(t_4) \\ y(t_5) \\ \cdot \\ \cdot \\ \cdot \\ y(t_{n_p})
\end{array}
\qquad
\begin{array}{cccccccc}
a_1 & 0 & 0 & 0 & 0 & 0 & \ldots & 0 \\
a_2 & a_1 & 0 & 0 & 0 & 0 & \ldots & 0 \\
a_3 & a_2 & a_1 & 0 & 0 & 0 & \ldots & 0 \\
a_4 & a_3 & a_2 & a_1 & 0 & 0 & \ldots & 0 \\
a_5 & a_4 & a_3 & a_2 & a_1 & 0 & \ldots & 0 \\
 & & \cdot & & & & & \\
 & & \cdot & & & & & \\
 & & \cdot & & & & & \\
a_m & a_m & a_m & a_m & a_m & a_m & \ldots & a_m
\end{array}
\qquad
\begin{array}{c}
u(t_o) \\ u(t_1) \\ u(t_2) \\ u(t_3) \\ u(t_4) \\ \cdot \\ \cdot \\ \cdot \\ u(t_{n_c})
\end{array}
\qquad 16.3.3
$$

Note that in this case, the number of predictions exceeds the number of coefficients in the SRM. For this case, $a_{m+i}=a_m$ because the process response does not change after the time to steady state is exceeded. The coefficient matrix ($\underline{\underline{A}}$) for Equation 16.3.3 is a lower diagonal matrix and is given by

$$
\underline{\underline{A}} \qquad
\begin{array}{cccccccc}
a_1 & 0 & 0 & 0 & 0 & 0 & \ldots & 0 \\
a_2 & a_1 & 0 & 0 & 0 & 0 & \ldots & 0 \\
a_3 & a_2 & a_1 & 0 & 0 & 0 & \ldots & 0 \\
a_4 & a_3 & a_2 & a_1 & 0 & 0 & \ldots & 0 \\
a_5 & a_4 & a_3 & a_2 & a_1 & 0 & \ldots & 0 \\
 & & & & \cdot & & & \\
 & & & & \cdot & & & \\
 & & & & \cdot & & & \\
a_m & a_m & a_m & a_m & a_m & & \ldots & a_m
\end{array}
\qquad 16.3.4
$$

$\underline{\underline{A}}$ is known as the **dynamic matrix** and is used directly to implement dynamic matrix control. Note that the dynamic matrix is constructed entirely from the coefficients of the SRM: each column contains the SRM coefficients in a staggered fashion. From Equation 16.3.3, knowing the SRM for a process, allows you to directly calculate the process response for a series of input changes using the dynamic matrix. The dimensions of $\underline{\underline{A}}$ are n_p by n_c where n_c is the number of input changes (**control horizon**) and n_p is the number of model predictions (**prediction horizon**). The amount of time into the future that the predictions are made is equal to $n_p \;\; T_s$. m, which is the number of coefficients in the SRM model, is also known as the **model horizon** because the response time (i.e., T_{ss}) based on the model is equal to $m \;\; T_s$.

Equation 16.3.3 can also be written in matrix notation as

$$\underline{y} \quad \underline{\underline{A}}\,\underline{u} \qquad 16.3.5$$

The value of Equation 16.3.5 stems from the fact that it can be used to calculate the dynamic behavior of y for a series of input changes assuming that the process is initially at steady state. Remember that Equation 16.3.5 is a linear model because it is based on linear SRM models.

Example 16.3 Development of a Dynamic Matrix from a SRM

Problem Statement. Develop the dynamic matrix corresponding to the following SRM (a_1=0.1, a_2=0.5, a_3=0.9 and a_4=1.0). Assume that $n_c=m$ and that $n_p=2m$.

Solution. From the SRM provided for this problem, m is equal to 4. Using Equation 16.3.4 and recognizing that $\underline{\underline{A}}$ is a (8×4) matrix,

$$\underline{\underline{A}} \quad \begin{array}{cccc} 0.1 & 0 & 0 & 0 \\ 0.5 & 0.1 & 0 & 0 \\ 0.9 & 0.5 & 0.1 & 0 \\ 1.0 & 0.9 & 0.5 & 0.1 \\ 1.0 & 1.0 & 0.9 & 0.5 \\ 1.0 & 1.0 & 1.0 & 0.9 \\ 1.0 & 1.0 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 & 1.0 \end{array}$$

Note that this dynamic matrix considers the effect of the input changes on the behavior of the output for 8 sample intervals into the future assuming that the system is initially at steady state.

Example 16.4 SRM Identification and Prediction

Problem Statement. For the step response shown in Figure 16.3.2, which is based on a 10% increase in heat duty for the heat exchanger for the endothermic CSTR (Example 3.6), predict the CV response for the input sequence shown in Figure 16.3.3. Assume that the CV is initially at steady state at 350 K.

Solution. First, the step test results should be used to identify the SRM parameters, a_i. The coefficients of the SRM are obtained by applying Equation 16.2.1 to the step response results. The coefficients of the SRM are listed in Table 16.5.

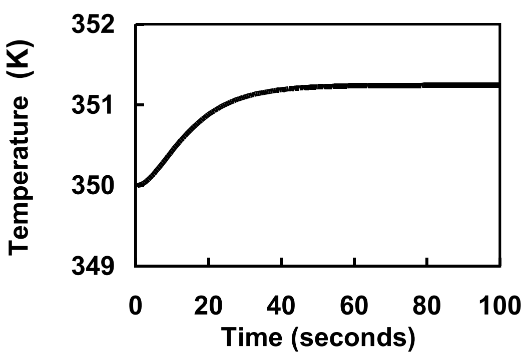


Figure 16.3.2 Open-loop response of the outlet CSTR temperature for a 10% increase in the heat duty.

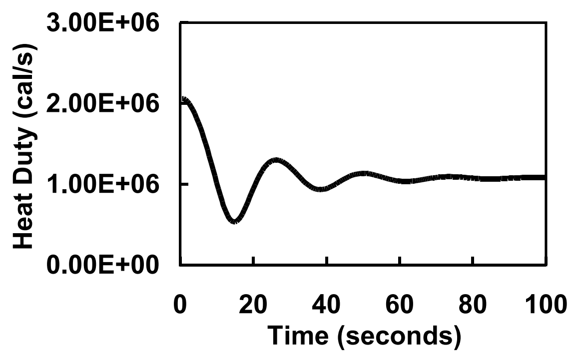


Figure 16.3.3 The input time history for Example 16.4.

Next, the input sequence is approximated by a series of step changes and the results are listed as $u(t_i)$ in Table 16.5. Then, Equation 16.3.3 is applied to predict the time behavior of the reactor temperature. The results are also listed in Table 16.5.

Table 16.5
Results for Example 16.4

Time (sec)	i	a_i	$u(t_i)$	$y(t_i)$
0	0	0	$1.07\ \ 10^{6}$	350.00
5	1	$2.07\ \ 10^{-6}$	$-7.45\ \ 10^{6}$	352.21
10	2	6.10	- 4.87	354.98
15	3	9.79	$4.09\ \ 10^{5}$	354.92
20	4	$1.26\ \ 10^{-5}$	3.43	354.41
25	5	$1.45\ \ 10^{-5}$	-7.37	354.56
30	6	1.57	-2.21	355.80
35	7	1.65	- 4.54	356.50
40	8	1.69	$1.23\ \ 10^{5}$	356.23
45	9	1.72	$6.55\ \ 10^{4}$	356.06
50	10	1.75	$-4.33\ \ 10^{4}$	356.40
55	11	1.76	-5.30	356.57
60	12	1.77	$9.30\ \ 10^{3}$	356.61
65	13	1.77	$3.61\ \ 10^{4}$	356.51
70	14	1.77	$8.04\ \ 10^{3}$	356.53
75	15	$1.77\ \ 10^{-5}$	1.10	356.64

16.4 Moving Horizon Controller

Figure 16.4.1 illustrates the key features of a moving horizon control algorithm. A moving horizon controller calculates the future control moves based on the predicted behavior of the process, but only implements the first calculated control move. Note that all the previous CV and MV values are fixed and known. The MV moves and the resulting CV values into the future remain unknown at this point. The moving horizon controller chooses the future MV values to regulate the CV to its setpoint using the SRM and the previous inputs, recognizing that the previous input will have an impact on the future values of the CV.

After one control interval has expired, a new CV value as well as the last change in MV value is now available. Once again, the controller recalculates the sequence of MV values into the future to meet the control objective using this new information. Although a sequence of control moves is calculated at each control interval, only the

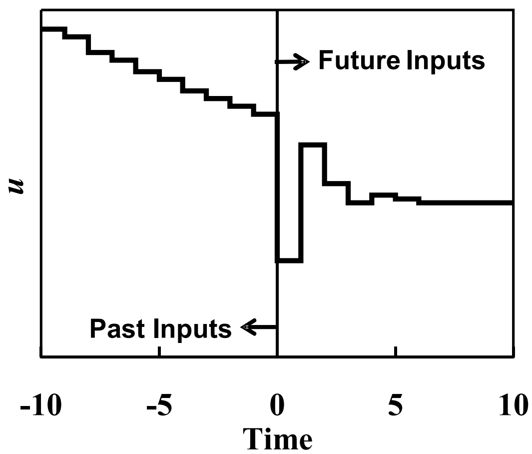


Figure 16.4.1a A plot of the time behavior of the MV for a moving horizon controller.

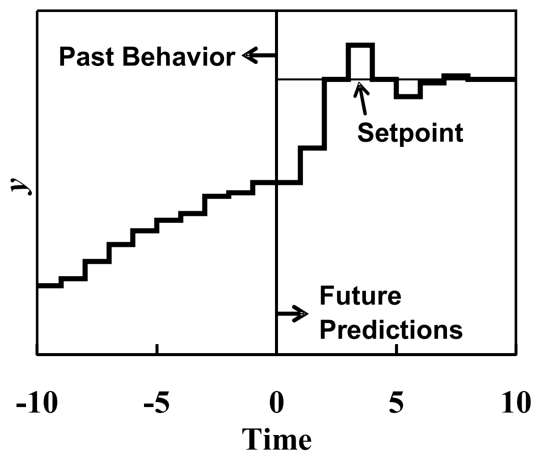


Figure 16.4.1b A plot of the time behavior of the CV for a moving horizon controller.

first control move is implemented on the process. In this manner, even though the complete sequence of control moves into the future is calculated at each control interval, only the first move is actually implemented before a new sequence of inputs is determined. The key feature of this approach is that, at each control interval, a sequence of control moves into the future, as well as the previous input sequence, is considered when determining the next change in the MV value. If the SRM model were perfect and there were no unmeasured disturbances, the first sequence of control moves after a setpoint change would, in fact, be implemented at each subsequent control interval. That is, after the first move, the next control move calculated by the moving horizon controller based on future predictions would be equal to the second control move calculated after the setpoint change. Likewise, the third control move would be equal to the third control move calculated after the setpoint change, and so on.

16.5 Prediction Vector

Up until this point, all the modeling that we have performed has assumed that $y(t_0)$ is at steady state and that $y(t)$ is only affected by MV changes that are made only for $t \geq t_0$. For a control application, this assumption is not realistic because MV changes at $t \leq t_0$ are likely to exist (e.g., Figure 16.4.1a). As a result, the effect of the previous input changes [$\Delta u(t)$ for $t < t_0$] must be taken into account to properly model the future behavior of the CV [$y(t)$ for $t > t_0$].

The **prediction vector, $\boldsymbol{y_P}$**, contains the values of $y(t)$ at discrete points in time for $t > t_0$ if no future MV changes are made [i.e., $\Delta u(t) = 0$ for $t \geq t_0$]. The prediction vector contains the effects of previous MV changes on future CV values. Because these input changes are fixed and known, the prediction vector is constant and can be determined using the SRM.

Assume that the process has a model horizon, m. That is, after m time steps, an input change has had its total steady-state effect on the process. For the SRM listed in Table 16.1, m is equal to 7; therefore, a_i is constant for $i \geq 7$. Applying Equation 16.3.3 to calculate the prediction vector at $t = t_1$ based on the previous inputs [i.e., $\Delta u(t_{-m})$, $\Delta u(t_{-m+1})$, $\ldots$ $\Delta u(t_{-1})$] and assuming that the process is at steady state at $t = t_{-m}$ results in

$$y_P(t_1) \quad y(t_{\ m}) \quad a_{m\ 1} \quad u(t_{\ m}) \quad a_m \quad u(t_{\ m\ 1}) \quad a_{m\ 1} \quad u(t_{\ m\ 2})$$
$$\ldots \quad a_3 \quad u(t_{\ 2}) \quad a_2 \quad u(t_{\ 1}) \quad a_1 \quad u(t_0)$$

where the negative subscripts indicate the number of sampling intervals before t_0. This formulation assumes that the process is at steady-state at t=t_{-m}; therefore, $y(t_{\ m})$ is added to the effects of the input changes to determine the future values of $y(t_i)$ resulting from the previous m input changes. Note that the coefficients of $u(t_{\ m})$ and $u(t_{\ m\ 1})$ are both a_m because $a_{m+1} = a_m$. $u(t_0)$ is zero for the prediction vector; therefore,

$$y_P(t_1) \quad y(t_{\ m}) \quad a_m \quad u(t_{\ m}) \quad a_m \quad u(t_{\ m\ 1}) \quad a_{m\ 1} \quad u(t_{\ m\ 2})$$
$$\ldots \quad a_3 \quad u(t_{\ 2}) \quad a_2 \quad u(t_{\ 1})$$

Similarly, $y_P(t_2)$ is given by

$$y_P(t_2) \quad y(t_{\ m}) \quad a_m \quad u(t_{\ m}) \quad a_m \quad u(t_{\ m\ 1}) \quad a_m \quad u(t_{\ m\ 2}) \quad a_{m\ 1} \quad u(t_{\ m\ 3})$$
$$\ldots \quad a_4 \quad u(t_{\ 2}) \quad a_3 \quad u(t_{\ 1})$$

In this manner, $y_P(t_n)$ is given by

$$y_P(t_n) \quad y(t_{\ m}) \quad a_m \quad u(t_{\ m}) \quad a_m \quad u(t_{\ m\ 1}) \quad \ldots \quad a_m \quad u(t_{\ 2}) \quad a_m \quad u(t_{\ 1})$$

where n_p is the number of T_s moves into the future that are modeled with $n_p > m$. The prediction vector, $\underline{\boldsymbol{y}}^P$, can be expressed in matrix form as

$$\begin{array}{c} y_P(t_1) \\ y_P(t_2) \\ \cdot \\ \cdot \\ \cdot \\ y_P(t_{n_p}) \end{array} \qquad \begin{array}{c} y(t_{\ m}) \\ y(t_{\ m}) \\ \cdot \\ \cdot \\ \cdot \\ y(t_{\ m}) \end{array} \qquad \begin{array}{cccccccc} a_m & a_m & a_{m\ 1} & a_{m\ 2} & \ldots & a_3 & a_2 \\ a_m & a_m & a_m & a_{m\ 1} & \ldots & a_4 & a_3 \\ & & & \cdot & & & \\ & & & \cdot & & & \\ & & & \cdot & & & \\ a_m & a_m & a_m & a_m & \ldots & a_m & a_m \end{array} \qquad \begin{array}{c} u(t_{\ m}) \\ u(t_{\ m\ 1}) \\ \cdot \\ \cdot \\ \cdot \\ u(t_{\ 1}) \end{array}$$

or

$$\underline{\boldsymbol{y}}_P \quad y(t_{\ m}) \quad \underline{\underline{\boldsymbol{A}}}_P \underline{\boldsymbol{u}}_P \qquad \textbf{16.5.1}$$

where

$$\underline{\underline{\boldsymbol{A}}}_P \qquad \begin{array}{cccccccc} a_m & a_m & a_{m\ 1} & a_{m\ 2} & \ldots & a_3 & a_2 \\ a_m & a_m & a_m & a_{m\ 1} & \ldots & a_4 & a_3 \\ & & \cdot & & & & \\ & & \cdot & & & & \\ & & \cdot & & & & \\ a_m & a_m & a_m & a_m & \ldots & a_m & a_m \end{array} \qquad \underline{\boldsymbol{u}}_P^T \quad u(t_{\ m}),\ u(t_{\ m\ 1}),\ \ldots\ ,\ u(t_{\ 1})$$

and where $y(t_{-k})$ is the value of the CV at $t \quad t_o \quad k \quad T_s$ and $\underline{\underline{A}}_P$ is dimensioned ($n_p \times m$).

Example 16.5 Coefficient Matrix for the Prediction Vector

Problem Statement. Based on the SRM given in Table 16.1, determine the coefficient matrix for the prediction vector ($\underline{\underline{A}}_P$) assuming that 8 predictions into the future are used.

Solution. In this case, because m is equal to 7, $\underline{\underline{A}}_P$ is dimensioned (8×7). Applying Equation 16.5.1 for this case yields

$$\underline{\underline{A}}_P \quad \begin{matrix} 1.0 & 1.0 & 0.99 & 0.98 & 0.95 & 0.87 & 0.63 \\ 1.0 & 1.0 & 1.0 & 0.99 & 0.98 & 0.95 & 0.87 \\ 1.0 & 1.0 & 1.0 & 1.0 & 0.99 & 0.98 & 0.95 \\ 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 0.99 & 0.98 \\ 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 0.99 \\ 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 \end{matrix}$$

16.6 DMC Controller

In this section, we will derive the DMC control law by combining the prediction vector (Equation 16.5.1) with the prediction of the response of the process to future moves (Equation 16.3.5). The values of $y(t)$ for $t \quad t_o$ can be calculated by combining the prediction vector with the effects of future control moves (Equation 16.3.5)

$$\underline{y} \qquad \underline{y}_P \quad \underline{\underline{A}}\,\underline{u} \tag{16.6.1}$$

Equation 16.6.1 is subject to a number of factors that undermine its accuracy: (1) errors in identifying the coefficients of the discrete-time SRM, (2) unmeasured disturbances in the past or in the future, (3) nonlinear behavior and (4) non-steady-state behavior at $t = t_{-m}$. If Equation 16.6.1 is used for control, offset results due to these sources of process/model mismatch.

Equation 16.5.1 is used to calculate $y_P(t_0)$ by

$$y_P(t_o) \quad y(t_{\ m}) \quad a_m \quad u(t_{\ m}) \quad a_{m\ 1} \quad u(t_{\ m\ 1}) \quad \ldots \quad a_2 \quad u(t_{\ 2}) \quad a_1 \quad u(t_{\ 1})$$

The error between the measured value of $y(t_0)$ and the predicted value, $y_P(t_o)$, can be used to adjust Equation 16.6.1 to make it more accurate. The error between the measured [$y_s(t_0)$]and predicted value [$y_P(t_0)$] is given as

$$y_s(t_0) \quad y_P(t_0)$$

Then, adding this correction to Equation 16.6.1 yields

$$\underline{y} \qquad \underline{y}_P \qquad \underline{\underline{A}}\,\underline{u} \qquad \underline{\ }^T \tag{16.6.2}$$

where $$\underline{\ } \quad [\qquad \mathrm{K} \quad]$$

so that the predicted value of $y(t)$ agrees with the latest measured value of the CV [$y_s(t_0)$]. **Adding the error to Equation 16.6.2 eliminates offset for a DMC controller.**

The DMC control law is based on minimizing an objective function, which is based on the error from setpoint and the size of MV changes, i.e., $u(t)$. The objective function, , is the sum of the square of the errors from setpoint and the sum of the squares of the input changes for the prediction horizon (i.e., n_p steps into the future).

$$\sum_{i\ 1}^{n_p} [y_{sp} \quad y(t_i)]^2 \quad Q[\ u(t_i)]^2 \tag{16.6.3}$$

where Q is the move suppression factor, which is the primary on-line tuning parameter for a DMC controller. When a small value of Q is used, tight control to setpoint is selected, which will result in relatively large changes in the MV. When a large value is Q is used, more gradual changes in the MV results, causing less aggressive control to setpoint.

Equation 16.6.2 shows that $y(t_i)$ is made up of three parts: the prediction vector (i.e., the effect of past inputs), the effects of future inputs and the process/model mismatch correction term, . Only the effect of future moves can be changed by the controller; therefore, lumping y_{sp}, $y_P(t_i)$ and into a single term, $E(t_i)$, yields

$$E(t_i) \quad y_{sp} \quad y_P(t_i) \tag{16.6.4}$$

Then Equation 16.6.3 becomes

$$\sum_{i\ 1}^{n_p} [E(t_i) \quad y_c(t_i)]^2 \quad Q[\ u(t_i)]^2 \tag{16.6.5}$$

where $$\underline{y}_c \quad \underline{\underline{A}}\,\underline{u}$$

The objective of the DMC controller is to choose the control moves, $u(t_i)$ for n_c moves into the future such that is minimized over the prediction horizon (n_p).

Perfect control (i.e., = 0 when Q=0), which is based upon assuming that $y_c(t_i)$ is the mirror image of $E(t_i)$, is given by

$$\underline{u} \quad \underline{\underline{A}}^{\ 1}\,\underline{E} \tag{16.6.6}$$

But, because it assumes that $y_c(t_i)$ can be moved instantaneously, this result is not realistic. Instead, we can choose the set of control moves that minimizes the sum of the squares of the errors from setpoint (i.e., Q=0). This solution can be obtained analytically by differentiating Equation 16.6.5 with respect to u and setting the result equal to zero,

$$\frac{\quad}{\underline{\;u}} \quad \underline{\underline{A}}^T(\underline{\underline{A}}\,\underline{\;u} \quad \underline{E}) \quad \mathbf{0} \tag{16.6.7}$$

Solving for $\underline{\;\boldsymbol{u}}$ gives

$$\underline{\;u} \quad (\underline{\underline{A}}^T \underline{\underline{A}})^{\;1} \underline{\underline{A}}^T \underline{E} \tag{16.6.8}$$

which is the control law for a DMC controller assuming that Q is equal to zero, i.e., no penalty for large changes in the MV. Note that $(\underline{\underline{A}}^T \underline{\underline{A}})^{\;1} \underline{\underline{A}}^T$ is equal to $\underline{\underline{A}}^{\;1}$ when $\underline{\underline{A}}$ is a square matrix. Therefore, once the SRM and the prediction vector are calculated, the DMC controller can be formulated directly using Equation 16.6.8.

Equation 16.6.8 results in very aggressive control because it is based on minimizing the deviation from setpoint without regard to the changes in the MV levels. That is, if Equation 16.6.8 is applied, excessively sharp changes in u result, which is not desirable operationally. Normal levels of process/model mismatch, combined with the aggressive nature of Equation 16.6.8, can easily yield unstable control performance. In addition, $(\underline{\underline{A}}^T \underline{\underline{A}})^{\;1}$ can be ill-conditioned due to process/model mismatch and deadtime in the process model. These problems can be overcome by using the full form of Equation 16.6.5 (i.e., $Q \quad 0$). Differentiating Equation 16.6.5 with respect to $\underline{\;\boldsymbol{u}}$ and setting the result equal to zero yields

$$\underline{\;u} \quad (\underline{\underline{A}}^T \underline{\underline{A}} \quad \underline{\underline{Q}}^2)^{\;1} \underline{\underline{A}}^T \underline{E} \tag{16.6.9}$$

where $\underline{\underline{Q}}$ is the move suppression matrix and is a diagonal matrix with positive elements, i.e.,

$$\underline{\underline{Q}} \quad \begin{matrix} Q & 0 & 0\ldots0 \\ 0 & Q & 0\ldots0 \\ 0 & 0 & Q\ldots0 \\ & & . \\ & & . \\ & & . \\ 0 & 0 & 0\ldots Q \end{matrix}$$

The following conclusions can be drawn from Equation 16.6.9

- For a specified value of Q, $(\underline{\underline{A}}^T \underline{\underline{A}} \quad \underline{\underline{Q}}^2)^{\;1} \underline{\underline{A}}^T$ is a constant matrix, which is explicitly determined from the coefficients of the SRM.

- The vector $\underline{\boldsymbol{E}}$, which takes into account the setpoint, the effect of previous inputs and the correction for mismatch, is updated after each control implementation.
- A full set of control moves is calculated at each control interval although only the first move is implemented.

DMC Control Design Guidelines. The implementation of a DMC controller requires the values of T_s, m, n_c, n_p and Q. The value of Q, which depends on the fidelity of the SRM model and the tuning criterion selected, is determined on-line under closed-loop conditions by trial-and-error tuning. On the other hand, the values of T_s, m, n_c and n_p can be set based on the following guidelines. First, assume that the time to steady state, T_{ss}, is known. Next, determine how frequently new information on the CV is available. For example, for an on-line GC, analyzer updates are available every 3-10 min. Therefore, for this case the fastest that the DMC controller should be executed is when a new analyzer update is available which generally sets T_s. Then, $m \quad T_{ss} \, / \, T_s$. On the other hand, continuous measurements are update approximately every 0.2 s. In this case, if the previous formula were used, m would be too large for practical use. Therefore, set m, such that 60<m<500 (note that the value 500 is based on current computer processing speeds) and calculate T_s, i.e., $T_s \quad T_{ss} \, / \, m$. Once m is determined, n_c and n_p can be calculated by the following guidelines.

$$
\begin{aligned}
n_c &\quad 0.5\,m \\
n_p &\quad 1.5\,m
\end{aligned}
\qquad \textbf{16.6.10}
$$

Example 16.6 Design of a DMC Temperature Controller

Problem Statement. Design a DMC controller for a temperature control loop that has a time to steady state equal to 20 min.

Solution. Because a temperature sensor provides continuous measurements, m is selected first. In this case, set m equal to 120. The sample interval, T_s, is equal to T_{ss}/m or 10 s. Therefore, using Equation 16.6.10, n_c is equal to 60 and n_p is equal to 180. As a result, the dynamic matrix for this case would be dimensioned 180×60.

Example 16.7 Design of a DMC Composition Controller

Problem Statement. Consider a SISO DMC controller applied to the overhead of a C_3 splitter (propylene/propane splitter). Assume that the time to steady-state is 7 h and that the overhead composition analyzer updates every 10 min. Determine the design specifications for this DMC controller.

Solution. Because new information on the CV is available only when the analyzer updates, T_s is set equal to the analyzer deadtime or 10 min. Then, m=(7 h×60 min/h)/10 min=42. From Equation 16.6.10, n_c=21 and n_p=63. As a result, the dynamic matrix for this case would be a 63×21 matrix.

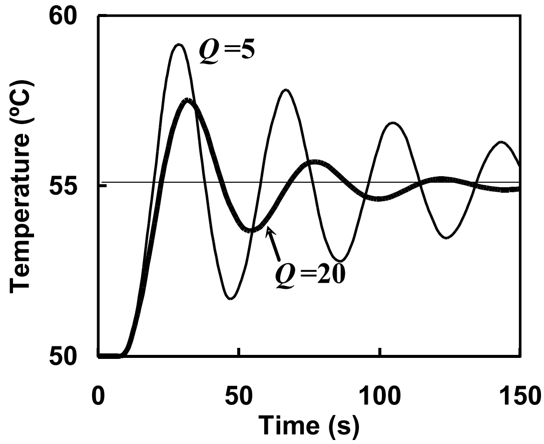


Figure 16.6.1a Closed-loop DMC results for the thermal mixing process for Q equal to 5 and 20.

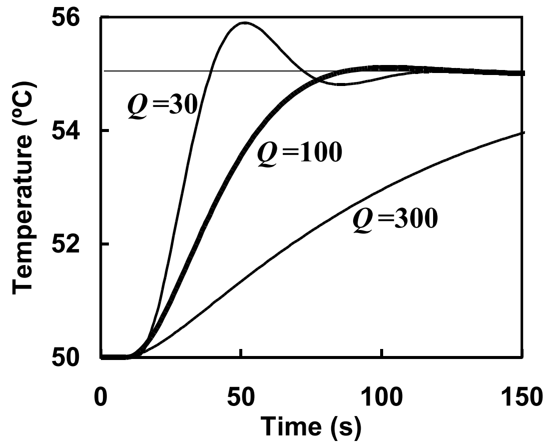


Figure 16.6.1b Closed-loop DMC results for the thermal mixing process for Q equal to 30, 100 and 300.

Example 16.8 Tuning a DMC Controller applied to the Thermal Mixing Process

Problem Statement. Evaluate the effect of the move suppression factor (Q) on the setpoint tracking performance for a 10% increase in the setpoint for a SISO DMC controller applied to the thermal mixing process (Example 3.3). Use a SRM with 10 model coefficients.

Solution. Using the average values for the SRM coefficients listed in Table 16.4, results for Q equal to 300, 100, 30, 20 and 10 are shown in Figure 16.6.1. The results for Q equal to 300 are severely overdamped and results in very sluggish control performance. For Q equal to 100, the closed-loop response is very nearly critically damped while for Q equal to 30 the response has a $1/28$ decay ratio with overshoot equal to 0.2. For Q equal to 20, the decay ratio is 1/10 with an overshoot of 50% while for Q equal to 10, the response is ringing with a decay ratio of 1/1.5 and the overshoot is equal to 0.8. Therefore, as Q is decrease the closed-loop response becomes more aggressive in a fashion similar to a P-only controller as K_c is increased. Of course, a P-only controller will exhibit offset while DMC controller is free of offset because a correction is applied at each control interval to remove mismatch between the predicted value of $y(t_0)$ and the measured value.

16.7 Extension to MIMO Processes

Extending DMC to MIMO processes is relatively straightforward. In the previous sections, it has been assumed that the system involved a single MV and a single CV. For the MIMO case, there are multiple MVs and multiple CVs, but the model of each MV-CV pair is handled as before. For the MIMO case, the SISO dynamic matrices are combined together in partitioned matrices while the SISO $\underline{\boldsymbol{E}}$ vectors are also combined into partitioned vectors. Partitioned vectors and matrices are used to apply DMC to MIMO processes using Equation 16.6.9. For example, the partitioned dynamic matrix, $\underline{\underline{\boldsymbol{A}}}$, is given by

$$\underline{\underline{A}} \quad \begin{matrix} \underline{\underline{A}}_{11} & \underline{\underline{A}}_{12} & \cdots & \underline{\underline{A}}_{1,j} \\ \underline{\underline{A}}_{21} & \underline{\underline{A}}_{22} & \cdots & \underline{\underline{A}}_{2,j} \\ & \cdot & & \\ & \cdot & & \\ & \cdot & & \\ \underline{\underline{A}}_{k,1} & \underline{\underline{A}}_{k,2} & \cdots & \underline{\underline{A}}_{k,j} \end{matrix}$$

where j is the number of MVs and k is the number of CVs. Consider a two-input/two-output process. $\underline{\underline{A}}_{11}$ is the dynamic matrix for y_1 as affected by u_1 and $\underline{\underline{A}}_{12}$ is the dynamic matrix of y_1 as affected by u_2, etc. For illustration purposes, consider

$$\underline{\underline{A}}_{11} \quad \begin{matrix} 1 & 2 \\ 3 & 4 \end{matrix} \qquad \underline{\underline{A}}_{12} \quad \begin{matrix} 5 & 6 \\ 7 & 8 \end{matrix}$$

$$\underline{\underline{A}}_{21} \quad \begin{matrix} 9 & 10 \\ 11 & 12 \end{matrix} \qquad \underline{\underline{A}}_{22} \quad \begin{matrix} 13 & 14 \\ 15 & 16 \end{matrix}$$

Then, the multivariable dynamic matrix is given by

$$\underline{\underline{A}} \quad \begin{matrix} 1 & 2 & 5 & 6 \\ 3 & 4 & 7 & 8 \\ 9 & 10 & 13 & 14 \\ 11 & 12 & 15 & 16 \end{matrix}$$

The MIMO MV vector is given by the following partitioned vector

$$\underline{\boldsymbol{u}} \quad \begin{matrix} \underline{\boldsymbol{u}}_1 \\ \underline{\boldsymbol{u}}_2 \\ \cdot \\ \cdot \\ \cdot \\ \underline{\boldsymbol{u}}_j \end{matrix}$$

For example, if

$$\underline{\boldsymbol{u}}_1 \quad \begin{matrix} 1 \\ 2 \end{matrix} \qquad \underline{\boldsymbol{u}}_2 \quad \begin{matrix} 3 \\ 4 \end{matrix}$$

then

$$\underline{u} \quad \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array}$$

For the application of the DMC control law, the $\underline{\underline{A}}$ matrix must be inverted, and the matrix can be quite large for most industrial DMC controllers. Consider the case with 10 MVs, 10 CVs and a prediction horizon of 100 sampling intervals. Based on the guidelines listed before, m is equal to 67 (i.e., 100/1.5) and n_c=33 (i.e., m/2). Therefore, the MIMO dynamic matrix is dimensioned 1000   330. There have been a large number of industrial DMC applications that are considerably larger than this example.

An additional issue that must be addressed when applying DMC to MIMO processes is how to prioritize the various control objectives. DMC controllers use CV weighting, which allows the user to assign a relative weighting to each of the CVs. A CV weighting matrix, $\underline{\underline{W}}$, is added to the DMC control law (Equation 16.6.9)

$$\underline{u} \quad (\underline{\underline{A}}^T \underline{\underline{W}}^2 \underline{\underline{A}} \quad \underline{\underline{Q}}^2)^{\ 1} \underline{\underline{A}}^T \underline{\underline{W}}^2 \underline{E} \qquad \textbf{16.7.1}$$

where $\underline{\underline{W}}$ is a partitioned matrix, which contains diagonal matrices, $\underline{\underline{W}}_i$. $\underline{\underline{W}}_i$ is a diagonal matrix, which contains the same element, w_i , on its diagonal. w_i is the relative weighting factor for the i-th CV.

16.8 Application of DMC for Constraint Control

One of the key advantages of DMC is that it can be readily applied for constraint control of multivariable processes. In fact, DMC was developed because conventional overrides were inadequate for maintaining the operation of an FCC unit at its optimum operating conditions[4]. For the optimal operation of an FCC unit, the operative constraints change over time. Each time there is a different combination of operative constraints, a different constraint control configuration with different controller tuning is generally required. As a result, for a complex processes like an FCC unit, there are an extremely large number of different constraint control configurations required, which is well beyond what is practical to maintain industrially. On the other hand, because DMC is a multivariable controller that has models for the effects of each MV on each constraint, it is able to automatically switch from controlling one combination of constraints to another.

DMC treats operative constraints as CVs. Because DMC is based on a least squares solution, constraints are added to the control law by adding them to the dynamic matrix. The DMC controller chooses the control action based on a compromise between meeting the objectives for the constraints and the CVs. The weighting factors (i.e., the diagonal elements for $\underline{\underline{W}}$ in Equation 16.7.1) for the CVs and the constraints are part of the controller design.

16.9 Combining a Linear-Based Optimization with DMC

The widespread industrial use of DMC and MPC is, in general, the result of their ability to operate processes in a more profitable fashion. That is, if DMC provided only reduced variability operation, its use by industry would be drastically reduced compared to its current use. The improved profitability of processes for DMC comes from its ability to operate processes at higher production rates for the more highly-valued products. In many cases, this can result by processing the largest feed rate to the process by maintaining the operation against the most advantageous set of constraints, i.e., constraint control.

A linear program (LP) assesses the constraints and the economics of the process and specifies to the DMC controller against which constraints to control the process. An LP determines optimum values of the decision variables for a linear economic objective function subject to a set of linear constraints. In the case of an LP, the optimum is located at a vertex, i.e., the intersection of n constraints where n is the number of decision variables. The LP determines which is the most favorable set of constrains against which to control. In this manner, as the operation of the process or the cost of the feeds and value of the products change, the LP ensures that the process maintains the most profitable operation. Because the process gains used by the LP are identical to the steady-state gains for the SRMs used by the DMC controller, the LP and DMC controller work together in a consistent fashion.

16.10 DMC Model Identification

In the previous examples, the SRM parameters were calculated directly from a step test. On the other hand, industrial processes are subject to nonlinearity and unmeasured disturbances. As a result, it is better to use a number of step tests to identify "average" coefficients for the SRM (i.e., see Example 16.2).

For model identification, the CV and MV values are known for a sequence of discrete times. The objective function for identification is given by

$$\quad_{i\ 1}^{k} \quad [y_s(t_i) \quad y(t_i)]^2 \qquad \textbf{16.10.1}$$

where k is the number of process measurements available for parameterizing the model, $y_s(t_i)$ is the measured value of the ith CV at t_i and $y(t_i)$ is the value of the predicted value of the ith CV at t_i calculated from the SRM coefficients. The better the SRMs of the process, the smaller the resulting value of . The values of the coefficients of the SRM are calculated such that is minimized. Due to the linear nature of $y(t_i)$, the a_i's can be calculated explicitly using matrix algebra.

When implementing the plant tests which are used for model identification, it is not necessary to make a set of complete open-loop step tests. In fact, for most MIMO systems, such an approach requires a prohibitive amount of time. Moreover, during plant tests, it is important to keep the CVs within specified operating ranges. One way to accomplish this is to make u changes at each of the following intervals:

$$\tfrac{1}{4} T_{ss},\ \tfrac{1}{2} T_{ss},\ \tfrac{3}{4} T_{ss},\ T_{ss},\ \tfrac{5}{4} T_{ss}$$

where T_{ss} is the open-loop time to steady state (the open-loop response time of the process). This approach also develops models for a range of input frequencies. Note that

$$T_{ss} \quad m \quad T_s$$

Mismatch between a process and its MPC models is affected by three factors:

- **Process nonlinearity**. MPC is a linear-based control technology. Therefore, as process nonlinearity increases, the fidelity of the models will generally decrease. An approach called piecewise linearization, which is based on linearizing the CV over a number of fixed intervals, can be used to correct for the effects of nonlinearity for mildly nonlinear processes.
- **Process variability.** Process variability is the variation in the measured value of a CV under relatively steady-state conditions. Process variability results from sensor noise, process fluctuations and from regulatory control loops that interact with one another. Detuning the regulatory control loops to the appropriate levels and servicing or replacing noisy sensors can reduce the process variability.
- **Unmeasured disturbances.** Unmeasured disturbances, such as, ambient temperature changes, are a normal part of industrial operations and must be averaged out during the period of collecting the testing data. On the other hand, if an operator opens a by-pass around a heat exchanger or retune a regulatory control loop, the process that is being tested can change significantly. Therefore, it is essential to remove any data from the test set used to identify the SRMs that contains abnormal changes to the process. Otherwise, the models that are identified using Equation 16.10.1 can be significantly corrupted, leading to a poorly performing MPC controller.

16.11 Organization of an Industrial MPC Application Project

The industrial implementation of MPC has evolved from its inception. The early MPC projects involved controlling a single process unit, such as a distillation column. As MPC has became the industrial standard for advanced control in the CPI and control computers have became much faster, the scope of MPC has grown rapidly. Today MPC applications involve a large number of unit operations spanning a major portion of a process. The larger the scope of an MPC project, the larger the potential economic improvement possible. To be able to successfully apply MPC to such large systems, it is necessary to organize the application project so that the MPC project has the greatest chance for success in spite of the fact that the engineers responsible for these applications generally have varying degrees of MPC application experience. By standardizing the implementation methodology, it is easier for less experienced application engineers to be able to successfully apply MPC Following is an organization of an MPC project that is based on a procedure developed by Bill Korkinski, formerly with AspenTechnology, Inc.

- **Understand the process**
- **Set the scope of the MPC controller**
- **Choose the control configuration**
- **Design the plant test**
- **Conduct the pretest.**
- **Conduct the plant test and collect the plant test data**
- **Analyze the plant test data and determine the MPC model**
- **Implement off-line tuning and testing of the MPC controller**
- **Commission the MPC controller and provide operator training**
- **Perform the post audit**

Understand the process. Process understanding is the single most important issue for a successful MPC application. Stated another way, if you do not fully understand the process and its preferred operation, it is highly unlikely that you will be able to develop a successful MPC application. Following are various approaches that can be used to ensure that you have adequate process knowledge for a successful MPC project.

- Study the process flow diagrams.
- Study the P&IDs for the process.
- Talk to the plant engineers to determine how they want it to work.
- Talk to the operators to determine how it really works.
- Interview the economic planners to determine how much the products are worth.
- Run steady-state process simulations if available.
- Read the plant operating manuals.
- Spend time on the process unit during the graveyard shift (typically 11 pm to 7 am).

After your studies are completed, you should be able to answer each of the following questions.

- What is the purpose of the plant?
- Where does the feed come from?
- Where does the products go?
- How much flexibility is associated with the feed supply and the product demand?
- How do the seasons affect the operation, the feed supply and the product demand?
- What are the 3 or 4 most important constraints that the operators are most concerned about?
- Where is energy used in the process and how expensive is it?
- What are the product specifications and what part of the plant most significantly affects meeting these products specification?
- Are there environmental or tax issues that affect the operation of the plant?

Set the scope of the MPC controller. When setting the scope of an MPC controller, you must decide how much of the process should be included in the controller to meet its objectives. To make this decision, you must discuss the objectives of the project with the operations personnel, the technical staff for that portion of the plant and the scheduling staff. These three groups can have vastly different points of view on the objectives of the project. Only by understanding each of these points of view and reconciling them can you be assured that you are solving the correct problem.

Choose the control configuration. Choosing the control configuration involves selecting the MVs and CVs for the MPC controller. One of the major issues in this step is to determine which regulatory controllers should be left closed and which ones should be opened. That is, when a PID loop is left closed, the MPC controller uses the setpoint to that control loop as an MV, e.g., a flow control loop for which the MPC controller determines the setpoint for the flow controller. PID loops offer certain advantages with regard to rejecting certain high frequency disturbances, e.g., pressure changes that affect flow through a control valve. When a PID is opened, the valve or flow rate in the loop becomes an MV and the CV of the PID loop becomes a CV for the MPC controller, e.g., a tray temperature control loop for which the steam flow rate is the MV and the tray temperature is the CV. Having the MPC controller responsible for certain CVs has the advantage that the MPC controller can provide feedforward compensation for measured disturbances and can provide some degree of decoupling from the other MVs in the process. Making these decisions and others is a challenging task that requires considerable process knowledge and control experience.

When selecting the MVs and CVs for an MPC controller, use good control engineering practice. (1) Make sure that each MV has a direct and immediate effect on at least one CV. (2) Use computed MV control to reject certain disturbances. (3) Use inferential measurements to reduce sensor deadtime. (4) Use transformations for certain CVs to linearize the overall response of the process.

Design the plant test. The design of the plant test involves gathering the required information for the test and planning the testing process. For example, for each selected MV, tabulate its tag number and description, nominal value and range of move sizes. The size of the moves of the MV during the test is a critical issue. If the moves are too small, the fidelity of the MPC model will be compromised. On the other hand, if the MV moves are too large, the process can violate product specifications and process constraints. Therefore, you should make as large MV changes as possible without upsetting the process (e.g., 1-10% MV changes). Make 5-15 step tests for each MV and vary the hold times for these changes before another input change based on a fraction of the time to steady state (Section 16.10). It is important to organize the sequence of MV changes so that the products remain on specification and the process constraints are not violated.

One approach that is used during this phase is to develop a roughed out gain matrix. A table is constructed with the CV's along the top of the x-axis and the MVs along the y-axis. For each CV-MV pair, a "+" is entered if the process gain is positive, a "-" is entered if the gain is negative and a "0" is entered if the gain is expected to be small. In this manner, the roughed out gain matrix can guide the application engineer as he/she designs the sequence of step changes while trying to keep the process on specification.

Conduct the pretest. The pretest is a set of activities designed to ensure that the process is ready to be tested. First, test all the control valves, sensors and regulatory control loops to ensure each is functioning properly. Remember that if you are in the testing phase and determine that one of these items is not function properly and repair it, you have changed the process and may have to retest. Next, make sure that each piece of equipment in the process is on-line and in proper operating order. Apply material and energy balances to the process as a consistency check. Finally, apply MV step changes and check the validity of the roughed out gain matrix. Reconcile any differences between the step tests and the roughed out gain matrix.

Conduct the plant test and collect the plant test data. The plant test is the most important phase of an MPC application project because if the model does not match the process, the controller reliability and controller performance will be poor. If the process changes during the test or large unmeasured disturbance occur during the test, the quality of the MPC models determined from the plant test can be poor.

Because after the MPC application project is completed, you would like the controller to perform well at the optimum operating conditions, it is beneficial to test the process as close to the optimum operating conditions as possible. Therefore, it is important to identify the optimum operating conditions by talking with an experienced process engineer or using a nonlinear optimizer based on process models.

During the testing phase, it is essential to identify abnormal operating periods so that data can be removed from the plant test data set, otherwise the identified models can be corrupted leading to poor MPC control performance. Abnormal operations can result from utility outages or instrument failures. In addition, operators can change the behavior of the process changing the lineup of the flow through the process, e.g., by opening the by-passing for a heat exchanger. In addition, if the tuning of the regulatory control loops is changed during the test, the response of the process can also change. For these reasons, it is usually advisable to have an MPC project engineer observing the process 24 hours per day during the plant test to ensure that corrupted data is not used to train the MPC model.

Analyze the plant test data and determine the MPC model. The accepted plant test data is supplied to MPC identification software, which determines SRM models for each input/output pair. Each input/output SRM should be plotted on a small graph and arranged into a matrix, similar to the roughed out gain matrix. That is, the CVs are listed across the top and MVs are listed along the left side and at each intersection of a CV and an MV, the plot of the corresponding SRM is mounted. In this manner, you are able to look at all of the SRM for the controller at one time. Thus, you can review the models to make sure that they are consistent.

The SRMs can be evaluated by plotting the residuals (i.e., the difference between the plant test data and the SRM prediction) for the complete set of accepted plant test data. By comparing the residuals for positive and negative step input MV changes, you can evaluate the nonlinearity of the process and determine where CV transformations can be applied. In addition, MPC identification packages offer statistical analysis of the SRMs that can guide the analysis of the MPC model. It may be necessary in certain cases to retest specific MVs. Therefore, it is advisable to begin the analysis of the plant test data before the plant test is completed.

Implement off-line tuning and testing of the MPC controller. For a successful application, it is important to accurately represent the economics of the process, properly weight the process constraints and select the proper level of move suppression for each MV. For the LP to function properly, accurate economic parameters are necessary, i.e., the incremental feed and utility costs and product values.

The best way to start the implementation and tuning process is to perform off-line testing of the MPC controller and LP. The MPC models can be used as the "process" and the MPC controller and LP can be interfaced with it in an off-line simulation environment. First, select the move suppression factors that provide reasonable control performance. These settings will help select the initial tuning settings during the on-line testing phase. Next, test the constraint controller to select the proper levels of equal concern errors for the process. Operate so that the controller moves from one set of operative constraints to another and back. Finally, check the operation of LP and the interface between the LP and the MPC controller. In this manner, you can develop preliminary settings for the MPC controller and LP as well as check the various software interfaces.

Commission the MPC controller and provide operator training. Before the MPC controller is applied to the process, it is essential that the operators undergo training for the MPC controller. Before closing the loop, make sure that the interface between the process and the MPC controller is functioning properly by applying the controller in the prediction mode. This will allow you to determine if the CV prediction and the change in the MV levels are reasonable.

When applying an MPC controller for the first time under closed-loop conditions, it is advisable to start slowly with very conservative settings. This approach will allow you to test the controller without creating a major upset. Initially, clamp the MVs using the upper and lower MV limits so that the range of MV changes is quite limited. Put the MVs into service one at a time and relax the MV constraints to determine if the controller is behaving properly. After each MV has been tested separately, begin slowly adding additional MVs to the controller. Ensure that the control performance is smooth and that the MVs are not changing too sharply. Also, determine if the LP is driving the process in the correct direction. In this manner, slowly adjust the aggressiveness of the controller to meet the expected control performance. In addition, operator training should be provided to ensure that the operators better understand what the MPC controller is attempting to do and how to interface with the controller.

Post audit. After the controller has been properly tuned, it is important to compare the performance of the MPC controller to the pre-project performance. This can be done by comparing the variance from setpoint (Section 9.2) for key CVs for the pre-project controller to the MPC controller. Also, it is important to determine whether the MV limits are set properly (i.e., not too tight or too loose). Finally, ensure that the controller is running the process at the economically optimum set of constraints most of the time. In addition, it is usually required to provide a complete set of documentation on the project to assist in the long-term maintenance of the application.

16.12 Summary

- Multivariable controllers use measured values of the MVs, the CVs and disturbances to simultaneously calculate all the MV levels for the control of MIMO processes. Multivariable controllers can provide decoupling, feedforward compensation and compensation for nonlinear behavior.
- DMC is the most popular form of MPC and uses SRMs of the process to calculate control action. The DMC controller uses the coefficients of the SRM, the previous input history and the latest measured value of the CV to calculate the next MV value using a moving horizon control approach. A move suppression factor is used to tune the DMC controller to provide reliable control performance.

16.13 Additional Terminology

Centralized controller - a controller that uses all the available process information to select control action for a MIMO process.

Control horizon (n_c) - the number of sampling intervals into the future that the control moves are calculated by the DMC controller.

Discrete-time step response model - the discrete-time behavior of a process to a unit step change in an input variable.

Dynamic matrix control (DMC) - a linear model predictive controller that uses step response models of the process to select control action.

Dynamic matrix - the $\underline{\underline{A}}$ matrix in the DMC controller, which is constructed from the coefficients of the SRM.

Model-based controller - a controller that chooses control action based upon a model of the process.

Model horizon (m) - the number of sampling intervals used to model a step response of the process.

Moving horizon algorithm - a control approach that calculates future control moves based on a process model and the input history.

Model predictive control (MPC) - a controller that uses a process model to calculate control action by determining a sequence of inputs that regulates the process to setpoint.

Move suppression factor (Q) - the tuning factor for a DMC controller that determines the aggressiveness of the controller.

Multivariable controller - a controller that uses all the available process information to select control action for a MIMO process.

Prediction vector (y^p) - the future values of the CV if no changes in the MV are made.

Prediction horizon (n_p) - the number of sampling intervals into the future for which the model is used to predict the behavior of the CV.

Process/model mismatch - the difference between the model predictions for the CV and the actual process behavior.

Sampling interval - the time period between the calculations of control action.

Single-step-ahead controller - a controller that calculates the current control action without predicting the future behavior of the process.

Step response model (SRM) - the discrete-time response of a process to a unit step change in an input variable.

16.14 References

1. Richalet, J., A. Rault and J. Papon, "Model Predictive Heuristic Control: Application to Industrial Processes", Automatica, Vol 14, No. 5, pp. 413-428 (1978).

2. Qin, S. J. and T. A. Badgwell, “An Overview of Industrial Model Predictive Control Technology”, *Proceedings of the Fifth International Conference on Chemical Process Control*, AIChE Symposium Series Number 316, Vol 93, p. 232 (1999).

3. Cutler, C. R. and B. L. Ramaker, “Dynamic Matrix Control - A Computer Control Algorithm”, *Proceedings of the Joint Automatic Control Conference*, San Francisco (1980).

4. Cutler, C.R., personal communication, Oct. 1999.

Chapter 17

Process Safety

Chapter Objectives

- Provide an overview of how safety is addressed in the process industries.
- Demonstrate the levels of protection that are used to ensure safe operation and mitigate problems when they occur.
- Show how to use the reliability of components to determine the overall reliability of a system.

17.1 Introduction

The first fundamental cannon of the National Professional Engineering Society (NSPE) is "1. Hold paramount the safety, health and welfare of the public."[1]. The formation of professional engineering societies resulted in large part because of a number of structural failures that occurred around the turn of the 20th century that resulted in significant loss of life and damage to communities. As a result, a key fundamental element of engineering societies is an emphasis on ensuring public safety. The American Institute of Chemical Engineers[2] (AIChE) was formed in 1908 and also emphasizes public safety in its code of ethics and programs.

Process engineering provides major benefits to society, but along with these benefits significant risks exist. In the textbook by Crowl and Louvar[3], they point out that the process industries has one of the best safety records in the manufacturing sector. Nevertheless, a number of large and destructive accidents have occurred. In order to indicate the scale of damage that can occur from industrial accidents associated with the process industries consider the following accidents.

- **Bhopal India Chemical Disaster,** known as the world's worst industrial disaster, was a chemical gas leak accident in 1984 at the Union Carbide pesticide plant. Over 30 tons of the toxic gas methyl isocyanate (and other gases) were released which eventually killed an estimated 15,000 people. The gas leak resulted from operating errors, design flaws, maintenance failures, training deficiencies and economy measures that endangered safety.
- The **BP Texas City Refinery Explosion**[4] occurred in 2005 and resulted in 15 deaths with over 170 injuries. The explosion was caused because by the failure a high level alarm on the bottoms of a distillation column. The distillation column was filled with hot hydrocarbon liquids and was eventually

discharged to the atmosphere when the pressure relief system released the hydrocarbon liquids to the atmosphere. A heavier than air cloud of hydrocarbon vapor formed and was ignited by a vehicle.

- The **Phillips Disaster**[5] occurred in 1989 in Pasadena, Texas and resulted in the deaths of 23 employees and caused 314 injuries. The initial explosion and fire took place on a polyethylene reactor in which a block valve was not properly installed resulting in a large discharge of hydrocarbon vapors that were ignited causing the explosion and subsequent fires.
- The **BP Oil Spill**[6] occurred in 2010 and resulted in the deaths of 11 workers on the platform. After the explosion on the rig, the well discharged crude oil into the Gulf of Mexico for 87 days before the well could be capped resulting in the discharge of 4.9 million barrels of oil. This spill is recognized as the largest offshore oil spill in US history.
- The **Texas City Disaster**[7] occurred in 1947 and was the deadliest industrial accident in US history. It occurred when a ship loaded with 2300 tons of ammonium nitrate fertilizer exploded after a fire broke out in the cargo area of the ship. As a result of the explosion and fires, at least 578 people died and over 3500 people were injured.
- **Fukushima Daiichi Nuclear Disaster**[8] was initially caused by a tsunami. The reactors automatically shutdown, but the tsunami took the backup cooling system off line so that three reactor melted downs occurred, leading to the release of radioactive materials second only to that of Chernobyl.
- **Haysville Grain Elevator Explosion**[9] occurred in Haysville, Kansas in 1998. A series of explosions of a mixture of grain dust and air resulted in the deaths of seven people and the destruction of the grain storage facility.

These examples of industrial accidents demonstrate that it is critically important to ensure safe industrial operations particularly when dealing with hydrocarbon systems, nuclear power operations, combustible dust and nitrogen-based fertilizers. Even when processing aqueous solutions (e.g., bioprocessing), there are safety issues, but the worst-case risk scenarios pale in comparison when processing hydrocarbons, for example.

Today, industry uses systems and operating procedures to ensure that their plants operate in a reliably safe fashion. For example, when a plant is designed, safety is a key issue in the design process. That is, the inherent safety of a design is a key issue in the preliminary design process. Later in the design process, hazard and operability (HAZOP) studies, as mandated by OSHA's Process Safety Management (PSM) Program, are conducted as part of the evaluation of each design option. In addition, after a process is in operation, HAZOP studies are periodically applied in order to identify potential problems. If design or operating changes are proposed for a process, a **Management of Change** (**MOC**) procedure is used to ensure any proposed significant change, whether for safety, environmental or other reasons, is appropriately reviewed, authorized, designed, tested, implemented and documented. And finally, if a serious safety incident occurs in the US, the Chemical Safety Board, which is an independent federal agency, will investigate the incident to determine the determine the cause and recommend action to prevent such a incident from occurring again.

The remainder of this chapter will address how process control and safety systems are intertwined in an effort to provide safe operation of plants in the process industries. Note that most of the chapter content has broader applicability than focusing just on safety. For example, the use of a Basic Process Control System (BPCS), alarm management, and risk analysis apply just as much to environmental issues and product quality issues as they do to safety. The chapter begins by considering the layer of protection approach that is used to accomplish safe operations (and other benefits).

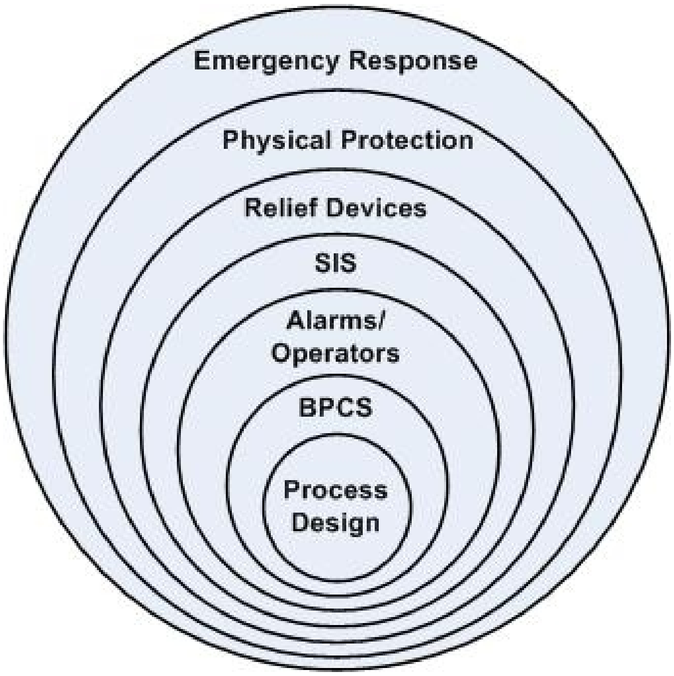


Figure 17.2.1 Schematic of the Layers of Protection approach to process safety.

17.2 An Overview of Plant Safety Systems (Layers of Protection)

The standard approach to safety system design in the process industries is to use the **Layers of Protection** (LOP) approach[10]. Figure 17.2.1 is a graphical representation of the Layers of Protection approach, which is based on a number of independent measures that are taken to ensure safe operation or mitigation of a safety incident. Moreover, the basic idea of this approach is that a safety incident should be handled by the lowest levels of the LOP as possible. For example, if the design and the BPCS (i.e., the process control systems) were unable to handle a safety incident, the alarm/operator function would come into action. If the alarm/operator system was unable to handle the incident, the SIS system would come into action, and so on.

Before discussing the individual layers, it is important to note the need for a critical source document, expected in project management, known as a process/project Functional Specification (sometimes called Functional Requirements) document. Preparing this document should precede any work on process design. That is, the Functional Requirements document specifies the **'whats'** of a process or process control system and the Process Design then identifies the '**hows'** (i.e., **how** the functional requirements will be met). The Functional Requirements document may also very well cover topics in several other layers of protection, such as the BPCS, alarming, and any need for a SIS.

The first layer of protection is the safety features built into the design of the process. Different designs have different susceptibilities to safety incidents; therefore, a well-conceived design will be less susceptible to key safety incidents. In addition, standard designs use a number of features that have safety implications. For example, Figure 17.2.2 shows a standard piping configurations for a control valve so that if the control valve fails, the bypass can be partially opened and the block valves on either side of the control valve can be closed, thus allowing for manual control of the flow. This would allow for at least a reasonable flow rate instead of having to rely on a malfunctioning control valve. In addition, the control valve can be replaced without having to shutdown the process with this configuration. Block valves, which are specified during the design process can also be used to isolate a portion of the process under abnormal operating conditions.

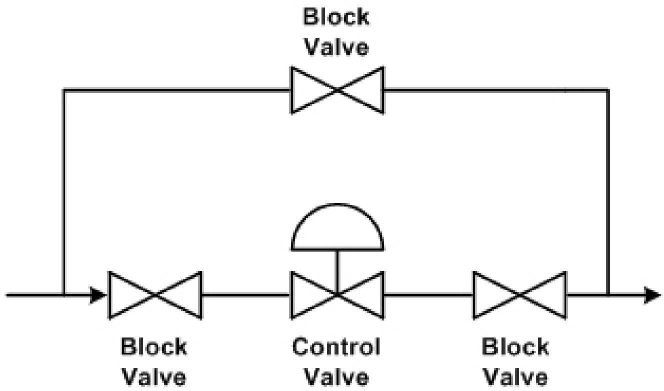


Figure 17.2.2 Typical piping configuration for a control valve.

The second level of protection, the **Basics Process Control System** (**BPCS**), involves the regulatory and supervisory control loops of the process and is addressed in more detail in the next section. The BPCS should be designed to maintain steady operation of the process in the face of upsets to the system, i.e., maintain levels, flows, temperatures, pressures and compositions in a stable operating range. A properly designed and implemented BPCS will handle large upsets gracefully and thus eliminate the need for further intervention in many cases. The implementation of the BPCS should consider extreme operational upsets as much as possible.

When the operation of the process is outside its normal operating window, alarms should be activated in the control room so that the operators are aware of the upset conditions and can take corrective action (i.e., the third level of protection). **Alarm systems** and **alarm management** are addressed in Section 17.4.

The fourth level of protection is the **Safety Instrumented System** (**SIS**), which is a normally dormant independent control system that takes corrective action only when an emergency results after first three layers of protection failed to handle the incident. For example, the SIS could shut off feed to an exothermic reactor and increase the cooling rate for the reactor to the maximum heat removal rate possible if the reactor temperature exceeds an upper limit. In certain cases, the SIS could shut the entire process down. A SIS, when utilized, is typically separate from the BPCS so that issues with the BPCS (e.g., computer crash) will not affect both the BPCS and SIS levels of protection. Also, a SIS is only utilized when needed. Many processes in industrial plants do not need a SIS layer. The SIS is considered in more detail in Section 17.5.

The fifth level of protection are relief devices, such as **relief valves** and **rupture disks**, which are designed to prevent the system pressure of a vessel from exceeding its rupture limits and thus prevent seriously damaging major pieces of equipment and many times prevent the environmental discharge of highly flammable materials. In most cases, the relief valves and the rupture disks are connected to discharge lines that lead to flaring furnaces or retention vessels for safe disposal or sequestering of the vented material.

The sixth level of protection involves the use of dikes around process units and storage vessels to retain any liquid discharge during a safety incident. The seventh level of protection is the emergency response, i.e., evacuation procedures, for employees and individuals in the surrounding communities.

Example 17.1 LOP for a Pressurized Tank Containing a Volatile and Flammable Material

Problem Statement. Develop a LOP approach for a pressurized tank that contains a volatile and flammable material and has a level controller applied.

Solution. The concern for this example is that if the tank were to overflow it could cause a safety incident. For the BPCS, a level controller regulates the level in the tank using a level sensor and a control valve in a standard arrangement (Figure 17.2.3). The next level for the LOP is a high level alarm. Note that in Figure 17.2.3, PHA stands for pressure high alarm. When the high level alarm sounds, it allows the operator to cut the feed to the tank. If the level continues to rise above the alarm level, the SIS would vent the tank to a flare. Note that different level sensors are used for the BPCS and SIS because if a single level sensor is used and it fails both systems would be inoperable.

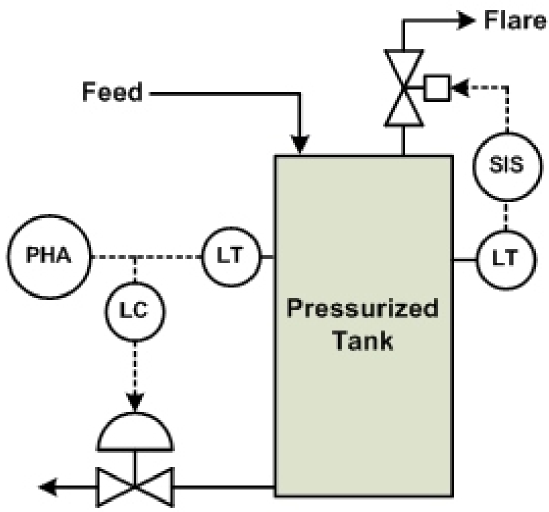


Figure 17.2.3 Schematic of LOP Example.

17.3 Basic Process Control Systems (BPCS)

The BPCS uses regulatory and supervisory control to maintain a process at its desired operating point (e.g., the specified levels for compositions, temperatures, pressures, levels, agitation rates and flow rates). As a result, the BPCS is designed to absorb normal disturbances (e.g., flow rate changes for the feed to a unit and day-to-night ambient air temperature changes). The primary components of the BPCS as pointed out in Chapter 2 are the sensor system, the actuator and the control computer and abnormal conditions can result when any of these components fails. Moreover, specific aspects of these system can be used so that the BPCS is more tolerant to component failures and major upsets.

For example, as pointed out in Section 2.3, the choice of the type of valve actuator used (i.e., air-to-open or air-to-close) will determine whether the valve fails open or closed when the instrument air is lost. Therefore, depending on the particular process, selecting the proper type of valve actuator for critical control loops can be the difference between a safety incident and a graceful transition when the instrument air is lost.

Override/select controls (Section 13.4) can be used for a variety of safety-related purposes. For example, cross-limiting firing controls are used to ensure that carbon monoxide, which is a toxin, is not emitted from a utility boiler. Select control is used to keep a distillation column from flooding, prevent damage to boiler tubes and control the maximum temperature inside a fixed bed reactor.

In addition, as pointed out in Section 14.2, windup can undermine control performance and in certain cases lead to a safety incident. By employing some form of anti-reset windup, this operational problem can be eliminated.

Finally, because of the vulnerability of the BPCS to hardware failures, using redundant components can significantly reduce this risk. This is particularly true for critical control loops, e.g., temperature control for a fluidized catalytic cracking reactor. Hardware redundancy is typically applied as "cold", "warm" or "hot"

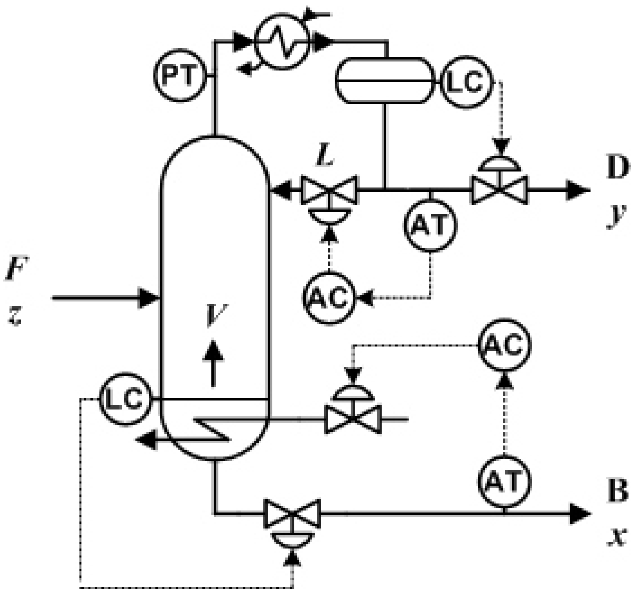


Figure 17.3.1 Schematic of a distillation column with the (*L*,*V*) configuration. Note that control valves represent flow control loops in this figure.

redundancy. Cold redundancy is used when there is adequate time to respond to a hardware failure. For example, from Figure 17.2.2, if a control valve fails on a slow responding process, the operators have plenty of time to open the bypass and replace the control valve. Warm redundancy occurs when the failure is more important but not critical. For example, if a controller were to fail but there is adequate time to manually bring on a backup control computer, this would be an example of warm redundancy. On the other hand, hot redundancy is used if the control process must not go down even for a short period of time. As an example, hot redundancy would be used for a control computer on an exothermic reactor that is susceptible to thermal runaway on a very short time basis. In the past, triple redundancy has been used for control computers on critical process units. For critical control loops, redundant sensors are used so that it is highly unlikely that the control loop will be without a good sensor reading.

As an example of a BPCS, consider the distillation column shown in Figure 17.3.1. Note that the overhead and bottoms compositions are controlled using composition controllers that are cascaded to flow controllers on the reflux and steam flow to the reboiler, respectively. In addition, level controllers are used to maintain the levels in the overhead receiver and the bottom of the column and are cascaded to flow controller on the overhead product and the bottoms product, respectively. Distillation columns can also apply temperature control using inferential tray temperature control (Section 13.2) and pressure control for the overhead pressure although these cases not shown in Figure 17.3.1.

The control computer for the BPCS comes in primarily three different forms and sizes. The smallest of these is a personal computer (PC) which, with certain software programs and I/O hardware, provides a relatively simple operator interface and control of several control loops. This form is often found in academic research laboratories and on industrial bench top equipment. They are not used very much in manufacturing plants due to their limited functionality, capacity, and ruggedness.

The most common form is a Programmable Logic Controller (PLC) which was developed, in part, for rugged industrial applications. It was originally designed to replace mechanical relays and utilized a programming language known as ladder logic. However, functionality has been expanded in recent years and PLCs now handle most monitoring and control aspects of individual unit operation equipment, with typical capacity of hundreds (or more) of input/output (I/O) points. It is very common for an equipment vendor to include a PLC with each piece of unit operation equipment (e.g., reactor, distillation column) that they provide to a customer.

The largest and most functional form is a Distributed Control Systems (DCS). These are typically installed in large pilot and manufacturing plants and contain a wide range of monitoring and control functions, often including embedded data historians and various advanced control algorithms. Such systems can easily handle many thousands of I/O points. A common industrial manufacturing plant configuration is for each major individual unit operation piece of equipment to be controlled by a dedicated PLC, with most or all PLCs then interfaced to a plant DCS. Such a configuration allows for most key plant information, e.g., alarms, process variable trends, to be standardized and centralized in one format in one place (i.e., in the DCS), thus freeing operators from having to interact with different systems with different formats, procedures, and user interfaces. It also enables engineers to optimize an entire plant, rather than just optimize individual unit operations.

17.4 Alarm Systems and Alarm Management

The national alarm standard (ANSI/ISA 18.2 [www.ansi.org]) and alarm best practice documents all agree that the definition of an alarm is "an indication of an abnormal event requiring a response". So, alarms are typically triggered when a process direct or virtual measurement (e.g., process variable) is outside a specified range in order to make the operator aware of the abnormal situation and to prompt a corrective response.

In the mid to late 1900s, when pneumatic and electronic controllers were prevalent, the number of alarms was modest since every alarm typically required wiring, other hardware, and significant time and effort for implementation. The advent of computer control changed this paradigm, as it became easy and cheap to add alarms. For example, an engineer could easily add an alarm to a control loop (with the controller now in software), or just about any other logic contained within application software, in less than 5 minutes at no capital cost. This caused an explosion in the number of alarms configured into a system and the evolution of bad habits by engineers such as using the alarm system for non-alarms (such as “alerts” and “notifications”), configuring redundant alarms, configuring chattering and other nuisance alarms, and creating alarms without a designated priority. What followed was often chaotic, including “information overload” for operators when real abnormal situations were occurring. Operators were often unable to keep up and make sense of all the information coming their way.

While alarm best practice publications (e.g., EEMUA 191 [www.eemua.org]) existed in the early 2000s to provide some guidance to alarm management, a national standard was deemed necessary. The International Society of Automation then created the national standard on alarm management (ANSI/ISA 18.2), originally published in 2009, which subsequently became an international standard (IEC 62682 [www.sis.se]). These standards require that an alarm philosophy document be developed to guide specific alarm management activities in a plant. The standards list the attributes associated with each alarm (setpoint, priority, deadband, time delay, class and type) and the need to agree on values for these attributes during a formal process called “rationalization”. Documentation of the values and the rationale for the values are expectations of the rationalization process. Some of the alarm attribute values (e.g., setpoint, priority) will be dependent on the step/phase of a process, especially for batch processes. Another aspect of an alarm is whether or not it is suppressed or unsuppressed for the various steps or phases of a process. Even continuous processes can be thought of as various phases (start-up, operation, and shut-down) with alarms which may be suppressed or have different attribute values during start-up and shut-down).

The standards also require specific documented activities regarding functional requirements, alarm identification, design, Human Machine Interface (HMI), advanced alarming, implementation, operation, maintenance, monitoring/assessment, A Management of Change procedure, and audit.

Regarding the alarm “Identification” portion of Alarm Management, the activities summarized in other parts of this chapter (e.g., HAZOP studies, FMEA) as well as government regulations and equipment manufacturer recommendations can be useful source documents in not only identifying what alarms need to be configured into a system, but also some of their attribute values (e.g., setpoint).

Regarding the "Functional Requirements" part of Alarm Management, these can be separate from or contained with the Functional Requirements of the overall control system (i.e., requirements for the BPCS and any SISs). These Requirements are the main source document for design activities and also drive testing activities (i.e., the purpose of testing is to prove that the system meets its requirements). So, requirements must be written in a way that they are testable.

Regarding the "Operations" part of Alarm Management, these activities include operator training as well as expectations on acknowledging alarms and appropriate responses to alarm activations.

In summarizing some of the most important alarm system recommendations, as identified in documents such as EEMUA 191, Seborg et. al. [12], the ANSI/ISA 18.2 standard (and its supporting technical reports) and ANSI/ISA 101 on HMIs:

- Every alarm should be an indication of an abnormal situation requiring an operator response
- Every alarm should have a priority (i.e., ranking of severity) associated with it (as a configured attribute)
- An alarm should continue until it is acknowledged
- Information overload should be avoided by judiciously avoiding redundancy, purging nuisance alarms, and not using the alarm system for routine alerts and notifications
- Each alarm should be recorded along with the resulting operator acknowledgment and response
- The operation of the alarm system should be monitored routinely and audited at least yearly. Monitoring can help identify any operator frustrations with the system. It can also identify the most frequently occurring alarms which, in turn, can drive process improvement projects.
- Changes to the alarm system and to alarm attribute values should be managed via a formal Management of Change procedure, including appropriate documentation.

For an indication of the frequency of alarms an operator is capable of handling, the existing literature (including the documents noted above) does contain some suggestive metrics (e.g., some suggest one alarm per 10 min. per operator is manageable). However, this is a controversial topic and not all the main guidance documents on alarm management agree on the metrics. Further it seems some of the numbers are meant for plants which incur thousands of alarms per day (common in many plants) and so the existence of large numbers of alarms and the need to respond to them is a major portion of operator duties. From another perspective, many of the alarms represent critical situations (e.g., safety, environmental, and product quality) where even one alarm per day is considered unacceptable in which case there needs to be active engineering and management efforts to help drive the frequency of at least the critical alarms to as close to zero as practical.

In many cases, the frequency of alarms is a testament to the quality of the control/automation system. That is, if a plant is highly automated and well controlled, things like product quality variability, environmental excursions, and alarm generation should be minimal.

17.5 Safety Instrumented System (SIS)

A Safety Instrumented System (SIS) is an independent emergency response system that performs specific control functions to maintain the safe operation of a critical process system and/or to take action when unsafe conditions are indicated. The action of the SIS is one of the last LOP responses to an emergency situation and often results in either putting the process in a safe state or shutting down the process. As pointed out in Example 17.1, it is important that the SIS is independent of the BPCS (in order to ensure SIS functionality is not compromised) and thus has its own control hardware (i.e., sensors, controllers and actuators). As part of the SIS, the Emergency Shut Down (ESD) system is designed to shut down the entire process in a systematic and controlled fashion.

A critical process system is one which, once running and an operational problem occurs, may need to be put into a "Safe State" to avoid adverse Safety, Health and Environmental (SH&E) consequences.

The need for a SIS usually results from having conducted a formal hazard identification process (HAZOP) by the project team engineers and other experts at the completion of the engineering design phase. A HAZOP study often reveals hazardous scenarios which require further risk mitigating measures.

A SIS typically requires sensors capable of detecting abnormal operating conditions. A logic solver (usually software) then receives the sensor input, makes appropriate decisions based on the nature of the signal(s), and change its outputs according to user-defined logic. The logic solver may use electrical, electronic or programmable electronic equipment, such as relays, trip amplifiers, or programmable logic controllers. The change of the logic solver output(s) results in the final element(s) taking action on the process (e.g. closing a valve) to bring it to a safe state.

International standard IEC61511 (www.sis.se) was published in 2003 to provide guidance to end-users on the application of Safety Instrumented Systems in the process industries. The applicable USA Standard regarding SIS is ANSI/ISA 84 (www.ansi.org).

Many processes in industry do not need or utilize a SIS LOP (e.g., most industrial fermentations), but many others (i.e., those that deal with nuclear or highly flammable or toxic chemicals) do.

Example 17.2 SIS for a Steam Boiler

Problem Statement. Develop SIS interventions for a steam boiler. The process schematic is shown in Figure 17.5.1. Fuel is combusted in the fire box and the hot combustion gases flow though tubes in the boiler transferring heat to the water producing steam. The primary failure modes for this process are (1) a loss of ignition in the fire box, (2) over filling the boiler and putting water into the steam system and (3) exposing the boiler tube due to a low water level in the boiler, which will damage the boiler tubes.

Solution. Figure 17.5.1 indicates how the SIS can be used to address each of the issues. A flame sensor (FS) is used to detect when a loss of ignition has occurred and the SIS shuts off the flow of fuel in this case. When the water level in the boiler is too high, the make-up water is shut off by the SIS. And when the water level in the boiler becomes close to exposing the boiler tubes, the SIS shuts off the fuel flow rate and some time after the fuel has been shut off, the flow rate of makeup water is shut off to prevent over filling of the boiler.

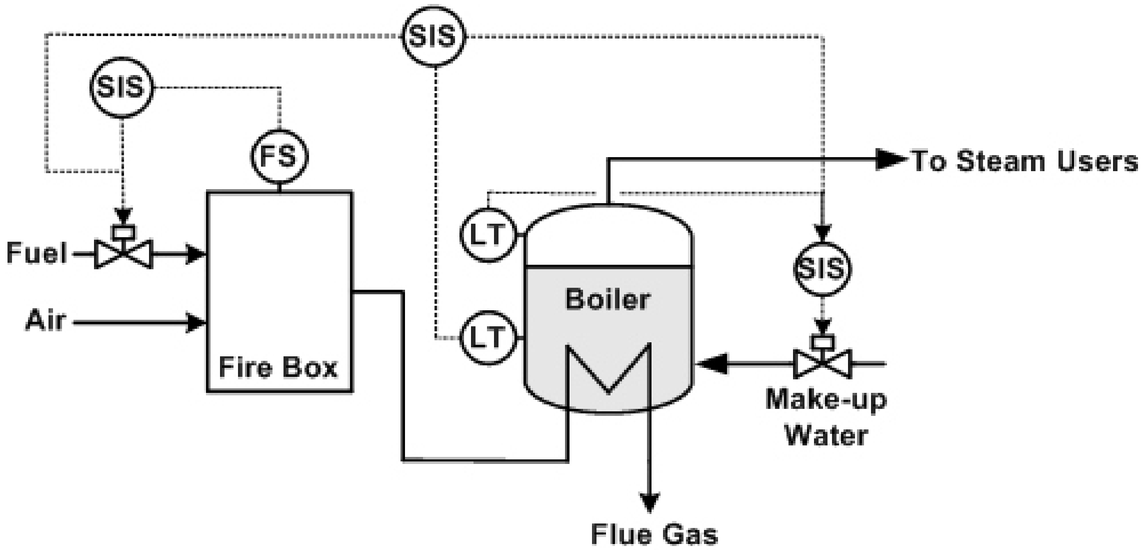


Figure 17.5.1 Schematic of steam boiler with SIS controls. FS-flame sensor.

17.6 Relief Devices

Relief devices prevent process vessels from rupturing due to high pressures by venting the contents to, e.g., a furnace flare, retention vessel or directly to the atmosphere. Both pressure relief valves and rupture disks are used as relief devices **in which the pressure relief valves takes the initial action and the rupture disk is a back up in case the valve is unable to prevent the pressure from rising.**. A pressure relief valve is a spring loaded valve that opens when the pressure is equal to the set pressure of the device and closes at a pressure typically 4-12% lower than the set pressure. A rupture disk is a metal disk that is flanged into a line that will rupture when the pressure exceeds its design pressure. The primary difference between them is that rupture disks are much less expensive, but relief valves terminate the venting process when the pressure has dropped to a safe level. In addition, under certain conditions relief valves can suffer for unstable oscillations (i.e., banging against the stops for closure and for full open) that can destroy the valve resulting is a discharge of the contents (e.g., hydrocarbon vapors) into the environment. If the high pressure conditions were due to a fire, this discharge would further feed the fire. Determining whether a relief valve can become unstable during a discharge event is a complex problem, but in general, relief valves tend toward instability as the length of the pipe from the vessel to the relief valve increases.

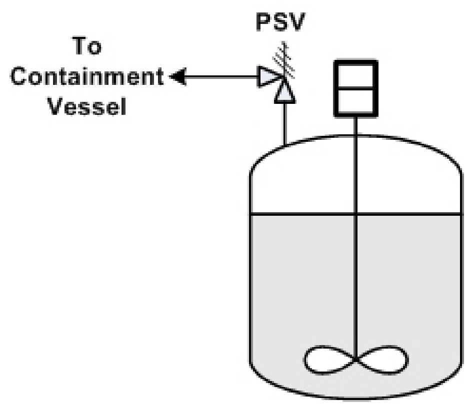


Figure 17.6.1 A pressure relief valve (PSV) applied to a batch reactor.

A common example of a pressure relief device is a relief valve applied to a batch reactor (Figure 17.6.1). It is essential to properly size the relief valve so that during an event involving a fire the relief valve is able to prevent the pressure in the reactor from exceeding the rupture pressure of the reactor. An undersized relief valve will cause the pressure during a fire to exceed the set pressure of the relief valve and cause the vessel to rupture.

Another common example of a pressure relief system is the use of several relief valves attached to the overhead condenser line of a distillation column (Figure 17.6.2 Note the same symbol for a pressure safety valve, PSV, is used in both Figures 17.6.1 and 17.6.2). From a cost standpoint, it is better to have several relief valves than to have a single large one.

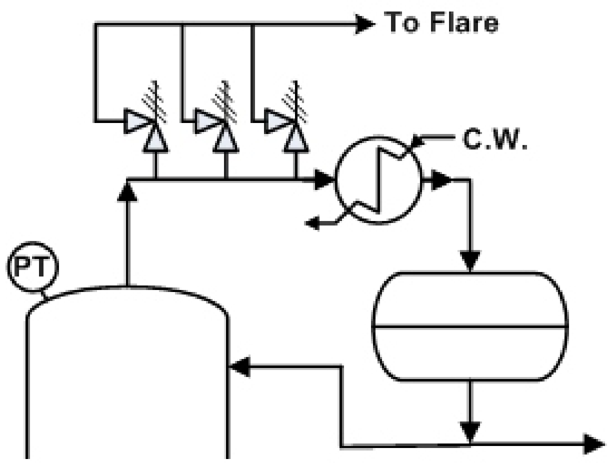


Figure 17.6.2 Three pressure relief valves applied to the overhead line of a distillation column.

17.7 Risk Assessment

Risk assessment in process safety includes incident identification and consequence analysis. Incident identification describes how an incident occurs. Consequence analysis describes the expected damage, which includes loss of life, damage to the environment or capital equipment, and days of outage. Risk management is the process of analysis and acceptance or mitigation the identified risk. Risk assessment and management involved in process control, focuses on evaluating the risk, and managing the risk concerning the failure of process control devices, such as controllers, actuators and sensors.

First, let's introduce several terms that will be used to implement risk assessment. The probability that a device will not fail in the reference time period is referred to as the **reliability**, R. Then, the probability of failure for a device in the reference time period, F is given by

$$F = 1 - R$$

When failure components are in series, i.e., a failure of any component will cause the system to fail, e.g., a control loop with a controller, a sensor and an actuator, the overall reliability R is a function of the reliability of the individual components in series, R_i,

$$R = \prod_{i=1}^{n} R_i \qquad \textbf{17.1}$$

For example, if a system is composed of two devices each with reliability, R_i, equal to 0.5, the overall reliability, R, is equal to the product of the R_i's or 0.25. Therefore, the probability of failure is equal to 0.75.

For components in parallel (e.g., a redundant sensor), the overall probability of failure, F, is given by

$$F = \prod_{i=1}^{n} F_i \qquad \textbf{17.2}$$

For example, for a redundant sensor with an individual probability of failure of 0.1 per year, the overall probability of failure of two redundant sensors is 0.01 per year.

Risk is the combined effect of the probability of failure with the economic and personnel/community consequences. Even a component with a low probability of failure can occasionally fail. When the probabilities of specific failures that lead to safety incidents are combined with the associated consequences of the incidents, the result is the risk.

Example 17.3 Probability of Failure of a Control Loop with and without Redundant Components

Problem Statement. Determine the overall reliability for a flow control loop with single components and double redundant components. The component reliabilities for a period of one year for the controller, flow sensor and control valve are 0.99, 0.8 and 0.7, respectively.

Solution. For the non-redundant case, Equation 17.1 can be applied directly,

$$R \quad (0.99)(0.8)(0.7) \quad 0.54$$

Therefore, the probability of failure is equal to 0.46

For the redundant case, Equation 17.2 is applied to each component before Equation 17.1 can be applied to determine the overall reliability. For the controller, F_1=(0.01)(0.01)=0.0001 and R_1=0.9999. For the sensor, F_2=(0.2)(0.2)=0.04 and R_2=0.96. For the valve, F_3=(0.7)(0.7)=0.049 and R_3=0.951.

Next, Equation 17.1 is applied to determine the overall reliability for the redundant system,

$$R \quad (0.9999)(0.96)(0.951) \quad 0.913$$

Therefore, the probability of failure for the redundant case is 0.087. As a result of the redundancy, the overall probability of failure during a one-year period was reduced from a 46% to 8.7% . Note that the redundancy of the controller did not significantly affect these results; therefore, redundancy for only the flow sensor and the flow measurement should be used.

Typically, risk assessment methods can be divided into three categories, qualitative method, semi-quantitative method and quantitative method. Constraints, such as time, money, skills, management perceptions and communication of risks to the public, affect the manner in which risk assessments are performed . When involved in risk assessment, the qualitative approach is considered to produce a subjective sense of risk, which may rank the risk of one scenario or group of scenarios to be greater than another scenario or a group of scenarios. In quantitative risk assessment, the risk regarding each scenario is evaluated numerically. This method not only allows people to identify the sequence of how a scenario could occur, but also measures risks of the chosen control system. Semi-quantitative risk assessment use values in the form of broad ranges of frequency or consequence levels.

For complex processes, graphical methods, such as fault tree analysis and event tree analysis, can be used to systematically perform a risk analysis. While this can appear to be a straight forward approach, these analysis tools are only as good as the process knowledge of the individuals using them. For example, if a significant failure mechanism is not considered, a risk analysis is likely to be flawed.

From a risk assessment perspective, the LOP approach is composed of a number of **somewhat** independent methods; therefore, Equation 17.2 is used to describe it. **That is, the more independent the elements of an LOP, the better the risk performance.**

Engineers should be cautious and conservative in considering any reliability data obtained from external sources, since the reliability of equipment operating in the manufacturing plant (which is often a hostile environment) is a function of many variables. Therefore, reliability will often be less than vendor data or the literature might suggest. For example, a vendor might supply Mean Time Between Failure (MTBF) data on their temperature sensors sold to customers which may have been obtained under ideal conditions in a test lab. However, for a real plant which experiences vibrations, ambient higher heat and humidity, occasional chemical spills, occasional voltage surges from local lightning strikes, and rough handling by operators, the real MTBF will likely be significantly less. Conservative estimates of reliability data should be used, to include some safety margin. Also, especially for plants that have a long history of operations, a plant's maintenance records can be a good source of reliability data.

At the completion of the initial process design, it is a standard best practice for a team to be organized to execute a Failure Modes and Effects Analysis (FMEA). Such an activity evaluates and scores all the failure modes of the process with respect to magnitude, expected frequency, and deductibility with the Risk Probability Number (RPN) then calculated as the product of these three numbers. Many, if not most, identified failure modes are safety related. The goal is perform a redesign for the process for those failures for which RPN is above a team/management determined threshold.

Example 17.4 Risk Analysis for Exothermic CSTR

Problem Statement

A plant FMEA team has established the following FMEA scoring matrices:

<u>Severity</u>

3 = likelihood of major equipment loss and/or personnel injury or death

2 = likelihood of moderate equipment damage and/or minor personnel injuries

1 = likelihood of minor equipment damage

<u>Frequency</u>

3 = MTBF < 1 year

2 = 1 year < MTBF < 3 years

1 = MTBF > 3 years

<u>Deductibility (of the failure)</u>

3 = no immediate deductibility (i.e., awareness by operators)

2 = operator awareness of the abnormal situation in 1-5 minutes

1 = operator immediately aware of the abnormal situation (< 1 minute)

A CSTR has only one control loop installed, that being for temperature control. Perform a risk analysis for a tank in which an exothermic reaction takes place with volatile toxic chemicals as a side product.

Solution

One (of many) possible failures is the failure of the temperature sensor. Should this happen, it is possible that the exothermic reaction would take off triggering a runaway reaction, raise temperature and pressure, and ultimately cause the rupture seal to break, releasing toxic gases to the surrounding area. Using the above matrices, the team scores this failure mode as 3 for severity, 2 for frequency, and 3 for deductibility.

The Risk Probability Number is then calculated as the product of these 3 numbers, which is 18.

The plant's engineering and management team has established a risk acceptability threshold of 10, meaning any RPN above 10 is unacceptable.

Therefore, a RPN of 18 then prompts a redesign. The design team decides to 1) add a pressure sensor with accompanying alarm to the reactor and 2) direct the output (vent) of the rupture disk to a scrubber unit. With this redesign, the severity score drops from 3 to 1 and the deductibility score drops from 3 to 2. The resulting RPN score of 4 is well below the threshold of 10, so the risk has been sufficiently reduced such that no further redesign is warranted.

Note that for simplicity, matrices with only 3 levels were used in this example. It is common for most FMEAs to have matrices with 5 or 10 levels of classification. Note also that most processes have many failure modes (often hundreds or more). The above analysis would be performed on each of the possible failure mode.

17.8 Summary

Safety is of paramount importance for the process industries from an economic and personnel/community safety point-of-view. The proper implementation of process safety involves a coordinated effort in terms of design, implementation and support. The LOP approach is the standard for safety in the process industries because it involves applying a number of independent methods for dealing with a safety incident, any one of which can, in certain cases, handle the problem by itself. Moreover, the LOP approach uses the simplest and least intrusive approach first and sequentially adds more aggressive methods to address a problem.

17.9 References

1. National Society of Professional Engineers code of ethics (www.nspe.org/resources/ethics/code-ethics)
2. American Institute of Chemical Engineers code of ethics (www.aiche.org/about/ethics-code)
3. Crowl, D.A. and J.F. Louvar, *Chemical Process Safety: Fundamentals with Applications*, 3rd Ed., Prentice-Hall, Englewood Cliffs, NJ, 2011.
4. Wikipedia: Texas City Refinery Explosion.
5. Wikipedia: Phillips Disaster.
6. Wikipedia: Deep Water Horizon oil spill.
7. Wikipedia: Texas City disaster.
8. Wikipedia: Fukushima Daiichi Nuclear Disaster
9. Wikipedia: List of Industrial Disasters.
10. AIChE Center for Chemical Process Safety, *Layers of Protection Analysis: Simplified Process Risk Assessing*, American Institute of Chemical Engineers, New York, New York (2001).
11. Engineering Equipment & Materials Users' Association: www.eemua.org .

12. Seborg, D.E., T.F. Edgar, D.A. Mellichamp, F.J. Doyle, *Process Dynamics and Control*, 3rd Ed., Wiley and Sons, New York, NY (2011).

17.10 Additional Terminology

alarm overload - when so many alarms are occurring that the operator is unable to make sense of what is happening.

BPCS - Basic Process Control System, the regulatory and supervisory control on a process.

ESD - Emergency Shut Down system.

HAZOP studies - hazard and operability studies, a qualitative evaluation of a process that includes an evaluation of the safety of the process.

Layers of Protection - an approach to improving process safety in which a series of methods are applied.

LOP - Layers of Protection

Management of Change - a systematic approach used to implement changes in operation of an organization.

MOC - Management of Change

probability of failure - the fractional probability that a component or system will fail within a specified time period.

reliability - the fractional probability that a component or system will not fail within a specified time period.

relief valves - spring load valve that open rapidly when the system pressure reaches a specified level.

risk -the combination of the frequency of occurrence of an incident with the associated harm of the incident.

rupture disks - thin disk that are flanged into a line and rupture at a specified pressure.

SIS - Safety Instrumented System, an independent control system that intervenes to counteract a safety incident.

Chapter 18

Batch and Discrete Process Control

Chapter Objectives

- Identify the key differences between batch, discrete and continuous processes
- Summarize process monitoring and control techniques used for batch processes
- Overview process monitoring and control techniques used for discrete processes

18.1 Introduction

To this point in the text, we have focused primarily on continuous processes operating at or near steady-state. As an example Figure 18.1.1 shows a endothermic CSTR that is operating isothermally. Note that feed is continuously added and that the product is continuously removed. As a result, this system would be expected to operate in a narrow operating range assuming that disturbances to the system are small. Therefore, the heat duty in this case, which is referred to as the **load** for this control loop, should be relatively constant. Because of relatively small deviations from setpoint, these systems can be analyzed and understood using a variety of methods from linear control theory (i.e., Chapters 4-7).

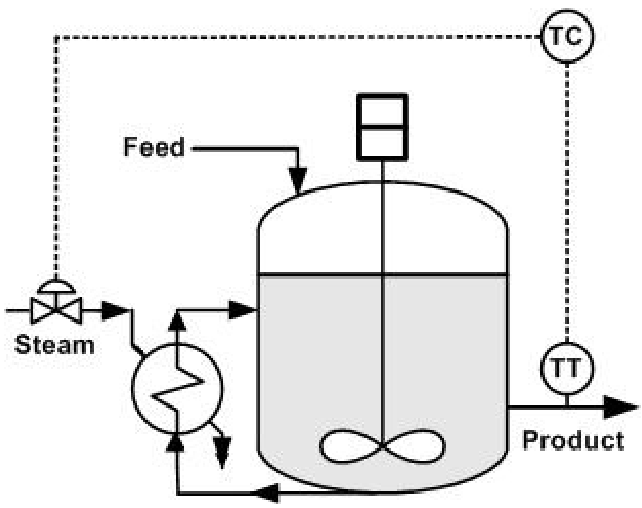


Figure 18.1.1 A continuous CSTR.

However, a significant percentage of chemical processes in industry, perhaps as many as 50% or more, are not continuous processes operating at or near steady-state. For example, the three general stages of a batch reactor are shown in Figure 18.1.2. Initially, the reactor is filled with reactant as shown in Figure 18.1.2a. Then the batch reaction is conducted at a specific temperature (Figure 18.1.2b). After the reaction reaches completion, the product is removed for further processing. Figure 18.1.3 shows the reactant concentration in the reactor during the reaction stage as a function of time for the reaction stage assuming a first-order reaction with constant temperature

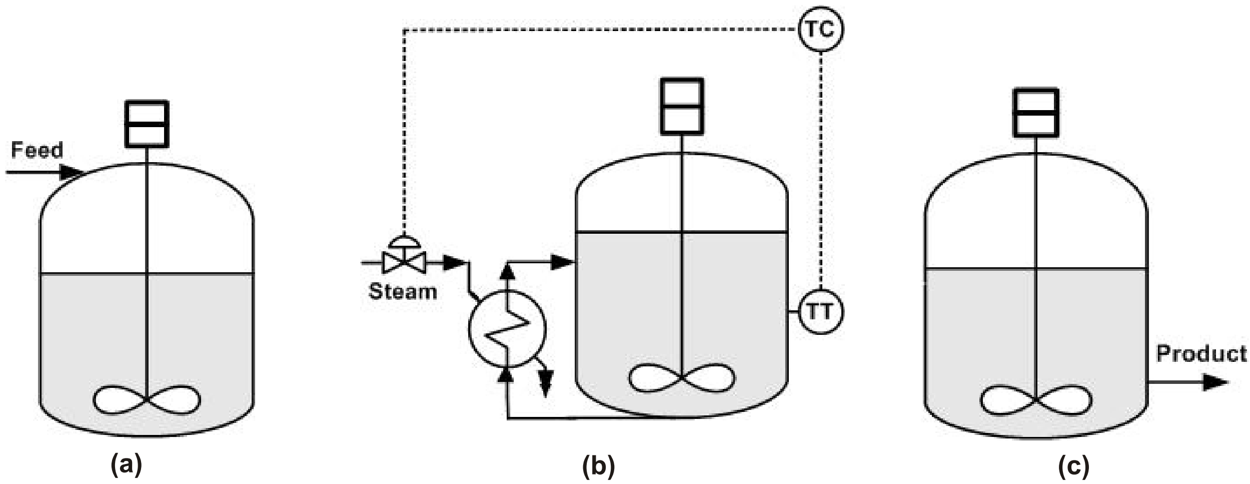


Figure 18.1.2 The sequence steps for a batch reactor: (a) filling phase; (b) reaction phase; (c) product removal phase.

conditions. Note that the reactant concentration exhibits an exponential decrease during the reaction. Therefore, initially the reaction rate is at a maximum rate requiring the largest amount of steam addition in order to maintain the reactor temperature, but near the end of the batch reaction period, dramatically less steam is required. Furthermore, the gain of this process as indicated by the slope of the concentration profile is large at the beginning of the reaction and much smaller later. This nonlinear gain with significant load changes present special problems for batch reactor control compared with control of continuous systems. Batch processes are especially common in such industrial sectors as specialty chemical, pharmaceutical and biotechnology. Smith[1] presents a thorough coverage of the topic of batch control. In addition, portions of chemical manufacturing processes often involve discrete operations (e.g., production lines inspecting large numbers of pharmaceutical tablets, processes controlling the size distribution of particles, processes manufacturing items such as medical devices).

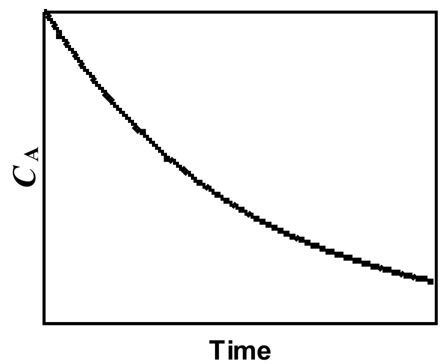


Figure 18.1.3 Reactant concentration versus time for reaction phase.

This chapter focuses primarily on the control aspects of batch processes with one section included on the monitoring and control of discrete portions of processes. Both of these types of processes can often benefit from different approaches regarding control than those used for continuous linear processes operating at steady-state.

18.2 The Application of Advanced PID Control to Batch Processes

The time-varying and non-linear characteristics of most batch processes frequently require the use of various advanced PID control techniques as part of the control strategy. A few of the more commonly used techniques are described as follows:

Adaptive Tuning. For some batch processes, the magnitude of changes in the process during the reaction stage are significant enough that adaptive tuning of the controller is appropriate. For example, in a fermentation process in which dissolved oxygen or the carbon source nutrient is being controlled, the cell mass typically changes over a thousand fold during cell growth and reproduction, therefore causing the demand for oxygen, nutrient feed and gas removal to change accordingly. For example, in a fermentation utilizing E.coli bacterial cells, which is used in making insulin, the cell mass doubles every 20-30 minutes during the several hour growth phase. Therefore, the optimum tuning parameters for the air flow control loop near the beginning of a fermentation, when cell concentration is very low, are different from the optimum tuning parameters at the end of the cell growth phase of the fermentation, when cell concentrations may be thousands of times greater. In the case of fermentations, it is true that such processes normally change slowly enough, even during active cell growth phases that one set of tuning constants could result in acceptable operation for the whole process. However, if process upset events such as loss of fermentor aeration or oxygen feed or agitation were to occur, the response to dissolved oxygen control would be far more catastrophic when cell mass is high compared to the beginning of the fermentation when cell mass is low. That is, a temporary loss of air supply during the first part of a fermentation would only cause a slow drop in dissolved oxygen concentration, which allows significant time for operator response and the process to recover, whereas the same loss of air when cell mass is high might immediately drop the dissolved oxygen to zero (due to very high oxygen demand), causing immediate changes in cellular metabolism (from aerobic to anaerobic) and accelerated cell death. So adapting tuning parameters such as increasing gain or reducing the integral reset constant as a function of increasing cell mass can improve the ability of the process to adequately recover from process upsets as the cell mass in the process increases.

Split Range controllers. Another reason that more advanced control techniques are sometimes needed for batch processes with significant load changes is that a far greater portion of the control valve's range is often needed with such processes. Most control valves are non-linear when considering their overall range. This is not usually a problem for continuous processes where the valve normally operates near the middle of its range. However, for valves which need to sometimes operate in the near closed or near open positions (which is not suggested as a good practice), the non-linearity of the valve at the outer edges of its range of motion, as well as its trying to operate near its fully open or fully closed position, can have a negative effect on process control. If a valve is operating at times near a full-open portion of its range, one tempting solution is just to use a larger control valve. However, doing so will often negatively influence the control loop's ability to tightly control a process at the desired setpoint. For example, in a fermentation, control of temperature is critical, usually requiring temperature to be within 0.25°C of setpoint . Using an oversized valve to ensure all possible loading conditions, can be accommodated, but can complicate efforts to maintain the temperature process variable at a precise setpoint due to problematic valve behaviors such as hysteresis, deadband, and stiction.

One solution, used in some applications, is to employ split range controllers (Section 14.4). That is, when using one control valve, it is a best practice to specify and use control valves such that they operate almost all the time between 20 % and 80 % of their range. When using split range controllers, the output of the controller is split and sent to two control valves. In most split range applications, the controller adjusts the opening of one of the valves when its output is in the range of 0 to 50% and the other valve when its output is in the range of 50% to 100%.

Control Loop Setpoint Ramping. A common software function utilitized in the control of time varying batch processes is to employ ramping of setpoint changes. The normal practice of making instantaneous setpoint changes or changes in discrete steps can sometimes place unnecessary mechanical forces on the equipment involved (e.g., motors, compressors) and may unnecessarily consume additional energy for periods of time when such additional energy is not needed.

As an example, the first third of many industrial fermentations is characterized by cell mass growing and multiplying, increasing over a thousand fold during this time. Most fermentors are equipped with a sterile air supply and mechanical agitation to supply oxygen to the growing cell mass and convert the oxygen from a gas to a "dissolved" state so that it can be utilized by the cells. As the fermentation begins, very little air and agitation are required, as the number of cells is small. As the cell mass increases, corresponding increases in air and agitation are needed to provide the necessary dissolved oxygen. So, an optimum economic control strategy, to avoid sudden changes to compressors and agitation motors, as well as only consume energy needed by the process, is to ramp up the agitation and aeration control loop setpoints in some way, e.g., linearly or exponentially. An exponential increase, if practical to achieve in the process control system, is theoretically preferential in this example as the increase in cell mass during the growth phase of a fermentation is exponential. Most modern process control systems contain embedded functionality to allow for easy configuration of ramp functions for applications such as ramping control loop setpoints.

Batch End-Point Control. A common process control challenge with many batch processes is determining when to end the batch, sometimes referred to as the batch end-point. Examples include:

- Ending operation in a batch chemical reactor when the reaction rate creating desired product is slowing while production of undesired side products is increasing.
- Ending a vacuum drying or freeze drying operation at the targeted value of product dryness. Remember that direct on-line measurements of moisture are not always practical to obtain and over drying can cause undesirable changes to certain product properties.
- Ending a fermentation when product production is decreasing and cell death is increasing.
- Determining when cells being grown in a seed bio-reactor are ready to be used in inoculating a fermentation.
- Determining the end-point of a batch crystallization process

As noted in the list above, running a batch too long can sometimes cause undesired side effects. In other cases, the process can last several days in duration, so end-pt. determination includes the desire to reduce cycle time, if practical to do, if that ends up increasing plant capacity (i.e., throughput). An optimization algorithm may be needed to deal with competing factors.

Due to significant lot to lot variability, using a fixed process duration time to end a batch lot can often be suboptimal and undesirable. It maybe satisfactory for some lots, but will be too early or too late for others.

Also, direct on-line measurements of critical parameters are sometimes not available or highly accurate and, therefore, indirect means (e.g., virtual sensors or models) are needed to determine end points.

18.3 Descriptions of Typical Batch Processes

Batch manufacturing processes are usually used to efficiently produce small quantities of chemicals, often at kilogram scale, including ones that need to meet unique customer requirements. Continuous processing, on the other hand, is typically used for producing large volumes (i.e., on a scale of tons) of chemicals (e.g., products derived from petroleum, such as gasoline and high volume chemical intermediates, such as acetic acid), designed to meet the needs of a broad range of end users.

A batch process is typically one that consists of several steps/phases with at least some of these stages that are time varying. Typical batch processes contain major stages of operation that are not linear and/or do not operate at or near steady state. As with all processes, including continuous processes, batch processes have start-up and shut-down steps/phases. However, processes described as batch usually contain several additional steps, occurring in sequence. Batch processes usually operate within a smaller time frame (hours or days) vs. continuous processes that often operate for up to weeks or months or years.

As an example, the steps involved in a typical industrial batch fermentation process are as follows.

- Clean the fermentor (e.g., automatically using a "clean in place" system).
- Charge the fermentor with water and initial nutrients to be used later by the culture of microorganisms.
- Steam sterilize the reactor (e.g., maintain a fermentor at 121^{o}C for 20 min.)
- Run tests to calibrate the Dissolved Oxygen (DO2) sensor and check the accuracy of the on-line pH probe.
- Inoculate the fermentor with a culture of living cells, which is typically transferred in a sterile manner from a separate batch process.
- Control the process for the first third of the fermentation phase while cells grow and multiply from an initial low cell mass concentration to a high cell concentration.
- Control a short induction process in which either the temperature of the fermentor contents is changed or nutrients are added to induce the cell mass to shift from a cell growth paradigm to one of producing the desired product (e.g., antibiotic, insulin).
- Control the last 2/3 of the process fermentation phase while cells produce the desired product.
- Terminate the fermentation and direct the contents of the fermentor to a separate system to harvest the cells (i.e., kill and break apart).

Note that there are a few processes known as semi-batch (see Example 3.7), which have attributes of both continuous and batch processes. In certain cases, they have dynamic start-up, shut-down, and perhaps one or two other time varying steps, but can include a significant amount of near steady state operation (e.g., separating cellular debris from broth in a continuous centrifuge following fermentor harvest operations).

A batch process typically produces a finite amount of product corresponding to what is called a "lot." All materials used in making a "lot" of product are also identified with their own "lot" numbers. Lot numbers are identified and maintained in manufacturing records and used for activities such as batch quality assurance reviews and as a basis of product recalls from the market if needed. The batch lot number, or a number allowing access to the lot number, is typically marked on the packaging label of products sold to customers.

In addition to start-up and shut-down steps, most batch processes contain one or more additional steps that are time varying. For example, the reaction rate (including heat generation) and reactant concentrations in a chemical reactor will change as raw materials are consumed as equilibrium is approached or as temperature changes. Also, certain attributes of life science processes (e.g., cell mass, oxygen demand, carbon dioxide evolution) will change in time as organisms grow and multiply. As a separate example, the composition of the output of industrial chromatography separation columns will change over time as the components of the complex mixture supplied to the column make their way through the column at different elution rates.

The time-varying nature of batch processes, which is often manifested by changing loads for control loops as well as process nonlinearity, often leads to use of advanced control techniques, such as split range controllers and advanced PID control algorithms (e.g., gain scheduling).

Many, if not most, batch processes in industry are semi-automated, rather than completely automated as are many continuous processes. This is primarily because the manufacture of a batch lot of product often requires certain manual operations to be integrated into the operation. For example, a batch process may require an operator to charge a reactor vessel with ingredients, or take a sample of the contents of a vessel and have it assayed and verified before alerting a process control computer (e.g., via a manual entry on a computer operator console) to proceed to the next step. Also, batch operation steps may involve multiple pieces of equipment and so, particularly in a campaigned multi-product facility, an operator may be needed to manually connect the appropriate unit operations equipment together, in accordance to a master schedule or to a prompt from a running batch recipe before processing of a current batch can proceed.

A common example of a semi-automated process is the steam sterilization of a bioreactor. A typical bioreactor has many pipes connected to it, used to supply air, acid/base addition for pH control, and nutrient feed lines, and also ones used to exhaust gases and enable liquid broth sampling. These lines, in addition to the bioreactor itself, must also be sterile to avoid being a source of contamination due to undesirable bacteria of the mixture that will undergo fermentation. It is sometimes not practical to put temperature sensors on all these pipes and so while the computer control system is heating up the overall system, operators typically check the above listed pipes with “temp sticks” or other devices to confirm that all the piping attached to the bioreactor exceeds a minimum threshold value. An operator’s entry to the computer indicates that all the attached piping is up to or above required temperature and so is ready for the next step in the sterilization process to proceed, which in turn, allows the sterilization timer (or time-temperature algorithm) for the bioreactor to commence.

Many automation engineers believe it is much more challenging and difficult to properly and consistently operate a semi-automated process (i.e., integrating manual and automated operations) than it is for a completely automated process. Note that the sources of variability in a semi-automated plant are more likely to be the manual portions than the automated portions.

Example 18.1 A Crystalization Process

Problem Statement. Figure 18.3.1 is a schematic of a crystalization process. Describe this process, overview its operation and indicate how setpoint ramping is used in its operation.

Solution. A crystallizer is a very common separations batch unit operation for both the CPI and bioprocesses. It works by precipitating out, as crystals, what initially is a nearly saturated solute in a solvent, and causing crystals to form by lowering the temperature of the mixture. The solution initially becomes supersaturated and then crystallization begins since solubility of the solute in a solvent decreases with decreasing temperature.

The use of crystallization is especially common in the pharmaceutical industry as the last step in the first phase of manufacturing produces what is known as the API (active pharmaceutical ingredient). API, in its highly structured crystal form, is subsequently combined perhaps after a milling operation to change particle size distribution with other compounds (known as excipients) in the formulation phase of manufacturing to produce the final oral dosage form (e.g., pills, capsules) of the pharmaceutical that is sold to customers. API crystals can have various properties (e.g., particle size distribution) and structures (e.g., polymorphs), which are critical product quality attributes (e.g., affecting product solubility, dissolution rate, and stability), which are determined, in large part, by the control of process parameters during the crystallization process (e.g, uniformity of mixing and temperature control). For example, the amount of supersaturation that takes place within the crystallizer reactor (largely determined by temperature conditions) will directly impact the particle-size distribution of the final product. An inadequate level of control of temperature during supersaturation and subsequent cooldown can lead to excessively large crystals, or undesirably fine ones.

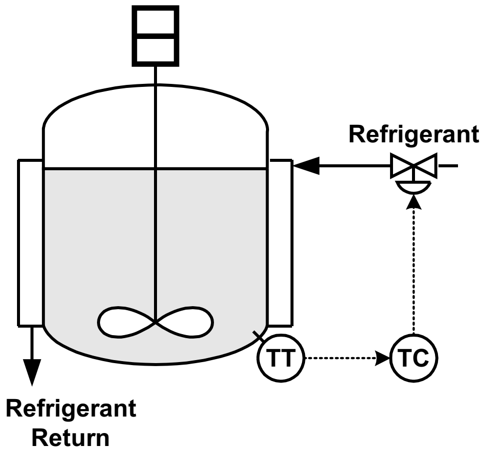


Figure 18.3.1 A schematic of a crystallizer with a temperature control loop.

While the time varying temperature profile that optimizes the generation of the proper form and size distribution of crystals is usually not linear, a decreasing linear ramping temperature setpoint is usually good enough as a starting point for process control analysis purposes. Therefore, the setpoint for the temperature controller in Figure 18.3.1 would be ramped down in order to initiate the precipitation process. A typical ramp rate for a crystallizer in the pharmaceutical industry is 0.5 ^{o}C/min.

Example 18.2 A Vacuum Drying Process

Problem Statement. A vacuum drier is a common separations batch unit operation for both the CPI and bioprocesses. It allows for the drying of hygroscopic (i.e., materials that tend to absorb water) and/or heat sensitive materials (e.g., many foods, pharmaceuticals) at lower temperatures than would exist in a more conventional dryer. Vacuum drying is based on the principle that at lower pressures liquids boil at lower temperatures. A schematic of one form of vacuum drier is shown in Figure 18.3.2. Typical batch processing times vary from a few hours to a few days. For a typical application, vacuum dryers operate in the pressure range of 0.03 to 0.06 atm. The boiling point of water in this pressure range is 25 to 30 deg. C.

The heat transfer in a drying operation is governed by the equation: $Q=UA\quad T,$ where Q is the heat transfer rate, U is the heat transfer coefficient, A is the surface area of material being dried, and T is the temperature difference between the liquid's boiling point and the temperature of the material being dried (i.e., the temperature inside the vacuum dryer).

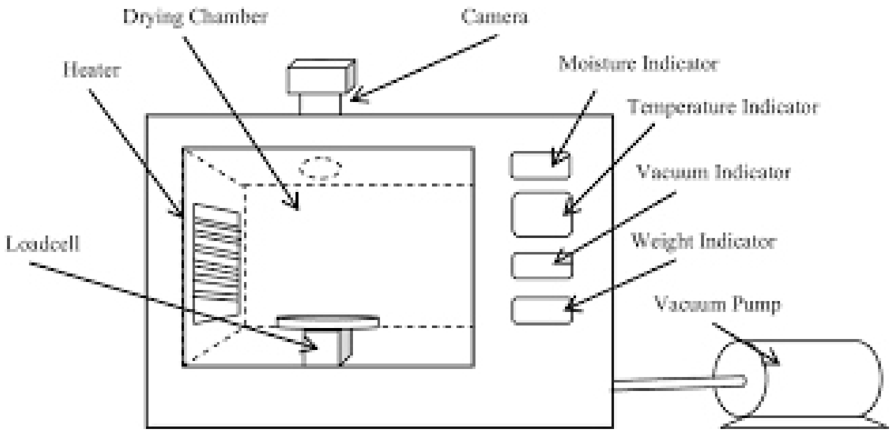


Figure 18.3.2 An example of a typical vacuum dryer.

1. Which of the above parameters in the equation for Q are relatively fixed and which is most likely to be amenable to adjustment to increase Q?

2. What changes to the operation of a vacuum dryer can be made to reduce the cycle time ?

3. What approaches might be employed to heat the vacuum dryer to a higher temperature (but still under the thermal instability temperature of the contents being dried)?

4. One of the most important process control objectives for many batch operations is determining the end point of the batch. For example, over drying a material can negatively alter certain properties of the material. What measurement options can you use to determine when to terminate a vacuum drying process?

Solution.

1. Most often, the material's properties and the dryer type effectively establish the U and A values for the process. Therefore, T is the parameter for which can most practically be used to maximize Q.

2. Q can be increased by increasing the temperature inside the vacuum dryer to a level still safely below when where thermal instability results in the material being dried. Also, higher vacuum pressure (i.e., lower pressure) can be used, which drives down the boiling point of the material further.

3. Heating of the contents of a vacuum dryer can be accomplished via controlling the temperature or flow of heating media to the jacket, heating vessel shelves holding the product, or microwaving.

4. Several methods are available to establish an end point for drying. [Note that a fixed batch time duration is normally not a satisfactory way to end a batch due to normal batch-to-batch variability (e.g., variable starting moisture content of the material to be dried).]

a. If the contents are on a weigh scale, drying can continue until the net weight plateaus.

b. A characteristic temperature curve may be established based on multiple batch runs, showing, for example, that the temperature increase late in the process when the evaporative cooling of unbound water is no longer significant (i.e., only bound water remaining). So, a point on the rising temperature part of the curve can sometimes be used as an endpoint.

c. An excellent technique is to interface a mass spectrometer to the vacuum dryer and measure a solvent to inert gas ratio of the gas in or leaving the dryer. For example, the water to argon ratio (mass 18/40) or water to nitrogen ratio (18/28) will drop as drying continues and the water concentration, as a percentage of gas in the dryer, goes down. Other analytical instruments (e.g., NIR spectroscopy) can also monitor extent of drying in real-time.

Example 18.3 A Freeze Drying Process

Problem Statement. Freeze drying is a unit operation commonly used for processing food and pharmaceutical products. Freeze drying is a low temperature process in which the product is frozen and the system pressure is reduced causing the water/solvent to vaporize (i.e., sublimate) by adding dry heat. During the freezing process, the material is cooled below the triple point of the water/solvent so that sublimation is guaranteed.

The first stage of the process is the freezing stage in which the temperature of the material is reduced below the triple point, usually in the range of -40 to -60 deg. C. This is followed by creating a deep vacuum in the vacuum dryer vessel, i.e., below the triple point of the solvent being removed (usually water). Note that the freezing phase is the most critical step in the freeze-drying process, as the freezing method impacts the size of the crystals being formed and, therefore, impacts certain final properties of the product, such as product stability and dissolution rate. Rapid cooling normally results in small ice crystals and slower cooing results in larger ice crystals. The freezing method also affects the duration of the freeze-drying cycle. Therefore, tight control of the temperature-time curve during freezing is important. The second stage requires the addition of dry heat to the material which increases the evaporation of the unbound water/solvent in the material. After the unbound water is removed from the material, the third stage begins as the temperature is further increased and the bound water/solvent begins to be removed.

Figure 18.3.3 is a schematic of a typical freeze dryer. Note that both heating and cooling are supplied to the product shelves in order to adjust the temperature of the drying chamber.

1. What is the primary difference between vacuum drying and freeze drying (both of which occur under vacuum)?

2. What are possible options in trying to reduce cycle time?

3. One of the most important process control objectives for many batch operations is determining the end point of the operation. For example, over drying material can negatively alter certain properties of the material. What measurement options can you think of as to when to terminate a freeze drying process?

Freeze Dryer

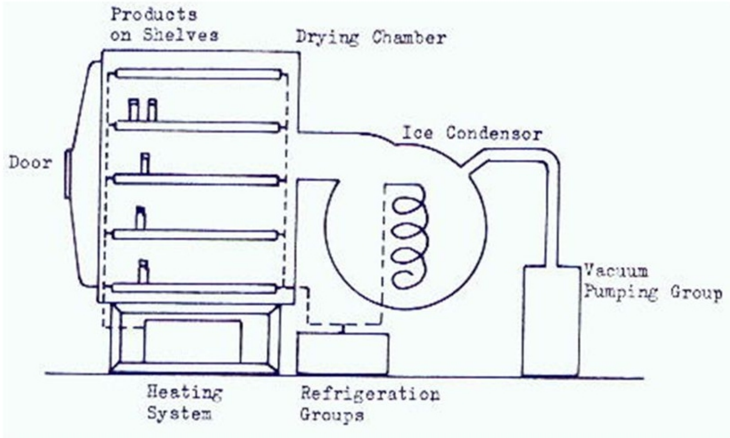


Figure 18.3.3 An example of a typical freeze drying system.

4. What might be the biggest source of variability with the freeze drying process?

Solution.

1. Vacuum drying boils the liquid in the drying material, so it involves changing the liquid phase of the water/solvent to the gas phase. Freeze drying utilizes sublimation, in which the solid form of solvent is created which then passes directly to the gas phase, avoiding the liquid phase.

2. The most dramatic way to reduce cycle time is to raise the shelf temperature as high as practical (after freezing) while maintaining sublimation conditions. It is estimated that each 1 degree C. increase in shelf temperature can reduce cycle time by over 10 %. Note that reducing temperature more quickly at the beginning of the process is not usually a good way of reducing cycle time because the benefit will be small and the risk of negatively changing certain final product characteristics (via changes in ice crystal size) could be significant.

3. One method commonly used in end-point control is based on placing thermocouples in various places in the product material (or in vials). Wireless thermocouple technology exists and can be helpful. The product temperature approaching the shelf temperature indicates that the associated phase of drying is near an end-point.

Another option is a dew point sensor. For example, the point where the dew point starts dropping indicates that the sublimation is essentially complete, i.e., the gas composition is changing from mostly water vapor to nitrogen. This measurement is especially useful when determining the end of primary drying (i.e., removal of unbounded water), after which temperature can be raised to promote secondary drying (i.e., removal of bound water). Note that fixed batch time duration is normally not a satisfactory way to end a batch due to normal batch-to-batch variability for a various reasons (e.g., variable starting moisture content of the material to be dried).

4. The biggest source of variability is usually the lack of temperature uniformity within the freeze dryer chamber and for material on chamber shelves, leading to non-uniform freezing rates in different parts of the drying chamber. Insuring the freeze dryer walls are well insulated (or even jacketed) can help provide a more uniform chamber temperature environment thereby enabling more uniform temperatures in the drying chamber.

18.4 Batch Process Data Recording and Analysis

Many, if not most, commercial process control systems offer the ability to log process data, but use a limited format, usually including a simple cryptic identifier (e.g., TIC101) and a calendar timestamp. This is satisfactory for many continuous processes, but is too limiting for efficient logging and analysis of batch process data. Users responsible for batch processes normally desire to call up their data by lot number, and perhaps even by the equipment used and batch step or phase, and then see their data as a function of relative time (i.e., time since the beginning of the batch). So data records ideally need to contain fields for process lot number, process step or phase, equipment tag, and specific data label. If not in the data record header (i.e., tag) itself, means need to exist in the data historian to create this information (e.g., relative time). An example of a recorded process variable (oxygen update) for a fermentation is shown below (Figure 18.4.1). The *x*-axis is relative time, i.e., the time since the start of the batch. This allows users to relate the oxygen uptake (and cell mass by inference) to the time into the batch. Use of calendar time, as exists in most commercial systems, would be less useful in analyzing the batch run as the significance of a particular value of oxygen uptake has value primarily in knowing when it occurred relevant to the beginning of the batch. Note also how time varying the oxygen uptake is, symbolic of the entire time varying nature of the batch process, which is typical of many batch processes.

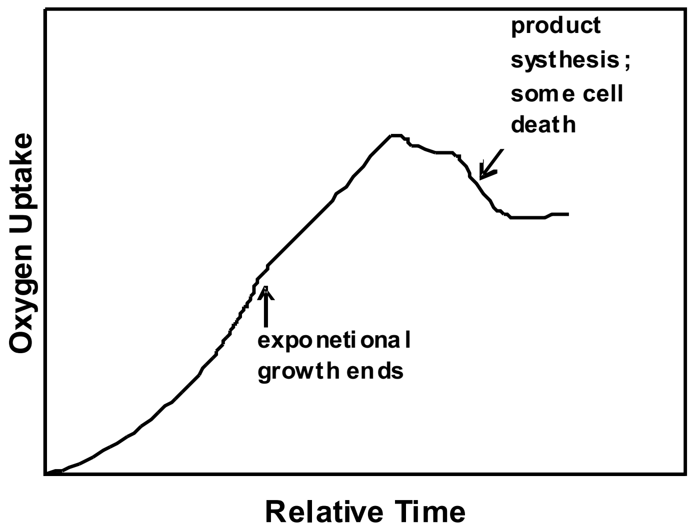


Figure 18.4.1 An example of a typical oxygen uptake for a fermenation reactor

Note that when FDA or internal auditors review the performance of a batch plant, they will almost always request that data from specific batch lots be provided for their review. Also, when company scientists and quality control employees compare batch records to one another, they do so by batch lot number and relative time so that trend plots can be easily overlayed with the origin of each plot always being the start of the batch rather than different calendar times. Also, formal batch records are almost always organized by batch lot number. Also, any product recalls are pursued as a function of certain batch lot numbers. So, the ability to record, store, and retrieve batch data, including process control parameters, by batch lot number is very important.

18.5 Batch Control Standards (ANSI/ISA 88)

Partly due to the semi-automated nature of many batch processes (i.e., need to coordinate with manual operations) and also partly due to frequently occurring complex aspects of many batch processes (e.g., scheduling constraints), it is appropriate to organize the control of batch processes into different states. States include Stop, Start, Run, Pause, Hold, Idle and Abort. That is, during many batch processes, conditions may be encountered such that the process needs to be temporarily paused and put on hold, (i.e., placed in a different state). In anticipating such situations, process control software needs to be organized and configured to enable a smooth transition between states. For example, putting a process in a Hold state may require changing setpoints of certain controllers, taking some controllers out of computer/auto mode, putting certain software timers on hold, putting certain valves into predetermined positions, and/or suppressing certain alarm functions. So control software benefits from being organized and configured as a function of process state.

To facilitate the configuration of process control software into states and to accommodate other needs, the International Society of Automation (ISA) developed the American Standard on Batch Control (ANSI/ISA 88 [www.ansi.org])[2], which has since become an international standard (IEC 61512 [www.sis.se])[3]. While developed specifically for batch processes, the standard can also successfully be applied to continuous and discrete processes that require significant flexibility.

The ANSI / ISA 88 standard indicates how process control software should be structured to accommodate batch processing. For example, it defines a physical model (i.e., representing hardware) of a plant in terms of Cells, Units, Equipment Modules and Control Modules. The highest levels of the Physical Model are the Enterprise (i.e., the company), Site (i.e., the plant location) and Area (i.e., the part of the plant). Within a process "cell", a product is made. A "unit" might be a reactor or a mixer.

Secondly, the standard defines a Procedural Model. This includes the "Recipe" which contains the instructions on how a particular product should be made. The Recipe includes the "formula" which contains the amount of each raw material to be used in executing the batch. The "recipe procedure" can be broken down into unit-procedures, with a unit-procedure then broken into operations, and an operation into phases. The role of procedures is, in part, to trigger actions handled by various physical modules.

As a result of organizing process control software per ANSI/ISA 88, a recipe can be written without knowing anything about how the equipment logic is implemented. For example, a procedure might change a temperature setpoint, but doesn't need to know any detail about how the relevant control module application will accomplish this. Also, the equipment logic can be implemented without knowing anything about how it will be used in different recipes. The two activities (i.e., physical equipment and procedures) can be completely separated. So, an important aspect of ANSI/ISA 88 is modularity. This makes process equipment and procedures developed in one application reusable in another application, resulting in savings of time and money. Readers can refer to the above standards to obtain more information on how to structure software for batch control.

18.6 Batch Process Alarm Management

Alarm management general principles are covered in Section 17.4. However, the nuances of alarm management specific to batch (and discrete) processes is important enough that the International Society of Automation (ISA) developed and published a technical report (ANSI/ISA 18.2 - TR6), “Alarm Management for Batch and Discrete Processes”, on this subject in support of the more general national alarm management standard (ANSI/ISA 18.2)[4]. A few of the important topics specific to batch process alarm management, discussed in TR6, are as follows:

Alarm Configuration as a Function of Process step. As noted previously, batch processes are typically a sequence of several steps. Often, individual alarms are important for some steps but not others. Therefore, when specifying values for specific alarm attributes, usually accomplished by an alarm rationalization team, separate values for each alarm need to be determined **for each batch step**. This includes when to enable an alarm and/or whether to suppress it for particular process steps. For example, in a fermentation batch operation, alarming pH is critically important for those steps during which living cells are growing and producing product in the fermentor. However, there is no need to have the alarm enabled during vessel cleaning, make-up, sterilization, and cell harvest activities. Therefore, the alarm should be configured as suppressed for those steps in which it is not needed, so as to avoid sending nuisance alarms to operators. Another example might be a pressure alarm on a chemical reactor, whereas pressure is monitored and alarmed at a low priority during the tank make-up step, but has its priority increased when the gas producing chemical reaction takes place in the following step. Configuring appropriate alarm priorities is important since, in the event of an abnormal situation, operators are trained to always deal first with the highest priority alarms.

Record tag information. The content discussed in Section 18.4 for batch process data recording also applies to alarm records. That is, it is helpful and efficient if alarm record tags include the batch lot number, batch step/phase, and a relative time stamp. This is information over and above what is normally expected for continuous process alarming. Relative time is important with regard to the importance of an alarm activation and the record should indicate when during the process of the batch the alarm occurred, rather than some arbitrary calendar time.

Time-varying alarm setpoints. While some process variables may be controlled at a specific value during a batch step, other process variables may not be. Then a challenge for engineer’s automating batch processes is how to alarm time-varying process variables. The traditional method used for continuous processes of comparing a process variable to a fixed alarm setpoint will not work for these cases. Some means of logically configuring the alarm setpoint in a time-varying way, in keeping with the normal expected behavior of the process variable, is needed.

One method used in fermentation batch processes in manufacturing plants by a leading pharmaceutical company is illustrated in Fig. 18.6.1. The trend plot follows the progress of dissolved oxygen in an E.coli fermentation making a protein, for which the dissolved oxygen normally starts out at near 100 % saturation at the beginning of the fermentation when cell concentration is very low, and then trends lower over time as the cell mass builds up, consuming ever more oxygen. If not for an increase in periodic aeration and/or agitation rate (one of which is shown at 7 hours into the batch), the dissolved oxygen would eventually approach zero. Further, it is noted that life science processes typically have higher variability than most non-life science processes. In the case of fermentations, the progress of oxygen consumption, and hence dissolved oxygen concentration during the first part of the fermentation, depends on how many cells the batch run started out with (which can be variable), as well as the viability of the cells (which can be variable). So the trend plots of dissolved oxygen for different batches are not identical quantitatively, but they are similar. In preparing to generate a plot such as shown in Fig. 18.6.1 (which is a copy of a real-time trend on an operator console), the dissolved oxygen trend plots for 25 successful historical batches were combined and averaged together (off-line), producing a time varying average value as well as computed +/- 2 standard deviation time varying lines. The +/- 2 standard deviation lines (shown in red) were than ported to the real time operator display in the control system's Human Machine Interface (HMI) for dissolved oxygen and became a permanent (background) part of dissolved oxygen trend plots. Thus, the upper and lower standard deviation plots were displayed on top of the current run's dissolved oxygen time varying value. So, when the current batch run's dissolved oxygen value drifts outside the historical +/-2 std. dev. red lines, this is an indication that the current dissolved oxygen represents an abnormal situation and an alarm is generated. The usual cause of such an alarm is when the fermentor becomes contaminated such that excess cell mass (e.g., from contaminating bacterial organisms) consumes additional oxygen, thus driving the dissolved oxygen abnormally low. A dissolved oxygen value abnormally high usually means excessive cell death is occurring, cell viability is low, or that a raw material (cell nutrient) was inadvertently left out of the original tank make-up.

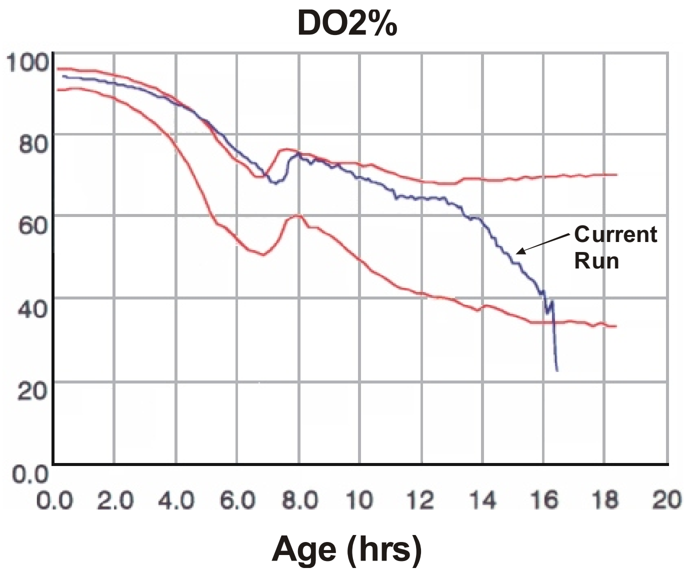


Figure 18.6.1 An example of a typical oxygen uptake for a fermenation reactor

Note that the ability to create such an operator display as described above is not simple and not available as a standard feature of most, if not all commercial automation systems. Such plots can be a custom application generated with the help of a real-time commercially available expert system. By comparing a plot of the current behavior of a batch to an earlier successful batch, the operator can many time identify when a batch has gone bad. Figure 18.6.1 demonstrates this point. Note that the current run starts to deviate from the form of the previous runs after about 14 hours into the run. The point here is that any special process monitoring and control requirements, such as creating the comparative plots should be identified when the System Functional Specification document is developed so that adequate time exists to work with vendors or company engineers to develop any custom applications or features of the system. The time to identify the need for special custom features is not when the vendor automation system is delivered and configuration work begins. By then, its usually too late.

18.7 Discrete Process Control

Another class of chemical manufacturing processes (or portions of processes that are primarily continuous or batch) are discrete operations. Discrete operations are often associated with different forms of sensors and process control than what is typically utilized for most continuous processes. Discrete operations include those dealing with individual finite entities (e.g., items, packages, pills, sampling systems, analytical assays) and are often associated with different forms of sensors, actuators, and control algorithms than what are typically utilized in most continuous and batch processes.

Examples of discrete portions of processes include:

- Ones where particulate (including biological cells) of certain size, concentration, or size distribution are required, for which sampling systems and on-line (or off-line) analytical systems are used to determine the desired measurements.
- Assembly or inspection lines where thousands of units per hour or other discrete products (e.g., pharmaceutical tablets, filled medical syringes) are being made and/or inspected.
- Inspection of final product packaging (including any labels, barcodes, RFID tags). Note that if identification scans indicate an error, a mechanism exists to reject the invalid item or packaged product.
- Process sampling systems, often requiring an automated sequence of discrete control operations, to properly collect, prepare, and transport samples from processes to on-line or at-line analyzers (e.g., on-line HPLCs).

In exploring selected applications in more depth:

• In some industries (e.g., pharmaceutical, biotech), FDA regulations require that companies prove that representative samples of their products, for every lot, comply with quality attribute criteria for purity, strength, and identity. Other criterion (e.g., dissolution time of powders) may also need to be satisfied. Such proof, which is a form of process control, often requires complex measurement and control systems, which can often be different from those used in common chemical processing unit operations. For example, mass spectrometers, HPLCs, and/or particle size analyzers are sometimes utilized.

• In bioprocesses, inoculating a fermentor or determining that a process can proceed to the next step may involve an analysis of living microscopic cells (a form of discrete entities) in a broth sample with the analysis conducted by a camera mounted on a microscope with application software counting the number and size of cells on a specified area of the microscope slide. Note that a few commercial cell counters exist that use a flow cell format such that continuous sampling can be used for counting cells. Applications similar to cell counting exist in many downstream pharmaceutical operations utilizing granulation and milling unit operations that produce particles in a particular size range. A variety of instruments are available to confirm that the target size range has been achieved. Samples can be accommodated that are either in powder form or as particles in liquids.

• The sensors and control devices associated with discrete processes are often different than those traditionally found on continuous and batch chemical processes. For example, the control of manufacturing of items (e.g., packaging containers) may involve use of robotic devices and the need to confirm with proximity sensors the location of robotic arms and to confirm that intended mechanical actions of the robots have been completed.

Figure 18.7.1 Part of a discrete system for filling bottles with tablets.

• For inspection applications, some form of camera based vision system is often employed. Such sensors are not simple. Imagine the complexity of application logic within a vision system that has thousands of discrete products passing within its 2 dimensional viewing area per hour and must, in real time, determine certain physical, color, or other attributes of each product, perhaps compute certain statistics, and report the results to an external computer system associated with control algorithms and decision making.

The controllers for such applications are not the usual PID controllers found in chemical manufacturing. Rather, statistical algorithms are often used that take input from vision systems and determine the defect rate for the large quantity of units passing before it. Once a pre-determined threshold is exceeded, alarms go off and perhaps the process is automatically paused while manual troubleshooting is pursued. Figure 18.7.1 is a photograph of discrete system that is filling bottles with tablets. Figure 18.7.2 shows a vial inspection system in which a variety of inspection processes are applied and observed using multiple cameras.

Some discrete process applications involve either manually taking samples of the process and/or use of automated sampled systems, where the follow-up assay result occurs at some significant time after the sample is taken (analogous to “dead-time) which complicates the control problem.

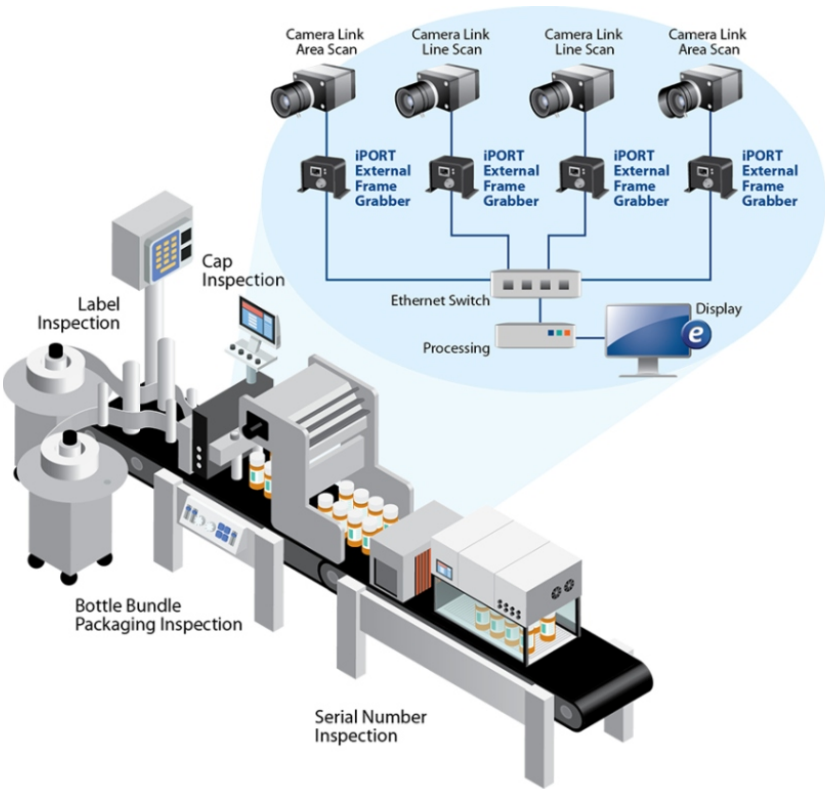


Figure 18.7.2 Schematic of vial inspection system.

The scope of this textbook limits the depth of coverage of discrete processes. Suffice it to say that discrete processes many times use a variety of less traditional (and usually more complex) sensors, analytical systems (with their sampling systems), control algorithms, final control devices and mathematical representations. Discrete processes often involve sampled data systems. One technique used to mathematically represent such systems involves Z transforms which can be considered as the discrete-time equivalent of the Laplace transform. Further information is available in the literature. Note that a process does not have to be all continuous, batch, or discrete, as many processes are a combination of the above.

18.8 References

1. Smith, C.L., *Control of Batch Processes*, Wiley and Sons, 2014.

2. ANSI/ISA 88, Batch Control (US Standard).

3. IEC 61512 Batch Control (International Standard).

4. ANSI/ISA 18.2, TR6, Alarm Management for Batch and Discrete Processes.

Chapter 19

Capstone Case Studies

Chapter Objectives

- To use heat exchangers, CSTRs and bio-reactors as examples to illustrate how process knowledge combined with the principles of process control presented in this text are important for the proper implementation of process control.

19.1 Introduction

This chapter is concerned with the control of heat exchangers, CSTRs and bio-reactors. The following material is based on various techniques developed in the text, viz., inferential control, selection of the proper mode for a PID controller and configuration selection, combined with the specific process characteristics of these unit operations.

General guidelines concerning recommended control configurations for these three classes of processes are presented in this chapter. The guidelines are based on the configuration selection analysis procedure presented in Chapter 15. Configuration selection should be based on coupling (for MIMO systems), dynamic response of the CV to MV changes and the sensitivity of the configuration to the key disturbances that affect the process.

19.2 Heat Exchanger Control

General Characteristics. Figure 19.2.1a shows a steam-heated heat exchanger for which the control objective is to control the outlet temperature of the process fluid being heated. Figure 19.2.1b shows a liquid/liquid heat exchanger (i.e, a heat exchanger that uses a coolant to extract heat from a hot process stream) for which the control objective is to control the outlet temperature of the process fluid being cooled. Steam-heated and liquid/liquid heat exchangers are the types of heat exchangers considered here. Both of these systems are self-regulating.

Section 13.3 introduced heat exchanger control by considering the control of a steam-heated heat exchanger. It was demonstrated that changes in the tube side flow rate of the process fluid caused significant changes in the effective process deadtime and process gain (see Table 13.2). As the flow rate decreases, the deadtime and gain both increase. As the number of passes used in the heat exchanger increases (i.e., the tube length increases), the

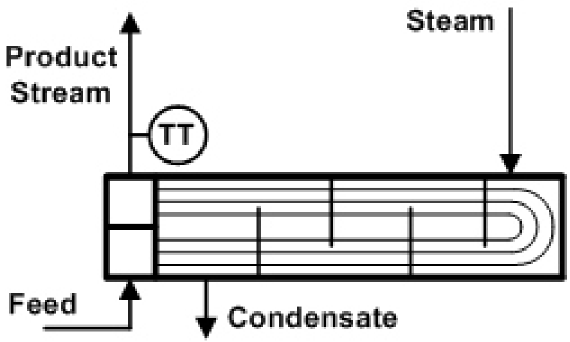


Figure 19.2.1a Schematic of a steam-heated heat exchanger.

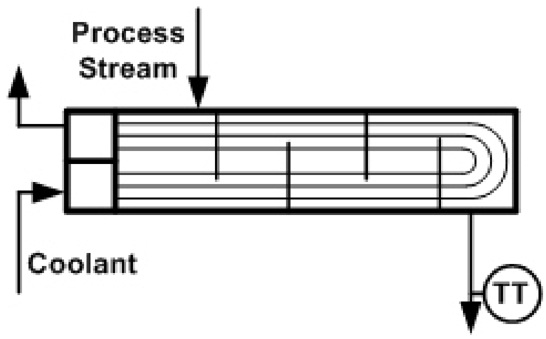


Figure 19.2.1b Schematic of a liquid/liquid heat exchanger.

deadtime increases. Industrial heat exchangers can have deadtime-to-time constant ratios approaching 4.0, depending on the heat exchanger design and the tube side flow rate. Figure 19.2.2 shows the steady-state temperature change of the outlet process fluid of a liquid/liquid heat exchanger as a function of coolant flow rate for different heat transfer areas. As the coolant flow rate increases, the magnitude of the process gain decreases continuously. Therefore, above a certain cooling water flow rate, no significant change in the product temperature results, i.e., the process gain approaches zero, which renders the process uncontrollable using the coolant flow rate as the MV. At a given flow rate, as the heat exchanger area increases, the magnitude of the steady-state gain increases.

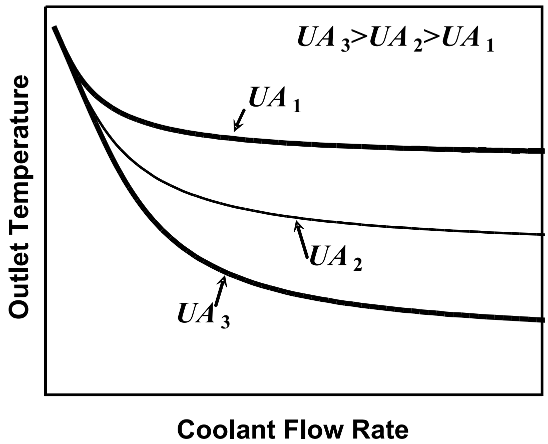


Figure 19.2.2 The static relationship between outlet temperature and coolant flow rate for several liquid/liquid heat exchangers with different heat transfer areas.

Steam-Heated Heat Exchangers. Even though these heat exchanger cases are SISO problems, the configuration selection problem is quite interesting and important. Figures 19.2.3a and 19.2.3b show two potential control configurations for steam-heated heat exchangers. The inferior configuration (Figure 19.2.3a) uses the steam flow as the MV for the temperature control loop while the preferred configuration (Figure 19.2.3b) uses the steam pressure in the heat exchanger as the MV. Steam pressure is the preferred MV because the steam pressure responds directly to changes in the pressure of the steam supply and to changes in the heat duty required to heat the process fluid. The steam pressure controller responds quickly to these disturbances and absorbs them efficiently. Moreover, a change in steam pressure inside the exchange quickly changes the overall process heat transfer rate. When the MV is the steam flow rate, a steam supply pressure upset results in a significant deviation in the CV from setpoint before corrective action can be taken. When the heat duty requirements change in response to changes in the feed flow rate or inlet temperature, the steam pressure inside the exchanger changes before the outlet temperature of the process stream begins to change. As a result, the steam pressure controller responds much more quickly to heat duty upsets than a steam flow controller. Because the steam pressure control loop responds quickly to the major upsets for a steam-heated heat exchanger, the control configuration based on steam pressure as the MV is preferred. In addition, for the preferred configuration, the steam pressure in the exchanger is a direct measure of

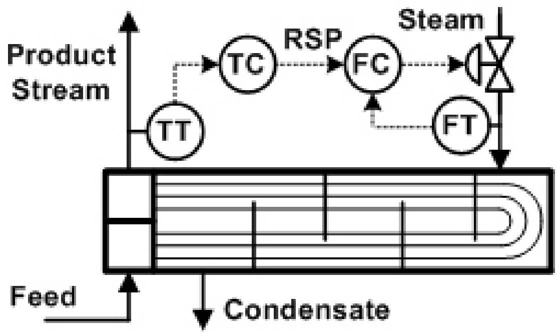


Figure 19.2.3a Inferior control configuration for a steam-heated heat exchanger, which is based on using steam flow as the MV.

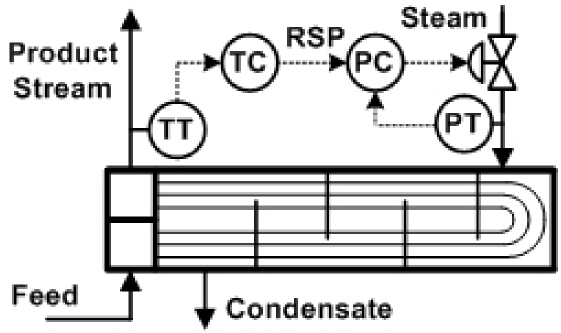


Figure 19.2.3b Preferred control configuration for a steam-heated heat exchanger, which is based on using steam pressure as the MV.

the degree of fouling, i.e., the higher the steam pressure for a fixed feed rate and inlet temperature, the more fouling that exists.

Figure 19.2.4 shows a modification to the preferred control configuration for the steam-heated heat exchanger. The condensate flow controls the level of condensate in the heat exchanger, which determines the heat transfer area available for condensing steam, which is the primary mode of heat transfer for this system. Because the condensate flow, instead of the steam flow, is manipulated, the steam pressure control loop is not as responsive as the steam pressure control loop in Figure 19.2.3b because the level responds slower than the pressure to changes in the respective valves. On the other hand, because the line size for the condensate is much smaller than for the steam feed, the size of the control valve is much smaller and less expensive. If the configuration using condensate flow as the MV is used, a steam trap is required downstream of the condensate valve. Choosing between the configurations shown in Figure 19.2.3b and Figure 19.2.4 involves a tradeoff between the capital costs and the control performance of these configurations. If control performance is a premium, the configuration shown in Figure 19.2.3b is preferred. Otherwise, economics will likely determine that the configuration shown in Figure 19.2.4 be chosen.

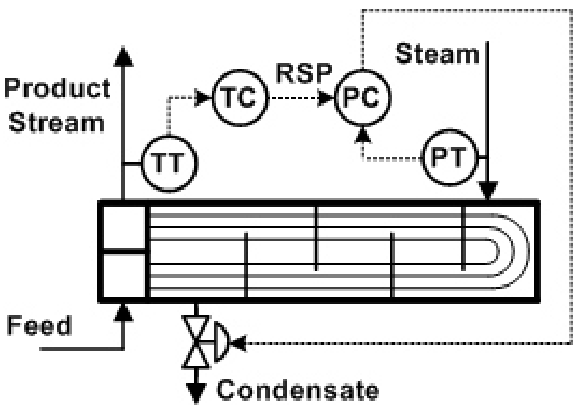


Figure 19.2.4 A modification to the preferred control configuration for the steam-heated heat exchanger in which the condensate flow is manipulated.

Because the gain of a heat exchanger tends to be nonlinear (e.g., Figure 19.2.2), the following scheduling of the PI tuning parameters provided by McMillan[1] can be used to compensate for a change in the flow rate of the process fluid

$$K_c \qquad K_{c0} \quad \frac{F}{F_0}^{\;2}$$

$$_{I0} \quad \frac{F}{F_0}$$

where K_{c0} and $_{I0}$ are determined for F_0. Note that the results presented in Chapter 13, which were obtained by tuning a PI controller for a steam-heated heat exchanger at different values of F (Table 13.2), agree with this scheduling function.

Liquid/Liquid Heat Exchangers. Figure 19.2.5a and 19.2.5b show two candidate configurations for controlling a liquid/liquid heat exchanger. One uses the coolant flow rate as the MV for the temperature controller while the other uses a bypass flow rate. When the coolant flow rate is used, the dynamic response of the process (changes in the temperature of the process fluid for changes in the flow rate of the coolant) is relatively slow because the dynamics of the heat transfer inside the exchanger and the transport delay for flow through the heat exchanger tubes largely determine the dynamic response of this configuration. It also suffers from the type of process nonlinearity described earlier in Figure 19.2.2 (variations in process deadtime and process gain). When feed bypass is used as the MV, the process dynamics are considerably faster, with low levels of process deadtime, because the dominant dynamic element for this process is the sensor. This configuration is also more linear. Changes in the bypass flow rate create changes in the feed flow rate to the heat exchanger, which cause disturbances for the temperature control loop. However, these disturbances occur at a relatively slow rate so that the temperature controller is able to efficiently absorb them. The feed bypass configuration has the advantage that the coolant flow rate can be maintained at a high value regardless of the feed flow rate, which can significantly reduce the tendency for fouling on the tube-side heat-transfer surface. As a result of these advantages, using the flow rate of the feed bypass as the MV for a liquid/liquid heat exchanger is generally preferred.

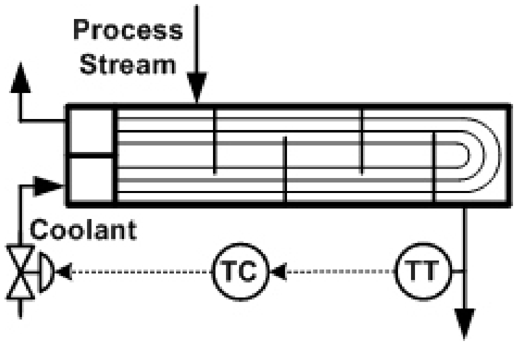


Figure 19.2.5a Inferior control configuration for a liquid/liquid heat exchanger with the flow rate of the coolant used as the MV.

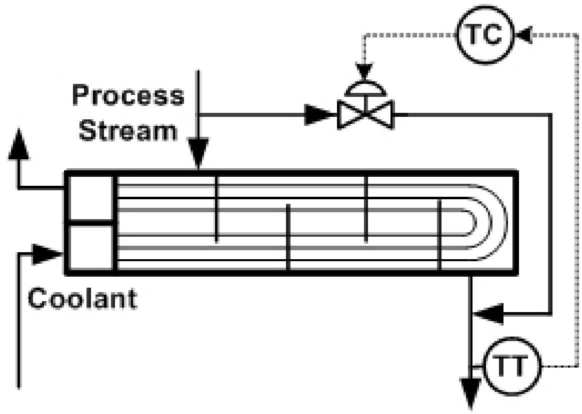


Figure 19.2.5b Preferred control configuration for a liquid/liquid heat exchanger with the flow rate of the feed bypass used as the MV.

Because the feed bypass configuration for liquid/liquid heat exchangers should have fast dynamics and very low deadtime-to-time constant ratios, derivative action should not be required and the tuning of this loop should be relatively straightforward. The reader is referred to McMillian[1] for more specific details on the application of temperature control to both steam-heated and liquid/liquid heat exchangers.

19.3 CSTR Temperature Control

General Characteristics. Because reactors convert reactants to products, reactors are the heart of chemical plants. This section considers temperature control for one class of reactors: CSTRs. Reaction rates and equilibrium relations govern the behavior of reactors. Equilibrium relations represent the upper limits on the conversion of a particular reaction system while the rate expressions represent the conversion rates of chemical reactions. Either the reaction rate or equilibrium can be controlling for a particular reactor. When the system is far removed from equilibrium constraints, equilibrium limitations can be neglected. When equilibrium limits the rate of reaction, the equilibrium limitations must be included in the rate expression. Catalysts do not participate in the chemical reactions or affect the chemical equilibrium, but do affect the rate of reactions.

One of the major challenges associated with CSTR temperature control is the nonlinearity of these systems. Example 5.12 shows that the gain and time constant of a CSTR are nonlinear functions of temperature for an irreversible first-order reaction. The following transfer function was derived by linearizing the nonlinear equation and assuming the reactant concentration, C_A, is constant.

$$G(s) = \frac{T(s)}{Q(s)} = \frac{\dfrac{1}{F C_p - \dfrac{V_r(-\Delta H) C_A k_0 E}{R T_0^2} \exp\left(\dfrac{-E}{R T_0}\right)}}{\dfrac{V_r \rho C_v}{F C_p - \dfrac{V_r(-\Delta H) C_A k_0 E}{R T_0^2} \exp\left(\dfrac{-E}{R T_0}\right)} s + 1} = \frac{K_p}{\tau_p s + 1}$$

As temperature increases, the gain and the time constant decrease because the exponential term increases more quickly than the RT_0^2 term. For the results for an endothermic CSTR shown in Section 9.6, the process rings when an upset causes the reaction temperature to increase, indicating that the decrease in the time constant outweighs the effect of the decrease in the gain. As the temperature decreases, the gain and time constant increase and the results in Section 9.6 indicate that, because the observed response was sluggish, the time constant effects again prevail. The nonlinear temperature dependence of reaction systems is a major challenge for reactor temperature control. As a rule of thumb for industrial systems, the reaction rate should approximately double each time the reaction temperature is increased by 10ºC (18ºF).

The most frequent disturbances with which a CSTR is faced are feed flow rate changes, changes in the feed temperature and changes in the enthalpy of the heating or cooling medium. One of the most difficult disturbances is a change in the heat-transfer coefficient between the heat-transfer medium and the reaction mixture. Even for a well-mixed CSTR, the linear velocity of the reaction mixture at the heat-transfer surfaces is rarely above 1 ft/s, making the heat-transfer surfaces prone to fouling.

Endothermic CSTRs. Endothermic CSTRs are generally much easier to control than exothermic CSTRs because endothermic reactors are self-regulating. Two candidate control configurations for an endothermic CSTR are shown in Figures 19.3.1a and 19.3.2b. Using the flow rate of steam as the MV (Figure 19.3.1a) makes the reactor susceptible to heat load changes and steam enthalpy changes that require corrective action by the temperature controller. This configuration has the advantage that it provides a direct measure of the heat load and, as a result, a measure of the conversion in the reactor.

As in the case of the steam-heated heat exchanger (Figure 19.2.3b), using the steam pressure in the reactor jacket as the MV (Figure 19.3.1b) more effectively absorbs reactor heat duty and steam enthalpy upsets because these disturbances cause immediate changes in the steam pressure. Controlling steam pressure linearizes the

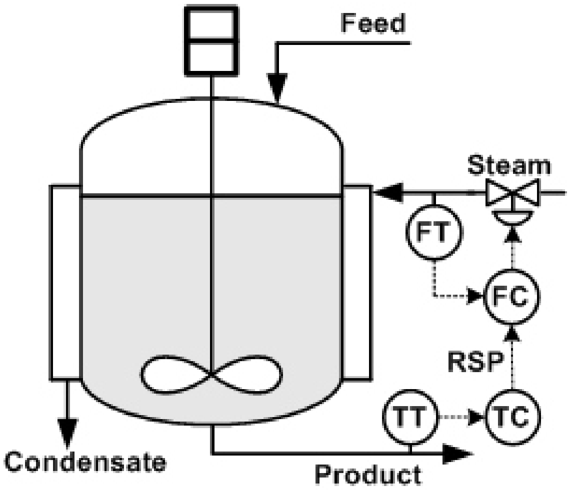


Figure 19.3.1a Schematic for temperature control of an endothermic CSTR using steam flow rate as the MV.

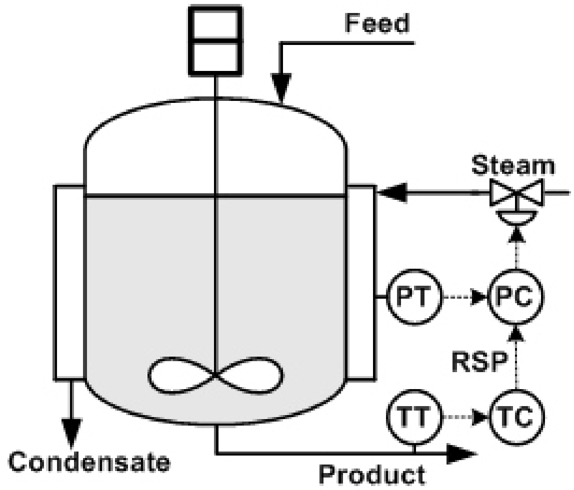


Figure 19.3.1b Schematic for temperature control of an endothermic CSTR using steam pressure as the MV.

temperature control system compared with controlling steam flow rate[1]. The major disadvantage of using the steam pressure as the MV is that it does not provide a direct measure of the heat load on the reactor, but this is usually a secondary consideration and an additional steam flow indicator can be installed if an on-line measure of the reactor heat duty is necessary.

Chapter 13 demonstrates that scheduling of the controller tuning is required when the system is strongly nonlinear. Excessive nonlinear behavior can be noted when ringing and sluggish behavior are observed for the same set of tuning parameters.

Exothermic CSTRs. Non-self-regulation in exothermic CSTRs can arise from the temperature characteristics of heat generation from reaction and heat removal by the coolant. Heat removal is essentially linear with temperature, whereas reaction heat generation exhibits a characteristic "S" shape versus temperature (e.g., Figure 8.3.2). Low reaction rates result in low heat generation at low temperatures. At intermediate temperatures, the slope of the heat generation curve is steep because of an abundance of reactant and a sufficiently high reaction rate. The slope of the heat generation curves decreases at high temperature because of depletion of the reactant.

Non-self-regulation can result when the slope of the heat generation curve exceeds the slope of the heat removal curve. Figure 19.3.2 shows the rate of heat generation by reaction and the rate of heat removal by the cooling system as functions of reactor temperature. The intersections of these two curves represent steady-state operating points for the reactor. The low- and high-temperature intersection points are stable operating points because, if the temperature of the reaction mixture increases, the rate of heat removal exceeds the rate of heat generation by reaction and, if the temperature decreases, the rate of heat removal is less than the rate of heat generation. In both cases, the reactor temperature returns to the original operating point.

The operating point corresponding to the intermediate temperature is open-loop unstable because, if the reactor temperature increases, the rate of heat generation by reaction exceeds the rate of heat removal, and the reactor moves to the stable high-temperature operating point. If the reactor temperature decreases, the rate of heat removal exceeds the rate of heat generation by reaction, and the reactor moves to the stable low-temperature

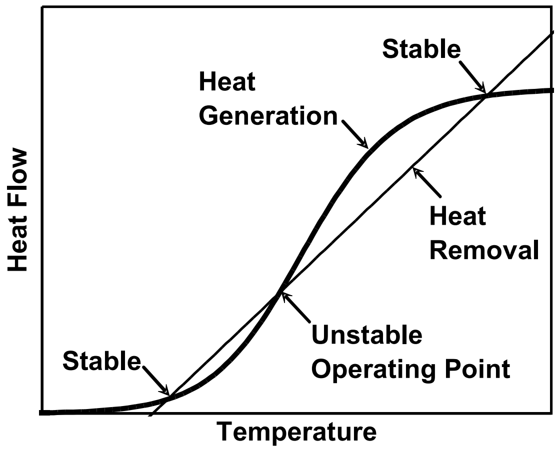


Figure 19.3.2 Demonstration of multiple steady state operating points for an exothermic CSTR.

operating point. Note that the slope of the heat generation curve is greater than that of the heat removal curve for the unstable operating point, and the slope of the heat generation curve is smaller for the stable points.

A feedback controller must be used to maintain operation at the unstable intermediate temperature operating point that, in many cases, is the desired operating point from an economic point of view. A feedback controller maintains a CSTR at an open-loop unstable operating point by increasing the heat removal rate so that it exceeds the heat generation rate at temperatures above the operating point and by decreasing the heat removal rate so that it is less than the heat generation at temperatures below the operating point.

A runaway in reactor temperature can occur when the rate of heat generation by reaction exceeds the maximum heat removal rate available from the coolant system and the conversion is much less than 100%. The likelihood of a temperature runaway for an exothermic CSTR can be reduced by replacing a liquid coolant with a boiling coolant (for which the heat-transfer coefficient is tripled), reducing the feed rate of reactants, reducing the concentration of the reactants in the feed or increasing the ratio of the heat-transfer area to the reactor volume. These changes also increase the controllability of the reactor.

Figures 19.3.3a and 19.3.3b show two recommended approaches for controlling the temperature of an exothermic CSTR by cascading the reactor temperature controller to a coolant temperature controller. Using the setpoint of the coolant temperature as the MV for temperature control of an exothermic CSTR is recommended because it results in a fast-acting system and linearizes the process gain[2] (the change in the outlet temperature of the process for a change in the coolant temperature).

In Figure 19.3.3a, the MV for the reactor temperature controller is the setpoint for the outlet temperature of the coolant. This configuration has the advantage of providing faster compensation for fouled heat-transfer surfaces, but responds more slowly to changes in the inlet temperature of the coolant or changes in the supply pressure of the coolant.

Figure 19.3.3b is a schematic of a configuration for controlling the reactor temperature using the setpoint of the inlet temperature of the coolant as the MV. This configuration has the advantages of a faster response to inlet coolant temperature changes and changes in the supply pressure of the coolant, but responds more slowly to fouling of the heat-transfer surfaces.

The choice between the configurations shown in Figures 19.3.3a and 19.3.3b depends upon which is more important: a fast response to coolant upsets or to fouling of the heat-transfer surface. As the circulation rate of the coolant increases, the differences between the approaches shown in Figures 19.3.3a and 19.3.3b diminishes. At high coolant circulation rates, these approaches provide equivalent performance for the full range of upsets. McMillan[1] presents an analysis of other possible configuration choices.

In certain cases, it is desirable to maximize the production rate of a reactor. Figure 19.3.4 demonstrates how one of the previous configurations (Figure 19.3.3b) can be extended to maximize production rate. A **valve position**

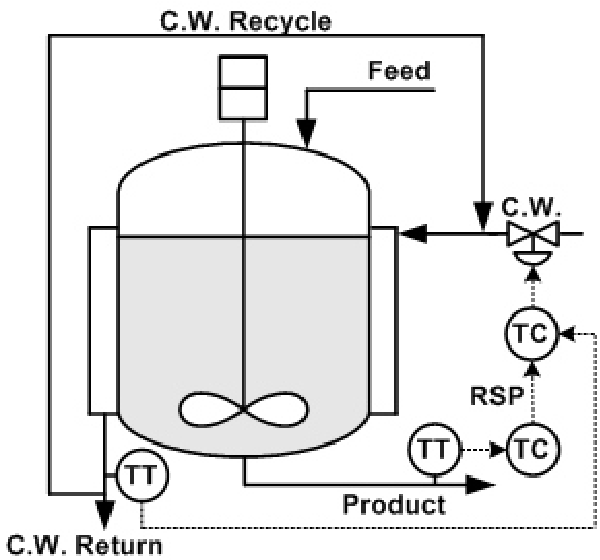


Figure 19.3.3a Schematic of a temperature controller for an exothermic CSTR in cascade with the outlet temperature of the coolant.

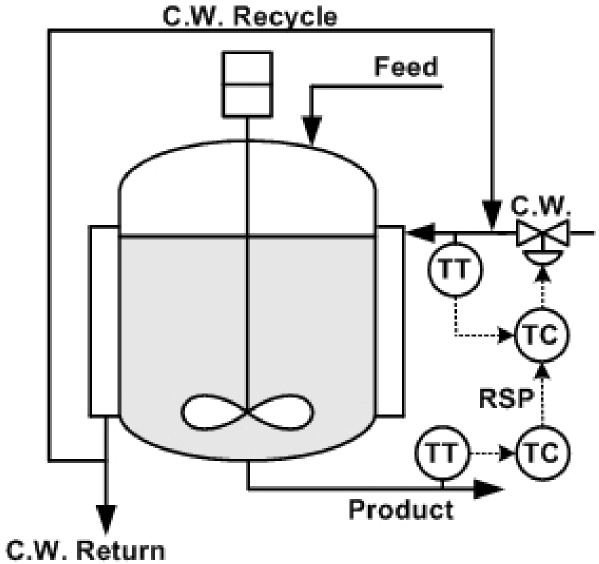


Figure 19.3.3b Schematic of a temperature controller for an exothermic CSTR in cascade with the inlet temperature of the coolant.

controller (VPC) manipulates the feed until the valve on the coolant makeup is sufficiently open, yet still able to reject disturbances (e.g., 80% open). An integral-only controller is usually used for the valve position controller so that gradual movement to the maximum production rate is ensured.

One of the undesirable aspects of the previous control approaches used to control exothermic CSTRs is that a separate heating system is required for startup. In certain cases, the evaporation of hot condensate can be used as the coolant for exothermic CSTRs. Figure 19.3.5 is a schematic of such an application. Note that the condensate heats the reactant mixture to initiate reaction and, when the reaction starts, evaporation of the condensate removes heat from the reactor. A steam pressure controller is used as the MV for the reactor temperature controller because it responds quickly to heat duty changes. The steam produced by cooling the reactor can be added to the low-pressure steam system, which can represent significant energy recovery in certain cases.

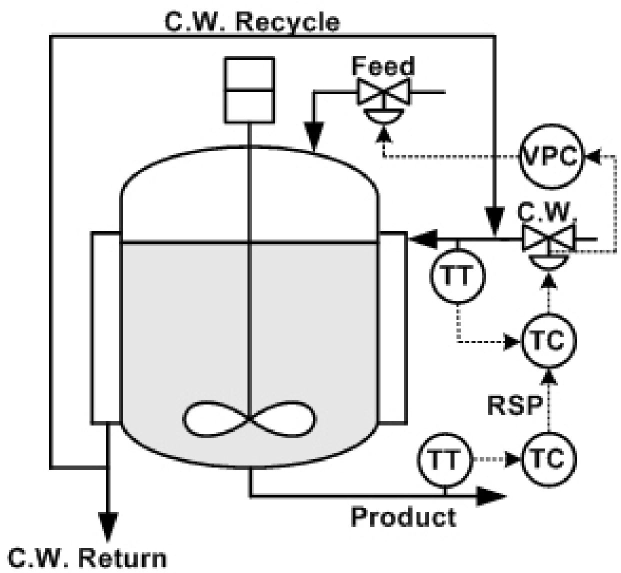


Figure 19.3.4 Schematic of control system for an exothermic CSTR that is designed to maximize production rate. Note that VPC is a valve position controller.

Typically, a range of controller gains provides stable operation of an exothermic CSTR operating at an unstable operating point. Because the reactor is open-loop unstable, a minimum amount of feedback control, which

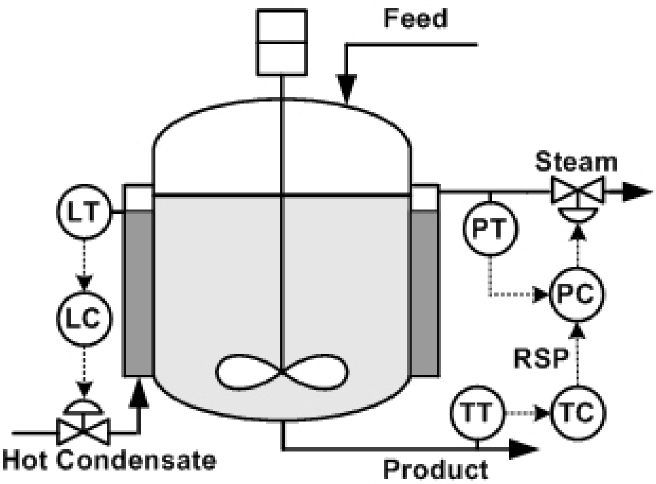


Figure 19.3.5 Schematic of a control configuration for an exothermic CSTR that uses hot condensate to startup the reactor and cool it.

corresponds to K_c^{min} , is necessary to stabilize the system. Likewise, there is an upper limit to controller aggressiveness, K_c^{max}, above which the closed-loop process becomes unstable. Therefore, there is a range or window of K_cs that provide stable operation, i.e., for stable operation,

$$K_c^{min} \quad K_c \quad K_c^{max}$$

This range of suitable values for K_c is a strong function of the effective deadtime-to-time constant ratio of the process. McMillan[1] indicates that, in general,

$$\frac{K_c^{max}}{K_c^{min}} \quad \frac{1}{{}_p / {}_p}$$

and recommends that a reactor not be designed for deadtime-to-time constant ratios exceeding 0.1. The effective deadtime of a temperature loop on a CSTR can be approximated by[1]

$$_p \quad T_{mix} \quad _{ht} \quad _{coolant} \quad _s$$

where T_{mix} is the mixing turnover time, which is approximated by the reactor volume divided by the sum of the feed rate, the recirculation rate and the pumping rate of the agitator; $_{ht}$ is the time constant for heat transfer, which is approximated by the product of the mass of metal in the heat-transfer surface and its heat capacity divided by the product of the heat-transfer coefficient and the area of the heat-transfer surface; $_{coolant}$ is the transportation delay of the coolant, which is approximated by the coolant jacket volume divided by the coolant recirculation rate; and $_s$ is the time constant for the temperature sensor system (e.g., the thermowell and RTD). Even though each of these elements is a time constant and not a deadtime, their combined effect on the high-order system behaves like deadtime (see Section 6.7). If the deadtime-to-time constant ratio is below 0.1, you can expect excellent reactor temperature control, i.e., peak errors can be expected to be less than the measurement error[1], which is largely offset from the true reading. If inadequate agitator circulation is applied, insufficient heat-transfer area is used, a low coolant flow rate is used or excessive sensor lag (e.g., a glass thermowell or a thermowell with an excessive air gap between the RTD and the walls of the thermowell) is present, poor control performance or possibly, in the extreme, an uncontrollable process results, irrespective of controller tuning efforts.

The controller tuning for temperature control loops applied to exothermic CSTRs is somewhat different from controller tuning for most other processes. Normal levels of integral action for these systems can lead to an unstable closed-loop system because this is an integrating process. It is recommended[1] that a reset time greater than 10 min or approximately 10 times that of normal tuning levels should be used in these cases. It is recommended that the gain be approximately doubled compared with normal tuning and that the derivative time be at least 0.5 min.

19.4 Bio-Reactor Control

Most industrial fermentations are aerobic and require that the level of dissolved oxygen be controlled within a defined range. Dissolved oxygen is the means by which microorganisms (much like fish in an aquarium) obtain oxygen. That is, they cannot directly utilize gaseous oxygen.

An industrial bioreactor typically has at least four main control loops: (temperature, inlet gas flow rate, tank back pressure, and stirrer agitation rate. Any of the above four control loops can be used by themselves or in combination to control dissolved oxygen level within a desired range. Also, a fifth means of controlling dissolved oxygen is available, that being to add pure oxygen to supplement the inlet gas (which is normally air).

A diagram of a bioreactor (Figure 19.4.1) noting the locations of the dissolved oxygen probe and possible manipulated streams is shown below:

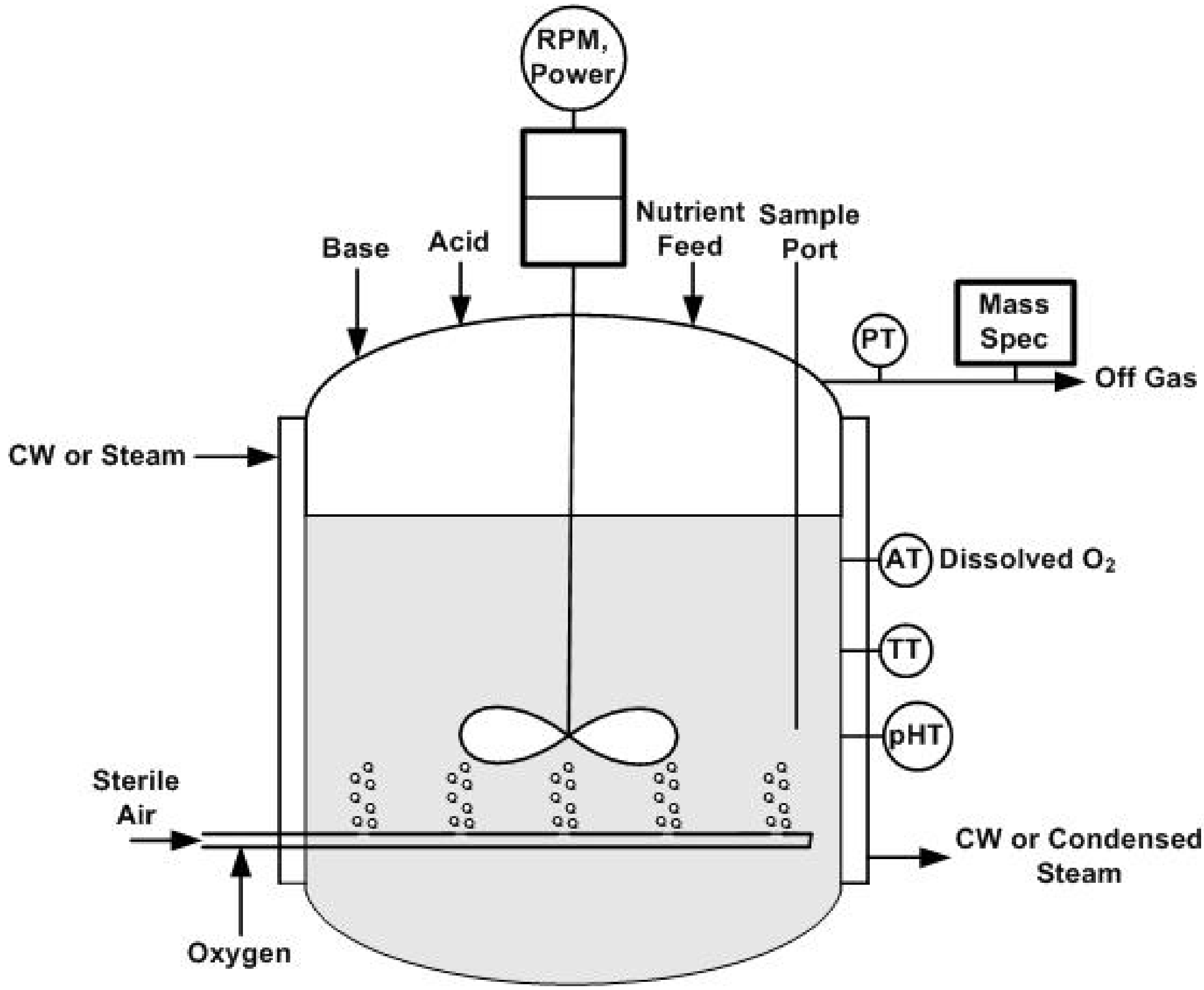


Figure 19.4.1 Schematic of a jacketed bio-reactor

A typical trend plot of dissolved oxygen (using either agitation or a pure oxygen feed as the control variable, starting at hour 1) is shown in Fig. 19.4.2

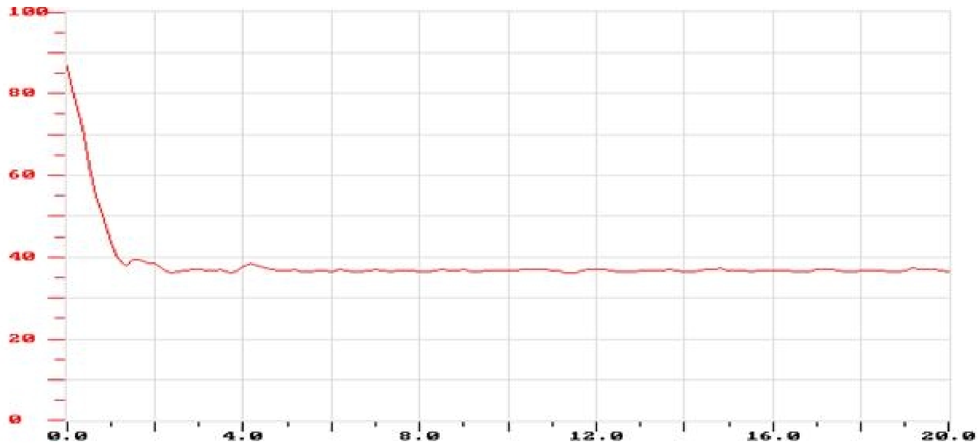


Figure 19.4.2 A typical plot of percent O2 versus time in hours.

Note that dissolved oxygen level typically starts out in the batch fermentation at a value close to 100 % saturation, due to the presence of only a small amount of cell mass drawing oxygen (i.e., from the inoculum). As cell mass increases (i.e., over 100 fold) during the fermentation, increasing amounts of oxygen are consumed, lowering the concentration of dissolved oxygen. When the value of dissolved oxygen arrives at a target value, a primary controller (cascaded to other controllers) starts up to control dissolved oxygen at the target value. Note that if the dissolved oxygen becomes too low, the culture will shift its metabolism to anaerobic and cause undesirable effects. When the dissolved oxygen level is too high, other undesirable effects occur.

The particular dissolved oxygen control strategy that is best to use will depend in large part on the attributes and limitations of the cell culture utilized, as well as available ranges of operation of controlled parameters. Considerations of the manipulated variables for the primary DO2 controller are discussed as follows:

A. **Use of temperature**. Reducing temperature will slow cellular metabolism and, hence, the oxygen consumption rate, causing the dissolved oxygen concentration to rise. The process time constant is likely to be relatively long (several minutes). Changing temperature is not usually desirable since changing temperature changes the rates of many, if not most, cellular metabolic reactions, some of which may be producing unwanted side products or otherwise affecting the mix of compounds in fermentation broth. Development scientists traditionally have optimized the fermentation at a fixed temperature, which has also been submitted in the New Drug Application to the FDA, so changing the temperature of a bioprocess may not only be detrimental to the process but may even violate conditions of FDA approval of the process.

B. **Use of gas inlet flow.** Gas inlet flow rate (air in most cases) is a common means of helping to control dissolved oxygen. The process time constant is fast (several seconds). However, the use of airflow has some limitations. Too much air will likely cause excessive foaming of the fermentation broth and too little air may make it challenging to hold positive pressure in the reactor and sweep out toxic gases (e.g., carbon dioxide). So, airflow usually is not adequate to be used by itself to control dissolved oxygen but is usually used in combination with one of the other control strategies ©, D, or E) listed here.

C. **Use of agitation**. Agitation is a common means of controlling dissolved oxygen for fermentation in which the cells are not shear sensitive. For example, E.coli, Streptomycetes, and fungal cells are typically not shear sensitive. The process time constant is relatively fast (a few seconds). One of agitation's two main purposes (the other being mixing) is to break up (i.e., by shear) the large bubbles entering the fermentation broth from the gas sparger. This significantly increases the total surface area of the total resulting smaller bubbles, hence increasing the rate of transfer of oxygen from the gas to dissolved state via a parameter known as $k_l a$, where "k_l " is the gas transfer coefficient and "a" is the surface area of the bubbles.

D. **Use of pure oxygen feed.** Mammalian cells are highly shear sensitive and so cannot tolerate the shear associated with a high speed agitator as well as the shear of a high airflow stream with bubbles collapsing at the surface. So, a low airflow and low agitation speed must be used. In fact, in many such bioprocesses, air enters the bioreactor via sintered metal devices so as to create tiny bubbles (without having to use the agitator to create small bubbles). Agitation (with marine low-shear blades) is still used, but only for broth mixing purposes. In these cases, where agitation and airflow are not highly effective in controlling dissolved oxygen, a pure oxygen feed is added to the inlet gas stream with the pure oxygen feed being the manipulated variable in controlling dissolved oxygen. The process time constant is relatively fast (seconds).

E. **Use of back pressure**. Increasing back pressure on the bio-reactor will increase the driving force in transferring oxygen from the gas state to the dissolved state, in keeping with Henry's Law. The process time constant is relatively fast (several seconds). However, increasing pressure will also increase the concentration of dissolved carbon dioxide (a waste product of cellular metabolism) in the broth, and carbon dioxide levels above certain values are known to be toxic to mammalian cell bioprocesses. Increasing pressure will also reduce the pressure difference associated with driving inlet gas flow to the bioreactor and thereby reducing the amount of air that be fed to the vessel.

Note, the relevant equation for dissolved oxygen based on assuming steady-state operation is:

$$F_{\text{Cell Uptake}} \quad F_{\text{Transport}} \quad k_1 a \;\; P_{O_2,\text{air}} \quad P_{O_2,\text{dissolved}}$$

where $F_{\text{Cell Uptake}}$ is the rate of oxygen consumption by the cells in the bio-reactor, $F_{\text{Transport}}$ is the rate of transport of oxygen from the bubble in the bio-reactor to the reaction mixture, k_1 is the gas transfer coefficient, a is the surface area of the gas bubbles, $P_{O_2,\text{air}}$ is the partial pressure of oxygen in the air injected into the sparger and $P_{O_2,\text{dissolved}}$ is the partial pressure of dissolved oxygen in the reactor.

The addition of pure oxygen or increasing pressure will increase the partial pressure of oxygen in the gas, hence increasing the driving force for oxygen transfer. The use of airflow or agitation mostly affects the "*a*" (surface area) part of $k_l a$.

Note that all terms in the above equation except $k_l a$ are known in real-time (i.e., oxygen uptake rate via mass spectrometry analysis of fermentation gases, dissolved oxygen via a DO2 probe in the fermentor, and the calculated partial pressure of oxygen in the sparging gas). Therefore, $k_l a$ (oxygen mass transfer coefficient) can be continuously calculated from available information as an on-line parameter in a running fermentation. $k_l a$ can then be compared with different fermentation conditions in different fermentors to determine what combination of sparger design, pressure, airflow, and agitation speed results in the best $k_l a$. Even more importantly, comparison of the on-line time varying calculated $k_l a$ to historical time varying values can be a great indicator of a faulty dissolved oxygen sensor. That is, the dissolved oxygen sensor is an electrochemical probe, which can drift from calibration, with a thin Teflon membrane that can sometimes foul and is normally replaced every few batch fermentations. Note that the sensor cannot normally be recalibrated once a fermentation begins. So, monitoring $k_l a$ on-line is a great on-line diagnostic check on the viability of the dissolved oxygen sensor.

Note that there is no direct DO2 control element. Rather, DO2 must be controlled by one of the available direct control loops, such as pressure, inlet airflow, agitation, or oxygen feed. That is, DO2 control is a cascade (primary/secondary) type loop (Section 12.1) where the outer DO2 slow loop is the master while the inner fast direct control loop is the secondary. Moreover, many times the primary controller requires adaptative controller tuning, such as, gain scheduling (Section 13.3) due to the nonlinearity exhibited by the system during the entire batch reaction phase. So, the challenge from a capstone design perspective is how to select the proper secondary control loop.

So, the reader can see that **many dissolved oxygen control strategies are possible, each with pros and cons, especially depending on the type of cells being used within the fermentor/bioreactor. So, knowledge of the process is key in determining which dissolved oxygen control strategy to use.**

19.5 Summary

These case studies have demonstrated that, to identify the proper control configuration, process knowledge must be used to accomplish each of the following tasks:

1. Identify the key process disturbances.
2. Determine the control configurations that most effectively reduces the impact of these disturbances.
3. Identify control configurations that provide fast dynamic response to upsets and have low deadtime-to-time constant ratios.
4. Evaluate the impact of coupling if present.
5. Choose a control configuration that represents the best combination of dynamic response to upsets and sensitivity to coupling.
6. For bioprocesses, when choosing a control configuration, strong consideration of the attributes and limitations of the cell culture being utilized is often needed.

19.6 References

1. McMillan, G.K., *Advanced Temperature Control*, Instrument Society of America, 1995, pp. 95-132.

Appendix A

Piping and Instrument Diagrams

A.1 Introduction

Piping and Instrument Diagrams (P&IDs also known as engineering flowsheets) are used by industry to document the control systems for their plants. Even though the Instrument Society of America has established a standard for symbols that are used in P&IDs (ANS/ISA-5.1-1984), the standard is general and, in some cases, open ended. As a result, operating and engineering companies in the CPI generally develop their own P&ID standards within their company. This appendix is intended to provide only a brief introduction to P&IDs. Only the elements of P&IDs that relate to the control schematics used in this text are presented. The reader is referred to the aforementioned ISA standard if more detailed information is required.

There are three general types of control diagrams. (1) Simplified diagrams (also known as process flow diagrams) are the simplest type and provide an overview of the process and the primary controls (i.e., the normally operating control loops) without identifying the sensors and controllers by number. (2) Conceptual diagrams present a complete representation of the primary controls, without numbering for sensors and controllers. In principle, these diagrams are closest to the control schematics used in this text. (3) P&IDs show all of the hardware, instruments and piping specifications, including block and bypass valves and all of the controls, including override and startup and shutdown controls. These detailed diagrams provide numbers for the controllers and sensors and all streams.

A.2 Loop and Sensor Convention for Acronyms

Table A.1 lists some of the P&ID acronyms for sensors and controllers. These acronyms appear inside a circle, used to represent instruments on a P&ID. Consider a tag for a temperature controller: TC-101. The first letter (i.e., the primary letter) indicates that temperature is involved, and the second letter specifies that it is a controller. The number following the "TC" is the loop number. Therefore, the temperature sensor/transmitter for this loop is represented by TT-101 because each instrument in the control loop has the same number.

One or more letters can follow the primary letter and indicate the particular function of the instrument. For example, consider a temperature controller that also records the temperature reading. In that case, the controller is represented by TRC-xxx. A temperature controller that has an indicator attached to it is given by TIC-xxx. FRC-xxx represents a flow controller that records the flow rate on a strip chart. The recorder designation has generally become obsolete for DCSs. For DCSs, FIC is used to represent a flow controller that provides indication of the flow rate.

Table A.1
Acronyms for Several Commonly Encountered Sensors and Controllers

Letter	First or Primary Letter	Succeeding Letters
A	Analysis	Alarm
C		Controller
D		Differential
F	Flow rate	Ratio
H		High
I		Indicator
L	Level	Low
P	Pressure	
R		Recorder
S		Switch
T	Temperature	Transmitter
Y		Computation block

FR-xxx indicates that a ratio of flow rates is formed by the denoted instrument. Switches (S) with high (H) and low (L) limits can also be specified for P&IDs. For example, a high-level switch is given by LSH-xxx. A high-pressure switch is given by PSH. The difference between two sensor readings is often used for control purposes. For example, a differential pressure sensor/transmitter (PDT-xxx) is used to indicate the onset of flooding in a column. Note that the schematics in this text represent differential pressure sensor/transmitters as DPTs, and they are known industrially as DP cells. The difference between two temperature readings that is transmitted to a controller and also recorded is given by TDRT-xxx.

One of the most widely used control instrument acronyms, and the most ambiguous, is the computation block (Y). It is used to represent a ratio operation (Section 12.3), a feedforward controller (Section 12.4) or a computed MV (Section 13.5). Sometimes a note attached to the computation block explains its function. **Usually, there is a sheet that accompanies P&IDs to define the symbols used.**

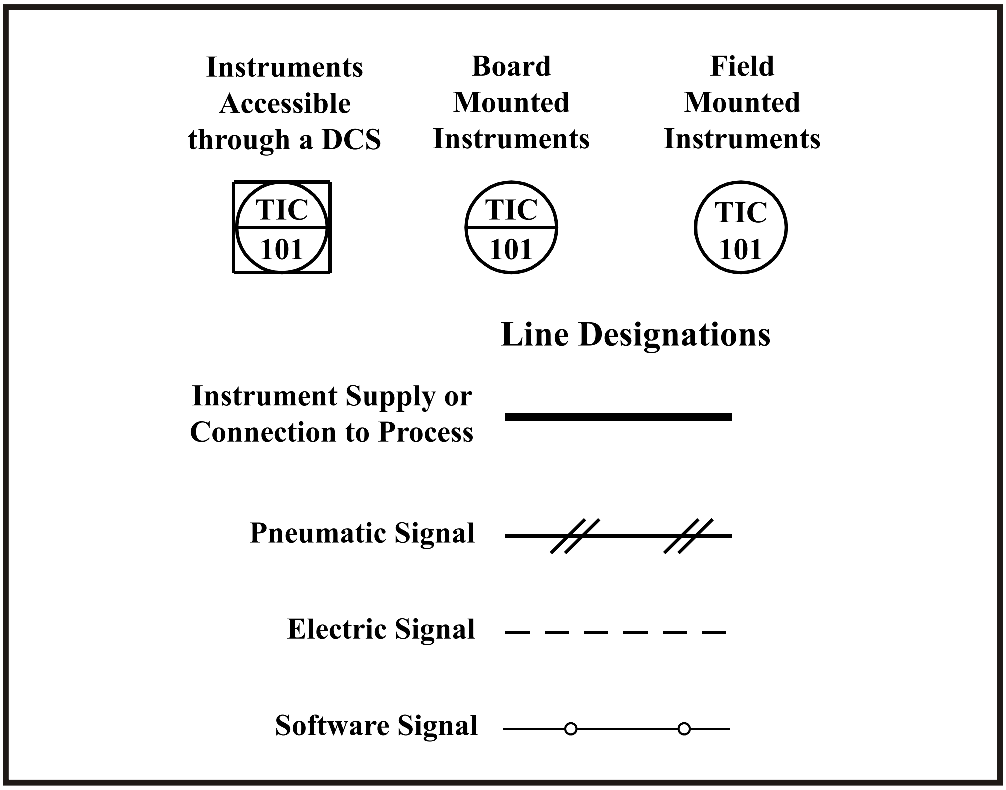


Figure A.1 Symbols for instruments and lines in a P&ID.

A.3 P&ID Symbols

A selected group of P&ID symbols is presented in Figure A.1. A "bubble" with a horizontal line through the middle indicates a sensor reading or a controller that is board-mounted in a control room, e.g., an analog electronic controller. The acronym for the instrument is placed inside the bubble and above the horizontal line, and the loop number associated with the instrument is placed below the line. If a bubble for an instrument does not have a horizontal line, it indicates that the instrument is mounted in the field. A bubble inside a square indicates that the instrument represented by the bubble can be accessed using a DCS or control computer.

A solid line without crosshatching represents instrument supply (e.g., instrument air lines) or a connection to the process (e.g., the pressure taps from a differential pressure sensor to the process). Pneumatic lines are given by a solid line with pairs of crosshatching. Electric lines are represented by dashed lines. Software signals, which are represented by a line connecting small circles, represent the logic flow within a control computer or DCS.

A.4 Example P&ID

Figure B.2 shows a P&ID for the stripping section of a column. The control loop number for the reboiler level is 328 and for the reboiler steam is 329. There are tags to show the connection between this P&ID and other P&IDs.

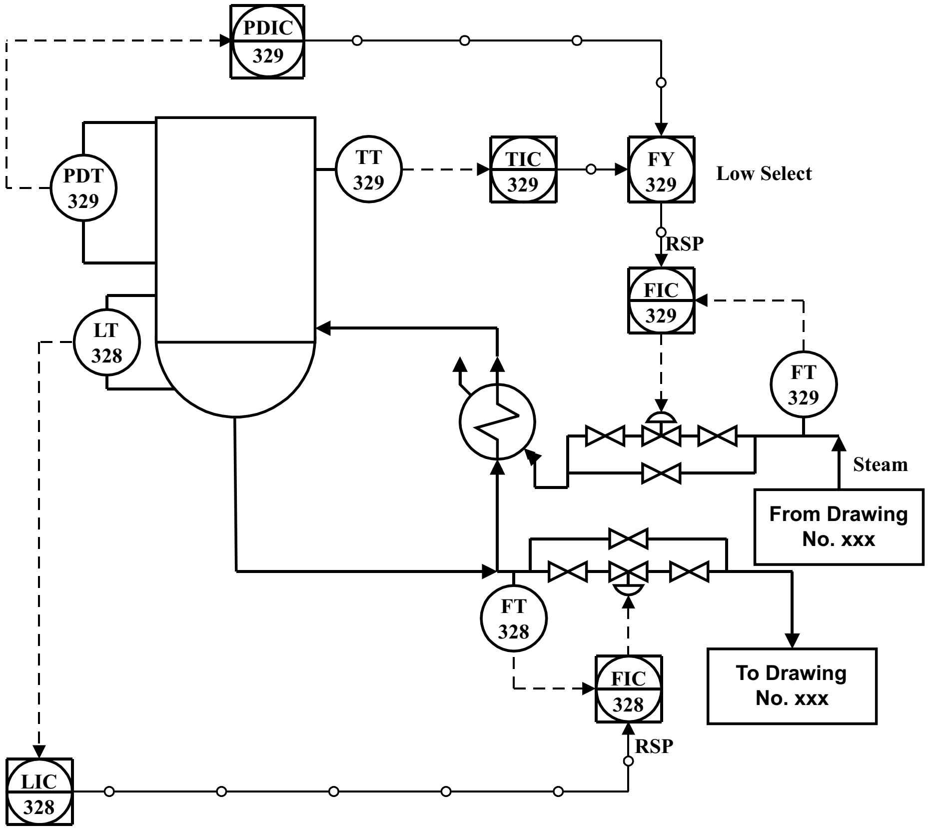


Figure A.2 Example of a P&ID

In most cases, a control engineer must use a number of P&IDs to determine the existing control configuration for a single distillation column.

There are field-mounted sensor transmitters for level, flow and differential pressure. The rest of the instrument functions reside in the DCS.

Appendix B

Pseudo-Random Number Generator

Application of the model for sensor noise presented in Chapter 3 requires a random number, x_n. Considerable work has been done in developing techniques to generate random numbers that have a very nearly perfectly random distribution. For modeling noise, the requirements for a random number generator are not nearly as demanding. A simple pseudo-random number generator[1] will suffice and is given by

$$x_{n+1} = 10^P C x_n - I(10^P C x_n)$$

where x_{n+1} is the calculated random number, which is between zero and unity, x_n is the previous random number, which is also between zero and unity, P is the number of significant figures used in x_n, $I(y)$ is the integer value of y and A and B are constants. To generate a random number, four values are required: A, B, P and x_n. A is any non-negative integer and B is any number from the set {3, 11, 13, 19, 21, 27, 29, 37, 53, 59, 61, 67, 69, 77, 83, 91}. C is calculated by the following expression

$$C = 10^{-P}(200A + B)$$

The initial random number (x_0), known as the seed, should be between zero and unity and contain the same number of significant figures as x_n such that

$$x_0 = 10^{-P} K$$

where K is any integer not divisible by 2 or 5 such that $0 < K < 10^P$.

1. Graybeal, W.T. and U.W. Pooch, *Simulation: Principles and Methods*, Little, Brown and Company, Boston, MA, Section 4.2.5 (1984).

Appendix C

Signal Filtering

C.1 Introduction

Some degree of noise is associated with all process measurements. This noise can be caused by electrical interference, mechanical vibration or changes in the process (e.g., variations resulting from turbulent flow). Noise is a high-frequency variation in the process measurement that is not associated with the true process measurement. A controller that is responding to the noise on a measurement makes high-frequency changes to the MV, causing short-term process variations in the CV. The noise can increase the variability of the CV about its setpoint, but the average value of the CV does not change. In effect, a controller using measured CV values with significant noise levels passes the noise into the process if preventive steps are not taken. Depending on the process gain and time constant, the noise level can be amplified by the controller. When derivative action is used, the feedback system is even more prone to amplifying the noise on the measurement of the CV. Filtering of process measurements is an effective means of reducing the effects of measurement noise.

C.2 Sampling

The measurement of a CV must be converted into a digital value by the A/D converter before it can be used by the control algorithm in the DCS. A continuous measurement (e.g., a temperature sensor) is sampled by the A/D converter at discrete points in time. After the measurement is sampled, its value is fixed at that level until the next sampling by the A/D converter. Figure C.1 shows a continuous measurement and the discrete sampled measurement. Because the sampled value is fixed at its last measured value, this sampling approach is called the **sample and hold** approach or a **zero-th order hold**. The time between sample points is called the **sampling period**, t. The **sampling rate** is 1/ t and the **sampling frequency** is 2 / t.

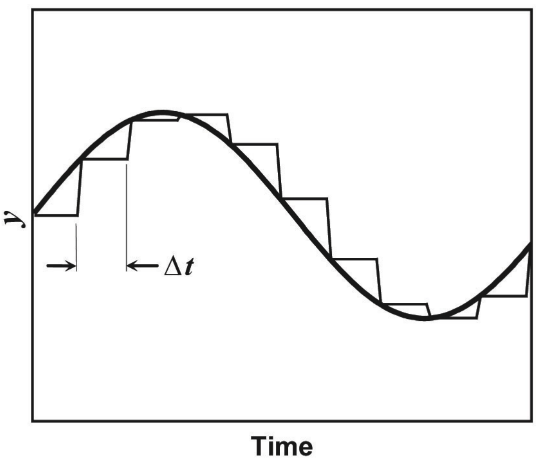


Figure C.1 Comparison between a continuous measurement and the corresponding sample and hold reading.

From Section 7.16, both the control interval and sampling period should be less than 0.05($_p$ $_p$) to approach continuous control performance. If this guideline is used, a sample and hold approach to sampling the CV should provide an accurate measure of the process behavior.

C.3 Signal Aliasing

A sampling period that is too long results in a loss of information. Significant variation of the signal can occur during the sampling period and knowledge of the true dynamic behavior is lost. Figure C.2 shows the original signal as a high-frequency sinusoid (thin line). A sampling period (t), which is too large to accurately measure the original signal, is applied to the original signal yielding the apparent signal (thick line). The apparent signal is also a sinusoid, but with a much lower frequency. This phenomenon is known as **signal aliasing**. According to Shannon's sampling theorem[1], to prevent signal aliasing, the sampling period must be less than one-half the period of the original signal to accurately reconstruct the original signal.

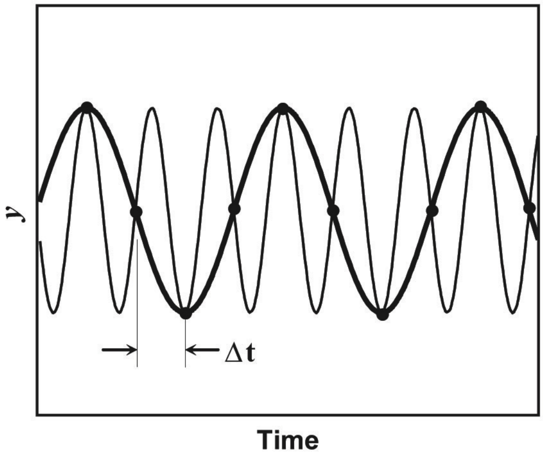


Figure C.2 Signal aliasing resulting from using too large a sampling period. Thin line- the original signal. Thick line- the apparent low-frequency signal resulting from signal aliasing. represents the sampled measurement.

An industrial sensor reading contains a full range of frequencies, i.e., low frequencies representing true process changes and high frequencies from sensor noise. As a result, the sampling period selected to follow the changes in the true process causes the high-frequency components of the sensor reading (i.e., those with periods twice t and smaller) to appear as lower-frequency components due to signal aliasing. The low-frequency components resulting from signal aliasing can affect control performance if their frequencies correspond to the frequencies of the process being controlled. The effects of signal aliasing on control performance can be handled by applying the proper filtering techniques to the sensor reading.

C.4 Filtering Process Measurements

The effects of high-frequency noise on a process measurement can be reduced by filtering the measurement signal. Filtering can be thought of as taking a running average of the measurement readings of the last *n* samples. In this manner, the high-frequency variations resulting from noise can be "averaged" out. The continuous form of a first-order filter is given by a first-order differential equation

$${}_f \frac{dy_f}{dt} \quad y_f \quad y_s$$

where y_s is the unfiltered sensor reading, y_f is the filtered value of the sensor reading, and ${}_f$ is the time constant for the filter. Using a first-order backward finite difference approximation of the first derivative yields the equation for a digital filter

$$y_f(t) = \frac{\frac{\tau_f}{\Delta t}}{\frac{\tau_f}{\Delta t}+1} y_f(t-\Delta t) + \frac{1}{\frac{\tau_f}{\Delta t}+1} y_s$$

where Δt is the sampling period. This equation can be written in the following simplified form

$$y_f(t) = [1 - f]\, y_f(t-\Delta t) + f\, y_s(t)$$

where f is the filter factor defined by
$$f = \frac{1}{\frac{\tau_f}{\Delta t}+1}$$

Consider the case for which $f = 0.05$. The filter for this case can be viewed as an average of the current sensor reading (y_s) and the previous 19 sensor readings because for $f = 0.05$, 5% of the new filtered value comes from the current measurement and 95% comes from the previous readings.

The smallest control interval for normal control loops on a DCS is 0.1 to 0.5 second. From Shannon's sampling theorem, the components of the noise with periods less than 0.5 second experience signal aliasing. Because analog filters can operate at much higher frequencies (smaller sampling periods) than A/D converters on a DCS, the continuous measurement signal is typically passed through an analog filter to remove the high-frequency noise before it goes to the A/D converter. Analog filters use resistors and capacitors to remove high-frequency noise from electrical signals. The output of the A/D converter has a digital filter which removes the lower-frequency noise. This signal processing sequence is shown schematically in Figure C.3.

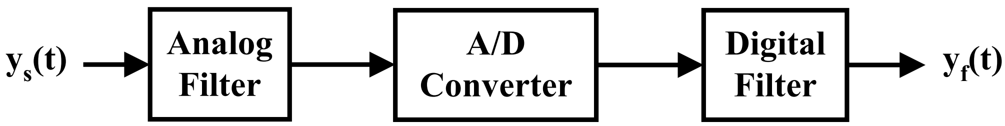


Figure C.3 Schematic showing the sequence of signal processing steps used to filter and sample a continuous measurement for use in a DCS.

1. Astrom, K.J. and B Wittenmark, *Computer Controlled Systems*, Prentice-Hall (1984).

Index

A

B

C

D

E

F

G

H

I

J

L

M

N

O

P

T

U

V

W

Z