Chemical Process Dynamics and Controls

Book I

(Chapters 1-9)

Welcome to the University of Michigan Chemical Engineering Process Dynamics and Controls Open Textbook.
This electronic textbook is a student-contributed open-source text covering the materials used at Michigan in our senior level controls course.

Follow this link to find more information about this course.

If you would like to suggest changes to these pages, please email rziff@umich.edu .

Click here for the 2007 version and here for the 2006 version of the text.

Table of Contents

Part I
Process Control Introduction

Chapter 1. Overview

Section 1. Introduction to Controls: Background and design methodology

1.1 Introduction

Process controls is a mixture between the statistics and engineering discipline that deals with the mechanism, architectures, and algorithms for controlling a process. Some examples of controlled processes are:

- Controlling the temperature of a water stream by controlling the amount of steam added to the shell of a heat exchanger.
- Operating a jacketed reactor isothermally by controlling the mixture of cold water and steam that flows through the jacket of a jacketed reactor.
- Maintaining a set ratio of reactants to be added to a reactor by controlling their flow rates.
- Controlling the height of fluid in a tank to ensure that it does not overflow.

To truly understand or solve a design problem it is necessary to understand the key concepts and general terminology. The paragraphs below provide a brief introduction to process controls as well as some terminology that will be useful in studying controls. As you begin to look at specific examples contained here, as well as elsewhere on the wiki, you will begin to gain a better grasp on how controls operate and function as well as their uses in industry.

1.2 Process Control Background

The role of process control has changed throughout the years and is continuously shaped by technology. The traditional role of process control in industrial operations was to contribute to safety, minimized environmental impact, and optimize processes by maintaining process variable near the desired values (1). Generally, anything that requires continuous monitoring of an operation involve the role of a process engineer. In years past the monitoring of these processes was done at the unit and were maintained locally by operator and engineers. Today many chemical plants have gone to full automation, which means that engineers and operators are helped by DCS that communicates with the instruments in the field.

What are the benefits of Process Control?

The benefits of controlling or automating process are in a number of distinct area in the operation of a unit or chemical plant. Safety of workers and the community around a plant is probably concern number one or should be for most engineers as they begin to

design their processes. Chemical plants have a great potential to do severe damage if something goes wrong and it is inherent the setup of process control to set boundaries on specific unit so that they don't injure or kill workers or individuals in the community.

1.3 The Objectives of Control

A control system is required to perform either one or both task:

1. Maintain the process at the operational conditions and set points

Many processes should work at steady state conditions or in a state in which it satisfies all the benefits for a company such as budget, yield, safety, and other quality objectives. In many real-life situations, a process may not always remain static under these conditions and therefore can cause substantial losses to the process. One of the ways a process can wander away from these conditions is by the system becoming unstable, meaning process variables oscillate from its physical boundaries over a limited time span. An example of this would be a water tank in a heating and cooling process without any drainage and is being constantly filled with water. The water level in the tank will continue to rise and eventually overflow. This uncontrolled system can be controlled simply by adding control valves and level sensors in the tank that can tell the engineer or technician the level of water in the tank. Another way a process can stray away from steady state conditions can be due to various changes in the environmental conditions, such as composition of a feed, temperature conditions, or flow rate.

2. Transition the process from one operational condition to another

In real-life situations, engineers may change the process operational conditions for a variety of different reasons, such as customer specifications or environment specifications. Although, transitioning a process from one operational condition to another can be detrimental to a process, it also can be beneficial depending on the company and consumer demands.

Examples of why a process may be moved from one operational set point to another:

1. Economics

2. Product specifications

3. Operational constraints

4. Environmental regulations

5. Consumer/Customer specifications

6. Environmental regulations

7. Safety precautions

1.4 Definitions and Terminology

In controlling a process there exist two types of classes of variables.

1. **Input Variable** – This variable shows the effect of the surroundings on the process. It normally refers to those factors that influence the process. An example of this would be the flow rate of the steam through a heat exchanger that would change the amount of energy put into the process. There are effects of the surrounding that are controllable and some that are not. These are broken down into two types of inputs.

a. *Manipulated inputs:* variable in the surroundings can be control by an operator or the control system in place.

b. *Disturbances:* inputs that can not be controlled by an operator or control system. There exist both measurable and immeasurable disturbances.

2. **Output variable**- Also known as the *control variable* These are the variables that are process outputs that effect the surroundings. An example of this would be the amount of CO_2 gas that comes out of a combustion reaction. These variables may or may not be measured.

As we consider a controls problem. We are able to look at two major control structures.

1. Single input-Single Output (SISO)- for one control(output) variable there exist one manipulate (input) variable that is used to affect the process

2. Multiple input-multiple output(MIMO)- There are several control (output) variable that are affected by several manipulated (input) variables used in a given process.

- **Cascade:** A control system with 2 or more controllers, a "Master" and "Slave" loop. The output of the "Master" controller is the setpoint for the "Slave" controller.
- **Dead Time:** The amount of time it takes for a process to start changing after a disturbance in the system.
- **Derivative Control:** The "D" part of a PID controller. With derivative action the controller output is proportional to the rate of change of the process variable or error.*
- **Error:** In process controls, error is defined as: Error = setpoint - process variable.
- **Integral Control:** The "I" part of a PID controller. With integral action the controller output is proportional to the amount and duration of the error signal.

- **PID Controller:** PID controllers are designed to eliminate the need for continuous operator attention. They are used to automatically adjust system variables to hold a process variable at a setpoint. Error is defined above as the difference between setpoint and process variable.
- **Proportional Control:** The "P" part of a PID controller. With proportional action the controller output is proportional to the amount of the error signal.
- **Setpoint:** The setpoint is where you would like a controlled process variable to be.

1.5 Design Methodology for Process Control

1. **Understand the process:** Before attempting to control a process it is necessary to understand how the process works and what it does.

2. **Identify the operating parameters:** Once the process is well understood, operating parameters such as temperatures, pressures, flow rates, and other variables specific to the process must be identified for its control.

3. **Identify the hazardous conditions:** In order to maintain a safe and hazard-free facility, variables that may cause safety concerns must be identified and may require additional control.

4. **Identify the measurables:** It is important to identify the measurables that correspond with the operating parameters in order to control the process.

Measurables for process systems include:

- Temperature
- Pressure
- Flow rate
- pH
- Humidity
- Level
- Concentration
- Viscosity
- Conductivity
- Turbidity
- Redox/potential
- Electrical behavior
- Flammability

5. **Identify the points of measurement:** Once the measurables are identified, it is important locate where they will be measured so that the system can be accurately controlled.

6. **Select measurement methods:** Selecting the proper type of measurement device specific to the process will ensure that the most accurate, stable, and cost-effective method is chosen. There are several different signal types that can detect different things.

These signal types include:

- Electric
- Pneumatic
- Light
- Radiowaves
- Infrared (IR)
- Nuclear

7. **Select control method:** In order to control the operating parameters, the proper control method is vital to control the process effectively. On/off is one control method and the other is continuous control. Continuous control involves Proportional (P), Integral (I), and Derivative (D) methods or some combination of those three.

8. **Select control system:** Choosing between a local or distributed control system that fits well with the process effects both the cost and efficacy of the overall control.

9. **Set control limits:** Understanding the operating parameters allows the ability to define the limits of the measurable parameters in the control system.

10. **Define control logic:** Choosing between feed-forward, feed-backward, cascade, ratio, or other control logic is a necessary decision based on the specific design and safety parameters of the system.

11. **Create a redundancy system:** Even the best control system will have failure points; therefore it is important to design a redundancy system to avoid catastrophic failures by having back-up controls in place.

12. **Define a fail-safe:** Fail-safes allow a system to return to a safe state after a breakdown of the control. This fail-safe allows the process to avoid hazardous conditions that may otherwise occur.

13. **Set lead/lag criteria:** Depending on the control logic used in the process, there may be lag times associated with the measurement of the operating parameters. Setting lead/lag times compensates for this effect and allow for accurate control.

14. **Investigate effects of changes before/after:** By investigating changes made by implementing the control system, unforeseen problems can be identified and corrected before they create hazardous conditions in the facility.

15. **Integrate and test with other systems:** The proper integration of a new control system with existing process systems avoids conflicts between multiple systems.

1.6 References

1. Romagnoli, Jose A. Introduction to Process Control. s.l. : CRC press, 2006.

Section 2. Introduction to DCS: Control system hardware

2.1 Introduction

Digital Control Systems(DCS) also known as Distributed Control System is the brain of the control system. It is used mainly for the automation of a manufacturing process and manages the logic that exist for major unit operations. A DCS in the past was tailor made for the process, plant or company that intended to use the structure to control and model it's process. Before the beginning of the DCS era there were pneumatic devices that controlled process and engineers manually turned valves on the site. Modeling of the systems was made possible by DCS as it allowed the ability to record and manage process from comfort of a computer screen. Because of DCS we are able to control processes remotely and gain a better understanding of how the process operate and how they can be improved to both increase safety and increase profit possibilities.

Control Systems are collectively named as "ICSS" Integrated Control and Safety System. Distinctly identified as "BPCS" Basic Process Control System. "SIS" Safety Instrumentation System. "F&G" Fire and Gas System.

DCS is employed in BPCS as well as used and prevalent control system. How does a DCS work?

In the field you have sensors and gauges that give and recieve information. They convert this information into a electric signal that is sent to a control room somewhere in the field. This control room has programmed logic that is able to converts the signal into a pressure, flow rate, concentration, temperature, or level. This logic also contains the information that controls the process and takes the signal compares it with the set point sent from the operator may or may not be in the field and sends a signal to the manipulated variables in the field. The DCS covers all of the computer logic from the operator screen to the field box that contain the logic.

Shutdown systems

Shutdown systems are the emergency setting of the logic to make sure the process can be contained and is environmentally safe. These setting are important for emergency response of the system. It is the job of the DCS to contain the logic for the shutdown system and be able to operate when a process exceed a certain limit.

Section 3. Current Significance: Process controls and you

Authors: (December 12, 2009) Steve Dzul, Steve Pankratz, Derrick Boroski

3.1 Introduction

Industrial processes are central to the chemical engineering discipline. Generally, processes are controlled in order to do things such as maximize safety, minimize cost, or limit effects on the environment. This course aims to help undergraduate engineering students understand the mechanisms used to moderate these processes, such as to control their output.

3.2 Automation

Generally, process controls are designed to be automated. This means that given a change in system response, the control system can act on its own to account for it. In order to minimize cost, automated systems have become widespread throughout industry. Before automation, a huge amount of labor would be required to run even the simplest processes. For example, a technician might be hired to monitor the temperature in a reaction vessel, and operate a valve to manipulate the cooling water flow rate in the jacket. In a sense, this technician operated as a control system. If the temperature reading is too high, the technician will manipulate the system in order to bring the temperature down. Via automation, this simple, arduous labor can be done by an algorithm.

By designing an effective control system, even the most complicated of processes can be run with minimal worker supervision. Telephone operators, for example, have largely been replaced by automated telephone switch boards. Removing the need for telephone operators decreases operating cost for phone companies, thereby allowing the general consumer to pay less for phone service. Automated process controls, therefore, are enormously important in the modern world.

Section 4. Failures in Process Control: Bhopal, Three Mile Island

Authors: (December 12, 2009) Steve Dzul, Steve Pankratz, Derrick Boroski

Process controls can have a huge impact on surrounding communities, as well as the environment. An engineer of a large-scale process, therefore, has an important ethical responsibility to operate a process safely and properly. These responsibilities extend well beyond the scope of merely the company for which they work. Catastrophic failures in process control remind us of the importance of control systems and engineering in today's world.

4.1 Bhopal, India Disaster

4.1.1 Overview

The Bhopal Gas Tragedy in Bhopal, India on December 3, 1984 was a large toxic gas leak that killed thousands of people in the surrounding area. A tank with 42 tons of methyl isocyanate(MIC) was contaminated with water. This in turn caused a run away reaction that greatly increased the pressure and temperatures in the tank, which forced the emergency venting of the toxic gases to the atmosphere.

This tragedy was largely due to the failure or lack of safety controls:

1. Runaway reaction as temperature and pressure increased without regulation
2. MIC was supposed to be cooled, however in the Bhopal plant the refrigeration system was not turned on. Temperature control on the tank could have greatly hindered the runaway reaction that ensued with the addition of water.
3. Flare tower to handle the leakage of toxic gases was not functional
4. The plant also had vent scrubbers, which were also not functional
5. Water curtain, which would neutralize some escaping gas, not designed properly. It was not tall enough to reach the top of the flare tower, making it essentially worthless.
6. Alarms that would have alerted to a malfunction in the tank had not been operational for 4 years

The figure below illustrates some of these failures:

1. Runaway reaction in tank 610
2. Cooling system not in use so fails to slow reaction
3. Flare tower not connected so gas not burnt off
4. Vent scrubber not operating so gas can't be neutralised with caustic soda. Gas escapes from 30m high vent
5. Water curtain set at 12-15m so can't neutralise gas with spray

SOURCE : India Today
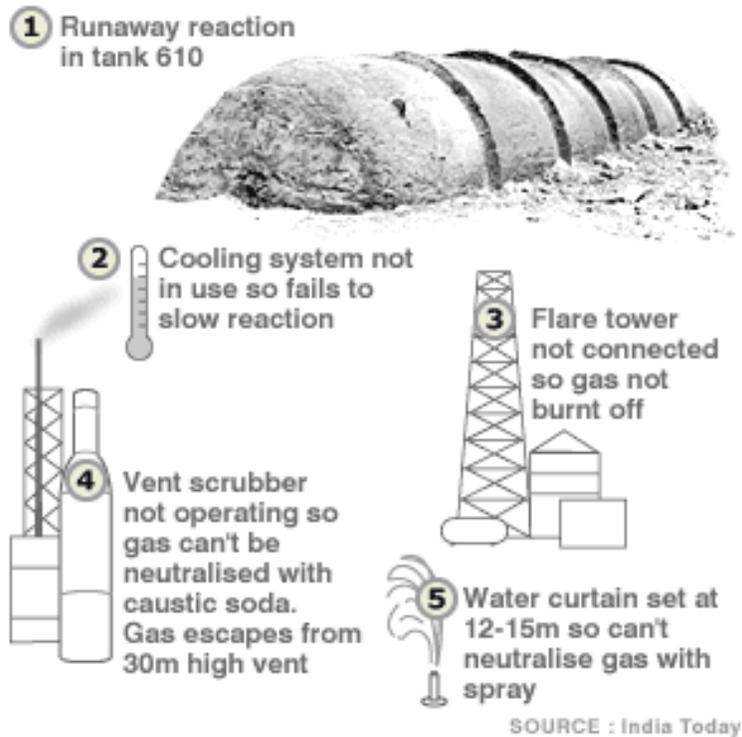
Had at least some of these been functioning the amount of toxic gas released would have been substantially reduced.

4.1.2 Results

From this tragedy we can see that if the plant had proper safety controls the effects of the disaster would have been greatly reduced. Therefore as a chemical engineer it is our responsibility to society to provide sufficient safety controls to chemical processes in order to prevent disasters such as the Bhopal Gas Tragedy from happening. Unfortunately, industrial negligence is still a problem in many third-world countries.

4.1.3 Reference

http://www.corrosion-doctors.org/Pollution/bhopal.htm

4.2 Three Mile Island Disaster

4.2.1 Overview

One of the largest and most far reaching plant failures in United States history took place at a nuclear power plant on Three Mile Island in March 1979. The event was caused by either a mechanical or electrical failure of the main feed water pumps causing the power plant to begin to overheat. As the heat increased, the control scheme caused the turbine

12

and reactor to shut down. This caused a pressure increase in the primary system (nuclear portion of the plant) and a relief valve automatically opened to release some of the pressure to prevent the reactor from blowing. All of these actions were well designed to prevent a significant event from happening. The problem was that the release valve did not close properly when the pressure in the reactor was relieved. As a result, when the reactor started back up, coolant in the core of the reactor was lost through the pressure relief valve. Because there was no control mechanism that measured the level of the coolant in the reactor, the operators, who only judged the water level by the pressure in the reactor, actually decreased coolant flow to the reactor.

The figure below is a simplified diagram of the TMI-2-plant:
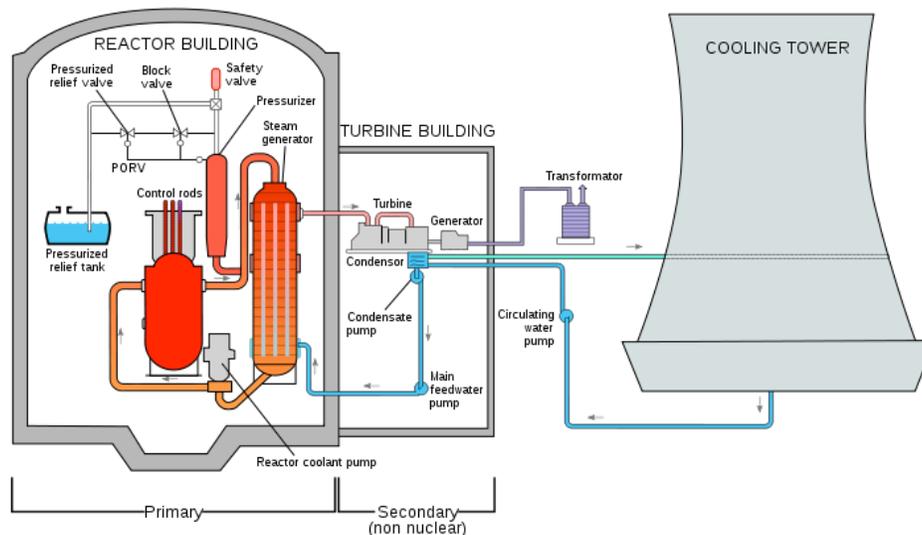
The result of the control design failure that prevented the operators from cooling the reactor was that the rods that held the nuclear fuel melted causing the fuel to also melt. This is the worst thing to have happen in a nuclear power plant and is what happened to cause the disaster at Chernobyl. Thankfully, the accident was largely contained and although the entire nation watched for 3 days as the threat of an explosion or breach of containment loomed, 0 deaths or injuries resulted. In fact, corrective steps were so successful that the average increase in radiation to the surrounding population was around 1% and the maximum increase at the boundary to the site is estimated to be less than 100% of the natural background radiation present in the region.

4.2.2 Results

The accident at Three Mile Island showed the importance of proper design of control systems. As a result the US Nuclear Regulatory Commission took steps to tighten their regulation and increase the safety requirements on Nuclear Power Plants. These included

revamping operator training as well as increasing the design and equipment requirements. This also brought the dangers of all industrial processes to the forefront and reminded people of the importance of the safety of the communities surrounding chemical and power plants.

Unfortunately, the incident also inspired intense fear of nuclear power in the general population and is partially responsible for the reduced build rate for new nuclear power plants since that time. Although control failures can be corrected fairly quickly, after one safety issue it is difficult to convince the general public that engineers have fixed the problem and that it will not happen again.

4.2.3 Reference

References: http://www.nrc.gov/reading-rm/doc-collections/fact-sheets/3mile-isle.html: The United States Nuclear Regulatory Commission

4.3 Texas City Disaster

Authors: (December 14, 2009) Virgil Humes, Fred Garner

4.3.1 Overview

In March of 2005, 15 people were killed at the of a refinery explosion in Texas City, TX. The refinery is the third largest in the country with the capability to process over 400,000 barrels of crude oil daily. On this particular day, the isomerization unit (whose purpose is to boost the octane rating of fuels) was being started up. As part of the normal start-up sequence, operators began feeding hydrocarbon liquid into the tower. However, a discharge valve that should have been opened was overlooked. A high level alarm was ignored and a second high-level alarm had not yet been re-enabled from the various maintenance activities that had been going on while the unit was down. Upon realizing that the tower was filling, a discharge valve was opened to release the hot fluid from the bottom of the vessel. This fluid then passed through a heat exchanger, warming the fluid that was still being pumped into the tower. The boiling that resulted in the bottom of the tower caused liquids to spill over the top of the stack and into a blow down drum with an atmospheric vent and no flare. The blow down drum could not contain the volume and began ejecting liquid and vapor hydrocarbon from the atmospheric vent. The vapors were soon ignited by a diesel truck with its ignition on. Fifteen people in a nearby trailer were killed in the resulting explosion. Inadequacies in the written start-up procedures, operator training, and the design of the safety relief system led to tragic and unnecessary loss of life. But examples such as these serve to make process environments even safer places to work in and around.

An animation detailing the circumstances surrounding the explosion was released by the U.S. Chemical Safety Board during the investigation. Some accounts of the incident vary

in a few details, but all are in agreement that the written procedures, operator training, and some aspects of the safety system design were to blame.

4.3.2 Results

After the incident at the refinery, representatives from BP said the company would eliminate all blow-down drums/vent stack systems in flammable service. Also, the Chemical Safety Board (CSB) recommended that BP commission an independent panel to investigate the safety culture and management systems of BP. The findings of the panel showed that BP management had not distinguished between occupational safety and process safety. This led to new implementations of process safety throughout BP North America.

4.3.3 Reference

CSB Final Report into March 2005 Texas City Disaster

Wikipedia - Texas City Refinery (BP)

Section 5. Process Controls in Everyday Life: Applying process control thinking to everyday situations

Written By: Andrew Sereno, Randy Tin (FA09)

5.1 Introduction

Have you ever wondered if the abstract and arbitrary subject of chemical engineering process controls has any relevance outside of Dow 1013? Below are a few examples of how the concepts learned in process controls can be applied to situations in your everyday life.

5.2 Examples of Process Control for the Common Man

5.2.1 Consumption

Spicing Up Soup

Pavlo LaBalle just finished scaling up a mixing tank for use in the manufacture of horse glue. Scaling up mixing tanks for horse glue is hard work - enough to make any man crave a hearty bowl of spicy liver soup. Fortunately, Mrs. LaBalle has a batch of liver soup ready for Pavlo's return home. Using his built-in composition sensor, the tongue, Pavlo realizes the liver soup spice concentration is below his specified setpoint - the soup is not spicy enough. Having never spiced his own soup, but not wanting to offend Mrs. LaBalle, Pavlo attempts to increase the spiciness of the soup using his favorite hot sauce, Sriracha. In order to prevent catastrophic overspicing, Pavlo adds a drop of Sriracha (a differential amount) and immediately begins sampling the soup after addition. In this way, Pavlo is able to determine the effect of a small disturbance (the introduction of Sriracha) on the process output (the soup's spiciness). Using his new knowledge of Sriracha's effect on the soup, Pavlo is able to optimize his addition of Sriracha so as to not catastrophically overspice the soup. Pavlo then adds a new amount of Sriracha estimated to bring the soup closer to the spice setpoint and samples the soup shortly after. By tasting the soup after the addition of spice, Pavlo is able to bring the soup closer to his desired spice setpoint through the use of feed backward spice control. By applying his education in chemical engineering process controls, Pavlo has enabled himself to enjoy his spicy liver soup.

Imbibing Alcohol

Jimmy Johnson just finished a presentation and a final report for Chemical Engineering 460. Unfortunately, during the presentation, Jimmy's personal record of 80 straight hours without sleep came to an abrupt end as he blacked out mid-line, and he was subsequently criticized for insufficiently displaying the iron resolve befitting of a Michigan engineer. Jimmy decided to drown his sorrows at Good Time Charley's through the careful ingestion of mood-improving fluids. Jimmy planned on drinking until he felt a moderate buzz, then leaving to finish his graduate school application to pursue a masters degree in English.

Jimmy, inexperienced with the process of satiating his thirst with such beverages, planned to continuously drink until he felt adequately affected. However, Jimmy failed to take into account the dead time inherent in his body's response to alcohol. Jimmy was unable to control his body's response as he had planned, resulting in a vast overshoot into drunken stupor.

Jimmy woke up with no pants the next morning. Fortunately, by applying his education in chemical engineering, he earned the necessary experience for tuning the amount he should drink in the future.

Purchasing Food at U-Go's

Rachel Malta was busy typing her technical report on the design space of a distillation column in a pilot horse glue plant. It was 11:50 p.m., and Rachel was making good progress until she felt the odd sensation of bubbling acids corroding through her stomach wall, at which point she decided to visit U-Go's to find something to eat.

Confronted with the shop's wide array of processed foods and saturated fats, a difficult problem lay before Rachel. If she purchased too much food, her hunger set point would be reached, but she would be left with bags of uneaten popcorn and sour patch kids. If she purchased too little, she would remain hungry and would be unable to return to purchase more food after the store closed. Rachel needed to make a prediction of how much food her digestion process would require to reach her desired fullness. Based on her previous experiences with U-Go's "food," her predictive control (her memory) decided to purchase a bag of barbecue chips, one apple, and a bagel. By applying her education in chemical engineering, Rachel was able to successfully predict how much food would satisfy her hunger without being forced to find more food.

5.2.2 Hygiene
Filling a Bathtub

Lan Ri has been in the Duderstadt Center for the past 3 nights working on his ChE 487 project. Having not bathed for 60 hours, Mr. Ri decides it is time to wash himself for the sake of his fellow Duderstadt dwellers, however Mr. Ri has forgotten how to stand up (he's been sitting at a computer for 60 hours). The ever creative Mr. Ri decides to take a bath instead of his usual standing shower. At the bath, Mr. Ri finds himself confronted with two flow rate controllers: one "hot water feed" controller and one "cold water feed" controller. Turning the controls, he realizes that they control the flow rate of the bathtub feed streams. Yet Mr. Ri has a problem - the hot water feed temperature, as measured by his built-in temperature sensor (his skin), is far too high. The cold water feed temperature, also measured using his skin-type temperature sensor, is too low for a comfortable bath. Accordingly, Mr. Ri performs a quick energy balance to determine expected ratio of the feeds necessary to achieve a bath temperature close to what his desired setpoint. However, because Mr. Ri has poor control of the actual flow rate of each stream (hot and cold), the actual resulting temperature is still below that which Mr. Ri desires. Mr. Ri then readjusts the hot-to-cold feed ratio by increasing the hot water feed

flow rate. As the bathtub fills up, Mr. Ri uses his built-in level sensor (his eyeballs) to turn off the valves controlling the feed stream flow when the bathtub level is at his calculated level. Mr. Ri was able to calculate the level to which the bathtub should be filled by using his knowledge of static fluid systems. By applying his education in chemical engineering process controls, Mr. Ri has enabled himself to wash away the smell of the Duderstadt Center.

[Can use some diagrams, pictures, and/or equations to describe these problems - RZ].

18

Chapter 2. Modeling Basics

Section 1. Verbal Modeling: process description, control specifications, and connections

Note: ⊞Video lecture available for this section!

Authors: (September 8, 2006) Brian McQuillan, Crystal Miranda, Brandon Quigley, and John Zhang

Stewards: (September 5, 2007) Kevin Luchi, Mike Nappo, Matt Neff, Lisa Schauman

- First round reviews for this page
- Rebuttal for this page

1.1 Introduction

Every process requires a great deal of planning in order to successfully accomplish the goals laid out by its designers and operators. In order to accomplish these goals, however, personnel that are not familiar with the design must fully understand the process and the functions of the control systems. Control systems consist of equipment (measuring devices, valves, etc.) and human intervention (plant operators and designers). Control systems are used to satisfy three basic needs of every process:

1. Reduce the influence of external disturbances
2. Promote the stability of the process
3. Enhance the performance of the process

Verbal modeling is used for creating and also understanding a process control system. Verbal modeling consists of first receiving and then gathering information about the process. A step-by-step process is then used to describe the control systems used to satisfy constraints and objectives that have been outlined. In the following sections you will read what requirements are generally outlined for process control and the step-by-step method used to meet these requirements.

1.2 Prerequisite Information Regarding a Process

For the sake of this article it is assumed that a process has already been designed and that certain restraints and criteria are provided by either a customer, management, or the government. The goal of this section is to classify the types of criteria that are usually given. These criteria then become the conditions that the control systems employed must

satisfy. In general, there will be five sets of criteria, often coming from different people and institutions. By gathering all of these criteria you will be able to describe the control system. If you do not have a complete list of these criteria you must research the process to determine these constraints before beginning the step-by-step process below.

Safety

The safe operation of a process is the biggest concern of those working in the plant and those that live in the surrounding community. The temperatures, pressures, and concentrations within the system should all fall within acceptable limits, and these limits can be dictated by either government agencies or company policy.

Production Objectives

The production objectives usually include both the amount and purity of the desired product. This criterion is generally set by the company or customer.

Environmental Regulations

These come in the form of restrictions on the temperature, concentration of chemicals, and flowrate of streams exiting a plant. State and federal laws, for instance, may dictate the exit temperature of a cooling water stream into a lake in order to prevent harm to aquatic wildlife.

Operational Constraints

Equipment found in the plant may have their own unique limitations, such as temperature or pressure that require proper control and monitoring. For instance, a thermocouple may be damaged at extremely high temperatures, thus the location of the thermocouple must be accounted for.

Economics

In general, a company will operate so that its profits are maximized. The process conditions that maximize these profits are determined by way of optimization. Many costs must be considered when optimizing process conditions. Some of these costs are fixed, or will not change with process variables (i.e. equipment costs) and others are variable, or do depend on process variables (i.e. energy costs). The overall process is usually limited by certain factors including availability of raw materials and market demand for the final product. Therefore, the economics of a process must be well understood before process changes are enforced.
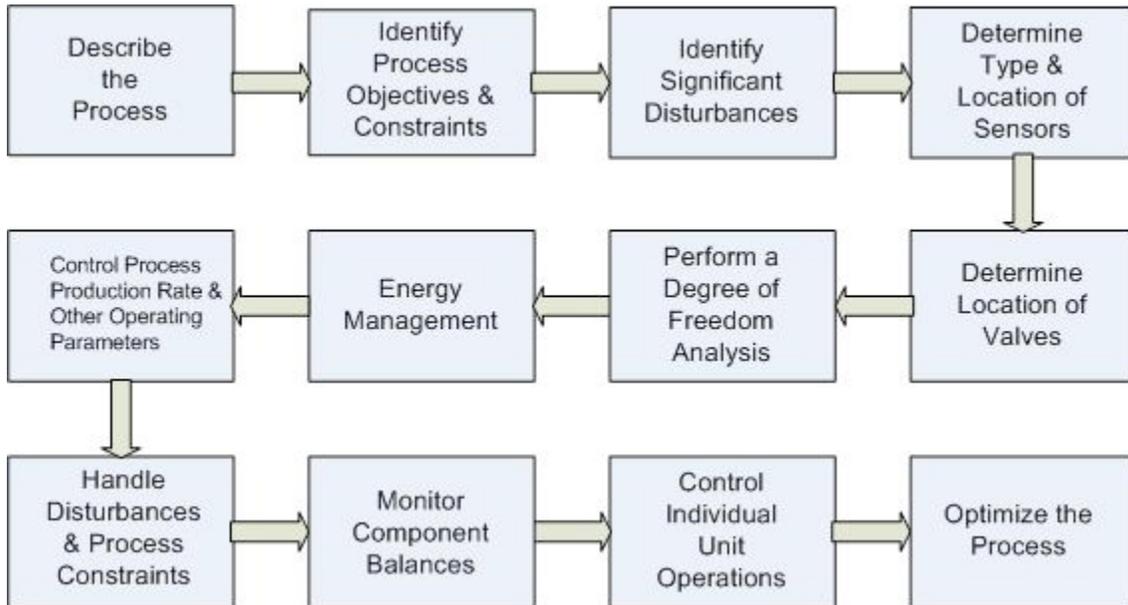
1.) Describe the Process

A brief description of the general process is needed while not dwelling on the details and calculations involved. The major steps of the process, as well as inputs and outputs of the process, should be stated. A simple diagram should be provided detailing the chemical process to help visualize the process.

2.) Identify Process Objectives and Constraints

The objectives and constraints of the process must be identified before process control actions can be performed.

The process objectives include the type, quantity, and quality of the product that is to be produced from the process. The economic objectives, such as the desired levels of raw material usage, costs of energy, costs of reactants, and price of products, should also be identified.

The process constraints include three different categories: operational, safety, and environmental limitations. *Operational constraints* refer to the limits of the equipment used in the process. For instance, a liquid storage tank can only hold a certain volume. *Safety constraints* describe the limits when the people or the equipment may be in danger. An example would be a pressure limitation on a reactor, which if exceeded, could result in an explosion. *Environmental constraints* limit how the process can affect the

immediate surroundings. For example the amount of harmful chemicals that can be released before damage is done to nearby water supplies. All of these constraints should be mentioned to build a robust control system.

Careful reading of the information provided to you by the customer, management, and government is required in order to properly identify each constraint and objective. Often times, the process objectives will be very clearly laid out by the needs of the customer or management. Operational constraints, or the limitations of the equipment being used, must be researched for each piece of equipment used in the process. Generally, by satisfying the operational constraints a good portion of safety constraints are satisfied as well, but additional safety constraints may exist and must be investigated by researching company policy and governmental regulations. Environmental regulations also have to be researched through resources such as the EPA and Clean Air Act. Satisfying the economic aspect is largely determined by manipulating additional variables after all other constraints and objectives have been met.

3.) Identify Significant Disturbances

Disturbances, in the sense of process description, are defined as inputs or external conditions from the surrounding environment that have certain properties that cannot be controlled by the plant personnel. Examples of disturbances include ambient air temperature, feed temperature, feed flow rate, feed composition, steam pressure changes, and cooling water temperature changes. Disturbances can drastically affect the operation of a unit. A control system should be able to effectively handle all process disturbances. As such, all possible disturbances must be identified and these disturbances need to be accounted for by the development of contingency plans within the process.

4.) Determine Type and Location of Sensors

A proper design must ensure that adequate measurements of the system are obtained to monitor the process. To meet this goal, sensors must be chosen to accurately, reliably, and promptly measure system parameters. Such parameters include temperature, flow rate, composition, and pressure. Placement of sensors is important both in the usefulness of measurements as well as the cost of the system. Sensors should be placed such that the measured quantities are appropriate in addressing control objectives.

5.) Determine the Location of Control Valves

Valves must be placed in a location to control variables that impact the control objectives. For example, control of the temperature of a reactor could be obtained by placing a valve on either the stream of heating / cooling fluids or by placing a valve on

the feed stream to the reactor. One must determine which streams should be manipulated to meet process objectives.

6.) Perform a Degree of Freedom Analysis

The degrees of freedom in a system are equal to the number of manipulated streams (determined in step 5) minus the number of control objectives and control restraints (determined in step 2). A degree of freedom analysis is used to determine if a system is being under- or over-specified by the process objectives. The degrees of freedom come from the number of knowns and unknowns that are specified within the system. If there are extra degrees of freedom present in a system, unused manipulated variables can be used to optimize the process. If there are negative degrees of freedom, a system is over-specified because more objectives and restraints exist than manipulated streams. In this case, all objectives cannot necessarily be met simultaneously and the least important objectives must be neglected. A system with zero degrees of freedom is fully specified. All objectives can be met, but there is no room for optimization.

7.) Energy Management

In any system with exothermic or endothermic reactions, distillation columns, or heat exchangers, energy management becomes a factor that must be accounted for. Heat must be removed from exothermic reactions in order to prevent reactor runaway, and heat must be supplied to endothermic reactions to ensure desired production rates. Strategies such as pre-heating feed streams with the excess heat from a product stream are helpful in maintaining efficient usage of energy, however, they also result in more complex processes that may require more intricate control systems.

8.) Control Process Production Rate and Other Operating Parameters

The production rate can be controlled by a variety of manipulated variables. One manipulated variable may be the feed rate. The plant feed rate can be changed and each subsequent unit can use its controls to accommodate this change, ultimately resulting in a change in the final production rate. Other manipulated variables may also include reactor conditions, such as temperature and pressure. Temperature and pressure affect reaction rates and can be used to alter the final production rate. It is important to choose the most suitable manipulated variable to control production rate.

In addition to the production rate, other control objectives must be effectively managed by manipulated variables. For example, temperature of an exothermic reactor may be controlled by the flow of a coolant stream passing over it in order to avoid dangerous high temperatures. The pressure of a reactor may be controlled by the flow of feed gas in order to comply with the pressure limitations of the vessel.

9.) Handle Disturbances and Process Constraints

The effects of disturbances should be minimized as much as possible, in order to maintain the system at desired conditions and meet all process objectives and constraints. Feedback or feedforward are specific control techniques and are common ways to overcome disturbances. A feedback control works by studying the downstream data and then altering the upstream process. The actions executed are reactive. Feedback can be viewed as an if-then statement: if a feed's temperature is detected to be lower than desired, then steam can be used to preheat the feed. Feedforward is a more proactive approach in that it adjusts a manipulated variable before the disturbance is felt in the process. Hence, if a sensor indicates low temperatures upstream of the feed, the feedforward control will counteract the effect of the cooler upstream temperatures by preheating the feed before the feed temperature is effected. Note that a disturbance must be detectable and measurable in order for the feedforward control to fix the anticipated disturbance before the system is effected.

Additionally, if constraints are reached during the process, controls should be implemented to avoid safety, operational, or environmental hazards. This can also be done with feedback and feedforward controls on manipulated variables.

10.) Monitor Component Balances

Every component within a process, whether it is inert or not, should be accounted for at every step of the system in order to prevent accumulation. This step is more crucial in processes that involve recycle streams. If such a stream is present, a purge stream is often necessary to remove unwanted components. In addition, component balances are used to monitor yield and conversion or reveal locations in the process where loss may be occurring. In order to monitor component balances, composition sensors are used.

11.) Control Individual Unit Operations

Most systems used today in industry employ the use of multiple unit operations. Each of these unit operations, however, needs to be fully controllable in the sense that it has a control system that can adjust manipulated variables in order to maintain other parameters. For instance, if an absorber is present, the system must be able to control the liquid solvent feed as some ratio to the gas feed. Another example is a crystallizer. The refrigeration load of the crystallizer must be controllable in order to control the temperature.

12.) Optimize the Process

In most cases, there will be certain aspects of a process that will not be dictated to a designer and can be changed to make the overall process more economical for the company. These are referred to as "unaccounted for" degrees of freedom and can be implemented as new control valves or adjustable controller setpoints.

1.4 Alternative Method of Verbal Modeling

This alternative method is also described in Peter Woolf's Recorded Lectures Lecture 3. For purposes of this description, it is scaled down to a single unit process. However, this method can easily be applied to describe an entire system of unit processes.

1. Describe the process in words

Some of the important questions to answer before delving deeper into a model are:

- What are the components entering the system?
- How do they enter? Separately? Combined stream? What physical states are they in?
- What happens inside the unit process and what comes out at each exit point?

Remember to keep this part simple. There is no need to include chemical formulations or equations of any sort. Just lay out the basic flow of material.

2. Define the primary goal of the process

The primary goal should be simple. Often, it is to maintain a specific measured variable above a minimum or below a maximum. In this step, the only thing that needs to be determined is what the main goal is, and a few supporting details about why this is an important goal to achieve.

For example, a primary goal could be to minimize the concentration of Compound Y in orange juice because studies show Compound Y gives the juice a bad aftertaste.

3. Identify secondary processes that influence the primary goal

In a typical unit process, the primary goal will be directly influenced by one or two other aspects of the system. These can include temperature, pressure, inlet conditions, and more and can occur at various points in the process.

The goal of this step is to determine which of these other process variables will be most likely to influence the primary goal and to step down from there.

For example, the temperature of the orange juice mixer could have the greatest influence on production of Compound Y.

4. Identify safety and environmental risks

Next, you need to identify all of the points in the process that represent any type of risk. This will be important later in determining which system variables need to be monitored.

Step through your process and identify any points that pose a significant risk of the hazards shown in the following figure.
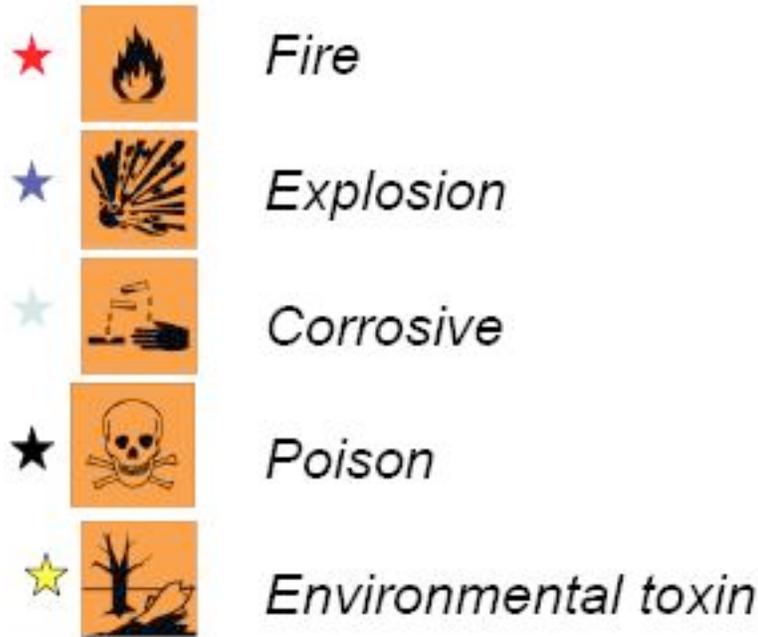
Examples include: Boilers represent fire and explosion risks. Any stream with a dangerous chemical can represent corrosive, poison, environmental, or all three risks.

5. Identify major costs associated with the process

How much something costs to produce is obviously a big deal in manufacturing. Identifying the largest sources of cost is critical in finding ways to reduce cost overall. Typical places to start identifying costs are at inlet streams (what is the cost of raw materials) and at any portion of the process where heat is added or removed.

It is important to include the high costs that can be associated with the risks identified in Step 4. Often the high cost of failure and risk exposure will determine what other seemingly costly steps must be taken to ensure the safety of the process.

6. Identify variables you can directly manipulate

The basics of the process have been laid out, and now it's important to determine what variables you can actually control. Typically, you only have direct control over the simplest of variables: switches and valves. Essentially, this just means that you cannot, in fact, choose a temperature for your system and implement it. What you can do, is control a valve or switch that activates heating or cooling to control the temperature.

During this step, you must decide where it is important to place these valves and/or switches. Use the information acquired previously about the primary goal and secondary effects to determine what variables are worth controlling. Remember that you don't need to put valves EVERYWHERE! Control valves are not costless and can also add unwanted complexity to your system. If you need to isolate equipment, you can install manual valves. Keep the control valves to the needed level.

7. Identify sources of variation

In order to write a control scheme, you need to know what values in your system will change and why. Some common causes of variation include:

- Environment: ambient temperature
- Other processes upstream or downstream: variable inlet conditions or outlet demand
- Economic forces: product worth, material costs
- Operators

Identifying what aspects of your process can be affected by these forces will allow you to assemble a more complete control scheme.

8. Describe your control system in words

Before you start trying to write everything out in computer code and mathematical equations, take the time to lay out your controls in words. This is much like preparing an outline before writing a paper. It can save you from many headaches later on.

One example of generic, simple syntax for verbal modeling is: Maintain [system variable] at specified level by adjusting [variable I can control].

1.5 The Barkel Method of Verbal Modeling

This method is an elaboration upon the steps outlined in Mr. Barry Barkel's lecture on September 29, 2009.

"If you design a system, you have the ethical responsibility to control it."

1) Understand the Process

Before you can control anything, you have to understand the process and how different parts of the system interact with each other. Make sure that the overall process is understood. This includes inputs and outputs as well as major steps, however specifics are not necessary. You should also be able to construct a diagram to help explain the process.

2) Identify Operating Parameters

Operating parameters can include temperature, pressure, flow, level, etc. Choose the parameters to manipulate in your system that will safely result in the desired output.

3) Identify hazardous conditions

Consider all possible dangerous aspects of your process when designing your system. This could include a chemical overheating or a vessel overflowing. It is imperative that you ensure the safety of your operators and being aware of all hazardous conditions can aid in this.

4) Identify measurables

The main three measureables addressed in this class are: temperature, pressure, and flow. However, there are many more measureables, some more common than others. Here are some more: pH, humidity, level, concentration, viscosity, conductivity, turbidity, redox/potential, electrical behavior, and flammability.

5) Identify points of measurement

It is important to place sensors in locations so that efficient and reliable measurements are taken of the system to monitor the process. For example, in a distillation column, temperature sensors will display different temperatures at different locations down the tower. The sensors must be positioned so that accurate readings of the system are given. Also, it is necessary to place sensors in an area of constant phase.

6) Measurement methods (thermo couple? choose for range)

After identifying what is to be measured and where it will be measured, you have to decide how it will be measured. For example, a thermocouple can be used to measure the temperature. When choosing the equipment, be sure to check that the conditions of use fall within the recommended range of operation for the equipment.

7) Select control methods

Decide whether to use feedback, feed-forward, cascade, or other types of control methods.

8) Select control system

A control system is a set of devices that will manipulate the actions of other devices in the system. A control system ranges from having an operator manually open and close a valve to running a system with feedback such as with PID controllers. Control systems can vary from relatively cheap to expensive. When picking a system, it would be most economic to choose the cheapest one that gets the job done. An example of a control system is a PIC, or programmable interface controller.

9) Select control limits

When choosing setpoints for controllers, you will also have to decide on a range that the values are allowed to fluctuate between before a corrective change is made. When selecting these limits, keep in mind that "equal" and "zero" do not exist due to the infinite number of decimals that electronics are now able to handle. So you must further define what "equal" means, when is it "close enough" to count as "equal" And when is the number small enough to count as "zero", is 0.1 or 0.01 or 0.0000000001 count as "zero"?

10) Define control logic

As every process is different, a customized code for each process must be written to tell the system what to do. For example, when a level control is a tank has reached a critically high point, the logic should spell out the necessary changes needed to bring the tank level back down. For example, this could be partially closing a valve upstream of the tank or partially opening a valve downstream of the tank.

11) Create redundancy system

In the real world, you must balance cost and efficiency/safety. On one hand, you don't want an out-of-control system if one control fails. But on the other hand, you can't afford to order two of everything. The critical point to keep in mind is to optimize the safety while minimizing the cost.

12) Define "fail-safe"

A fail safe is a set up in the control logic to ensure that in the event of a failure of a control method, the system will automatically reach a safe condition so that there is little to no harm done to other equipment or personnel.

13) Set lead/lag criteria

Valves and other equipment do not necessarily open/close or turn on/off at the exact instant a button is pressed or the control logic kicks in. There is often lag time associated with each controller. You must determine how long this lag time is so that you can take it into account.

14) Investigate effects of change before/after

Be sure to investigate effects of changing each controller. For example, what are the effects of closing/opening this valve?

15) Integrate all systems

Ensure that all systems are working together and that there are no holes in the system. Make sure that information does not fall through any cracks in the system.

1.6 Common Errors

1. **Impossible direct manipulations** e.g., Change the concentration of salt in a tank

2. **Missing the forest for the trees** e.g., Sacrificing product quality for tight level control on a tank

3. **Excessive or insufficient control** e.g., Control every variable because you can or ignore the possibility of significant disturbances

1.7 Worked out Example 1

PLEASE NOTE: ALL VALUES IN THE FOLLOWING PROBLEMS ARE FICTICIOUS BUT MEANT TO BE LOGICAL

A heat exchanger uses steam to heat a stream of water from 50ºF to 80ºF. The water enters at a flow rate of 20 gallons per minute from a nearby lake. The process costs $65 per hour and yields a profit of $2 per gallon of product. The steam is provided by the plant and is 1000ºF, which is a temperature that the pipes can sustain. For safety reasons, the exchanger may only run for 12 consecutive hours and requires 4 hours to cool down. Using more than 10,000 gallons of water per hour would cause an environmental disturbance to the water source. The diagram is shown below. Verbally model this system.
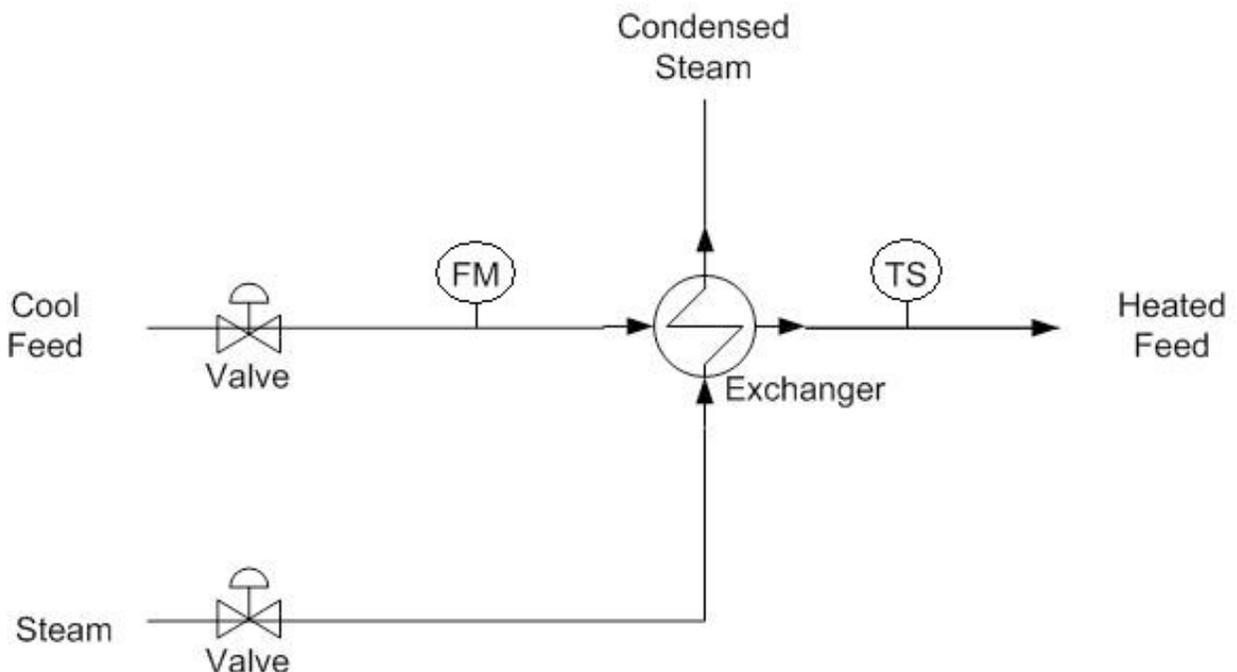
== Solution ==

1.) Describe the Process

The purpose of the process is to heat an incoming stream of water from a temperature of 50ºF to a temperature of 80ºF. The main equipment involved is a shell-and-tube heat exchanger.

2.) Identify Process Objectives and Constraints

The product specification of the process is water at a flow of 20 gallons per minute and a temperature of 80ºF.

Economically, the process costs $65 per hour to operate. There are no costs for the raw materials, as the only inputs to the system are water and steam. The finished product produces a profit of $2 per gallon. The economic objective is to reduce process costs while producing sufficient product.

The operational constraints and safety concerns are due to the pipes. The pipes can only sustain a temperature of 1000ºF. Safety is a concern because attempting to heat the incoming water to a certain temperature may cause the heat exchanger to malfunction, leading to equipment damage and possible burn injuries to nearby personnel. The system may only operate for 12 consecutive hours, after which the system will need to be cooled down for 4 hours to avoid the aforementioned hazards. A simplified assumption is that there are no constraints on steam because it is provided by the plant and causes no safety issues. The only environmental constraints involve the incoming water stream. The incoming water is gathered from the nearby lake, and a stream of greater than 10000 gallons per hour would cause a disturbance in the equilibrium of the lake.

3.) Identify Significant Disturbances

Significant disturbances can be found in the ambient air temperature, variable flow rates of the feed, and the temperature of the steam.

4.) Determine the Type and Location of Sensors

A flow sensor (FM) is placed at the incoming water stream. A temperature sensor (TS) is located on the product water stream. A flow sensor is not needed for the steam stream for

this problem because this value is not needed for control. A sensor could be placed here but the information is not needed for this problem.

5.) Determine the Location of Control Valves

A flow valve is placed at the entrance of the incoming water stream. A flow valve is placed at the entrance of the steam.

6.) Perform a Degree-of-Freedom Analysis

There are two manipulated variables: the flow of the water feed stream and the flow of the incoming steam. There are two control objectives: the flow of the feed stream, monitored by the flow sensor, and the temperature of the product, monitored by the temperature sensor. Therefore the system has zero degrees of freedom.

7.) Energy Management

The incoming steam is used to transfer heat to the cool water feed. The temperature sensor on the product stream determines the applicable setting on the steam flow valve.

8.) Control Process Production Rate and Other Operating Parameters

The process production rate is controlled by the flow valve on the entering water stream. The water temperature is controlled by the flow valve on the incoming steam.

9.) Handle Disturbances and Process Constraints

Changes in the ambient air temperature can be detected by the temperature sensor, and can be corrected by the flow valve on the incoming steam stream. Variable flow rates of the water feed stream can be detected by the flow sensor and compensated by adjustments on the flow valve on the water feed stream. Changes in the temperature of the steam can be detected by the temperature sensor. The flow valve on the steam stream can be adjusted to increase or decrease the flow of steam and subsequently the amount of heat exchanged.

10.) Monitor Component Balances

A vent is located on the heat exchanger to release excess steam from the system. Aside from that, any accumulation is unlikely and can be neglected.

11.) Control Individual Unit Operations

The outlet temperature of the product stream is controlled by the flow valve on the steam feed stream. The flow of the incoming water stream is controlled by the flow valve on the water feed stream.

12.) Optimize the Process

One might notice that the process is only using 1,200 gal/hr of water, well below the 10,000 gal/hr environmental constraint. If the profit of the process is linear with the flow-rate of water, then increasing the flow-rate of water will increase the profits for the company. (With the constraints specified, this is a Linear Programming optimization problem. The optimal setpoint falls on a boundary condition.) However, the flow-rate of water entering the system is already specified, which results in zero degrees of freedom. (Zero degrees of freedom implies there are no further control valves or setpoints.) Further investigation should be conducted to determine the reason for the flow-rate specification. When considering increasing the flow-rate of water into the system, one should also check that the other constraints are not violated.

1.8 Worked out Example 2

A process converts phenol into salicylic acid through a series of two reactors. Phenol and NaOH are fed in the liquid phase into the first reactor where it reacts with gaseous carbon dioxide that is pumped in. Assume constant fresh feed temperature and that the feed flow rate is within operational constraints. Management has dictated that salicylic acid production must be 200 moles per hour. Also, management would like the product stream to have a molar composition of 80% salicylic acid. Due to environmental concerns, a maximum flow rate of 10000 gallons per hour of cold water can be used to cool the first reaction chamber. The valve controlling the flow of cold water does not allow a flow rate in excess of 7500 gallons of water per hour. The salicylic acid product is used in another process to produce aspirin, which has a market value of $10 per mole. The first reactor can be operated at pressures up to 200 atm, while the second can be operated at pressures up to 10 atm. The first reaction is exothermic, while the second reaction is assumed to generate negligible heat. A diagram of this process is shown below, as well as the reaction scheme. Verbally model this system.
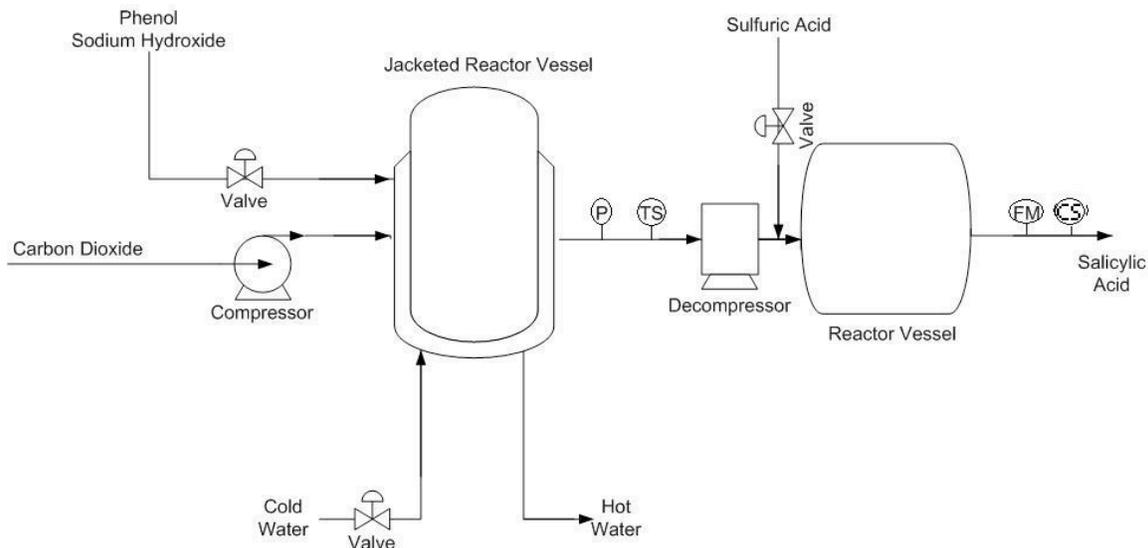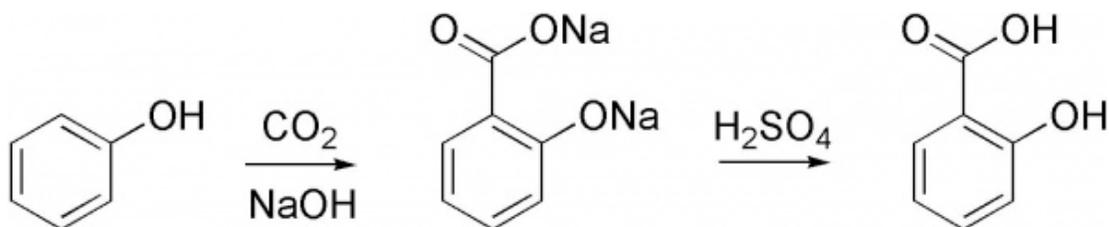
== Solution ==

1.) Describe the Process

The purpose of the process is to convert phenol into salicylic acid, the precursor for aspirin. First, the phenol is reacted with gaseous carbon dioxide (CO_2) and sodium hydroxide (NaOH) under high pressure and temperature in the first reactor. The product of this reaction is then combined with sulfuric acid to form salicylic acid in the second reactor. The reaction scheme is shown above.

2.) Identify Process Objectives and Constraints

The process is expected to produce 200 moles per hour of salicylic acid. The product stream must contain at least 80% by moles of salicylic acid. The equipment used in the process dictates the operational limitations. The first reactor vessel can be operated up to a pressure of 200 atm, while the second reactor vessel has a 10 atm upward pressure limit. As such, pressures in excess of these limits must be avoided. Since the first reactor will generate a significant amount of heat, the heat must be removed to avoid damage to equipment and possible runaway reactions. Therefore, a heat exchanger (in the form of a reactor jacket in this case) with cool water should be included to decrease the temperature of the reactor. Economic concerns demand that phenol, sodium hydroxide, and sulfuric acid should not be used in extreme excess. The costs of these materials and the energy costs required to process them affect the overall profitability, so these compounds should not be wasted. Environmental regulations limit the use of water to cool the reactor at 10000 gallons per hour, however the valve constraints limits the amount of water to only 7500 gallons per hour.

3.) Identify Significant Disturbances

The amount of cold water available to cool the reactor can be considered a disturbance because it comes from a reservoir outside of our control. The ambient temperature is also a disturbance. If it drastically increases, the amount of cold water needed to cool the reactor would need to increase as well. Composition of the feed streams will be assumed to be constant in this example. Therefore, they are not considered disturbances.

34

4.) Determine the Type and Location of Sensors

A temperature sensor (TS) and pressure sensor (P) are located on the stream exiting the first reactor vessel. A flow meter (FM) is located on the product stream leaving the second reactor. A composition sensor (CS) will also be located on the product stream leaving the second reactor. The pressure drop can be controlled through the decompressor and thus is a control.

5.) Determine the Location of Control Valves

Control valves are located on the feed stream containing the phenol and sodium hydroxide, the incoming cold water to the first heat exchanger, and the sulfuric acid feed stream. There is also a pump located on the carbon dioxide stream that enters the reactor.

6.) Perform a Degree of Freedom Analysis

There are 3 valves, 1 pump, and 1 decompressor but 5 objectives. This results in zero degrees of freedom. The valve located on the sulfuric acid feed stream is meant to meet the composition constraint placed on the product stream leaving the second reactor. The valve located on the feed stream carrying the reactants is set to satisfy production requirements. The valve on the cold water stream is used to maintain reactor temperature, which satisfies an operational constraint. The pump is to ensure the correct pressure is achieved in the reactor, also satisfying an operational constraint. The decompresser is to maintain a pressure of less than 10 atm in the second reactor, thus satisfying another operational constraint.

7.) Energy Management

The heat from the exothermic reaction in the first reactor is transferred to the cold water stream. The hot water stream exiting the reactor vessel jacket could be used to heat streams on other processes. The second reactor is assumed to generate negligible heat during the reaction, thus any release of heat from the reactor will be considered safe to release into the environment surrounding the process.

8.) Control Process Production Rate and Other Operating Parameters

The production rate is measured by the flow sensor on the product stream and this signals the control valve on the feed stream through a feedback mechanism to change the production rate as necessary.

9.) Handle Disturbances and Process Constraints

If the temperature sensor on the reactor exit stream exceeds a certain level due to a diminished cold water supply, the feed stream valve would decrease the amount of reactants entering the reactor. The amount of feed would also be decreased if more than 7500 gallons per hour of cooling water were needed, as this is an operational constraint. If the pressure gauge controlling the pump begins to read higher than allowed pressures, the pump would decrease the flow of the carbon dioxide entering the reactor. Also, if the pressure gauge reads out a pressure that will be too high for the second reactor, the decompresser will be allowed to disperse more pressure. If ambient air temperature drastically increases, the temperature sensor would open the cold water valve allowing more cooling water to enter the reactor vessel jacket. If the composition of the product stream falls below 80 mole percent of salicylic acid, then the valve controlling the sulfuric acid feed would allow more sulfuric acid into the second reactor to increase the conversion of reactants.

10.) Monitor Component Balances

The composition sensor and flow meter on the product stream leaving the second reactor will account for every species to ensure that there is no accumulation or loss within the system.

11.) Control Individual Unit Operations

The first reactor vessel's pressure is fully controlled by the pressure gauge and pump system and its temperature is fully controlled by the temperature sensor which controls the reactant feed valve and the cool water valve. The second reactor's pressure is fully controlled by the same pressure gauge and the decompresser system, and its temperature will be highly dependent on the amount of cooling water used to cool the product exiting the first reactor.

12.) Optimize the Process

Since there are no unaccounted degrees of freedom, there are no valves to adjust in order to optimize the process. It should be noted, however, that if there was no constraint on the composition of the product stream, the sulfuric acid feed valve would have become an unaccounted for degree of freedom. If this had been the case, the valve could be adjusted to maximize the profit of the process. In order to maximize the profit, the benefits of having higher conversion and more product would have to be weighed against the increase costs of using more sulfuric acid feed.

1.9 Sage's Corner

An example of verbal modeling
http://www.youtube.com/watch?v=YTs_IWccnvw&feature=player_embedded

Powerpoint Slides from the example

1.10 References

Luyben, William L., Tyreus, Bjorn D., and Luyben, Michael L. "Chapter 8: Eastman Process" in Plantwide Process Control, McGraw-Hill, New York, pp. 251-272.

Luyben, Michael L., Tyreus, Bjorn D., and Luyben, William L., "Plantwide Control Design Procedure" in AIChE Journal Dec. 1997, Vol. 43, No. 12 pp. 3161-3174.

Riggs, James B. and Karim, M. Nazmul. "Chapter 17: Multiunit Controller Design" in Chemical and Bio-Process Control, Ferret Publishing, pp. 491-504.

Stephanopoulos, George. Chemical Process Control: An Introduction to Theory and Practice, Prentice Hall, New Jersey, pp. 1-41.

Section 2. Degrees of Freedom: importance, calculation procedure, and examples

Authors: (13 December 2009) Jon Butler, Matthew J LaVelle

2.1 Introduction

In control engineering, a degree of freedom analysis is necessary to determine the regulatable variables within the chemical process. These variables include descriptions of state such as pressure or temperature as well as compositions and flow rates of streams.

2.2 Definition

The number of process variables over which the operator or designer may exert control. Specifically, control degrees of freedom include:

1. The number of process variables that may be manipulated once design specifications are set
2. The number of said manipulated variables used in control loops
3. The number of single-input, single-output control loops
4. The number of regulated variables contained in control loops

The following procedure identifies potential variables for manipulation.

2.3 The Process

The method we will discuss is the Kwauk method, developed by Kwauk and refined by Smith. The general equation follows:

Degrees of freedom = unknowns - equations

Unknowns are associated with mass or energy streams and include pressure, temperature, or composition. If a unit had **Ni** inlet streams, **No** outlets, and **C** components, then for design degrees of freedom, C+2 unknowns can be associated with each stream. This means that the designer would be manipulating the temperature, pressure, and stream composition.

This sums to an equation of

Total Unknowns = Ni*(C+2) + No*(C+2)

If the process involves an energy stream there is one unknown associated with it, which is added to this value.

38

Equations may be of several different types, including mass or energy balances and equations of state such as the Ideal Gas Law.

After **Degrees of Freedom** are determined, the operator assigns controls. Carrying out a DOF analysis allows planning and understanding of the chemical process and is useful in systems design.

2.4 Applications

Single phase systems

- All outlet streams have the same composition, and can be assumed to have the same temperature and pressure

Multiple phase systems

- An additional (C-1) composition variable exists for each phase

Complete Process

- When connecting units which share streams, one degree of freedom is lost from the total of the individual units

2.5 Example 1: Blender

This example (from the ECOSSE Control Course web page listed below) investigates degrees of freedom in a simple vapor mixing unit. Two gaseous streams enter a vessel and exit as a single well-mixed stream (Figure 1). We will apply the above equation to determine degrees of freedom.
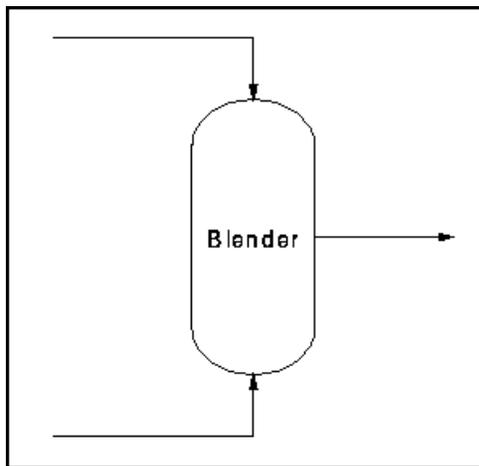
Figure 1: Blender Schematic

Here, there are 3 streams, each with C+2 unknowns for a total of **3C+6 Unknowns**.

We have **C** mass balances and **1** energy balance for a total of **2C+1** equations. We also know composition, pressure, and temperature for the incoming streams. This adds **2C+2** to the equation. Putting everything together gives:

Degrees of freedom = 3C+6 - (2C+1 + 2C+2)

Hence, the system has **3** degrees of freedom. Therefore, we can fix outlet composition, pressure, and flow rate. Figure 2 shows an example control scheme:
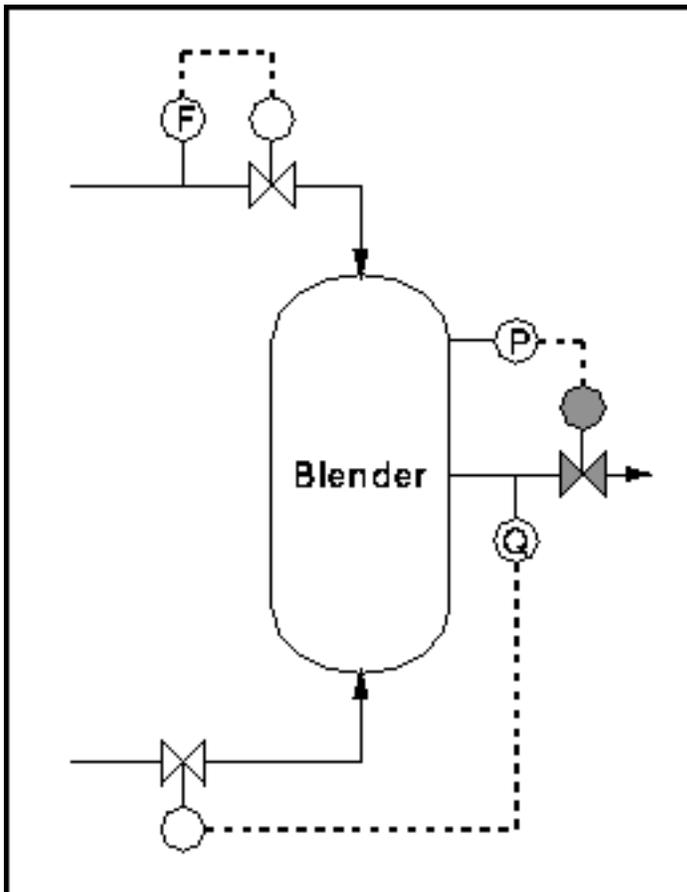
Figure 2: Blender with controls

2.6 References

Ponton JW, 1994, Degrees of Freedom Analysis in Process Control, Chemical Engineering Science, Vol. 49, No. 13, pp 1089 - 1095.

http://eweb.chemeng.ed.ac.uk/courses/control/restricted/course/third/course/formal.html

Section 3. Incidence Graphs: interpretations, consistency, and inconsistency

Note: ⊞Video lecture available for this section!

Authors: (September 26, 2006) Christopher Garcia, Anwar Stephens, Winardi Kusumaatmaja, Meng Yang Ng

Stewards: (September 5, 2007) Alexander Voice, Andrew Wilkins, Ibrahim Oraiqat, Rohan Parambi

- First round reviews for this page
- Rebuttal for this page

3.1 Introduction

Incidence graphs are useful tools for chemical engineers to model various relationships in a process. They are used to systematically map an entire chain of processes and controllers to describe the effect, which each element has on the others. This helps to visualize the possible process pathways or a chain of effects.

Below is an example of an incidence graph. The circles are nodes which represent a particular device or controller, e.g. a temperature sensor. The arrows indicate the directional pathway of influence. For instance, a temperature sensor node would be connected to an adjacent node (such as a heat exchanger). This heat exchanger node would in turn be connected to other nodes representing devices or controllers. This type of diagram could be extremely useful in identifying the redundancies within a control system. A more elaborate application of incidence graphs will be discussed in the Worked Out Examples section.
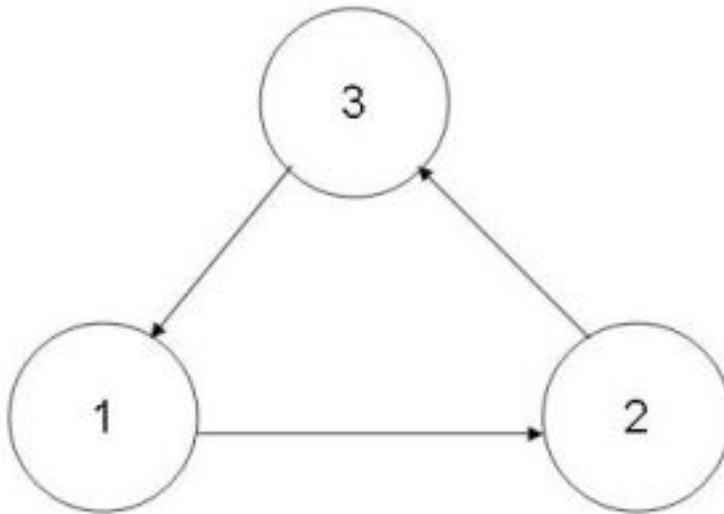
3.2 Monotonicity

Before proceeding to the next few sections, it is imperative to understand the basics of monotone systems. A monotonic function defines a function that preserves a given order. Monotonic functions are classified as monotonic increasing or monotonic decreasing. Taking the example of a function f(x), for a monotonic increasing function an increase in variable 'x' would result in an increase in f(x). On the other hand, for a monotonic decreasing function an increase in variable 'x' would result in a decrease in f(x).

Monotonic Increase : If $x \geq y$, then $f(x) \geq f(y)$
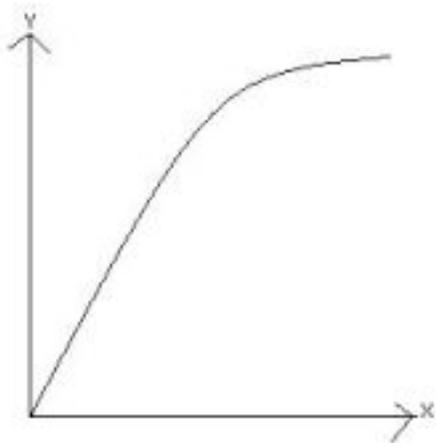
Monotonic Decrease : If $x \geq y$, then $f(x) \leq f(y)$
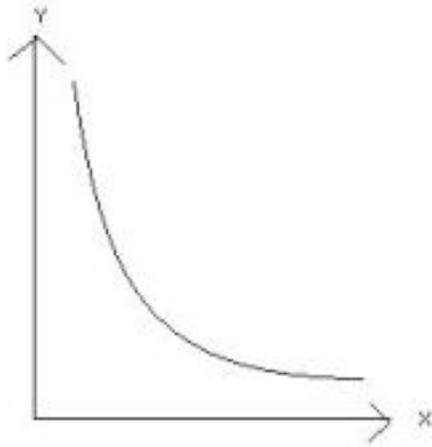
As long as a function is either always monotone increasing or monotone decreasing, it is referred to as a monotone function.

Monotonic functions are ideal systems to control because changes in one variable/device lead to known changes in other variables/devices. A non-monotonic function on the other hand, does not yield a straightforward output for a particular input. In these cases IF-THEN-ELSE statements are used to illustrate the controls for the nodes. For instances, IF node 1 is activated, THEN node 2 is repressed and so on.

An example of a possible non-monotonic graph
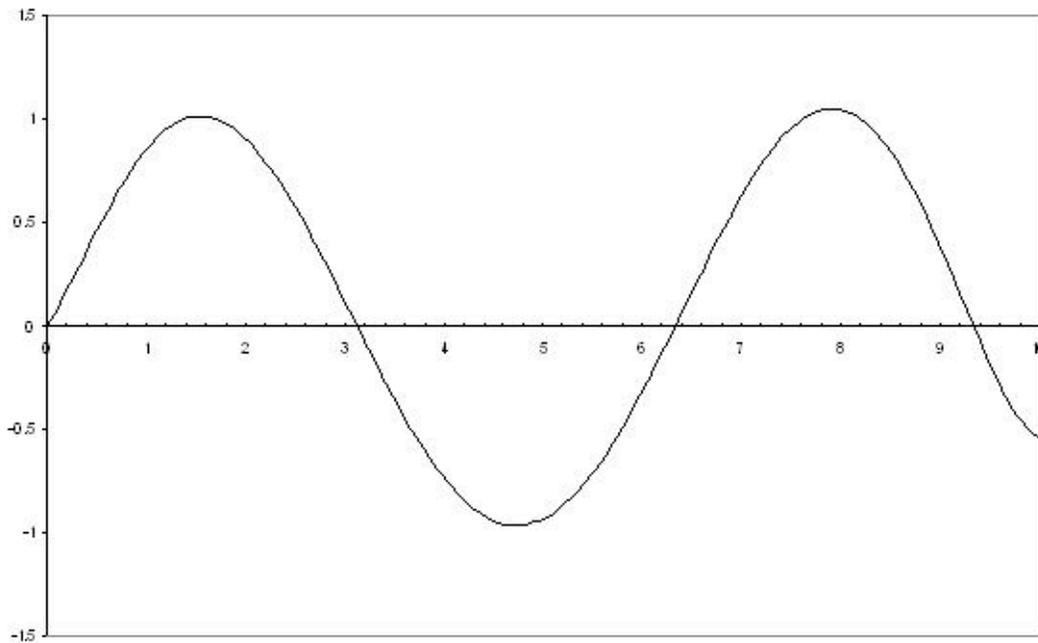
3.3 Incidence Graphs

Incidence graphs (also "Levi" or "Causation" graphs) consist of monotonic connections between nodes, whereby each connection indicates a particular effect of one factor on another. The term "monotonic" means that a change in the initiating device influences only one aspect of the target device. This simplification facilitates describing how control systems work through the use of incidence diagrams.

In reality, not all interactions are monotonic. Some control devices can simultaneously affect several aspects of other devices or variables. This may result in a non-monotonic relationship, in which one factor increases another over one region, but diminishes it over another (imagine a parabola, which is a non-monotonic curve). Temperature often has such a relationship with the rate of a reaction: for an exothermic reversible reaction, increasing the temperature will increase yield up to a certain point; however, further increase will diminish the yield.

Incidence graphs can only be used to represent monotonic relationships. Therefore, over simplification of a system can cause problems in constructing an incidence graph. Similarly a non-monotonic relationship can be resolved by breaking down a factor into separate components, each of which has only monotonic relationships with other factors.

When two factors are connected by multiple signal paths, incoherence between the results of different paths leads to *inconsistency*. In other words, the outcome cannot be predicted because the incidence graph does not communicate quantitative information. For instance, in the reaction rate example, reaction rate could be separated into the variables which affect it: concentrations, rate constant, and equilibrium constant. An incidence graph can then be constructed to demonstrate that temperature and reaction rate have an *inconsistent* relationship.

A diagram is consistent (to be discussed later in detail) if all signal paths between two nodes, for every pair of nodes, are in agreement with one another (i.e. they have the same result).

3.3.1 Traversing

Every incidence graph establishes a particular way to depict the direct effects of one device on its neighbors. Typically, a line will connect a node to its neighbors--these lines are terminated by either a bar or an arrowhead.

An arrowhead at the end indicates that the initiator (from which the line is coming) is either increasing the value of the target (where the arrow is pointing) or simply activating it (turning it on). The context of the connection will determine whether something is being increased or turned on. For example, it would be illogical to assume that a drop in temperature would cause a temperature controller to *deactivate* a pressure controller. Instead, it would be more prudent to assume that the temperature controller would *adjust* the pressure sensor.
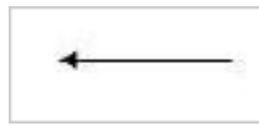
arrowhead bar

A perpendicular bar indicates that the initiating device is either decreasing the value of the target device or inhibiting it (turn off). Note that these indications are not necessarily strictly followed. Once again, context of the situation determines which interpretation should be used.

perpendicular bar

Graphs consisted of the arrowhead bar and the perpendicular bar allow you to see the result of an increase/decrease in one aspect of the initial device on a designated device in the system. Everything done after the initial device comes as a result of changing the

45

initial device. In addition, these graphs provide the possibility of simplifying complicated relationships between devices into one conceptual unit. The relationship between the initiating device and target nodes can be described by the following set --> {initiator,target1,target2,etc}. It is read as "the increase or decrease in the initiating device affects the target nodes in the following manner... (+ for increase/activate and - for decrease/inhibit). The initiator can be any device or variable in the system as long as the arrows support the designation. In other words, you can show the effects of any device on any other device, as long as the pathway is supported by the diagram. There are often multiple routes that a signal can travel.

2 helpful tips to keep in mind as you interpret an incidence graph:

1. Always travel from **tail to head** (or source to target) when traversing the graph from the initiating device to the designated device. Keep in mind that it is possible to have arrowhead and/or perpendicular bars pointing in both directions in between two devices.

2. Arrowhead bars don't always mean increase as perpendicular bars don't always mean decrease. Arrowhead bars indicate that the effect of performing an action to the device at the tail of the bar would yield the **same** effect to the device at the head of the bar. Penpendicular bars indicate that the effect of performing an action to the device at the tail of the bar would yield the **reverse** effect to the device at the head of the bar.

The following is a simple example of an incidence graph. We will refer to the circles below as "nodes" from this point onward. Nodes are the selected devices or variables in a control system.
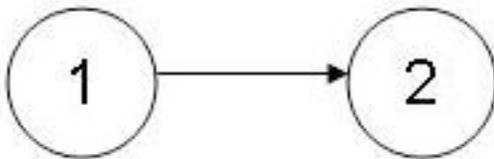
This incidence graph indicates that:

a. A change in an aspect of 1 will change some aspect of device 2 in the same direction -- otherwise written as **{+,+}**

The ends indicate the direction of influence as well as the type of influence.

46
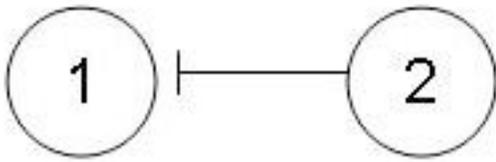
This incidence graph indicates that:

a. increase in an aspect of 2 (initiator) decreases some aspect of 1 (target)

b. indicated by {+,-}

This next example builds on the previous two...
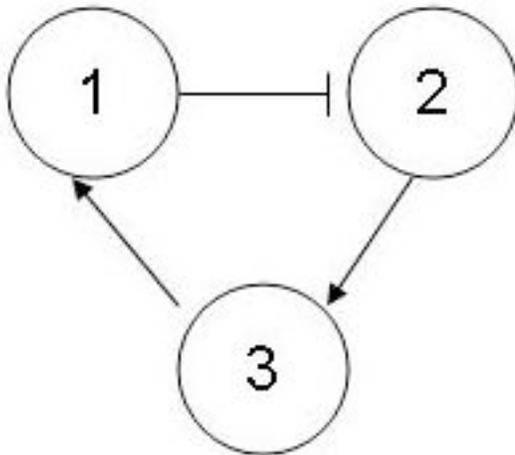
This incidence graph indicates that:

Given an arbitrary 3,1,2 order, an increase in 3 (initiator) will make 1 (target) go up. 1's increase will then make 2 decrease. The symbolic representation for this would be {+,+,-}. One can see that this diagram shows how node 3 influences everything directly and indirectly connected to it.

Now, given a 1,2,3 order, node one becomes the initiator. The symbolic representation then changes to {+,-,-}. **NOTE: the last node is negatively affected because 2's impact on 3 is shown to increase when 2 increases. But since 1 decreases 2, then the opposite is true!**

The main objective of these diagrams is to determine the effect that a change in the starting node has on another node in the system. However, in some cases, you can follow a pathway that ends at the starting node in order to check for ambiguity. For example, the 1,2,3,1 order yields a result of {+,-,-,-}. Since increasing node 1 cannot cause a decrease in itself, it is ambiguous.

Incidence graphs for complex controls systems will usually indicate that there are several ways for a device to influence another device in the system.

Multiple paths

Remember these are done with respect to increases in a node's value. So, if a path leads you to a point where you decrease a node, then you have to do the opposite of what the diagram is indicating for all steps after that (for that route).
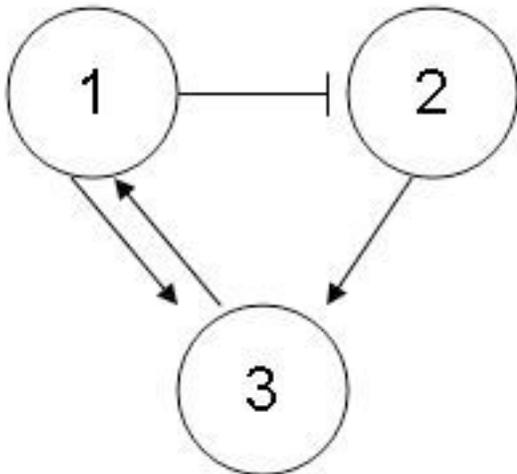
The incidence graph indicates:

1. For an order {1,2,3}: you observe {+,-,-}... remember not to count beginning node twice…

2. For an order {1,3}: you observe {+,+}

3.4 Consistent, Inconsistent, and Partially Consistent

When designing controls for systems it is always important to keep in mind the need for redundacy and feedback control. Redundancy in a system will ensure that a system is monitored by multiple controllers. So, if one controller was to fail there would be another

to ensure the same desired outcome. Consistent graphs can be used to describe redundancy. Feedback control is also very important and it is not good to have the system overreact to external stimuli. A controlled response is desired. Inconsistent graphs can be used to describe feedback control. Additionally, systems exist that both need to be redundant and feedback controlled and these can be described with partially consistent graphs.

3.4.1 Consistent Graphs

Consistent pathways happen when all pathways from the same starting node, that lead to the same target node, give the same output.

For example,

All possible pathways from node 1 to target node 4 yield the same output. These are called consistent pathways.
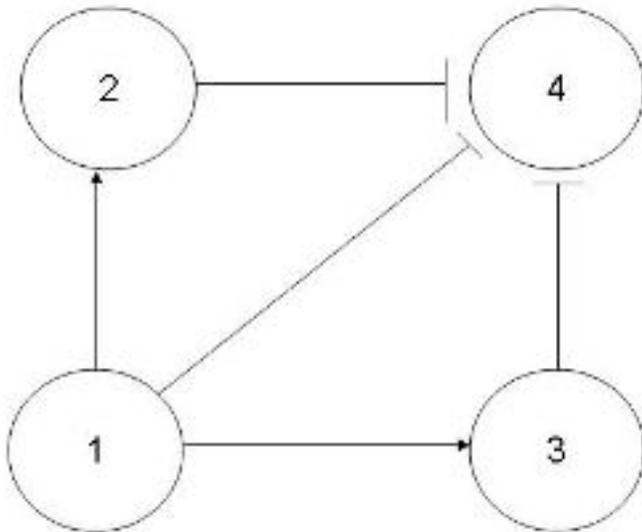
Paths(s)	Sign series
1,4	(+,-)
1,2,4	(+,+,-)
1,3,4	(+,+,-)

From the table and graph above, all pathways from node 1 to node 4 yield the same outcome. Node 4 is being decreased.

A consistent graph is made up of only consistent pathways for all possible pathways between any combination of two nodes. The graph below is an example of a consistent graph.
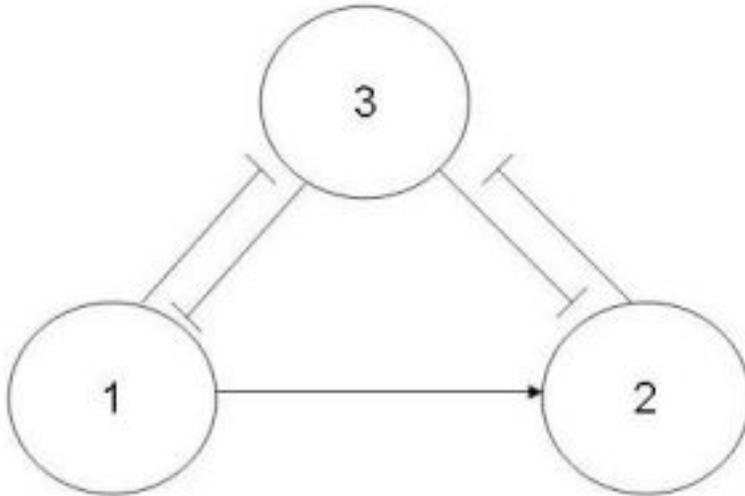
Paths(s)	Sign series
1,2,3	(+,+,-)
1,3	(+,-)
3,1,2	(+,-,-)
3,2	(+,-)
1,3,1	(+,-,+)
1,2,3,1	(+,+,-,+)
3,2,3	(+,-,+)
3,1,3	(+,-,+)
3,1,2,3	(+,-,-,+)

All pathways leading from node 1 to node 3 -

1,2,3 : (+,+,-)
1,3 : (+,-)

Since the two pathways cause the same change to node 3, this sub-pathway is consistent.
All pathways leading from node 3 and to node 2 -

3,1,2 : (+,-,-)
3,2 : (+,-)

Since the two pathways cause same change to node 2, this sub-pathway is consistent.

50

All pathways leading from node 1 and back to node 1 -

1,3,1 : (+,-,+)
1,2,3,1: (+,+,-,+)

Since the two pathways result in the same change to node 1, this sub-pathway is consistent.

All pathways leading from node 3 and back to node 3 -

3,2,3 : (+,-,+)
3,1,3 : (+,-,+)
3,1,2,3: (+,-,-,+)

Since the three pathways give the same effects to node 3, this sub-pathway is consistent.

Because all sub-pathways of this incidence graph are consistent, the entire process is consistent.

The table shown below gives examples of many different incidence diagrams and shows whether or not they are consistent.

Number	Graph	Consistent?	Paths(s)	Sign series	Interpretation
1.		Yes	{1}	+	1 influences itself
2.		Yes	{1,2}	+,+	1+ influences 2
3.		Yes	{1,2}	+,-	1 - influences 2
4.		No	{1,2,1}	+,+,-	influence of 1 on 1 via 2 is ambiguous
5.		Yes	{1,2,1}	+,+,+	1 + influences 2, which + influences 1
6.		Yes	{1,2,1}	+,-,+	1 - influences 2, which - influences 1
7.		Yes	{3,1,2} {3,2}	+,-,- +,-	both paths from 3 to 2 show that 3 - influences 2
8.		No	{3,1,2} {3,2}	+,-,+ +,-	influence of 3 on 2 is ambiguous
9.		No	{3,1,2} {3,2} {1,3,1} {1,2,1} {3,2,1} {3,1}	+,-,- +,- +,-,+ +,+,- +,-,+ +,-	influence of 1 on 1 via 2 is ambiguous and influence of 3 on 1 is ambiguous
10.		Yes	{1,4,5,3} {1,3} {6,4,5,3,2} {6,5,2} (and many many more possible paths)	+,-,-,+ +,+ +,+,+,-,+ +,+,+	Each variable 1,2,3,4,5,6 net has a simple relationship, monotonic between any other variable in the set.

Another type of consistent graph is one in which the valves of the system are not dependent on one another for feedback. In other words, there is no proportion of the output signal from a particular valve of a system is passed to the input of another valve. An example of this would be to consider the following process diagram:
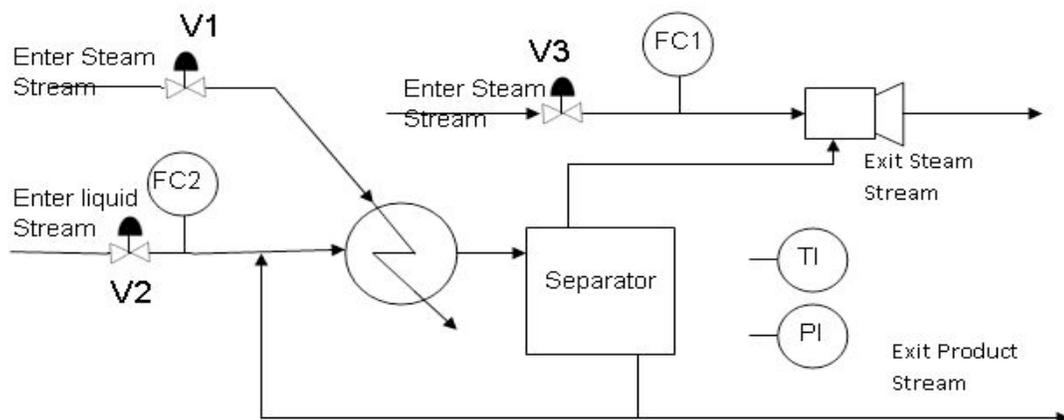
This process is of a distillation column in which the bottom columns collect high intensity liquids and the liquid is removed from a series of evaporators. The evaporator is a single-effect evaporator equipped with a simple heat exchanger, a separation vessel and a steam jet ejector to supply the vacuum. The product is then sent to an oven where it is dried and sent to the manufacturer. It is the chemical engineer's task to create a consistent incidence graph in which the valves in the diagram do not depend on each other's feedback. An example of this is below:
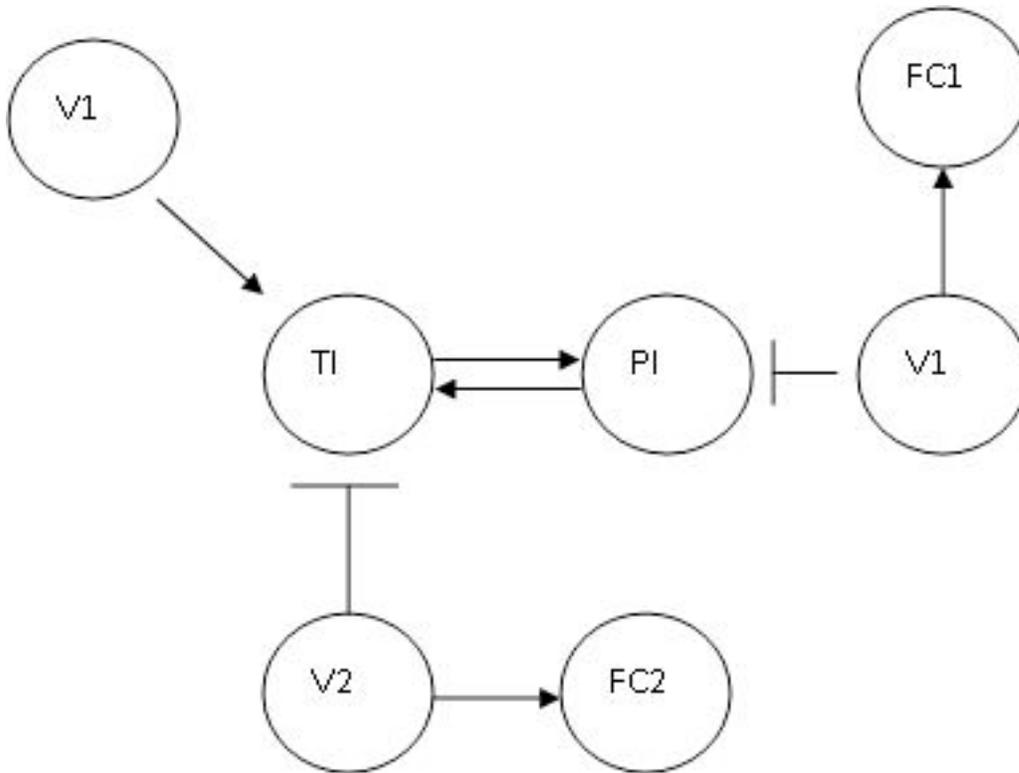
Paths(s)	Sign series
V1, TI, PI, TI	(+,+,+,+)
V2, TI, PI, TI	(+,-,-,-)
V3, PI, TI, PI	(+,-,-,-)

Based on the table above, the incidence graph is consistent because the valves are not dependent on one another for feedback.

3.4.2 Inconsistent Graphs

Inconsistent pathways happen when pathways from the same starting node, that lead to the same target node, give different output.

For example,

Pathways from node 1 to target node 4 do not yield the same output. These are called inconsistent pathways.
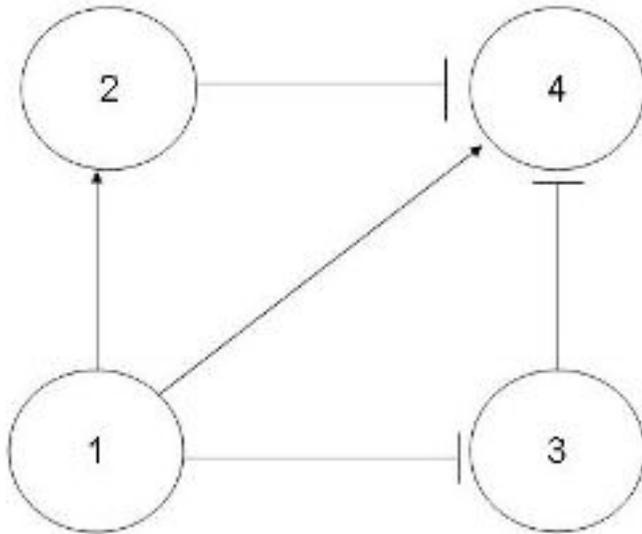
Paths(s)	Sign series
1,4	(+,+)
1,2,4	(+,+,-)
1,3,4	(+,-,+)

From the table and graph above, all pathways from node 1 to node 4 do not yield the same outcome. Node 4 is being decreased while simultaneously being increased.

An inconsistent graph is made up of only inconsistent pathways for all possible pathways between any combinations of two nodes. The graph below is an example of an inconsistent graph.
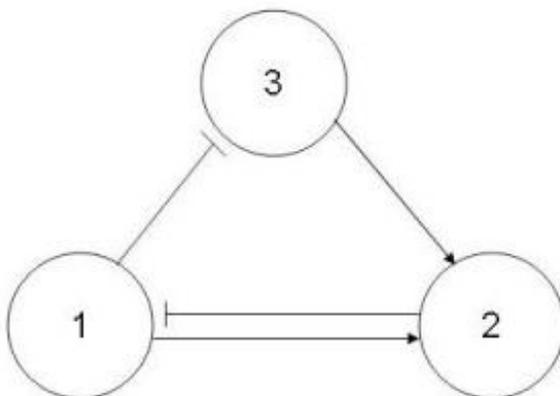
Paths(s)	Sign series
1,2	(+,+)
1,3,2	(+,-,-)
1,2,1	(+,+,-)
1,3,2,1	(+,-,-,+)

All pathways leading from node 1 to node 2 -

1,2 : (+,+)
1,3,2 : (+,-,-)

Since the two pathways cause different changes to node 2, this sub-pathway is inconsistent.

All pathways leading from node 1 and back to node 1 -

1,2,1 : (+,+,-)
1,3,2,1: (+,-,-,+)

Since the two pathways cause different changes to node 1, this sub-pathway is inconsistent.

Because all sub-pathways in this incidence graph are inconsistent, the incidence graph is inconsistent.

3.4.3 Partially Consistent Graphs

Partially consistent graphs are made up of both consistent and inconsistent pathways. If you increase node 1, all paths leading to node 4 cause a decrease in the node which means it is a consistent pathway. If node 1 is increased, one path leading to node 2 causes an increase and another path causes a decrease in the node. This is an inconsistent pathway. Since this model contains both at least one consistent and one inconsistent pathway, it is a partially consistent model.
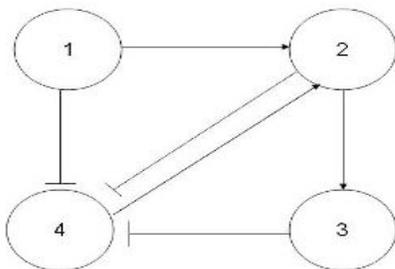
Paths(s)	Sign series
1,2,3,4	(+,+,+,-)
1,2,4	(+,+,-)
1,4	(+,-)
1,2	(+,+)
1,4,2	(+,-,-)

All pathways leading from node 1 to node 4 -

1,2,3,4: (+,+,+,-)
1,2,4 : (+,+,-)
1,4 : (+,-)

Since the three pathways cause the same change to node 4, this sub-pathway is consistent.

All pathways leading from node 1 and to node 2 -

1,2 : (+,+)
1,4,2 : (+,-,-)

Since the two pathways cause different changes to node 2, this sub-pathway is inconsistent.

In this graph there are sub-pathways that are consistent and inconsistent, therefore the incidence graph is partially consistent.

3.4.4 Summary

Just because a process is inconsistent does not mean that the desired results cannot be achieved. All it means is that you cannot consistently achieve the desired results. Take the example of the flow of a final product. In a consistent process, no matter which path is taken, the flow of the final product will always respond the same, whether desirable or not. In an inconsistent process, changing the path will not always cause the flow of the final product to change the same way, but rather depends on the path taken.

In all, inconsistent processes are not bad. Often, processes will be very complex, containing many variables, and inconsistent processes will be unavoidable. Care must just be taken with complex, inconsistent processes to achieve desirable results.

Feed stream

CSTR (jacketed)

Heating coil

Cooling water feed

Cooling water exit

Product stream
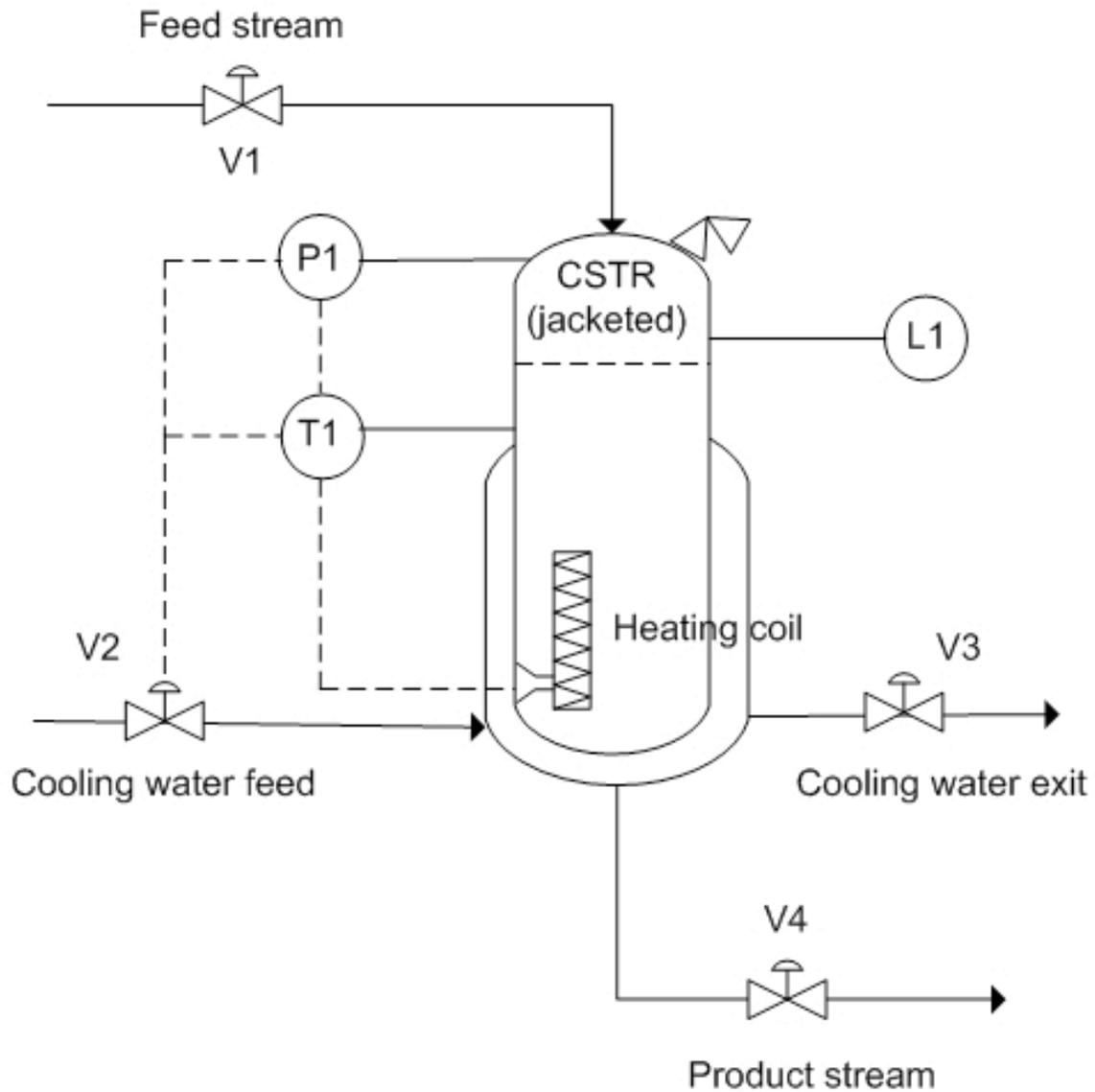
You are a process engineer and you have been assigned to devise a process control for a new reactor. The optimum operating condition for the reaction inside the reactor is 220 °C ± 10°C. The reaction is exothermic and often goes way past the optimum temperature. It is imperative that the reaction should not go above 250°C. The reactor has a built in heater (H1) to keep temperature at its optimum temperature and a cooling water feed to cool the reactor in case temperature goes past 250°C. The reactor also has temperature sensor installed to monitor the temperature. If temperature goes above 250°C, the desired mechanism is to increase the flow of water in the cooling jacket surrounding the reactor and to turn off the heating element inside the reactor until it reaches its optimum temperature. As a control redundancy mechanism, a pressure sensor is also installed inside the reactor. The pressure control will send signal to the cold water feed if pressure reaches 1.2 atm which is the critical pressure. Draw a causation graph governing the

situation when the temperature gets too hot inside the reactor. Base your graph on the following variables:

Note: (There may be more than one correct solution to the assigned problem)

a)temperature

b)temperature sensor (T1)

c)cooling water feed (V3)

d)heating element (H1)

e)pressure

f)pressure sensor (P1)
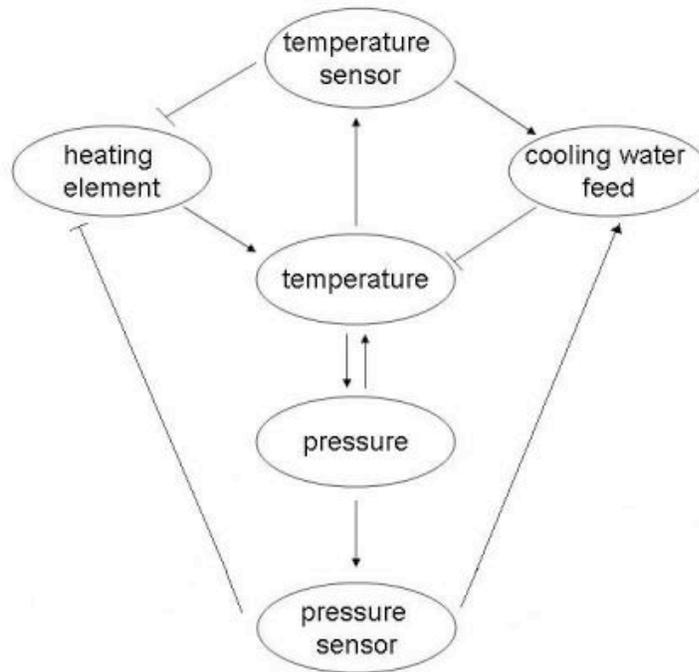
High temperature will 'activate' the temperature sensor, telling it that the temperature has exceeded the optimal temperature and an action need to be done to bring it back to the desired temperature. The temperature sensor would then open the valve for the cooling water feed which would cause the water to flow through the jacket. This results in the temperature inside the reactor to decrease. The reverse happens for the heating element. An activation of the temperature sensor would translate to an inhibition of the heating

element in which the heating element would stop increasing the temperature of the reactor. Temperature and pressure are directly related by ideal gas law and thus affects each other in both ways. Thus, for the control redundancy mechanism, an increase in pressure would activate the pressure sensor which would then activate the cooling water feed to cool the temperature for the reactor.

Logic control:

- IF T1 =< 220 °C, THEN H1 is on, ELSE H1 is off
- IF T1 >= 250 °C, THEN V2 and V3 is open, ELSE V2 and V3 is close
- IF P1 >= 1.2 atm, THEN V2 and V3 is open, ELSE V2 and V3 is close

3.6 Worked out Example 2

After completing the assignment of devising the temperature control, your boss asked you to devise a mechanism to control the flow of raw materials and products. Draw a causation graph showing negative feedback relations between L1, V1 and V4 when the level inside the reactor is increasing (this is an example of an inconsistent pathway). The normal operating conditions are shown below:

V1 is 50 gal/min, V1 is usually open

V4 is 50 gal/min, V4 is usually open

L1 is 20 meters, L1 should not go above 30 meters and should not be below 10 meters
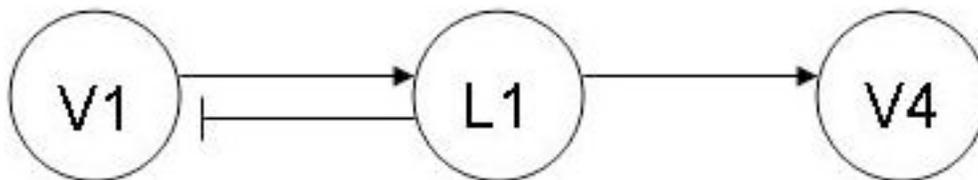
The level sensor monitors the fluid height inside the jacketed CSTR. Assuming the level inside the reactor is increasing steadily, the level sensor will send a negative feedback signal to V1 to decrease the input flow.

Logic control:

IF L1 > 30 meters, THEN V1 is closed

IF L1 < 10 meters, THEN V4 is closed

Determine whether the following causation graph is consistent, inconsistent or partially consistent. Include your reasoning.
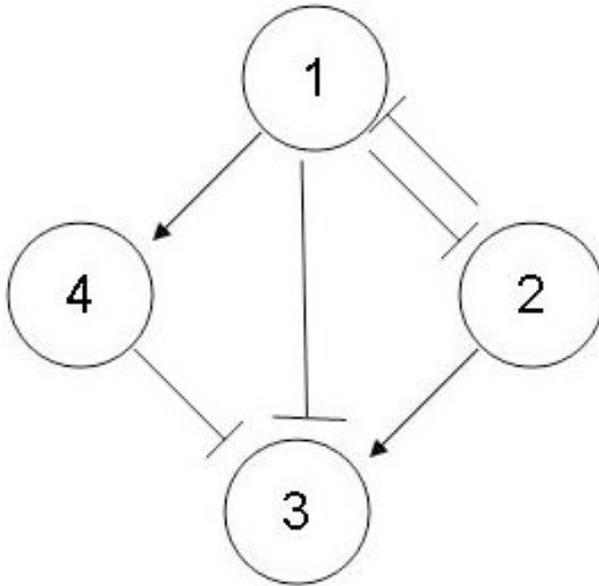
Solution:

All relationships are consistent. Each variable 1,2,3,4 has a simple relationship which is monotonic among the other variables in the set.

Paths(s)	Sign series
{1,2}	(+,-)
{1,2,3}	(+,-,-)
{1,3}	(+,-)
{1,4,3}	(+,+,-)
{1,4}	(+,+)
{2,1}	(+,-)
{2,3}	(+,+)
{2,1,3}	(+,-,+)
{2,1,4}	(+,-,-)
{4,3}	(+,-)

3.8 Sage's Corner

A Brief Narration for Better Understanding of the Incidence Graphs
http://video.google.com/googleplayer.swf?docId=857592852338575559

slides for this talk

3.9 References

- Woolf, Peter (2006). A Mini-Guide to Consistent Graphs.
- Woolf, Peter (2006). Terms Review.
- Angeli, David & Sontag, Eduardo D. (2003). Monotone Control Systems.

Section 4. Excel modeling: logical models, optimization with solver for nonlinear regression, sampling random numbers

Note: Video lecture available for this section!

Authors: (12 September 2006 / Date Revised: 19 September 2006) Curt Longcore, Ben Van Kuiken, Jeffrey Carey, Angela Yeung

Stewards: (13 September 2007) So Hyun Ahn, Kyle Goszyk, Michael Peterson, Samuel Seo

- First round reviews for this page
- Rebuttal for this page

4.1 Introduction

Microsoft Excel program is one of the most popular and useful computer programs for a wide variety of numerical applications. Excel features several different functions, interfaces and graphing tools which can be applied in many fields. The following functions might be especially useful for logical programming in Excel:

- Logical functions (ex: IF, OR and AND): These functions can be used in control analysis, particularly in safety regulation (for example, if the temperature exceeds X degrees, shut down the reactor and send cooling water to the jacket)

- Solver: This function can be used to maximize, minimize or try to obtain an input value in a cell by varying referenced cells.

- Random Number Generator. This can be used as a tool for probability or randomly selecting data points for analysis from a large pool of points.

While it might be less accurate than other modeling programs, the Excel model provides good estimates, and facilitates observation and understanding of process behavior while requiring minimum computer knowledge. The following sections will provide you with a better understanding of these common Excel functions, followed by examples in engineering applications.

4.2 Logical Methods

A logical test is any expression that can be evaluated as TRUE or FALSE with specified input conditions. Commonly used logical tests are IF, OR and AND statements. All three statements can be used in conjunction with each other for more complex and complete modeling.

Note: all functions and formulas in Excel require an equal (=) sign in front of the syntax.

4.2.1 Excel's IF statement

The IF function in Excel tests values and formulas in the target cells and returns one specified value "if" the condition input evaluates to be TRUE and another value if it evaluates to be FALSE. The IF statement in Excel is called out with the following syntax:

- IF(logical_test,value_if_true,value_if_false)

Logical_test: any expression or correlation of values that can be evaluated as TRUE or FALSE.

Value_if_true: the value or expression that is returned IF the condition in the logical test returns TRUE.

Value_if_false: the value or expression that is returned IF the condition in the logical test returns FALSE.

The **value_if_true** and the **value_if_false** part in the syntax may vary and can be numerical values, expressions, models, and formulas. If either of the values is left blank, the test will return 0 (zero) correspondingly.

Sample coding:

- IF(1+1=2, "Correct!", "Incorrect!")

For the logical test part, 1+1 is equal to 2, and the iteration will return TRUE. The **value_if_true** is called for the TRUE response, and "Correct!" will be output in the corresponding cell.

It is possible to nest multiple IF functions (called **Nested IF**) within one Excel formula. Such a formula is useful in returning multiple answers, compared to a a single IF function which can return only two possible answers.

A Nested IF statement has the following syntax:

- IF(logical_test1, value_if_true1, IF(logical_test2, value_if_true2, value_if_false))

The following example shows how a Nested IF function is used to make a spreadsheet that assigns a letter grade to different scores.

Suppose that grade distribution is:

@ A= Greater than or equal to 80

@ B= Greater than or equal to 60, but less than 80

@ C= Less than 60

This logic function can be written in several different ways. If the score is tested to be the letter grade C first, then checked to see if it is a B or an A, the syntax could be written as follows:

- IF(A5<60,"C",IF(A5<80,"B","A"))

Note that the Nested IF statement can be placed in either the **value_if_false** OR the **value_if_true** positions. In the example above, it was placed in the **value_if_false** position because of how the conditional was written.

Additional letter grades could be added by including more nested IF functions inside the existing function. One formula can include up to 7 nested IFs, which means you can return up to 8 different results.

4.2.2 Excel's OR statement
The OR statement in Excel is similar to the IF statement. It returns TRUE if **any** argument or expression of values is TRUE, and it returns FALSE if **all** arguments or expression of values are FALSE. The syntax for OR is:

- OR(logical1,logical2,...)

logical1: there can be up to 30 arguments for the logical expression to test and return either TRUE or FALSE. If all of the arguments are false, the statement returns FALSE for the entire expression.

Sample coding:

- OR(1+1=2, 2+2=5)

In the logical1 part, 1+1 is equal to 2 and will return TRUE. In the logical2 part, 2+2 is not equal to 5 and will return FALSE. However, since one of the test returns TRUE (logical1), the function returns TRUE even if logical2 is FALSE. **TRUE** will be output in the corresponding function cell.

4.2.3 Excel's AND statement

The AND statement returns TRUE if **all** its arguments or values are TRUE and returns FALSE if one or more argument or value is FALSE. The AND statement syntax is:

- AND(logical1,logical2,...)

logical1: similar to OR statement with up to 30 arguments to test that could return either TRUE or FALSE.

Sample coding:

- AND(1+1=2, 2+2=5)

Logical1 will return TRUE for the correct calculation while logical2 will return FALSE for the incorrect calculation. However, since one of the expressions is FALSE, the AND logical test will return FALSE even if one of the tests is true.

It is common and useful to understand how to implement different logical functions into each other. A good example would be:

- IF(AND(2+2=4,3+3=5),"Right Calculation","Wrong Calculation")

During iteration, the AND statement should return FALSE since one of the expression/calculation is incorrect. The FALSE return will output "Wrong Calculation" for the IF statement iteration.

4.2.4 Other Useful Excel Functions

Excel has many useful functions that can help any engineer who is using Excel to model or control a system. An engineer may want to know the average flow rate, number of times the liquid in a tank reaches a certain level, or maybe the number of time steps that a valve is completely open or completely closed. All of these situations and many more can be addressed by many of the simple built-in functions of Excel. Below are some of these functions and a brief definition of what they compute and how to use them. They are followed by an image of a sample excel file with an example of each function in use. To see a list of all the built-in functions Excel has to offer, click shift-enter on the cell.

AVERAGE(number1,number2,...) - Returns the average or arithmetic mean of the arguments. The arguments can be numbers, arrays, or references that contain numbers.

CEILING(number1, significance) - Rounds a number up to the nearest integer or multiple of significance.

CHIDIST(x, degrees_freedom) - Returns the probability of the chi-squared distribution.

COUNT(value1,value2,...) - Counts the number of cells that contains numbers.

COUNTIF(range, criteria) - Counts the number of cells within the range that meet the given criteria.

COUNTIFS(criteria_range1,criteria1,criteria_range2,criteria2,...) - Counts the number of cells within multiple ranges that meet multiple criteria.

FLOOR(number1,significance) - Rounds a number down toward zero, to the nearest multiple of significance.

FREQUENCY(data_array,bins_array) - Calculates how often a value occurs within a range of values (bins). The function will return a vertical array of numbers having one more element than the bins_array. Once the formula is entered into the target cell, highlight the target cell making a vertical array one cell larger than bins_array. Once selected press F2 to open the formula of the target cell and then press CTRL+SHIFT+ENTER to execute your function.

LARGE(array,k) - Returns the k-th largest value in a data set. For example, the 2nd largest number.

MAX(number1,number2,...) - Returns the largest number in a set of values.

MEDIAN(number1,number2,...) - Returns the median of a set of values.

MIN(number1,number2,...) - Returns the minimum value in a set of values.

MODE(number1,number2,...) - Returns the most frequently occurring value in an array or range of values.

PERCENTILE(array,k) - Returns the k-th percentile of values in a range

ROUND(number1, num_digits) - Rounds a number to the specified number of digits.

SMALL(array,k) - Returns the k-th smallest value in a data set. For example, the 7th smallest number.

STDEV(number1,number2,...) - Returns the standard deviation of a given set of values ignoring text and logical values.

SUM(number1,number2,...) - Adds all of the values in a range of cells.

SUMIF(range,criteria,range,...) - Adds values of cells in a range meeting criteria.

VAR(number1,number2,...) - Estimates the variance of a given set of values.

	A	B	C	D	E
1					
2				Range	Bins
3	AVERAGE(D3:D26)	4.125		1	2
4	COUNT(D3:D26)	24		1	5
5	COUNTIF(D3:D26,3)	3		1	9
6	FREQUENCY(D3:D26,E3:E6)	7		1	
7		10		2	
8		7		2	
9		0		2	
10	LARGE(D3:D26,3)	7		3	
11	MAX(D3:D26)	8		3	
12	MEDIAN(D3:D26)	4		3	
13	MIN(D3:D26)	1		4	
14	MODE(D3:D26)	1		4	
15	SMALL(D3:D26,6)	2		4	
16	STDEV(D3:D26)	2.251810865		5	
17	SUM(D3:D26)	99		5	
18	VAR(D3:D26)	5.070652174		5	
19				5	
20				6	
21				6	
22				6	
23				7	
24				7	
25				8	
26				8	
27					
28					

4.2.5 Nesting

Though not complex or difficult, placing logical statements within one another can be a powerful practice. This is referred to as **nesting**.

An example demonstrating the use of a simple **IF** statement **nested** within another **IF** statement is shown below:

If one were to ask three individuals whether or not they agree that energy has a velocity, these individual will hypothetically respond in one of three ways: "Yes","No", or (other). In this case, if the individual were to respond "Yes", then the questioner would respond "I agree!". If the individual were to respond "No", the questioner would bitterly respond with a "Bah". And if the individual were to respond with a nonsensical answer, then the questioner would respond "what .." in confusion.

The important point to get out of this example is that **nested** statements gives us the potential to process more information in one cell (or line of programming code).

	A	B	C	D	E	F	G	H	I	J
1										
2					=IF(F6="Yes","I agree!", IF(F6="No","Bah", "what? .."))					
3										
4										
5	Me - "I don't think energy has a velocity."					You - "Yes" or "No?"		How I feel about your response…		
6	Do you agree?				-->	Yes	-->	I agree!		
7	Do you agree?				-->	No	-->	Bah		
8	Do you agree?				-->	5	-->	what? ..		

4.3 Excel's Data Analysis Tools

Excel has several useful built in add-ins that can make data analysis much easier. Of these we will cover the solver tool, as well as the ANOVA analysis.

4.3.1 Solver Tool

The **solver** in Excel is a built-in tool with multiple capabilities. It could be used to optimize (find the maximum or the minimum value) a specific cell (the target cell) by varying the values in other cells (the adjustable cells). It may also be used to solve a system of non-linear equations. Solver formulas embedded in the spreadsheet tie the value in the target cell to the values in the adjustable cells. The user can also specify numerical range constraints for the cells involved.

Optimization Model

In optimization, one seeks to maximize or minimize the value of a real function. Just as a student in a calculus class might use optimization techniques to find a local maxima or minima of a complex function, a CEO might use similar techniques to find how much product their company should manufacture to maximize profits. There are numerous optimization techniques, many specialized for specific types problems. For the most part, MS Excel uses gradient-based techniques.

The solver tool is located in the **Tools** menu. If it does not appear there, it must be added in by selecting **Add-ins** and selecting the appropriate check box.

The solver window contains the following fields:

• **Set Target Cell:** This is the cell containing the value you want to optimize.

- **Equal To:** Choose whether to maximize the value, minimize the value, or set it to a specific number.
- **By Changing Cells:** Here you specify all the cells that can be varied when optimizing the target cell value.
- **Subject to the Constraints:** Constraints are optional, but can be added by clicking **Add** and defining each constraint.
- **List of Available Constraints**
  - **<= (Less than or equal to)** Stipulates that a selected cell must be less than or equal to a certain value.
  - **= (Equal to)** Stipulates that a selected cell must be equal to a certain value.
  - **>= (Greater than or equal to)** Stipulates that a selected cell must be greater than or equal to a certain value.
  - **int (Integer)** Stipulates that a selected cell must be an integer
  - **bin (Binary)** Stipulates that a selected cell must be equal to 1 or 0.

The window where these constraints can be entered is shown below.
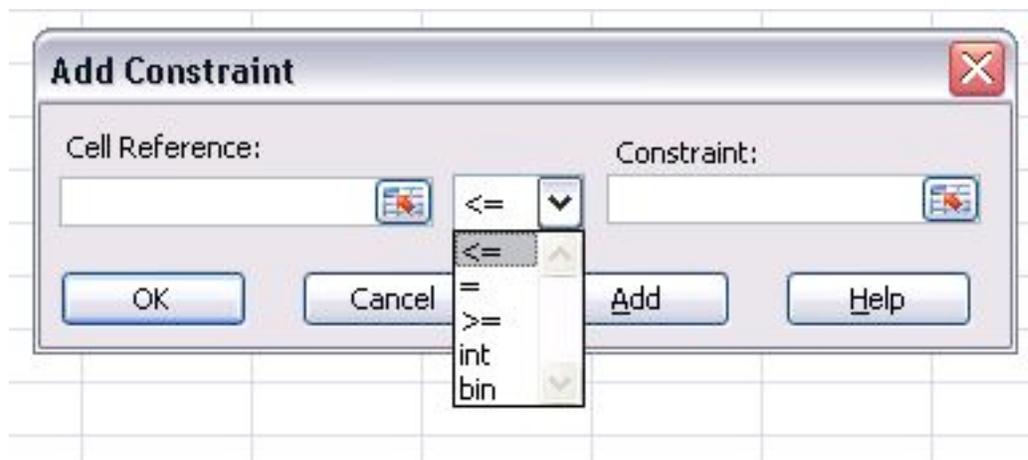
It is important to make sure that constraints do not violate each other otherwise solver will not work without reporting an error. For example, if a cell must be great than 2 and less than 1 Solver will not change anything.

Once all the fields in the Solver window are completed appropriately, click **Solve**. If solver can find a solution a window will pop up telling you so, and the values will appear in the target and adjustable cells. If solver cannot find a solution it will still display the values it found but will state that it could not find a feasible solution. Be aware that Excel solver is not a powerful enough tool to find a solution to nonlinear equations, where programs such as Matlab might be a better choice.

* Guided Example: Suppose you have an Excel spreadsheet set up as shown below, with cell C6 depending on cell B6 according to the quadratic relationship shown, and you want to minimize the value in C6 by varying the value in B6.
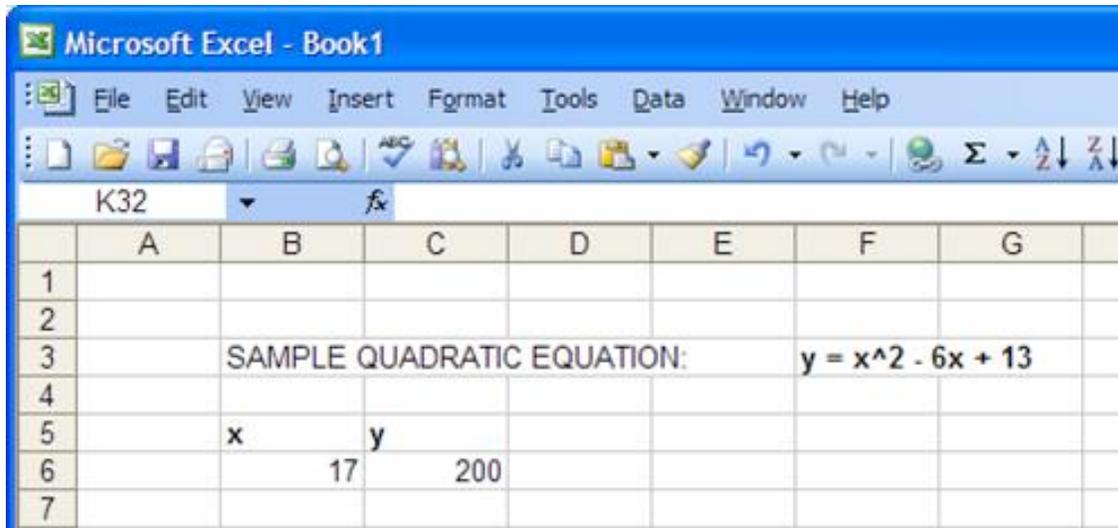
1) Open up the solver window.

2) Input C6 in the **set target cell** field, check the **min** button, and input B6 in the **by changing cells** field:
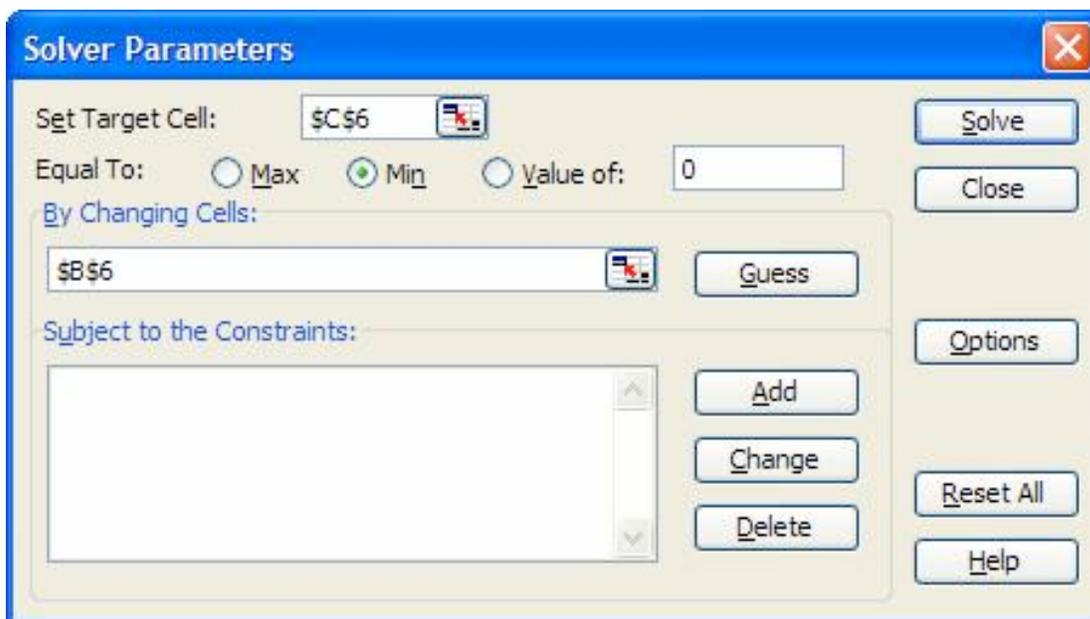
3) Click **solve** to find the solution.

You can place further constraints or further specify how you want your task solved under "Options." Some of the most commonly used specifications include:
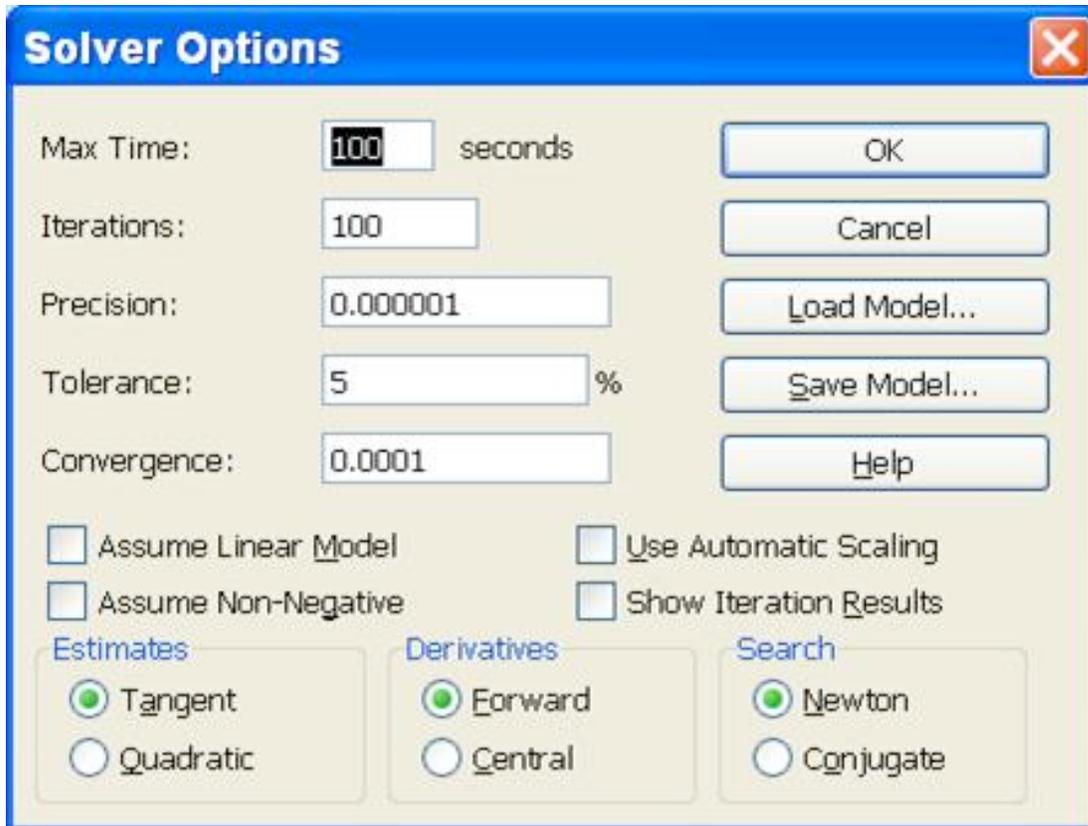
* Max Time: This is the maximum amount of time excel will spend trying to find a converging value before giving up.

* Iterations: This is the number of iterations excel will perform to converge to a value.

* Precision: This is related to how close excel has to get before a value is "acceptable." In general, making this value several orders of magnitude smaller will yield a more accurate result, but will take more time to do so, and may have a harder time converging.

* Tolerance: This is also related to how accurate a solution is, this is related to the percent error. Making this smaller should also yield a more accurate result, but with more difficulty finding a solution.

* Convergence: Controls the amount of relative change that occurs during the last five iterations that Solver performs. The smaller the number of the convergence the less change that will occur.

Box Options Click on the box to select the following options:

* Assume Linear Model: This will speed up the time to find a solution because it will make Solver assume that the model is linear.

* Assume Non-Negative: Solver will make the lower limit of adjust zero for all the cells that are programmed to be adjusted in the model that are not limited by a chosen constraint.

* Use Automatic Scaling: Used for large differences in magnitude to scale the inputs and outputs.

* Show Iteration Results: Solver will pause between each iteration to show the results of that iteration in the model.

Estimates Specifies to Solver on how to use initial estimates for iterations

* Tangent: Uses linear extrapolation by utilizing a tanget vector.

* Quadratic: Uses quadratic extrapolation. Best to use for nonlinear models.

Derivatives Specifies what type of differencing is used for partial derivatives for both the objective and constraints.

* Forward: Constrained values change slowly between iterations.

Nonlinear Regression

Excel's solver function can also be used to find a solution for two-variable non-linear regression. The description of the data by a function is carried out by the process of iterative nonlinear regression. This process minimizes the value of the squared sum of the difference between the actual data and the predicted value (residual).

There are three basic requirements for using solver including:

1) Raw data

2) A model equation to find predicted values

3) Initial guesses for varying values

Before using the solver function, the raw data should be organized into separate columns for the independent variable (first column) and dependent variable (second column). A third column will show the predicted value that is obtained by using the model equation mentioned earlier, and must reference various single cells that have initial values that Excel will use as a starting point for the regression. A fourth column should be created with the square of the difference between the predicted value and the actual experimental value (raw data); this is also known as the square of the residual. A cell at the bottom of this fourth column should include the sum of all the squared difference values above it. Two more cells anywhere else in the worksheet should be designated as the initial guesses for the values to be varied to find the solution.

It is important to choose a reasonable value for the initial guess. This will not increase the chance that Excel arrives at a global optimization rather than a local optimization, but also reduces the time it takes for Excel to solve the system of non-linear equations.

Now, solver can be applied to the spreadsheet to optimize the regression. Refer to section 3.1 for how to access and use solver. The bottom most cell of the fourth column that has the sum of the squares of the residuals will be the target cell. Since the residuals are essentially the error, the sum of the residuals squared should be as small as possible, so the min circle should be selected. The initial guesses should be entered into the "by changing cells". Constraints can be entered if the values of the cells are limited to a certain range. Solver will produce the values of the varying parameters that yield the best fit to the nonlinear data.

- Guided Example: A screen shot of a simple population problem is shown below. The yellow highlighted cells are the initial guesses for the variables. In this case, the initial population (P0, cell E2) is entered as a constraint.
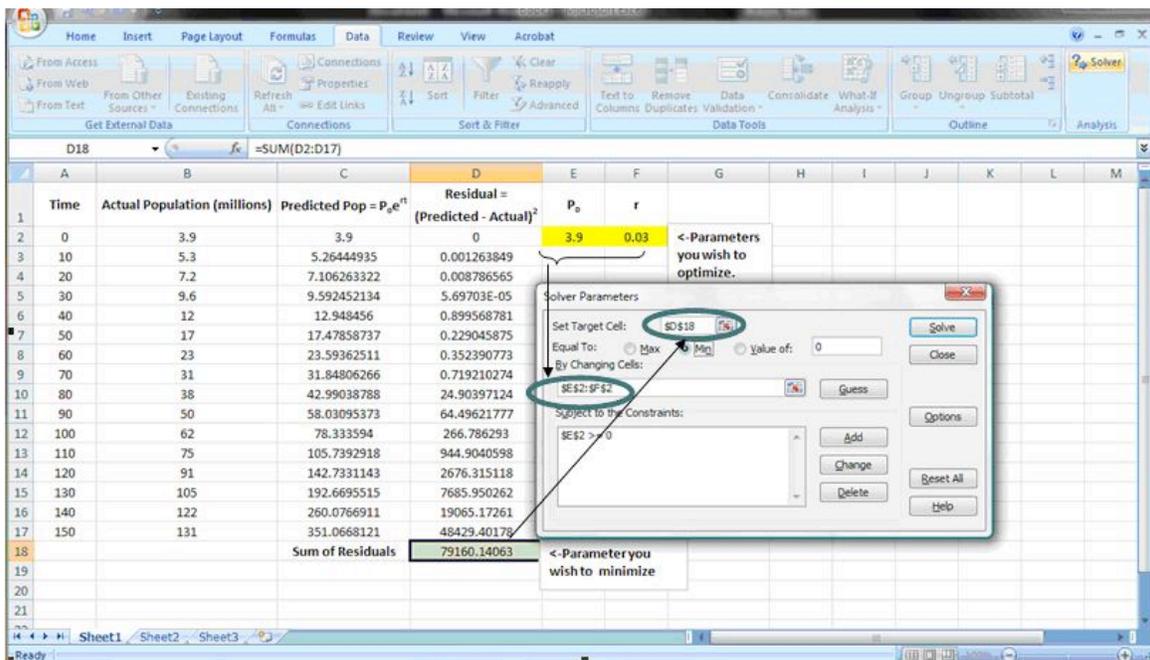
4.3.2 ANOVA Analysis

ANOVA is another tool on Excel that can be used for data analysis purposes. It is used primarily to compare continuous measurements to determine if they are sampled from the same or different distributions. This is usually done on set of data that consists of numerous samples and/or groups. Single factor ANOVA analyzes the difference between "groups" which can be different units, sensors, etc. On the other hand 2 way ANOVA analyzes the differences on the basis of both samples (trials) and groups, allowing you to compare, which one most affects the system. Two-way ANOVA can be employed for two different cases: data with or data without

replicates. Data without replicates is used when collecting a single data point for a specified condition, while data with replicates is used when collecting multiple data points for a specific condition.

The telling result of ANOVA analysis is a p-value telling you if there is a statistically significant difference between data sets. A more in depth discussion of the uses and implementation of ANOVA can be seen here.[1]

4.4 Random Number Sampling

A useful numerical technique is the random number generator. This generator uses random number sampling to create a predictive model. This can be especially useful in probability problems.

A typical random number generator produces any number between 0 and 1 with equal frequency. At each step a random number is generated that picks a possible outcome. A typical Excel statement would appear as the following: **IF (RAND()<X, value1, value2)**. This means that X percent of the time, you would get a value1 and 1-X percent of the time, you would get a value2. For example, suppose a situation exists with two possible outcomes, A and B, where A takes place 70% of the time. If a random number is generated that is greater than 0.3, then the outcome is A. Inversely if the random number is less than or equal to 0.3, then B is the outcome. This would be represented as IF(RAND()<0.7, A, B). Each step within the generator does not affect the next step, i.e. each step is determined individually. A process in which previous outcomes do not affect later outcomes is known as a Stochastic method or "Monte Carlo" method. A Monte Carlo simulation uses a model that takes random input that follows an assumed distribution (in this case, 70% A and 30% B) and produces output that models some phenomenon. The following is a simple example of such a simulation:

- A botanist is conducting an experiment that involves growing 15 plants from seed in a low-light room. There is a 30% chance that a given seed will germinate. The experiment is to be repeated three times. How many of the seeds germinate in each of the experiments?

- *In this example, the random input for each trial would be 15 random numbers between 0 and 1. By tabulating how many of the numbers are greater than 0.3 and how many are less than or equal to 0.3 you know how many of the seeds sprouted.*

The Excel is equipped with a random number generator which can be called out using the **RAND** function. A random number between 0 and 1 can be generated in a cell by the following command:

1. **=RAND()**

In Microsoft Excel 2007 the **RAND** function returns a new random number every time the worksheet is calculated or when F9 is pressed. If the same random number is needed to be referred to more than once in an Excel IF, AND, OR statement, one can create a column of random numbers using **=RAND()**, and refer back to these cells in an adjacent column.

There are many ways to manipulate the random numbers that Excel produces. A random number can be generated within a different range by multiplying the **RAND** function by a constant, i.e. multiply by 100 to change a decimal to percent. Another option to generate numbers within a different range is the **RANDBETWEEN** function. **RANDBETWEEN** generates a random integer within a specified range. The following command generates a random integer between -6 and 21:

- **=RANDBETWEEEN(-6,21)**

It is also possible to modify your random numbers by nesting the **RAND** function within another function. For example, one can square the results of one distribution to create a different distribution of numbers.

Another way to utilize the **RAND** function is to take a random sampling of a data set. If one has a spreadsheet with several hundred data points, and would like to analyze only 100 random points, another column can be created titled "Random Number" or something similar. In the first cell of this column, type the **=RAND()** function. Then drag this function all the way to the end of the data set. At this point, the user can go to the DATA drag-down menu and select Sort... In this menu, select your random number column in the SORT BY drag down menu. Either ascending or descending can be selected, as desired. Once the spreadsheet is sorted by the random numbered cells, the first 100 data points can be used as a statistically random cross-section of the data.

The sort tool can be useful in cases when you are trying to find out the probability of having certain values above a certain threshold. For example, if the 100 data points that you have represent the temperature variations occurring within a system, you can use sort tool to determine how many temperature values exceed your threshold. From that, you can calculate approximately the probability of the system temperature exceeding the threshold. Two screenshots of the sort process are shown below. Microsoft Excel 2007 uses a different menu, allowing for one to choose between sorting from Largest to Smallest or vice versa. A screenshot below shows this new menu.

Sorting Setup:

Results to Randomization:

Data Point	X-Value	Y-Value	Random #
130	260	1.1	0.34241
29	58	2.1	0.829027
173	346	2.8	0.775198
202	404	3.9	0.548976
71	142	2.9	0.804771
172	344	4	0.102228
169	338	8.5	0.275583
223	446	1.4	0.26525
159	318	2.1	0.593537
2	4	3.7	0.373848
281	562	4.8	0.539074
84	168	3.6	0.800056
260	520	1.1	0.622218
245	490	4.7	0.064465
9	18	2.5	0.142237
92	184	3.1	0.350258
205	410	7.2	0.782673
166	332	5.4	0.875973
163	326	6.5	0.442182
24	48	2.2	0.161209
167	334	4.7	0.92963
41	82	1.4	0.501325
113	226	2.5	0.128892
160	320	8.5	0.593928
22	44	3.4	0.301458
266	532	3.6	0.540075
103	206	4.6	0.75285
158	316	3.7	0.03427
214	428	3.6	0.283944
212	424	8.5	0.061719
14	28	3.1	0.7407
82	164	8.5	0.484194

Microsoft Excel 2007:
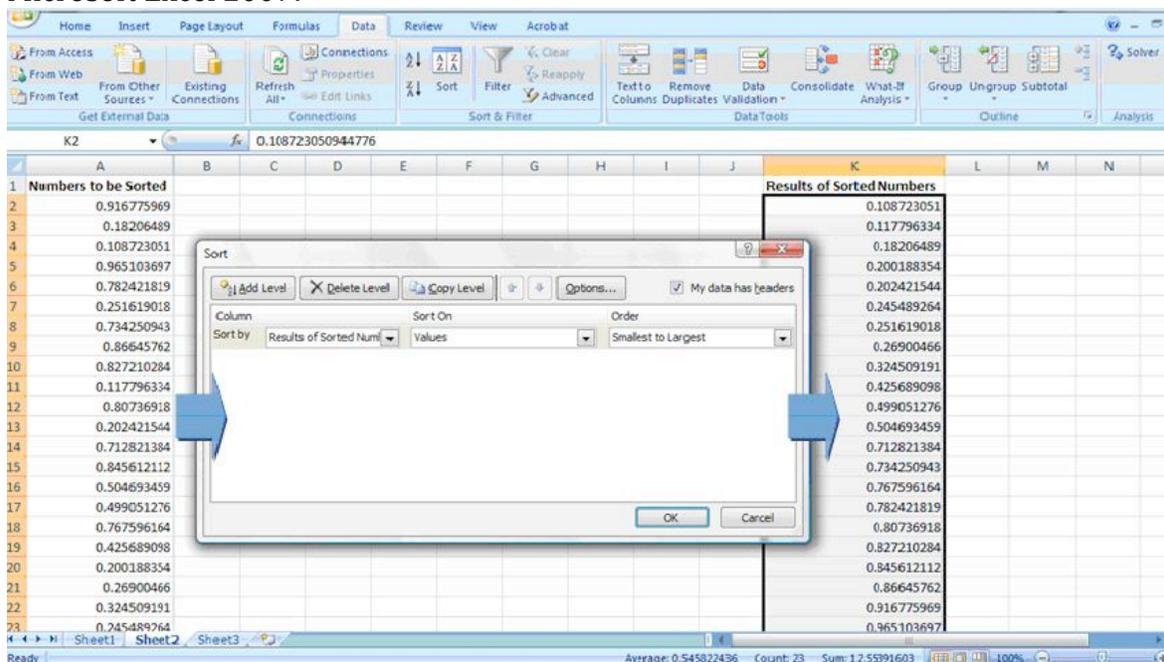
4.5 Worked out Example 1

The Arrhenius equation defines the relationship between a reaction rate constant k and temperature:

$$k(T) = Ae^{-E/RT}$$

where T is the absolute temperature, A is the frequency factor, E is the activation energy, and R is the universal gas constant.

It is frequently desirable to be able to predict reaction rates at a given temperature, but first you need to know the values of A and E. One way to obtain these values is to determine the reaction rate constant at a few temperatures and then perform a nonlinear regression to fit the data to the Arrhenius equation. Given the following rate constant data for the decomposition of benzene diazonium chloride, determine the frequency factor and activation energy for the reaction.

$k\ (\mathrm{s}^{-1})$	$T\ (\mathrm{K})$
0.00043	313.0
0.00103	319.0
0.00180	323.0
0.00355	328.0
0.00717	333.0

78

Solution: The following Excel file contains a solution to the problem that uses Excel's solver tool to perform the nonlinear regression and determine the values of *A* and *E*:

Example 1

The spreadsheet is set up as described in the **nonlinear regression** section above.

1. Open solver (Tools / Solver)
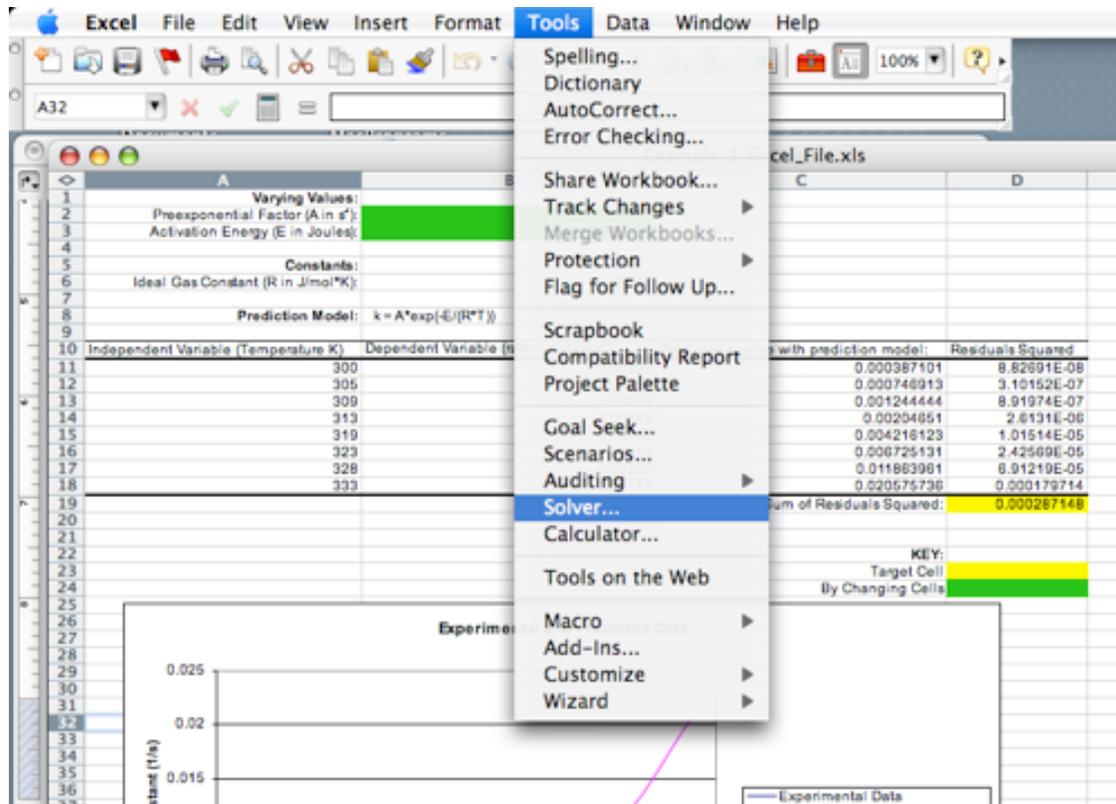
2. Set the Sum of Residuals Squared value cell as the target cell
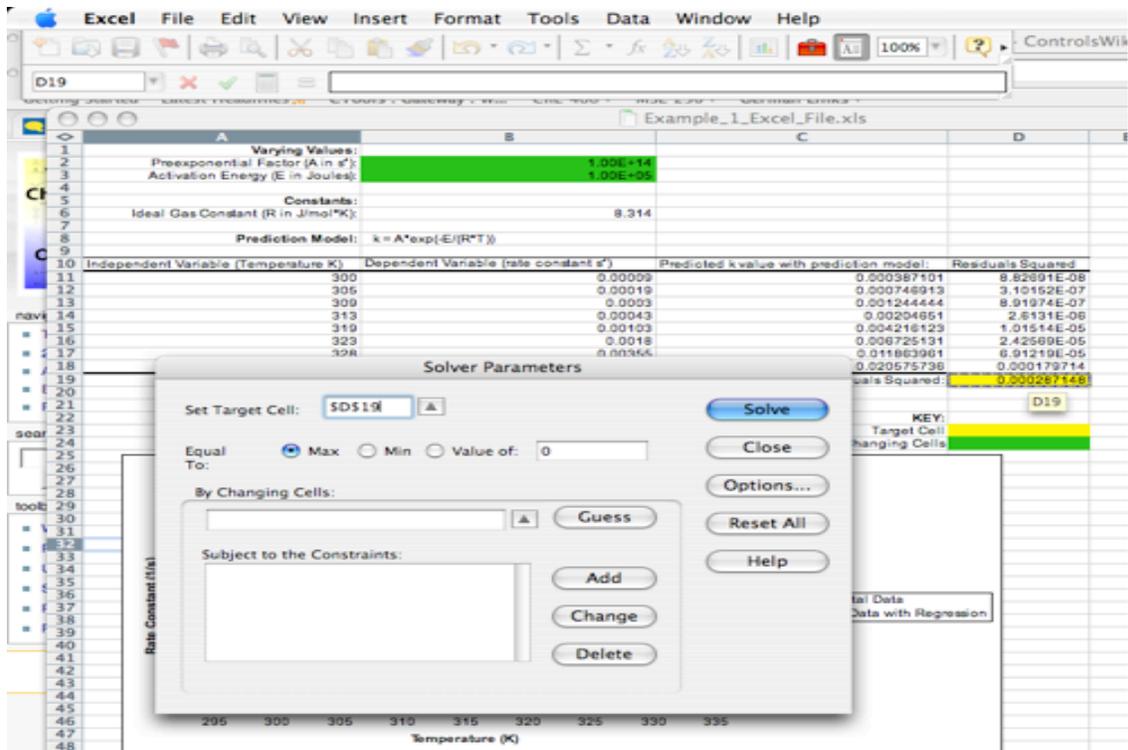
3. Click the Min radio button
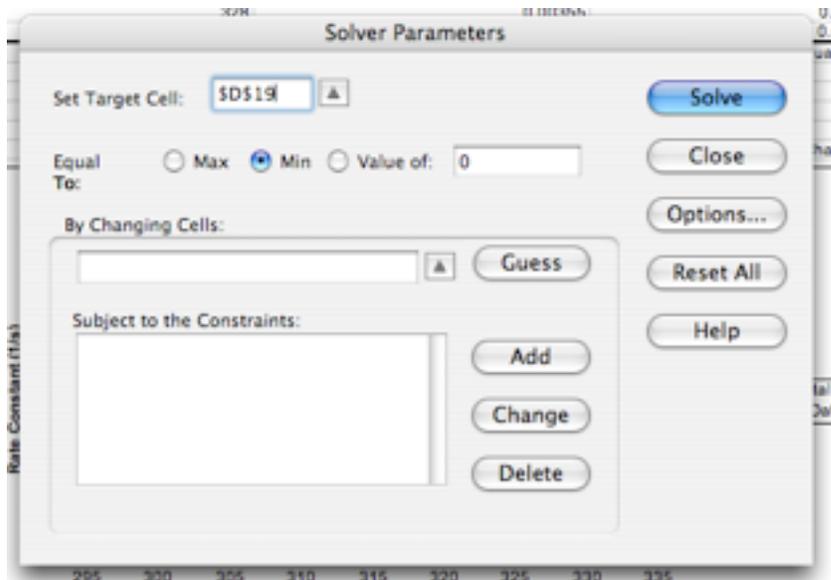
4. Set Pre-exponential Factor and Activation Energy as the adjustable cells. For this problem, keep *A* between 1E+13 and 1E+14 s^-1 and *E* between 9.5E+4 and 1.05E+5 J

5. Sometimes, depending on what version of excel you are using, a message will come up like the one below. If this happens, click OK. The graph will update with the new curves.
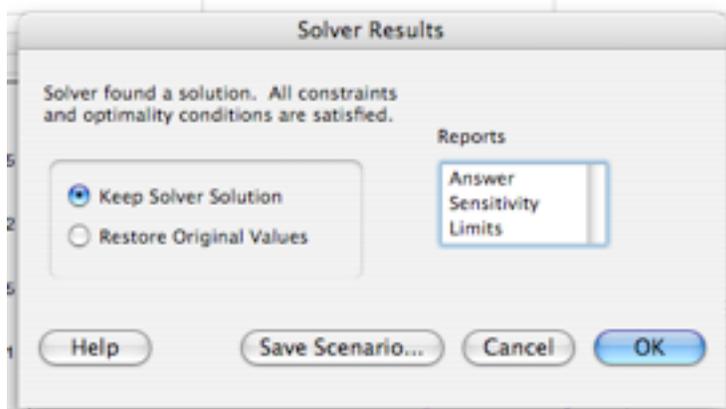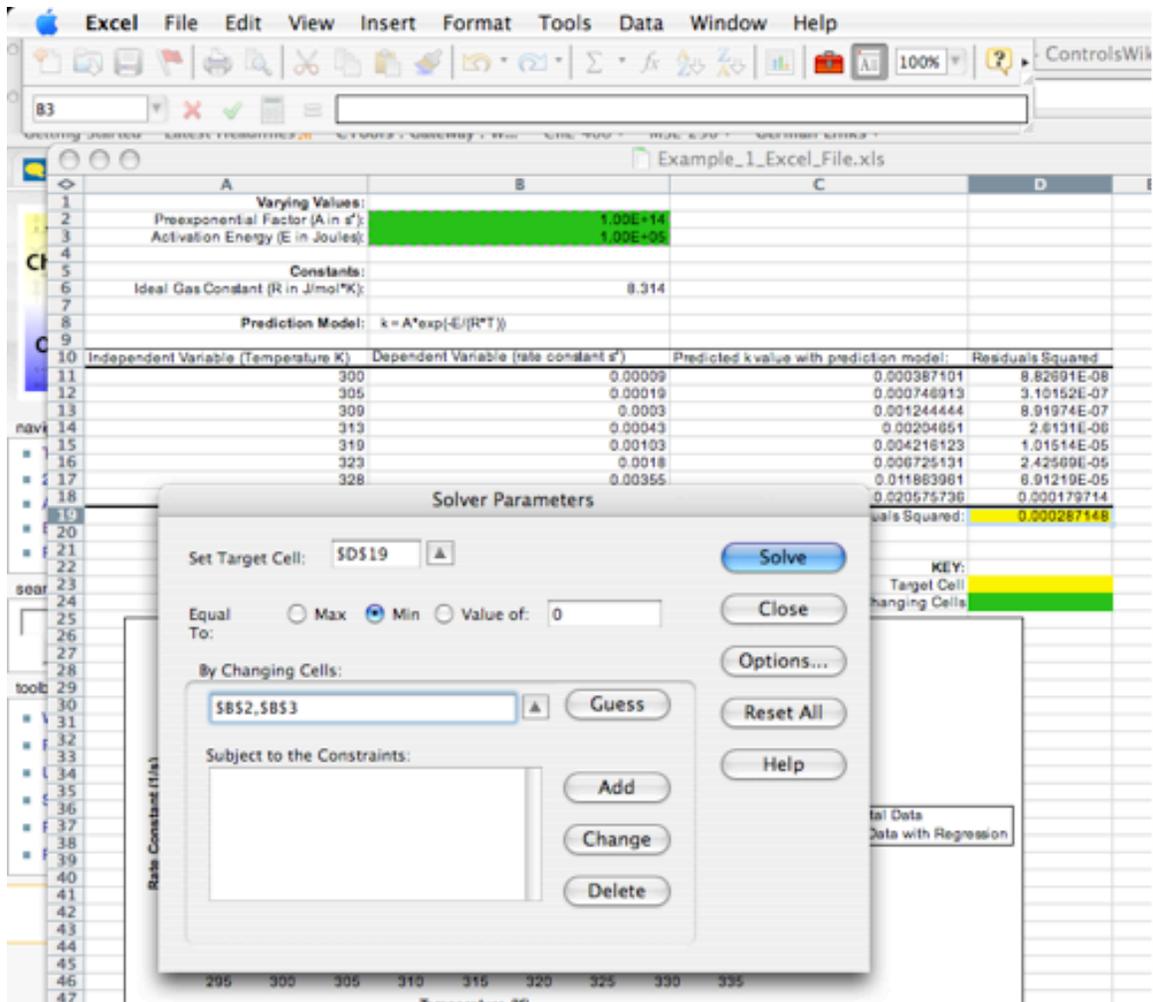
6. Click Solve and observe the values of A and E calculated by Solver and the changes to the plot showing the rate constant as a function of temperature.

NOTE: If solver is not fitting the data well, click the **options** button in the solver window and make the precision value smaller by several orders of magnitude.

4.6 Worked out Example 2

Kinetics can be used to determine the probability of a reaction taking place between gas-phase molecules. Consider molecules A and B, which have a 60% chance of reacting upon collision. Create an Excel spreadsheet that utilizes a random number generator and an if-statement to model an individual collision.

Solution

In this Excel file, sample coding for Random number generator and if-statement is shown.

- Collision Example Excel File:

Collision_ex

Random number generator is used to output numbers between 0 and 1 for four trials, ten trials, and one hundred trials on sheets one, two, and three, respectively. Formula for the random number generator is employed in the cell where you want the number to be output (in this case, B6 is the cell). An IF logical test is then employed to test whether the number follows the requirement or not. The IF-statement cell will then output the corresponding result based on the logic test. In this collision case, if the random number function generated a number greater than 0.4, the logic test will return TRUE and then output "Reacts" as the result. Assuming that the random number generator produces an even distribution between 0 and 1, the random number will be greater than 0.4 sixty percent of the time. Coding of the IF-statement is shown:

- =IF(B6>=0.4,"Reacts","Does not React")

On the other hand, if the number generated is less than 0.4, the logic test returns FALSE and "Does not React" will be shown as the result. Different numbers are generated each time the spreadsheet is updated, thus different result will return for the IF function. You can also modify the numbers and return statement to better understand the operation of the functions.

The number of reactions is summed up and divided by the total number of trials to compare the predicted reaction probability with the value given in the problem. It can be seen by comparing sheets one, two, and three that increasing the number of trials decreases the variation in the predicted reaction probability.

On many computers, excel tools such as solver and the data analysis package are not already included and must be installed before used. This section will go over how to access these add-ins for those using Windows opposed to Macs.

1) Click on the windows symbol in Excel 2007. At the bottom of the opened window, select the Excel options box.
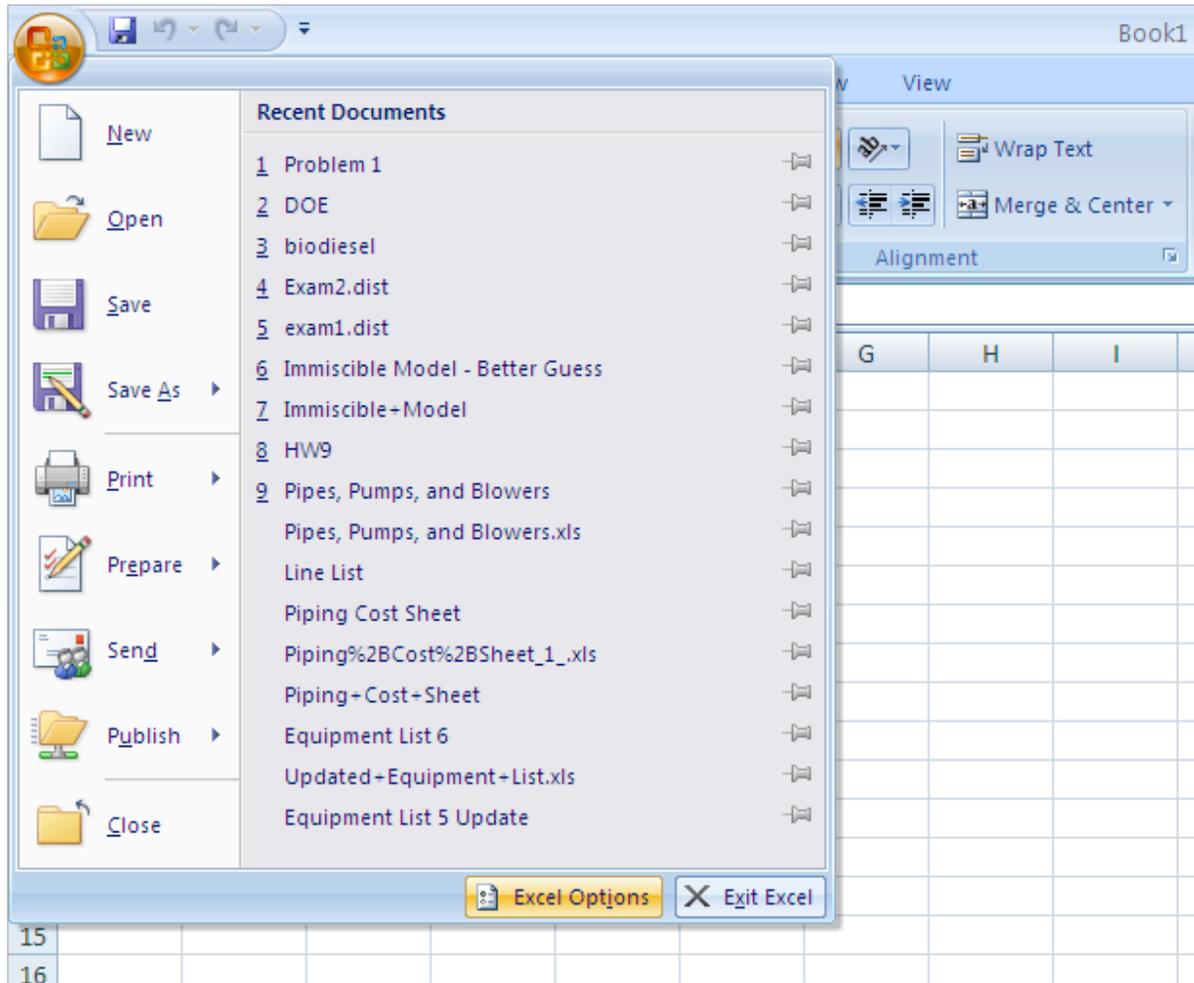
2) On the left side panel you will see a section labeled "Add-ins". Select this.
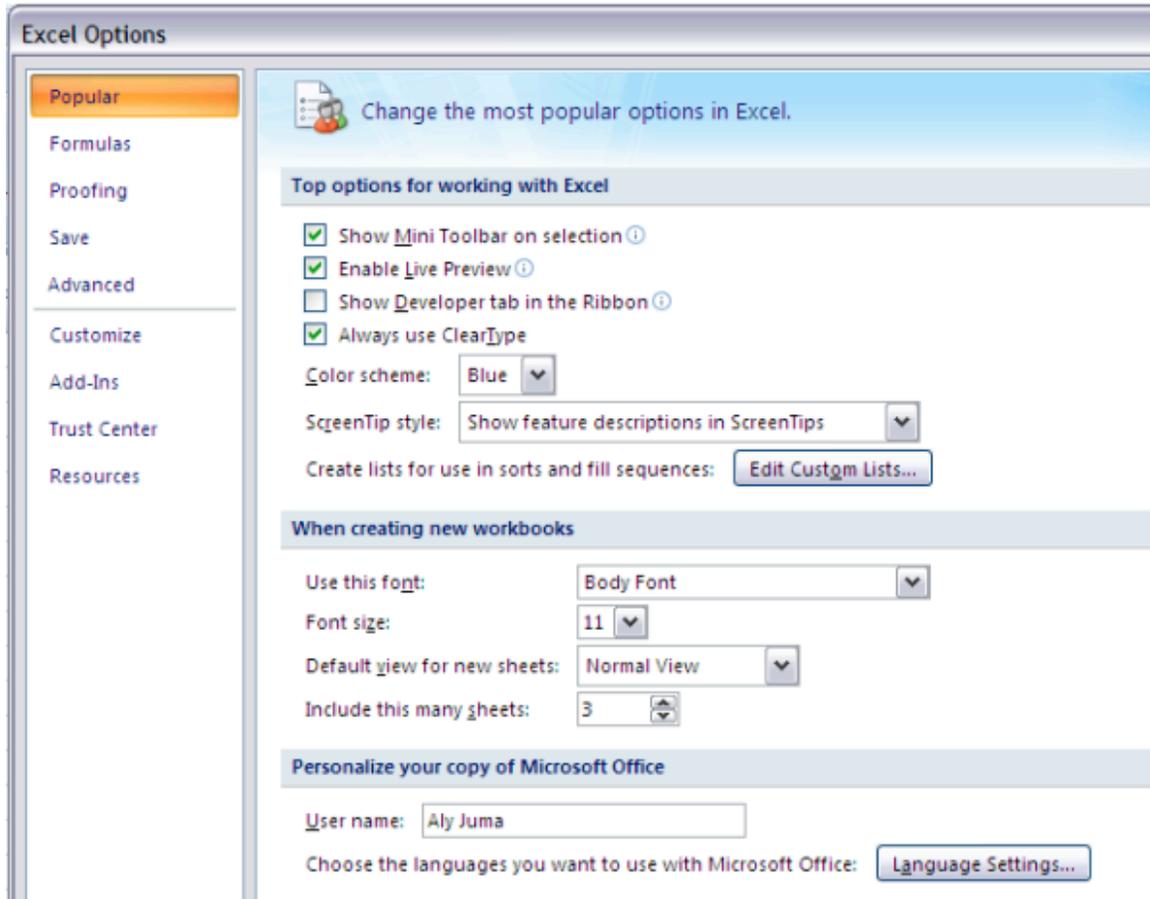
3) Here you will see a list of "Inactive Applications" that are available for excel. Of these the ones we are interested in are the "Solver Add-In" and the "Analysis ToolPak." Select one of these and click "Go" at the bottom of the window.
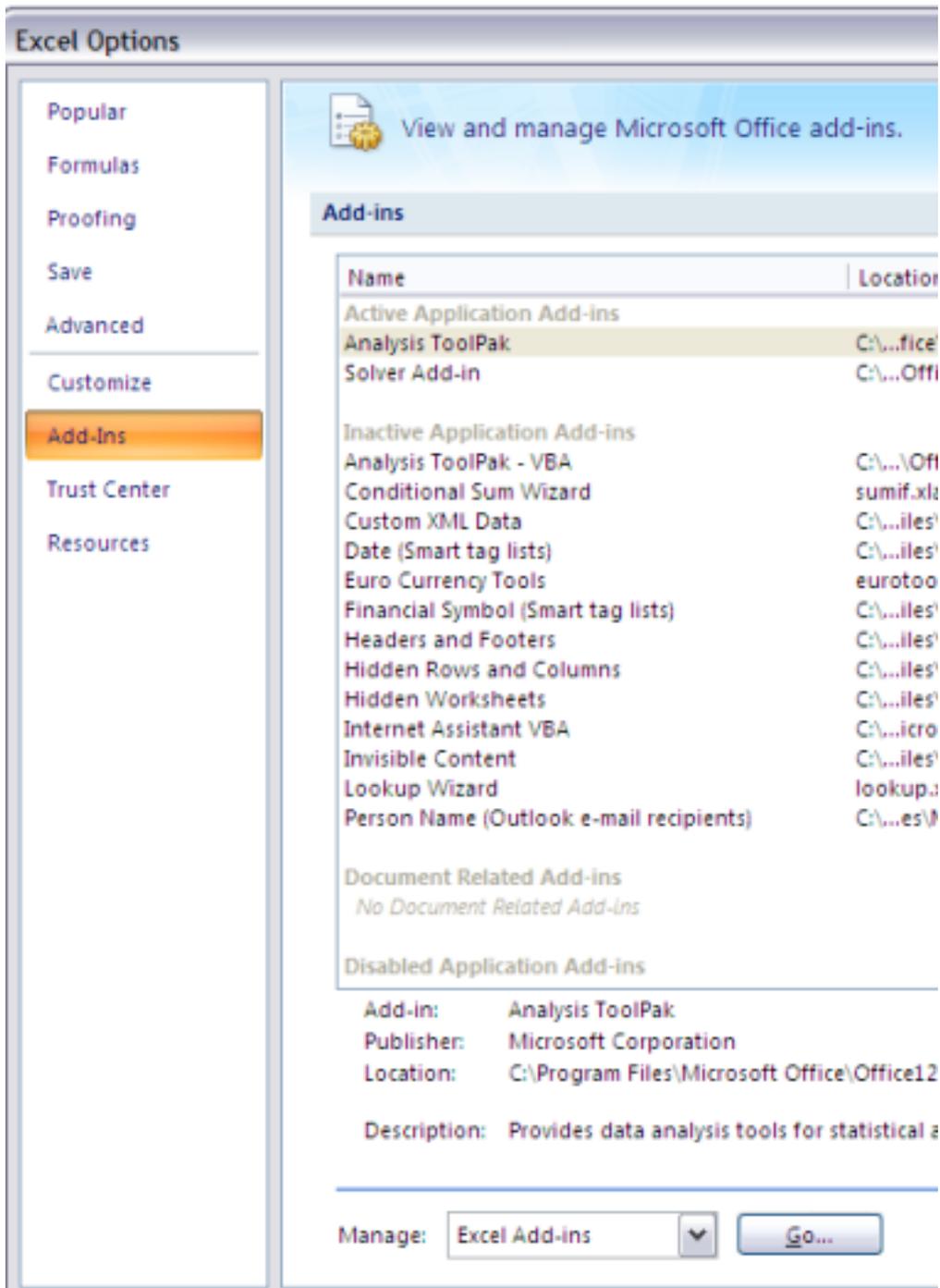
4) A new window will appear asking you which components of the package you would like to install. Select the needed tools (solver and analysis toolpak). Click "Ok" and windows should install the corresponding tools into excel.
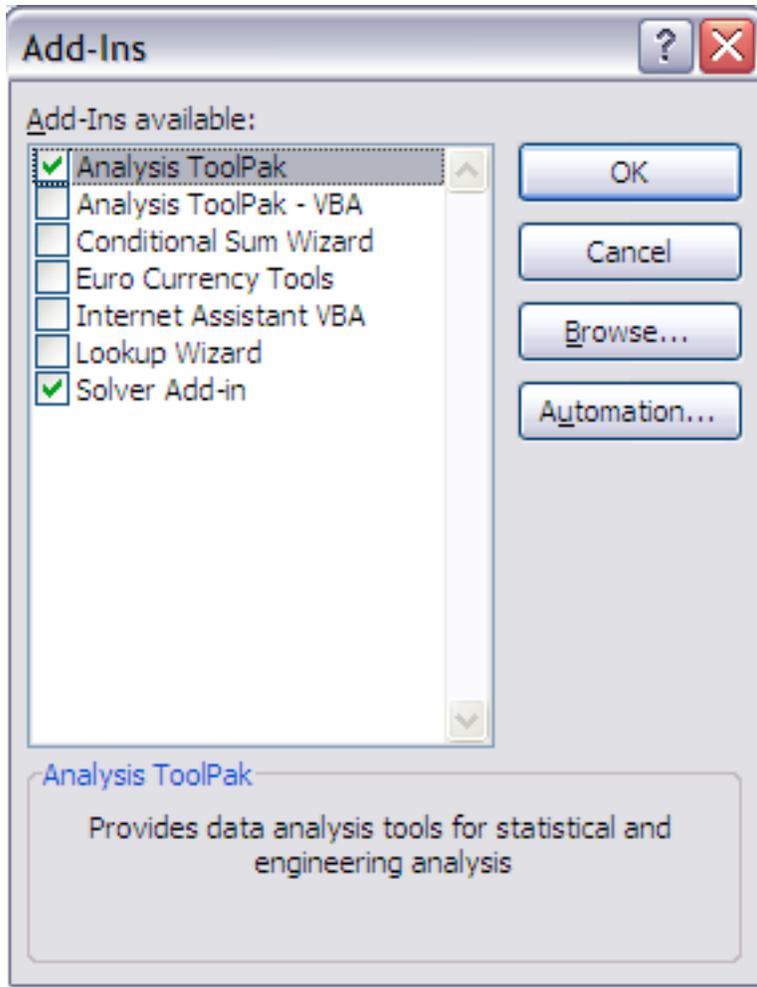
5) Once done, these tools will be found under the data section of Excel, on the far right of the window.
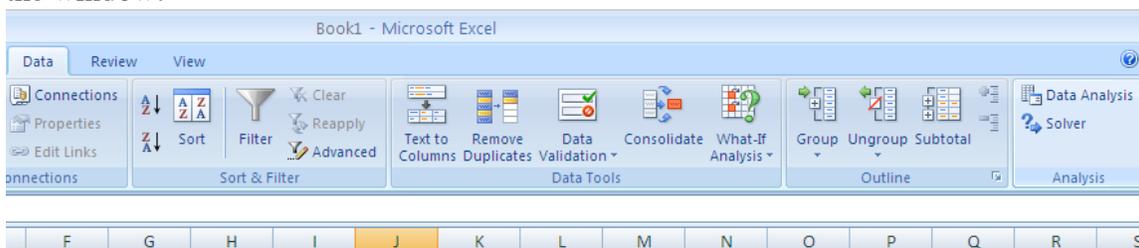
4.8 Sage's Corner

Example of Non-linear regression
http://video.google.com/googleplayer.swf?docId=-6226344938424937989
Media:NonLinearRegression_No_narration.ppt slides

Monte Carlo Integration in Excel
http://video.google.com/googleplayer.swf?docId=5563937806029403905

Powerpoint Slides with No Narration
Excel Spreadsheet Used in Monte Carlo Integration Example

4.9 References

- Bender, E.A. *An Introduction to Mathematical Modeling,* Mineola, NY: Dover Publications, Inc.
- Fogler, H.S. (2006). *Elements of Chemical Reaction Engineering,* Upper Saddle River, NJ: Prentice Hall Professional Technical Reference. ISBN 0-13-047394-4
- Microsoft, Excel, Proprietary EULA, www.microsoft.com.

Section 5. Noise modeling: more detailed information on noise modeling: white, pink, and brown noise, pops and crackles

Note: ⊞Video lecture available for this section!

Authors: Danesh Deonarain, Georgina Mang, Teresa Misiti, Carolyn Ehrenberger

Date Presented: September 08, 2006 Date Revised: September 19, 2006

- First round reviews for this page
- Rebuttal for this page

5.1 Introduction

Noise is all around us in all sorts of forms. In the common use of the word, noise refers to sound. However, noise can more accurately be thought of as random variations that are always present in one or more parts of any entity such as voltage, current, or even data. Rather than thinking of noise only as an acoustic term, it should be thought of more as a random signal. Noise can be the inherent fluctuations in some part in a system (ie. temperature at a given point) or it can be the unavoidable interference on a measurement from outside sources (ie. vibrations from a nearby generator blur measurements from a pressure transducer). The static interference on your radio, the 'snow' on your television, and the unresolved peaks on an infrared spectroscopy report are all examples of noise.

Chemical engineers can use statistical properties to characterize noise so they can understand a current process or develop an optimal process. By characterizing the noise and determining its source, the engineer can devise methods to account for the noise, control a process, or predict the path of a system. For example, when chemical engineers design plants, they use characterizations of noise to determine the best control scheme for each process. Mathematical modeling can be used to characterize and predict the noise of a given system. In modeling, to simplify the representation of general trends that reoccur, noise is classified in two major categories: frequency based and non-frequency based. **Frequency based noise** consists of the colors of noise, and **non-frequency based noise** includes pops, snaps and crackle.

The purpose of the following sections is to give you a qualitative understanding of the sources of noise and how it can be characterized and handled. Examples will be given to help you gain a quantitative understanding of how noise relates to controlling and modeling chemical engineering processes.

5.2 General Types of Noise

Three general types of noise can be categorized in describing a process, and are as follows:

88

1) Gaussian noise is usually not dependent on time, meaning that it is random and not systematically planned. The amplitude of the frequency can vary, making a crackling notation or sound (see #Crackles below). Some examples of Gaussian noise can include splashing in a tank or unplanned interruption in a sensing device.
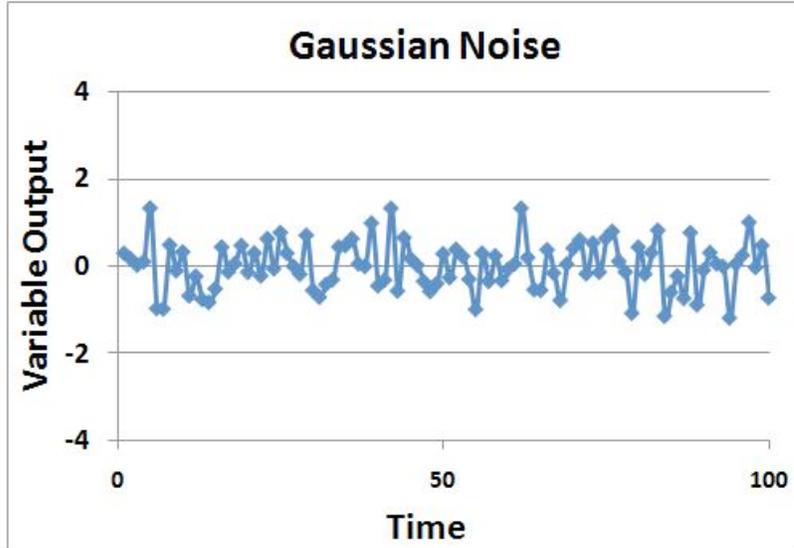
2) Drift noise is correlated to time, and has random movement. Examples of drift noise can include fouling or catalyst decay, in which the potency of the substance may decline over time as decaying occurs. Stock price fluctuations also model somewhat like drift noise.
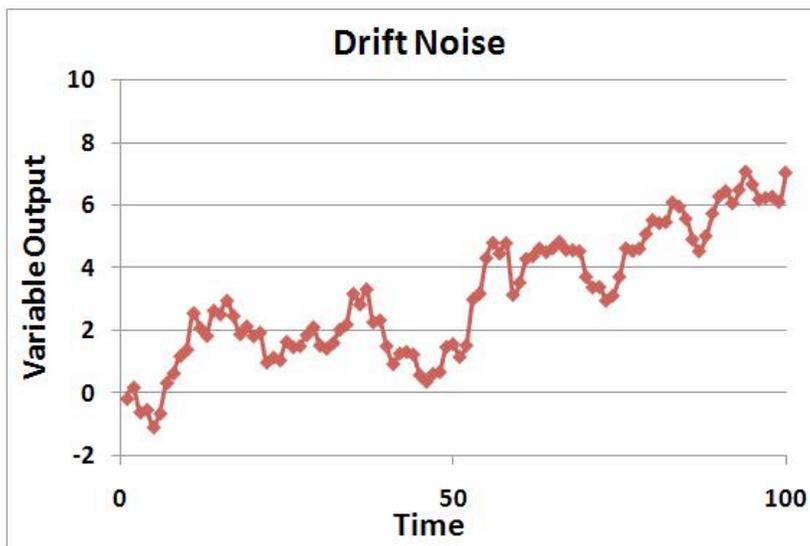
3) Shot noise may be defined as sporadic and short bursts of noise, in which the amplitude is similar among the bursts of noise. Shot noise can be correlated to pops (see

#Pops below), in which at random times, the same shot noise is witnessed with the same amplitude. Examples of shot noise include partial clogging or jamming in the process in which the same amplitude will be seen by the noise whenever the clogging or jamming in the process occurs. Another example is customer demand of the product. If the control system (when trying to optimize it) depends on the customer order, and customer demand is not consistent at all times (meaning downtimes for orders) but the order amount is the same when orders are placed, then it will be effected by the shot noise described by the customer order (customer demand).
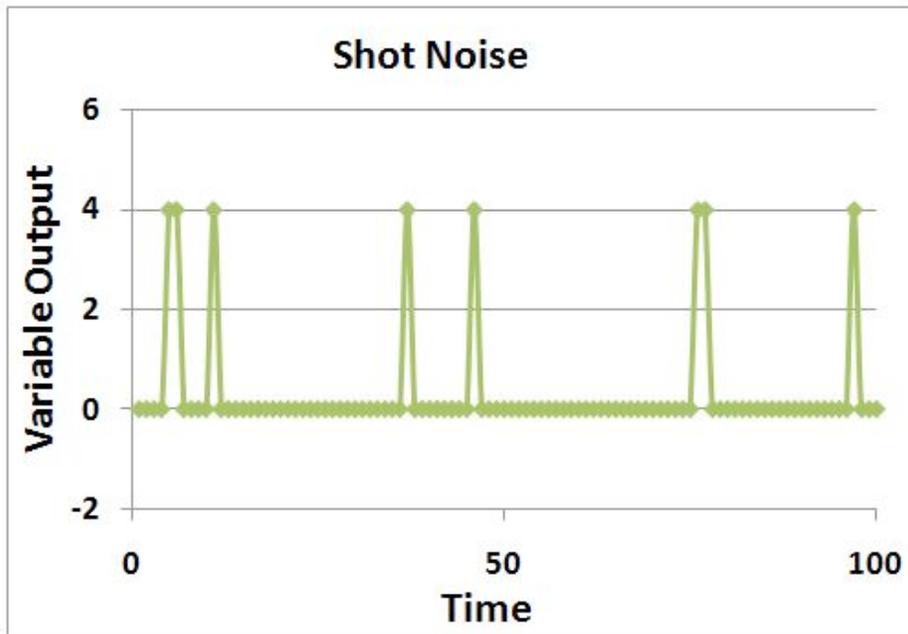
Microsoft Excel Formulas for 3 Types of Noise

Noise Type	Excel Formula
Gaussian Noise	= [*Original value*] + RAND() - RAND() + RAND() - RAND()[1]
Drift Noise	= [*Previous value*] + RAND() - RAND() + RAND() - RAND()[1]
Shot Noise	= if(RAND()>0.9, [*Original value*]+[*single error*], [*Original value*])

[1]The more add/subtract cycles of RAND() you use, the closer approximation you'll get for a Gaussian distribution
of random errors.

90

As you would see these in an Excel file:

\	A	B	C	D
1	Time	**Gaussian Noise**	**Drift Noise**	**Shot Noise**
2	1	**original value**	**original value**	**original value**
3	2	= B2 + RAND() - RAND() + RAND() - RAND()	= C2 + RAND() - RAND() + RAND() - RAND()	= if(RAND()>0.9, D2+[*single error value*], D2)
4	3	= B2 + RAND() - RAND() + RAND() - RAND()	= C3 + RAND() - RAND() + RAND() - RAND()	= if(RAND()>0.9, D2+[*single error value*], D2)

Note: These formulas can be extended for any length of time. Longer periods of time are useful for visualizing drift noise.

5.2.1 Combined Types of Noise

Sensors, processes, demands, etc. often do not behave with simple noise models. However, most noise can be derived from the three general types of noise. An example of a process with random and shot noise is shown in the thumbnail.
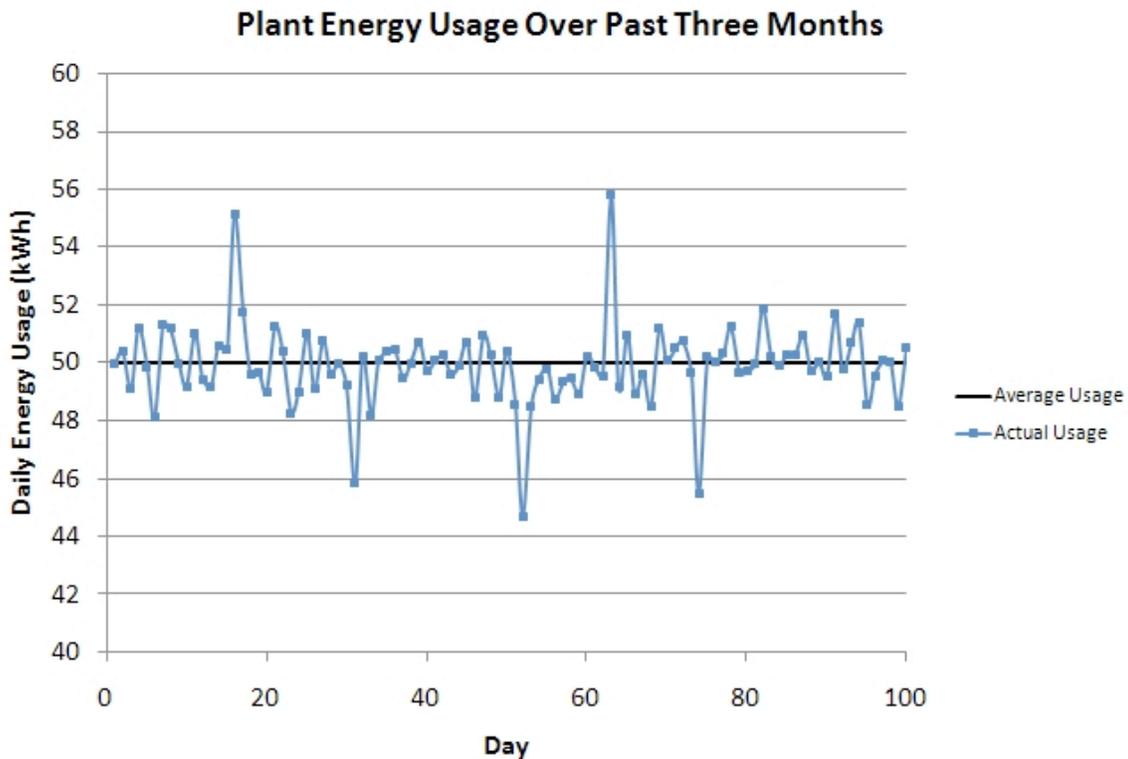
This is a graph of the total daily electricity usage for a plant during a normal operation period. As we can see, there are minor fluctuations occurring every day, but some days there are large changes of electricity usage. These large changes can be seen as shot noise when the consumption is viewed on this daily scale.

Possible reasons for the minor fluctuations could be due to electric heaters kicking on and off. Changes in the operation of smaller pumps can cause these changes in electricity demand.

The large shots could be due to an energy-intensive part of the process that only operates when needed. The positive shots could be due to this process kicking on and the negative shots could be due to a large process kicking off when not needed.

With this graph, we can say that there is shot noise present with the large changes in consumption, but there is also random noise present with the minor daily fluctuations. We would not say there is drift noise present because the consumption remains around a constant mean value and does not drift away from there.

However, the total consumption when viewed from a different scale may show other forms of noise. If we viewed the hourly usage of electricity on one of the days with shot noise, it may resemble drift noise on the hourly scale.

5.3 Colors of Noise

The first category of noise is frequency-based noise, classified by the colors of noise. In understanding the colors of noise, we need to understand the process of converting a given set of data to a point where it can be classified as some color of noise. The section below, "Modeling Noise," goes into more detail as to how this is done. At this point, it is important to understand the main colors of noise which are possible in a system. Noise is classified by the spectral density, which is proportional to the reciprocal of frequency (f) raised to the power of beta. The power spectral density (watts per hertz) illustrates how the power (watts) or intensity (watts per square meter) of a signal varies with frequency (hertz).

$$PSD \propto \frac{1}{f^\beta} \quad \text{where } \beta \text{ is } \geq 0.$$

Just like colors of light are discriminated by their frequencies, there are different colors of noise. Each beta value corresponds to a different color of noise. For example, when beta equals zero, noise is described as white and when beta equals two, noise is described as brown.

5.3.1 White Noise

One major type of noise is white noise. A defining characteristic of this type of noise is it has a flat power spectral density, meaning it has equal power at any frequency. β equals zero for white noise.

White noise and white light have similar properties. White light consists of all visible colors in the spectrum while white noise is created by combining sounds at all different frequencies.

Pure white noise across all frequencies cannot physically exist. It requires an infinite amount of energy and all known energy is finite. White noise can only be created within a specific and defined range of frequencies. Again making an analogy to white light, for a small band of frequencies, *visible* white light has a flat frequency spectrum.

Another property of white noise is that it is independent of time. This factor also contributes to the idea that white noise cannot physically exist since it is impossible to have noise be completely independent of time.

A sample power spectral density chart of white noise is shown below. As can be seen from the chart, the power over all displayed frequencies is essentially the same, suggesting the PSD is equal to 1.

It is important to note that white noise is not always Gaussian noise. Gaussian noise means the probability density function of the noise has a Gaussian distribution, which basically defines the probability of the signal having a certain value. Whereas white noise simply means that the signal power is distributed equally over time. For gaussian white noise one finds the number of measurements greater than the power spectral line as measurements that are less than the power spectral line with majority falling at the spectral line. Gaussian white noise is often used as a model for background noise in satellite communication. White noise can also come from other distributions, such at the Poisson distribution. Poissonian white noise will look like a normal distribution that has been shifted to the left for small number of measurements. For larger number of measurements Poissonian white noise will look like the normal distribution (also known as gaussian white noise). A Poissonian white noise model is useful for systems where things happen in discrete amounts such as photons being transmitted or the number of cars that go through an intersection.
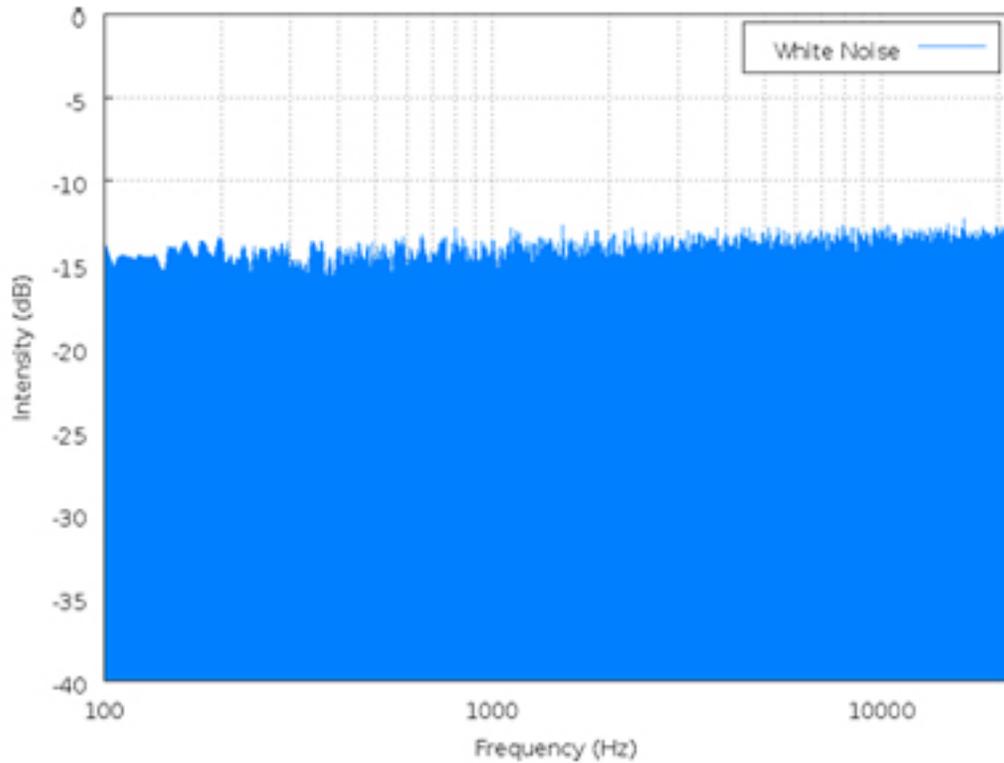
image from http://en.wikipedia.org/wiki/White_noise on 9/18/2006

5.3.2 Pink Noise

Pink noise is a signal whose power spectral density decreases proportionally to the inverse of the frequency, where the β value is equal to one. Because of its decrease in power at lower frequencies when compared to white noise, pink noise often sounds softer and damper than white noise.

Pink noise is actually what is commonly considered audible white noise and is evident in electrical circuits. Pink noise becomes important to measure in things such as carbon composition resistors, where it is found as the excess noise above the produced thermal noise.

Below is a power spectral density chart for Pink noise.

94

Pink Noise Power Spectral Density
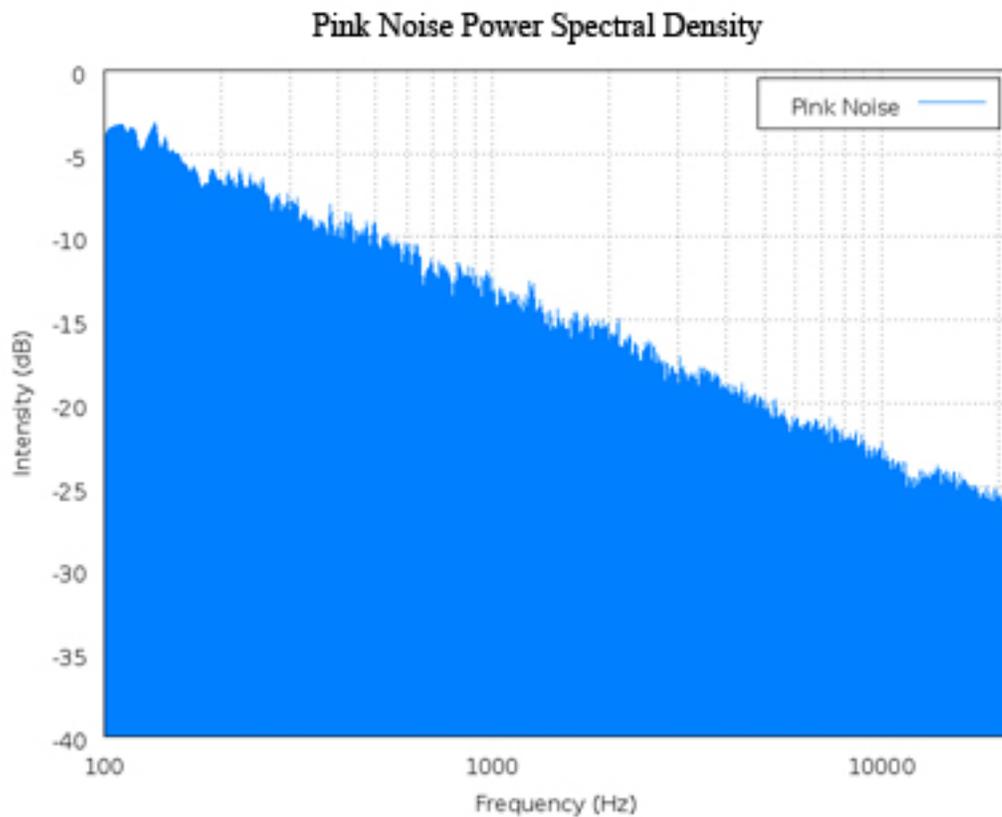
image from http://en.wikipedia.org/wiki/Pink_noise on 9/18/2006

5.3.3 Brown Noise

When β equals 2, the noise is Brownian. Brown noise is signal noise created by Brownian, or random, motion. It is a form of unavoidable interference in the transmission of information and can be compared to a random walk that does not have a clearly patterned path.

Brown noise has more energy at lower frequencies. Compared to white or pink noise, brown noise is even more soft and damped. It can be generated by integrating white noise. Interference from outside sources, such as vibrations from nearby machinery or background light, in instrument readings usually have a brown noise pattern.

Below is a power spectral density chart for brown noise. From the charts of brown noise and pink noise, it can be observed that brown noise loses power as frequency increases at a much faster rate than that of pink noise.

Brown Noise Power Spectral Density
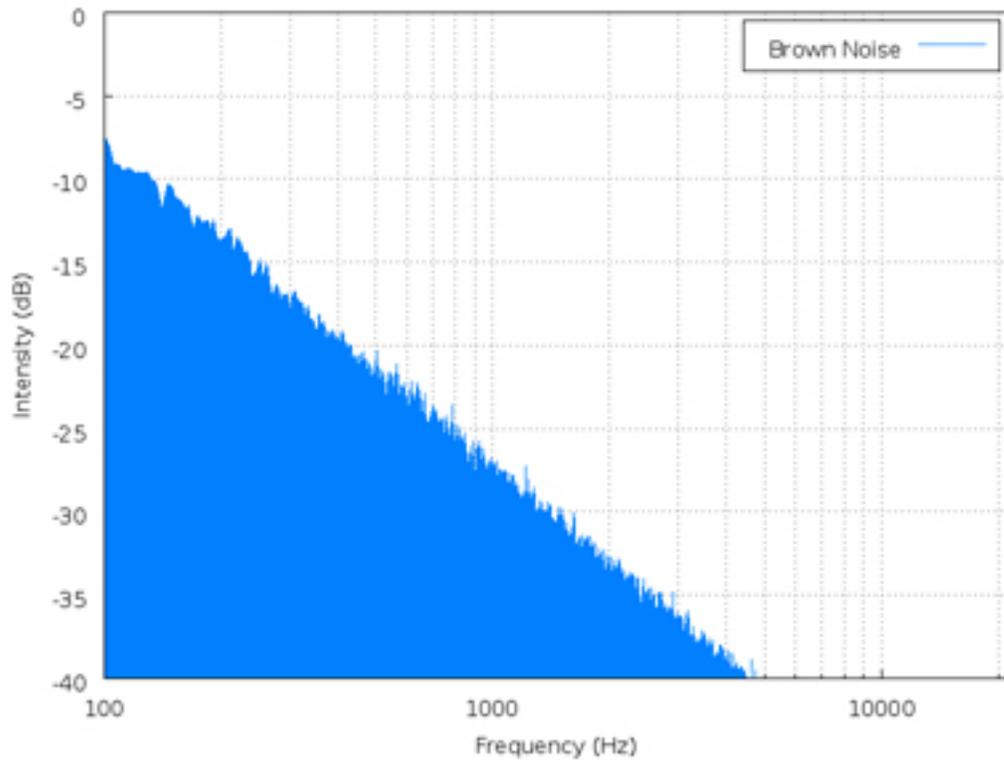
image from http://en.wikipedia.org/wiki/Brown_noise on 9/18/2006

5.3.4 Blue Noise

Blue noise is a signal whose power spectral density increases proportionally to the frequency, where the β value is equal to negative one.

An example of would be a reversable, exothermic, batch reaction where the products are being accumulated at an overall constant rate, but the reverse reaction occurs when the temperature gets too high. Thus it is overall increasing in product, but has small fluctuations as it increases.

Below is a power spectral density chart for blue noise.

96
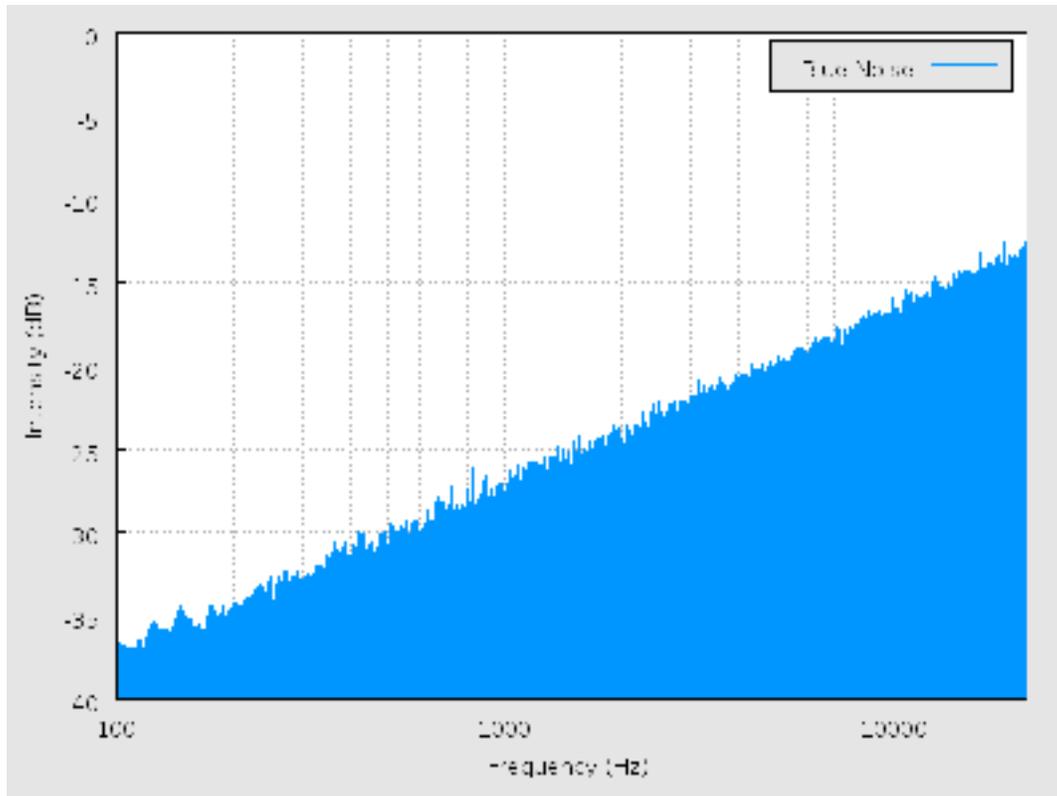
5.3.5 Purple Noise

When β equals -2, the noise is violet. Purple noise has more energy at higher frequencies. It can be generated by differentiating white noise.

Below is a power spectral density chart for purple noise. From the charts of purple noise and blue noise, it can be observed that purple noise gains power as frequency increases at a much faster rate than that of blue noise.

5.3.6 Modeling Colors of Noise

Characterizing noisy signals is important to a chemical engineer so that he or she can determine the sources of noise and how to account for it. Sometimes, the noise will be a characteristic of the system. In these cases, the system can not have a very sensitive control because if the sensor for the controlling device is too sensitive, it will respond to the noise of your system and end up controlling your system in a noisy way. Other times, the noise will be from unavoidable disturbances to the system. In these cases, the noise can be damped or accounted for in an appropriate manner.

In order to apply the knowledge of the colors of noise we need to understand the process of converting a given set of data to a point where it can be classified as some color of noise. The general process of classifying noise follows.

Data $\Longrightarrow$ *Curve to Fit Data* $\Longrightarrow$ *Fourier Transform* $\Longrightarrow$ *Power Spectral Density* $\Longrightarrow$ *Classification of Noise Color*

The classification of noise color becomes important when we need to have an estimate of future data. In other words, given the trends in our current data we can use it to estimate future data. This is done as follows,

Noise Color $\Longrightarrow$ *Power Spectral Density* $\Longrightarrow$ *Fourier Transform* $\Longrightarrow$ *Curve* $\Longrightarrow$ *Generate Data From Curve*

The power of a noise signal is detected at a certain frequency. Then a plot of the log(power) vs. the log(frequency) can be constructed, and the slope of the line gives the beta value. Following a backward thought process, one can produce a certain color of noise by creating frequency components which have a value generated by a Gaussian distribution and then scaling by the appropriate beta power of frequency.

A general method to characterize and model noise is explained below.

1. DATA

Data is what the signal transmits. The signal is dependent on what you are measuring or modeling. The data can be collected, for example, if you're measuring temperature in a reactor, then your data is the temperature readings from a thermocouple at a certain position in the reactor over a period of time.

2. CURVES

After you collect the data, you plot the data and find a best fit equation, $x(t)$, for that set of data. A math program can be used to find a best fit equation. Microsoft Excel can be used for a simple model and programs such as Polymath or Mathematica can be used for more complex models. A simple equation could be in the form of $Asin(Bt)$. The coefficients A and B can be varied to fit the data. Additional terms can be added for more accurate modeling of complex data. The equation you find can then be used to predict and model future signal noise.

3. FOURIER TRANSFORMS

A Fourier transform is a tool used to convert your data to a function of $\frac{1}{f^\beta}$. In this form, the noise can be more easily characterized. Apply a Fourier transform to the curve, $x(t)$, you fit to your data to generate a relation for the power spectral density. The Fourier transform can be performed by a computer depending on the complexity of $x(t)$ (or see "Simplifying the Fourier Transform" below). The transform is the integral shown below.

$$X(w) = \int_{-\infty}^{\infty} x(t)e^{-jwt}\, dt$$

Where;

- $x(t)$ *is the equation of the curve to fit the data.*

- e^{-jwt} *is the exponential form of writing the relation*
$cos(wt) - jsin(wt)$ *. The j in the second term here indicates that that term is imaginary.*

- w *is the frequency*

4. POWER SPECTRAL DENSITY

This value is attained by simplifying and squaring the Fourier Transform. Since the series of simplifications needed are fairly complex, the resulting power spectral density equation is displayed below.

$$PSD = \sqrt{\left(\sum x(t)cos(wt)\right)^2 + \left(\sum x(t)sin(wt)\right)^2}$$

At this point we attain a numerical value for the PSD at the particular frequency, w. These numerical PSD values can be plotted versus frequency to obtain the PSD chart.

5. CLASSIFICATION OF NOISE

The summation is repeated over and over for different frequencies, w. A plot of the PSD vs. frequency, w is made with these values. Once this is done a best fit line is

applied to the data points which will give you the characterization relation

$$PSD \propto \frac{1}{f^\beta}$$

. Based on this we can then classify the noise as a color of noise.
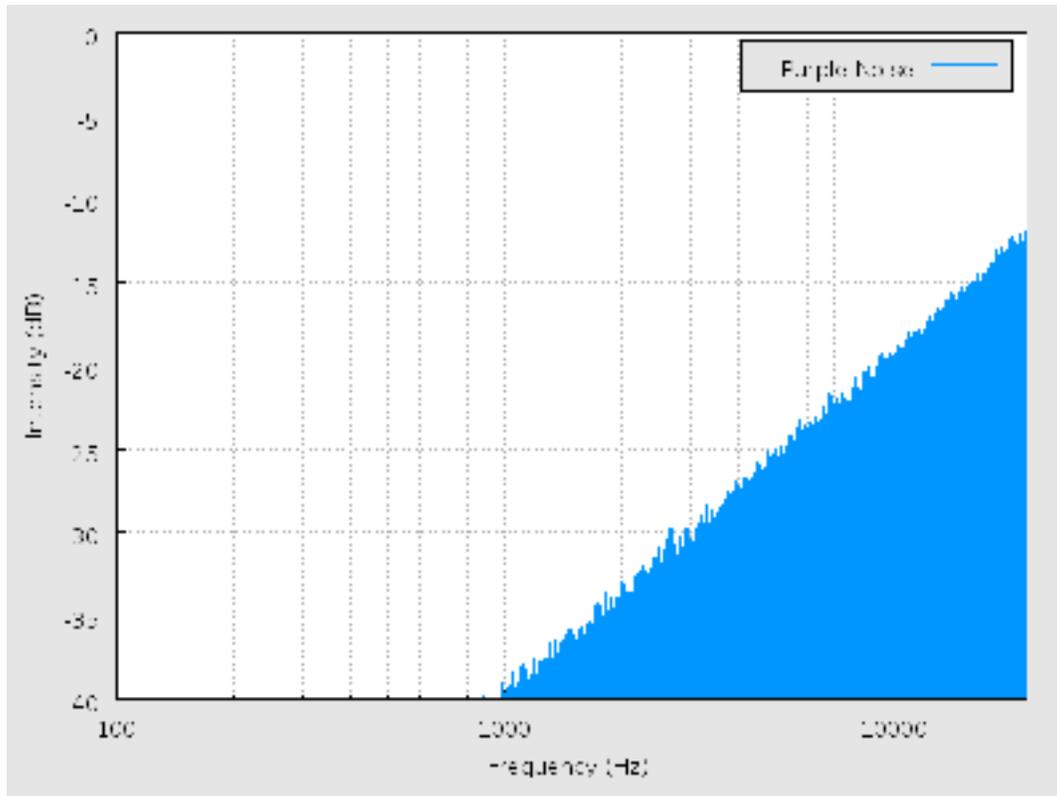
THE REVERSE PROCESS

Knowing how to convert data to a color of noise is only half the problem. What if we know what type of noise is possible and we need data from it for a given process? Knowing the noise color means that we know the power spectral density relation to the frequency. From here onwards we follow the reverse route as that taken to get to the noise color by using the inverse Fourier transform instead.

Inverse Fourier Transform

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(w)e^{jwt} \, dt$$

Where;

- $x(t)$ *is the equation of the curve to produce future the data.*

- e^{jwt} *is the exponential form of writing the relation* $cos(wt) + jsin(wt)$ *. The j in the second term here indicates that that term is imaginary.*

- w *is the frequency*

We can either use it to produce a curve or, by making a similar simplification as we did for the Fourier transform, to generate data directly. This reverse process should be trivial to someone who worked through the forward process.

5.4 Pops, Snaps & Crackles

The second category of noise is non-frequency based noise. Three examples of this non-frequency noise are pops, snaps and crackles. The study of these noise types is fairly new, and not a lot is known about how to deal with them when they arise in our research or studies. However, one can successfully distinguish and classify these noise types by noticing certain properties which are typical of them.

5.4.1 Pops

At one end of the extreme of non-frequency noise is what is defined as pops. Pops are infrequent random spikes in noise of approximately the same amplitude. An example of this would be a loose wire that is usually completing the circuit. However, it randomly disconnects for a split second and then recompletes the circuit for normal operation. Chemical engineering processes may inherently have pops that are unpredictable. Safety features can be added to a process, such as pressure relief valves, to handle the pops when they do occur.

The following is an image of what pops might look like in a system:
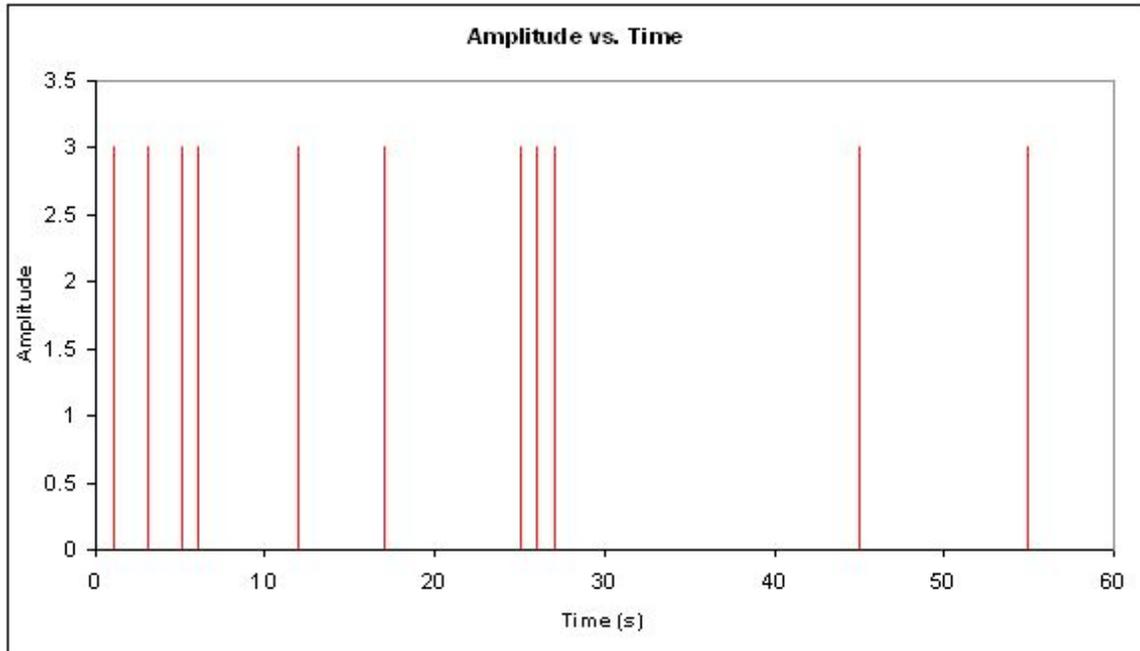
5.4.2 Snaps

On the other end of the non-frequency noise spectrum there are snaps. Snaps are single independent events that occur only once. When a pencil is bent and it breaks under the force is an example of snapping. Another example would be the snapping of a piece of chalk or exploding a bomb. The "popping" of one's knuckles is really snapping since it is an independent event, unless you snap all your knuckles all at once or one after the other really fast. Just like pops, snaps are unpredictable, and safety features should be added to the system to handle the snaps.

The following is an example of a snap:
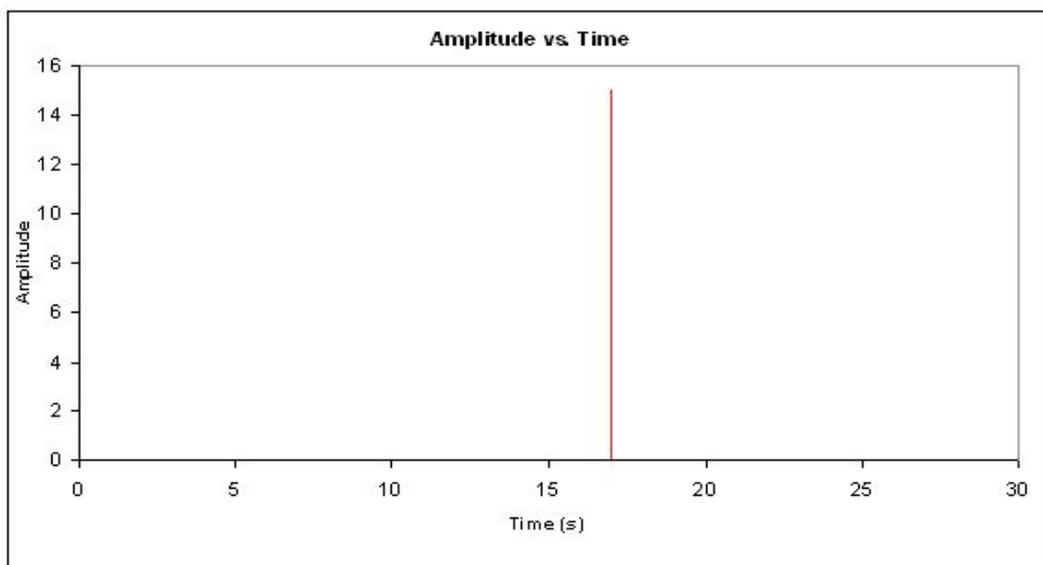
5.4.3 Crackles

In between popping and snapping there is crackling. A very common example for crackling is the sound heard coming from a burning piece of wood. Like popping, there is a non-frequency or irregularity in which the crackles occur. In addition, there is also an irregularity of the amplitude of the crackle. In the case of the fire, not only can you not predict when the crackle sound would be heard, but you cannot predict how loud it will be either. Furthermore, there is a universality condition associated with crackling. Regardless of the scale, the similar randomness in repetition and amplitude should be observed.

In dealing with this universality condition the concept of a critical exponent arises. For example, if we are looking at the same crackling effect, S, over a larger period of time the two would have to be equal after scaling the larger one.

$$< S >_{small} (T) = A < S >_{large} \left(\frac{T}{B}\right)$$ for some scaling factors A and B.

Where;

$< S >$ is S average.

$$B = \left(\frac{1}{(1 - \delta)}\right)$$

If the time scale is expanded by a small factor then the rescaling of the size will also be small, say $1 + a\delta.$

$$< S > (T) = (1 + a\delta) * < S > (1 + \delta T)$$

For small δ we have $$a < S >= T \frac{d < S >}{dT}$$

Solving gives

$$< S > (T) = S_o T^a$$

The exponent a is called the critical exponent and is a universal prediction of a given theory.

There are other concepts which are used as a check for universality for times when the critical exponent cannot be used but the most common one is the critical exponent.

The following is an example of what a crackle might look like in a system:
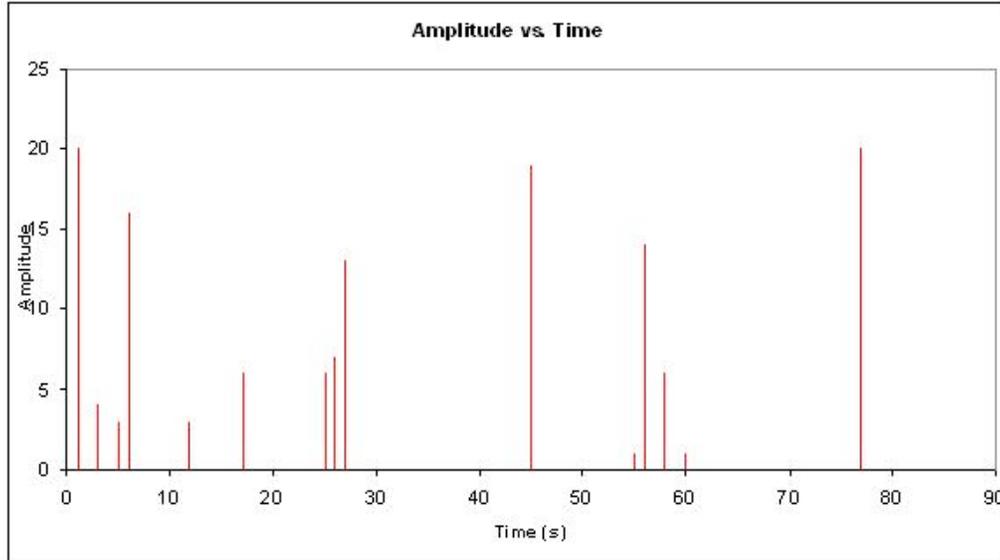
5.5 Summary

Noise is all around us and is an unavoidable characteristic of monitoring signals. When noise is frequency based, like the colors of noise, it can be characterized and modeled using mathematical tools. When noise is not frequency based, like pops, snaps, and crackles, it can be recognized, but ways to model it are still being devised. Characterizing and modeling noise is important for chemical engineers to accurately handle their data, control processes, and predict future trends.

5.6 Crackle Example

'*Note that this is similar to but importantly not a real life example. The data is made up to demonstrate the principle of Crackling*'

A lead engineer monitoring the instantaneous flow rate of a coolant used to cool an exothermic reactor gathers data set 'a' over a 30hr period. Later on that week he gathers data set 'b' for the same coolant flow rate. For both sets of data he determined that the way the instantaneous coolant flow rate is reacting to the exothermic reactor is optimal to the reaction process. For his records he would like to represent the noise in the data in a compact form. He wants you to first characterize the noise and then provide any other information, which will generalize the trend in the data. Once you have completed these tasks, report back to him.

Time (hr)	Flow Rate (m^3/s)				
0	0	11	0	21	0
1	0	12	0	22	0
2	0	13	10	23	0
3	0	14	0	24	0
4	0	15	0	25	3
5	2	16	0	26	5.1
6	3.5	17	0	27	0
7	0	18	7.4	28	0
8	0	19	12.8	29	0
9	4	20	0	30	0
10	0				

Data set 'a'

Note that between 0 and 30hr for all intervals of time not represented in this table there is a zero assigned as the instantaneous flow rate.

Time (hr)	Flow Rate (m^3/s)				
0	0	16	0	32	0
1	0	17	0	33	0
2	0	18	5.5	34	0
3	0	19	14	35	0
4	0	20	0	36	0
5	3	21	0	37	0
6	4.5	22	0	38	6
7	0	23	0	39	16
8	0	24	0	40	0
9	4	25	3	41	8
10	0	26	7	42	3
11	0	27	16	43	5
12	0	28	0	44	0
13	8	29	0	45	0
14	0	30	0	46	0
15	0	31	5	47	0
				48	0
				49	0
				50	0

Data set 'b'

Note that between 0 and 50hr for all intervals of time not represented in this table there is a zero assigned as the instantaneous flow rate.

——Solution——

Each set of data is plotted, Flow Rate vs. Time. The two graphs show crackling trends.
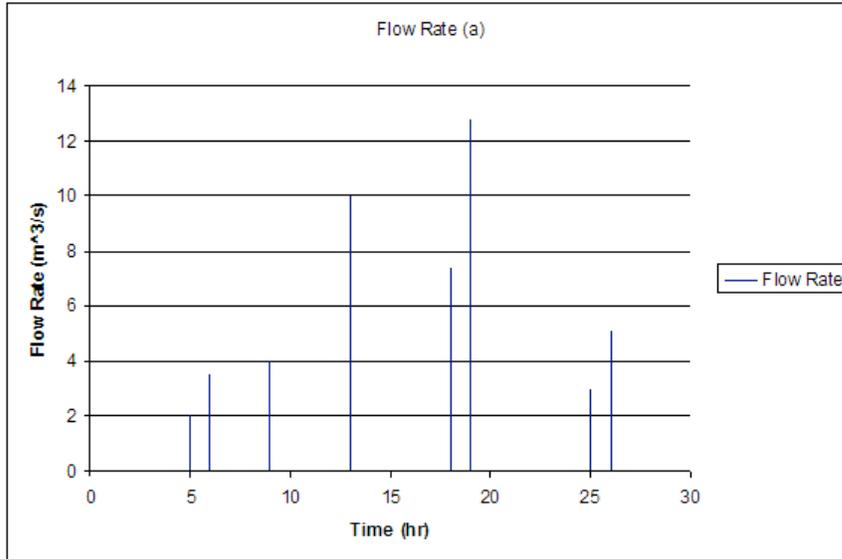
Graph for data set 'a'
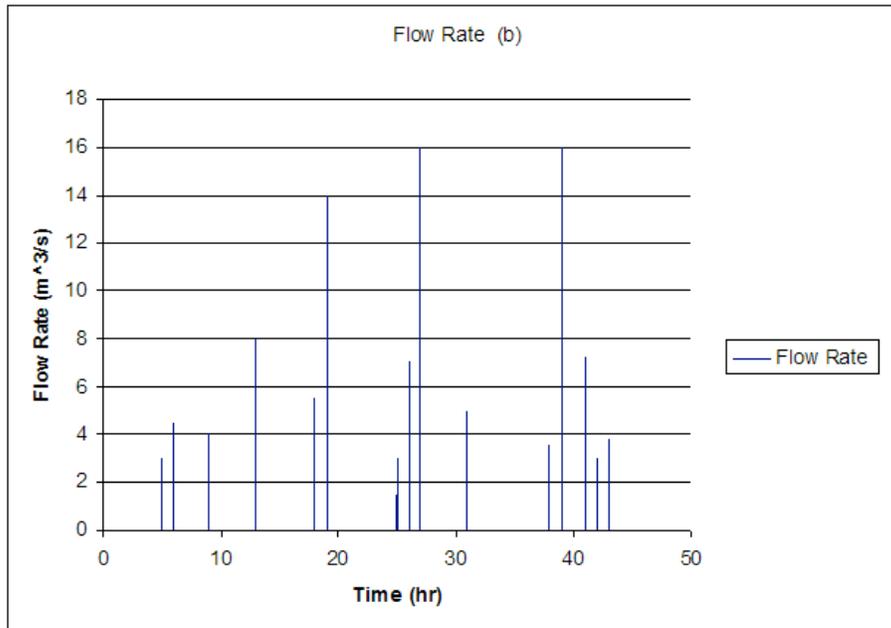
Graph for data set 'b'

Upon observing this, the general noise trend is crackling. To provide a final proof we apply the universality relation $<S>(T) = S_o \, T^a$.

We first determine the critical exponent, a, for data set a.

We see that;

$$<S> = 1.54$$

$$S_0 = 2$$ this value can be the first value attained or an average of the first 2 or 3 values.

$$T = 30$$

Solving for the critical exponent using the relation given above we get;

$$a = -0.08$$

Once again we carry out the same calculation for data set 'b'.

$$<S> = 2.12$$

$$S_0 = 3$$ this value can be the first value attained or an average of the first 2 or 3 values.

$$T = 50$$

Solving for the critical exponent using the relation given above we get;

$$a = -0.08$$

The similarity of the two critical exponent gives further proof that the data the instantaneous flow rate of the coolant over a period of time will show crackling.

5.7 Colors of Noise Example

A chemical engineer is reading flow rates from flow meter. Every 0.1 day for 8 days a reading was taken and the data is given here: Colors of Noise Example. The data displays the fluctuations from the set flow rate of 3000 liters per hour at a wastewater treatment plant. The specifications for the plant say that the max flow rate is 6000 liters per hour, or the pipes will burst. Also, the flow rate cannot fall below 200 liters per hour, or the system will automatically shut down.

The chemical engineer notices that there were some readings close to these limits. It is the chemical engineer's job to determine if the readings are accurate flow rates or if there is an error with the flow meter. By characterizing the type of noise, the chemical engineer can determine the source of the noise and so take appropriate preventative measures. What type of noise is present and what protective measures should be taken?

——Solution——

1) Plot data. From the data presented, a flow rate vs. time chart was graphed to gauge degree of fluctuation in the flow rate.
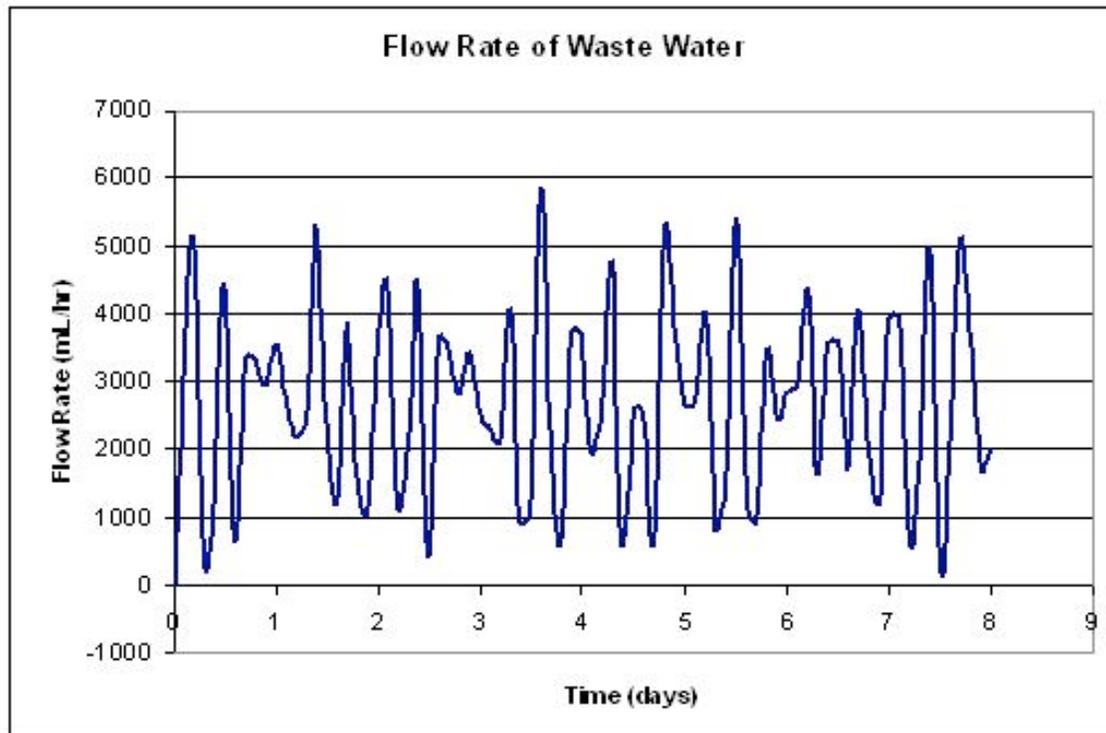
2) Calculate the power spectral density data. Using the simplified integral derived to calculate the power spectral density, a table was created with the resulting PSD values at varying frequencies. These frequencies were self defined and were chosen to encompass a broad range. More detailed calculations can be found here: Colors of Noise Example under the worksheet titled "PSD CALC".

3) Plot the power spectral density. The power spectral density for each frequency was then plotted against the frequency, creating the power spectral density plot shown below.

108

PSD for Waste Water Plant

$y = -0.023x + 14874$

$R^2 = 0.0982$
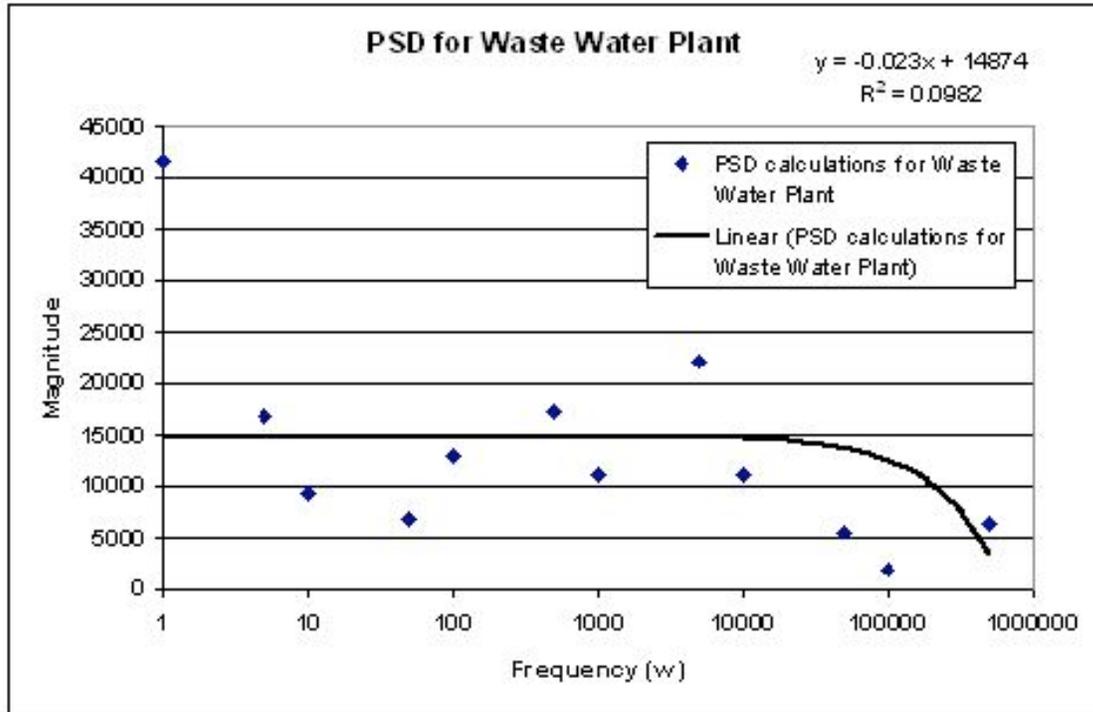
4) Characterize the noise. To determine the β value for the data, a linear trend line was taken on the data. This trend line can be seen in the power spectral density plot above. The slope of this trend line is the β value for the data. In this case, the β=0.023. Since this value is not that of white noise (β=0) nor that of pink noise (β=1), we can say that this noise is somewhere between white and pink noise.

5) Determine the source of noise.

There are two possible major sources of noise. They are the liquid motion in the pipe and noise in the flow meter caused by itself or outside sources. Since it is found earlier that β=0.023, the source of noise is probably from the flow meter. It is *not* from the motion of liquid in the pipe because liquid motion tends to produce brown noise (β=2).

6) Protective measures that can be taken. Knowing this, a correction can be made on the calculations to determine the actual flow rate. A full step by step solution can also be found here: Colors of Noise Example under the worksheet titled "SOLUTION".

Note: The data used is fabricated from a complex sine equation and was not actually collected.

5.8 References

- Kosko, Bart. "White Noise Ain't So White", Noise. ISBN 0670034959
- Ziemer, Rodger E. Elements of Engineering Probability and Statistics, New Jersey: Prentice Hall. ISBN 0024316202

- Papoulis, Athanasios. Probability, Random Variables, and Stochastic Processes, New York: McGraw – Hill Book Company. ISBN 0071199810
- Peebles, Peyton Z. Jr. Probability, Random Variables, and Random Signal Principles, New York: McGraw – Hill, Inc. ISBN 0071181814
- Sethna, James P. Crackling Noise, Nature 410, 242-250 (2001). Also available here.
- http://en.wikipedia.org

Section 6. Numerical ODE solving in Excel: Euler's method, Runge Kutta, Dead time in ODE solving

Note: ▣Video lecture available for this section!

Authors: (Presented: 9/8/06 /Date Revised: 9/19/06) Aaron Bennick, Bradley Anderson, Michael Salciccioli

Stewards: (9/5/07) Sean Gant, Jay Lee, Lance Dehne, Kelly Martin

- First round reviews for this page
- Rebuttal for this page

6.1 Introduction

This article focuses on the modeling of ordinary differential equations (ODEs) of the form:

$$\frac{dy}{dx} = f(x, y)$$

In creating a model, a new value y_{i+1} is generated using the old or initial value y_i, the slope estimate ϕ, and the step size h. This general formula can be applied in a stepwise fashion to model the solution. All stepwise models will take the following general form:

$$y_{i+1} = y_i + \phi h$$

The modeling methods discussed in this article are Euler's method and the Runge-Kutta methods. The difference between the two methods is the way in which the slope ϕ is estimated.

Choosing The Right Model and Step Size

The proper numerical modeling method heavily depends on the situation, the available resources, and the desired accuracy of the result. If only a quick estimate of a differential equation is required, the Euler method may provide the simplest solution. If much higher accuracy is required, a fifth-order Runge-Kutta method may be used. Engineers today, with the aid of computers and excel, should be capable of quickly and accurately estimating the solution to ODEs using higher-order Runge-Kutta methods.

In all numerical models, as the step size is decreased, the accuracy of the model is increased. The tradeoff here is that smaller step sizes require more computation and therefore increase the amount of time to obtain a solution. A balance between desired

accuracy and time required for producing an answer can be achieved by selecting an appropriate step size. One suggested algorithm for selecting a suitable step size is to produce models using two different methods (possibly a second and third order Runge-Kutta). The steps size can then be systematically cut in half until the difference between both models is acceptably small (effectively creating an error tolerance).

6.2 Euler's Method

Euler's method is a simple one-step method used for solving ODEs. In Euler's method, the slope, ϕ, is estimated in the most basic manner by using the first derivative at x_i. This gives a direct estimate, and Euler's method takes the form of

$$y_{i+1} = y_i + f(x_i, y_i)h$$

For demonstration, we will use the basic differential equation

$$\frac{dy}{dx} = 3x^2 + 2x + 1$$

with the initial condition $y(0) = 1$. If a step size, h, is taken to be 0.5 over the interval 0 to 2, the solutions can be calculated as follows:

x	y_Euler	y_Actual	Error
y(0)	1	1	0
y(0.5)	1 + [3(0)^2 + 2(0) + 1](0.5) = 1.5	1.875	0.375
y(1)	1.5 + [3(0.5)^2 + 2(0.5) + 1](0.5) = 2.875	4	1.125
y(1.5)	2.875 + [3(1)^2 + 2(1) + 1](0.5) = 5.875	8.125	3.25
y(2.0)	5.875 + [3(1.5)^2 + 2(1.5) + 1](0.5) = 11.25	15	3.75

The y_actual values in this table were calculated by directly integrating the differential equation, giving the exact solution as: $y = x^3 + x^2 + x + 1$

6.2.1 Use of Microsoft Excel

Calculating an ODE solution by hand with Euler's method can be a very tedious process. Fortunately, this process is greatly simplified through the use of Microsoft Excel. Creating a spreadsheet similar to the one above, where the x values are specified and the y_Euler values are recursively calculated from the previous value, makes the calculation rather simple. Any step size and interval can be used.

6.2.2 Drawbacks of Euler's Method

There are several drawbacks to using Euler's method for solving ODEs that must be kept in mind. First of all, this method does not work well on stiff ODEs. A stiff ODE is a differential equation whose solutions are numerically unstable when solved with certain numerical methods. For stiff equations - which are frequently encountered in modeling chemical kinetics - explicit methods like Euler's are usually quite inefficient because the region of stability is so small that the step size must be extremely small to get any accuracy. In a case like this, an implicit method, such as the backwards Euler method, yields a more accurate solution. These implicit methods require more work per step, but the stability region is larger. This allows for a larger step size, making the overall process more efficient than an explicit method. A second drawback to using Euler's Method is that error is introduced into the solution. The error associated with the simple example above is shown in the last column. This error can be seen visually in the graph below.
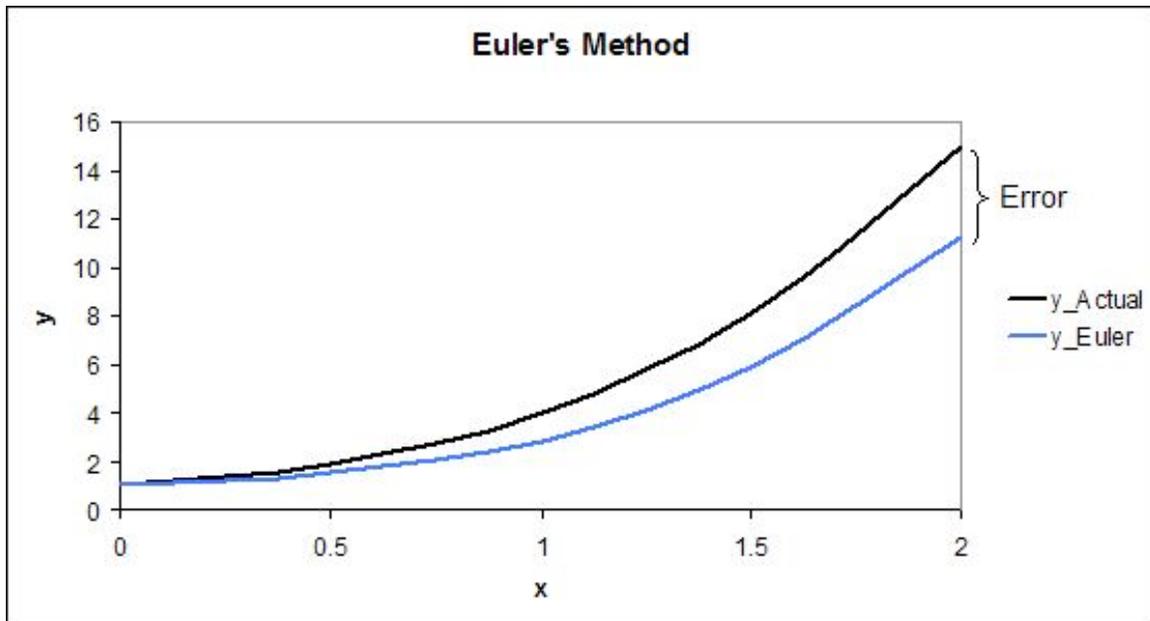
It can be seen that the two values are identical at the initial condition of $y(0)=1$, and then the error increases as the x value increases and the error propagates through the solution to $x = 2$. The error can be decreased by choosing a smaller step size, which can be done quite easily in Excel, or by opting to solve the ODE with the more accurate Runge-Kutta method.

6.3 Runge-Kutta Methods

The Runge-Kutta method for modeling differential equations builds upon the Euler method to achieve a greater accuracy. Multiple derivative estimates are made and, depending on the specific form of the model, are combined in a weighted average over the step interval. The order of the Runge-Kutta method can range from second to higher, depending on the amount of derivative estimates made. The second-order Runge-Kutta

method labeled Heun's technique estimates derivatives by averaging endpoint measurements of the step size along a function. This averaged value is used as the slope estimate for x_{i+1}. Third and higher power Runge-Kutta methods make mid-point derivative estimations, and deliver a weighted average for the end point derivative at x_{i+1}. As the Runge-Kutta order increases, so does the accuracy of the model.

The general form of the Runge-Kutta method is

$$y_{i+1} = y_i + \phi(x_i, y_i, h)h$$

Where $\phi(x\_i,y\_i,h)$ now represents a weighted average slope over the interval h.

$$\phi(x_i, y_i, h) = a_1k_1 + a_2k_2 + ... + a_nK_n$$

where the a's are constants and the k's are

$$k_1 = f(x_i, y_i)$$

$$k_2 = f(x_i + p_1h, y_i + q_{11}k_1h)$$

$$k_3 = f(x_i + p_2h, y_i + q_{21}k_1h + q_{22}k_2h)$$

$$k_n = f(x_i + p_{n-1}h, y_i + q_{n-1,1}k_1h + q_{n-1,2}k_2h + ... + q_{n-1,n-1}k_{n-1}h)$$

The constants a, p, and q are solved for with the use of Taylor series expansions once n is specified (see bottom of page for derivation). The resulting set of equations have one or more degrees of freedom. This means that for every order of Runge-Kutta method, there is a family of methods. Below are some of the more common Runge-Kutta choices.

6.3.1 Second-Order Runge-Kutta Methods (n = 2)

Every second order method described here will produce exactly the same result if the modeled differential equation is constant, linear, or quadratic. Because this is typically not the case, and the differential equation is often more complicated, one method may be more suitable than another.

Heun's Technique

The second-order Runge-Kutta method with one iteration of the slope estimate , also

$$a_1 = a_2 = \frac{1}{2}$$

known as Heun's technique, sets the constants and

$p_1 = q_{11} = 1$. Huen determined that defining a_1 and a_2 as $\frac{1}{2}$ will take the average of the slopes of the tangent lines at either end of the desired interval, accounting for the concavity of the function, creating a more accurate result. When substituted into the general form, we find

$$y_{i+1} = y_i + \left(\frac{1}{2}k_1 + \frac{1}{2}k_2\right)h$$

with k_1 and k_2 defined as

$$k_1 = f(x_i, y_i)$$

$$k_2 = f(x_i + h, y_i + hk_1)$$

For demonstration of this second-order Runge-Kutta method, we will use the same basic differential equation $\frac{dy}{dx} = 3x^2 + 2x + 1$ with the initial condition $y(0) = 1$. If a step size, h, is taken to be 0.5 over the interval 0 to 2, the solutions can be calculated as follows:

x	k_1	k_2	y_Heun	y_Actual	Error
y(0)	3(0)^2 + 2(0) + 1 = 1	3(0.5)^2 + 2(0.5) + 1 = 2.75	1	1	0
y(0.5)	3(0.5)^2 + 2(0.5) + 1 = 2.75	3(1)^2 + 2(1) + 1 = 6	1 + [0.5(1) + 0.5(2.75)](0.5) = 1.9375	1.875	-0.0625
y(1)	3(1)^2 + 2(1) + 1 = 6	3(1.5)^2 + 2(1.5) + 1 = 10.75	2.375 + [0.5(2.75) + 0.5(6)](0.5) = 4.125	4	-0.125
y(1.5)	3(1.5)^2 + 2(1.5) + 1 = 10.75	3(2)^2 + 2(2) + 1 = 17	5.375 + [0.5(6) + 0.5(10.75)](0.5) = 8.3125	8.125	-0.1875
y(2.0)	3(2)^2 + 2(2) + 1 = 17	3(2.5)^2 + 2(2.5) + 1 = 24.75	10.75 + [0.5(10.75) + 0.5(17)](0.5) = 15.25	15	-0.25

When compared to the Euler method demonstration above, it can be seen that the second-order Runge-Kutta Heun's Technique requires a significant increase in effort in order to produce values, but also produces a significant reduction in error. Following Runge-Kutta methods can be worked through a similar manner, adding columns for additional k values. Below is a graphical description of how slope is estimated using both Euler's method, and Heun's technique. Observe the increase in accuracy when an average slope across an interval of 0.5 is used instead of just an initial estimate.
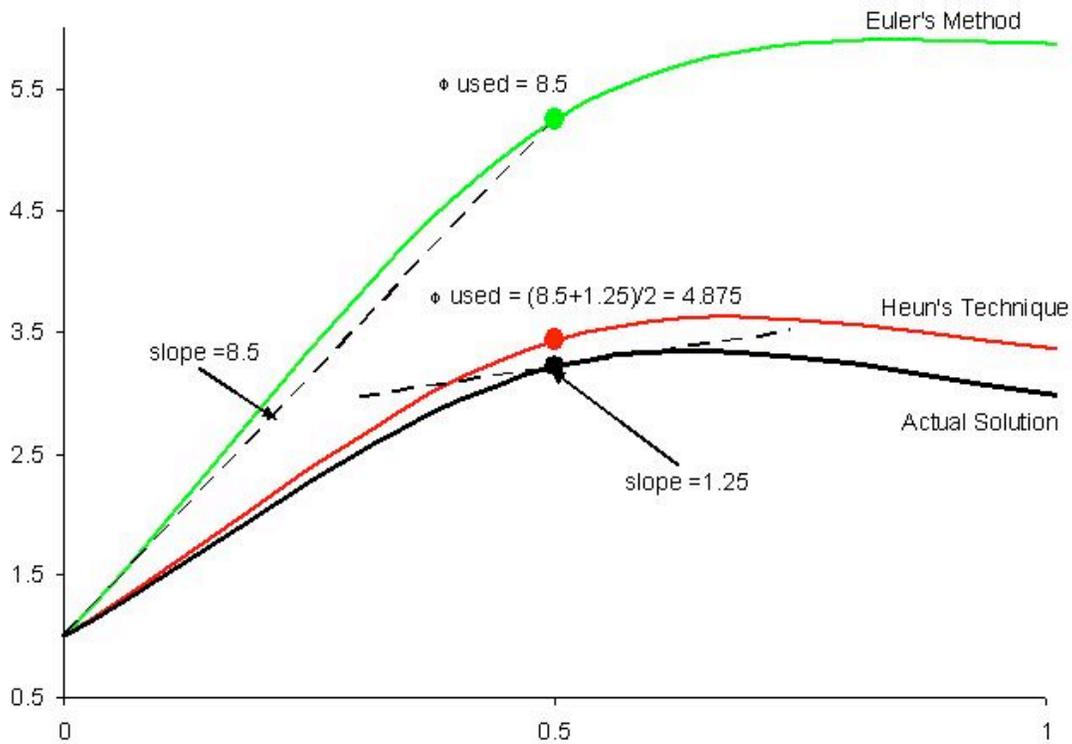
Improved Polygon Method

Using the improved polygon method, a_2 is taken to be 1, a_1 as 0, and therefore

$$p_1 = q_{11} = \frac{1}{2}$$

. The general form then becomes

$$y_{i+1} = y_i + k_2 h$$

with k_1 and k_2 defined as

$$k_1 = f(x_i, y_i)$$

116

$$k_2 = f(x_i + \frac{1}{2}h, y_i + \frac{1}{2}hk_1)$$

Ralston's Method

The Ralston method takes a_2 to be $\frac{2}{3}$. Therefore $a_1 = \frac{1}{3}$ and $p_1 = q_{11} = \frac{3}{4}$. It has been determined by Ralston (1962) and Ralston and Rabinowitz (1978) that defining a_2 as $\frac{2}{3}$ will minimize truncation error in second-order Runge-Kutta methods. The general form becomes

$$y_{i+1} = y_i + (\frac{1}{3}k_1 + \frac{2}{3}k_2)h$$

with k_1 and k_2 defined as

$$k_1 = f(x_i, y_i)$$

$$k_2 = f(x_i + \frac{3}{4}h, y_i + \frac{3}{4}hk_1)$$

6.3.2 Third-Order Runge-Kutta Methods (n = 3)

The third-order Runge-Kutta methods, when derived, produce a family of equations to solve for constants with two degrees of freedom. This means an even more variable family of third-order Runge-Kutta methods can be produced. A commonly used general third-order form is

$$y_{i+1} = y_i + [\frac{1}{6}(k_1 + 4k_2 + k_3)]h$$

with

$$k_1 = f(x_i, y_i)$$

$$k_2 = f(x_i + \frac{1}{2}h, y_i + \frac{1}{2}hk_1)$$

$$k_3 = f(x_i + h, y_i - hk_1 + 2hk_2)$$

6.3.3 Fourth-Order Runge-Kutta Methods (*n* = 4)

The family of fourth-order Runge-Kutta methods have three degrees of freedmon and therefore infinite variability just as the second and third order methods do. The fourth-order versions are most favored among all the Runge-Kutta methods. What is know as the *classical fourth-order Runge-Kutta method* is

$$y_{i+1} = y_i + [\frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)]h$$

with

$$k_1 = f(x_i, y_i)$$

$$k_2 = f(x_i + \frac{1}{2}h, y_i + \frac{1}{2}hk_1)$$

$$k_3 = f(x_i + \frac{1}{2}h, y_i + \frac{1}{2}hk_2)$$

$$k_4 = f(x_i + h, y_i + hk_3)$$

6.3.4 Fifth-Order Runge-Kutta Methods (*n* = 5)

Where very accurate results are required, the fifth-order Runge-Kutta *Butcher's* (1964) *fifth-order RK method* should be employed:

$$y_{i+1} = y_i + [\frac{1}{90}(7k_1 + 32k_3 + 12k_4 + 32k_5 + 7k_6)]h$$

with

$$k_1 = f(x_i, y_i)$$

118

$$k_2 = f(x_i + \frac{1}{4}h, y_i + \frac{1}{4}hk_1)$$

$$k_3 = f(x_i + \frac{1}{4}h, y_i + \frac{1}{8}hk_1 + \frac{1}{8}hk_2)$$

$$k_4 = f(x_i + \frac{1}{2}h, y_i - \frac{1}{2}hk_2 + hk_3)$$

$$k_5 = f(x_i + \frac{3}{4}h, y_i - \frac{3}{16}hk_1 + \frac{9}{16}hk_4)$$

$$k_6 = f(x_i + h, y_i - \frac{3}{7}hk_1 + \frac{2}{7}hk_2 + \frac{12}{7}hk_3 - \frac{12}{7}hk_4 + \frac{8}{7}hk_5)$$

The integration of *k's* within other *k* values suggests the use of a spreadsheet. As with all Runge-Kutta methods, the calculation of values for the fifth-order version would be greatly assisted through the use of Microsoft Excel.

6.4 Interactive Model Comparison in Excel

To give a better understanding of the impact between different Runge-Kutta methods, as well as the impact of step size, the interactive excel sheet below will allow you to enter the step size h into a set of models and observe how the models contour to match an example differential equation solution.

Observe the relationship between model type, step size, and relative error.

[ODE model comparison interactive spreadsheet](#)

6.5 Dead Time

Dead time, or delay differential equation, occurs when there is a delay or lag in the process of a real life function that is being modeled. This means that the numerical model is not accurate until the delay is over. This concept can come into play for the start up of a reaction process. For example, at the start of a reaction in a CSTR (Continuous Stirred Tank Reactor), there will be reagents at the top of the reactor that have started the reaction, but it will take a given time for these reactant/products to be discharged from the reactor. If one was modeling the concentration of reagents vs. time, time t=0 would have started when the tank was filled, but the concentrations being read would not follow a standard model equation until the residence time was completed and the reactor was in

continuous operational mode. The dead time is the time it would take for the readings to start meeting a theoretical equation, or the time it takes for the reactor to be cleared once of the original reagents. Dead time can be determined experimentally and then inserted into modeling equations.

The solution of such a delay differential equation becomes problematic because in order to solve the equation, information from past times (during the delay) is needed in addition to the current time. For example if t_o is the lag time for the given scenario, then the value to use in the ODE becomes $(t-t_o)$ instead of t. When modeling this in excel, $(x-t_o)$ is substituted in for the x value. With this substitution, Euler's method can be used again in the same way to approximate the solution. This substitution will be carried into the differential equation so that the new dead time solution can be approximated using Euler's method. To model the reactor before the initial deadtime is completed, piece-wise functions are often used. This way for the dead time, a given model is used not characteristic of the reactor at normal operating conditions, then once the dead time is completed the modeling equation is taken into effect.

The link below will help to show how to include dead time in a numerical method approximation such as Euler's method. As seen in the excel file, the dead time that is specified by the user in the yellow box will change the delay in the model. The more dead time, the further shifted from the theoretical equation the new model is. To take dead time into account in excel, the x value is simply substituted out for (x-t) where t is equal to the dead time. If you look at the equations entered in the Y cells, you will see that the x value inserted into the differential equation is (x-t), where t is the user specified dead time. It can be seen through this example spreadsheet that the effect of dead time is a simple horizontal shift in the model equation.

Dead Time Interactive Spreadsheet

6.6 Error

There are two types of error associated with solving ODEs using stepwise approximation methods in Excel. These errors are also present using other methods or computer programs to solve ODEs. The first, discretization, is the result of the estimated y value that is inherent of using a numerical method to approximate a solution. Discretization errors, also called truncation, occur proportionately over a single step size. Truncation error will propagate over extended results because the approximation from previous steps is used to approximate the next. In essence, a copy of a copy is being made. The accuracy of the next point is a direct result of the accuracy of the previous. Just as the quality of a second copy is dependant on the quality of the first. This error can be reduced by reducing the step size.

Please see ODE model comparison interactive spreadsheet to better learn how step sizes can influence error.

120

The second types of errors are rounding errors. These errors are dependent on the computer's capacity for retaining significant digits. The more significant digits that a computer can hold, the smaller the rounding error will be.

6.6.1 Estimating Error in Euler's Method

To mathematically represent the error associated with Euler's method, it is first helpful to make a comparison to an infinite Taylor series expansion of the term y_{i+1}. The Taylor series expansion of this term is

$$y_{i+1} = y_i + f(x_i, y_i)h + f'(x_i, y_i)\frac{h^2}{2} + f''(x_i, y_i)\frac{h^3}{3} + ... + f^n(x_i, y_i)\frac{h^n}{!n}$$

When this expansion is compared to the general form of Euler's method it can be seen that Euler's method lacks every term beyond $f(x_i, y_i)h$. These missing terms, the difference between the Euler approximation and an infinite Taylor series (taken to be the true solution), is the error in the Euler approximation. Mathematically respresenting the error in higher order Runge-kutta methods is done in a similar fashion.

6.6.2 Estimating and Minimizing Error in Runge Kutta Method

Uniform time steps are good for some cases but not always. Sometimes we deal with problems where varying time steps makes sense. When should you change the step size? If we have an nth order scheme and and (n+1)th order scheme, we can take the difference between these two to be the error in the scheme, and make the step size smaller if we prefer a smaller error, or larger if we can tolerate a larger error. This is fairly simple with Runge Kutta, because we can take a fifth order method and a fourth order method using the same k's. Only a little extra work at each step.

Another way of estimating error in the Runge-Kutta method is to reverse directions at each step of the advancing solution and recompute the previous ordinate. By considering the difference between the newly computed previous ordinate and the originally computed value, you can determine an estimate for the truncation error incurred in advancing the solution over that step. This is a bit more tedious, but does give a good estimate of truncation error.

6.7 Euler's Method for Systems of ODEs

In chemical engineering and other related fields, having a method for solving a differential equation is simply not enough. Many real world problems require simultaneously solving systems of ODEs. For problems like these, any of the numerical methods described in this article will still work. The only difference is that for n ODEs, n initial values of y are needed for the initial x value. Then the numerical method of choice is just applied to every equation in each step before going on to the next. This further complicates the step-by-step problem solving methodology, and would require the use of

Excel in nearly every application. The ODE solved with Euler's method as an example before is now expanded to include a system of two ODEs below:

$$\frac{dy_1}{dx} = 3x^2 + 2x + 1, y_1(0) = 1$$

$$\frac{dy_2}{dx} = 4y_1 + x, y_2(0) = 2$$

Step size is again 0.5, over an interval 0-2.

x	y_1Euler	y_2Euler	y_1Actual	y_2Actual	y_1 Error	y_2 Error
y(0)	1	2	1	2	0	0
y(0.5)	1 + [3(0.5)^2 + 2(0.5) + 1](0.5) = 2.375	2 + [4(1) + 0.5](0.5) = 4.25	1.875	5.875	-0.5	1.625
y(1)	2.375 + [3(1)^2 + 2(1) + 1](0.5) = 5.375	4.25 + [4(2.375) + 1](0.5) = 9.5	4	18.5	-1.375	9
y(1.5)	5.375 + [3(1.5)^2 + 2(1.5) + 1](0.5) = 10.75	9.5 + [4(5.375) + 1.5](0.5) = 21	8.125	51.875	-2.625	30.875
y(2.0)	10.75 + [3(2)^2 + 2(2) + 1](0.5) = 19.25	21 + [4(10.75) + 2](0.5) = 43.5	15	124	-4.25	80.5

It should be observed that there is more error in the Euler approximation of the second ODE solution. This is because the equation also has y_1 in it. So there is the error introduced by using the Euler approximation to solve the 2nd ODE, as well as the error from the Euler approximation used to find y_1 in the 1st ODE in the same step!

Exact solutions were again obtained from directly integrating the ODEs:

$$y_1 = x^3 + x^2 + x + 1 \quad \text{and} \quad y_2 = 4y_1 x + \frac{x^2}{2} + 2$$

The elementary liquid-phase reaction A --> B is to be carried out in an isothermal, isobaric PFR at 30 degrees C. The feed enters at a concentration of 0.25 mol/L and at a rate of 3 mol/min. The reaction constant is known experimentally to be 0.01 min^{-1} at this temperature. You have been given the task of building a reactor that will be used to carry out this reaction. Using Euler's method with a step size of 0.05, determine how large the reactor must be if a conversion of 80% is desired.

Simplified Design Equation:
$$\frac{dV}{dX} = \frac{F_{A0}}{kC_{A0}(1-X)}$$

Lumping the given flow, concentration, and reaction constant together gives:

$$\frac{dV}{dX} = 1200\frac{1}{1-X}$$

Since no volume is required for a conversion of zero, the initial condition needed is V(0)=0.

Now the information simply has to be entered into Excel. A column for the conversion, X, going from 0 to 0.8 in 0.05 increments is used for the step size, and the first value, V(0), is known to be zero. To get the volume, simply add the previous volume to the constants multiplied by the step size and 1/(1-X), or:

$$V_{i+1} = V_i + 1200 * 0.05 * \frac{1}{1-X}$$

Copying this formula down the column to the final conversion value of 0.8, gives the results shown in the table below:

X	V
V(0)	0
V(0.05)	63
V(0.1)	130
V(0.15)	200
V(0.2)	275
V(0.25)	355
V(0.3)	441

V(0.35)	533
V(0.4)	633
V(0.45)	743
V(0.5)	863
V(0.55)	996
V(0.6)	1146
V(0.65)	1317
V(0.7)	1517
V(0.75)	1757
V(0.8)	2057

The final reactor volume obtained is approximately 2057 L. This compares reasonably well to the exact value of 1931 L. The Excel file used to obtain this solution, along with the exact solution, can be downloaded below.

Example 1

6.9 Second-order Runge-Kutta derivation

The following example will take you step by step through the derivation of the second-order Runge-Kutta methods. Setting $n = 2$ results in a general form of

$$y_{i+1} = y_i + (a_1 k_1 + a_2 k_2)h$$

with k_1 and k_2 defined as

$$k_1 = f(x_i, y_i)$$

$$k_2 = f(x_i + p_1 h, y_i + q_{11} k_1 h)$$

The constants in the general form must be defined. To do this we will employ a second-order Taylor series expansion for y_{i+1} in terms of y_i and $f(x_i, y_i)$. This Taylor series is

$$y_{i+1} = y_i + f(x_i, y_i)h + f'(x_i, y_i)\frac{h^2}{2}$$

124

Expanding $f'(x_i, y_i)$ with the chain rule, and substituting it back into the previous Taylor series expansion gives

$$y_{i+1} = y_i + f(x_i, y_i)h + \left(\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial x}\right)\frac{h^2}{2}$$

The next step is to apply a Taylor series expansion to the k_2 equation. The applied Taylor series expansion rule is $g(x+r, y+s) = g(x,y) + r\frac{\partial g}{\partial x} + s\frac{\partial g}{\partial y} + ...$, and the resulting equation is

$$f(x_i + p_1 h, y_i + q_{11}k_1 h) = f(x_i, y_i) + p_1 h\frac{\partial f}{\partial x} + q_{11}k_1 h\frac{\partial f}{\partial y} + O(h^2)$$

where $O(h^2)$ is a measure of the truncation error between model and true solution.

When this Taylor series expansion result of the k_2 equation, along with the k_1 equation, is substituted into the general, and a grouping of like terms is performed, the following results:

$$y_{i+1} = y_i + [a_1 f(x_i, y_i) + a_2 f(x_i, y_i)]h + [a_2 p_1 \frac{\partial f}{\partial x} + a_2 q_{11} f(x_i, y_i)\frac{\partial f}{\partial y}]h^2 + O(h^3)$$

Setting this result equal to the substituted general form will allow us to determine that, for these two equations to be equivalent, the following relationships between constants must be true.

$$a_1 + a_2 = 1$$

$$a_2 p_1 = \frac{1}{2}$$

$$a_2 q_{11} = \frac{1}{2}$$

It should be noticed that these three equations, relating necessary constants, have four unknowns. This means that to solve for three constants, one must first be chosen. This results in a family of possible second-order Runge-Kutta methods.

6.10 Sage's Corner

Example of usage of Euler's Method:

http://video.google.com/googleplayer.swf?docId=1095449792523736442

Narrated example of using the Runge-Kutta Method:

http://video.google.com/googleplayer.swf?docId=-2281777106160743750

Unnarrated example of Using the Runge-Kutta Method:
File:Numerical Solving in Excel, Unnarrated.ppt

6.11 References

- "Delay Differential Equation", Wolfram MathWorld, Online: September 8, 2006. Available http://mathworld.wolfram.com/DelayDifferentialEquation.html
- Chapra, Steven C. and Canale, Raymond P. "Numerical Methods for Engineers", New York: McGraw-Hill.
- Franklin, Gene F. et al. "Feedback Control of Dynamic Systems", Addison-Wesley Publishing Company.
- R. England. "Error Estimates for Runge-Kutta Type Solutions to Systems of Ordinary Differential Equations", Research and Development Department, Pressed Steel Fisher Ltd., Cowley, Oxford, UK. October 1968.
- Call, Dickson H. and Reeves, Roy F. "Error Estimation in Runge Kutta Procedures", ACM, New York, NY, USA. September 1958.

Section 7. Solving ODEs with Mathematica: How to find numerical and analytical solutions to ODEs with Mathematica

Note: ⊞Video lecture available for this section!

Authors: Matthew Baumgartner, Olyvia Dean, Viral Patel, Joel Schweitzer, and Eric Van Beek

Stewards: Brian Hickner, Lennard Gan, Addison Heather, Monique Hutcherson

Date Released: September 6, 2006 /Date Revised: September 8, 2007

7.1 Introduction

Mathematica is an advanced mathematics solution program created by Wolfram Research, Inc. One of the most powerful software packages of its kind, Mathematica is able to easily and conveniently solve complicated mathematical formulae, including differential equations. This article focuses on the modeling of first and higher order Ordinary Differential Equations (ODE) in the following forms:

$$\frac{dy}{dx} = f(x, y)$$
(Basic ODE)

$$\frac{dy^z}{d^z x} = f(x, y)$$
(Higher Order ODE)

Like all software, Mathematica uses a specific language in order to perform calculations. The names of all functions must be capitalized- this applies to both mathematical functions (such as **Sin** and **Cos**) and built-in functions (such as **Plot** and **DSolve**). For this reason it is common for users to write their own functions in minimized letters. This decreases the chance of overriding or redefining a Mathematica function. Square brackets always follow function calls, and the function's parameters are always enclosed in curved brackets. For example, if a user wished to plot $sin(x)$ from $x = 0$ to $x = 1$, they would type: **Plot[Sin[x],{x,0,1}]**. The user must type "Shift"+"Enter" to input the function. Typing only "Enter" will add a line to the formula. . For full PC keyboards, the "Enter" key on the far right is the equivalent of "Shift"+"Enter". If the user does not want to see the output line, it can be suppressed by typing a semi-colon at the end of the expression. Mathematica also recognizes common mathematical constants such as pi (**Pi**), Euler's constant (**E**), and imaginary numbers (**I**). Note that these constants must also be capitalized when entered.

Mathematica features two functions for solving ODEs: **DSolve** and **NDSolve**. **DSolve** is used when the user wishes to find the general function or functions which solve the differential equation, and **NDSolve** is used when the user has an initial condition. The prompts for the functions are quite similar. **Note:** Two equal signs must be used to denote equality in an equation. Using one equal sign assigns a value to a variable.

Example:

$f(x) = 5x^2 + 7$ This input creates a variable named $f(x)$ that has a value of $5x^2 + 7$.

$f(x) = = 5x^2 + 7$ This input creates a function $f(x)$ which is defined by $5x^2 + 7$.

Mathematica will sometimes store user created functions or variables into its cache. This causes some (potentially) correct commands to output errors. It is recommended to quit the local kernel and reinitialize the commands after any changes to your code.
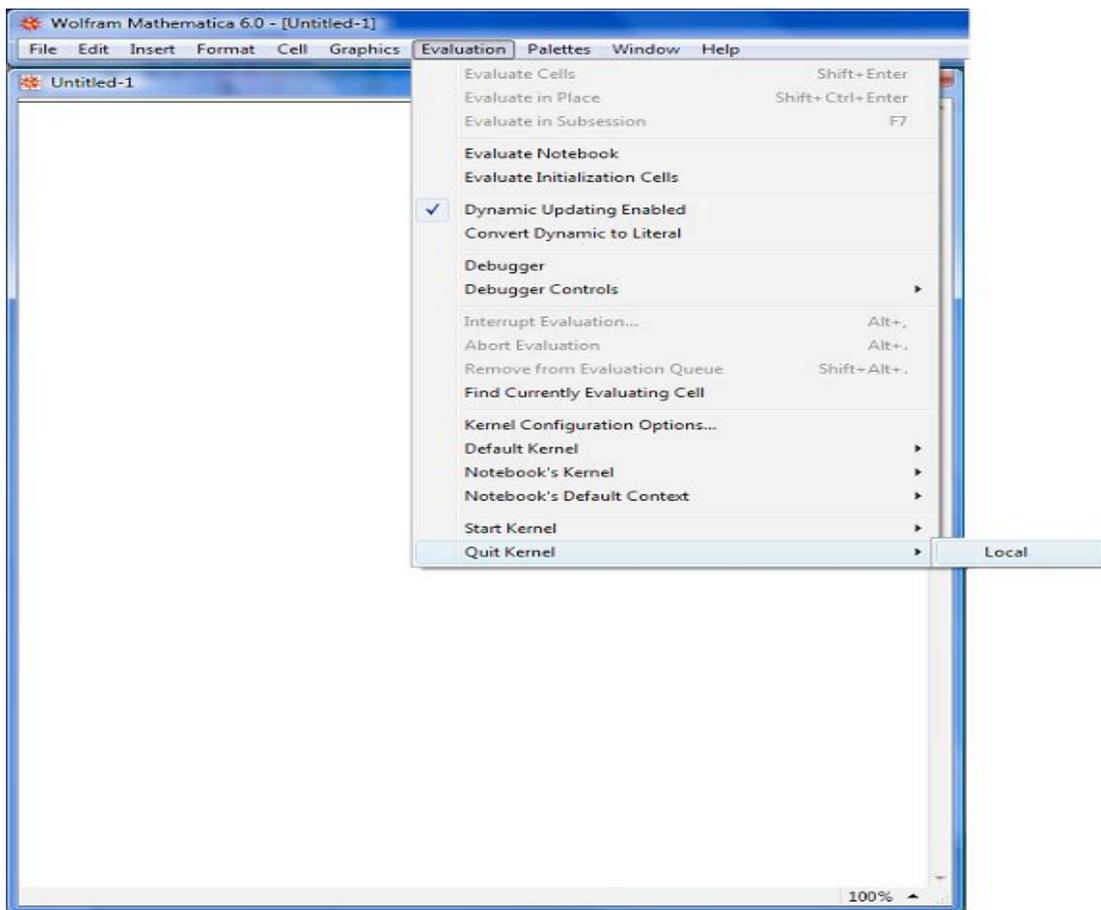
Also, check that every command in the code is being executed. Clearing all the outputs in the cell may be helpful to figure out which commands had not been executed. This can be done by going to the "Cell" option found on the top and choosing the "Delete All Output" Example:

OUTPUT IS NOT CLEARED

CLEARING THE OUTPUTS

CODE WITH CLEARED OUTPUTS

As seen in the examples, it is easier to troubleshoot and debug your program when it looks less confusing. Clearing the extra outputs helps you focus on just the code that you have written.

7.2 First Order ODEs

Notation: In the following examples, *eqn* represents the ODE, *y* represents the function being solved for, *i* represents the initial condition, and *x* and *t* are independent variables.

7.2.1 ODEs With Initial Conditions

NDSolve will not display a numerical value, but rather an "Interpolating Function" which can be displayed graphically. The easiest way to display this graph is to assign the solution to a variable (called *Solution* in the example) in the input line, and then use the "Plot" function to display it.

Input for One ODE: *Solution* = **NDSolve[{*eqn*,*y*[0] == *i*},*y*,{*x*,*xmin*,*xmax*}]**

Input for Multiple ODEs: *Solution* = **NDSolve[{*eqn1*,*eqn2*,...,*y1*[0] == *i1*,*y2*[0] == *i2*,...},{*y1*,*y2*,...},{*x*,*xmin*,*xmax*}]**

Input for Partial Differential Equations: *Solution* = **NDSolve[{*eqn*,*y*[0] == *i*},*y*,{*x*,*xmin*,*xmax*},{*t*,*tmin*,*tmax*}]**

After Mathematica solves the ODE, plot the solution by typing: **Plot[Evaluate[*y*[*x*] /. Solution],{*x*,*xmin*,*xmax*}]**, where "/." is a notation used by Mathematica meaning "following the rule of". It basically recalls the function stored as the variable in the

130

previous input line. Depending on the needs of the user, the functions "ParametricPlot" or "Plot3D" may also be useful. The notation for these functions is the same as "Plot".

Here is a simple example of what you would type into Mathematica:

- **Solution = NDSolve[{y'[x]==y[x]*Cos[x+y[x]],y[0]==1},y,{x,0,30}]**

Mathematica will output:

- Output[1]= {{y->InterpolatingFunction[{ { 0.,30.} },<>] } }

To plot this function you would type:

- **Plot[Evaluate[y[x]/.Solution],{x,0,30}]**

Note: Remember to type "Shift"+"Enter" to input the function

7.2.2 ODEs Without Initial Conditions

Input for One ODE: **DSolve[*eqn,y,x*]**

Input for Multiple ODEs: **DSolve[{*eqn1,eqn2,...*},{*y1,y2,...*},*x*]**

Input for a Partial Differential Equation: **DSolve[*eqn,y,{x1,x2,...}*]**

Here is a simple example of what you would type into Mathematica:

- **Solution = DSolve[y'[x] ==4*x-2*x*y[x],y[x],x]**

Mathematica will output:

- Output[1]= {{y[x]->2+ C[1]}}

7.3 Second Order and Higher ODEs

Input for Higher Order ODEs: *Solution = NDSolve[{eqn,y[0] == i1, y' [0] == i2,...},y,{ x,xmin,xmax}]*

Plotting the solution: **Plot[Evaluate[{y[x], y' [x],...} /. Solution],{x,xmin,xmax}]**

Here is a simple example of what you would type into Mathematica:

- **Solution = NDSolve[{y[x] +Sin[y[x]]+y[x]==0,y[0]==1,y'[0]==0,y,{x,0,30}]**

Mathematica will output:

- Output[1]= {{y->InterpolatingFunction[{ {0.,30.} },<>] } }

To plot this function you type:

- **Plot[Evaluate[{y[x],y'[x],y*[x]}/.Solution],{x,0,30}]**

Mathematica should output a graph.

7.4 Algorithms Used by Mathematica

Mathematica uses two main algorithms in order to determine the solution to a differential equation. These algorithms are the Adams method and the Gear method.

The Adams and Gear methods are forms of linear multistep methods. An example of these would be the following:

$$y_{n+1} = y_n + h(b_0 f(t_n, y_n) + b_p f(t_{n-p}, y_{n-p}))$$

In the example above, h denotes the step size and the coefficients $b_0, \ldots, b_s$ are determined by the method used. Multistep methods are expansions of more familiar single-step methods used to solve differentials (i.e. Euler, Runge-Kutta, Taylor). Each of these methods requires an initial point in order to calculate the following point. Similarily, multistep methods also require initial points in order to solve the ODE. The number of initial points required depends on which method is used to solve the ODE. Multistep methods typically produce less error than the single-step methods because of multiple initial points.

In order to determine what method to use one must first find the stiffness of the function.

7.4.1 Adams Method

Euler, Taylor and Runge-Kutta methods used points close to the solution value to evaluate derivative functions. The Adams-Bashforth method looks at the derivative at old solution values and uses interpolation ideas along with the current solution and derivative to estimate the new solution [4]. In order to solve an ODE using this method, $f(t,y)$ must be continuous and satisfy Lipschitz condition for the y-variable which states [5]:

$$|f(h) - f(0)| <= B|h|^\beta$$

for all $|h| < \varepsilon$ where B and β are independent of h, $\beta > 0$ and α is an upper bound for all β for which a finite B exists

132

This is a basic form of the Adams-Bashforth method. Note that two initial points are required for this method.

$$y_{n+1} = y_n + h\big[\beta_1 f(y_n, t_n) + \beta_2 f(y_{n-1}, t_{n-1})\big]$$

There is another Adams method that requires three initial points. The method is solved the same way, however the equation varies a little bit and is referred to as the Adams-Moulton Method.

$$y_{n+1} = y_n + h\big[\beta_0 f(y_{n+1}, t_{n+1}) + \beta_1 f(y_n, t_n) + \beta_2 f(y_{n-1}, t_{n-1})\big]$$

The coefficients/constraints, β can be solved for using knowledge of ODE's and other math tools. It was stated earlier that $\dfrac{dy}{dt} = f(t, y)$. We can let $f(t,y) = \lambda y$, therefore $\dfrac{dy}{dt} = \lambda y$. We can also let $y_n = \sigma^n y_0$ if there is a constant step size and σ represents a polynomial. Through substitution we find [6]:

$$-\sigma^2 + (\beta_1 h\lambda + 1)\sigma + \beta_2 h\lambda = 0$$

We can expand the quadratic using another math identity and ultimately solve for constraints β_1 and β_2. Another method for solving for coefficients β_1, β_2 is mentioned below:

In order to find the coefficient β_j one must first use polynomial interpolation to find the polynomial p of degree $s - 1$ such that:

$$p(t_{n-i}) = f(t_{n-i}, y_{n-i}), \text{ for } i = 1, \ldots, s.$$

From this the Lagrange formula for polynomial interpolation yields

$$p(t) = \sum_{j=1}^{s} \frac{(-1)^{s-j} f(t_{n-j}, y_{n-j})}{(j-1)!(s-j)!h^{s-1}} \prod_{\substack{i=1 \\ i \neq j}}^{s} (t - t_{n-i}).$$

Now the polynomial p is a locally good approximation of the right-hand side of the differential equation $y' = f(t,y)$ that is to be solved. Now we must consider the equation $y' = p(t)$ instead. This equation can be solved exactly by simply taking the integral of p.

133

$$y_n = y_{n-1} + \int_{t_{n-1}}^{t_n} p(t)\, dt.$$

The Adams–Bashforth method arises when the formula for p is substituted. The coefficients b_j turn out to be

$$\beta_j = \frac{(-1)^{s-j}}{(j-1)!(s-j)!} \int_0^1 \prod_{\substack{i=1 \\ i \neq j}}^{s} (u+i)\, du.$$

```
The Adams-Bashforth method is typically used for Linear and Non-liner
ODE's with dense systems.
```

7.4.2 Gear Method

Taking

$$\beta_i = 0$$

for $i \geq 1$

The Gear method, also known as the **backward differentiation** formulae (BDF, a.k.a. Gear's formulae) is another multi-step method but is generally used for multiple equations. In order to use the gear method your function must have a stiffness greater than 500, meaning that the function is stiff.

$$y_{n+1} = h\beta_0 f(y_{k+1}, t_{k+1}) - \sum_{i=1}^{m} \alpha_i y_{k+1-i}$$

As a particular cases, taking $a_i = 0$ for $i \geq 2$ and optimizing the remaining coefficients to maximize the accuracy of the resulting scheme recovers the Implicit Euler method. Taking $a_i = 0$ for $i \geq 3$ gives

$$y_{n+1} = -\alpha_1 y_n - \alpha_2 y_{n-1} + h\beta_0 f(y_{n+1}, t_{n+1})$$

134

We now focus on this case in particular. Applying this method to the scalar model problem $dy/dt = \lambda y$ and assuming constant h and a solution of the form $y_n = \alpha^n y_0$, we find the following quadratic equation for σ

$$(1 - \beta_0 h\lambda)\sigma^2 + \alpha_1\sigma + \alpha_2$$

the two roots of which are given by

$$\sigma = \frac{-\alpha_1 + \sqrt{\gamma(1+\epsilon)}}{2(1-\Sigma)} \text{ or } \frac{-\alpha_1 - \sqrt{\gamma(1+\epsilon)}}{2(1-\Sigma)}$$

$$\gamma = \alpha_1^2 - 4\alpha_2$$

$$\epsilon = (4\alpha_2\beta_0/\gamma)h\lambda$$

$$\Sigma = (\beta_0)h\lambda$$

Applying the identities, we may expand both roots in terms of powers of h. By our assumed form of the solution, it follows that $y_n = \sigma_+^n y_0, + \sigma_-^n y_0$ .The leading-order term in the expansion in h of σ_- (a "spurious root") is proportional to h. For small h, σ_-^n quickly decays to zero, and thusmay be neglected. The leading-order terms in the expansion in h of σ_+ (the "physical root") resemble the Taylor-series expansion of the exact solution over a single timestep.

$$\sigma_+ = (-\frac{\alpha_1}{2} + \frac{\sqrt{\gamma}}{2}) + \beta_0(-\frac{\alpha_1}{2} + \frac{\sqrt{\gamma}}{2}[1 + \frac{2\alpha_2}{\gamma}])\lambda h + \beta_0^2(-\frac{\alpha_1}{2} + \frac{\sqrt{\gamma}}{2}[1 + \frac{2\alpha_2}{\gamma} - \frac{2\alpha_2}{\gamma^2}])\lambda^2 h^2 + ...$$

Matching coefficients with the expansion of the exact solution

$$\sigma = e^{\lambda h} = 1 + \lambda h + \frac{\lambda^2 h^2}{2} + \dots$$

as indicated by underbraces in the above expression, and applying the definition $\gamma = \alpha_1^2 - 4\alpha_2$, we arrive at three equations for $\alpha_1, \alpha_2 and \beta_0$ to achieve the highest order of accuracy possible with this form. It is easily verified that

$$\alpha_1 = -4/3, \alpha_2 = 1/3, and \beta_0 = 2/3$$

satisfy these three equations.

The leading-order error term of this method is proportional to h^3. Thus, over a single timestep, the scheme is "locally third-order accurate"; more significantly, over a fixed time interval [0,T], the scheme is globally second-order accurate. The resulting method,

$$y_{n+1} = \frac{4}{3} y_n - \frac{1}{3} y_{n-1} + h \frac{2}{3} f(y_{n+1}, t_{n+1}),$$

is thus referred to as BDF2, and may be viewed as a implicit alternative to AM3 that, for the same number of steps into the past, p = max(m,n), has reduced order of accuracy but greatly improved domain of stability. Higher-order BDFs may be derived in an analogous fashion; BDF3, BDF4, BDF5, and BDF6 in particular are found to have excellent stability properties as compared with their AM counterparts with the same number of steps.

7.5 Worked out Example 1

You have been assigned a non-isothermal reactor with the following material and enthalpy balances.

$$\frac{dc_a}{dt} = c_a(t)^2 - c_b(t) - T(t)^2 \qquad \text{Species a Material balance}$$

$$\frac{dc_b}{dt} = c_b(t)^3 - c_a(t) - T(t)^2 \qquad \text{Species b Material balance}$$

$$\frac{dT}{dt} = c_a(t) - c_b(t) - T(t)^2 \qquad \text{Enthalpy balance}$$

When is the concentration of species a and species b equal? The initial conditions for the reactor are as follows:

$T_0 = 1$

$c_a0 = 2.0$

$c_b0 = 1.8$

$t_f = 0.2$

s = NDSolve[{x'[t] == x[t]^2 - y[t] - z[t]^2, y'[t] == y[t]^3 - x[t] - z[t]^2, z'[t] == x[t] - y[t] - z[t]^2, x[0] == 2, y[0] == 1.8, z[0] == 1}, {x, y, z}, {t, 0.2}]

Plot[Evaluate[{x[t],y[t]} /. s], {t, 0, 0.2}]
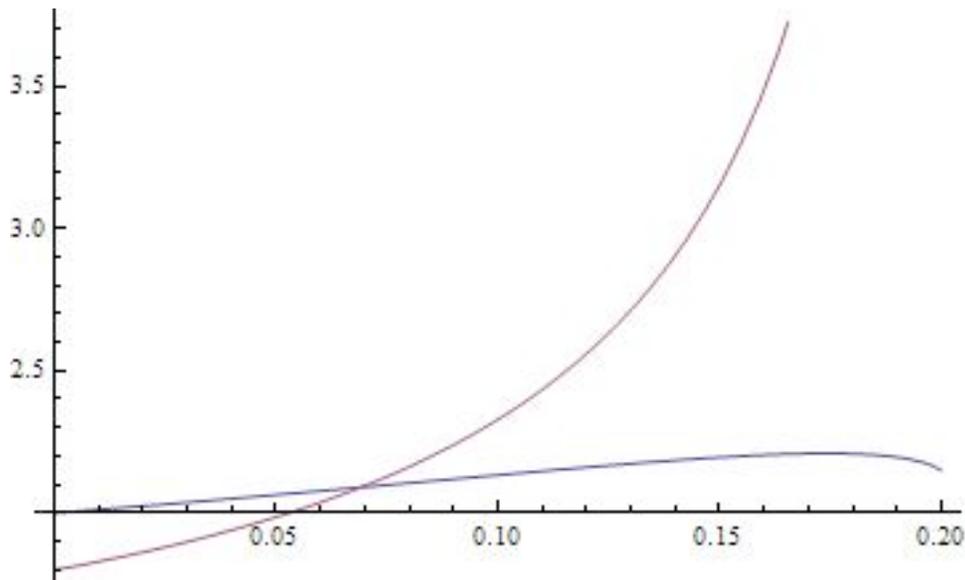
7.6 Worked out Example 2

You have been asked to study a semi-batch reactor for the reaction $A \rightarrow B$. The rate law is

$$-r_a = \frac{kC_a}{1 + C_b}$$

, where k = 2.7. Other parameters are: initial volume = 5, volumetric flow rate = 0.05, initial concentration of A = 2500. Use *Mathematica* to create a Conversion versus Time plot for 100 time units.

1. Mole Balance:

$$\frac{dX}{dt} = -\frac{r_a V}{N_{A0}}$$

2. Rate Law:

$$-r_a = \frac{kC_a}{1 + C_b}$$

3. Stoichiometry:

$$C_A = \frac{N_{A0}(1 - X)}{V_0 + v_0 t}$$

$$C_B = \frac{N_{A0} X}{V_0 + v_0}$$

4. Combine:

$$\frac{dX}{dt} = \frac{k(1 - X)(V_0 + v_0 t)}{N_{A0} X}$$

If the semi-batch reactor runs for 100 time units, the conversion is about 0.8. This can be seen in the plot below or by running the two commands that follow into *Mathematica*.

```
s = NDSolve[{y'[t] == 2.7 (1 - y[t]) (5 + .05 t)/(2500 y[t]), y[0] == 0.0001}, y, {t, 0, 100}];
```

```
Plot[Evaluate[y[t] /. s], {t, 0, 100}, PlotRange -> All]
```
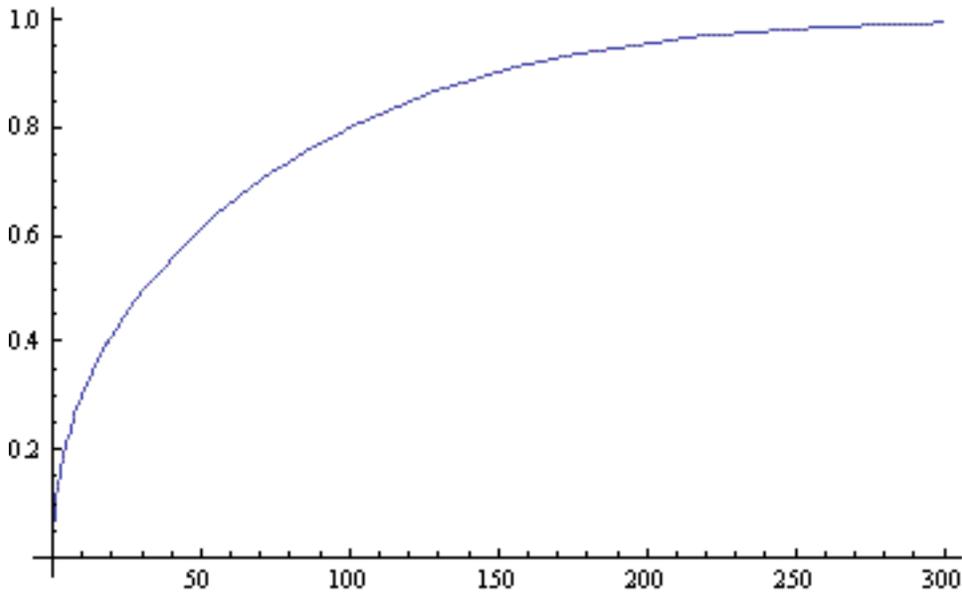
This is the *Mathematica* notebook file for the example: Media:MathematicaEx2.nb

Note: The parameters used in this problem are fabricated and are intended to illustrate the use of *Mathematica* in solving ODEs.

7.7 Formatting Plots in Mathematica

Here are some useful tips to formatting plots in Mathematica:

Label Plot: e.g. Plot[Sin[x], {x,0,2Pi}, **PlotLabel->"Volume of Tank vs. Time"**]

Label Axes: e.g. Plot[Sin[x], {x,0,2Pi}, **AxesLabel->{"time(min)","volume(L)"}**]

Color Plot: e.g. Plot[Sin[x], {x,0,2Pi}, **PlotStyle->Red**]. You can also make the plot thick, dashed, etc. e.g. Plot[Sin[x], {x,0,2Pi}, **PlotStyle->{Red,Thick,Dashed}**]

Insert Legend: e.g. **Needs["PlotLegends`"]** <--MUST insert this BEFORE your plot command. Plot[{Sin[x],Cos[x]}, {x,0,2Pi}, **PlotLegend->{"sine", "cosine"}**]

7.8 Sage's Corner

Solving ODEs with Mathematica Video

http://video.google.com/googleplayer.swf?docId=5007050669435611182

slides for this talk

Mathematica is a powerful computing tool, however the syntax can be a bit difficult to understand. Here are some notes for troubleshooting in Mathematica.

1. Check to make sure that your variable names and signs are consistent.

- Ex) Make sure you use xI everywhere instead of xI and x1 or xl.
- Ex) Functions, including the ones you create, are usually followed by brackets such as Sin[x] or y[x]. However, brackets are not necessary when you are solving for a function in a set of differential equations such as NDSolve[eqns, {y}, {x, 0, 50}];
- Ex) Check to see if your parentheses are aligned such that you are actually entering the function you think you're entering. Recall order of operations and the distributive property. x*(a+b) is NOT equal (x*a) + b. This seems simple, but often gets overlooked when dealing with several lines of code.

2. You may find it easier to define all of your variables together at the beginning of your code. This makes it easy to go back and change the values assigned to each variable if trying to determine the impact one variable has on a system of equations. For instance, say you are trying to determine the effects the flow rates into two different tanks (F1, F2) will have on the tank volumes over ten time steps. The differential equations governing the situation are:

$$\frac{dV_1}{dt} = 4F_1 - V_1 \quad \text{and} \quad \frac{dV_2}{dt} = F_2 - \frac{1}{3}V_2$$

, where F1 = 2, F2 = 12, V1(0) = 0, V2(0) = 0.

If you write the ODEs in Mathematica by directly substituting in F1 = 2 and F2 = 12, you will have to change the ODEs each time you change the values of F1 and F2. Below is an example of what the Mathematica equations would look like.

s = NDSolve[{V1'[t] == 8 − V1[t], V2'[t] == 12 − (1/3)*V2[t], V1[0] == 0, V2[0] == 0},{V1,V2},{t,0,10}]

Another option is to define F1 and F2 before your equations, and then call them when solving for your ODEs. This allows you to easily substitute in new values for your variables without changing the differential equations, and reduces the odds of making simple computational errors in your ODE in the process. The Mathematica code would look as shown below.

variables = {F1 -> 2, F2 -> 12};

s = NDSolve[{V1'[t] == 4*F1 − V1[t] , V2'[t] == F2 − (1/3)*V2[t], V1[0] == 0, V2[0] == 0} **/. variables**,{V1,V2},{t,0,10}]

3. Be Aware of the Kernel

The Mathematica Kernel stores all of the information about variable definitions. When you define a variable , the definition is stored there. The kernel is automatically started when a new Mathematica session is started. You may also start it manually. To start the kernel manually go to Evaluation -> Start Kernal -> Local. Once the kernel is started and you wish to go back and change a variable definition you must "Quit the Kernal" before you see the change occur. In order to "Quit the Kernal" you must go to Evaluation -> Quit Kernal -> Local. Terminating the Mathematica Kernal, erases all of the definitions previously entered. For this reason, after you "Quit the Kernal" and enter in new definitions for your variables you must re-enter all of your code. The images below show how to "Start Kernal" and how to "Quit the Kernel."
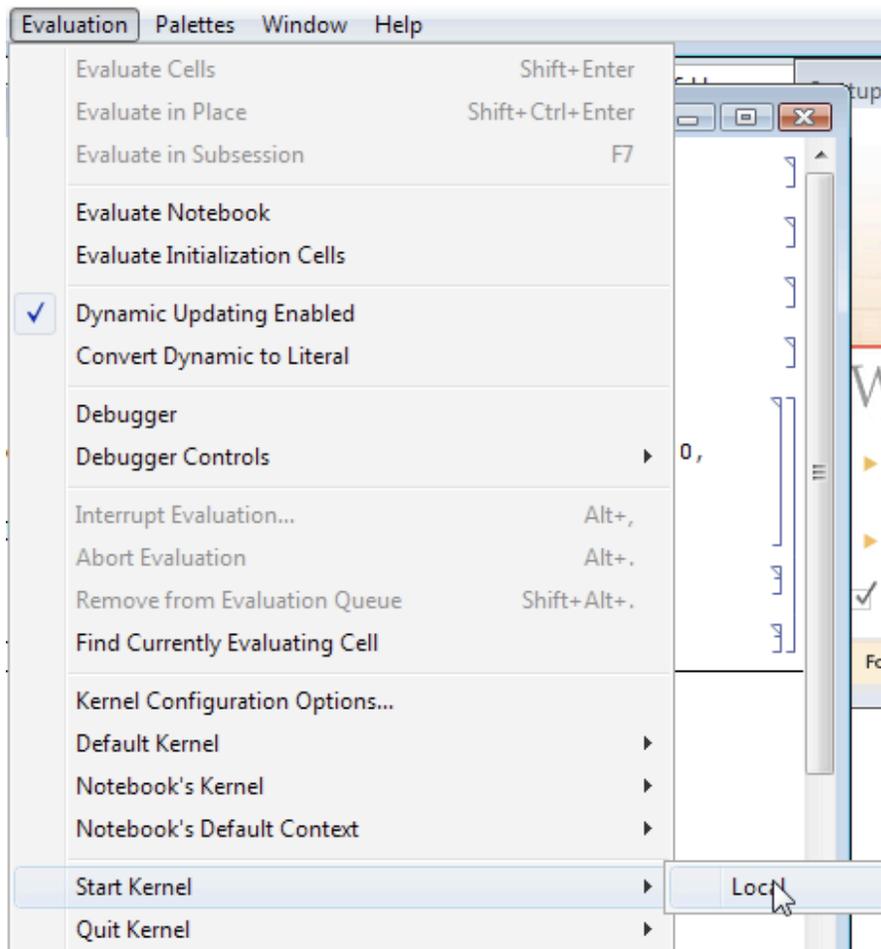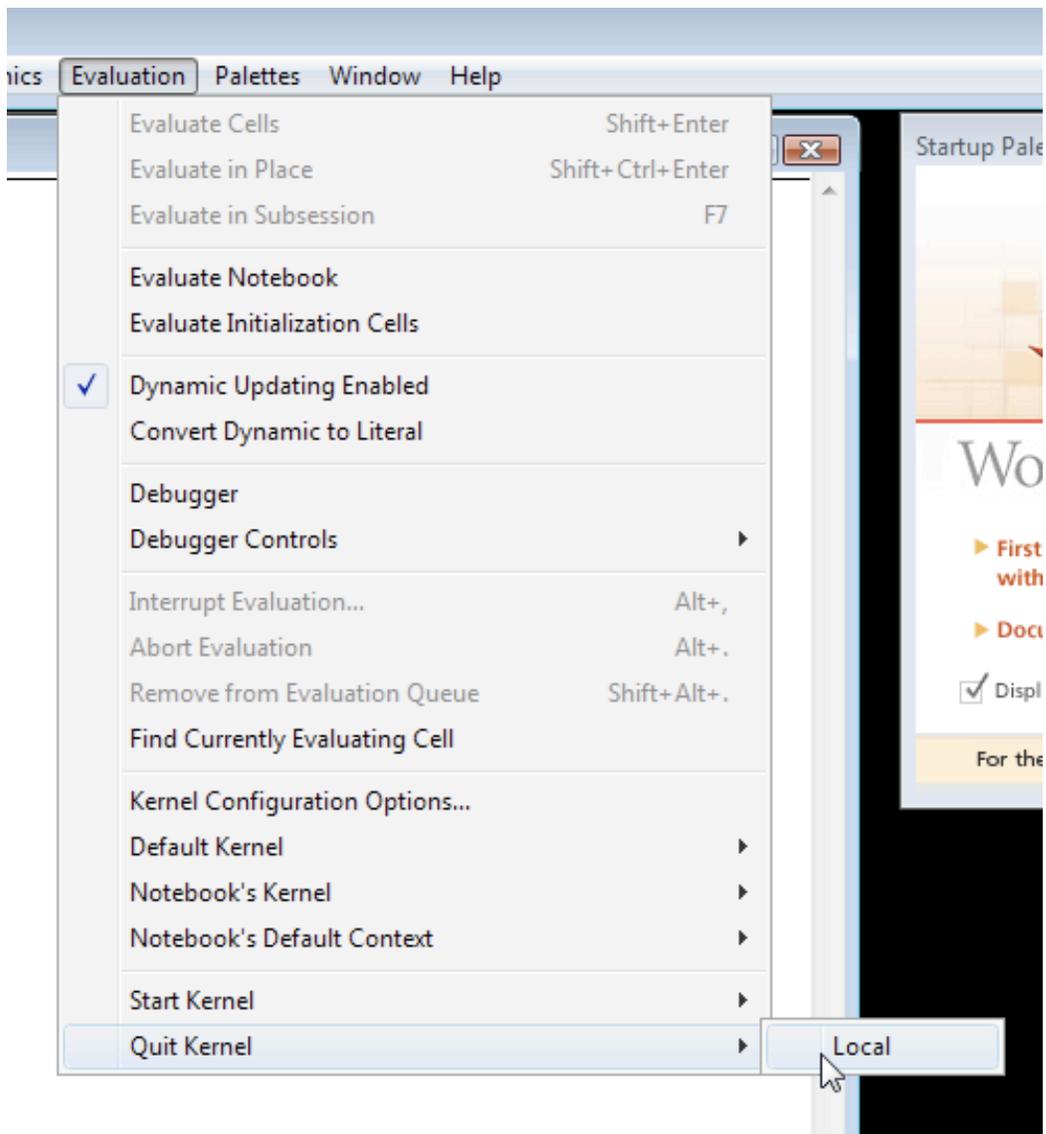
Example:

An example of when you would want to "Quit the Kernal" when using Mathematica for Controls is when you are finding a steady state point using a PID controller. We start by defining the variables for vset, taui, taud, and Kc. Mathematica solves this equation using the defined variables and shows that (x->3.9, V->10) is the steady state point. This is shown in the figure 1 below.

```
In[6]:= vars = {vset → 10, taui → .1, taud → 1, Kc → 10};
     eqnss =
       {(.5 + .5 * Kc * vset - .5 * Kc * V + (.5 / taui) * x - 2 * V) / (1 + .5 * taud) == 0,
         vset - V == 0}
     s = Solve[eqnss, {V, x}] /. vars
```

$$\text{Out[7]= } \left\{ \frac{0.5 - 2\,V - 0.5\,Kc\,V + 0.5\,Kc\,vset + \frac{0.5\,x}{taui}}{1 + 0.5\,taud} == 0,\ -V + vset == 0 \right\}$$

Out[8]= {{x → 3.9, V → 10.}}

Now, say you want to solve for the steady state point, but you do not want to define Kc because you will be varying it later to see its impact. You again start by imputing the variables and definitions for vset, taui, and taud, but not Kc. As you see in Figure 2, you still receive the identical answer as in Figure 1 when you defined Kc even though you have not defined it here.

```
In[2]:= vset = 10;

In[3]:= taui = .1;

In[4]:= taud = 1;

In[5]:= eqnss =
       {(.5 + .5 * Kc * vset - .5 * Kc * V + (.5 / taui) * x - 2 * V) / (1 + .5 * taud) == 0,
         vset - V == 0}
     s = Solve[eqnss, {V, x}]
```

Out[5]= {0.666667 (50.5 - 7. V + 5. x) == 0, 10 - V == 0}

Out[6]= {{V → 10., x → 3.9}}

This is because the Kernel is still storing the definition for this variable. In order to terminate this definition you must "Quit the Kernel." Once you do this, you must also re-enter your code and you will receive the correct answers. Figure 3 shows the output once you have "Quit the Kernel" and re-entered the Mathematica code.

```
In[3]:= vset = 10;

In[4]:= taui = .1;

In[5]:= taud = 1;

In[6]:= eqnss =
        {(.5 + .5 * Kc * vset - .5 * Kc * V + (.5 / taui) * x - 2 * V) / (1 + .5 * taud) == 0,
         vset - V == 0}
        s = Solve[eqnss, {V, x}]

Out[6]= {0.666667 (0.5 + 5. Kc - 2 V - 0.5 Kc V + 5. x) == 0, 10 - V == 0}

Out[7]= {{x → -0.1 - 10. (-0.4 - 0.1 Kc) - 1. Kc, V → 10.}}
```

Now you see that the definition for Kc has been deleted because the steady state points are in terms of Kc. To find the impact of Kc you can use this solution.

4. Define functions or formulas that are used often

If you will use a function often and it is not already defined in Mathematica by default, define your own. This can be especially helpful if you intend repeat the same function multiple times while only changing variables. A scenario where defining your own function could save you time and error is when comparing P-Fisher's values while keeping the margins constant. Only the variables will change but the function will remain the same. The new function will only last as long as each session in Mathematica.

To define a function, begin by naming the function followed by an open bracket. List all the variables with an underscore after each name and separate each variable with a comma. Close the bracket when all the variables have been listed. To let Mathematica know that this is a newly defined function, after closing the bracket place a semi-colon and equal sign. Now define what the formula is using the variable names. When you have finished, click shift-enter simultaneously. To test the new formula, start with naming the function and open bracket, then list all the variables numerical values separated by comma's in the same order that was listed when defining the function. Close the bracket after all the variables have been listed and click shift-enter simultaneously. The output will be the next line in Mathematica with the answer.

Example:

It is also possible to use this tool to put equations in terms of unknown variables. To do this begin by defining the function the same way as before. When plugging in the numerical values of the variables, leave the unknown variables as their variable name. The Mathematica output will provide the answer in terms of the unknown variables.

Example:

5. "Make it pretty"

Oftentimes when people program in any language, they tend to crowd lines, almost as if they were trying to save paper. Well, since this does not apply when you are typing on a screen, don't worry about it.

- Ex) You could write:

vars = {V -> 1000, Cao -> 0.2, UA -> 20000, F -> 2000, DE1 -> 30000, k1 -> 1.0 10^13, DH1 -> 50000, rcp -> 60, R -> 1.987, To -> 500, Tf -> 500, Tjin -> 500, Fj -> 1000, Vj -> 100, Caf -> 0.2, DH2 -> -192000, DE2 -> 30000, k2 -> 1.0 10^13};

eqns = {Ca'[t] == (Caf - Ca[t]))/V - Ca[t]*k1*Exp[-DE1/(R*T[t])], Cb'[t] == (0 - Cb[t])/V +k1*Ca[t]*Exp[-DE1/(R*T[t])] - k2*Cb[t]*Exp[-DE2/(R*T[t])], T'[t] == (Tf - T[t])/V + (-DH1/(rcp))*k1*Ca[t]*Exp[-DE1/(R*T[t])] + (-DH2/rcp)*k2*Cb[t]*Exp[-DE2/(R*T[t])] - (UA (T[t] - Tj[t]))/(V *rcp), Tj[t] == (Fj (Tjin - Tj[t]))/Vj + (UA (T[t] - Tj[t]))/(Vj *rcp), Ca[0] == 0.2, Cb[0] == 0, T[0] == 500, Tj[0] == 500};

sol = NDSolve[eqns /.vars, {Ca, Cb, T, Tj}, {t, 0, 50}]; Plot[{Ca[t]} /. sol, {t, 0, 50}, AxesLabel -> {t, Ca}, PlotRange -> Full];Plot[{Cb[t]} /. sol, {t, 0, 50}, AxesLabel -> {t, Cb},PlotRange -> Full];Plot[{T[t]} /. sol, {t, 0, 50}, AxesLabel -> {t, T}, PlotRange -> Full];Plot[{Tj[t]} /. sol, {t, 0, 50}, AxesLabel -> {t, Tj}, PlotRange -> Full]

- but it looks much better if you do this:

```
vars = {V -> 1000, Cao -> 0.2, UA -> 20000, F -> 2000, DE1 -> 30000,
k1 -> 1.0 10^13, DH1 -> 50000, rcp -> 60, R -> 1.987, To -> 500,
Tf -> 500, Tjin -> 500, Fj -> 1000, Vj -> 100, Caf -> 0.2,
DH2 -> -192000, DE2 -> 30000, k2 -> 1.0 10^13};
```

```
eqns = {
Ca'[t] == (F (Caf - Ca[t]))/V - Ca[t]*k1*Exp[-DE1/(R*T[t])],
Cb'[t] == (F (0 - Cb[t]))/V + k1*Ca[t]*Exp[-DE1/(R*T[t])] - k2*Cb[t]*Exp[-DE2/(R*T[t])],
T'[t] == (F (Tf - T[t]))/V + (-DH1/(rcp))*k1*Ca[t]*Exp[-DE1/(R*T[t])] + (-DH2/rcp )*k2*Cb[t]*Exp[-DE2/(R*T[t])] - (UA (T[t] - Tj[t]))/(V *rcp),
Tj'[t] == (Fj (Tjin - Tj[t]))/Vj + (UA (T[t] - Tj[t]))/(Vj *rcp), Ca[0] == 0.2, Cb[0] == 0, T[0] == 500, Tj[0] == 500};

sol = NDSolve[eqns /.vars, {Ca, Cb, T, Tj}, {t, 0, 50}];
```

```
Plot[{Ca[t]} /. sol, {t, 0, 50}, AxesLabel -> {t, Ca}, PlotRange -> Full]
Plot[{Cb[t]} /. sol, {t, 0, 50}, AxesLabel -> {t, Cb}, PlotRange -> Full]
Plot[{T[t]} /. sol, {t, 0, 50}, AxesLabel -> {t, T}, PlotRange -> Full]
Plot[{Tj[t]} /. sol, {t, 0, 50}, AxesLabel -> {t, Tj}, PlotRange -> Full]
```

These thoughts will hopefully help you later on and make it easier to read your code.

6.) "Check the colors!"

Here is a list of font color that Mathmatica will output for a specific reason.

Local Variables in a certain Font Color

Local variables of Module and With in Green

1. Example:

$$\text{Module}\left[\{x = 2, y = 3, \text{exponent}\}, \text{exponent} = 2; \{x, y\}^{\text{exponent}+x}\right]$$

Function arguments and pattern names in *Green (Italics)*

$$\text{func}[x\_, power\_] := \{x, x^2, x^{power}\}$$

1. Example:

Variables made special by use in arguments in Turquoise

$$\text{Table}[x^{power}, \{x, 1, 5\}, \{power, 1, 5\}]$$

- Example:

146

Errors and Warnings in a certain Font Color

Syntax Errors in Purple

- Example: `Table[triangle[{x+, y, z]]`

Emphasized Syntax Errors in Dark Red with Yellow Background

- Example: `Table[triangle[{x+, y, z]]`

Missing arguments in Bright Red

1. Example: `{N[‿], Rule[expr‿]}`

Excess arguments in Bright Red

1. Example: `{Sqrt[5, 7], Log[base, a, wrong]}`

Possible unwanted assignments in Bright Red

- Example: `If[x = 4, y = 9]`

Unrecognized option names in Bright Red

- Example: `Plot[Sin[x], {x, 0, 2 π}, PlotArea → {-1, 1}]`

Local Scope conflicts in Dark Red with Yellow Background

- Example:
  `f[x_, conflicted_] := Module[{ok, x, conflicted}, ok + x]`

Variables that will go out of scope before being used in Bright Red

- Example: `Table[Dynamic[var], {var, 1, 10}]`

Shadowing in multiple contexts in Bright Red

- Example:

```
{Global`Sin, Sin}
```

Other in a certain Font Color

Comments in Light Grey

- Example:

```
(* A comment explaining something *)
```

Strings in Dark Grey

- Example:

```
"A string of characters"
```

Global symbols that have no value assigned in Bright Blue

- Example:

```
{Table, undefinedsymbol, qq1}
```

7.10 Accessing Mathematica from your personal computer

If you prefer to use Mathematica from your home computer rather than one on campus, you can remotely log into a CAEN computer through U-M's Virtual Sites service. Follow the steps outlined below:

2. Visit http://virtualsites.umich.edu/ [[1]]. The U-M WEBLOGIN page will appear. Log in using your unique name and Kerberos password.
3. The Virtual Sites page should appear on your screen. Click on the large, orange **Connect Now** button.
4. Under the heading **What software do you need?**, select **Engineering (CAEN) Legacy XP**. Then click on the **Request Connection** button toward the bottom of the page.
5. Follow the instructions under the heading **Launch Virtual Sites** by downloading and opening the file in the Internet Explorer information bar. A window will pop up on your screen titled **Remote Desktop Connection**. Click the **Connect** button.
6. You will then see the screen that appears when logging into a CAEN computer. Your unique name should already be included. Type your Kerberos password and click **OK** to log in to the CAEN computer.
7. Click **OK** on the **Notice** window that pops up and your personal settings will be applied. Also click **OK** if a **System Requirements Wizard** window appears.
8. If you scroll down using the scroll bar on the right-hand side of the screen, you will see the **Start** button. Click **Start -> All Programs -> Math and Numerical Methods -> Wolfram Mathematica -> Mathematica 6**. Once

148

Mathematica opens on the computer screen, you may use it as though you were actually sitting in front of a CAEN computer on campus.

9. To log out, click **Start -> Log Off**. Under the **Log Off Windows** window, click **Log Off**.

7.11 References

- "Adams Method", Wolfram MathWorld, Online: August 5, 2007. Available http://mathworld.wolfram.com/AdamsMethod.html [1]
- "Gear Predictor-Corrector Method", Davidson University, Online August 5, 2007. Available http://webphysics.davidson.edu/Projects/SuFischer/node47.html [2]
- Fogler, H. Scott (2006), Elements of Chemical Reaction Engineering, New Jersey: Pretince Hall PTR. ISBN 0-13-047394-4 [3]
- "Adams-Bashforth", Pitt Math, Online: September 8, 2007. Available: http://www.math.pitt.edu/~sussmanm/2071Spring07/lab02/index.html#AdamsBashforth Methods [4]
- "Lipschitz Condition", Wolfram MathWorld, Online: September 7, 2007. Available http://mathworld.wolfram.com/LipschitzCondition.html [5]
- "Adams-Bashforth (AB) Methods", University of California-San Diego, Online: September 7, 2007. Available: http://renaissance.ucsd.edu/chapters/chap11.pdf [6]
- "Gears Method", University of California-San Diego, Online: September 7, 2007. Available: http://renaissance.ucsd.edu/chapters/chap11.pdf [7]
- Use the Harvard Referencing style for references in the document.
- For more information on when to reference see the following wikipedia entry.

Section 8. Fitting ODE parameters to data using Excel: Using regression to fit complex models in Excel

Note: Video lecture available for this section!

Authors: Anthony Campbell, Victoria Cardine, David Hines, Stephen Kerns /Date Revised: 14th September 2007

Stewards: Jeff Byrd, Khek Ping Chia, John Cruz, Natalie Duchene, Samantha Lyu

8.1 Introduction

One of the most important abilities of an engineer is to model measured values of a system. Modeling of a system aids in optimizing objectives, as well as predicting future operations. This can be especially helpful when changes cannot be immediately implemented. In general modeling terms, a set of data can always fit a set of variables. However, an excessive number of degrees of freedom usually exist, creating multiple solutions. Therefore, the physical significance of the variables is an important consideration.

As time proceeds, various values may need to be changed to accommodate variations within the system. The following sections will discuss the use of Microsoft Excel© as a tool to adjust the values of variables given a set of experimental data. The models presented below take into account the design equations of the system being modeled, and the rate laws, where applicable.

8.2 Adding in the Solver Application in Excel 2007

In order to use "Solver" in Excel 2007, one must first add the Excel tool-pack. To add the tool-pack, follow these instructions:

- Click on the Microsoft Office symbol at the top left corner of the screen
- Click on "Excel options" at the bottom right corner of the drop box
- Click on "Add-Ins" in the left hand column
- Click on the "Solver Add-In" option at the bottom of the list
- Click on "GO" at the bottom right corner
- Check the "Solver Add-in" box at the bottom of the list
- Click on "OK" at the top of the right column

To access Solver, click on the "Data" tab at the top of the screen. The Solver tool is located in the "Analysis" box on the right side of the screen.

The user interface of Excel 2007 differs from those of previous versions. The following images illustrate the process just described.
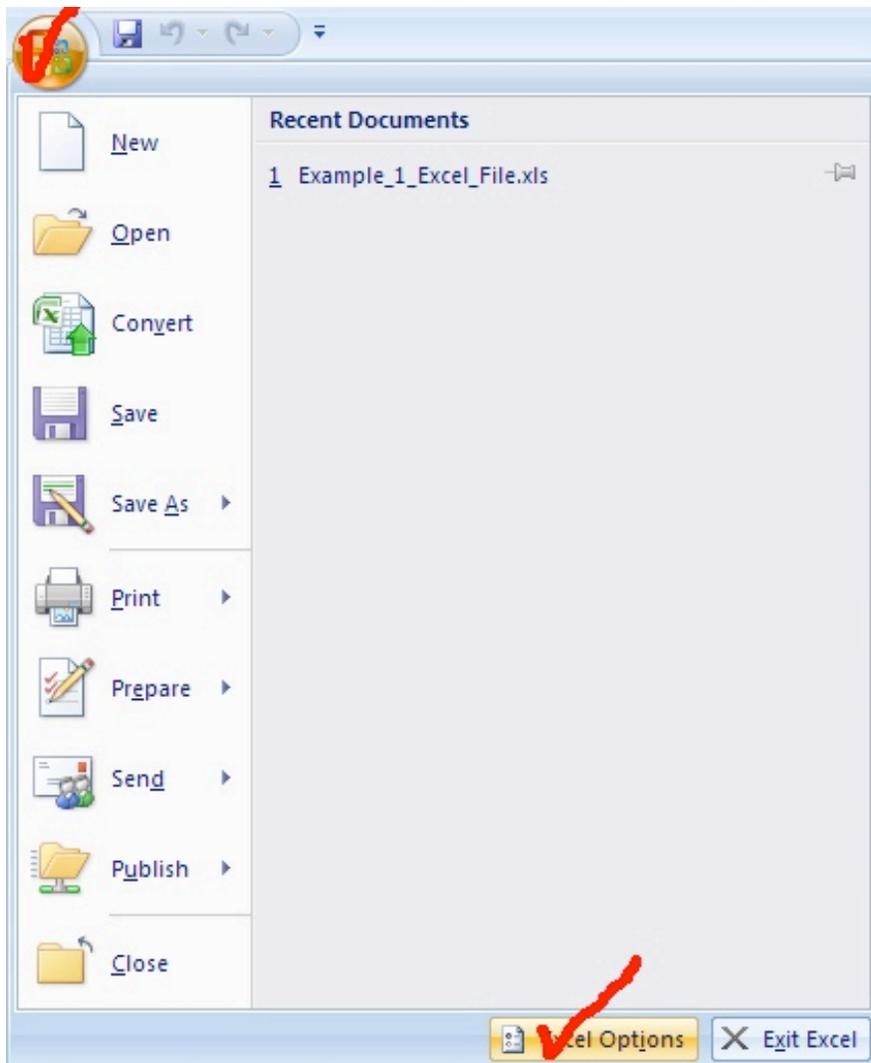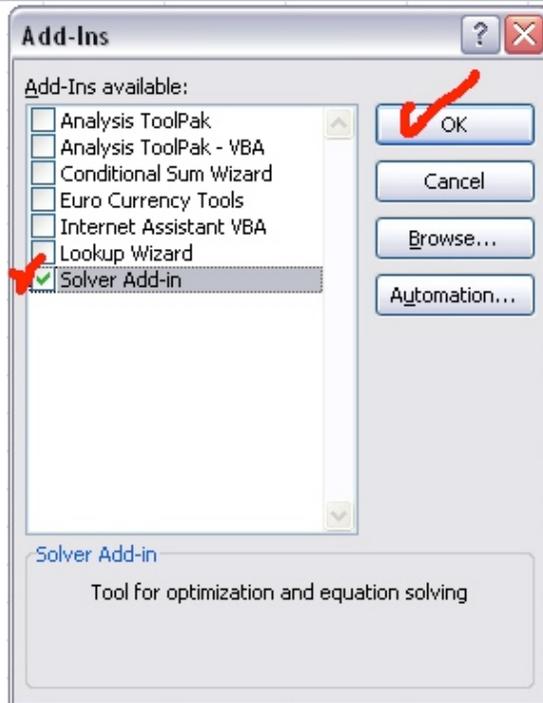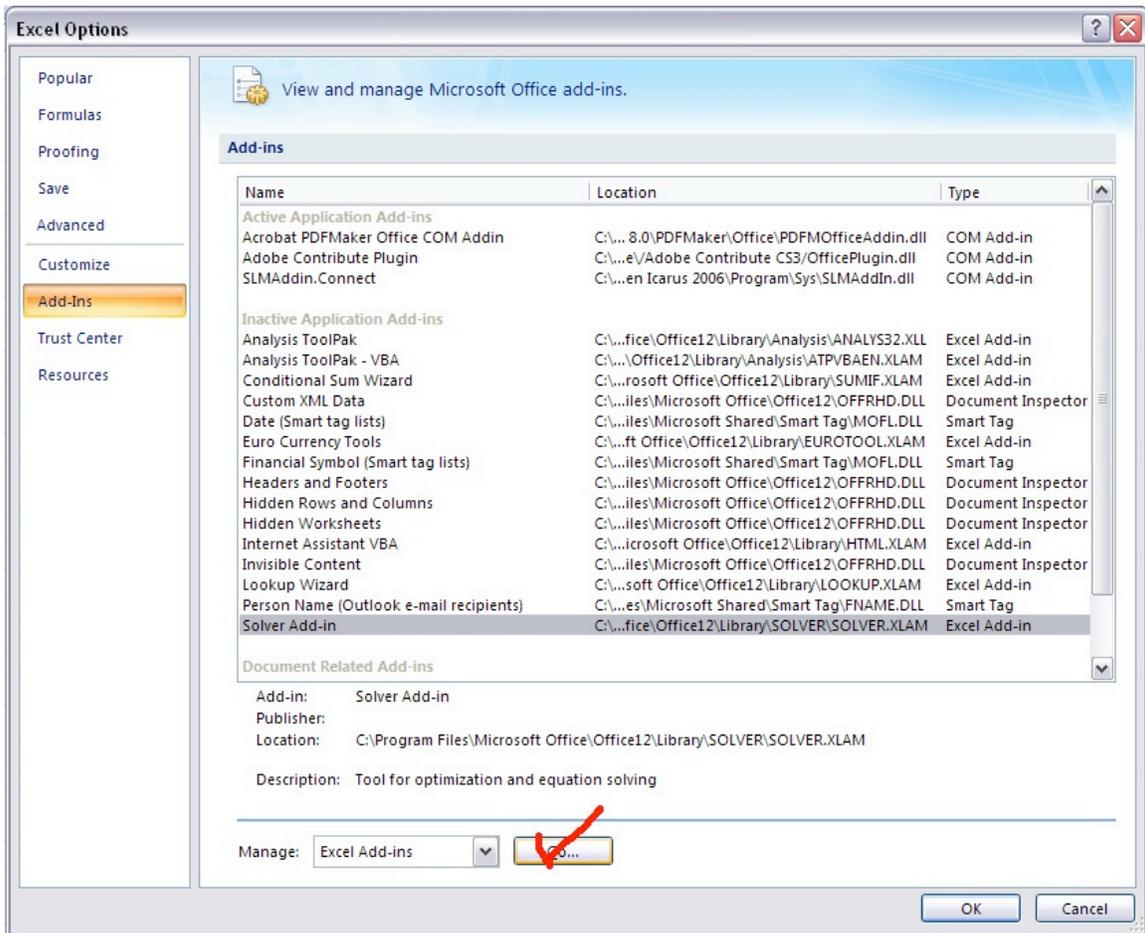
In order to use the Solver function, click on Data and then the Solver button in the Analysis box.
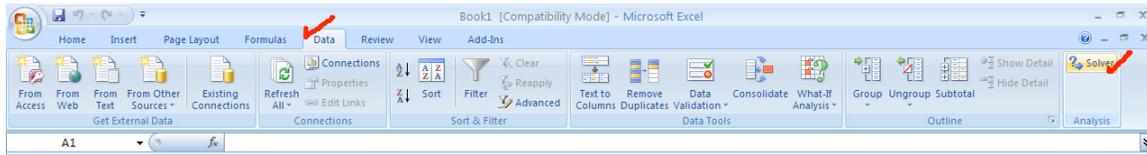
NOTE: If using Microsoft Office 2003, the Solver Application can be added by clicking on the "Tools" tab at the top of the screen. The same directions can be followed after this step.

NOTE: If you receive an error message while trying to access Solver, it may already be added in Excel. Uninstall Solver (by following the same steps, but unchecking "Solver Add-in") then reinstall it.

8.3 Adding in the Solver Application in Excel 2008

Excel 2008 for Macs does not come with the solver add-in. The software can be downloaded at Solver.com

This solver application requires Excel 12.1.2. If this is not the current version of Excel being used, it can be updated by running Microsoft AutoUpdate or by opening excel, going to help and then clicking on "Check for Updates". Make sure when updating Excel to close all Microsoft Applications.

This application allows the user to have a solver program that can be opened and used while running Excel. It works the same way as the solver add-in.

8.4 Using Excel Solver to Fit ODE Parameters

Excel Solver is a tool that can be used to fit function variables to given experimental data. The following process can be used to model data:

- Define the data set by entering values into an Excel spreadsheet
- Define the model that you want to fit to the data
- Define the sum of least squares in one of the spreadsheet blocks
- In Solver, have Excel minimize the sum of least squares by varying the parameters in the model

Accurate modeling is dependent on two factors; initial values and verifying results.

Initial values: As stated in the introduction, many data fitting problems can have multiple "solutions". In numerical methods, given a set of initial parameters the data will converge to a solution. Different initial values can result in differing solutions. Therefore, if the initial values are not set by the problem statement, multiple initial guesses should be entered to determine the "best" value for each variable.

Verification of curve fit: When fitting a curve to given data points, it is important to verify that the curve is appropriate. The best way to do this is to view the data graphically. It is fairly simple to use the Chart tool to graph the data points. Adding a trend line will show the mathematical relationship between the data points. Excel will generate a function for the trend line and both the function and the R-squared value can be shown on the plot. This can be done by clicking "Options" next to "Type" of trend and checking the empty boxes that are labeled "Display equation on chart" and "Display R-squared value on chart".

There are various types of trendlines that can be chosen for your given data points. Depending on the trend that you may see, you might want to try a few of the options to obtain the best fit. Sometimes the squared residual value will be better for a certain type of trend, but the trend is not necessarily "correct". When you have obtained sample data, be aware of the trend you should be seeing before fitting a certain trend type to the data points (if this is possible). The polynomial trend type gives you the option to change the order of the equation.

Typically, you will not acquire a squared residual value of zero. This value is a simple analysis of the error between the trend line, and the actual data. Once the function and squared residual values are generated, you can begin to evaluate the generated solution.

During your solution analysis, it is important to check for variables or parameters that may be unnecessarily influencing the data. Excel can generate a "best" solution, which is not correct due to background calculations or Excel's misinterpretation of the modeled system. To assess this possibility, you can graph data generated by using the solved for variables against the given data.

Additional information about the Excel Solver Tool can be found in *Excel Modeling.*

8.5 Worked out Example 1: Mass Balance on a Surge Tank

In this example we will walk through the scenario in which you want to model a non-heated surge tank in Microsoft Excel©. This model would be very critical in a process where smooth levels of process flow are required to maintain product specifications.

Therefore, an engineer would like to know whether they will need to periodically refill or drain the tank to avoid accidents. Let's look at this simple example:
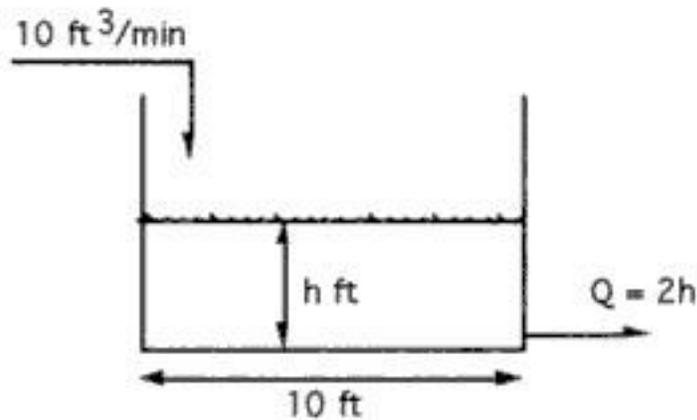
Problem statement:

Suppose water is being pumped into a 10-ft diameter tank at the rate of 10 ft3/min. If the tank were initially empty, and water were to leave the tank at a rate dependent on the liquid level according to the relationship Qout = 2h (where the units of the constant 2 are [ft2/min], of Q are [ft3/min], and of h are [ft]), find the height of liquid in the tank as a function of time. Note that the problem may be, in a way, over specified. The key is to determine the parameters that are important to fit to the experimental surge tank model.

Solution:

The solution with fitted parameters can be found here: Surge tank solution This Excel file contains this scenario, where given the actual data (the height of the liquid at different times), the goal is to model the height of the liquid using an ODE. The equations used to model h(t) are shown and captioned. In order to parameterize the model correctly, Solver was used to minimize the sum of the least square differences at each time step between the actual and model-predicted heights by changing the parameters' values. The model predicts the actual height extremely well, with a sum of the least squared differences on the order of 10^-5. A graph showing the modeled and actual height of the tank as a function of time is included as well.

8.6 Worked out Example 2: Fitting a heated surge tank model(a heated CSTR) parameters to data in Excel

Problem Introduction: In this example we will analyze how we can fit parameters used for modeling a heated CSTR to data stored in Excel. The purpose of this problem is to learn how to vary ODE parameters in Excel in order to achieve a square difference between the calculated and experimental data of zero. Some questions to consider are, "How do I model this? Is this a first order ODE? Are there any coupled ODE's? How

many variables need to be considered?" This example will use an Excel file to help you in this process.

Problem Statement: Download the heated CSTR problem file.<u>CSTR Problem</u> Take the experimental data produced from the heated CSTR model in the Excel spreadsheet*. You are modeling a heated CSTR. It would be wise and is highly recommended to review both the heated CSTR problem and a wiki on using Excel's solver.

You know this is a heated CSTR...

2. Numerically solve the ODEs through modifying the given wrong set of parameters.
3. While changing the parameters, use Excel to calculate the error between the calculated model and the experimental data.
4. Use the Excel solver function to solve for some of the parameters to minimize error to less than 5%
5. How would your approach be different if our heat transfer area needed to be as small as possible due to capital constraints? Explain.
6. How would your approach be different if our mass feed rate has to be in the range of 3-3.5 kg/s? Explain.

Problem Solution: The solution with fitted parameters can be found here.<u>CSTR Solution</u>

With the model and data given in the example the error must first be defined. Here there error is the squared residual or [(model observation at time i)-(data observation at time i)]^2 over (all i).

Chemical engineering intuition must be used to reasonably approximate some variables. In real life we would double check that "parameters normally known, but erroneous" are correct. (Due to problem constraints, we will assume these are accurate, but a true chemical engineer would check regardless). We must now modify "parameters often needing fitting". For part 3, we can modify the parameters as shown in the solution.

156

The spreadsheet shows the following data:

	A	B	C	D (Description)
1		Given Information		Description
2		Parameters often needing fitting		
3	E	166000	J/mol	Activation energy
4	k_0	1.97E+20	s^{-1}	Rate constant
5	UA	7000	J/s	Heat Transfer Coefficient and Area
6	ΔH_{Rx}	-200	J/mol	Heat of reaction
7	m	4	kg/s	Mass feed rate and product rate
8				
9		Parameters normally known, but possibly erroneous		
10	C_{A0}	9	mol/L	Initial Feed Concentration
11	$C_{A0, 2}$	9	mol/L	Change Initial Feed Concentration (leave the same if this does not occur)
12	t for $C_{A0, 2}$	0	s	Time when Initial Feed Conc changes
13	T_0	400	K	Feed temperature
14	$T_{0, 2}$	400	K	Change Initial Feed Temperature (leave the same if this does not occur)
15	t for $T_{0, 2}$	0	s	Time when Initial Feed Temperature Changes
16	T_c	350	K	Coolant temperature
17	$T_{c, 2}$	350	K	Change Initial Coolant Concentration (leave the same if this does not occur)
18	t for $T_{c, 2}$	0	s	Time when Initial Coolant Temperature Changes
19				
20		Parameters that do not normally vary		
21	V	3000	L	Reactor volume
22	ρ	1	kg/L	Density of reactor feed and product (assumed constant)
23	ΔC_p	1000	J/kg-K	heat capacity of the reactor feed and product
24				
25		Parameters that do not effect actual results*		
26	Δt	1	s	Time Step *Change only time steps are not equal between experimental and calculated values
27	R	8.314	J/mol-K	Ideal gas constant
28	E/R	19966.32	K	Normalized activation energy

$$\frac{dC_A}{dt} = \frac{m}{\rho V}(C_A$$

$$\frac{dT}{dt} = m \Delta C_p (T_0 -$$

The original parameters have been modified to produce an error of **zero**.

CSTR_Heat_Effects.xls [Compatibility Mode] - Microsoft Excel

Home | Insert | Page Layout | Formulas | Data | Review | View | Add-Ins

$$\frac{dC_A}{dt} = \frac{m}{\rho V}(C_{A0} - C_A) - k_0 C_A e^{-E/RT}$$

$$\frac{dT}{dt} = \frac{m\Delta C_p(T_0 - T) - V\Delta H_{rxn}C_A k_0 e^{-E/RT} + UA(T_C - T)}{V\rho C_p}$$

$$C_{Ai} = C_{A(i-1)} + \frac{dC_A}{dt}\Delta t$$

$$T_i = T_{i-1} + \frac{dT}{dt}\Delta t$$

Total square difference is equal to zero (0)

Calculated and Experimental Data match (Parameters have a good fit)

	Given Information		Description
	Parameters often needing fitting		
E	166000	J/mol	Activation energy
k_0	1.97E+20	s^{-1}	Rate constant
UA	7000	J/s	Heat Transfer Coefficient and Area
ΔH_{Rx}	-200	J/mol	Heat of reaction
m	4	kg/s	Mass feed rate and product rate
	Parameters normally known, but possibly erroneous		
C_{A0}	9	mol/L	Initial Feed Concentration
$C_{A0, 2}$	9	mol/L	Change Initial Feed Concentration (leave the same if this does not occur)
t for $C_{A0, 2}$	0	s	Time when Initial Feed Conc changes
T_0	400	K	Feed temperature
$T_{0, 2}$	400	K	Change Initial Feed Temperature (leave the same if this does not occur)
t for $T_{0, 2}$	0	s	Time when Initial Feed Temperature Changes
T_c	350	K	Coolant temperature
$T_{c, 2}$	350	K	Change Initial Coolant Concentration (leave the same if this does not occur)
t for $T_{c, 2}$	0	s	Time when Initial Coolant Temperature Changes
	Parameters that do not normally vary		
V	3000	L	Reactor volume
ρ	1	kg/L	Density of reactor feed and product (assumed constant)
ΔC_p	1000	J/kg-K	heat capacity of the reactor feed and product
	Parameters that do not effect actual results*		
Δt	1	s	Time Step *Change only time steps are not equal between experimental and calculated values
R	8.314	J/mol-K	Ideal gas constant
E/R	19966.32	K	Normalized activation energy

Calculated

t (s)	T_0	T (K)	T_c	C_{A0}	C_A	dCa/dt	dT/dt
0	400	400	350	9	9	0.00E+00	0
1	400	400	350	9	9	-3.72E-01	-0.04227
2	400	399.9577	350	9	8.627994	-3.54E-01	-0.04556
3	400	399.9122	350	9	8.273736	-3.37E-01	-0.04869
4	400	399.8635	350	9	7.936446	-3.21E-01	-0.05167

Experimental

(s)	T0	T (K)	Tc	CA0	CA	dCA/dt	dT/dt
0	400	400	350	9	9	0	0
1	400	400	350	9	9	-0.37200623	-0.042265421
2	400	399.957735	350	9	8.62799377	-0.35425749	-0.045560993
3	400	399.912174	350	9	8.27373628	-0.33729001	-0.048692964
4	400	399.863481	350	9	7.93644627	-0.32108402	-0.051665678

	total square difference		0
	t (s)		T0
	total square difference	0	0

square

t (s)	T0
0	0
0	0
0	0
0	0
0	0

If we have restrictions on heat exchange area then we should try to fit the "heat transfer coefficient and area" as low as possible. Likewise, if we have restrictions on feed rate, we must keep this above 4 while fitting other parameters as best as possible to achieve a small error. These have certain industrial applications as costing of heat exchangers and limitations on mass flow rates need to be modeled before a reactor design is implemented.

Side Note: The example above assumes a first order rate law. This is not always the case. To change the equations to fit a new rate law, one must derive a new dCa/dt equation and dT/dt using their knowledge of reactor engineering. Once the new equations are derived write the formula in an excel format in our dCa/dt and dT/dt column and then drag it to the end of the column. All other columns should stay the same.

8.7 Sage's Corner

A brief narration for a better understanding of Example 1 - Surge Tank Model and tips for effectively picking a model equation can be found here:
http://video.google.com/googleplayer.swf?docId=70222262469581741
A copy of the slides can be found here:

[Unnarrated Slides](#)

An example of fitting an ODE parameter to data using excel - based on a simple radioactive decay rate ODE:
http://video.google.com/googleplayer.swf?docId=8229764325250355383
A copy of the slides can be found here: [Unnarrated Slides](#)
A copy of the excel file can be found here: [Excel File](#)

An example of modelling a reaction rate ODE with rate constant and rate order variable parameters:
http://video.google.com/googleplayer.swf?docId=-4220475052291185071
A copy of the slides can be found here: [Unnarrated Slides](#)
A copy of the excel file can be found here: [Excel File](#)

8.8 References

Holbert, K.E. "Radioactive Decay", 2006, ASU Department of Electrical Engineering, Online: October 1, 2007. Available http://www.eas.asu.edu/~holbert/eee460/RadioactiveDecay.pdf

Section 9. Helpful Mathematica Syntax: Hints on how to use Mathematica to model chemical processes

Title: Useful functions in Mathematica as pertaining to Process controls

9.1 Introduction

Mathematica has many different functions that are very useful for controlling chemical engineering processes. The Mathematica syntax of these functions is often very complicated and confusing. One important resource is the Mathematica Documentation center, which is accessed through the program help menu. The Documentation Center provides an extensive guide of every function in the program including syntax and sample code. A "First Five Minutes with Mathematica" tutorial is also present in the menu, which provides a quick guide to get started with Mathematica. This article will briefly outline several useful Mathematica functions and their syntax.

9.2 Mathematica Basics

Here are some basic rules for Mathematica.

Notebook, Cells and Evaluation

The standard file for Mathematica is called a notebook. The notebook consists of two kinds of cells: input and output. The inputs are typed into one cell and the outputs are displayed in another (See Below).
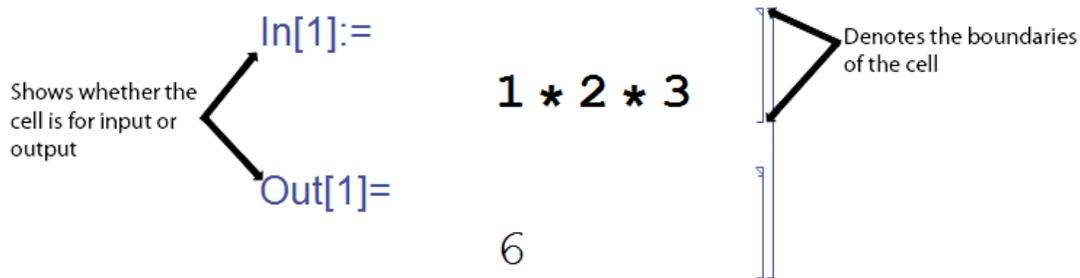
In order to have Mathematica calculate your inputs, you must evaluate the cells. Do this by pressing **Shift+Enter**. This will only evaluate the current cell that your cursor is located. To evaluate all the cells of a notebook press Ctrl+A, to select all, and then press Shift+Enter.

160

Parentheses vs. Brackets

Brackets, [], are used for all functions in Mathematica. The only thing parentheses, (), are used for is to indicate the order of operations.

Equality, = vs. ==

The single equal sign (=), is used to assign a value or a function to a variable. The double equal sign (==) is used to set two values equal to each other, such as solving for a value, or to test equalities. When testing equalities, Mathematica will output 'True' or 'False.'

If you recieve 'True' or 'False' as an output when you are expecting it, you likely used a == somewhere that you should have used a =.

Semicolon Use

Placing a semicolon after a line of code will evaluate the expression; however, there will be no output shown.

Mathematica Functions

All Mathematica functions start with a capital letter, eg. *Sin[x], Exp[x], Pi* or *Infinity*.

Assigning and Inserting Variables or Parameters

To assign values to variables, use -> rather than an equals sign. For example,

```
param = {Kc → 0.5, Ca0 → 1}

{Kc → 0.5, Ca0 → 1}
```

To insert a set of parameters to a function use /. This symbol applies a rule or list of rules in an attempt to transform each subpart of an expression.

For example, if you want to enter the above parameters into the expression $y = Kc * x + Ca0$, enter the following in Mathematica:

```
y = Kc * x + Ca0 /. param

1 + 0.5 x
```

Variables are case sensitive. For example, X is different than x. Also, if you define a variable, it will remain defined until you redefine it as something else or clear it. You can clear a variable, for example, x, by using the Clear[x] command. You can also quit the Kernel, which will clear all variables.

Forcing a Numerical Expression

To force Mathematica into giving a numerical solution, use the function *N[expression]*. For example,

```
In[7]:= (5 * 10 * 20 - 3) / (3! * 5)

        997
Out[7]= ---
        30


In[8]:= N[(5 * 10 * 20 - 3) / (3! * 5)]

Out[8]= 33.2333
```

Another method to do this is to place a decimal place in any of your numbers (i.e. "5." instead of just "5")

9.3 Integration

Mathematica can do some very complex integrations with fairly little coding. Integration can be very helpful for many engineering applications, but when the functions become very complex they become increasingly difficult to integrate. Mathematica will integrate any function that is possible to integrate.

To integrate a function *f(x)* the Mathematica, the command is *Integrate[]*.

For an indefinite integral the correct coding is *Integrate[f(x),x]* where x is the independent variable.

For a definite integral with limits of *y* and *z* the correct coding is *Integrate[f(x),{x,y,z}]*

For example:

We can integrate a function like the one below:

$f(x) = Sin(5 * x / 8) * x$

We can find the indefinite integral as seen here:

In[45]:= **Integrate[Sin[5 x / 8] * (x), x]**

Out[45]= $-\dfrac{8}{5} x \, Cos\left[\dfrac{5\,x}{8}\right] + \dfrac{64}{25} Sin\left[\dfrac{5\,x}{8}\right]$

We can find the definite integral from 0 to π as seen here:

In[46]:= **Integrate[Sin[5 x / 8] * (x), {x, 0, Pi}]**

Out[46]= $\dfrac{8}{25}\left(8\,Cos\left[\dfrac{\pi}{8}\right] + 5\,\pi\,Sin\left[\dfrac{\pi}{8}\right]\right)$

9.4 Solver

Mathematica's *Solve* function attempts to solve an equation or set of equations for the specified variables. The general syntax is *Solve[eqns, vars]*

Equations are given in the form of *left hand side == right hand side*.

A single variable or a list of variables can be also be specified. When there are several variables, the solution is given in terms of a list of rules. For example,

```
Solve[{y == 5 * x + 3, y == 6 * x - 2}, {y, x}]

{{y → 28, x → 5}}
```

When there are several solutions, *Solve* gives a list of them. For example,

```
z = 0;

Solve[z == w^2 - 4 * w + 3, w]

{{w → 1}, {w → 3}}
```

Solve will not always be able to get explicit solutions to equations. It will give the explicit solution if it can, then give a symbolic representation of the remaining solutions in terms

163

of root objects. If there are few symbolic parameters, you can the use *NSolve* to get
numerical approximations to the solutions.

9.5 Plotting

Mathematica contains a very powerful plotting interface and this overview will go over
the basics of plotting. Using Mathematica's Plot[ ] function there is a basic syntax to be
used:

Plot[function, {variable, variable_min, variable_max}]

For example, the command "Plot[y=x, {x,0,10}]" will output a graph of the function y=x
for a range of x from 0 to 10.
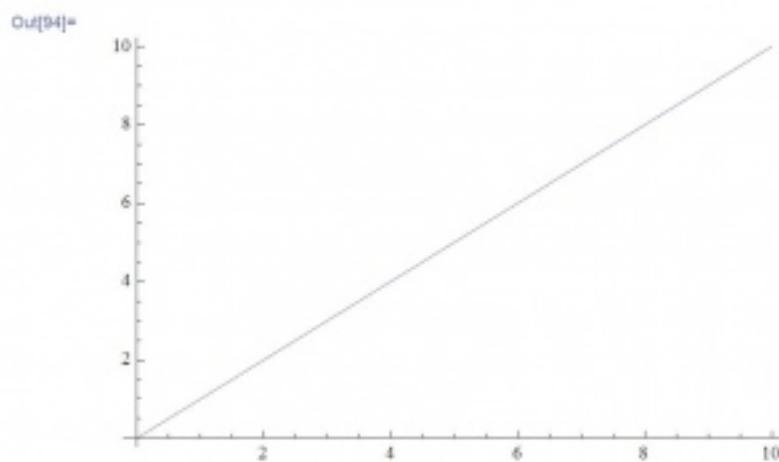
The function could also be previously defined and then referenced in the Plot [ ]
command, this functionality provides a lot of simplicity when referencing multiple
functions and helps prevent a typo from affecting your Plots.

y[x]==x

Plot[y[x],{x,0,10}]

To plot multiple functions on the same plot, use an array of functions using curly brackets
{ }, in place of the single one:

Plot[{y=x,y=2*x}, {x,0,10}]

164
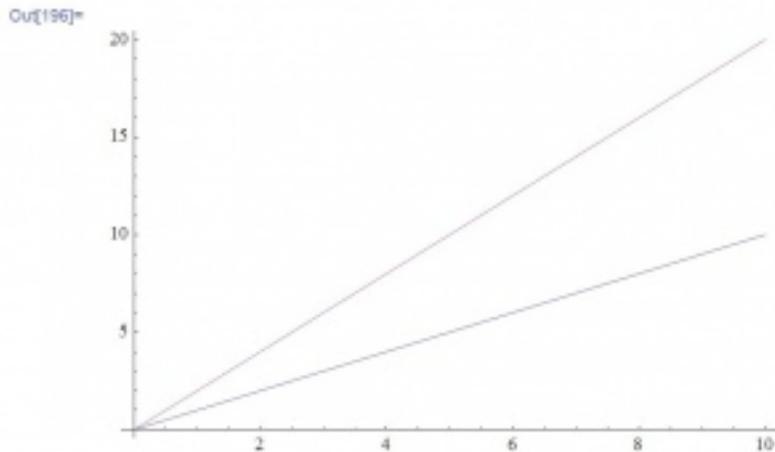
Another useful plotting option in Mathematica is the ListPlot function. To use this function, the following formatting should be used: ListPlot[Table[f(x),{x,min,max}]
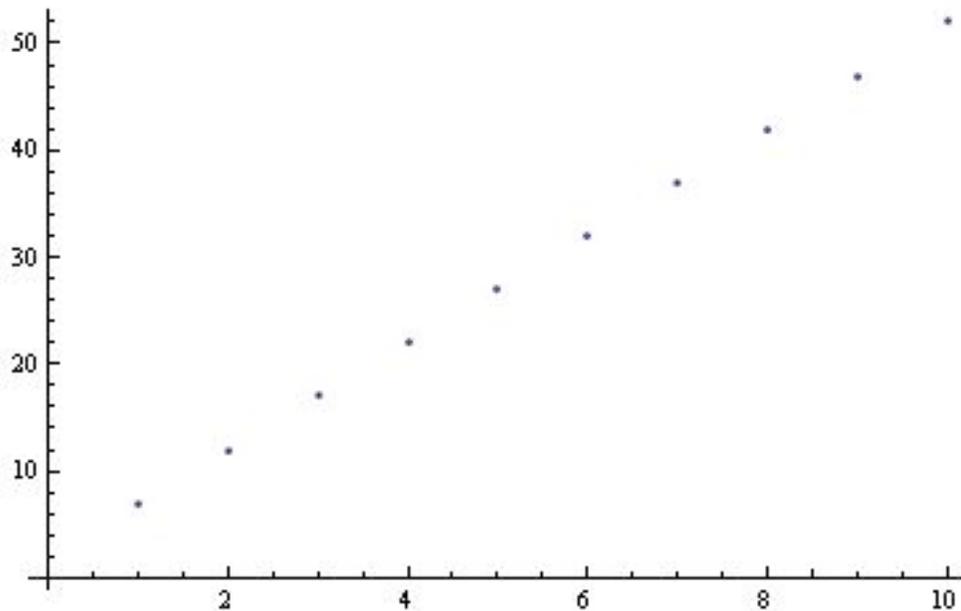
This is obviously only the basics, Mathematica provides many more options including colors, legends, line styles, etc. All of these extra features are well documented at Wolfram's Reference Site.

An additional piece of information: Type the following command on the line after the Plot command to determine the max value plotted. It will display the number associated with the absolute max of the function within the plotting range.

Max[Last/@Level[Cases[%,_Line,Infinity],{-2}]]

Within Mathematica, you have the option of solving ODEs numerically or by having Mathematica plot the solution.

Explicit solution:

The built-in function *DSolve* solves a differential equation or a list of differential equations for either one indepedent variable or a list of independent variables, which solves a partial differential equation. The most basic syntax for the function *DSolve* is *DSolve[eqn,y,x]*. For example,

```
DSolve[y'[x] + 3 y[x] + 4 == Sin[x], y[x], x]
```

$$\left\{\left\{y[x] \to e^{-3x} C[1] + \frac{1}{30} (-40 - 3 Cos[x] + 9 Sin[x])\right\}\right\}$$

The syntax for solving a list of differential equations is *DSolve[{eqn1,eqn2,...},{y1,y2,...},x]*, while the syntax for solving a partial differential equation is *DSolve[{eqn, y, {x1,x2,...}]*. For example,

```
DSolve[{y'[x] == 2 * y[x], z'[x] == 3 * z[x] + 4 * Sin[x]}, {y[x], z[x]}, x]
```

$$\left\{\left\{y[x] \to e^{2x} C[1], z[x] \to e^{3x} C[2] - \frac{2}{5} (Cos[x] + 3 Sin[x])\right\}\right\}$$

Note that differential equations must be stated in terms of derivatives, such as *y'[x]*. Boundary conditions can also be specified by giving them as equations, such as *y'[0]==b*.

Plotted Solution:

Solutions given by *DSolve* sometimes include integrals that cannot be carried out explicitly, thus *NDSolve* should be used. *NDSolve* finds a numerical solution to an ordinary differential equation within a specified range of the independent variable. The result will be given in terms of *InterpolatingFunction* objects, and thus the solution should be plotted. For example,

166

```
s = NDSolve[{y'[x] == 2 * y[x], y[0] == 1}, y, {x, 0, 30}]

{{y → InterpolatingFunction[{{0., 30.}}, <>]}}

Plot[Evaluate[y[x] /. s], {x, 0, 30}]
```

Note that a place holder was defined to store the *NDSolve* solution in order for the solution to be plotted. The general syntax for plotting is *Plot[Evaluate[y[x] /. s], {x, xmin, xmax}]*, where *s* is said placeholder. The plot below shows the solution to this example.
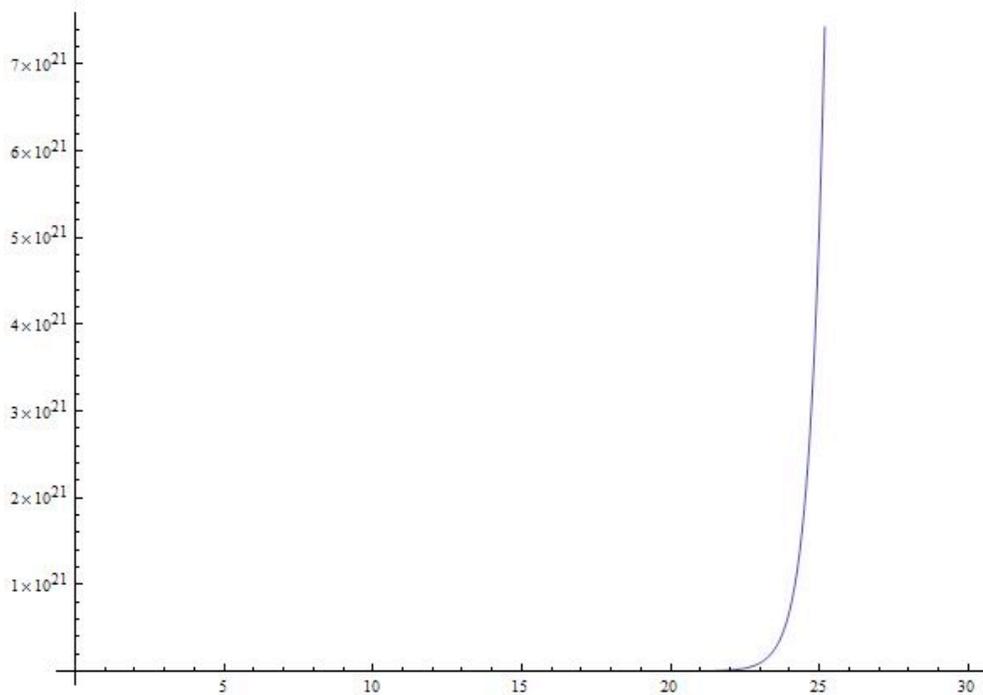
9.7 Another Example Solving ODEs

In this example the water level of a tank is controlled by adjusting the feed rate of water into the tank. The system is modeled by the following equation:

$$\frac{dV}{dt} = \frac{Fin}{2} - 2 * V$$

There is a PID controller on the valve but is not important in explaining the Mathematica syntax. Consider it just a different way to define Fin.

Parameters are assumed to be Vset = 10, Kc = 10, tauI = 0.1, and tauD = 1

For this problem the differential equation needs to be modeled with the given parameters in Mathematica. The problem asks what the steady state of the system is. This is solved for by substituting in a PID expression for Fin (not important in this context), setting the derivatives equal to zero, and solving for V and Xi(a made up variable for the PID controller). The following Mathematica code does all said things.

```
Fin = 1 + Kc (Vset - V[t]) + 1 / ti * xi[t] - td * V'[t];
eqns = {V'[t] == 0.5 * Fin - 2 * V[t], xi'[t] == Vset - V[t], V[0] == 0, xi[0] == 0};
param = {Kc → 10, ti → 0.1, td → 1, Vset → 10};
tmax = 10;
sol = NDSolve[eqns /. param, {V[t], xi[t]}, {t, 0, tmax}];
Plot[{V[t], xi[t]} /. sol, {t, 0, tmax}]
```
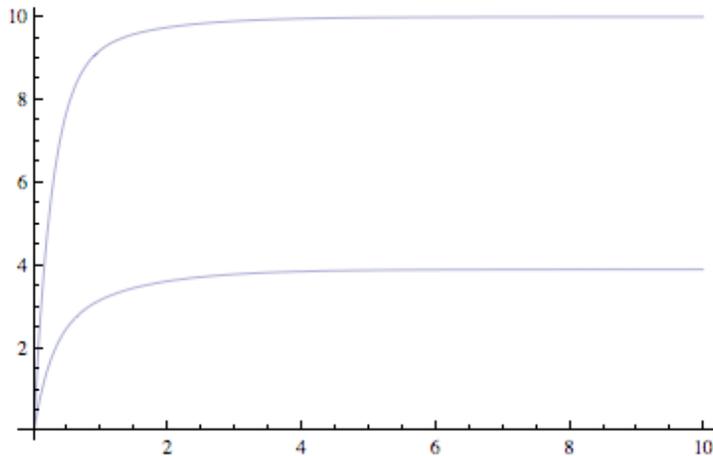


```
s = DSolve[eqns /. param, {V[t], xi[t]}, t];
Limit[s[[1]][[1]][[2]], t → Infinity]
Limit[s[[1]][[2]][[2]], t → Infinity]
(*OR*)
Solve[eqns /. param /. {V'[t] → 0, xi'[t] → 0}, {V[t], xi[t]}]
```

10.

3.9

{{xi[t] → 3.9, V[t] → 10.}}

The final line shows that the steady state values for the problem are V = Vset or 10 units of volume, and Xi (the made up variable for the PID controller) has a steady state value of 3.9

In this example there is syntax for defining parameters, applying those parameters to equations, setting up equations, and solving equations for steady state values. There is also syntax for plotting multiple equations on the same graph and how to evaluate multiple functions over time periods. Overall very useful for solving Controls Engineering problems in Mathematica.

168

It is very simple to create a matrix in Mathematica. Simply pick a name for the matrix (for this example we will use a) then simply enter each row inside of curly brackets {} with the individual values separated by commas, and the the row separated by a comma. The function MatrixForm[] will display a matrix in its matrix form rather than as a list of numbers.

EX: a = {{5, 2, 3}, {7, 9, 11}, {8, 13, 9}}

MatrixForm[a] displays:

5 2 3

7 9 11

8 13 9

You can refer to any specific position in the matrix by using the system:

matrix[[row],[column]]

EX: For the matrix above a[[1],[2]] a value of 2 would be returned.

9.9 Jacobians and Eigenvalues

The Jacobian is a helpful intermediate tool used to linearize systems of ODEs. Eigenvalues can be used with the Jacobian to evaluate the stability of the system. Mathematica can be used to compute the Jacobian and Eigenvalues of a system by the following steps.

1. Enter the system as individual equations

$$\frac{dx}{dt} = 3x^2 + 2y$$

$$\frac{dy}{dt} = x^2 - 3y^2$$

entered as

eq1 = 3*x^2+2y

eq2 = x^2-3*y^2

2. Compute the Jacobian with Mathematica's Derivate function

$$Jac = \begin{pmatrix} 6x & 2 \\ 2x & -6y \end{pmatrix}$$

computed by

Jac = {{D[eq1,x],D[eq1,y]},{D[eq2,x],D[eq2,y]}}

3. Use Mathematica's Eigenvalues function to evaluate the stability of the system

Begin by entering a desired value of {x,y} into the matrix (i.e. steady state values of x and y).

$$Jac = \begin{pmatrix} -2.31 & 2 \\ -0.77 & 1.33 \end{pmatrix}$$

Then call the Eigenvalue function.

Eigenvalues[Jac]

Result: λ={-1.82, 0.85}

The Jacobian can be computed as long as the matrix is square. Many other function types are acceptable when entering eq1 and eq2, including trigonometric and exponential functions.

9.10 User-Defined Functions

Mathematica also supports user-defined functions. Defining your own functions helps to simplify code, especially when lengthy expressions are used consistently throughout the code. The generic Mathematica syntax for defining a function is as follows:

f[a_,b_,c_]:=(a+b)/c

170

Here, the function name is "f" with three arguments: a, b, and c. The syntax ":=" assigns the expression on the right hand side to the function "f." Make sure that each of the arguments in the brackets on the left hand side are followed by underscores; this identifies the variables being called by the function as local variables (ones that are only used by the function itself). For this example, the function "f" will take three arguments (a, b, and c) and calculate the sum of a and b, and then divide by c.

After defining the function, you can use the function throughout your code using either real numbers, as the first example shows, or any variables. This is done by calling the function with the correct number of arguments. When referencing variables, if the variable already has an assigned value the value will be used, if it does not, Mathematica will display your function's results with the variables displayed. A few examples are shown below.

$$f[1, 3, 2] = \frac{1 + 3}{2} = 2$$

$$f[x, y, z] = \frac{x + y}{z}$$

9.11 Nonlinear Regression

Mathematica can be used to solve multiple variable, nonlinear regression problems. This form of regression can be useful to create models of how temperature or pressure sensors depend on different valves in a system. The NonLinearRegress function in Mathematica can be used to find a best-fit solution to a user defined functional form.

The NonLinearRegress function is not part of the standard package in Mathematica. For this reason, this function must be loaded into Mathematica before an attempt to use.

After this step, the NonlinearRegress is loaded and can be used to solve for a set of data. To use the this function, enter NonlinearRegress[data, expression, parameters, variables]. The best fit values for the parameters will be returned along with statistical information about how well this form fits the given data. An example of Nonlinear Regression in Mathematica is shown in the image below.

```
In[81]:= Needs["NonlinearRegression`"]
        NonlinearRegress[{{1, 2, 2}, {3, 2, 4}, {5, 4, 5}, {6.5, 4, 3}},
        a * x1 + b * x2 + c, {a, b, c}, {x1, x2}]|
```

Out[82]=

$\Big\{$BestFitParameters → {a → 0.16, b → 0.2, c → 2.28},

$$\text{ParameterCITable} → \begin{array}{c|ccc} & \text{Estimate} & \text{Asymptotic SE} & \text{CI} \\ \hline a & 0.16 & 1.12 & \{-14.0709, 14.3909\} \\ b & 0.2 & 2.32164 & \{-29.2992, 29.6992\} \\ c & 2.28 & 3.69345 & \{-44.6498, 49.2098\} \end{array}',$$

EstimatedVariance → 3.92,

$$\text{ANOVATable} → \begin{array}{l|ccc} & \text{DF} & \text{SumOfSq} & \text{MeanSq} \\ \text{Model} & 3 & 50.08 & 16.6933 \\ \text{Error} & 1 & 3.92 & 3.92 \\ \text{Uncorrected Total} & 4 & 54. & \\ \text{Corrected Total} & 3 & 5. & \end{array},$$

$$\text{AsymptoticCorrelationMatrix} → \begin{pmatrix} 1. & -0.904534 & 0.530669 \\ -0.904534 & 1. & -0.822871 \\ 0.530669 & -0.822871 & 1. \end{pmatrix},$$

$$\text{FitCurvatureTable} → \begin{array}{l|c} & \text{Curvature} \\ \hline \text{Max Intrinsic} & 0 \\ \text{Max Parameter-Effects} & 0 \\ 95. \% \text{ Confidence Region} & 0.0680875 \end{array}\Big\}$$

9.12 Probability Density Function

Often in controls, it is necessary to calculate the probability that a certain event will occur, given a particular known distribution. In these scenarios, it is helpful to know how to use the built-in probability density function in Mathematica. The syntax for using the probability density function to calculate the probability of observing a value x for a generic distribution *dist* is shown below.

PDF[*dist*,*x*]

A variety of distributions can be used in conjunction with the probability density function. These distributions include the normal distribution and the binomial distribution.

Normal Distribution

For a normal distribution with a mean μ and a standard deviation σ, use the following syntax to specify the distribution.

NormalDistribution[μ,σ]

172

To calculate the probability of observing a value *x* given this distribution, use the probability density function as described previously.

PDF[NormalDistribution[μ,σ],*x*]

To calculate the probability of observing a range of values from *x1* to *x2* given this distribution, use the following syntax.

NIntegrate[PDF[NormalDistribution[μ,σ],*x*],{*x,x1,x2*}]

For the theory behind the probability density function for a normal distribution, click here.

Binomial Distribution

For a binomial distribution with a number of trials *n* and a success probability *p*, use the following syntax to specify the distribution.

BinomialDistribution[*n,p*]

To calculate the probability of observing a successful trial exactly *k* times given this distribution, use the probability density function as described previously.

PDF[BinomialDistribution[*n,p*],*k*]

To calculate the probability of observing between a range of *k1* number of successes and *k2* number of successes given this distribution, use the following syntax.

NIntegrate[PDF[BinomialDistribution[*n,p*],*k*],{*k,k1,k2*}]

Alternatively, binomial distribution can be found using a user defined function. Define the equations as follow: binom[nn_,kk_,pp_]:= nn!/(kk! (nn-kk)!) pp^kk (1-p)^(nn-kk)

where:

number of independent samples = nn

number of events = kk

probability of the event = pp

To find the odds of getting 5 heads out of 10 coin tosses, assuming the probability of head = 0.5, substitute the values for nn, kk and pp as following: binom[10,5,0.5]

Using the same scenario, but instead for the odds of getting 5 or more heads in the 10 tosses, the function can be used as: Sum[binom[10,i,0.5],{i,5,10}]

For the theory behind the probability density function for a binomial distribution, click [here](#).

9.13 Manipulate Function

Often in controls, it is useful to manipulate certain variables dynamically and see the result in a plot or table. This is particularly useful in setting the constants in tuning PID controllers. The Manipulate function in Mathematica lets you do this. You can use the manipulate function in conjunction with the Plot command. The general syntax for using the manipulate function is as follows:

Manipulate[expr,{u,umin,umax}]

Where *expr* is the function that you want to add controls to and *u* is the variable you wish to manipulate.

Manipulate a function:

You can change the variable *x* in the function *s*. Moving the bar lets you change the variable *x*.

```
Manipulate[s = x^3 + 2 x^2 - 3 x + 6, {x, -10, 10, 1}]
```


Manipulate a plot:

You can also manipulate a plot. The slider bar lets change the variable *n* and the plot changes accordingly.

`Manipulate[Plot[Tan[n*x-1], {x, 0, 10}], {n, 0, 1}]`
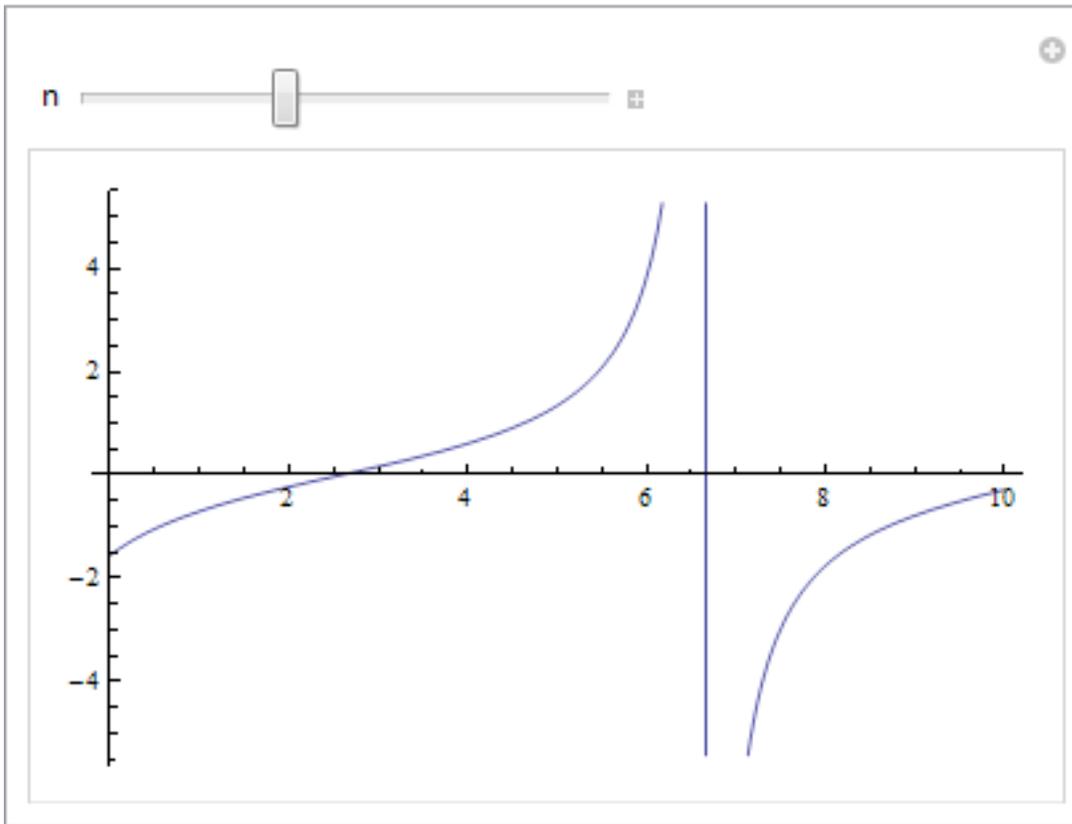
The Mathematica file used to create this example, can be downloaded below.

Manipulate Command Example

For more information regarding manipulate, please see the Wolfram's Reference Site.

[Note - need to add manipulate for the Bode plots - R.Z.]

9.14 Adding Gridlines

```
"PLOTTING WITH GRIDLINES
Make your graphs more readable and attractive!

TO ADD STANDARD GRIDLINES TO YOUR PLOT:
        use GridLines->Automatic

Other gridline options:
IF YOU HAVE A SINUSOIDAL GRAPH:
        use Gridlines->{{Pi,2Pi},{-1,-.5,.5,1}}
IF YOU WANT TO DRAW GRIDLINES AT SPECIFIC PLACES:
        use Gridlines->{{x1,x2},{y1,y2}}"

y1 = t^2
Plot[{y1}, {t, 0, 10}, PlotStyle → Thick, PlotRange → Full, Frame → True, GridLines → Automatic]
```

```
PLOTTING WITH GRIDLINES
Make your graphs more readable and attractive!

TO ADD STANDARD GRIDLINES TO YOUR PLOT:
        use GridLines->Automatic

Other gridline options:
IF YOU HAVE A SINUSOIDAL GRAPH:
        use Gridlines->{{Pi,2Pi},{-1,-.5,.5,1}}
IF YOU WANT TO DRAW GRIDLINES AT SPECIFIC PLACES:
        use Gridlines->{{x1,x2},{y1,y2}}
```

t^2
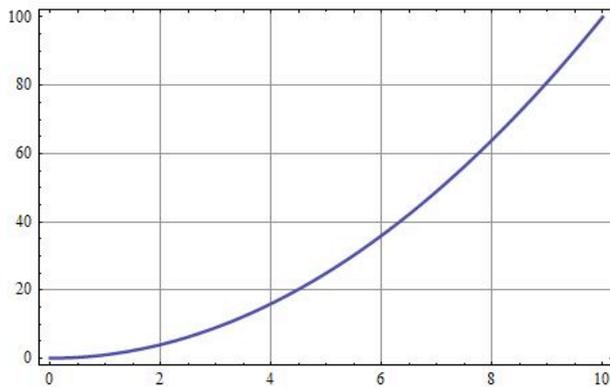
9.15 Plotting Log-Log Graphs

```
"CREATING LOG-LOG PLOTS

Sometimes when graphs are in the form of y = a*x^b, it is more visually appealing for
  the graph to be plotted on a log-log plot. A log-log plot uses a logarithmic scale
  for both axes. It is so simple to create log-log plots in Mathematica; use the simple
  steps below to create beautiful, functional graphs!"

omega = 4 * t^5;
LogLogPlot[AR, {t, 0.01, 100}, PlotStyle → {Thick}, GridLines → Automatic]
```
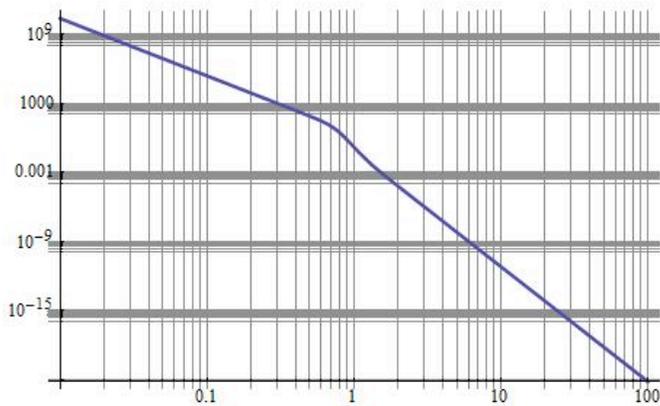
```
CREATING LOG-LOG PLOTS

Sometimes when graphs are in the form of y = a*x^b, it is more visually
  appealing for the graph to be plotted on a log-log plot. A log-log plot uses
  a logarithmic scale for both axes. It is so simple to create log-log plots in
  Mathematica; use the simple steps below to create beautiful, functional graphs!
```

"Additional Features of Mathematica"
"The following code demonstrates how to plot multiple curves in the same
  space in Mathematica. Details are also outlined for customizing axis labels,
  plot ranges and finding minimum/maximum values of a function for a certain range."

"Define the functions. Here 'a', 'b' and 'c' are the dependent variables and
  are functions of 'x'"
a = x^2 + 1
b = (x - 10)^2 + 1
c = x^2 + 10

Define the functions. Here 'a', 'b' and
  'c' are the dependent variables and are functions of 'x'

"To plot multiple functions in one graphing space, list the function names and
  separate with commas inside { }. Here the PlotRange command takes arguments
  for the 'x' range and the 'y' range in { }. The axes can be labeled using
  'AxesLabel' and large, bold style can be used."

Plot[{a, b, c}, {x, -10, 10}, PlotStyle → Thick, PlotRange → {{-10, 15}, {0, 100}},
  Axes → True, AxesLabel → {Style[x, Large, Bold], Style[y, Large, Bold]}]

"FindMinimum takes as arguments a function, then an independent variable range"
FindMinimum[b, {x, -10, 10}]
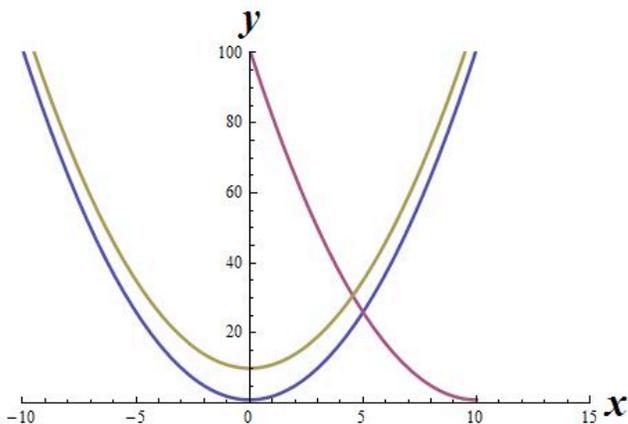
"Results from FindMinimum:"

{1., {x → 10.}}

"Conversely, use FindMaxmimum to find the maximum value of a function for a given range."

[It would be good to replace this images with actual text - R.Z.]

Chapter 3. Sensors and Actuators

More information on sensors and actuators at [ECOSSE](#)

Section 1. Control systems: Measurement Devices

Christian Hatfield, Varun Kaushik, Alon Mandel

1.1 Control Systems: Industrial Applications

Many control systems are used today in a large number of industries consisting of applications from all kinds. The common factor of all control types is to sustain a desired outcome that may change during a chemical reaction or process. The most common control type used today in industry is a PID controller (proportional, integral, derivative), which allows the operator to apply different control techniques that can be used to achieve different settings in an experiment or process. A PID controller can be used in two main control mechanisms that include feed back and feed forward. The purpose of this article is to provide examples of common industrial control systems that apply different control architectures.

1.2 Temperature control: Thermocouple

A thermocouple is a device to measure and control temperature within a system. They are used in a wide variety of industrial applications (gas turbines, chemical reactors, exhaust, chemical manufacturing etc) due to their low cost and portability. The fundamental working principle for thermocouple operations is the Seebeck Effect. Mostly, thermocouples operate in a P or a PID control mode. In order to measure temperature between two points, the thermocouple employs two metallic ends (made from different alloys). When the two conductor ends are exposed to a thermal gradient, they generate a voltage between them. This voltage drop gives rise to the temperature measurement output that a thermocouple provides. Depending on the types of alloys in both conductor ends, and the magnitude of the thermal gradient, thermocouples can measure temperature differences between 1-23000C. They can operate in feedback loops or feed forward loops. Thermocouples are mostly digital control units. Thermocouple prices started around $100 and cost up to $2500 for more accurate, and self-calibrated models.

1.3 Pressure Control: Pressure Switch

A pressure switch is a device that controls systems against pressure drops or pressure spikes. The most basic types of pressure switches work on an ON-OFF basis, but can also be manufactured to work in PID mode. The fundamental method of operation is to set the "Set-Pressure" to a given quantity. This deactivates the pressure switch from the circuit connecting it to the control valve upstream. If, at any point during the process, the pressure rises past the set-point, the switch is activated and completes the circuit, thus shutting off the control valve. Pressure switches can be hydraulic, or pneumatic based (air-based pressure). One common application of a pressure switch in the industry is to protect PD (positive-displacement) pumps from over-pressurization. A PD pump can generate very high pressures if not controlled by a pressure switch; thus setting a pressure switch inline with a PD pump will prevent over pressurization since it shuts off the control valve. Pressure switches are common in any industry since all require pressurization of certain components during manufacturing, processing or refining stages. They sell between $200-$2000 depending on the magnitude of the set-point required for protection.

1.4 Composition Control: Ratio Control

A ratio controller is used to ensure that two or more process variables such as material flows are kept at the same ratio even if they are changing in value. Control modes can be operated in different types, but mostly feedback PI controller is used for ratio control. In industrial control processes, ratio control is used in the following processes: burner/air ratio, mixing and blending two liquids, injecting modifiers and pigments into resins before molding or extrusion, adjusting heat input in proportion to material flow. The prices for industrial ratio controllers start around $500 and increase depending on the sensitivity of the unit to different magnitudes of compositional changes and size of the equipment to be annexed to.

1.5 Level Control: Level Switches

Level controls are used to monitor and regulate the liquid level in industrial vessels. There are many various sensors used in level control including ultrasonic, lasers, and floatation sensors. They all work on the same general principle. A sensor measures the distance from the base of the vessel to the top of the liquid level, mainly by using sonar waves or a laser beam. Based on the time it takes for the wave or for the laser beam to return to the emitting source, the controller sends information to change or maintain the level. The magnetic float control reads on a sensor located on the wall of the vessel and sends that information to the controller. Some examples of level control in industry are: maintaining the liquid level in a distillation column, protecting from overflow, and measuring the amount of product in storage tanks. Level sensors/controls vary in price based on the type and the accuracy required. Below is a picture of a magnetic float level control.

1.6 Flow Control: Flow Meters

Flow controls are used to regulate the flow of a liquid or gas through a system. The main form of flow control is a valve. There are many different types of valves, but they all change flow rates by opening or closing based on what amount is needed. A flow sensor reads the flow rate, and a controller will operate the valve to increase or decrease the flow. Some basic types of flow sensors are rotameters and digital mass meters. These controls are used in all forms of industry to control flows including water treatment, product measurement, and fuel flow to furnaces. The low price end for flow sensors is typically around $500, and depending on the size of the process, and the process material, the price can range into the thousands. Below are two pictures. The first is a manual control valve that can be opened or shut to regulate flow. The second is a valve that can be used in an automatic control system to regulate flow.

Section 2. Temperature Sensors

Note: ⊞Video lecture available for this section!

Authors: (14 September 2006) Ardemis Boghossian, James Brown, Sara Zak

Stewards: (11 September 2007) Arthur L. Edge III, Kathryn Meintel, Renu Rao, Kaveh Saba

- First round reviews for this page
- Rebuttal for this page

2.1 Introduction

Temperature sensors are vital to a variety of everyday products. For example, household ovens, refrigerators, and thermostats all rely on temperature maintenance and control in order to function properly. Temperature control also has applications in chemical engineering. Examples of this include maintaining the temperature of a chemical reactor at the ideal set-point, monitoring the temperature of a possible runaway reaction to ensure the safety of employees, and maintaining the temperature of streams released to the environment to minimize harmful environmental impact.

While temperature is generally sensed by humans as "hot", "neutral", or "cold", chemical engineering requires precise, quantitative measurements of temperature in order to accurately control a process. This is achieved through the use of temperature sensors, and temperature regulators which process the signals they receive from sensors.

From a thermodynamics perspective, temperature changes as a function of the average energy of molecular movement. As heat is added to a system, molecular motion increases and the system experiences an increase in temperature. It is difficult, however, to directly measure the energy of molecular movement, so temperature sensors are generally designed to measure a property which changes in response to temperature. The devices are then calibrated to traditional temperature scales using a standard (i.e. the boiling point of water at known pressure). The following sections discuss the various types of sensors and regulators.

2.2 Temperature Sensors

Temperature sensors are devices used to measure the temperature of a medium. There are 2 kinds on temperature sensors: 1) contact sensors and 2) noncontact sensors. However, the 3 main types are thermometers, resistance temperature detectors, and thermocouples. All three of these sensors measure a physical property (i.e. volume of a liquid, current

through a wire), which changes as a function of temperature. In addition to the 3 main types of temperature sensors, there are numerous other temperature sensors available for use.

Contact Sensors

Contact temperature sensors measure the temperature of the object to which the sensor is in contact by assuming or knowing that the two (sensor and the object) are in thermal equilibrium, in other words, there is no heat flow between them.

Examples (further description of each example provide below)

- Thermocouples
- Resistance Temperature Detectors (RTDs)
- Full System Thermometers
- Bimetallic Thermometers

Noncontact Sensors

Most commercial and scientific noncontact temperature sensors measure the thermal radiant power of the Infrared or Optical radiation received from a known or calculated area on its surface or volume within it.

An example of noncontact temperature sensors is a pyrometer, which is described into further detail at the bottom of this section.

2.2.1 Thermometers

Thermometers are the most common temperature sensors encountered in simple, everyday measurements of temperature. Two examples of thermometers are the Filled System and Bimetal thermometers.

Filled System Thermometer

The familiar liquid thermometer consistsof a liquid enclosed in a tube. The volume of the fluid changes as a function of temperature. Increased molecular movement with increasing temperature causes the fluid to expand and move along calibrated markings on the side of the tube. The fluid should have a relatively large thermal expansion coefficient so that small changes in temperature will result in detectable changes in volume. A common tube material is glass and a common fluid is alcohol. Mercury used to be a more common fluid until its toxicity was realized. Although the filled-system thermometer is the simplest and cheapest way to measure temperature, its accuracy is limited by the calibration marks along the tube length. Because filled system thermometers are read visually and don't produce electrical signals, it is difficult to implement them in process controls that rely heavily on electrical and computerized control.

Bimetal Thermometer

In the bimetal thermometer, two metals (commonly steel and copper) with different thermal expansion coefficients are fixed to one another with rivets or by welding. As the temperature of the strip increases, the metal with the higher thermal expansion coefficients expands to a greater degree, causing stress in the materials and a deflection in the strip. The amount of this deflection is a function of temperature. The temperature ranges for which these thermometers can be used is limited by the range over which the metals have significantly different thermal expansion coefficients. Bimetallic strips are often wound into coils and placed in thermostats. The moving end of the strip is an electrical contact, which transmits the temperature thermostat.

2.2.2 Resistance Temperature Detectors

A second commonly used temperature sensor is the resistance temperature detector (RTD, also known as resistance thermometer). Unlike filled system thermometers, the RTD provides an electrical means of temperature measurement, thus making it more convenient for use with a computerized system. An RTD utilizes the relationship between electrical resistance and temperature, which may either be linear or nonlinear. RTDs are traditionally used for their high accuracy and precision. However, at high temperatures (above 700°C) they become very inaccurate due to degradation of the outer sheath, which contains the thermometer. Therefore, RTD usage is preferred at lower temperature ranges, where they are the most accurate.

There are two main types of RTDs, the traditional RTD and the thermistor. Traditional RTDs use metallic sensing elements that result in a linear relationship between temperature and resistance. As the temperature of the metal increases, increased random molecular movement impedes the flow of electrons. The increased resistance is measured as a reduced current through the metal for a fixed voltage applied. The thermistor uses a semiconductor sensor, which gives a power function relationship between temperature and resistance.

RTD Structure

A schematic diagram of a typical RTD is shown in Figure 1.
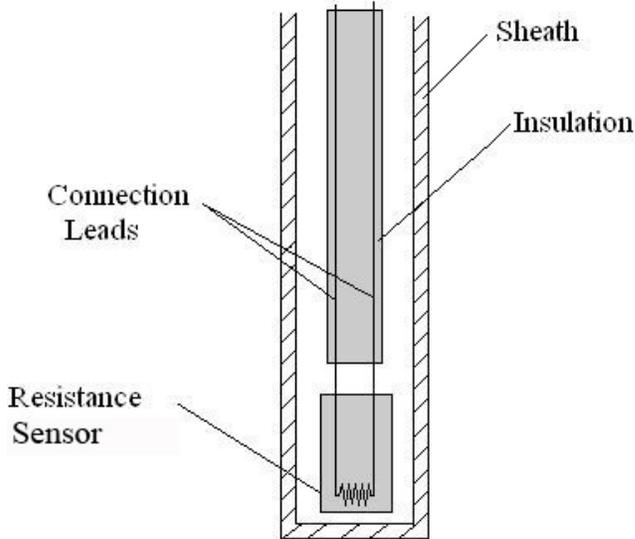
Figure 1. Schematic Diagram of Resistance Temperature Structure

As shown in Figure 1, the RTD contains an outer sheath to prevent contamination from the surrounding medium. Ideally, this sheath is composed of material that efficiently conducts heat to the resistor, but resists degradation from heat or the surrounding medium.

The resistance sensor itself is responsible for the temperature measurement, as shown in the diagram. Sensors are most commonly composed of metals, such as platinum, nickel, or copper. The material chosen for the sensor determines the range of temperatures in which the RTD could be used. For example, platinum sensors, the most common type of resistor, have a range of approximately -200°C – 800°C. (A sample of the temperature ranges and resistances for the most common resistor metals is shown in Table 1). Connected to the sensor are two insulated connection leads. These leads continue to complete the resistor circuit.

Table 1. Common Metal Temperature and Resistance Ranges

Element Metal	Temperature Range	Base Resistance	TCR(Ω/Ω/°C)
Copper	-100 – 260 °C	10 Ω at 0 °C	0.00427
Nickel	-100 – 260 °C	120 Ω at 0 °C	0.00672
Platinum	-260 – 800 °C	100 Ω at 0 °C	0.003916

There are 4 major categories of RTD sensors. There are carbon resistors, film thermometers, wire-wound thermometers and coil elements.

Carbon resisters are the most commonly used. They are inexpensive and are accurate for low temperatures. They also are not affected by hysteresis or strain gauge effects. They are commonly used by researchers.

Film thermometers have a very thin layer of metal, often platinum, on a plate. This layer is very small, on the micrometer scale. These thermometers have different strain gauge effects based on what the metal and plate are composed of. There are also stability problems that are dependent on the components used.

In wire-wound thermometers the coil gives stability to the measurement. A larger diameter of the coil adds stability, but it also increases the amount the wire can expand which increases strain and drift. They have very good accuracy over a large temperature range.

Coil elements are similar to wire-wound thermometers and have generally replaced them in all industrial applications. The coil is allowed to expand over large temperature ranges while still giving support. This allows for a large temperature range while decreasing the drift.

RTD Operation

Most traditional RTD operation is based upon a linear relationship between resistance and temperature, where the resistance increases with temperature. For this reason, most RTDs are made of platinum, which is linear over a greater range of temperatures and is resistant to corrosion. However, when determining a resistor material, factors such as temperature range, temperature sensitivity, response time, and durability should all be taken into consideration. Different materials have different ranges for each of these characteristics.

The principle behind RTDs is based upon the Callendar – Van Dusen equation shown below, which relates the electrical resistance to the temperature in °C. This equation is merely a generic polynomial that takes form based upon experimental data from the specific RTD. This equation usually takes on a linear form since the coefficients of the higher-order variables (a_2, a_3, etc.) are relatively small.

$$R_T = R_0(1 + a_1 T + a_2 T^2 + a_3 T^3 + a_4 T^4 + ... + a_n T^n) \qquad (1)$$

R_T: Resistance at temperature T, in ohms
R_0: Resistance at temperature = 0°C, in ohms
a_n: Material's resistance constant, in $°C^{n-1}$

Another type of RTD is the thermistor, which operates based upon an exponential relationship between electrical resistance and temperature. Thermistors are primarily composed of semiconductors, and are usually used as fuses, or current-limiting devices.

Thermistors have high thermal sensitivity but low temperature measuring ranges and are extremely non-linear. Instead of the Callendar - Van Dusen equation, the thermistor operates based upon the nonlinear equation, equation (2), shown in degrees K.

$$R_T = R_0 exp(b(\tfrac{1}{T} - \tfrac{1}{T_0})) \qquad (2)$$

T_0: Initial temperature, usually set at 298K
b: Material's temperature coefficient of resistance, in K

Errors associated with resistance thermometers will occur due to the individual or collective efforts of: defective insulation, contamination of the resistor, or insecure lead wire connections.

2.2.3 Thermocouples

Another temperature sensor often used in industry is the thermocouple. Among the various temperature sensors available, the thermocouple is the most widely used sensor. Similar to the RTD, the thermocouple provides an electrical measurement of temperature.

Thermocouple Structure

The thermocouple has a long, slender, rod-like shape, which allows it to be conveniently placed in small, tight places that would otherwise be difficult to reach. A schematic diagram of a typical thermocouple is shown in Figure 2.

Figure 2. Schematic Diagram of Thermocouple Structure

As illustrated in Figure 2, the thermocouple contains an outer sheath, or thermowell. The thermowell protects the contents of the thermocouple from mechanical and chemical damage.

188

Within the thermowell lies two metal wires each consisting of different metals. Various combinations of materials are possible for these metal wires. Three common thermocouple material combinations used for moderate temperature measurements are the Platinum-Rhodium, Iron-Constantan, and Chromel-Alumel metal alloys. The metal alloys chosen for a thermocouple is based upon the emf value of the alloy pair at a given temperature. Sample emf values for the most common materials at various temperatures are shown in Table 2. For a given pair of materials, the two wires are connected at one end to form a junction. At the other end, the two wires are connected to a voltage measuring device. These ends of the wires are held at a different reference temperature.

Table 2. Common Metal Temperature and Emf Values

Alloy Type	Emf Value at 20 °C	Emf Value at 50 °C	Emf Value at 100 °C
Platinum-Rhodium	0.113 mV	0.299 mV	0.646 mV
Iron-Constantan	1.019 mV	2.585 mV	5.269 mV
Chromel-Alumel	0.798 mV	2.023 mV	4.096 mV

Various methods are used to maintain the reference temperature at a known, constant temperature. One method consists of placement of the reference junction within either an ice bath or oven maintained at a constant temperature. More commonly, the reference temperature is maintained electronically. Though not as stable as an ice bath, electronically controlled reference temperatures are more convenient for use. Reference temperatures could also be maintained through temperature compensation and zone boxes, which are regions of uniform temperature. The voltage difference across the reference junction is measured and sent to a computer, which then calculates the temperature with this data.

Thermocouple Operation

The main principle upon which the thermocouple function is based on is the difference in the conductivities of the two wire materials that the thermocouple is made of, at a given temperature. This conductivity difference increases at higher temperatures and conversely, the conductivity difference decreases at lower temperatures. This disparity results in the thermocouples being more efficient and useful at higher temperatures. Since the conductivity difference is small at lower temperatures and thus more difficult to detect, they are inefficient and highly unreliable at low temperatures.

The conductivity difference between the two wires, along with a temperature difference between the two junctions, creates an electrical current that flows through the thermocouple. The first junction point, which is the point at which the two wires are connected, is placed within the medium whose temperature is being measured. The second junction point is constantly held at a known reference temperature. When the temperature of the medium differs from the reference temperature, a current flows through the circuit. The strength of this current is based upon the temperature of the medium, the reference temperature, and the materials of the metal wires. Since the reference temperature and materials are known, the temperature of the medium can be determined from the current strength.

Error associated with the thermocouple occurs at lower temperatures due to the difficulty in detecting a difference in conductivities. Therefore, thermocouples are more commonly used at higher temperatures (above -125°C) because it is easier to detect differences in conductivities. Thermocouples are operable over a wide range of temperatures, from -200°C to 2320°C, which indicates its robustness and vast applications. Thermocouples operate over this wide range of temperatures, without needing a battery as a power source. It should be noted that, the wire insulation might wear out over time by heavy use, thus requiring periodical checks and maintenance to preserve the accuracy of the thermocouple.

To determine the temperature of the medium from the current strength, the emf or voltage values of the current and of the wire materials at the reference temperatures must be known. Often, the measured temperature can be found by using standard thermocouple tables. However, these tables are often referenced at 0°C. To correct for this different reference temperature, equation (3) can be used to calculate the temperature from a given current.

$$\xi_{T_1,T_3} = \xi_{T_1,T_2} + \xi_{T_2,T_3} \qquad (3)$$

ξ : emf of an alloy combination generated at two different temperatures
T_1: temperature of the medium whose temperature is to be determined
T_2: reference temperature of the thermocouple
T_3: reference temperature of the standard thermocouple table, which in this case is 0°C

Once the emf between two alloys is calculated relative to a reference temperature when T3 is 0°C, the standard thermocouple table can be used to determine the temperature T1 of the medium. This temperature is usually automatically displayed on the thermocouple.

Apart from the common occurrence of the thermocouples being placed in the fluid to measure temperature change, thermocouples can be also embedded in solids with excellent results. This is highly effective while establishing the different thermal

190

properties for a solid. The heat transfer to the thermocouple will now be in the form of conductive heat transfer. As a result, this setup would be very similar to heat conduction in series, since the thermocouple is almost always made from a different material then the actual solid. Such discrepancies depend on the manner in which the thermocouple is embedded in the solid and should be taken into account when the thermal properties are being calculated and analyzed. One example is shown in the photo below.
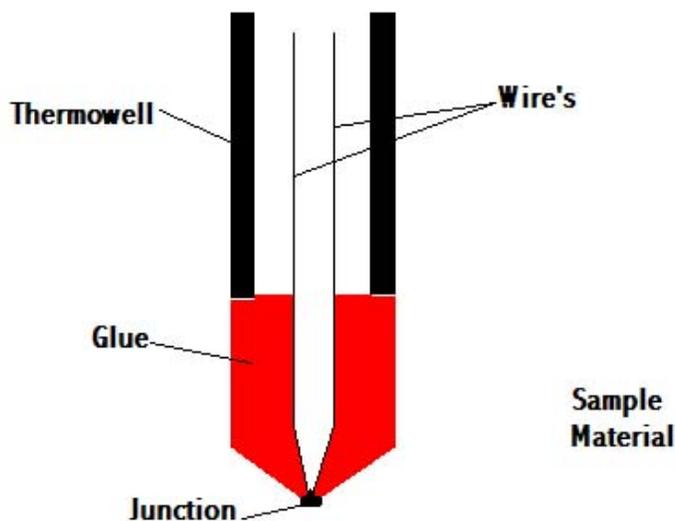
Figure 3. Schematic diagram of how the thermocouple function

Laws for thermocouples

Law of homogenous material

o If all the wires and the thermocouple are made of the same material, temperature changes in the wiring do not affect the output voltage. Thus, need different materials to adequately reflect the temperature.

Law of intermediate materials

o The sum of all the thermoelectric forces in a circuit with a number of dissimilar materials at a uniform temperature is zero. This implies that if a third material is added at the same temperature, no net voltage is generated by the new material.

Law of successive or intermediate temperatures

o If two dissimilar homogeneous materials produce thermal emf1 when the junctions are at T1 and T2 and produce thermal emf2 when the junctions are at T2 and T3 , the emf generated when the junctions are at T1 and T3 will be emf1 + emf2 .

Application

Steel industry

o Monitor temperature and chemistry throughout the steel making process

Heating appliance safety

o Thermocouples in fail-safe mode are used in ovens and water heaters to detect if pilot flame is burning to prevent fire and health hazard

Manufacturing

o Used for testing prototype electrical and mechanical apparatus

Process plants

o Chemical production plants and refineries use computer programs to view the temperature at various locations. For this situation, a number of thermocouple leads are brought to a common reference block.

2.2.4 Pyrometers

Unlike the thermometer, RTD and the thermocouple, pyrometers (non-contact temperature sensors) measures the amount of heat radiated, rather than the amount of heat conducted and convected to the sensor. Various types of pyrometers, such as total radiation and photoelectric pyrometers, exist. Below is a schematic of an optical pyrometer in Figure 4.

Optical Wedge-Type Pyrometer
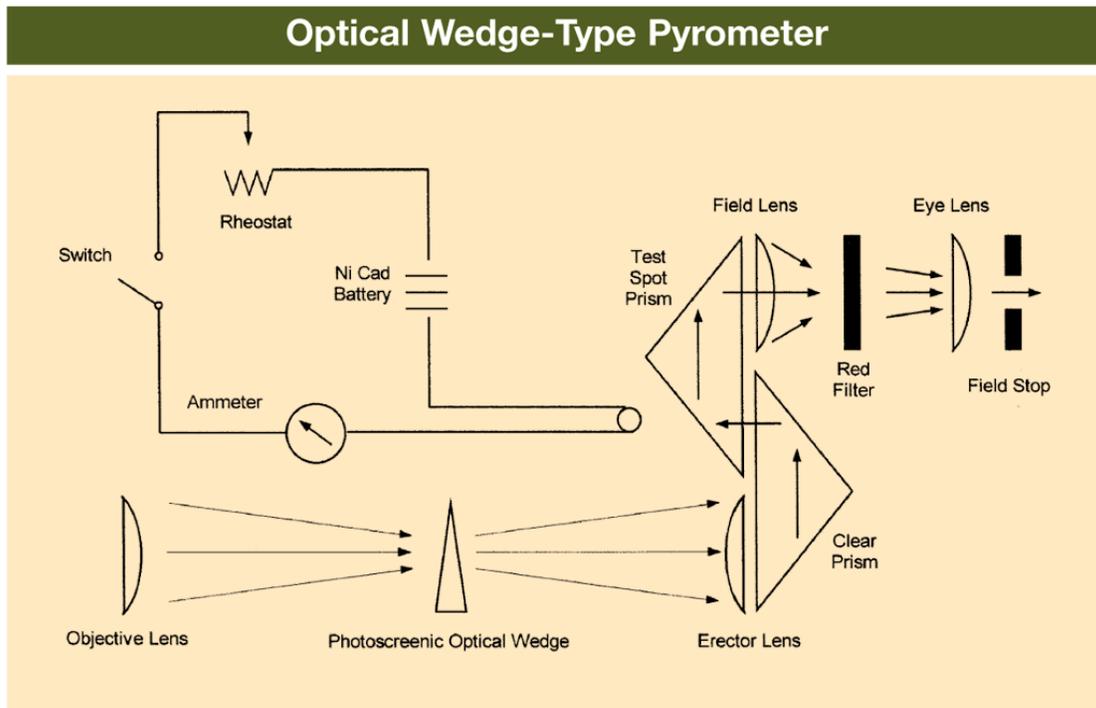
Figure 4. Schematic diagram of an optical pyrometer

These pyrometers differ in the type of radiation they measure. There are many factors that influence the amount of radiated heat detected, thus there are many assumptions that must be made regarding the emissivity, or the measure of the manner in which heat is radiated, of the object. These assumptions are based upon the manner in which heat is radiated as well as the geometry of the object. Because temperature is dependent on the emissivity of a body, these assumptions regarding the emissivity introduce uncertainties and inaccuracies in the temperature readings. Therefore, because of the error associated with them, pyrometers are not often used in industry.

Table 3. Summary of Temperature Sensors

	Type	Function	Operating Range (C)	Advantage	Disadvantage
Thermometer	Liquid Filled	Calibrated with markings along the tube to provide a visual reading of the temperature	–200 C to 260 C	Cheap, good for everyday usage	Limited range, can't connect to temperature regulation systems. Visual readouts lead to limited accuracy.
	Bimetal		–10 C to 110 C	Good for quick checks on temperatures, no batteries, inexpensive	
RTD	Traditional (uses metallic sensing elements)	Provides an electrical means of temperature measurement, making it compatible with a computerized system	–200C to 700C	Easy to recalibrate, provides readings with two significant figures, remain stable for long periods of time.	Smaller temperature range than thermocouples, higher start up costs, sensitive to environmental disturbances
	Thermistor (uses a semiconductor sensor)				
Thermocouple		Consists of two metal wires connected to a voltage source to provide an electrical means of	–200C to 2320 C depending on the metals chosen for the two wires	No batteries needed, quick, can be used for wide temperature ranges	Can lose their accuracy after being used for a long time due to wear in the insulation between wires

193

There are a few different types of pyrometers. There are optical and radiation pyrometers.

How Optical Pyrometers Work:

• Compares the color of visible light given off by the object with that of a electrically heated wire

• The wire can be preset to a certain temperature

• The wire can be manually adjusted to compare the two objects

How Radiation Pyrometers Work:

• This sensor works by measuring the radiation (infrared or visible light) that an object gives off

• The radiation heats a thermocouple in the pyrometer which in turn induces a current

• The larger the current induced, the higher the temperature is

Pyrometers are usually used at very high temperatures, but can be used at colder temperatures as well. There are lots of industrial applications to pyrometers. Plant operators can use pyrometers to get a sense of what temperature certain processes are running at. The downside to pyrometers is that they are not very accurate as thermocouples or RTD sensors are. This is because they rely on quantifying colors of light.

2.3 Temperature Regulators

Temperature regulators, also known as temperature control valves (TCVs), physically control, as well as measure, temperature. Temperature regulators are not capable of directly maintaining a set value; instead, they relate the load (in this case the valve opening) with the control (temperature measurement). These regulators are most useful when temperature is correlated to a flow of a substance. For example, a TCV may be used to control the temperature of an exothermic reaction that requires constant cooling. The TCV measures the temperature of the reaction and, based upon this temperature, either increases or decreases the flowrate of cooling fluid to adjust the temperature of the reaction. Similarly, the regulator could be used to adjust the flow amount of steam, which is typically used to heat a substance. Therefore, by adjusting flowrate, the regulator can indirectly adjust temperature of a given medium.

2.3.1 Regulator Structure

The structure of a typical thermal regulator consists of four main parts, as shown in Figure 3. The temperature detecting element, which in most cases is a temperature sensor,

as described above, sends either an electrical or mechanical signal through the connector to the actuator. The actuator then uses this signal to act upon the power source, which determines the position of the valve. (This will be further described in the next section.)
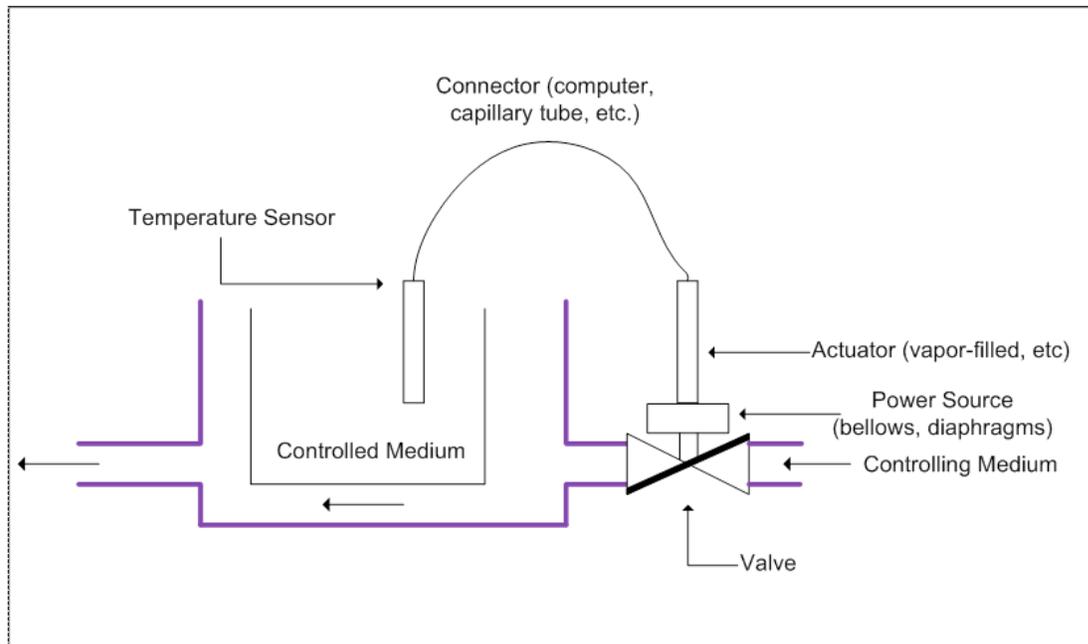
Note: The actuator consists of the capillary tubing, the power source, and the adjustment.

Figure 3. Schematic Diagram of Temperature Regulator Structure

2.3.2 Regulator Operation

The temperature regulator operates based upon a mechanical means of temperature control. As previously mentioned, the bulb of the regulator is typically filled with a heat conducting substance. Due to the thermal expansion properties of this substance, the substance expands as the temperature increases. This expansion causes a change in the pressure of the actuator, which correlates to the temperature of the medium. This pressure change repositions a valve on the regulator, which controls the flowrate of a coolant. The temperature of the medium is then altered by the change in the flowrate of this coolant.

2.3.3 Types of Temperature Regulators

Though all regulators have the same basic build and purpose, they exist in a variety of forms. In particular, these regulators vary in four primary ways: temperature detecting elements, temperature detector placement, actuator type, and valve type.

Temperature Detecting Elements

Most temperature regulation systems use thermocouples or RTDs as temperature sensing devices. (Described above) For these systems, the connector is a computer. The sensors send an electric signal to the computer, which calculates the temperature. The computer then compares the temperature measured by the sensor to a programmed set-point temperature, thus determining the required pressure in the actuator. The pressure in the actuator changes position of the power source (diaphragm or bellows), which consequently changes the flowrate through the valve.

Some temperature regulation systems use a filled bulb as a temperature sensor. Based on the thermal expansion properties of the material within the bulb, the material expands as the temperature increases. This expansion causes a change in the pressure of the actuator. The pressure change actuator then repositions power source. Again, the change in the power source changes the flowrate through the valve.

Temperature regulation systems using thermocouples or RTDs as temperature sensing devices are much more common than regulation systems using filled bulbs.

Temperature Detector Placement: Internal and Remote Detection

Temperature detection can be done with internal or remote elements. For internal temperature detectors, the thermal actuator and temperature detector are located entirely within the valve. For remote temperature detectors, the primary temperature detecting element is separate from the actuator and valve, and is connected to the actuator with either electrical wiring or capillary tubing, depending on the mechanism of the temperature sensor. Remote temperature detectors are more common, as internal temperature detectors are limited in use. Internal temperature detectors can only measure the temperature of the fluid flowing through the valve and not the temperature of the process.

Actuator Type: Thermal Systems

There are four main categories of thermal actuators used in temperature regulators. Thermal actuators produce power and work, proportional to the measured temperature of the process, on the power source. Actuator types include the vapor-filled system, the liquid-filled system, the hot chamber system, and the fusion-type or wax-filled system. Of all the thermal systems mentioned, liquid-filled systems are the most common, because they relate temperature and pressure change in a linear fashion.

Vapor-Filled Systems

In the vapor-filled system, the thermal actuator is partially filled with a volatile liquid. As the temperature of the sensor increases, the vapor pressure of the liquid also increases. This increases the pressure on the power source, and adjusts the flowrate through the valve.

Liquid-Filled Systems

In liquid-filled systems, the thermal actuator is filled with a chemically stable liquid, such as a hydrocarbon. As the temperature increases, the liquid expands, which produces a force on the power source.

Hot Chamber Systems

In hot chamber systems, the thermal actuator is partially filled with a volatile fluid. An increase in temperature of the system forces some of this fluid into the power unit, where the heat of the unit causes this liquid to turn into a superheated vapor. The pressure increase produces a force on the power source.

Fusion-Type (Wax-Filled) Systems

Of all the systems mentioned, the fusion-type system is the least common. In the fusion-type system, the thermal actuator is filled with special waxes such as hydrocarbons, silicones, and natural waxes. The wax contains large amounts of copper, which increases the heat-transfer quality of the wax. As temperature increases, the wax expands, producing a force that repositions the power source.

Valve Type: Direct and Pilot Actuated

The two main types of valves used in thermal regulators are the direct and pilot actuated valves. In all such thermal regulators, there is a power source (such as bellows and diaphragms) that provides the force required to reposition the valve to control the temperature. These power sources rely on a change in the pressure of the actuator in order to properly regulate temperature. In direct-actuated TCVs, this power unit is directly connected to the valve, which provides the force required to open and close the valve. In pilot-actuated TCVs, the thermal actuator moves a pilot valve, which then transfers energy in the form of pressure to a piston, which then provides the work necessary to reposition the main valve.

Direct-actuated TCVs are often much simpler in structure than pilot-actuated TCVs, and therefore they are also much cheaper. In addition, they respond better to smaller changes in temperature and more accurately reflect the temperature of the medium. Thus, if the exact temperature of the system is essential to ensure correct operation, a direct-actuated TCV should be used. Pilot-actuated TCVs usually have much smaller temperature sensing devices, a faster response time, and the ability withstand much higher pressures through the regulating valve. Therefore, at high pressures or rapid temperature changes, a pilot-actuated TCV should be used.

2.4 Resistance Temperature Detector Example

Note: This example problem has been made up for demonstration purposes.

A newly hired chemical engineer at Hypothetical Industries is responsible for monitoring and maintaining temperatures for one of the company's exothermic reactions. The

process uses a platinum resistor thermometer to measure the temperature of the process, the properties of which are given below. The reaction's ideal range is between 250°C – 350°C. Below 250°C, the catalyst ceases to function, and above 350 °C, it can be classified as a runaway reaction. The engineer can control steam and cooling water to regulate the process temperature.

Describe what changes the engineer should make to either the flowrate of the coolant or the steam to keep the system functioning optimally.

Case I: $R_T = 25 ohms$

Case II: $R_T = 13.9 ohms$

Case III: $R_T = 19.4 ohms$

GIVEN DATA:

$R_T = R_0(1 + a_1 T + a_2 T^2)$

$R_0 = 10$ ohms
$a_1 = 3.91 x 10^{-3} (°C)$
$a_2 = -6.72 x 10^{-8} (°C^{-2})$

SOLUTION:

Substitute the given value of R_T for each of the 3 cases into the mathematical calculations below to get the answers, we have worked out Case I with the corresponding numbers below.

Case I:

1. Factor R_0 throughout the right side of the equation.
2. Subtract R_T from both sides of the equation.
3. Solve for T using the quadratic equation.

$$25 = 10(1 + 3.19 \times 10^{-3}T + -6.72 \times 10^{-8}T^2)$$
$$25 = 10 + 3.19 \times 10^{-2}T + -6.72 \times 10^{-7}T^2$$
$$0 = -15 + 3.19 \times 10^{-2}T + -6.72 * 10^{-7}T^2$$

$$T = \frac{-3.19 \times 10^{-2} \pm \sqrt{(3.19 \times 10^{-2})^2 - 4 \times (-6.72 \times 10^{-7}) \times (-15)}}{(2 \times -6.72 \times 10^{-7})}$$

$$T = 474.97°C$$

ALTERNATE SOLUTION:
1. Since the constant a_2 is so small $(x10^{-7})$ we can just neglect that term.
2. Now the problem is just a simple linear equation, which can be solved by isolating T on one side of the equation.

$$25 = 10(1 + 3.19 \times 10^{-3}T)$$
$$25 = 10 + 3.19 \times 10^{-2}T$$

$$15 = 3.19 \times 10^{-2}T$$

$$T = \frac{15}{3.19 \times 10^{-2}}$$

$$T = 470.21°C$$

ANSWERS:

```
Case I.    The engineer should start flowing cooling water
into the reaction process because the temperature is ~ 500°C
which is
 above the appropriate range.
 Case II.    The engineer should increase steam feed to the
reaction process because the temperature is ~ 125°C which is
below
 the appropriate range.
 Case III.   The engineer does not have to do anything
because the temperature is in the appropriate range ~ 300°C.
```

2.5 Temperature Sensor Example

Note: This example has been made up for demonstration purposes.

You are a chemical engineer at Hypothetical Industries and you are responsible for monitoring and regulating the temperature for one of the company's reactions. Determine

which temperature sensor should be used to optimally measure and regulate the reaction's temperature in each of the following situations.

Case I. T=900°C

Case II. T=500°C but sensor location in a large continuous reactor makes it difficult to repair

Case III. T=50°C and you are estimating the current temperature of a lab scale reaction

SOLUTION:

```
Case I.  We know that the reaction occurs at 900°C. Taking
this optimal temperature as the
 only necessary parameters, the temperature sensor would be
a thermocouple. The temperature is outside of the range for
resistance
 thermometers. Regular thermometers do not send electric
signals; therefore they can not be used for this process.
 Case II.  Although this temperature is within the
operating ranges of both thermocouples and RTDs, since the
sensor may not
 be readily accesible for repair, we must use an RTD
because thermocouples lose accuracy after extended usage.
 Case III.  Since the temperature is within the range of
all three sensors, and we
 only need a rough estimate of temperature, we can use a
thermometer which will be much less costly than the
alternative sensors.
```


2.6 Sage's Corner

A brief narration for better understanding of temperature sensors.

http://video.google.com/googleplayer.swf?docId=-1336940015427451435
A copy of the slides can be found here: [[1]]

2.7 References

- Dougherty, E.P. ,*Temperature Control Principles for Process Engineers*, Hanser Publishers, 1993 ISBN 0344615980
- Liptak,B.G. ,*Instrument Engineer's Handbook*, 4th Edition,Volume 2. Taylor and Francis Group,2005 ISBN 0849310814
- Perry,R.H. ,*Perry's Chemical Engineer's Handbook*, 7th Edition, McGraw-Hill,1997 ISBN 0070498415

- Richardson,J.F. and Peacock,D.G. ,*Chemical Engineering*, 3rd Edition,Volume 3. Butterworth-Heinemann,1994 ISBN 0080410030
- Moffat,R.J. ,*Notes on Using Thermocouples*, 1997
- **Error! Hyperlink reference not valid.**
- Pyrosales
- Cole parmer Suppliers
- Temperature.com

Section 3. Pressure Sensors

Authors: (09/14/06 / Date Revised: 09/20/06) Wen Chung Lim, James Bennett, Jamila Grant, Ajay Bhasin

Stewards: (date) Names

- [First round reviews for this page](#)
- [Rebuttal for this page](#)

3.1 Introduction

Pressure must be considered when designing many chemical processes. Pressure is defined as force per unit area and is measured in English units of psi and SI units of Pa. There are three types of pressure measurements:

1. Absolute pressure - atomospheric pressure plus gauge pressure.
2. Gauge Pressure - absolute pressure minus atmospheric pressure.
3. Differential Pressure - pressure difference between two locations.

There are various types of pressure sensors that are available in the market today for use in industry. Each functions best in a certain type of situation.

3.2 Sensor Selection Criteria

In order for a pressure controlled system to function properly and cost-effectively, it is important that the pressure sensor used be able to give accurate and precise readings as needed for a long period of time without need for maintenance or replacement while enduring the conditions of the system. Several factors influence the suitability of a particular pressure sensor for a given process: the characteristics of the substances being used or formed during the process, the environmental conditions of the system, the pressure range of the process, and the level of precision and sensitivity required in measurements made.

3.2.1 Process

The pressure sensing element (elastic element) will be exposed to the materials used in the process, therefore materials which might react with the process substances or degrade in corrosive media are unsuitable for use in the sensor. Diaphragms are optimal for very harsh environments.

3.2.2 Environment

The environment (as in the system -- pipes, vibrations, temperature, etc.) in which the process is carried out also needs to be considered when choosing a suitable pressure sensor. Corrosive environments, heavy vibrations in the piping units, or extreme

202

temperatures would mean that the sensors would have to have an added level of protection. Sealed, strong casing materials with interior liquid containing glycerine or silicone are often used to encase the internal components of the sensor (not including the sensing element), protecting them from very harsh, corrosive environments or frequent vibrations.

3.2.3 Pressure Range

Most processes operate within a certain pressure range. Because different pressure sensors work optimally in different pressure ranges, there is a need to choose pressure gauges which are able to function well in the range dictated by the process.

3.2.4 Sensitivity

Different processes require different levels of precision and accuracy. In general, the more precise the sensor, the more expensive it is, thus it is economically viable to choose sensors that are able to satisfy the precision desired. There is also a compromise between precision and the ability to detect pressure changes quickly in sensors, hence in processes in which the pressure is highly variable over short periods of time, it is unadvisable to use sensors which take a lot of time to give accurate pressure readings, although they might be able to give precise pressure values.

3.3 Pressure Measuring Methods

Several pressure measuring methods have been developed and utilized; these methods include visual inspection of the height of liquid in a column, elastic distortion, and electrical methods.

3.3.1 Height of Liquid in Column

The height of a liquid with known density is used to measure pressure. Using the equation $P = \varrho g h$, the gauge pressure can be easily calculated. These types of pressure measuring devices are usually called manometers. Units of length may be used to measure the height of the liquid in the column as well as calibrated pressure units. Typically water or mercury is used as the liquid within these columns. Water is used when you desire greater sensitivity (its density is much less than liquid mercury, so its height will vary more with a pressure change). Mercury is used when you desire higher pressure measurements and not as great sensitivity.

3.3.2 Elastic Distortion

This pressure measuring method is based on the idea that deformation of an elastic material is directly proportional to the pressure being measured. There are mainly three sensor types that are used in this method of measuring pressure: Bourdon-tubes, diaphragms and bellows. (refer to "Types of Sensors" Section)

3.3.3 Electrical Methods

Electrical methods used for measuring pressure utilize the idea that dimensional changes in a wire affect the electrical resistance to the conductor. These devices that use the change in resistance of the wire are called strain gauges. Other electrical sensors include capacitive sensors, inductive pressure transducers, reluctive pressure transducers, potentiometric transducers, piezoresistive transducers and piezoelectric sensors. (refer to "Types of Sensors" Section)

3.4 Types of Sensors

There are many different pressure sensors to choose from when considering which is most suitable for a given process, but they can generally be placed into a few categories, namely elastic sensors, electrical transducers, differential pressure cells and vacuum pressure sensors. Listed below each general category are specific internal components, each functioning best in a certain situation.

3.4.1 Elastic Sensors

Most fluid pressure sensors are of the elastic type, where the fluid is enclosed in a small compartment with at least one elastic wall. The pressure reading is thereby determined by measuring the deflection of this elastic wall, resulting in either a direct readout through suitable linkages, or a transduced electrical signal. Elastic pressure sensors are sensitive; they are commonly fragile and susceptible to vibration, however. In addition, they tend to be much more expensive than manometers, and are therefore preferentially used for transmitting measured data and measuring pressure differences. A wide variety of flexible elements could conceivably be used for elastic pressure sensors; the majority of devices use some form of a Bourdon tube, bellows, or diaphragm.

Bourdon Tube Gauges

The principle behind all Bourdon tubes is that an increase in pressure on the inside of the tube in comparison to the outside pressure causes the oval or flat shaped cross-section of the tube to try to achieve a circular shape. This phenomenon causes the tube to either straighten itself out in the c-type or spiral cases or to unwind itself for the twisted and helical varieties. This change can then be measured with an analog or digital meter connected to the tube. Tube materials can be changed accordingly to suit the required process conditions. Bourdon tubes can operate under a pressure range from 0.1-700 MPa. They are also portable and require little maintenance; however,they can only be used for static measurements and have low accuracy.

Types of Bourdon tubes include C-type, spiral (a more coiled C-type tube), helical and straight tube Bourdon tubes. C-type gauges can be used in pressures approaching 700MPa; they do have a minimum recommended pressure range, though -- 30 kPa (i.e., it is not sensitive enough for pressure differences less than 30 kPa).

204

Bellows

Bellows elements are cylindrical in shape and contain many folds. They deform in the axial direction (compression or expansion) with changes in pressure. The pressure that needs to be measured is applied to one side of the bellows (either inside or outside) while atmospheric pressure is on the opposite side. Absolute pressure can be measured by evacuating either the exterior or interior space of the bellows and then measuring the pressure at the opposite side. Bellows can only be connected to an on/off switch or potientiomenter and are used at low pressures, <0.2 MPa with a sensitivity of 0.0012 MPa.

Diaphragms

Diaphragm elements are made of circular metal discs or flexible elements such as rubber, plastic or leather. The material from which the diaphragm is made depends on whether it takes advantage of the elastic nature of the material, or is opposed by another element (such as a spring). Diaphragms made of metal discs utilize elastic characteristics, while those made of flexible elements are opposed by another elastic element. These diaphragm sensors are very sensitive to rapid pressure changes. The metal type can measure a maximum pressure of approximately 7 MPa, while the elastic type is used for measuring extremely low pressures (.1 kPa - 2.2 MPa) when connected to capacitative transducers or differential pressure sensors. Examples of diaphragms include flat, corrugated and capsule diaphragms. As previously noted, diaphragms are very sensitive (0.01 MPa) . They can measure fractional pressure differences over a very minute range (say, inches of water) (elastic type) or large pressure differences (approaching a maximum range of 207 kPa) (metal type).

Diaphragm elements are very versatile -- they are commonly used in very corrosive environments or with extreme over-pressure situations.

Examples of these elastic element pressure sensors are shown here.

Pressure → Motion
Flat diaphragm

Pressure → Motion
(b) Corrugated diaphragm

Pressure → Motion
(c) Capsule

Motion
Pressure
(d) Bellows

Motion
Pressure
(e) Straight tube

Motion
Pressure
(f) C-shaped Bourdon tube

Motion
Pressure
(g) Twisted Bourdon tube

Motion
Pressure
(h) Helical Bourdon tube

Motion
Pressure
(i) Spiral Bourdon tube
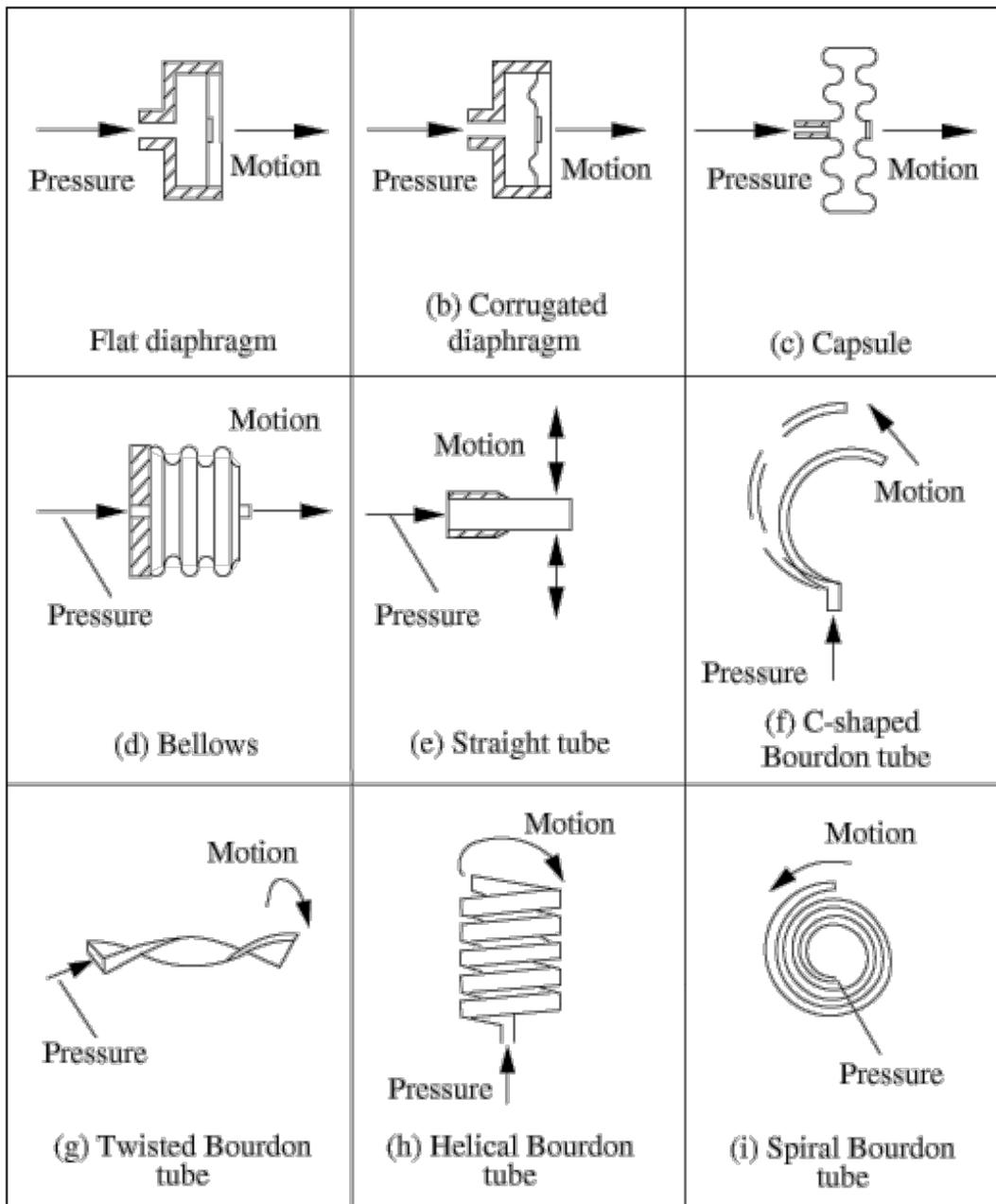
3.4.2 Electric Sensors

Sensors today are not neccessarily only connected to a gauge meter needle pointer to indicate pressure, but may also serve to convert the process pressure into an electrical or pneumatic signal, which can be transmitted to a control room from which the pressure reading is determined. Electric sensors take the given mechanics of an elastic sensor and incorporate an electrical component, thus heightening the sensitivity and increasing the amount of instances in which you could utilize the sensor. The types of pressure

transducers are capacitive, inductive, reluctive, piezoelectric, strain gauge, vibrating element, and potentiometric.

Capacitive

A capacitive sensor consists of a parallel plate capacitors coupled with a diaphragm that is usually metal and exposed to the process pressure on one side and the reference pressure on the other side. Electrodes are attached to the diaphragm and are charged by a high frequency oscillator. The electrodes sense any movement of the diaphragm and this changes the capacitance. The change of the capacitance is detected by an attached circuit which then outputs a voltage according to the pressure change. This type of sensor can be operated in the range of 2.5 Pa - 70MPa with a sensitivity of 0.07 MPa.

An example of a capacitive pressure sensor is shown on the right.
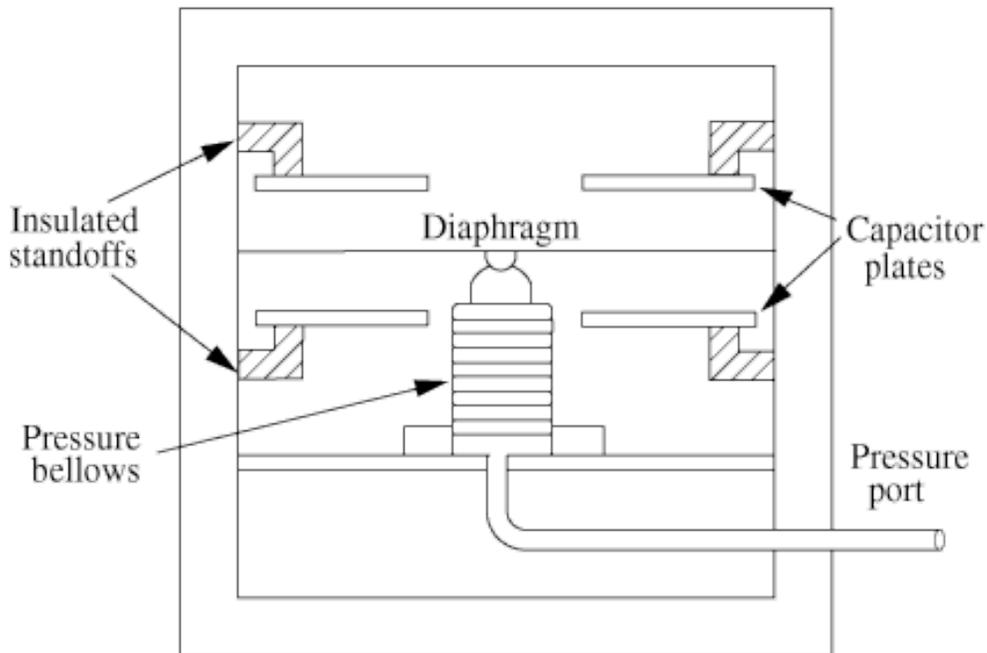
Inductive

Inductive pressure sensors are coupled with a diaphragm or a Bourdon tube. A ferromagnetic core is attached to the elastic element and has a primary and 2 secondary windings. A current is charged to the primary winding. When the core is centered then the same voltage will be induced to the two secondary windings. When the core moves with a pressure change, the voltage ratio between the two secondary windings changes. The difference between the voltages is proportional to the change in pressure.

An example of an inductive pressure sensor utilizing a diaphragm is shown below. For this kind of pressure sensor, taking Chamber 1 as the reference chamber with a reference

pressure P_1 coming into the chamber and the coil being charged with a reference current. When the pressure in the other chamber changes, the diagphragm moves and induces a current in the other coil, which is measured and gives a measure of the change in pressure.

These may be used with any elastic element (though, it is typically coupled with a diaphragm or a bourdon tube). The pressure reading generated will be determined by voltage calibration. Thus, the range of pressure in which this sensor may be used is determined by an associated elastic element but falls in the range of 250 Pa - 70 MPa .
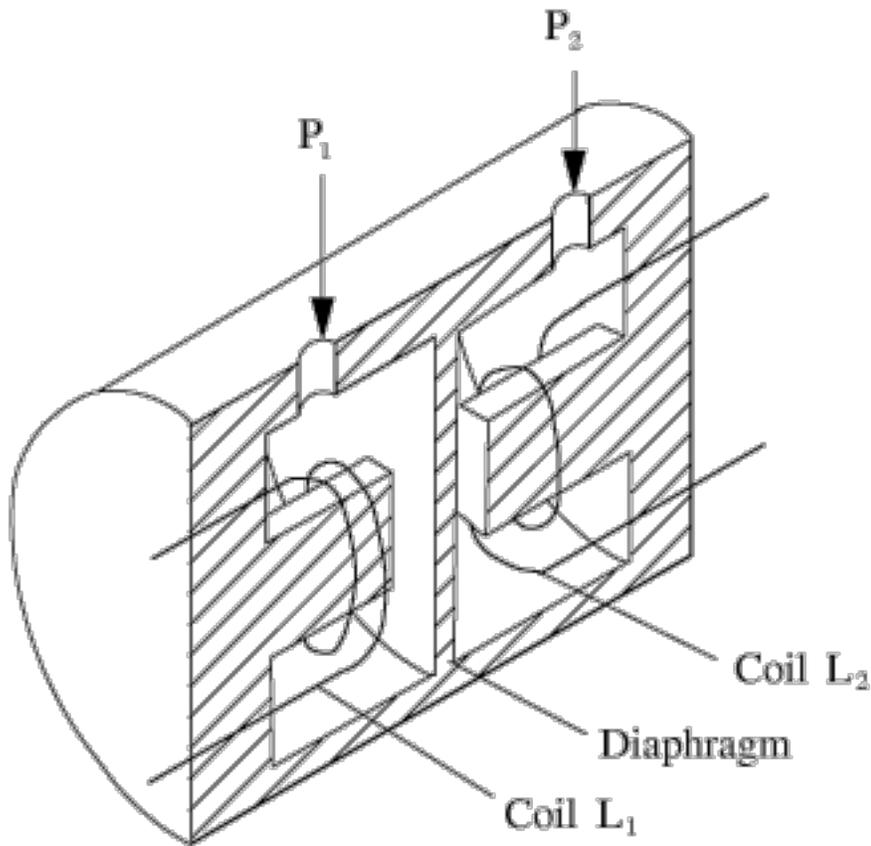
Reluctive

Reluctive pressure sensors also charge a ferromagnetic core. When the pressure changes, the flexible element moves a ferromagnetic plate, leading to a change in the magnetic flux of the circuit which can be measured. The situations in which one would use a reluctive electric element is one in which the inductive sensor does not generate a precise enough measurement. The pressure range is 250 Pa - 70 MPa with a sensitivity of 0.35 MPa.

An example of a reluctive pressure sensor can be seen on the right.

Additional coil-core set
(optional)

Armature

Coil L_2

Twisted Bourdon
Tube

Coil L_1

Magnetic "E" core
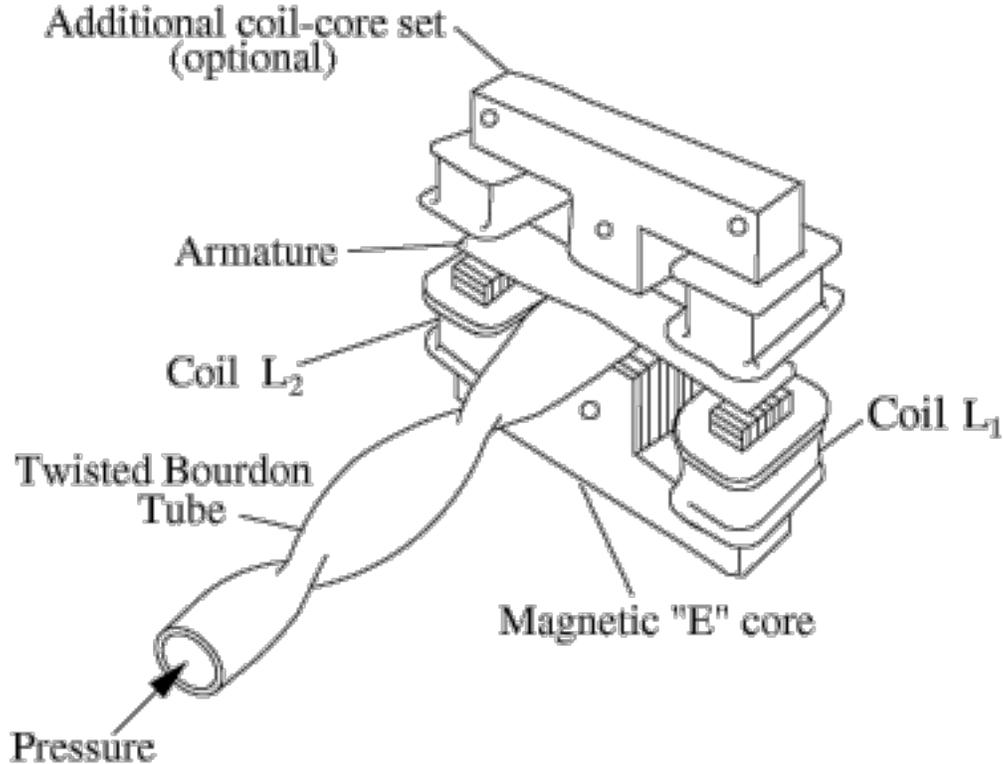
Pressure

Piezoelectric

Piezoelectric sensors use a crystal sensor. When pressure is applied to the crystal, it deforms and a small electric charge is created. The measurement of the electric charge corresponds to the change in pressure. This type of sensor has a very rapid response time to constant pressure changes. Similar to reluctive electric element, the piezoelectric element is very sensitive, but responds much, much faster. Thus, if time is of the essence, a piezoelectric sensor would be desired. The pressure range is .021 - 100 MPa with a sensitivity of 0.1 MPa.

On the right is an example of a piezoelectric pressure sensor.

Nut

Housing

Lead wire

Crystal

Disc

Diaphragm
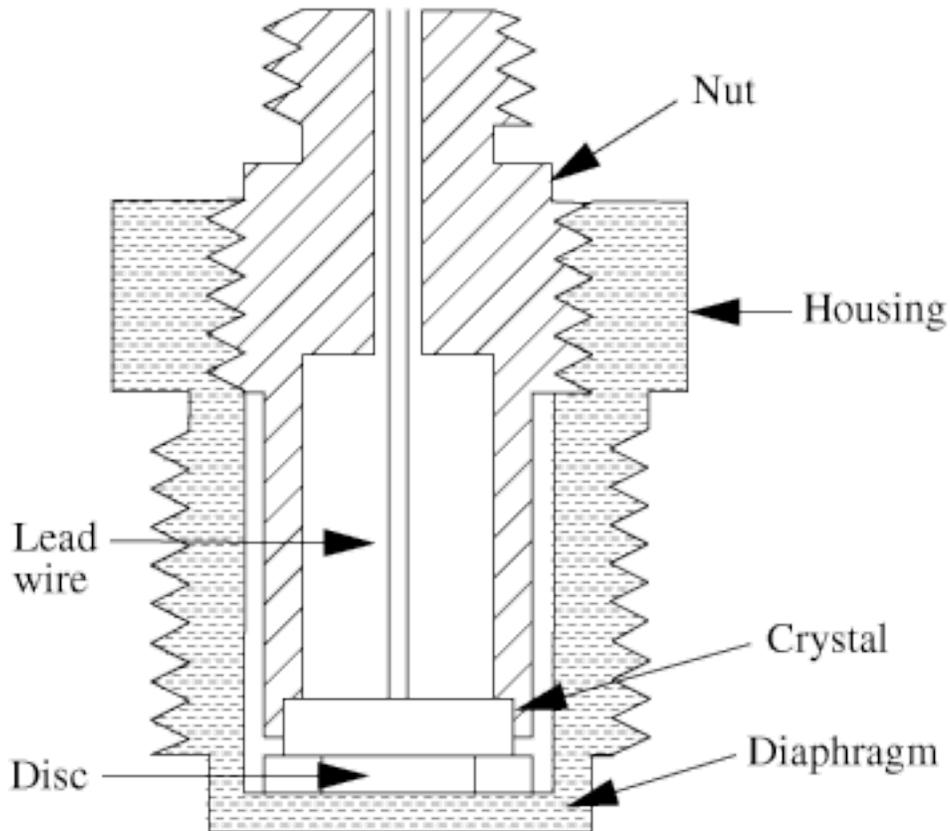
Potentiometric

Potentiometric sensors have an arm mechanically attached to the elastic pressure sensing element. When pressure changes, the elastic element deforms, causing the arm to move backwards or forwards across a potentiometer and a resistance measurement is taken. These sensing elements do posess an optimum working range, but are seemingly limited in their resolution by many factors. As such, they are low end sensors that aren't used for much. With a low sensitivity and working range, they may be best suited as a cheap detector evaluating a coarse process.The pressure range is 0.035 - 70 MPa with a sensitivity of 0.07 -0.35 MPa.

An example of a potentiometric pressure sensor is shown on the right.
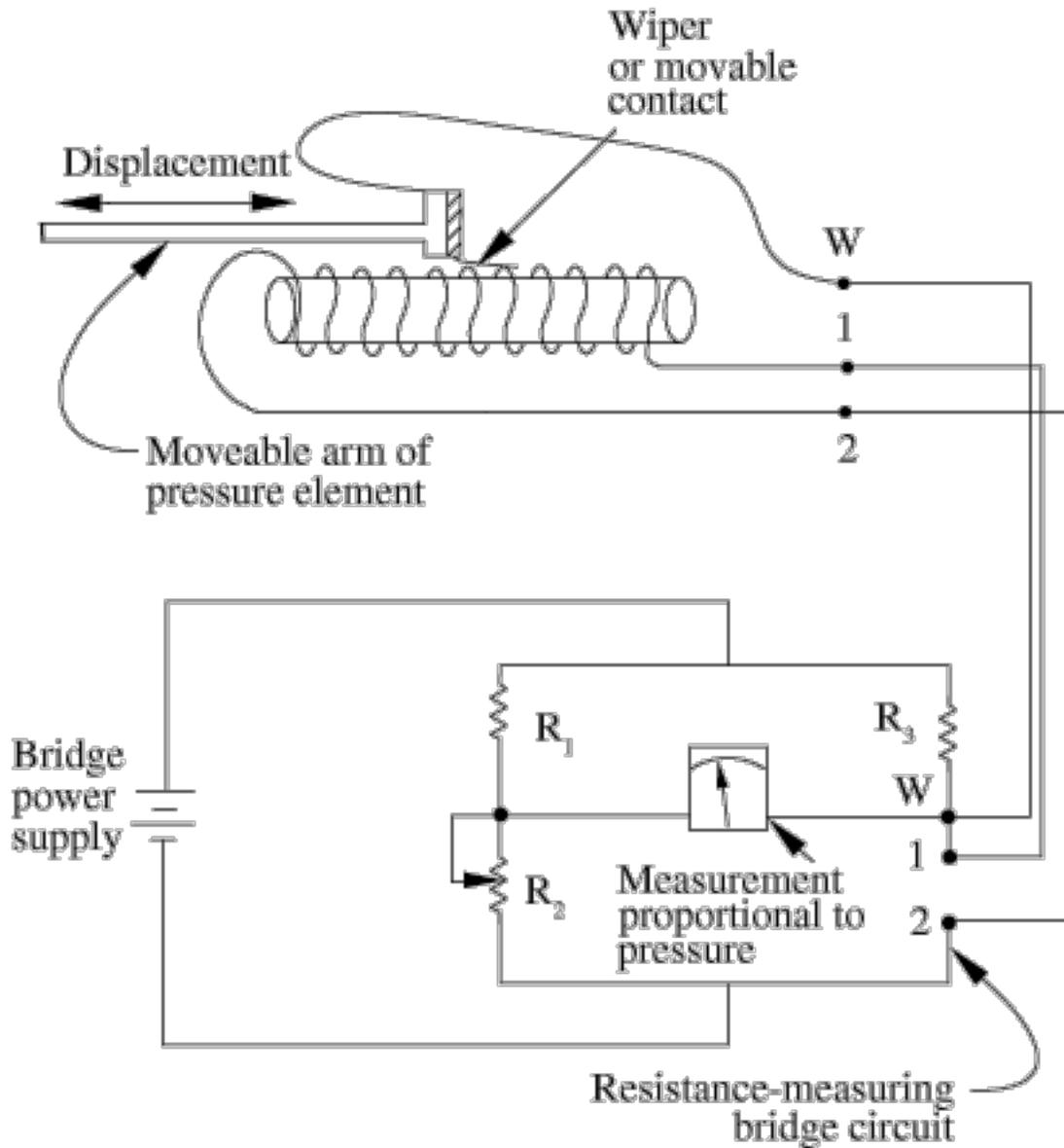
Strain Gauge

The strain gauge detects changes in pressure by measuring the change in resistance of a Wheatstone bridge circuit. In general, this circuit is used to determine an unknown electrical resistance by balancing two sections of a bridge circuit such that the ratio of resistances in one section ($\frac{R_3}{R_2}$) is the same as that in the other section($\frac{R_4}{R_1}$), resulting in a zero reading in the galvanometer in the center branch. One of the sections contains the unknown component of which the resistance is to be determined, while the other

section contains a resistor of known resistance that can be varied. The Wheatstone bridge circuit is shown below:
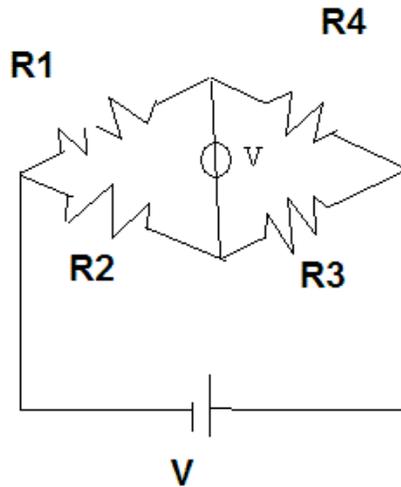
The strain gauge places sensors on each of the resistors and measures the change in resistance of each individual resistor due to a change in pressure. Resistance is governed by the equation

$$R = \rho \frac{L}{A}$$

where ρ = resistivity of the wire, L = length of the wire, and A = cross-sectional area of the wire. A pressure change would either elongate or compress the wire, hence a compression sensor is needed on one resistor and an elongation sensor on the other. To control the effects of temperature (a wire would also either elongate or compress with a change in temperature), a blank sensor would be placed on the remaining two resistors. These gauges are frequently a type of semiconductor (N-type or P-type). Thus, their sensitivity is much greater than their metal counterparts; however, with greater sensitivity comes a more narrow functional range: the temperature must remain constant to obtain a valid reading. These gauges are affected greatly by variations in temperature (unlike the other types of electrical components). The pressure range is 0 - 1400 MPa with a sensitivity of 1.4 - 3.5MPa.

An example of an unbonded strain gauge is shown below. This makes use of strain-sensitive wires one end fixed to an immobile frame and the other end attached to a

212

movable element, which moves with a change in pressure.
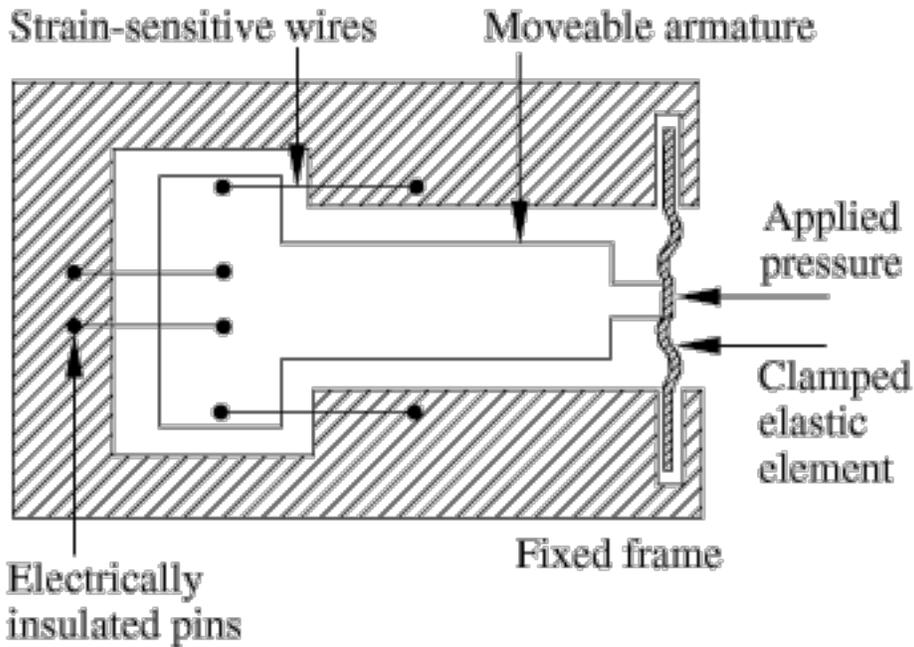
An example of a bonded strain gauge can be seen below. This is placed on top of a diaphragm, which deforms with change in pressure, straining the wires attached to the diaphragm.
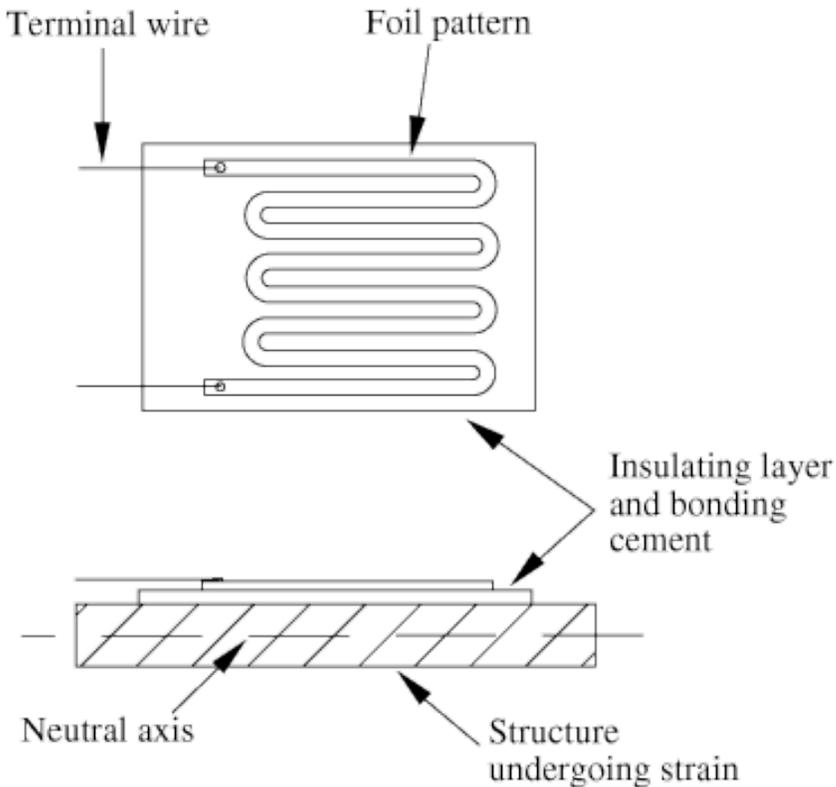
Vibrating Element

Vibrating element pressure sensors function by measuring a change in resonant frequency of a vibrating element. A current is passed through a wire which induces an electromotive force within the wire. The force is then amplified and causes oscillation of the wire. Pressure affects this mechanism by affecting the wire itself: an increase in pressure decreases the tension within the wire and thus lowers the angular frequency of oscillation of the wire. The sensor is housed in a cylinder under vacuum when measuring absolute pressures. These absolute pressure measuring sensors are very efficient: they produce repeatable results and are not affected by temperature greatly. They lack sensitivity in meausurement, though, so they would not be ideal for a process in which minute pressures need monitoring. The pressure range is 0.0035 - 0.3 MPa with a sensitivity of 1E-5 MPa.

A vibrating wire pressure sensor is shown below.
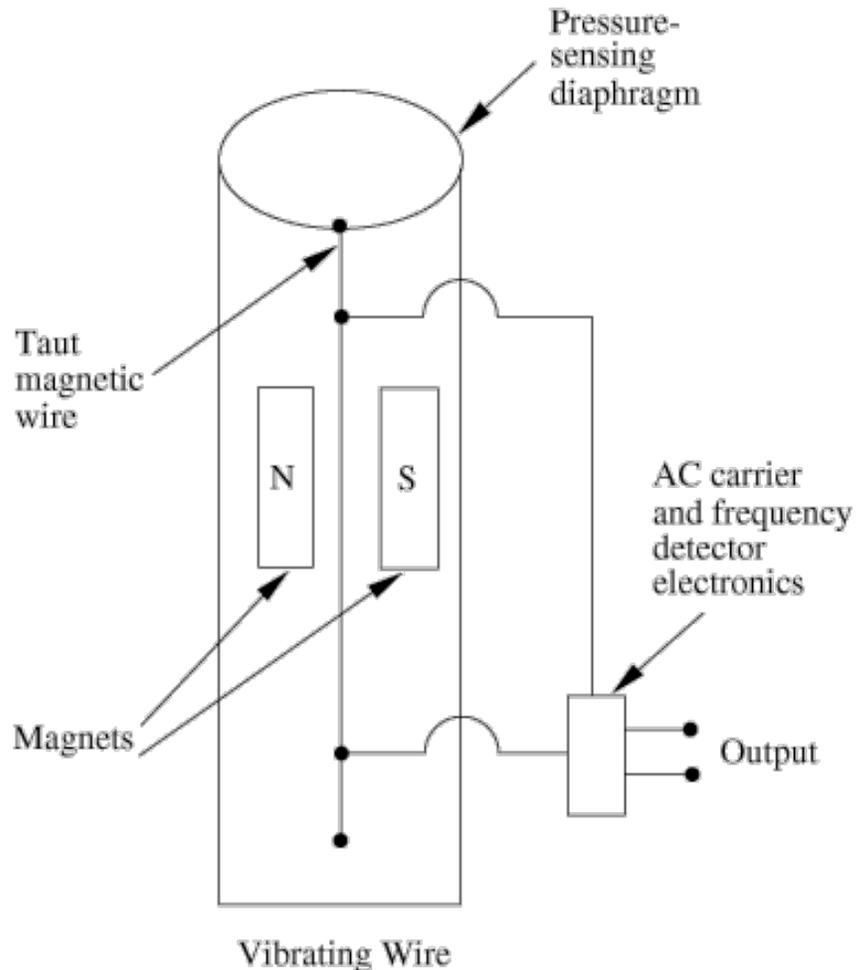
A vibrating cylinder pressure sensor (for absolute pressures) is shown below.
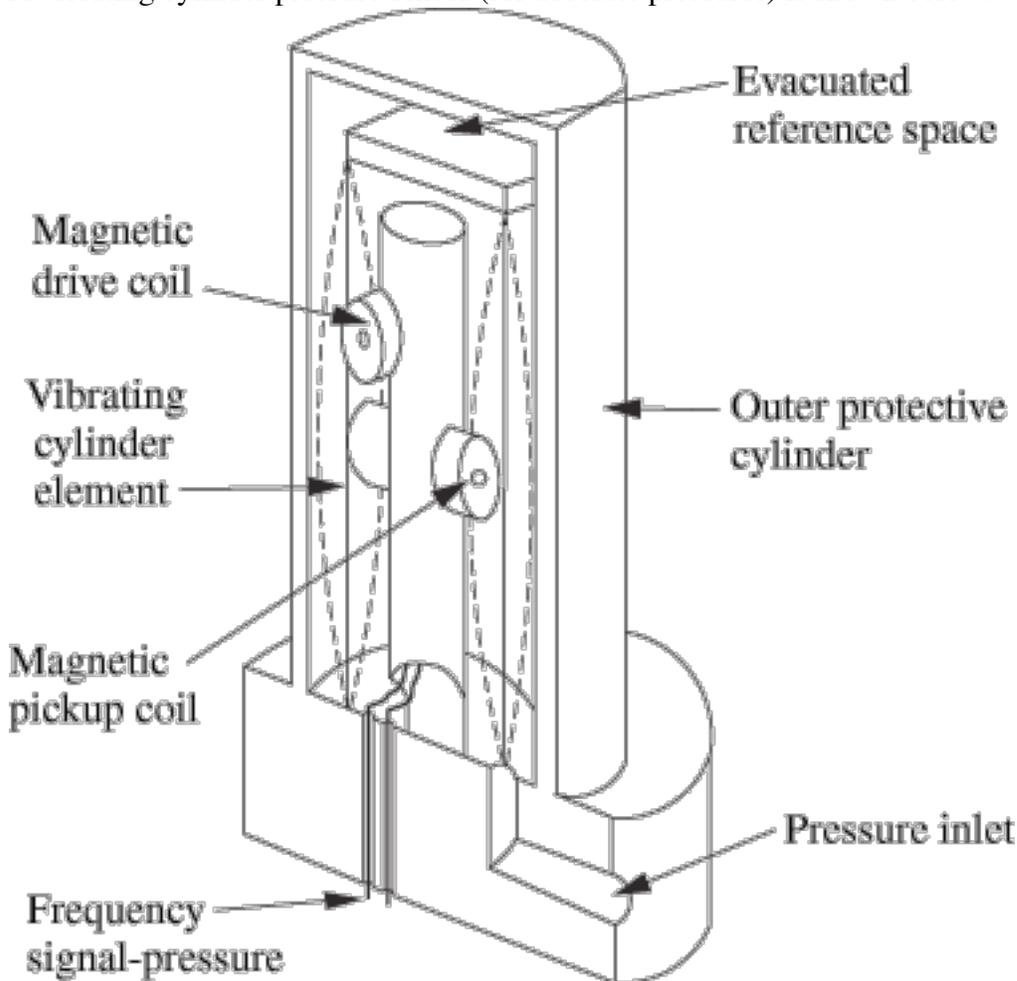
3.4.3 Differential Pressure Cells

Differential pressure cells are used with various kinds of sensors in which measurement of pressure is a result of a pressure differential such as orifice plates, flow nozzles, or venturi meters. The differential pressure cell converts the pressure differential into a transmittable signal. Where the differential pressure (DP) cell is placed depends on the nature of the fluid stream that is being measured. A typical DP cell is minimally invasive (an external component attached across the point of measurement); it is commonly employed with a capacitive element paired with a diaphragm that allows the capacitive body to separate or move together, generating a signal (via change in capacitance) that can be interpreted to a pressure drop. They are often used to detect small differences in large pressure drops. Its placement is similar to connecting a voltmeter in parallel to a resistor to measure its voltage "drop" (analgous to the pressure drop).

An example of a differential pressure cell using bellows can be found here: http://www.tpub.com/fluid/ch2s.h49.gif

The range of pressure measured and sensitivity of a differential cell depends upon the electric and elastic components used in the cell itself. It is a great sensor to use when measuring a pressure drop; however, for all other applications, it is fairly useless.

3.4.4 Vacuum Sensors

Such sensors are able to measure extremely low pressures or vacuum, referring to pressures below atmospheric pressure. Besides diaphragm and electric sensors designed to measure low pressures, there are also thermal conductivity gauges and ionization sensors.

Thermal Conductivity Gauges

The principle involved here is the change in gas thermal conductivity with pressure. However, due to deviation from ideal gas behavior in which the relationship between these two properties is linear, these kind of gauges, which are also called Pirani gauges, can only be used at low pressures, in the range of (0.4E-3 to 1.3E-3) MPa. They are amazingly sensitive elements as well (can detect changes of 6E-13 MPa).

In these gauges, a coiled wire filament has a current flowing through it, which heats up the coil. A change in pressure changes the rate of heat conduction away from the filament, thereby causing its temperature to vary. These changes in temperature can be detected by thermocouples in the gauge, which are also connected to reference filaments in the gauge as part of a Wheatstone bridge circuit.

An example of a Pirani gauge can be found here:
http://www.bama.ua.edu/~surfspec/vacbasics_files/image046.jpg

Ionization Gauges

There are two categories for these types of gauges: hot cathode and cold cathode. For hot cathode gauges, electrons are emitted by heated filaments, while for the cold cathode gauges electrons are released from the cathode due to collision of ions. Electrons hit the gas molecules entering the gauge forming positive ions, which are collected and cause an ion current to flow. The amount of cation formation is related to the gas density and consequently the pressure to be measured, as well as the constant electron current used, hence the ion current flow is a direct measure of the gas pressure. These both are highly sensitive instruments and thus most suited for fractional pressures. The hot cathode gauges are even more sensitive than cold cathode gauges and are able to measure pressures near 10^{-8}Pa. Their sensitivity ranges from (1E-16 to 1E-13) MPa

An example of an ionization gauge can be found here:
http://www.bama.ua.edu/~surfspec/vacbasics_files/image049.jpg

216

3.4.5 Summary

Types of elements	Press. Range	Sensitivity	Advantages	Disadvantages
Bourdon tube	0.1-700 MPa	0.03 MPa	Portable Low-Maintenance	Static measurements Low accuracy
Bellows	< 0.2MPa	0.0012 MPa	Can be used at low press.	Can only be connected to on/off switch or potentiometer
Diaphragms	0.1-2.2 MPa	0.01 MPa	Quick response time Very accurate Generally good linearity Can be exposed to corrosives	Very expensive
Capacitive	2.5 Pa-70 MPa	0.07 MPa	Used for low pressure or vacuum ranges Robust	Electronically complex Capacitor plates can stick together
Inductive	250Pa – 70 MPa	0.35 MPa	High sensitivity	Limited by elastic elements More coarse than reluctive
Reluctive	250Pa – 70 MPa	0.35MPa	High sensitivity	Requires external AC source
Piezoelectric	0.021- 100 MPa	0.1 MPa	Very rapid response time	Affected by high temp. and static forces
Potentiometric	0.035- 70 MPa	0.07 – 0.35 MPa	Can be made very small	Low sensitivity and working range
Strain Gauge	0-1400 MPa	1.4-3.5 MPa	Very high sensitivity Can be used on mobile parts	Extremely slow response time Very small signal output
Differential	Depends on other elements	Depends on other elements	Used for pressure drop	Can only be used for press. drop
Thermal Conductivity	0.4E-3 – 1.3E-3 MPa	6E-13 MPa	Measures vacuum ranges	Has linearity only at low pressures
Ionization	1.3E-13 – 1.3E-8MPa	1E-13 – 1E-16 MPa	Highly sensitive Can measure high and ultra-high vacuum ranges	Limited by photoelectric effect
Vibrating	0.0035MPa - 0.3MPa	1E-5 MPa	Very accurate Not affected by temp. changes	Can't be used at high pressures

3.5 Semi-batch Reactor Problem

Suppose you have a semi-batch reactor (1,000 L) with 50 kg of zinc within the reactor at 1 atm pressure and 25 C. 6M hydrochloric acid is flowing in at 1 L / min to react with the zinc to produce zinc chloride for use in another process.

A) What factors should be considered?

B) Say the valve fails at an operating pressure of 4 atm (i.e., it will not close and the reactor will be flooded with HCl) At what pressure would you (safely) set the shutdown point?

C) What type of sensor should be used?

Ans:

3.5.1 The factors that should be considered are:

1. **Process**
   1. Hydrochloric acid is very, very caustic (especially at such a high molarity), and thus whatever sensor you choose, it should be able to withstand the corrosive nature of hte process.
2. **Range of pressures**
   1. Initially the reactor is at 1 atm of pressure. Considering the reaction 2 $HCl_{(aq)}$ + $Zn_{(s)}$ --> $H_{2(g)}$ + $ZnCl_{2(aq)}$, you're producing one mole of hydrogen gas in addition to the existing air pressure in the vessel. As

the reaction proceeds, the pressure would increase substantially within the vessel. Modeling the pressure of $H_{2(g)}$ ideally, $P = nRT / V$

$$P = \frac{(6MHCl)(1L/min)(1mol\,H_{2(g)}/2mol\,HCl_{(aq)})(0.08206\,Latm/molK)((25+273)K)}{1000L}$$

  2.
   = 0.073 atm / min.
  3. After about 1 hour, the pressure of $H_{2(g)}$ would increase to 4.38 atm, putting the total pressure in the vessel at 5.38 atm (past failing pressure for the valve).
3. **Environment**
   1. There is no danger of high temperatures or excessive vibration due to high flow / reaction.
4. **Sensitivity**
   1. Since this is a moderately dangerous process, we need to have the sensor output to a computer so an engineer can monitor the process as a fail-safe. We assume the sensor will signal the HCl valve to close once the operating pressure becomes 3 atm, however devices do fail from time to time. We also need high sensitivity, so electric components will be preferrable (i.e., we don't want the process to runaway as it potentially could if the sensor isn't highly sensitive to incremental changes).

3.5.2 The shutdown point

Considering the rapid increase in pressure, as evaluated in (2), and the failure of the valve at 4 atm, the shutdown point should be approximately 3 atm.

3.5.3 The type of sensor:

1. Considering the types of sensors discussed, we can automatically eliminate vacuum sensors as they operate at very low pressures (almost vacuum, hence the name). We can also eliminate differential pressure sensors as we are not looking for a pressure drop across the vessel.
2. Since we desire high sensitivity, we would like to use electric components. Considering the range of pressures (3 atm max ~ .3 MPa) a capacitive element would be optimal because it is robust and functions well in low pressure system.
3. Noting the corrosiveness of the HCl in the system, a diaphragm is likely to be used as the elastic element. Diaphragms are also fairly robust and offer quick response times.
4. This combination is likely to be filled in a strong, sealed, glycerine/silicone filled case so as to protect the sensor from degradation.

So, overall, the sensor we choose will utilize a diaphragm as the elastic component, a capacitive element as the electrical component, and an anti-corrosive body.

3.6 Worked out Example 2

Your supervisor told you to add a pressure sensor to a very expensive and important piece of equipment. You know that the piece of equipment is operating at 1MPa and at a very high temperature. What sensor would you choose?

Solution

Because the piece of equipment that you are dealing with is very expensive, you need a sensor that has the greatest sensitivity. An electric sensor would be desired because you could have it connected to a computer for quick and easy readings. Also, you need to pick a sensor that will operate at 1 MPa and a sensor that will be able to withstand high temperatures. From the information given in this article you know that there are many sensors that will operate at 1 MPa, so you need to decide on the other factors. One of the most sensitive electric sensors is the capacitive sensor. It has a sensitivity of 0.07MPa. The capacitive sensor usually has a diaphragm as the elastic element. Diaphragms have a rapid response time, are very accurate and work at 1 MPa.

3.7 Sage's Corner

If you are a sage for this page, please link your narrated PowerPoint presentation here.

3.8 References

- 2005/2006 Cole-Parmer General Catalog, 2005
- Liptak, Bela G.(2005). *Instrument Engineers Handbook*, 4th Edition, Volume 2. Taylor & Francis CRC Press. ISBN 0849310814
- Perry, Robert H.(1997), *Perry's Chemical Engineers' Handbook*, 7th Edition. McGraw-Hill. ISBN 0070498415
- Richardson, J.F. and Peacock, D.G. (1994), *Chemical Engineering*, 3rd Edition, Volume 3. Butterworth-Heinemann. ISBN 0080410030
- Freeman, Lawrence A., Carpenter, Michael C., Rosenberry, Donald O., Rousseau, Joseph P., Unger, Randy and McLean, John S. *Use of submersible pressure transducers in water-resources investigations*. http://pubs.usgs.gov/twri/twri8a3/#N10423
- "Pressure Sensors," http://www.data-acquisition.us/industrial_electronics/input_devices_sensors_transducers_transmitters_measurement/sensors/pressure-sensors.html
- Wikipedia, http://en.wikipedia.org/wiki/Ion_gauge

Section 4. Level Sensors

Note: Video lecture available for this section!

Authors: (September 14, 2006) Andrew MacMillan, David Preston, Jessica Wolfe, Sandy Yu

Stewards: (September 10, 2007) Yoo Na Choi, Yuan Ma, Larry Mo, Julie Wesely

- First round reviews for this page
- Rebuttal for this page

4.1 Introduction

Level sensors allow for the level control of fluid in a vessel. Examples of where these sensors are installed include reactors, distillation columns, evaporators, mixing tanks, etc. Level sensors provide operators with three important data for control: (1) the amount of materials available for processing, (2) the amount of products in storage, (3) the operating condition. Installing the correct level sensor ensures the safety of the operator and the surrounding environment by preventing materials in vessels from overflowing or running dry.

There are several different types of level monitors, including:

- Visual
- Float
- Valve Controlled
- Electronic
- Radiation

These different types of sensors can also be grouped into categories of process contact and non-process contact. As the name suggests, process contact sensors are within the tank, in physical contact with the material. Non-process contact sensors transmit various types of signals to reflect off of the material and thus measure the level. This sensor design can maintain its integrity within a potentially corrosive material and/or be positioned such that it can monitor the level changes from above the tank.

4.2 Visual Level Sensors

Visual level controls were the earliest developed level sensor. These types of monitoring devices can be something as simple as looking into an open container or inserting a marked object such as a dipstick. This type of sensor is the simplest and possibly the most reliable. These devices do not provide a way of connecting to a control device. They

require human input with no way of automation; however, electronic issues will not be a possible problem.

4.2.1 Common Uses

Visual indicators can be found in many places and are not limited to chemical engineering applications. Examples of these are: dipsticks found monitoring the oil levels in a car and measuring cups with markings indicating different volumes. In chemical plants, visual indicators are used to measure changes in level as well as for high and low level alarms.

4.2.2 Benefits of Visual Level Sensors

Visual sensors are generally less expensive than other types of sensors. They are more reliable because of the simplicity of design.

4.2.3 Restrictions

Visual indicators do not always accurately measure how much volume is in the tank. They do not allow for any digital connections to process control systems. In the case of the sight tubes, they must be affixed directly to the tank and cannot be read remotely without a camera or other transmission tool. Tanks placed in hard-to-reach places would also cause problems contributing to inaccuracies in measurement, such as not reading markers on eye level or interpolating between marker intervals. Also, the connections between the tube and the tank are subject to residue buildup which could prevent the readings from being accurate. Weather is also a concern with sight tubes. External temperature changes could affect the fluid in the tube and thus measurement inaccuracy. For example, the fluid in the sensor could freeze, or the sensor could become clogged. This type of indicator requires modification to the vessel wall and would therefore be optimally installed at the time of initial construction. If considered at the time of the initial capital installation of the tank a sight tube will not add a large cost to the project. Yet, later modification to include this in the design and associated tank changes are potentially very costly. Despite these shortfalls, a sight tube is a reliable and appropriate choice in many common level reading applications.

4.2.4 Sight Tube Indicators

Sight tube indicators allow operators to monitor levels with precision while keeping the tank sealed. This type of monitoring device is comprised of a vertical tube equal in height to the actual vessel. This tube is connected in at least two places directly to the vessel so that its contents can flow into the monitoring tube. This ensures that the height of liquid in the tank will be equal to the height of liquid in the sight tube. Markings can be affixed to the indicator for both calibration and volume readings. There are many styles of this indicator. Some models have a float ball contained in the tube so that it will float on top of the liquid. Other models of the indicator have many paddles on fixed posts through the length of the tube. These paddles float

horizontally when submerged and are vertical when not submerged. This results in a color change allowing for an easy identification of the tank volume.

A sight tube.

More pictures of this type of gauge can be found on [Seetru Limited's website](#).

4.3 Float Type Level Sensors

Float Types of level sensors are based on the principle of buoyancy which is the upward force produced on a submerged object by the displaced fluid. This force is equal to the weight of the displaced fluid. Float sensors take their measurements at the interfaces of materials, where the movement of the float and/or the force on the float are caused by the differing densities of the float and the fluid. There are two broad categories of Float Type level sensors: Buoyancy and Static.

4.3.1 Buoyancy Types

Buoyancy level sensors are less dense than the fluid and thus change position along with the fluid level. The movement of the float transmits the level information through some mechanical linkage to an output such as a valve or operator observation. There are basic three types of mechanical linkage - chain / tape sensors, lever / shaft mechanisms and magnetically coupled devices.

222

- Chain / Tape sensors – The linkage is by a flexible chain or tape.

- Lever / Shaft mechanisms – The linkage is a rigid shaft.

- Magnetically coupled devices – These devices are similar to the Chain/Tape sensors, except a magnet is attached to the float and another is acted upon by the floating magnet moving a tape like the chain/tape type devices. The moving magnet can be sequestered from the float attached magnet for use in corrosive media.

Schematic of a float type level sensor is shown below. As the fluid level rises or falls, buoyant force is transferred through mechanical linkages to your output device.

Mechanical Linkage

Float

Fluid

Valve or Meter
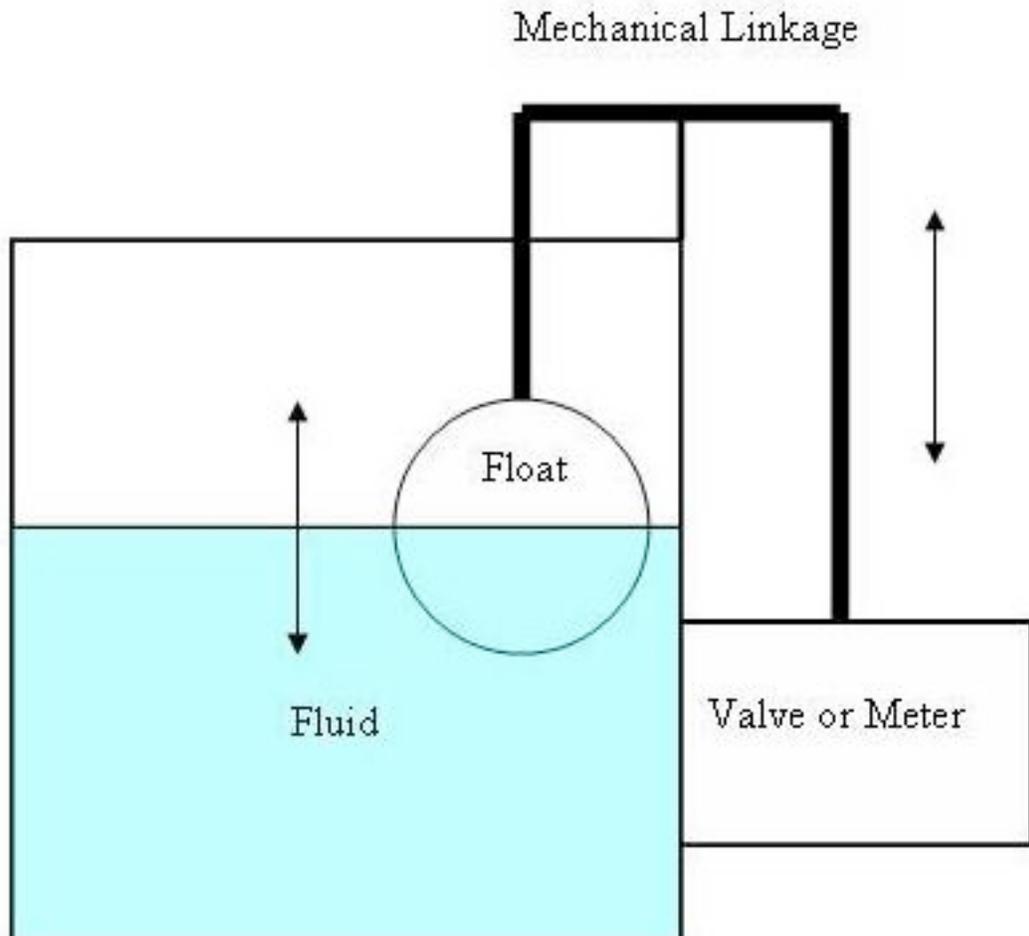
4.3.2 Static Types

Static level sensors are denser than the media being measured and thus do not move. As the level changes, the buoyant force acting on the "float", which is actually a weight, changes. The change in weight is measured by a scale. The level of the tank is calculated by measurements of the weight change in the float, not its actual change in position.

4.3.3 Formulas

Changes in the volumetric rate of material flow, resulting from changes in such things as the pressure of your pump will affect the operation of a float type level sensor. The same movement of the float will have the same effect on the valve, but as valve is moderating a different flow it has a different proportional sensitivity Kc.

$$K_c = \frac{(q)}{k}$$

where q is the flow rate and k is the height the float moves to completely traverse the valves operating range.

This is important because the given change in fluid height used to previously design the system will still change the valve's position in the same manner, but won't affect the same change in flow rate. If your pressure has dropped, your valve will have to be opened wider to match the same flow rate. The amount wider that it has to be opened is called an offset. Manipulating the valve response to the float movement, also known as the gain:

$$Gain = \frac{(valve_{response})}{float_{movement}}$$

can be used to alleviate this problem, but it won't work exactly the same way it did before the change in load.

Operation power is the amount of power the float has available to perform actions to control the system. Changes in the specific gravity of the fluid (or the float) will affect buoyancy of the float and thus will affect the amount of force exert on your system. A temperature is a very common parameter that can change your fluid's specific gravity. Warmer fluids tend to be less dense, and so your float will generate less buoyant force for a given change in fluid level.

224

Zero buoyancy force level is the level at which the weight of the float is exactly as buoyant as the fluid it replaces. It can be approximated by the following formula:

$$ZeroBuoyantForceLevel = D\frac{(SG_{float})}{SG_{fluid}}$$

where D is the diameter of the float and SG is the specific gravity.

4.3.4 Common Uses

Float Type Level Sensors regulate how much water is in the reservoir of a flush toilet; the float is attached to a lever which has a rotating axis that stops when the flow of water reaches a certain level.

- Chain / Tape sensors - regulating the level on storage tanks at atmospheric pressures.

- Lever / Shaft Mechanisms - regulating the level on vessels under pressure.

4.3.5 Benefits

Float Type level sensors do not require external energy sources to operate. Since they are simple robust machines it is easy to repair. The cost of these units makes them on the economical side, ranging from approximately $20 (homedepot.com) for the apparatus in a standard toilet to a few hundred dollars for a 3/4" cast iron valve with float (Liptak 2005). Prices would increase with higher quality materials and valve size. Floats could be made out of plastics or metals (steel, stainless steel, etc.). Material selection will depend on the application; more chemically resistant materials would be used for corrosive mediums.

4.3.6 Restrictions

Float type level sensors should only be used in clean fluids. Fluids that are a suspension of solids or slurries could foul the operation of the machine. Anything that could increase friction on the mechanical linkages could increase the dead band, which is the delay the sensor experiences due to the excess of force required to overcome the static friction of a non-moving system. An excess of force is required to be built up, so a float type level sensor will not respond immediately to changes in level.

Valve-based level sensors not only measure the fluid level, but also cause the fluid level to change accordingly. Two basic types of valves that will be discussed are altitude valves and diverter valves.

4.4.1 Altitude Valves

A simple altitude valve uses a spring that opens and closes different ports and lines when pressure changes due to the changing fluid levels. When fluid levels exceed the setting of the spring, a diaphragm connected to the spring lowers, closes the drain port, and opens the main line pressure. This turns off the main valve and stops fluid from flowing into the tank. When fluid levels decrease, the diaphragm raises, opens the drain port, and closes the main line pressure. This turns on the main valve and more fluid is supplied to the tank.

An example of an altitude valve is shown in the bottom left corner of the following website: http://www.gaindustries.com/html/01a_SAVV.htm

Common Uses

Altitude valves are on/off controls and can be found in supply lines connected to basins, tanks, and reservoirs. The main job of these valves is to prevent overflow of the fluid and to hold fluid level constant.

Benefits

There is no external power source needed when operating altitude valves because they are controlled by the pressure of the process fluid. More complicated altitude valves can also be used for other purposes other than an on/off control. These features include the following:

- Open the valve when pressure drops to a predetermined point
- Delayed opening so the valve only opens when fluid level drops a certain amount below the set point
- Close the valve slowly to eliminate pressure build-up
- Two-way flow to allow fluid to return when the level has dropped below the set point
- Open the valve while maintaining constant inlet pressure and plant distribution pressure
- Reduce pressure for outlet streams. This may be needed when working with equipment that has high supply pressures, such as aerator basins. Aerator basins are typically used in wastewater treatment because they have the capability of holding large amounts of water.

- Check valves can be installed to allow the valve to close if the pressure reaches a predetermined low point (which could be the cause of equipment failure elsewhere in the plant)

Restrictions

Altitude valves should not be used in pressurized vessels. These valves are constructed from a limited number of materials, such as cast iron. This restricts the number of fluids that can be used with the valves. Therefore, altitude valves are usually used for water service. Since these valves are usually used to operate a large amount of fluid, they are subsequently large with multiple functions that make them very expensive. A 30 inch altitude valve can cost over $50,000. In addition, frequent maintenance is required because of the many components and moving parts. These valves are operated at ambient temperatures and should be checked for freezing of stationary sensing lines.

4.4.2 Diverter Valves

Diverter valves are connected to a dip tube that is submerged in a controlled tank. When the fluid level of the tank drops, the dip tube is exposed to atmospheric pressure. Once air enters the dip tube, it utilizes the Coanda effect. Coanda effects occur when a curved surface exists. Gases tend to follow the nearest curved surface while pushing other fluids in a different direction. In the case of the diverter valve, air from the dip tube follows the curved surface of the control port back to the storage tank and pushes fluid flowing out of the storage tank to the other wall of the valve. This fluid then flows into the controlled tank to adjust the level.

227
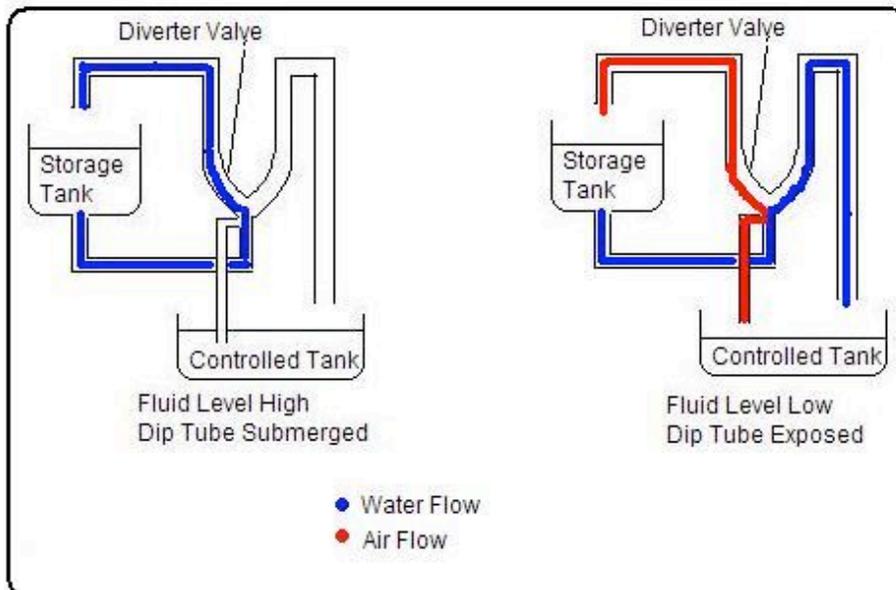
Common Uses

Diverter valves control fluid level by acting as an on/off switch. They can also be modified to be used for throttling. Diverter valves are used as low level indicators so they should only been used when low levels of a tank are a concern.

Benefits

Diverter valves need no external power source for operation. They have no moving parts, plugs, or packing, so they require minimal maintenance. Diverter valves can be operated at any temperature and are not affected by vibration. They can be constructed from many materials such as metals, plastics, and ceramics. Therefore, there is limited restriction to the type of liquid used with the valves. However, when using fluidized solids and slurries, the dip tube must remain open, so build-up does not occur. Any build-up near or in the tube should be removed.

Restrictions

Diverter valves should not be used when operating under conditions other than atmospheric pressure. They are also not for use when dealing with hard to handle processes. Outlet pressure cannot be greater than atmospheric pressure or back pressure will occur and the valve will not operate correctly.

4.5 Electrical Level Sensors

4.5.1 Conductive Level Sensing

Example of a conductive level sensor.

Conductive level sensors work by applying a low-voltage across two electrodes at different levels in a vessel. When both electrodes are immersed in a conductive liquid, a current flows. This type of electrical conductivity setup is best applicable for point level detection (level detection at a specific point in the material.) They are typically made of titanium, Hastelloy B, or stainless steel.

228

Common Uses

Conductive level sensors are commonly used to measure both conductive and corrosive liquids. A common conductive liquid is water, whereas some common corrosive liquids are nitric acid, ferric chloride, and hydrochloric acid.

Benefits

This method is considered extremely safe due to the low voltages and currents used. Conductive level sensors are also known for their easy installation and use.

Restrictions

The greatest concern with conductive sensors is maintenance. The probe needs to be monitored for buildup on the sensor. Residues from wet or sticky fluids that cause build up can be prevented by coating the sensors with Teflon or polyethylene-based materials.

4.5.2 Capacitance Level Sensing

A capacitor is made up of two conductors (electrodes/plates) that are electrically separated by a nonconductor (dielectric). In the case of level sensing, one of the electrodes is typically a vertically positioned rod while the other is the metallic vessel wall. The dielectric between them is the material being measured in the vessel. The principle of capacitance level sensing is based on the formula below:

Capacitance = Dielectric constant x (Area of the plates ÷ Distance between plates)

If the dielectric is a liquid, the capacitance probe can measure the combined capacitance of both the liquid and gas. When the liquid level rises or falls, the total capacitance value will change. Since the dielectric constant and distance between the plates are constant (the rod and the vessel are stable), the only value changing the capacitance is the area of the plates immersed in the liquid. Total capacitance changes approximately proportionally to the liquid rise or fall in the column. Consequently, the liquid level can be calculated by the change in capacitance.

Common Uses

The list below shows just a few applications of capacitance level sensors found in industry and the materials they sense.

- Chemical/Petrochemical - Oil, clay, soda ash
- Food - Flour, Powered Milk, Sugar
- Charcoal - Wood, Charred sawdust
- Pharmaceuticals - Various powders and liquids
- Mining - Various minerals, metals, stone,

Benefits

Capacitance level sensing is useful in its ability to sense a wide variety of materials such as solids, organic and aqueous solutions, and slurries. For example, materials with dielectric constants as low as 2.1 (petroleum oil) and as high as 88 (water) or more can be detected with capacitance level sensing. The equipment, typically made of stainless steel, is simple to use, clean and can be specifically designed for certain applications. For example, they can be made to withstand high temperatures and pressures or have built-in splashguards or stilling wells for environments prone to turbulence.

Restrictions

There are limitations to using capacitance level sensors. One major limitation for capacitance probes is found when using tall bins that store bulk solids. Probes are required to extend for the length of the vessel, so in a lengthy vessel the long cable probes can be subject to mechanical tensions and breakage. Another limitation is from build up and discharge of a high-voltage static charge that can result from the rubbing and movement of low dielectric materials, but this danger can be eliminated with proper design and grounding. Also, abrasion, corrosion, and build up of material on the probe can cause variations in the dielectric constant of the material being measured. To reduce this issue, capacitance probes can be coated with Teflon, Kynar, polyethylene or other materials.

4.6 Radiation-based Level Sensors

Radiation-based level sensors are based on the principle of a material's ability to absorb or reflect radiation. The common types of radiation used in continuous level gages are ultrasonic, radar / microwave and nuclear.

4.6.1 Ultrasonic (Sonic) Level Sensors

Ultrasonic level sensor transmitters emit high frequency ultrasonic acoustic waves which are reflected back by the medium to the receivers. By measuring the time it takes for the reflected echo to be received, the sensor can calculate the actual distance between the receiver and the fluid level. These sensors can be accurate from a distance of 5mm to 30m.

Common Uses

Ultrasonic level sensors are commonly used for point level detection. They are best used for viscous liquids, slurries and bulk solids.

Benefits

The combination of ultrasonic level sensors' high functionality for relatively low prices makes them a popular choice for non-contacting level sensing.

Restrictions

Since the speed of sound in air can fluctuate in different conditions, ultrasonic sensors are not suitable for use in all applications. Environments that have varying moisture and temperatures will influence the sensors readings. Turbulence, steam, and foam prevent waves from reflecting properly and distort readings. In addition, the level sensor must be mounted properly so that it correctly senses the distance between the transmitter and the fluid level.

4.6.2 Microwave / Radar Level Sensors

Microwave / radar level sensors are similar to ultrasonic level sensors in that they require a transmitter and receiver. In addition to these materials, radar sensors also need an antenna and operator interface to use electromagnetic waves to calculate level distance.

Common Uses

These sensors are frequently used for non-contacting situations that require level sensing in varying temperature and pressure environments.

Benefits

Microwave / radar sensors have an advantage over ultrasonic sensors in that they are able to operate in high pressure and high temperature environments. They can also sense solids with grain sizes larger than 20mm.

Restrictions

When fluid level is under a thick layer of foam or dust, these sensors may not detect the fluid level, but instead detect the level of the dust or foam. Microwave / radar level sensors are also more costly than ultrasonic level sensors.

4.6.3 Nuclear Level Sensors

Nuclear level sensors rely on gamma rays for detection. Although these gamma rays can penetrate even the most solid of mediums, the intensity of the rays will reduce in passage. If the gamma ray emitter and the detector are placed on the top and bottom of a vessel, the thickness of the medium (level) can be calculated by the change in intensity.

Common Uses

These sensors are usually used when the material being measured presents a risk to human life or the environment. These include, for example, materials that are toxic, carcinogenic, or explosive.

Benefits

Nuclear radiation has the ability to detect level even through solid tank walls. Since they appear to "see" through walls, the nuclear gage may be modified and/or installed while the process is running and avoid expensive down time.

Restrictions

Nuclear level sensors are typically the last resort when choosing a level sensor. They not only require a Nuclear Regulatory Commission (NRC) license to install, but are extremely expensive in comparison to other level sensors.

4.7 Summary of Benefits and Restrictions

	Benefits	Restrictions
Visual	- Economical, less expensive - Not subject to much technical/electronical issues	- Require human operator - No connection to process control systems - Cannot be read remotely - Connections between tube and tank subject to clogging, freezing, etc - Require modification to vessel wall
Float	- Do not require external energy source to operate - Economical - Easy to repair	Should be used in clean fluids
Valve Controlled	- Do not require external energy source to operate - Can be hooked up to process control systems - Allow for two-way flow - Eliminate pressure issues	- Should not be used in pressurized vessels - Require frequent maintenance - Costly
Electronic	**Conductive Level Sensing** - Considerably safe, low voltage and current - Easy installation **Capacitance Level Sensing** - Able to sense a wide variety of materials - Flexible in design	- Maintenance - Probes must extend for full length of vessel, prone to mechanical damage
Radiation	**Ultrasonic Level Sensors** - Non-contact level sensing - Low cost **Microwave/Radar Level Sensors** - Able to operate in high pressure, high temperature, and vacuum environments - Able to sense solids with grain size > 20 mm **Nuclear Level Sensors** - Able to detect even through solid tank walls - Able to be modified while process is running to avoid expensive down time	- Speed of sound fluctuates depending on conditions - Readings can be influenced by environment - More costly - Cannot work if liquid level is under thick layer of airy foam or dust - Require licensing - Extremely expensive

Imagine that you are responsible for maintaining a certain amount of chemical in a storage vessel for use later on downstream in your process. Your chemical has a specific gravity of 1.2. Assume that it is effectively independent of temperature. At maximum flow rate, a valve is feeding 80 gallons per minute into a storage vessel. 2 inches of valve travel is required to completely shut off the flow through the valve. The spherical float you are given to work with is 8 inches across, chemically compatible with your chemical, and its specific gravity can be altered. Your process dictates that you have to maintain about 400 gallons in this tank for use downstream, and that corresponds to a height of twenty inches from the bottom of the tank. At this height, the ball is 4 inches submerged. The ball has to travel 5 inches to completely close your valve. Calculate:

a) the specific gravity the float should be

b) the gain of the regulator

c) the proportional sensitivity of your system

d) assume the specific gravity of your chemical increases dramatically. Explain qualitatively what that should do to the gain, zero buoyant force level and dead band of the sensor

e) was a float type valve appropriate to be used in this problem?

4.8.1 Answers

a) Using the formula for Zero Buoyant Force level

$$ZeroBuoyantForceLevel = D\frac{(SG_f)}{SG_p}$$ and plugging in values

from the problem statement, we come up with the following

$$4in = 8in\frac{(SG_f)}{1.2}$$ We used 4 inches, and not 20 inches because the zero buoyant force level indicates the level the float is submerged, not the level of the fluid. Isolating the value for SGf, we arrive at a specific gravity of 0.6 for the float.

$$Gain = \frac{(valve_{response})}{float_{response}}$$

b) Using the formula for the gain of the regulator and values from the problem statement, we come up with the following

$$Gain = \frac{(2in)}{5in}$$ proceeding with the calculation we arrive at a value of 0.4 for the gain.

c) Using the formula for the proportional sensitivity of the system $K_c = \frac{(q)}{k}$ and values from the problem statement, we come up with the following

$$K_c = \frac{(80gpm)}{2in}$$ proceeding with the calculation we arrive at a value of

$$K_c = \frac{(40gpm)}{in}$$

d) Float type level sensors work on the principal of buoyancy. If the density of the liquid being displaced increases, the buoyant force on the float increases for the same change in liquid height. The gain of the sensor refers to the amount of travel in the valve versus the amount of travel in the float, so for a given change in fluid height, the valve will still change the same amount as it did. However, since the buoyant force is increased due to the increased process fluid density, the dead band should decrease because the force of an incremental increase in height should enable the sensor to overcome friction more easily.

e) A float type valve was appropriate to be used in this problem, because the requirements were that "about 400 gallons" be maintained and the fluid was chemically compatible with the float. Due to unavoidable dead band, or the time it takes the float to respond to a given change in fluid, it won't maintain exactly 400 gallons, but it will stay close to that.

4.9 Example 2 - Sensor Selection

Please answer the questions for the four hypothetical scenarios regarding selection of an appropriate level sensor. Give reasons for your selections.

1) A commonly used tank contains Dichloromethane which requires constant monitoring to ensure that no leaks exist. The precise volume will be regulated to monitor any decrease in level when a known transfer is not taking place. This tank will be filled to ~5,000 gallons and not be refilled until the level is less than 1,000 gallons of dichloromethane. What type of sensor should you install?

2) Under what circumstances would you select a radiation sensor?

3) Toilets have a Float Level Sensor to indicate when a flush has occurred and the bowl needs to be refilled. Which type of Float Sensor does the toilet have?

4) When would be a good time to use a visual sensor?

4.9.1 Answers

1) The most appropriate selection of level sensor for this application is the capacitance sensor or the ultrasonic sensor. These are both electrical sensors that allow for measurements of changes in height, whereas other types offer a binary result: either the level is above or below a set point. The capacitance and ultrasonic sensors differ only by the way they are installed. Capacitance sensors require contact with the liquid while ultrasonic sensors do not. In this specific case for dichloromethane, which is non-corrosive, it is not necessary to monitor the volume without contacting the liquid, and therefore the capacitance sensor is also an acceptable selection.

2) A radiation-based level sensor is most appropriate when sensors cannot be brought into direct contact with the measured substance. This is only a possible option and often an expensive one.

3) Toilets have a lever/shaft mechanism. This float is less dense than the water which fills the tank. It will drop quickly when the toilet is flushed and will slowly rise as the tank fills with water. When the tank is full the float will rotate the shaft upwards enough to close off the flow of water.

4) Visual sensors should only be used when an electrical sensor is not required or as a backup. This is because they cannot be monitored automatically as visual sensors do not provide a digital response. It is also a good idea to provide a secondary way of obtaining level information when installing a new tank, and manual checks are often a good idea.

4.10 Sage's Corner

Automotive Applications of Level Sensors

http://www.youtube.com/v/AlMK92Ojoa4
Unnarated Powerpoint Presentation

4.11 References

- [Image Available] Seetru, http://www.seetru.com/index_main.html, 12 Sept 2006.
- Liptak, Bela G., Instrument Engineers Handbook, 4th Edition, Volume 2. Taylor & Francis CRC Press, 2005.
- Perry, Robert H., Perry's Chemical Engineers' Handbook, 7th Edition. McGraw-Hill, 1997.
- Richardson, J.F. and Peacock, D.G., Chemical Engineering, 3rd Edition, Volume 3. Butterworth-Heinemann, 1994.

- Transactions in Measurement & Control: Volume 4, 2006. Omega Engineering, Inc. Retrieved September 10, 2006, from http://www.omega.com/literature/transactions/volume4/
- Wikipedia. Retrieved September 9, 2007, from http://en.wikipedia.org/wiki/Level_sensor
- Considine, Douglas M., Process/Industrial Instruments & Controls Handbook, 4th Edition, McGRAW-HILL, INC., 1993.

Section 5. Flow Sensors

Authors: (September 14, 2006) Ashwini Miryala, Kyle Scarlett, Zachary Zell, Brandon Kountz

Stewards: (September 5, 2007) Jeremy Goldfarb, Daniel McClure, Lilian Ngobi, Daniel Wheeler

- [First round reviews for this page](#)
- [Rebuttal for this page](#)

5.1 Introduction

Flow is defined as the rate (volume or area per unit time) at which a substance travels through a given cross section and is characterized at specific temperatures and pressures. The instruments used to measure flow are termed flow meters. The main components of a flow meter include the sensor, signal processor and transmitter. Flow sensors use acoustic waves and electromagnetic fields to measure the flow through a given area via physical quantities, such as acceleration, frequency, pressure and volume. As a result, many flow meters are named with respect to the physical property that helps to measure the flow.

Flow measurement proves crucial in various industries including petroleum and chemical industries. Consequently, flow measurement becomes a major component in the overall economic success or failure of any given process. Most importantly, accurate flow measurements ensure the safety of the process and for those involved in its success.

Before reading about the intricate details of various flow meters it's a good idea to think about aspects other than the design, governing equations, and the mechanism a flow meter uses in identifying a flow profile. To gain more from this section consider the following questions when learning about each flow meter because when choosing a flow meter the two main things to think about are cost and application:

1. What mechanism does this flow meter use to measure flow?

2. How expensive is it?

3. What impact does it have on the system? (How much power does it withdraw from the system and does the disturbance of flow cause significant problems up or downstream?

4. What are the accuracy limits of the tool? (Increased accuracy and precision results in higher cost especially for those with automated noise filtering and signal amplifying capabilities.

Sometimes you may be given an assignment to purchase a tool for your company's system and given with it a list of qualities to look for in order of importance, sometimes

you're going to have to make this list yourself. The following table is a component of the TRIZ method for developing a list of qualities in descending order of importance on the left, perpendicular to the factor it has an impact on, to the right. For a given system you figured out that the cheapest solution for accurately measuring the flow rates of various pipe lines is by placing multiple flow meters in a series of positions. From a few calculations using propagation of error you find a moderate range of accuracy limits necessary for your system to be considered well monitored. You know the accuracy limits are less important than the total cost of all the flow meters because your boss told you not to waste any money. On the table you can see that **Cost** is at the top of the left column so **Cost** is the main concern. **Accuracy** is in the farthest left column on the top row so **Cost** most dramatically exacerbates the **Accuracy** of the flow meter when it's minimized. In the intersecting box you see that the lowest price flow meters are the ones with the least amount of accuracy, generally speaking. The next most important quality of your assignment is to get flow meters with the appropriate accuracy, thus **Accuracy** is below **Cost** in the left column. Looking to the top row you see that the **Accuracy** of the flow meter most greatly effects the **Impact on the system**. If you have a low quality flow meter it may be due to the side of the mechanism used to measure flow or the power the meter draws from the system (through mixing, frictional losses, increase in the turbulence of the flow, or buoyant effects caused from heat transfer). Completing the rest of the table you can decide if there are contradictions to what you thought were the most important qualities based on the inputs in the intersecting cells.

	Accuracy	Impact on system	Cost	Maintenance
Cost	Low accuracy/precision meters costs the least	Lower costs will result in larger meters that increase it effects on flow		Lowest cost means the flow meter will need frequent Maint.
Accuracy		Higher accuracy low cost flow meters have more surface area	What is the right price? How much error can I afford?	How long will it be reliable before it needs to be cleaned and recalibrated?
Maintenance	Application dependent, more Maint. Increases reliability	Down time, accumulation of deposits, disturbing flow	Down time, buildup on meter increases drag	
Impact on system	Lower impact may mean lower accuracy		Are there other costs of impact from installing the flow meter	How intricate is the meter, can it be quickly cleaned and recalibrated

An example where this model is important is for a system containing suspended particles in a fluid. If you want to measure the flow rate it might be cheapest to use a pitot tube yet the increased cost of maintaining the flow meter extends into down time of the system, more work from the technicians for a relatively small aspect of the process and lower profits as a result. For this system, maintenance would be the most important factor followed by accuracy, impact on the system and cost. If cost were least important factor you could afford to install a couple doppler meters, gather accurate measurements and the impact of the measurements on the system would be relatively low.

5.2 Common Types of Flow Meters

The flow rate as determined by the flow sensor is derived from other physical properties. The relationship between the physical properties and the flow rate is derived from fundamental fluid flow principles, such as Bernoulli's equation.

239

These sensors work according to Bernoulli's principle which states that the pressure drop across the meter is proportional to the square of the flow rate.

$$-\Delta p \propto V^2$$

Using the pressure drop across a pipe's cross section is one of the most common manners to determine a flow measurement. As a result, this property has heavy implications for industrial applications. Flow meters characterized by differential pressure come in several different varieties and can be divided into two categories, laminar and turbulent. Differential pressure sensors operate with respect to Bernoulli's principle. Bernoulli's principle states that the pressure drop across the meter is proportional to the square of the flow rate.

Orifice Meter

Orifice plates are installed in flow meters in order to calculate the material balances that will ultimately result in a fluid flow measurement on the sensor. An orifice plate is placed in a pipe containing a fluid flow, which constricts the smooth flow of the fluid inside the pipe. By restricting the flow, the orifice meter causes a pressure drop across the plate. By measuring the difference between the two pressures across the plate, the orifice meter determines the flow rate through the pipe. The larger the pressure drop, the faster the flow rate would be. There are two types of orifice meters that are usually used in industry, they are the orifice-square edge and the orifice-conic edge. The orifice-square edge has insignificant friction at the interface between the fluid and the orifice plate. These types of orifice plates are recommended for smooth fluid flows, particularly clean liquids and gases such as steam. Generally, drain holes are incorporated in the design so that liquids and gases do not accumulate inside the pipes. Multi-phase fluids are not recommended for the orifice-squared edge because clogging becomes a significant problem as time progresses. The orifice-conic edge is similar to the orifice-square edge, the primary difference being that the orifice-conic edge has a gradually expanding cross-sectional area, and this cross-sectional area is circular in shape. A conic-edge design is often a better choice for low velocity, high viscosity flows. Both types operates best under comparable temperature and pressure conditions, pipe sizes and provide similar accuracies.

Orifice meters used in conjunction with DP (Differential Pressure) cells are one of the most common forms of flow measurement. In addition, an orifice meter can be used to measure flows when there is a significant difference in pressure in the pipe, like between the upstream and downstream sides of a partially obstructed pipe, which is exactly what the orifice meter does on its own. The plate offers a precisely measured obstruction that essentially shrinks the pipe and forces the flowing substance to constrict. A DP cell allows the comparison of pressure on the upstream (unobstructed) side and the downstream (constricted) side. A greater rate of fluid flow would usually result in a

larger pressure drop, since the size of the orifice remains constant and the fluid is held longer building potential energy on the upstream side of the orifice. Some of the other types of orifice plates include concentric, eccentric and segmental plates, each having different shapes and placements for measuring different processes. These plates are available in varied shapes so that the meter has the optimum structure for different applications. Moreover, the density and viscosity of the fluid, and the the shape and width of the pipe also influences the choice of plate shape to be used.

Orifice-Square Edge   Orifice-Conic Edge

Courtesy of
Superior Products, Inc.
www.orificeplates.com
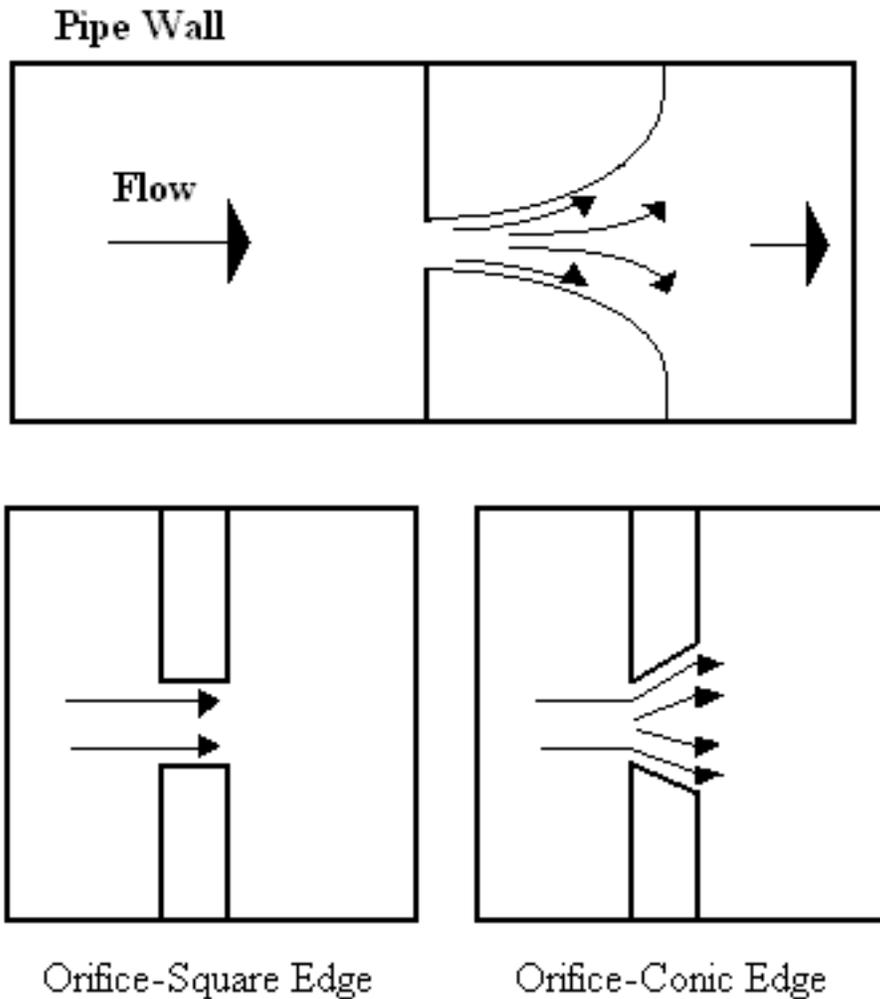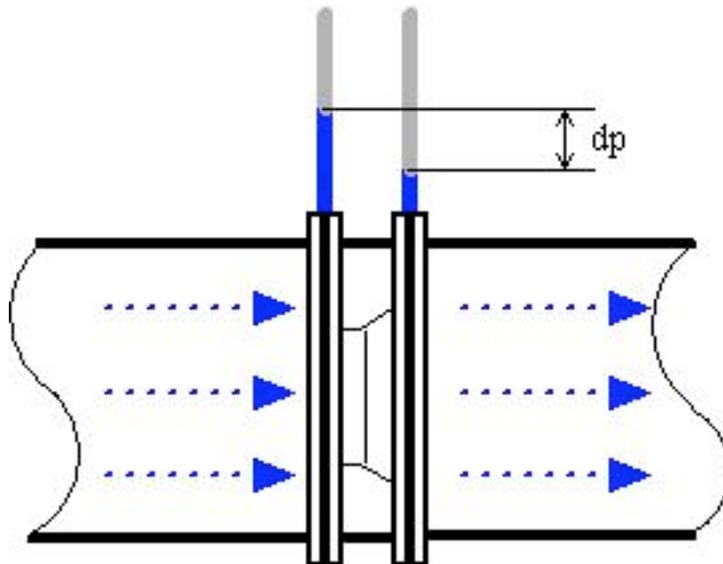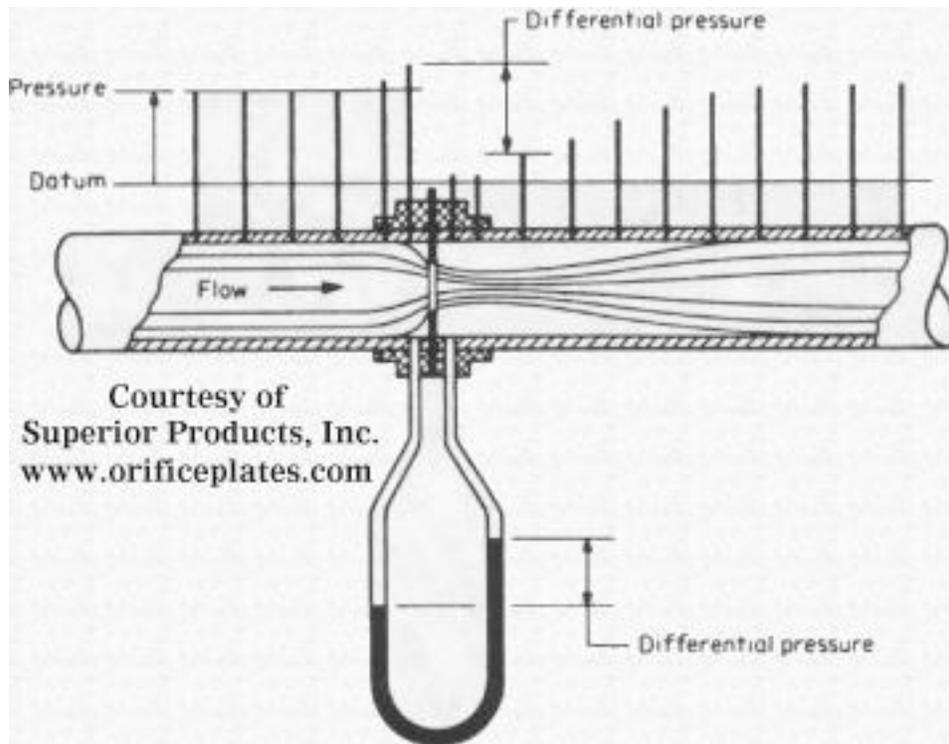
www.EngineeringToolBox.com

Such a pressure drop across the plate is then related to the flow rate using basic fluid mechanics principles that incorporate parameters such as density of the fluid and size of the pipe. The flow rate Q, given by the orifice meter, is usually modeled with the following equation:

242

$$Q = \frac{C_d A_2}{\sqrt{1 - (A_2/A_1)^2}} \sqrt{\frac{2(p_1 - p_2)}{\rho}}$$

Where $p_1 - p_2$ is the pressure drop across the plate, ρ is the fluid density, A_1 is the pipe cross-sectional area, A_2 is the orifice cross-sectional area, and C_d is the discharge coefficient (usually about 0.6). C_d is used to account for energy losses within the system.

The orifice meter is one of the most commonly used flow meters, since it is inexpensive to install and operate, it is uncomplicated and easy to construct, and it is highly robust and long lasting. Orifice meters are not only simple and cheap, they can also be delivered for almost any application and be made of any material. This simplicity of its design and function is one of its paramount advantages, with the meter essentially consisting of just a modified plate. This not only reduces its initial price but also shrinks its operating costs, maintenance expenses, and spare parts expenditure and availability. Lower flow rates reduces their accuracy, whereas higher flow rates combined with high quality, unworn orifice plates increases it. The orifice plate is best when a sharp edge is present towards the upstream side of the meter. Wear reduces the accuracy of orifice plates. The turndown rate of orifice plates are generally less than 5:1. More information about turndown rates is mentioned in a later section.

Venturi Meter

Venturi meters can pass 25 – 50% more flow than an orifice meter. In a Venturi meter setup, a short, smaller diameter pipe is substituted into an existing flow line. Because the Venturi meter is insensitive to changes in the velocity profile of the fluid flow, the pipe design does not need to be straight like the orifice meter. Though initially expensive, the Venturi meter has relatively low maintenance and operation costs.

In the Venturi Tube the fluid flowrate is measured by reducing the cross sectional flow area in the flow path, generating a pressure difference. After the constricted area, the fluid is passes through a pressure recovery exit section, where up to 80% of the differential pressure generated at the constricted area, is recovered.

There are two main types of Venturi meters. The first one, known as the classical Herschel Venturi meter, is a very long meter characterized below. Pressure readings at different points in the meter are combined to provide an average pressure reading. Cleaning the classical Venturi meter is limited. The second type of Venturi meter is known as the short form Venturi meter. This differs from its longer counterpart by reduced size and weight.

By Bernoulli's principle the smaller cross-sectional area results in faster flow and therefore lower pressure. The Venturi meter measures the pressure drop between this constricted section of pipe and the non-constricted section.
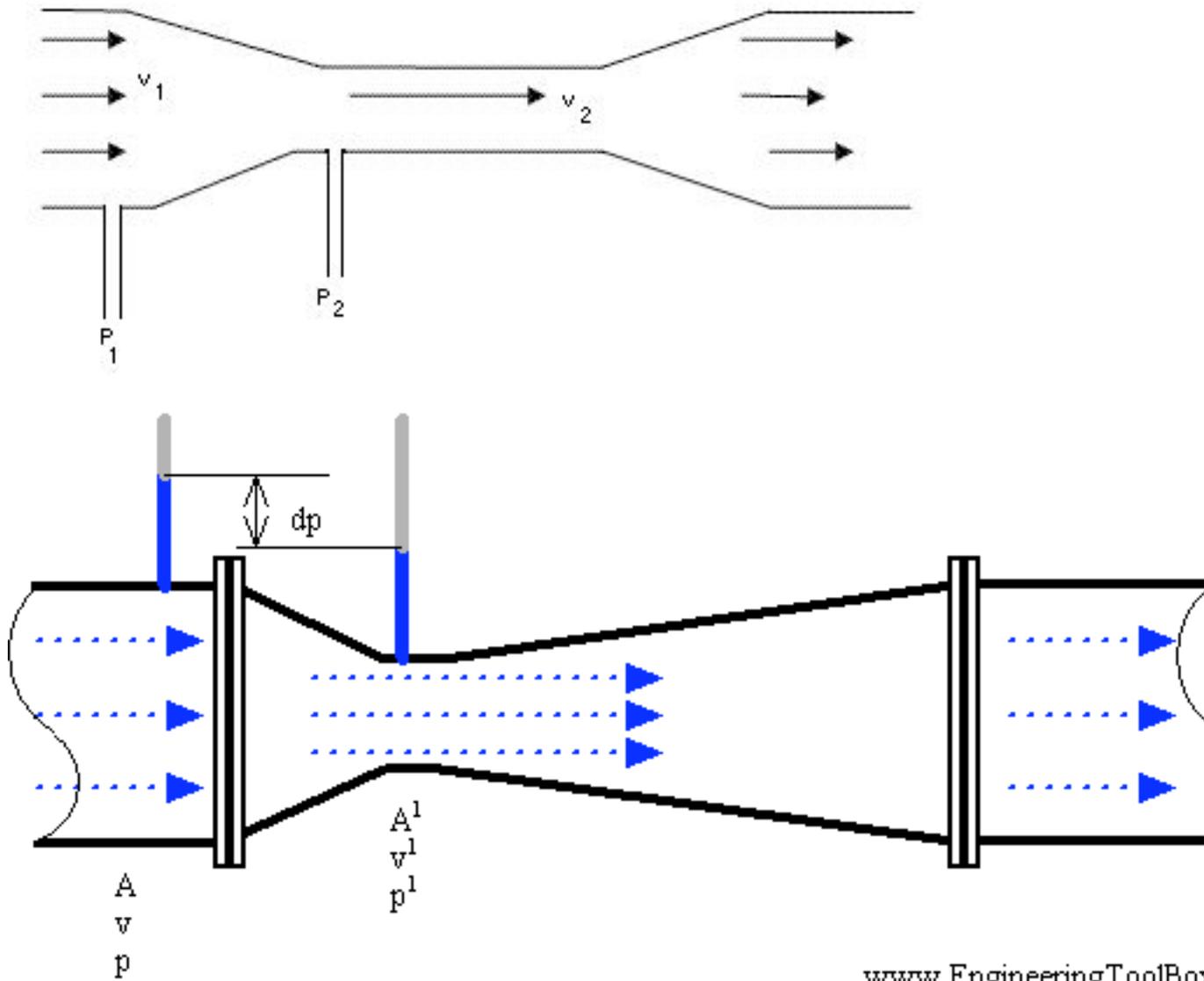
www.EngineeringToolBox

The discharge coefficient for the Venturi meter is generally higher than that used for the orifice, usually ranging from .94 to .99.

The Venturi meter is most commonly used for measuring very large flow rates where power losses could become significant. The Venturi flow meter has a higher start up cost than an orifice, but is balanced by the reduced operating costs.

Due to simplicity and dependability, the Venturi tube flowmeter is often used in applications where higher turndown ratiosor lower pressure drops than orifice plates can provide are necessary. With proper instrumentation and flow calibrating the venturi meter flowrate can be reduced to about 10% of its full scale range with proper accuracy. This provies a turndown ratio of around 10:1.

Flow Nozzle

Another type of differential pressure flowmeter is the flow nozzle. Flow nozzles are often used as measuring elements for air and gas flow in industrial applications. At high velocities, Flow Nozzles can handle approximately 60 percent greater liquid flow than orifice plates having the same pressure drop. For measurements where high temperatures and velocities are present, the flow nozzle may provide a better solution than an orifice plate. Its construction makes it substantially more rigid in adverse conditions and the flow coefficient data at high Reynolds numbers is better documented than for orifice plates. Liquids with suspended solids can also be metered with flow nozzles. However, the use of the flow nozzles is not recommended for highly viscous liquids or those containing large amounts of sticky solids. The turndown rate of flow nozzles is similar to that of the orifice plate. The flow nozzle is relatively simple and cheap, and available for many applications in many materials.
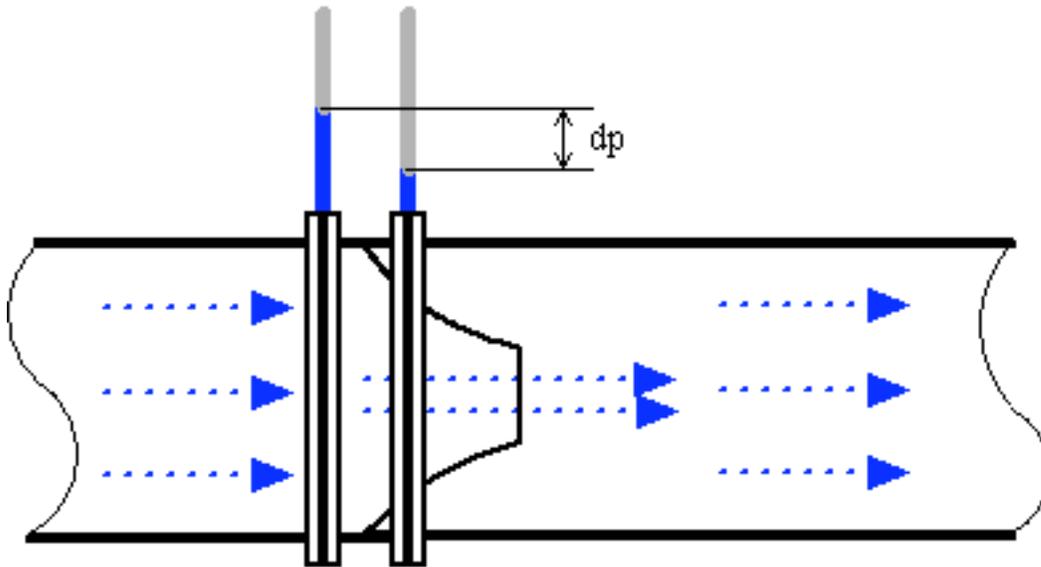
www.EngineeringToolBox.com

The Sonic Nozzle- Critical(Choked)Flow Nozzle

One type of flow nozzle is the sonic nozzle. The Sonic Nozzle is a converging-diverging flowmeter. It consists of a smooth rounded inlet section converging to a minimum throat area and diverging along a pressure recovery section or exit cone.
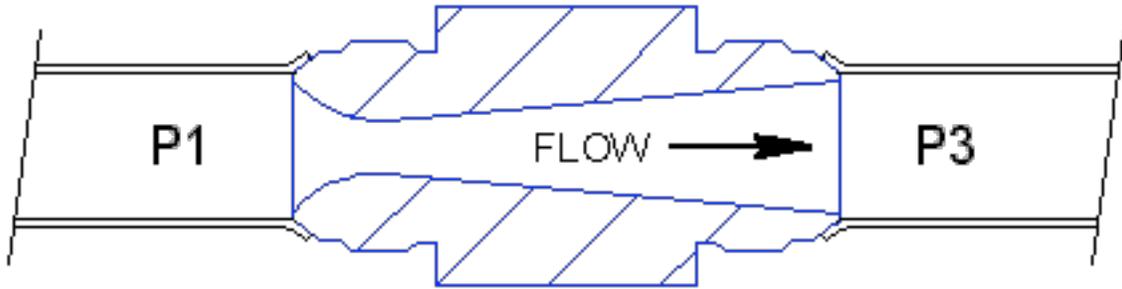
The Sonic Nozzle is operated by either pressurizing the inlet (P1) or evacuating the exit (P3), to achieve a pressure ratio of 1.2 to 1 or greater, inlet to outlet. When a gas accelerates through a nozzle, the velocity increase and the pressure and gas density decrease. The maximum velocity is achieved at the throat, the minimum area, where it breaks Mach 1 or sonic. At this point it's not possible to increase the flow by lowering the downstream pressure. The flow is choked. Pressure differences within a piping system travel at the speed of sound and generate flow. Downstream differences or disturbances in pressure, traveling at the speed of sound, cannot move upstream past the throat of the Nozzle because the throat velocity is higher and in the opposite direction.

Sonic Nozzles are used in many control systems to maintain fixed, accurate, repeatable gas flow rates unaffected by the downstream pressure. If you have a system with changing or varying gas consumption downstream and you want to feed it a constant or locked flowrate, a Sonic Nozzle is an excellent way to achieve this.

Pitot Tubes

Pitot tubes measure the local velocity due to the pressure difference between points 1 and 2 in the diagrams below. Unlike the other differential flow meters, the pitot tubes only detect fluid flow at one point rather than an overall calculation. The first diagram shows a simple pitot tube configuration while the second shows a compact pitot tube configuration.
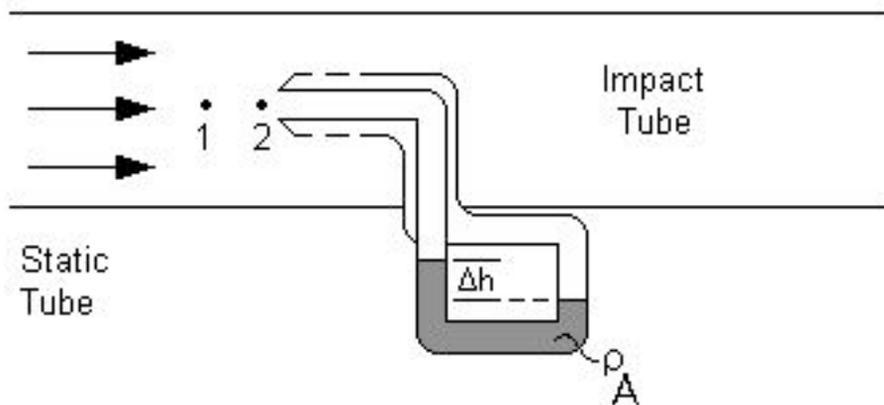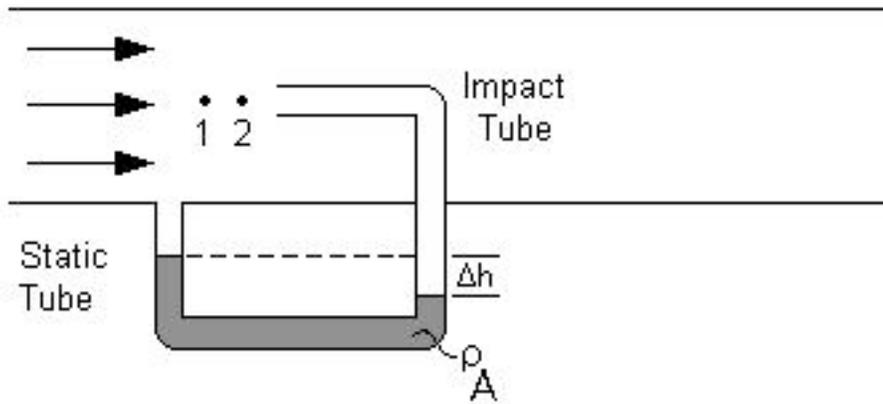
Both tubes work in a similar manner. Each pitot tube has two openings, one perpendicular to the flow and one parallel to the flow. The impact tube has its opening perpendicular to the fluid flow, allowing the fluid to enter the tube at point 2, and build up pressure until the pressure remains constant. This point is known as the stagnation point. The static tube, with openings parallel to the fluid flow gives the static pressure and causes a sealed fluid of known density to shift in the base of the tube. Pressure drop can be calculated using the height change along with the fluid densities and the equation below.

$$\Delta p = \Delta h (\rho_A - \rho) g$$

with Δp as the pressure drop, ρ_A as the known fluid density, ρ as flowing fluid's density, and g as the acceleration due to gravity.

This pressure drop can be related to the velocity after accounting for the losses throughout the piping in the system, given by C_p. This dimensionless coefficient is found through accurate calibration of the pitot tube. The equation below describes this relationship.

$$v = C_p \sqrt{\frac{2(p_1 - p_2)}{\rho}}$$

with v as the fluid velocity, C_p as the loss coefficient, p_1 as the pressure at point 1, p_2 as the pressure at point 2, and ϱ as the flowing fluid's density.

By placing the tube at the exact center of the pipe, the maximum velocity can be measured and the average velocity can be calculated via the Reynolds number. The governing equation and chart are below.

$$Re = \frac{D v_{max} \rho}{\mu}$$

with Re as the Reynolds number, D as the pipe diameter, v_{max} as the maximum velocity, ϱ as the flowing fluid's density, and μ as the flowing fluid's viscosity.
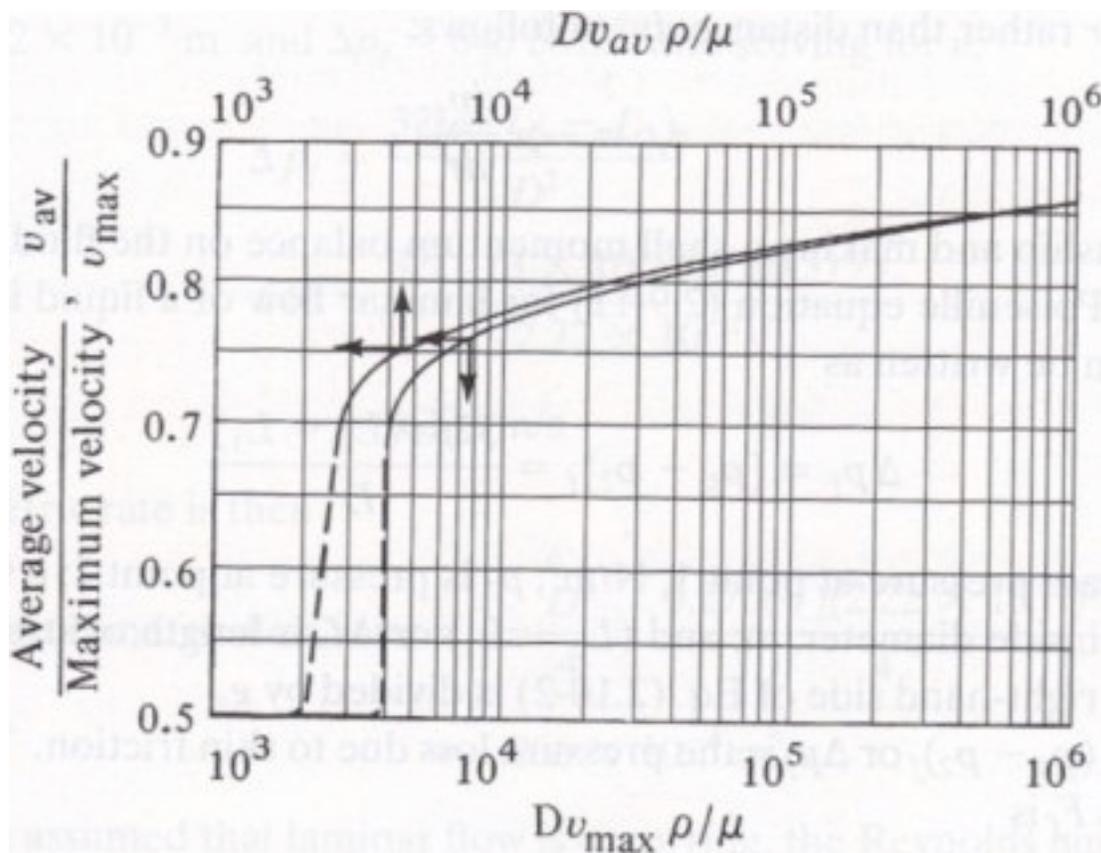
Adapted from Geankoplis.

Finally, the flow rate can be found by accounting for the area of the pipe.

$$Q = v_{avg}\frac{\pi}{2}r^2$$

with Q as the volumetric flow rate, v_{avg} as the average velocity, and r as the pipe's radius.

It should be noted that all the equations apply to incompressible fluids only, but these can be used to approximate gas flows at moderate velocities. This flow meter must also be placed at least 100 pipe diameters in distance, downstream of the nearest flow obstruction. This ensures no unwanted pressure fluctuations and accurate pitot tube readings. Furthermore, fluids containing large particles can clog the pitot tube and should be avoided.

5.2.2 DIRECT FORCE

These flow meters are governed by balancing forces within the system.

Rotameter

A rotameter is a vertically installed tube that increases in diameter with increasing height. The meter must be installed vertically so that gravity effects are easily incorporated into the governing equations. Fluid flows in through the bottom of the tube and out through the top. Inside the glass tube there is a float that changes position with the flow rate. When there is no liquid flow, the float rests in the bottom of the meter.
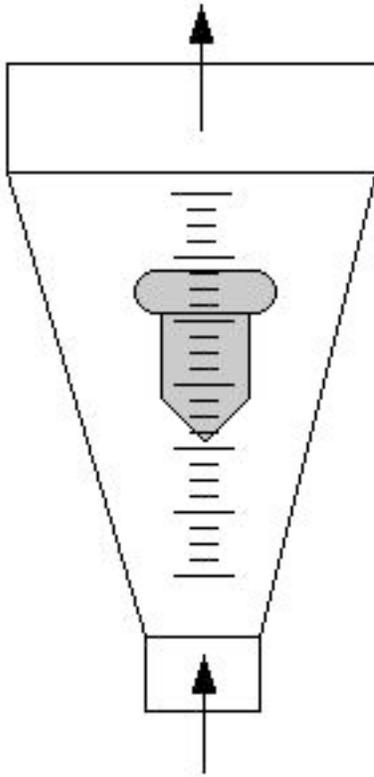
The applied concept for the rotameter is differential area. As the flow rate of the fluid changes, the position of the float changes and annular area change directly, keeping a constant pressure drop across the meter. Changes in float position and annular area are approximately linear with changes in flow rate. Upon achieving a stable flow rate, the vertical forces are balanced and hence the position of the float remains constant. The volumetric flow is given by the following equation:

$$Q = CA_b \sqrt{\frac{2g\left(\frac{V_f(\rho_f - \rho)}{A_f} - \rho h_f\right)}{\rho\left[1 - \left(\frac{A_b}{A_a}\right)^2\right]}}$$

with C being the discharge coefficient, A_b being the cross sectional area of the top of the float, V_f volume of the float, ϱ_f the density of the float, ϱ the density of the fluid h_f the height of the float, A_a the cross sectional area of the bottom of the float.

Generally, rotameters are inexpensive and simple to use. This allows them to be used in many plant applications.

250

For the full derivation of the above equation, refer to the attachment.

[1]

Turbine Meter

A turbine wheel is placed in a pipe that holds the flowing fluid. As the fluid flows through the turbine, the turbine is forced to rotate at a speed proportional to the fluid flow rate. A magnetic pick-up is mounted to the turbine wheel, and a sensor records the produced voltage pulses. Voltage information can then be translated into the actual flow meter reading.
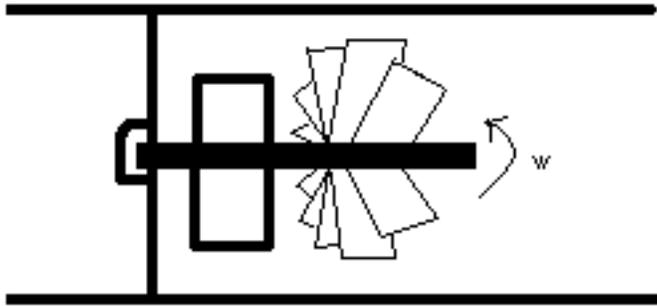
The following is the equation used to model the turbine meter:

$$Q = VA = \frac{\omega \left(\bar{r}\right)^2 A^2}{\bar{r} A tan\beta - 0.037 Re^{-0.2} n \left(R_o + R_i\right) D sin\beta}$$

with A the pipe area, $\bar{r}$ the root mean squared radius, ω rotational speed, β the angle between the flow direction and the turbine blades, R_o the outer blade radius, R_i the inner radius, and D the distance between blades.

There are two main advantages of the tubine meter over conventional differential head devices

1) The extended are more accurate registatrion of flow in the low flow range of process operation. This results from the registration being proportional to the velocity rather than the velocity square

2) The comparatively low head loss across the meter

Another advantage to using this type of flow meter is reliability. Extensive testing has proven these results. Additionally, the turbine flow meter does not have a high installation cost. However, due to the turbine wheel motion, a low to medium pressure drop can result. Turbine wheel replacement may also be required due to abrasion caused by particles within the fluid.

For a more detailed derivation of the above equation, please look at the following:

[2]

Propeller Flow Meter

Propeller flow meters have a rotating element similar to the wheel in turbine meters. Again, rotation is caused by fluid flow through the propeller, and voltage pulses are created as the propeller passes a magnetic or optical sensor. Similarly, the frequency of the pulses is proportional to flow rate of the fluid and the voltages can be directly correlated with the fluid flow rate.
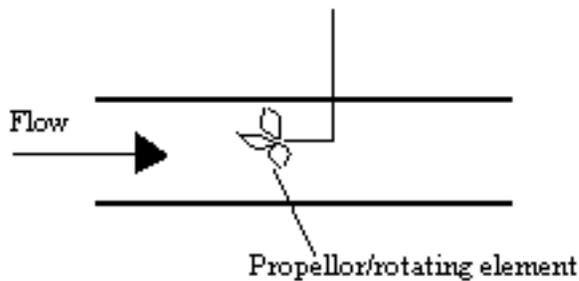
Propeller flow meters are often used specifically with water, though other fluids may also be used. Low cost coupled with high accuracy make propeller flow meters a common choice in many applications.

Paddle Wheel Sensors

A kind of propeller sensor is the paddle wheel sensor. Paddle wheel sensors are similar to turbine sensors, except for one thing. The shaft of the paddle wheel sensor is perpendicular to the flow of the fluid while the turbine sensor's shaft is parallel to the flow of the fluid. This adds several advantages for the paddle wheel flow sensor. Due to the shaft being perpendicular to the flow, it sustains less axial from the fluid, and thus less friction. Paddle wheel sensors also have a smaller number of blades, so there is less force needed to turn the paddle wheel. This means that a paddle wheel can be accurate at lower flows, have a high efficiency, as well as a longer lifetime.

There are two kinds of paddle wheel sensors, insertion and inline sensors. There is more than one design for an insertion sensor, but one popular design has the bearing built into

252

the rotor and the shaft sliding through it as the center axis of spin. The blade sticks out and is inserted into the pipe with the flowing fluid.

An inline paddle wheel sensor is used for smaller pipes. It contains a rotor assembly with the magnet sealed inside and a main body.

For detailed mathematics of paddlewheels, see [Paddlewheel Flow Sensors: The Overlooked Choice](#)

Coriolis Mass Flow Meter

A Coriolis flow meter harnesses the natural phenomenon wherein an object will begin to "drift" as it travels from or toward the center of a rotation occurring in the surrounding environment. A merry-go-round serves as a simple analogy; a person travelling from the outer edge of the circle to its center will find himself deviating from his straight-line path in the direction of the ride's rotation.

Coriolis flow meters generate this effect by diverting the fluid flow through a pair of parallel U-tubes undergoing vibration perpendicular to the flow. This vibration simulates a rotation of the pipe, and the resulting Coriolis "drift" in the fluid will cause the U-tubes to twist and deviate from their parallel alignment. This Coriolis force producing this deviation is ultimately proportional to the mass flow rate through the U-tubes.

$$MassFlow = \frac{F_c}{2wx}$$

where Fc is the Coriolis force observed, w is the angular velocity resulting from rotation, and x is the length of tubing in the flow meter.
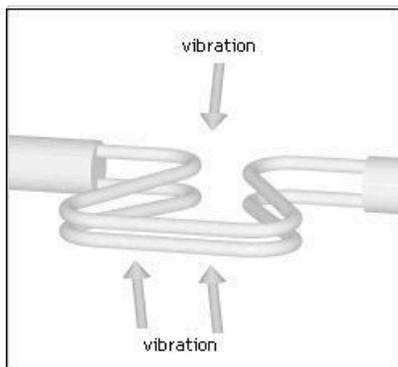
Coriolis flow meter undergoing no flow.

Adapted from Wikipedia's entry on Coriolis Mass Flow Meters.
http://en.wikipedia.org/wiki/Mass_flow_meter
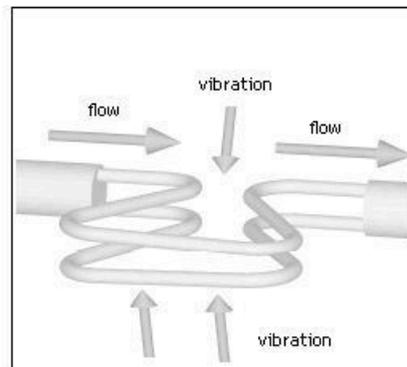
Coriolis flow meter exhibiting deflection as a result of mass flow

253

Because the Coriolis flow meter measures the mass flow rate of the fluid, the reading will not be affected by fluctuations in the fluid density. Furthermore, the absence of direct obstructions to flow makes the Coriolis flow meter a suitable choice for measuring the flow of corrosive fluids. Its limitations include a significant pressure drop and diminished accuracy in the presence of low-flow gases.

To understand how the mass flow rate is measured with this device, refer to the following attachment.

[3]

5.2.3 FREQUENCY

These flow meters use frequency and electronic signals to calculate the flow rate.

Vortex Shedding Flow Meter

A blunt, non-streamline body is placed in the stream of the flow through a pipe. When the flow stream hits the body, a series of alternating vortices are produced, which causes the fluid to swirl as it flows downstream. The number of vortices formed is directly proportional to the flow velocity and hence the flow rate. The vortices are detected downstream from the blunt body using an ultrasonic beam that is transmitted perpendicular to the direction of flow. As the vortices cross the beam, they alter the carrier wave as the signal is processed electronically, using a frequency-to-voltage circuit. The following diagram shows the basic principle of the vortex-shedding flow meter:
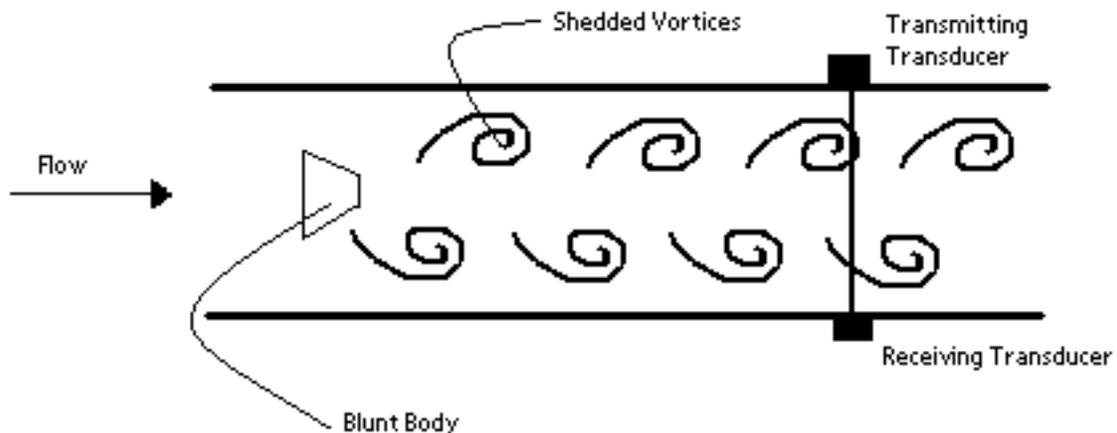
Vortex-shedding flow meters are best used in turbulent flow with a Reynolds number greater than 10,000. One advantage of using this type of flow meter is its insensitivity from temperature, pressure, and viscosity. The major disadvantage to using this method is the pressure drop caused by the flow obstruction.

254

There are two types of Ultrasonic meters, the transit time/time of flight and Doppler models, both of which have unique equations representing the principles behind them. The basis for these meters is monitoring ultrasonic waves in fluid passing through a pre-configured acoustic field. These meters are based on the technique of sound waves that change.

5.3.1 Transit Time/Time of Flight Flow Meters

Transit time meters have two opposing transducers outside of the pipe to measure the time of a signal sent from a transducer upstream to a transducer downstream and vice versa.
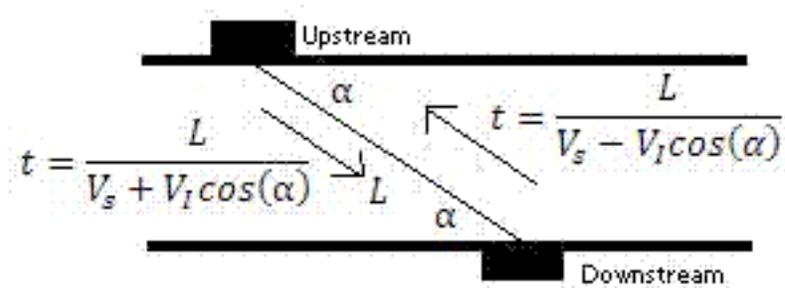
This allows the average velocity and hence the flow rate, Q, to be determined.

$$Q = K \left(\frac{\pi d^3 \tan\alpha}{8} \right) \left(\frac{1}{t_{UD}} - \frac{1}{t_{DU}} \right)$$

where d is the diameter of the pipe, α is the angle between direction of the flow and the pipe, t_{UD} is the time for the signal to reach downstream transducer from the upstream transducer, and t_{DU} is the time for signal to reach upstream transducer from the downstream transducer.

With the Time of Flight Ultrasonic Flowmeter the time for the sound to travel between a transmitter and a receiver is measured. This method is not dependable on the particles in the fluid.
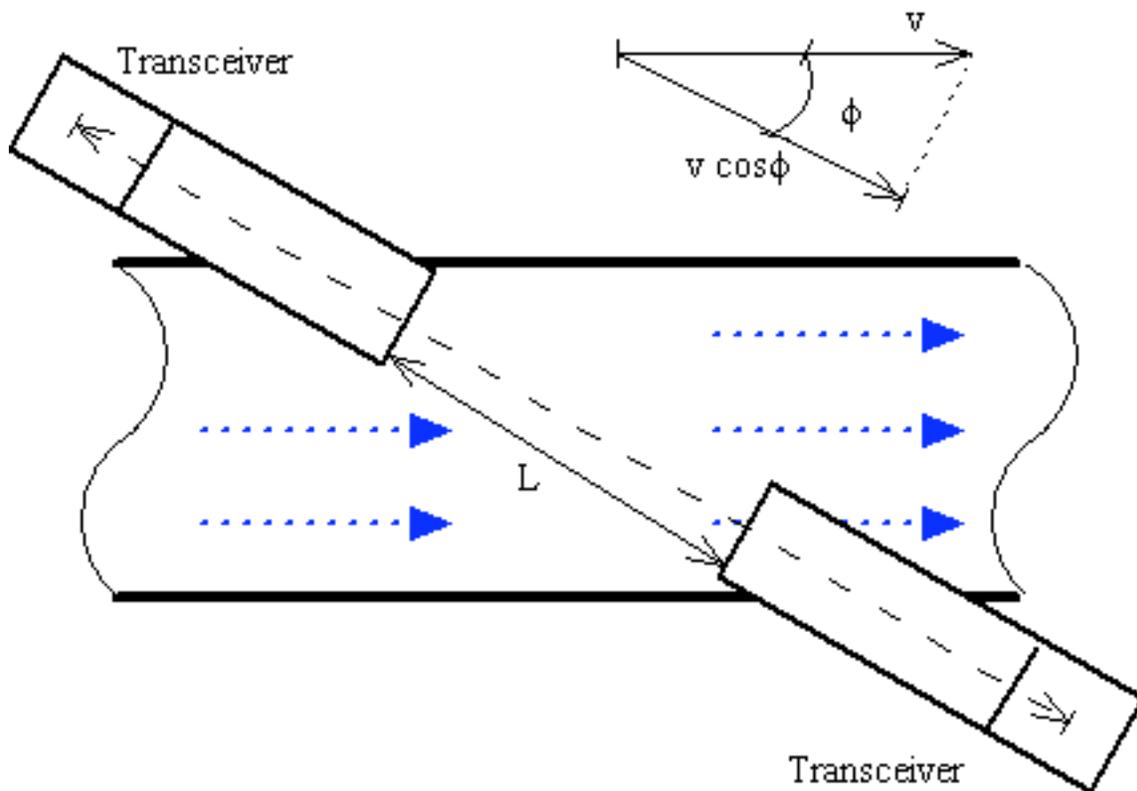
www.EngineeringToolBox.com

Two transmitters / receivers (transceivers) are located on each side of the pipe. The transmitters sends pulsating ultrasonic waves in a predefined frequency from one side to the other. The difference in frequency is proportional to the average fluid velocity.

Downstream pulse transmit time can be expressed as

$$td = L / (c + v \cos\Phi)$$

where td = downstream pulse transmission time L = distance between transceivers

Downstream pulse transmit time can be expressed as

$$tu = L / (c - v \cos\Phi)$$

where tu = upstream pulse transmission time

Since the sound travels faster downstream than upstream, the difference can be expressed as

$t = t_d - t_u$ $t = 2 \, v \, L \, \cos\Phi \, / \, (c^2 - v^2 \cos2\Phi)$ $t = 2 \, v \, L \, \cos\Phi \, / \, c^2$ (4) (since v is very small compared to c)

5.3.2 Doppler Meters

Doppler meters use the frequency shift of an ultrasonic signal when it is reflected by suspended particles or gas bubbles (discontinuities) in motion. The Doppler Effect Ultrasonic Flowmeter uses reflected ultrasonic sound to measure the fluid velocity. By measuring the frequency shift between the ultrasonic frequency source, the receiver, and the fluid carrier, the relative motion are measured. The resulting frequency shift is named the *Doppler Effect*.

www.EngineeringToolBox.com
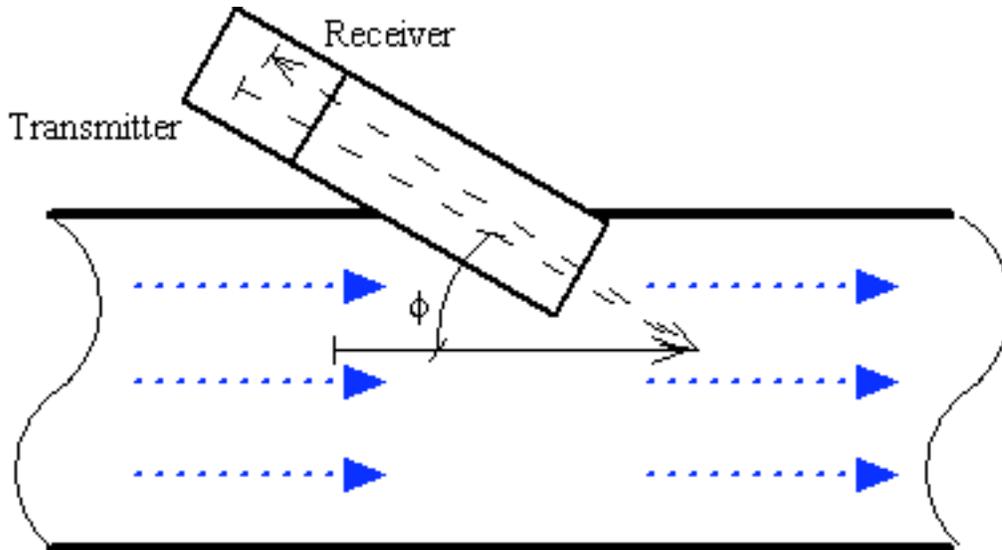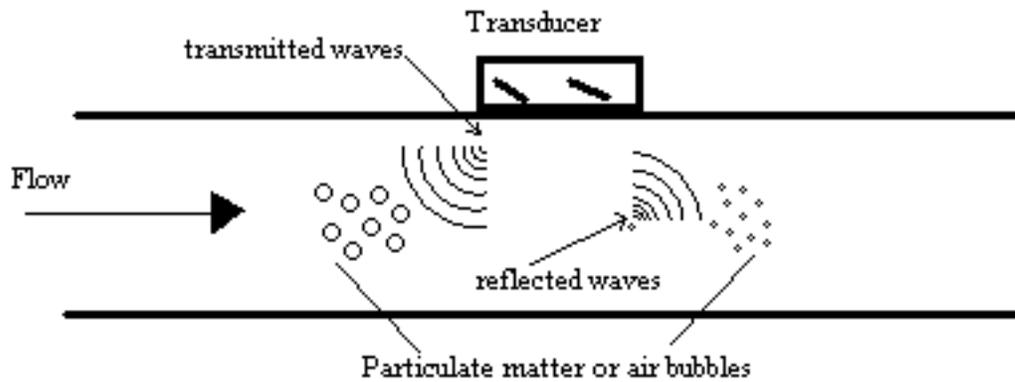
These reflected signals travel at the velocity of light.

$$V_i = \frac{V_s \Delta f}{2 f_{actual} \cos\alpha}$$

where f is the actual frequency and Δf is the change in frequency or frequency shift.

The fluid velocity can be expressed as

v = c (fr - ft) / 2 ft cosΦ

where

fr = received frequency

ft = transmission frequency

v = fluid flow velocity

Φ = the relative angle between the transmitted ultrasonic beam and the fluid flow

c = the velocity of sound in the fluid

This method requires that there are some reflecting particles in the fluid. The method is not suitable for clear liquids.

Advantages with the Doppler Effect Ultrasonic Flowmeter

Doppler meters may be used where other meters don't work. This could be in liquid slurries, aerated liquids or liquids with some small or large amount on suspended solids. The advantages can be summarized to:

- Obstruct less flow

- Can be installed outside the pipes

- The pressure drop is equal to the equivalent length of a straight pipe

- Low flow cut off

- Corrosion resistant

- Relative low power consumption

Limitations with Doppler Effect Ultrasonic Flowmeters

- Doppler flowmeters performance are highly dependent on physical properties of the fluid, such as the sonic conductivity, particle density, and flow profile.

- Non uniformity of particle distribution in the pipe cross section may result in a incorrectly computed mean velocity. The flowmeter accuracy is sensitive to velocity profile variations and to the distribution of acoustic reflectors in the measurement section.

- Unlike other acoustic flowmeters, Doppler meters are affected by changes in the liquid's sonic velocity. As a result, the meter is also sensitive to changes in density and temperature. These problems make Doppler flowmeters unsuitable for highly accurate measurement applications.

5.3.3 Benefits with Ultrasonic Flowmeters As A Whole

- Obstruction less flow

- Pressure drop equal to an equivalent length of straight pipe

- Unaffected by changes in temperature, density or viscosity

- Bi-directional flow capability

- Low flow cutoff

- Corrosion-resistant

- Accuracy about 1% of flow rate

- Relative low power consumption

Both meters are effective in measuring open channels and partially filled pipes but are very sensitive to flow conditions and hence should be calibrated with care. Also, there is no pressure drop since there are no obstructions in the flow path.

5.3.4 Limitations with Ultrasonic Flowmeters as A Whole

- The operating principle for the ultrasonic flowmeter requires reliability high frequency sound transmitted across the pipe. Liquid slurries with excess solids or with entrained gases may block the ultrasonic pulses.

- Ultrasonic flowmeters are not recommended for primary sludge, mixed liquor, aerobically digested sludge, dissolved air flotation thickened sludge and its liquid phase, septic sludge and activated carbon sludge.

- Liquids with entrained gases cannot be measured reliably.

The following link will help to show how both types of the Ultrasonic meter works and how the above equations are derived.

[4]

5.3.5 OTHER TYPES
Magnetic Flow Meter

One magnetic model flow meter positions electric coils around the pipe of the flow to be measured. A pair of electrodes is set up across the pipe wall. The fluid flowing has a minimum value of electrical conductivity, the movement of the fluid through the pipe acts as a conductor moving across the magnetic field. There is an induced change in voltage between the electrodes, which is proportional to the flow velocity.

Induced Electric Field

Voltmeter

Magnetic Field
(perpendicular to flow)
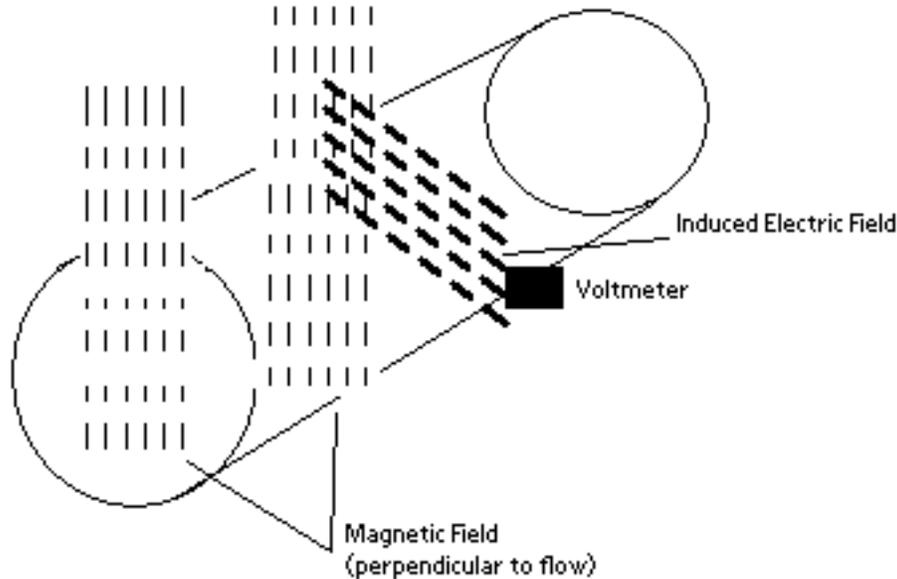
The flow velocity is found by measuring the changes of induced voltage of the conductive fluid passing through a controlled magnetic field at right angles. According to Faraday's Law, the magnitude of the voltage induced is directly proportional to the product of the magnetic flux, distance between probes and the velocity of the medium (fluid).

$$E = -N\frac{d\phi}{dt} = -NB\frac{dA}{dt} = -NBD\frac{dz}{dt} = -NBDv$$

where E is the voltage of induced current, N is the number of turns, B is the external magnetic field, ϕ is the magnetic flux, D is the distance between electrodes and v is the velocity of the fluid.

Some of the advantages are minimum pressure drop because of minimum obstructions in flow path; low maintenance cost because of no moving parts. One of the disadvantages is that it usually requires a fluid with an electrical conductivity of higher than $3\,\mu$S/cm.

Calorimetric Flow Meter

This type of flow meter is suitable for monitoring the flow rates of liquid and gaseous mediums. The calorimetric principle is based on two temperature sensors in close quarters of one another but thermally insulated from one another.

261
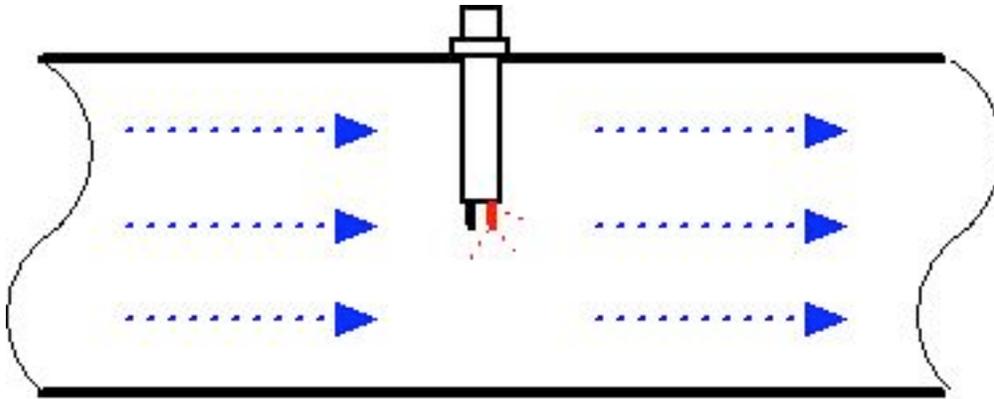
One of the two temperature sensors is constantly heated and the cooling effect of the flowing fluid is used to monitor the flow rate. In a stationary phase fluid condition there is a constant temperature difference between the two temperature sensors. When the fluid flow increases, heat energy is extracted from the heated sensor and the temperature difference between the sensors are reduced. The reduction is proportional to the flow rate of the fluid. The calorimetric flow meter can achieve relatively high accuracy at low flow rates.
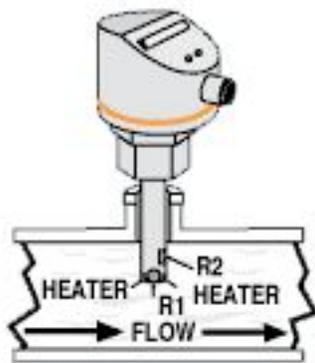
Common applications: air compression, argon compression, nitrogen compression, carbon dioxide compression and flow detection of all fluids (liquids and gases)

Gear Flow Meter

This type of flow meter has oval shaped gears with fitting teeth which control the amount of fluid passing through. The flow rate is calculated by number of times the gears are filled and emptied. These meters have high accuracy and are used for measuring low flow and for high viscosity fluids. It is very easy to install these types of meters because it requires no pipe.

262
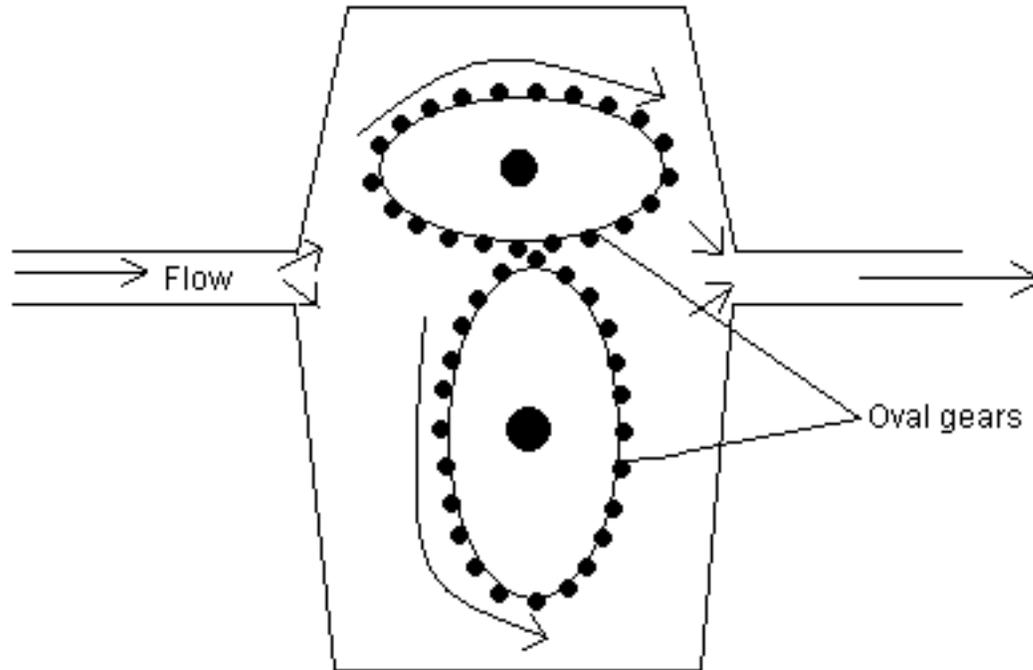
Thermal Flow meters

These flow meters take advantage of the thermal properties of the fluid to measure the flow of the fluid in a pipe. In most thermal flow meters, a measured amount of heat is applied to the heater of the sensor. Portions of this heat is lost to the fluid as it flows. Accordingly, as the fluid flow increases more heat is lost. The amount of heat lost is tracked using a temperature measurement instrument in the sensor. Then an electronic transmitter uses the heat input and temperature measurements to calculate the fluid flow, taking into account the thermal properties of the fluid.

Common applications of thermal flow meters are measuring the flow rate of clean gases like air, nitrogen, hydrogen, helium, ammonia, and argon. Most thermal flow meters are used to measure pure gases that would be used in laboratory experiments or semiconductor production. Mixtures like flue stack flow can also be measured but the mixture compositions must be known to use the appropriate thermal properties of each component in the mixture. The key advantage of this technology is its dependence on the thermal properties which are mostly independent of the gas density.

However caution should be taken if this type of flow meter is used to measure a fluid of unknown or varying composition. Additionally, thermal flow meters should not be used to measure abrasive fluids because they may damage the sensor. Some fluids can coat the sensor causing it to be inoperable and must be regularly cleaned to be useful.

5.3.6 Table of Flow Meters

The following table outlines specifics for most of the flow meters. This may be helpful in answering example 1. Accuracy is given in terms of FSD (full scale deflection) which is the deflection of meter's pointer to the farthest point on the scale. This implies the highest measurement.

Type of Flowmeter	Operating Temperature of Sensor (K)	Pressure Change (Comparative)	Pipe Size (mm)	Accuracy	Measurables	Advantages	Disadvantages
Orifice-Square Edge	<800	High	>40	±1-2% FSD	Clean gases, steam, clean liquids, most dirty liquids and most corrosive liquids	Negligible friction at the boundary between the fluid and the orifice plate	Not suited for multi-phase fluids because clogging becomes a significant problem as time progresses
Orifice-Conic Edge	<800	High	>40	±2% FSD	Clean liquids, viscous liquids, mostly dirty liquids, and most corrosive liquids	Has a gradually expanding cross-sectional area and is good for low velocity, high viscosity flows	Not recommended for high velocity, low viscous flows
Venturi	<800	80-90% of Energy Recovered	>50	±1-2% FSD	Clean gases, most dirty gases, steam, clean liquids, most viscous liquids, most dirty liquids, most corrosive liquids, and most fibrous, abrasive slurries	Pipe design does not need to be straight, relatively low maintenance and operation costs	Insensitive to changes in the velocity profile of the fluid flow, relatively expensive to orifice flow meters
Pitot Tube	<800	Low	>10	±1-2% FSD	Clean gases, most dirty gases, steam, clean liquids, most viscous liquids, most dirty liquids, most corrosive liquids, and most fibrous, abrasive slurries	Detects fluid flow at one point from pressure drop	Clogs occur from fluids containing large particles
Electromagnetic	230-450	Low	Jun-00	±0.2 flow rate to 1.0% FSD	Clean liquids, viscous liquids, dirty liquids, corrosive liquids, fibrous slurries and abrasive slurries	Minimum P-drop because of few obstructions in flow path; low maintenance cost because of no moving parts	Requires a fluid with an electrical conductivity > 3 μS/cm
Turbine	10-530	Medium	6-600	±0.5 flow rate	Clean gases, clean liquids, and most corrosive liquids	Relatively high-tested reliability and low installation cost	May cause low to medium P-drop from turbine wheel motion, may require turbine wheel replacement due to abrasion
Ultrasonic-Transit Time	100-530	Low	>12	±1 flow rate to ±2% FSD	Clean liquids, most viscous liquids, and corrosive liquids	Effective in measuring open channels and partially filled pipes	Very sensitive to flow conditions and thus need to be calibrated with care
Doppler	100-400	Low	>12	±1 flow rate to ±2% FSD	Most viscous liquids, dirty liquids, corrosive liquids, fibrous and abrasive slurries	Effective in measuring open channels and partially filled pipes	Very sensitive to flow conditions and thus need to be calibrated with care
Rotameter	Glass < 400 Metal < 800	Low	<=75	±0.5 flow rate to ±1% FSD	Clean gases, clean liquids, viscous liquids, and most corrosive liquids	Relatively inexpensive and simple to use	Typically fragile, must specifically be installed vertically with fluid flowing upwards
Vortex	80-680	Low	15-400	±0.5 flow rate to ±1 flow rate	Clean gases, most dirty gases, steam, clean liquids, and most corrosive liquids	Insensitive to temperature, pressure, and viscosity	Causes pressure drop due to flow obstruction
Coriolis	Insensitive to Temperature	Low	10-100	±1% FSD	All. Exception – Both clean and dirty gases can be measured only at high pressures	Effectively measures flow rates of high viscous liquids at a low pressure drop, insensitive to temperature, pressure, corrosive liquids, and fluctuating densities	Significant pressure drop and diminished accuracy in the presence of low-flow gases

5.4 Flow Profile Distortion

In the real world, the flow profile is not always symmetrical. Pipe fittings such as elbows, tee-pieces, and reducers can change the flow profile. One example of a fitting that alternates the flow profile is a sharp elbow which causes pure swirls throughout the fluid. Some flow meters are more sensitive to particular types of flow distortion. More complex flow conditions produce better velocity profile but there is a trade off since they are more expensive and give higher pressure drops.

5.5 Turndown Ratio

The turndown ratio is a term used to describe the range of accurate operability of a specific flow meter. This rangeability is critical when selecting flow meters for specific applications. Typically, in a plant setting gas flow may not be constant, but accurate measurement of gas flow is needed ranging from no flow to full flow.

For example if nitrogen gas is being used in a plant conducting multiple batch reactions, sometimes the little nitrogen (100 m3/min) will be needed and other times full nitrogen flow (1000 m3/min) will be needed. For this system, the turndown ratio is 10:1. Accordingly, a flow meter must have a turndown ratio of at least 10:1.

264

For each type of flow meter, the turndown ratio is limited by the theoretical and physical constraints. For example, for an orifice meter the accuracy may be compromised near the limits of the rangeability. Orifice meters create a pressure drop in the measured fluid which is proportional to the velocity squared. If the range of differential pressures becomes too large the overall accuracy of the flow meter at its range limits may be inconsistent.

Turndown ratio can be expressed as:

TR = qmax / qmin (1)

where

TR = Turndown Ratio

qmax = maximum flow

qmin = minimum flow

The table below shows a list of typical turndown ratios for different flow meters.

Type	Turndown Ratio
Orifice Meter	3:1
Turbine Meter	10:1
Rotary Displacement Meter	10:1-80:1
Diaphragm Meter	80:1
Ultrasonic Meter	50:1
Thermal Meter	10:1-100:1

Example - Turndown Ratio for an Orifice Meter

The turndown ratio - TR - for an orifice meter with maximum flow of 12 kg/s and a minimum flow of 3 kg/s can be calculated as:

TR = (12 kg/s) / (3 kg/s) = 4 - normally expressed as turndown ratio of 4:1

This is a typical turndown ratio for a orifice plate. In general a orifice plates has turndown ratio between 3:1 and 5:1.

Turndown Ratio and Measured Signal

www.EngineeringToolBox.com
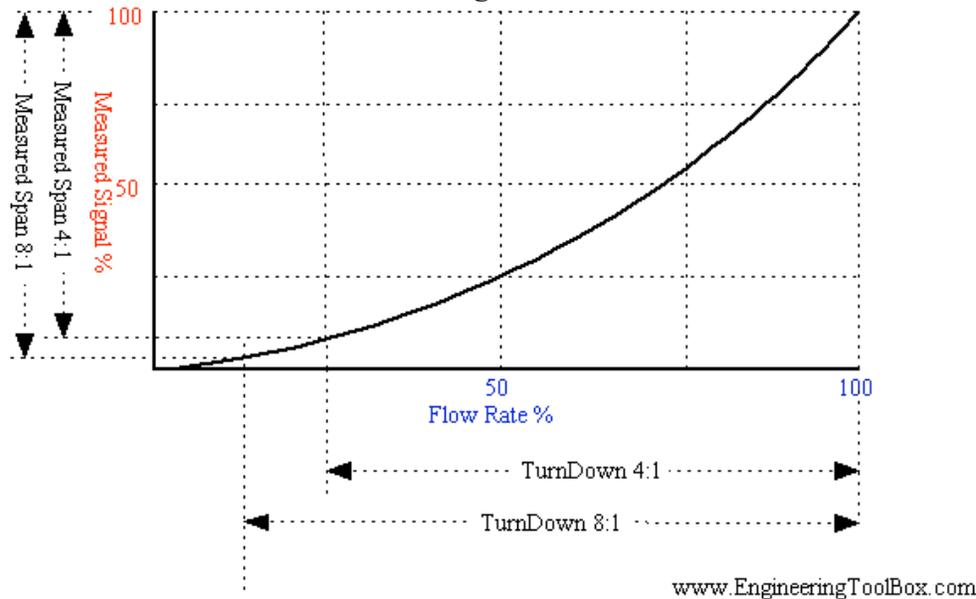
The graph above shows how the turndown ratio effects the measured signal % and flowmeter %. With an increased turndown ratio there is a larger range that the flowmeters can operate within. In a flow meter based on the orifice or venturi principle, the differential pressure upstream and downstream of an obstruction in the flow is used to indicate the flow. According the Bernoulli Equation the differential pressure increases with the square of flow velocity. A large turndown ratio will cramp the measurement signal at low flow rate and this is why venturi and orifice meters are not accurate at low flowrates.

5.6 Example 1

a) Oil is flowing in a medium sized pipe which is insensitive to temperature and pressure. Pressure drop measured is relatively low. Which type of flow meter is recommended?

Coriolis flow meter because it effectively measures flow rates of high viscous liquids at a low pressure drop. Coriolis flow meters are also insensitive to temperature and pressure.

(b) Why is Doppler only recommended for liquids containing suspended solids or gases?

The Doppler sensor transmits sound into the flow stream, which must be reflected back to the sensor to indicate flow velocity. Gas bubbles or suspended solids act as reflectors for the Doppler signal. Doppler flow meters are recommended for liquids containing solids or bubbles with a minimum size of 100 microns and a minimum concentration of 100 ppm.

266

(c) When would you use Vortex flow meters?

Vortex is mainly used to measure steam flow rates. They are also used to measure gases and liquids with low viscosity that contains hardly any solids.

(d) A conductive liquid is flowing in a pipe (1000 mm size) at the temperature of 430K. Which type of flow meter is recommended to measure the flow rate?

Electromagnetic flow meter is used in this case as it measures flow rates of conductive liquids.

5.7 Example 2

Kerosene at 20 deg C flows at 18 m^3 / h in a 5 cm diameter pipe. If a 2 cm diameter thin plate orifice with corner taps is installed, what will the measured drop be in Pa? Why is orifice meter used in this case?

Data given: C_d=0.6030; μ = 192E-3 kg/ (m s) for kerosene; ϱ = 804 kg/m^3

Solution: Orifice meter is used in this case because Kerosene is a clean and non viscous liquid.

Usually, pipe velocity is calculated using the velocity formula to calculate Re and C_d values. The pipe velocity is:

$$v = \frac{Q}{A} = \frac{\left(\frac{18}{3600}\right)}{\left(\frac{\pi}{4}(0.05^2)\right)} = 2.55m/s$$

;

In this case, we already have the C_d. So,using orifice pressure drop formula, calculate

$$Q = \frac{18}{3600} = 0.6030\frac{(2\pi)(0.02)^2}{\sqrt{1 - \frac{(0.04)^2}{(0.02)^2}}}\sqrt{\frac{2\Delta p}{804}} = 0.005m^3/h$$

Final step is to solve for Δp to give 273 kPa.

$$\Delta p = \frac{(0.005)^2 * \left(1 - \frac{(0.04)^2}{(0.02)^2}\right)(804)}{(0.6030)^2(2\pi)^2(0.02)^4(2)} = 273kPa$$

5.8 Sage's Corner

Selection of an Industrial Flow Sensor

5.9 References

- Cole-Palmer Catalog, 2005-2006
- [5]Engineering fundamentals' website
- Perry, Robert H., Perry's Chemical Engineers' Handbook, 7th Edition. McGraw-Hill, 1997.
- Richardson, J.F. and Peacock, D.G., Chemical Engineering, 3rd Edition, Volume 3. Butterworth- Heinemann, 1994.
- [6]Process Measurement and control Products
- [7]Universal Flowmeters Presents Flowmeters.com
- [8]Transactions Volume 4: Flow and Level Measurement
- [9] Wikipedia entry on Coriolis Mass Flowmeters
- [10]Omega Literature: Flow and Level Measurement
- Geankoplis, Christie John, Transport Processes and Separation Process Principles, 4th Edition, Pearson Education, 2003.
- McKetta, John J., Cunningham, William A., Encyclopedia of Chemical Processing and Design: Volume 43 - Process Control: Feedback Simulation to Process Optimization, CRC Press, 1993.

Authors: (September 14,2006) Nicole Blan, Jessica Nunn, Pamela Anne Roxas, Cynthia Sequerah

Stewards: (September 27, 2007) Eric Gauthier, Kevin Lin, Matt Naughton, Rebecca Rutishauser

Anchors: (September 26, 2007) Megan Boekeloo, Stacy Young

- [First round reviews for this page](#)
- [Rebuttal for this page](#)

6.1 Introduction

There are many useful analytical tools, such as photometric analysis, electrometric analysis, chromatography, mass spectrometry, thermal conductivity, and various physical property measurements (density and specific gravity), which can be used to determine the composition of mixtures. A wide array of methods to measure composition are available, so it is important to choose the best method given a set of conditions. The first step is to decide between measuring physical or chemical properties to determine the composition. If you choose to measure a physical property, make sure that it is unique to the desired component of the mixture and will accurately allow you to determine the composition. The goal of this article is to explain the various analytical methods and tools used to determine the composition of a given sample. After reading this, you should be able to determine which method of composition measurement is most appropriate for a given circumstance.

6.1.1 Types of Testing: On-line v. Off-line

On-line analysis is the continuous monitoring of the composition of a sample, which is under the influence of a control system and directed by an actuator which can respond and regulate the operating conditions in real time such that the desired set points are maintained. On-line testing can either be performed in-line or by slip stream testing. In on-line testing, the sensor is attached directly to the line and provides feedback via a transmitter. On the other hand, in slip stream testing, a side stream of the process runs alongside the main line. Such an apparatus closely resembles the set up of a bypass. The slip stream process conditions can be continuously manipulated to make the measurements easier to obtain. Similar to in-line testing, the sensor is directly attached to the slip stream and provides feedback through a transmitter. Advantages of on-line analysis include an immediate and continuous feedback responding to changes in process conditions. The main disadvantage of on-line testing is that it is usually much more complicated and more expensive than off-line testing. Also, off-line testing is more robust and has more varied applications, while on-line testing may not work in every

situation. On-line testing prevents the continued production of undesired product, with an immediate response and correction of the flawed material.

Off-line analysis involves the extraction of a sample from the process or reaction, and its subsequent testing in a machine that may be situated at a location far away from the process line in a lab. In this case, a sample is manually removed which is later sent to the composition analyzer. The results of the analysis are examined and then they are sent to the control system or actuator to make the appropriate adjustments. On-line analysis is comparable to off-line testing, the primary difference being that in on-line testing, the samples are analyzed on a machine that is next to the process line. This greatly reduces time lost in transporting the sample, though it still permits the prospect of introducing contaminants into the sample. In off-line analysis, some of the disadvantages include sample dead-time, which is the time lost during transportation, variability of sample testing locations, and lag time for adjustments to be made to the process. In off-line analysis, the lag time between removing the sample from the system and receiving the results of the test could cause significant losses to the company, since the defective product is produced and the process is not corrected until the results are received. This prolonged defective processing could result in the accumulation of losses worth thousands of dollars, which could have been avoided if an on-line analysis system was set up in place.

While both methods are widely used in industry, there is a push towards more on-line testing. A paper published in 2003 urging drug companies to adopt more in-process analytical testing such as on-line testing, endorsed by the Food and Drug Administration reflects this trend. This has been a part of a broader effort by the FDA to encourage companies to move towards better manufacturing practices. While on-line testing maybe comparatively expensive to install at first, the savings from the process would not only recoup the costs associated with installation, but also prevent significant losses. Many companies have adopted lean manufacturing techniques, and one of its hallmarks is to correct a defect as early in the production process as possible, not allowing defective products to move on to the next processing step. Allowing defective products to perpetuate through the system may even contaminate other non-defective materials in the production process.

6.1.2 Standards and Calibration Curves

Before discussing the different ways to measure the composition of a sample, one needs to understand that composition sensors use standards or calibration curves to measure an unknown composition against a known one(s). The standard can either be an internal standard against which all other measurements are compared, such as a reference cell or a series of standards used to create a calibration curve. With a calibration curve, a series of known standards are measured with the sensor and the signal produced by the sensor is graphed on the y-axis of a plot; the known information about each sample is graphed on the x-axis. From this information, a relationship can be developed between the signal output from the sensor and the known quantity that you are trying to measure. The data measured for the unknown sample can then be compared to the calibration curve

graphically or using the equation written to describe the curve. Using this information, the measurement made by the sensor can be interpreted as a useful composition measurement of weight percent, volume percent, mole percent, etc.

6.2 Photometric Analysis

Photometric analysis is the measurement of the intensity of visible light and other electromagnetic (EM) waves. Through the measurement of these values, the composition of samples or flows can be determined in various different ways. A wide range of photometric devices based on many differing principles are used in the chemical engineering industry for this purpose. A few of the most common instruments are covered in this article.
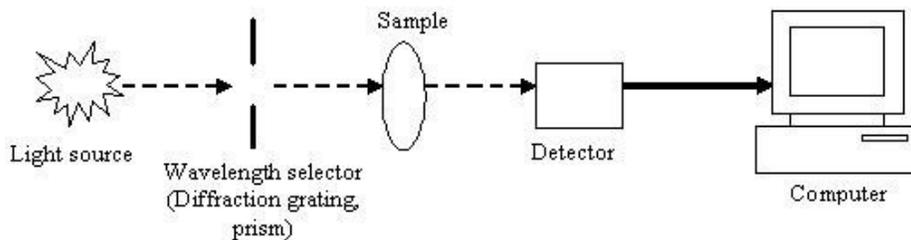
6.2.1 Types of Photometers

A photometer is any type of instrument used to measure characteristics of light. There are two broad categories of photometers: dispersive and non-dispersive.

Dispersive Photometers
In dispersive photometers, the light from the source is dispersed and a narrow spectral band is selectively directed to the sample and detector. A monochromator is usually the tool that performs this function. The dispersing element of the monochromator can be a diffraction grating, prism, and interference filters. Diffraction gratings are the most common; for each position of the grating a narrow band of dispersed light passes through an exit slit. Dispersive devices can scan across a spectrum and make measurements at several wavelengths. This capability allows them to be used for the analysis of multiple components.

Non-dispersive Photometers
Non-dispersive photometers use a narrow-band-pass filter to block out a large amount of undesired radiation. They make measurements at selected discrete wavelengths. The filter passes radiation at the selected reference and measurement wavelengths. The reference wavelength filter selects a band where none of the components present in a process stream absorb radiation; the measurement wavelength filter is selected to match the absorption band of the component being

analyzed. The ratio of the transmitted light at the reference and measurement wavelengths is measured by the photometer. Non-dispersive photometers are usually used to measure a single component in the process stream.

An example of such a device is the Non-Dispersive Infrared (NDIR) analyzer that is frequently used to measure the concentrations of certain gases in a mixed gas flow. The NDIR analyzer uses a reference non-absorbent gas such as O_2, Cl_2 and N_2 and compares that to a sample gas like CO, CO_2, SO_2, CH_4, etc. Two beams of infrared radiation are used: a reference beam for the non-absorbent gas and an analyzing beam for the sample.

6.2.2 Spectrophotometers

Spectrophotometers are widely used as composition sensors. Spectrophotometers are photometers that measure the intensity of light at specific wavelengths. Many different types of spectrophotometers exist, and they are typically categorized based on the ranges of wavelengths they monitor: ultraviolet (280-380 nm), visible light (400-700nm), infrared (0.7-1000μm), or microwave (1-300mm). Other important distinctions are the measurement technique, the method of obtaining the spectrum, and the source of intensity variation.

The emission spectrum recorded by the spectrophotometer typically consists of a series of narrow peaks of varying heights, also known as spectral lines. The height or amplitude of those lines can be related to the concentration or abundance of the specific material through a calibration curve and their position on the spectrum distinguishes the components of the material. While there are numerous methods for obtaining a spectrum, the most common application of a spectrophotometer is for the measurement of light absorption.

Absorption

The degree of absorption of electromagnetic waves exhibited by the components in the sample is a distinctive feature that can be used to reveal the composition of the sample. Absorption is related to the inverse of transmittance by:

$$A = log\left(\frac{1}{T}\right)$$

A = Absorbance (Absorbance unit, AU)

T = Transmittance (as percentage)

Spectrophotometers based on absorption have a radiation source that will emit a specified range of EM waves to be absorbed by the sample by measuring the intensity of light that

passes through the sample. A useful relationship used to quantitatively determine the concentration of an absorbing species in solution is the *Beer-Lambert Law*:

$$A = -log\left(\frac{I}{I_0}\right) = \epsilon * c * L$$

I = Intensity of incident light at a particular wavelength

I_0 = Intensity of transmitted light

epsilon = molar absorptivity or extinction coefficient (1 / *M* * *cm*)

c = concentration of the absorbing species

L = path length through the sample

According to the *Beer-Lambert Law*, the absorbance of a solution is directly proportional to its concentration. The constant *epsilon* is a fundamental molecular property for a given solvent. It can be found from references tables or determined through the construction of a calibration curve. The *Beer-Lambert Law* is useful for characterizing many compounds, but does not hold for all species. For example, very large, complex molecules such as organic dyes often exhibit a 2nd order polynomial relationship between absorbance and concentration.

Two widely-used devices that measure the absorbance of samples are *Ultraviolet* and *Infrared* spectrophotometers.

Ultraviolet (UV) Spectrophotometers

UV spectrophotometry is a useful method for detecting and quantifying substances in solution. Certain compounds in the gaseous or liquid state absorb ultraviolet radiation, resulting in absorption peaks that can be correlated with specific functional groups. These species include aromatic and carbonyl compounds in organic substances and transition metal ions in solution. This property is useful for categorizing and measuring the composition of these types of samples, however this property cannot be used as a specific test for a given compound. The nature of the solvent, the pH of the solution, temperature, high electrolyte concentrations, and the presence of interfering substances can influence the absorption spectra of compounds, as can variations in slit width in the spectrophotometer. Many organic solvents have significant UV absorption and therefore will add peaks to the species dissolved in them. Solvent polarity, high electrolyte concentration, and pH can effect the absorption spectrum of an organic compound by causing protonation or deprotonation of the compound. Temperature can also cause

complex molecules to change in conformation and can also alter the properties of the solvent used. Variations in the slit width will alter the effective bandwidth, distorting wavelength measurements.

The basic components of the overall analyzer are the UV light source, sample holder, optical filters, and detector (typically a phototube). Optical filters (typically a monochromator or diffraction grating) separate the wavelengths of light going through the sample and ensure that only the desired wavelengths reach the detector.

Most UV analyzers are dispersive photometers that also function in the visible light spectrum and are called UV-Vis spectrophotometers. Since these instruments are generally found in laboratories, they are generally used for offline analysis. However, there are inline UV machines and sensors available for real-time inline measurements of process flow streams. UV spectrophotometers used throughout many industries for various applications.

Infrared (IR) Spectrophotometers

Infrared spectroscopy is based on the fact that there are specific frequencies at which molecular bonds rotate or vibrate when exposed to discrete energy levels of radiation. A beam of infrared light is passed through the sample, either one wavelength at a time with a monochromatic beam or using a Fourier transform instrument that measures all wavelengths at once. Inspection of the transmitted light reveals how much energy was absorbed at each wavelength, this data is then plotted (see example below). This plot allows the frequency of the vibrations to be associated with a particular bond type. The characteristic IR absorption spectra obtained for a sample can be used to identify the compound, while the amplitude of absorbance is proportional to the analyte concentration. IR analyzers are used to analyze gaseous or liquid compounds that can absorb IR waves. The system components are generally similar to those of an UV spectrophotometer, with the exception of the IR light source.

IR instruments have been used regularly in a fixed laboratory setting for decades now. Both dispersive and non-dispersive IR techniques exist. More recently, manufacturers of instruments have significantly reduced their size and power requirements. In addition to inline IR sensors for process streams, portable instrumentation is available for field analysis such as the monitoring of stack gases. Another relatively modern technology, the Fourier Transform Infrared (FTIR) Spectrometer, is widely used in industry today because of its high sensitivity and broad range of application. For example, FTIR analysis is used in the quantification of harmful organic peroxides in lubricants and fuels.

Infrared Spectrum of Silicone (polydimethylsiloxane)
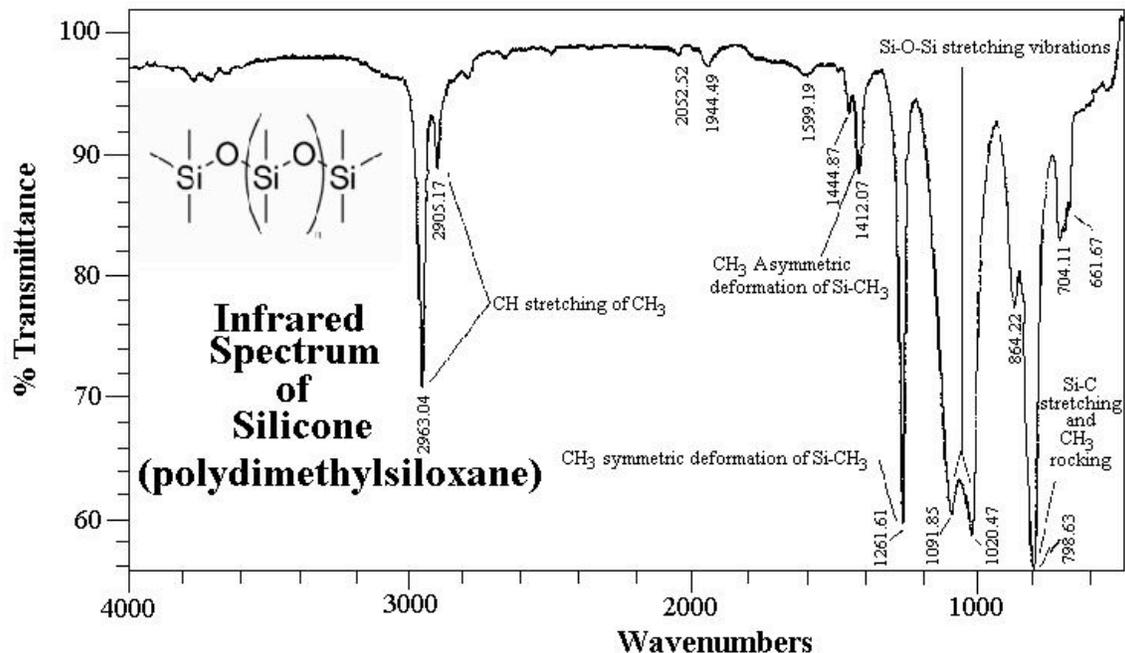
Radiation

Instead of measuring the absorption of light, many spectrophotometers measure the radiation of EM waves. Samples may radiate EM waves in many forms such as heat, luminescence, fluorescence, and phosphorescence. The measurement of such radiation is often employed in biological and pharmaceutical areas because of its high sensitivity and specificity. For example, fluorescence spectrophotometers are used to assay the presence and concentration of target biological compounds. Fluorescent probe molecules are introduced to the sample that will attach selectively to the biological molecule of interest and the device will then detect and measure the amount of fluorescent radiation. In these experiments, the appropriate standards must be measured to properly calibrate the radiation measured with sample concentration. Additionally, contamination of the sample can affect the radiation patterns.

Examples of other techniques utilizing the radiation EM waves include flame emission spectroscopy, particle-induced X-ray emission, and X-ray fluorescence. These methods are strictly used for laboratory analysis.

6.2.3 Photometry Using Visible Light

The manner in which certain materials interact with visible light can be used to analyze the composition of the sample material. Refractometers, turbidimeters, and opacity meters are all analyzers which utilize this principle.

Refractometry

The theory behind refractometry is that visible light bends (or refracts) when it passes two opaque media of differing densities. Snell's Law relates the angle of incidence α to the refractive index η by:

$$\frac{n_1}{n_2} = \frac{\alpha_2}{\alpha_1}$$

α_1, α_2 = angles of incidence

η_1, η_2 = refractive indices of the two media (unitless)

Alternatively, the critical angle may be used to analyze the composition. The critical angle is defined as the angle of incidence that produces a 90° refraction with respect to normal of the interface i.e. $\alpha_2 = 0^o$. Thus, Snell's Law becomes:

$$\frac{n_1}{n_2} = \frac{1}{sin\alpha_c}$$

α_c = critical angle

Please refer to the diagram below.
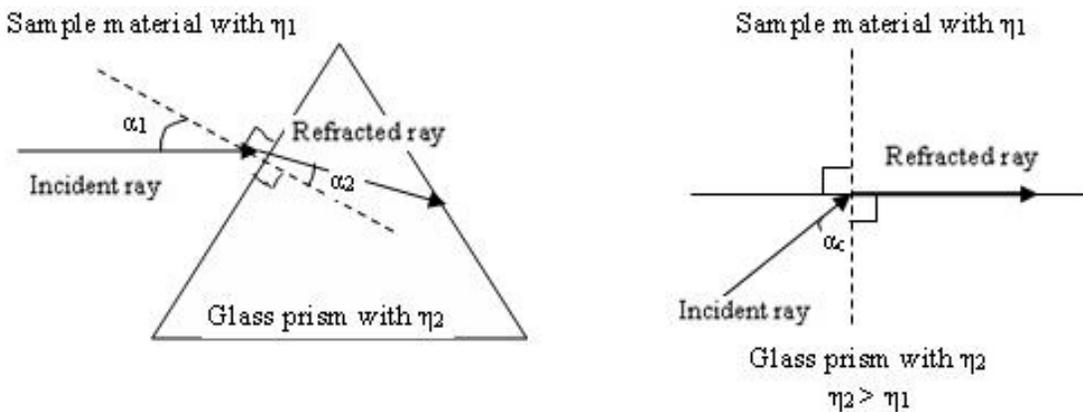
Figure 1: Light refraction at a glass-medium interface

Different materials will have different refractive indices; the experimental refractive index measured can be compared to a list of refractive index values at a data bank.

Continuous process or online refractometers generally measure the changes in the critical angle at a glass-sample interface caused by the variations in composition concentrations. Refractometers are used in numerous industries such as food, chemical, metalworking industry because of its high accuracy and repeatability, low maintenance and easy installation. An example is the use of a refractometer to monitor the amount of juice concentrate in juice.

Turbidimeters

Turbidimeters measure the turbidity, or sample clarity, of fluids. A light beam passes through the analyzer and if the water, or other chemical that is to be tested, is "pure," then the light would travel undisturbed. However if the sample contains suspended solids or particles, the light would interact with the particles causing the particles to absorb the light and reflect it into different directions. An example is smoke in air that gives the air a hazy look.

Most modern turbidimeters are nephelometers. Nephelometers or nephelometric turbidimeters measure the *scattered* light reflected off the suspended particles in the sample. This instrument contains a light source, some sort of container to hold the sample, and photodetectors to sense the scattered light. A tungsten filament lamp is most commonly used as a light source. If the suspended particles, typically measured in nanometers, are smaller than the wavelength than the incident light, the reflected light is scattered equally in all directions. If the particles are larger than the wavelength of the incident light, the pattern of scattering will be focused away from the incident light.

Nephelometers are usually used when solids are present in the sample in small concentrations. When there is high turbidity in a sample, multiple scattering occurs and the turbidimeter may no longer be able to analyze the fluid. Disadvantages of nephelometers include being prone to fouling and incapability of analyzing fluids that have high concentrations of solid particles.

Turbidimeters are important to industry, specifically for chemical plants, oil refineries, and waste water treatment plants. For example, nephelometers are used in detecting the pollutants in water in water quality monitoring. There are specific EPA design requirements for turbidimeters that can be found at:
<http://www.epa.gov/ogwdw/mdbp/pdf/turbidity/chap_11.pdf>

Opacity Monitors

Opacity monitors measure the *attenuation* of light due to *scattering and absorption* by the sample. Attenuation is defined as the decrease in intensity and amplitude of the measured light beam when compared to the emitted light beam. Opacity monitors differ from turbidimeters in that they analyze samples by measuring the percentage of transmission of the light source that passes through the sample and NOT the scattering of the light source. A common application is the measurement of smoke density in chimney

stacks. The density of the pollutant or particle is expressed as percent opacity, percent transmittance, or optical density.

% opacity = 100 x opacity = 100 – % transmittance

6.3 Electrometric Analysis

Electrometric analysis uses the principles of electrochemistry to analyze the composition of different samples.

6.3.1 Conductivity Cells

Conductivity cells measure a liquid's ability to conduct an electric current. Since liquids conduct current due to the presence of ions, the composition can be determined by measuring the concentration of the ions present. Conductivity cells determine concentration from the definition of molar conductivity (lambda). Molar conductivity is a property that can be influenced by temperature. It also varies with concentration if the chemical being measured is a weak electrolyte. It is defined by the following equation:

$$\Lambda = \frac{\kappa}{C}$$

Λ = molar conductivity [Sm^2/kmol]

κ = conductivity [S]

C = concentration [kmol/m^3]

Since it is difficult to directly measure conductivity, conductivity cells measure the resistance (R) of a sample liquid between two electrodes when a small current is applied. Resistance is related to conductivity by the following equation:

$$\kappa = \left(\frac{A}{x}\right)\left(\frac{1}{R}\right)$$

κ = conductivity [S]

A = area of electrode [m^2]

x = distance between electrodes [m]

278

R = resistance [ohm, Ω]

Note that the electrode area, A, and distance between the two electrodes, x, are fixed for a particular sensor and are often expressed as a cell constant (k_c).

The following flow chart shows how composition is determined in terms of concentration:

Measure Resistance	Calculate Conductivity	Calculate Concentration
R	$\kappa = \dfrac{A}{x} \cdot \dfrac{1}{R}$	$C = \dfrac{\kappa}{\Lambda}$
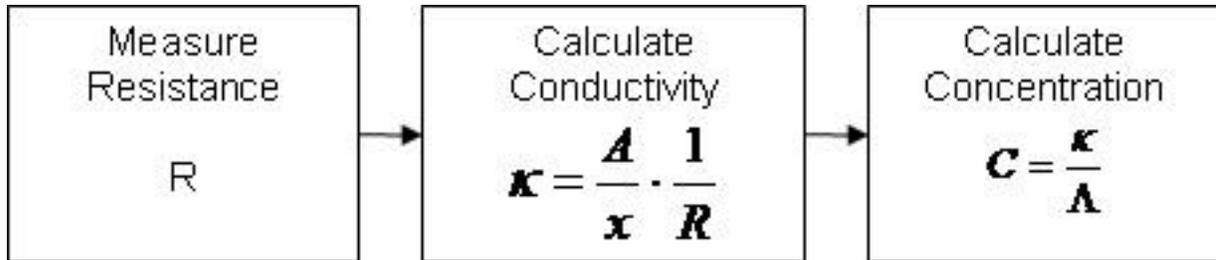
Conductivity cells can be used in flow system and installed directly into the system of pipelines. Conductivity cells can come in three types: two electrode probe, multiple electrode probe, or a non-contact, electrode-less probe (toroidal). The two electrode probe is one of the most common types and typically uses carbon or stainless steel as the electrodes. The multiple electrode probed has either two, four, or six electrodes. This probe has the advantage of reducing errors due to polarization (electrochemical reaction on the electrode surface) and fouling. Finally, the toroidal probe has no electrodes and instead measures the AC current in a closed loop of solution using two embedded toroids (a magnetic core with wire wrapped around it to make an inductor). The electrode-less probe is good for solutions that tend to have problems with polarization or oily fouling. In-line conductivity cells are suitable for temperatures up to 480K and pressures up to 1700 kN/m_2.

Since concentration is found from molar conductivity, which is a property of the fluid that is strongly dependent on temperature, a temperature sensor is often paired with a conductivity cell so temperature effects can be accounted for in concentration calculations.

Conductance cells are best used in applications where a single type of ion is present in a dilute solution. Higher concentrations lead to complications that cannot be described by the simplistic linear relation between concentration and conductivity, since molar conductivity starts to vary with higher concentrations of weak electrolytes.

6.3.2 Ion Selective Electrodes

pH Electrode

The pH electrode is commonly used in both the lab and industrial setting. This electrode is sensitive to the concentration of the H^+ ions. The pH electrode basically measures the potential difference between a reference solution and the sample. Please refer to the pH sensor section for examples and more information on relevant theory.

Other Ion Selective Electrodes

Electrodes can also be specific to other ions such as: Na^+, K^+, Ag^+, ClO_4, BF_4, SO_2, and CO_2. An electrode's ion sensitivity depends mainly on the method of separating ions in the solution. The biggest problem with ion selective electrodes is the problem of not enough selectivity for the desired ion, and subsequent interference by other ions with similar chemical properties to the desired ion. Most electrodes use a membrane to separate a specific ion, though other techniques exist. Selective ion electrodes can be divided into the following three categories based on the mode of separation: solid membrane (glass or crystalline), liquid membrane (classical ion exchanger with a neutral or charged carrier), and membrane in a special electrode (gas sensing or enzyme electrode). Examples of ion selective electrodes used in industry include nitrate sensors for sewage treatment or river quality monitoring, hydrazine and sodium sensors for in boiler water, and choride and fluorine sensors for in water treatment plants.

Oxidation-Reduction Potential (ORP) Sensors

ORP sensors, or also called Redox sensors, can determine concentration of ions that exist in two different oxidation states in solution. For example, this type of electrode can be used to analyze the composition of a solution with Fe^{2+} and Fe^{3+} ions. This type of electrode measures the potential caused by a solution's reduction or oxidation ability.

This type of sensor directly measures electric potential which indicates the reducing or oxidizing strength of a solution. The reducing or oxidizing power of a solution depends on whether the higher or lower oxidation state is preferred by the ion.

When an ion favors the higher oxidation state (or has reducing properties), the ions tend to give up electrons and the electrode becomes more negatively charged.

When an ion favors the lower oxidation state (or has oxidizing properties), the ions tend to gain electrons and the electrode becomes more positively charged.

When purchasing an ORP sensor from a vendor, many models come already packaged with a pH sensor, since these two measurements are often both monitored to keep a process running at the right conditions. For instance, in the pool and spa industry, its important to monitor both the concentration of sanitizers in the water with an ORP sensor, as well as measure the pH of the water.

Polarographic Sensors

Polagraphic sensors work by performing electrolysis with two types of electrodes: one which is polarizable and one which is not. They operate very similarly to voltameters, but with voltameters the potential is changed in a controlled manner and the voltage measured. The oldest system of polarographic measurement used dropping mercury electrodes. Here the principle is that that the surface tension of mercury changes with how much charge is applied between the top and bottom of the electrode, and this surface tension change can be seen in how the weight of the mercury drops varies with time.

More modern types of polarography use two different types of metals which are connected in a solution with electrolytes, and electrons will move towards the more positively charged metal. This movement of electrons induces a current, which eventually stops once the charges are balanced (polarized). If all the charges are balanced within the system, no current will flow. Therefore a small potential is applied to a polarographic sensor so that a current exists for the sensor to measure.

The sample being analyzed needs to have the ability to depolarize the system, or tip the balance of charges. The ion of interest crosses a selective membrane and reacts by an oxidation/reduction reaction with a metal surface (often called the measuring electrode). The reaction between the ion and electrode causes an uneven distribution of charges. Once a system is depolarized, the potential of the system increases since a more positive and a more negative region exists separately. The increase in potential causes current to flow which is the signal measured by the sensor. This creates a graph showing the relationship between applied current versus measured potential. A disadvantage to this type of measurement tool is that this graph requires a calibration curve in order to interpret the data, and so analysis of composition using polarography can't be done in-line. The advantage of this type of method is that it is highly accurate and very reproducible.

Below is a basic schematic of a polarographic sensor.
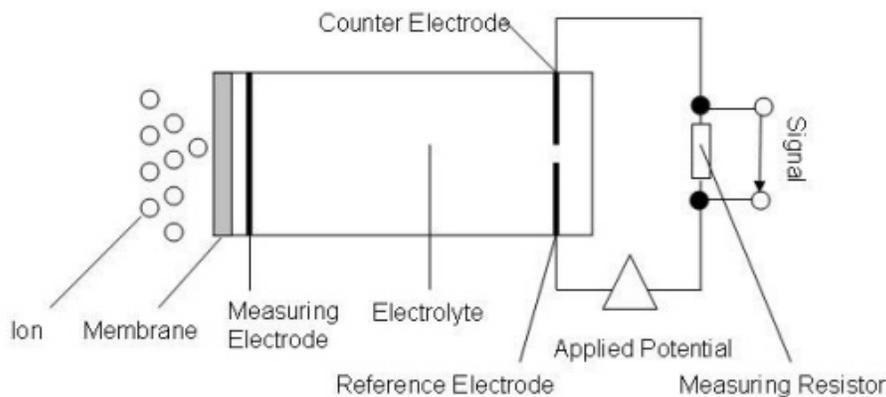
Polagraphic sensors are most commonly used assessing pollutants in air such as Cl_2, HCl, HBr and HF. It an also be used to measure O_2 and other inert gases in flue.

High Temperature Ceramic Sensors

These types of sensors have a heated section of zirconium oxide (ZrO_2), also known as zirconia, that is slightly doped with yttria (Y_2O_3). This stable lattice structure is able to measure oxygen ion conduction at high temperatures. This sensor is exclusively used for measuring O_2 and is often used to measure O_2 composition in flue gases.

6.4 Chromatography

Chromatography is a set of physical methods used to separate and thus analyze complex mixtures. Chromatography consists of a mobile phase, usually a gas or a liquid, and a stationary phase, typically a solid or a liquid. These two phases are combined in a column of the stationary phase based on the type of chromatography one wishes to complete (i.e. Gas-liquid, Gas-solid, Liquid) in order to separate the mixture. The retention time of a component is the time before the component elutes from the column. The retention time and elution order are both based on the relative interaction between each solute and each phase. The stronger the solute's interaction with the mobile phase, relative to the stationary phase, the sooner the solute will elute from the column. Solute-phase interaction is a function of the charges of the molecules, the absorption of the stationary phase, and the relative solubility.

Chromatography by itself does not detect substances, but it is routinely coupled with a detection unit. Each substance is detected as leaves the chromatograph. The retention time for each substance can then be compared to a standard to determine what substances were in the mixture. For unknown substances, the elution order can determine some properties of the substance. For example, a substance that initially elutes after ethanol in a nonpolar column is less polar than ethanol. However, other composition detectors are far better suited for analysis of unknown substances.

Chromatography is the most widely used composition detecting process for on-line analysis of samples within a complex mixture. Due to its precision, chromatography is commonly used to separate or purify samples within the chemical or biochemical industries. Depending on the type of chromatography used, even delicate mixtures like proteins can be analyzed. However, chromatography is primarily useful when there are some expectations for the components in a mixture; chromatography is not best for a completely unknown sample, but it is useful for detecting the composition of a process stream where most components are known and a suitable standard exists.

Gas-Liquid Chromatography (GLC)

GLC is one of the more common types of chromatography. It is commonly referred to as GC (gas chromatography). In a typical run, the liquid mixture to be analyzed is added to the system and vaporized before entering the column. A carrier gas, often Helium, carries the vaporized sample through a column. The column can either be packed with stationary phase, or the column can have a very small diameter in which case it is lined with stationary phase (capillary tube). The gas mobile phase runs through the stationary phase within the column at various rates determined by relative volatility and affinity to the stationary phase. The longer a sample molecule spends in the gas phase, the faster it will elute from the column. For this reason, temperature and the chemical identity of the coating on the stationary phase are two important variables which may need to be manipulated to receive clear separation of the components in the mixture. In general, lower temperature will result in better separation but longer elution times and sample spreading due to flow effects. Efficient use of this apparatus requires a balance between these competing effects.

GC is relatively simple for a technician to use; a calibrated gas chromatograph can run several samples without a large amount of sample preparation. Gas chromatography works well on substances that vaporize below 300°C and are free of ions; however, it does not work well on substances that decompose below 300°C.
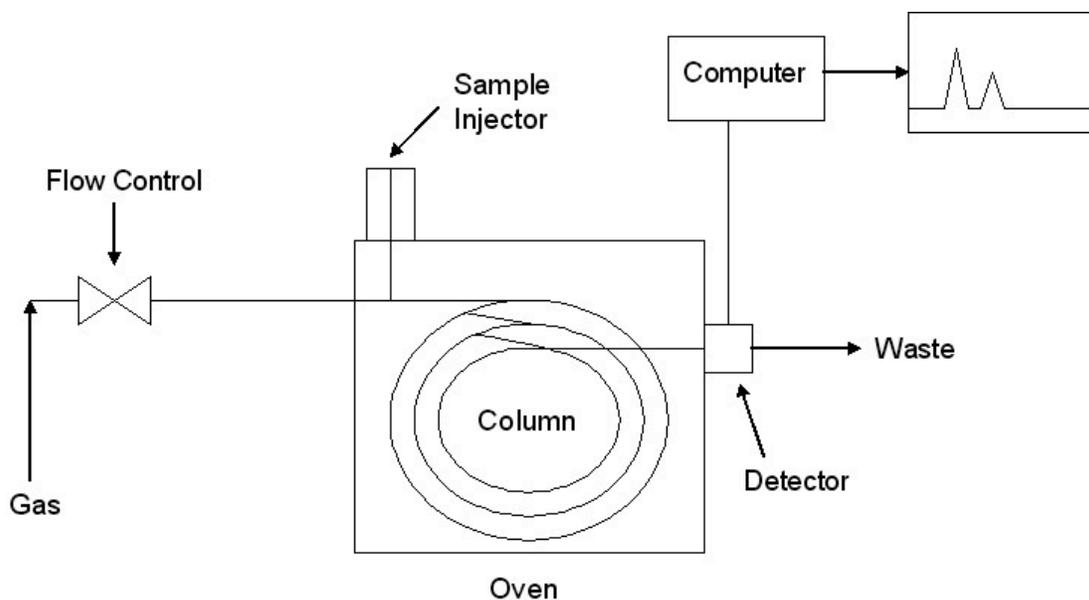
Figure: Diagram of Gas Chromatography

Gas-solid Chromatography (GSC)

GSC uses a gas mobile phase which runs through an absorbent solid (e.g. silica gel or alumina) stationary phase within the glass column as determined by absorption. This method operates along the same principles as gas-liquid chromatography. It is most commonly used with gases (as opposed to a vaporized substance) and with low molecular weight hydrocarbons.

Liquid-Chromatography (LC)

LC analyzes low volatility liquids within a solid stationary phase. Separation occurs from adsorption/desorption due to the solutes having different degrees of attraction to the stationary phase, typically due to the stationary phase having different polarity than the mobile phase. It is common to use a mixture of solvents to accurately separate mixtures. Liquid chromatography is used in several industries, including the pharmaceutical and petrochemical industries.
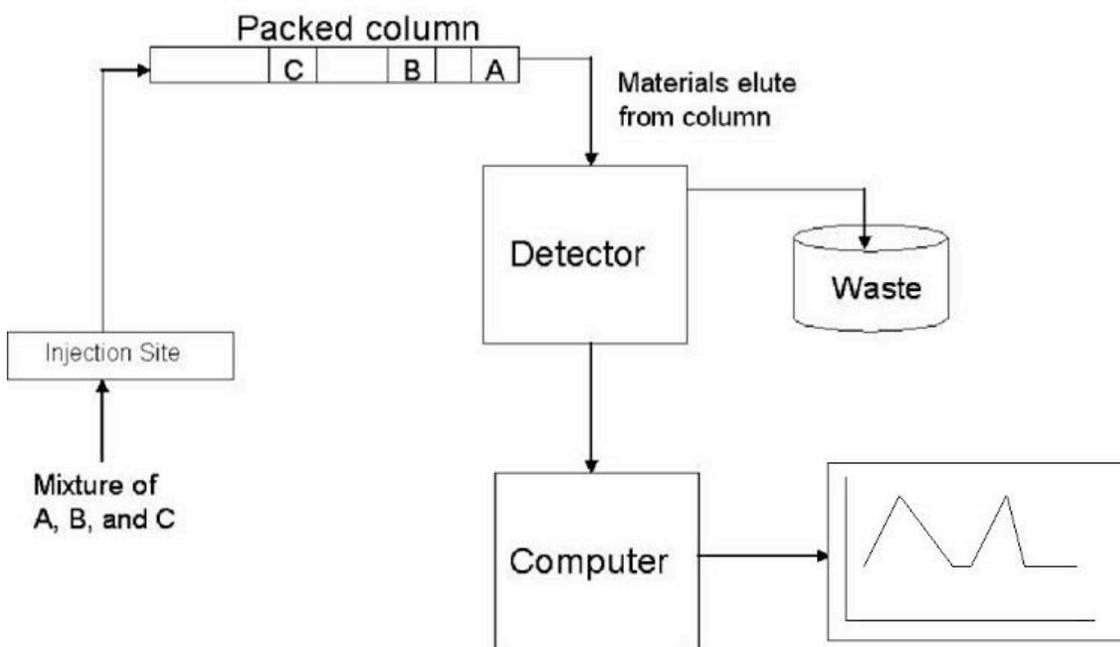
Figure: Example of Liquid Chromatography

6.5 Mass Spectrometry

Mass spectrometry is among the most precise compositional sensing tools, and, accordingly, one of the most expensive. The basic idea is illustrated below:

284

Ion Accelerator

Detector

Flight Tube

Lightest Molecules Travel Fastest
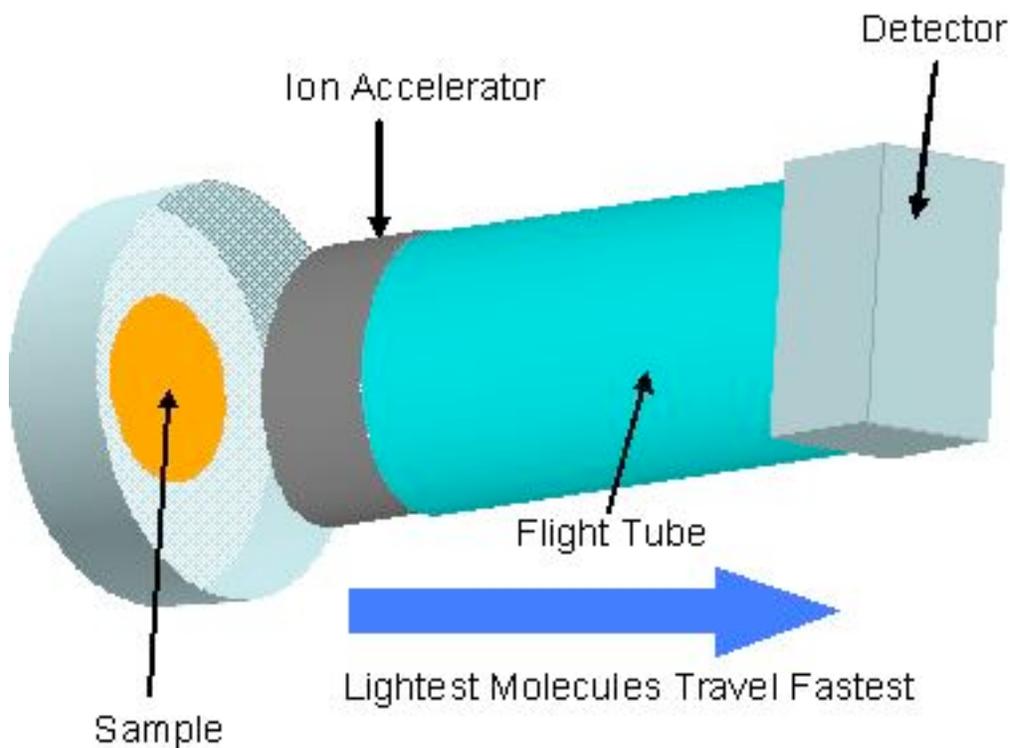
Sample

A sample is ionized—this may be performed by a number of techniques—the ions are subjected to an electrical force and accelerated through a tube. Because the electrical force applied to each molecule is the same, a molecules acceleration through the tube depends on its mass (F=ma). The acceleration is also dependent on any charge on the molecule due to magnetic attractive or repulsive forces. At the end of the tube, a detector calculates the time of flight for each of the molecules.

Mass spectrometry is essentially a tool to create a spectrum of distinct mass/charge ratios. It is very often used after chromatography separation techniques to serve as a molecule identification technique.

6.5.1 MS Components

There are 3 fundamental components for mass spectrometry—an ion source, a mass analyzer, and a detector.

The ion source is the component responsible for ionizing the sample. There are many different methods including chemical, electrical, or laser ionization. Selection of the proper ion source depends on the characteristics of the sample (phase, biologically active, etc.).

The mass analyzer is the technique and tool used to differentiate the ions in a spectrum based on the mass/charge ratio. The technique described earlier is the Time-of-Flight

technique. Another common method is a sector field mass analyzer where the molecules are accelerated through a tube and a magnetic force is applied perpendicular to the direction of flight of the molecules. The molecules will deflect at different magnitudes depending on their size.

Magnetic Field—Perpendicular to ion flow

Increasing Mass/Charge Ratio
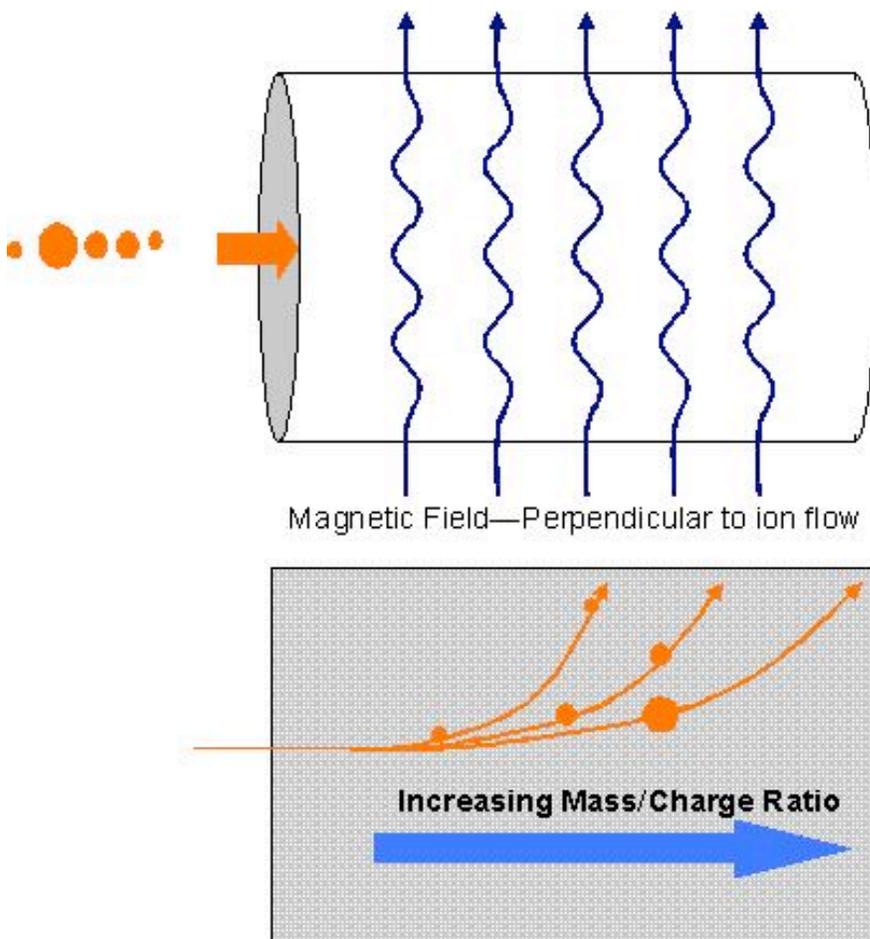
The detector generally amplifies changes in an electrical current that occur when molecules hit the detector.

6.6 Thermal Conductivity for Gases

Changes in thermal conductivity may be used to detect variations in a mixture of gases. When a heat source is present, gases and vapor will conduct heat. The change in heat between the source and gases will stabilize to a common temperature. The temperature is mainly dependent on the thermal conductivity and therefore the composition of the gas.

Typical equipment for thermal conductivity analyzers includes: a reference cell, a sample cell, a combined heat source (wire filaments or thermistors), and detector. The sample cell and reference cell are usually placed in a holder where the detector may be mounted. The reference is an identical cell of the sample cell, through which a known gas will flow. The reference-detector resistance will be constant and the sample-detector

286

resistance will vary depending on the composition. Resistance is a function of temperature and the output from the detector bridge will be a function of sample composition.
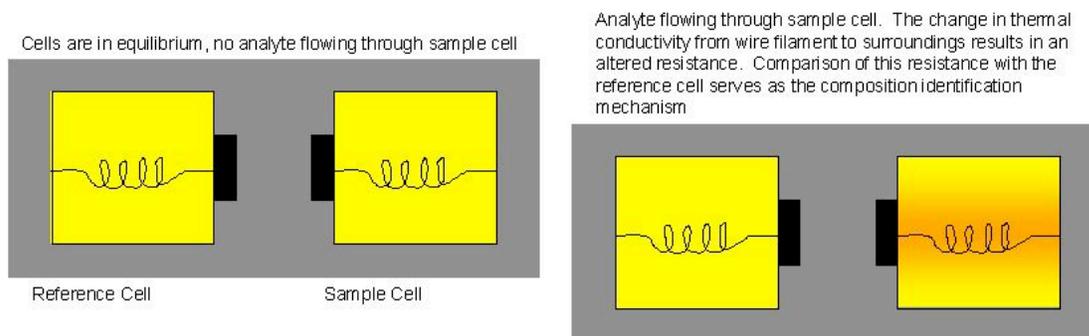
Thermal conductivity, like mass spectrometry, is often used in combination with gas chromatography

6.7 Physical Property Measurements

Physical properties can be used to indirectly measure the composition, but this technique is not as versatile as those described above, since it only applies to a limited number of circumstances. The advantage to using physical properties to indirectly measure composition is that these types of measures are often cheaper than the types of sensors described above.

6.7.1 Density and Specific Gravity

Density and specific gravity measurements require a binary mixture with at least one fluid phase to make accurate measurements. Given the temperature and pressure, the density of a gas can be found using the ideal gas law or some other relationship, and since density is a function of composition, the composition can be measured indirectly. See Thermodynamicsfor more information. Specific gravity is simply the ratio of density of water over the density of the non-compressible component at the same physical conditions, which means that it is simply a dimensionless measurement of density.

Four different measuring devices for finding density are described below.

Liquid Column
In a liquid column, the pressure measurements in a column are used to determine the density. The column may be open to the atmosphere, where the gauge pressure at the bottom of the fixed-height column is measured. Or, the column may be closed, in which case the differential pressure measurement is made between the bottom of the fixed-height column and at the top of the column immediately below the liquid surface.

Displacement – Hydrometer
A hydrometer is usually made of glass and consists of a cylindrical stem and a bulb weighted with mercury or lead shot to make it float upright. The liquid to be tested is poured into a tall jar, and the hydrometer is gently lowered into the liquid until it floats freely. The point at which the surface of the liquid touches the stem of the hydrometer is noted. Hydrometers usually contain a paper scale inside the stem, so that the specific gravity can be read directly.

The operation of the hydrometer is based on the Archimedes principle that a solid suspended in a liquid will be buoyed up by a force equal to the weight of the liquid displaced. Thus, the lower the density of the substance, the lower the hydrometer will sink.
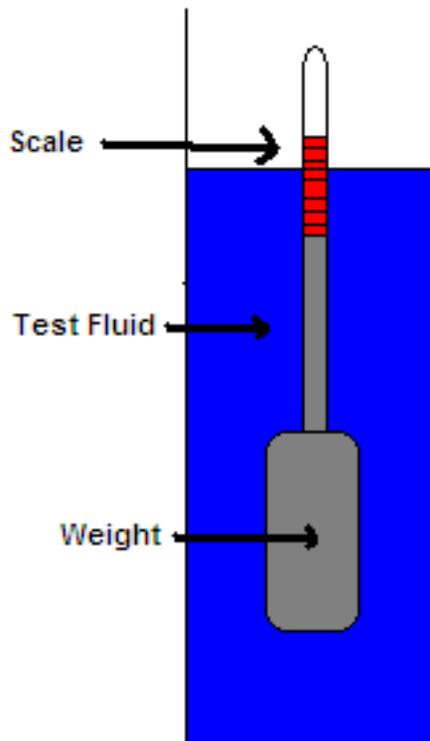
Direct Mass Measurement
With this technique, fluid is continuously passing through a U-shaped tube, where the tube is oscillating at its natural frequency. An electrochemical device strikes the tube periodically at the curved portion of the tube. The change in frequency between the electrochemical strike and the natural frequency is proportional to the changes in density which correlates to the composition of the substance.

Radiation-Density Gauges
Gamma radiation inside a pipe/vessel is used to determine the density of the liquid.
Below is a diagram of the setup.
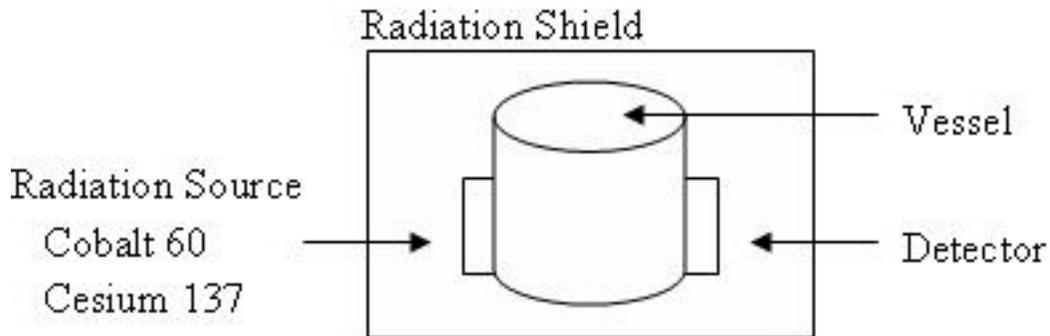
If the path-length for the radiation source is under 610 mm then cesium 137 is used.
Above 610mm, cobalt 60 is used. The detector is usually an ionization gauge. The
density is a function of the absorption of the radiation being detected. Decay of the vessel
must be taken into consideration when calibrating

6.7.2 Viscosity

Please refer to the Viscosity Sensor section of the process control website for in-depth
information.

6.8 Comprehensive Charts

- **For simple, easy to read charts summarizing the advantages and
  disadvantages of each type of sensor, please follow these links for a PDF
  version of each chart.**

Media:Photometric sensors.pdf

Media:Electroanalytical Measuring Techniques.pdf

Media:Other Types of Measuring Techniques.pdf

- **For complete summary charts including more sensors not discussed in
  the article, please follow the links here for a PDF version of these
  summary charts.**

Media:CompositionSensorsGeneral.pdf

Media:CompositionSensorsPhysical.pdf

Media:CompositionSensorsChemical.pdf

Media:CompositionSensorsElectrical.pdf

6.9 Applied Example: Mars Pathfinder Spectrometer

The Mars Pathfinder uses an Alpha Proton X-Ray spectrometer to identify the chemicals that comprise the Martian surface. The incredible tool shoots X-Rays at the surface of the planet and uses a backscatter detector to collect the alpha particles, protons, and X-rays characteristic of specific elements. These atom specific scattering events allow NASA scientists to determine the surface composition.
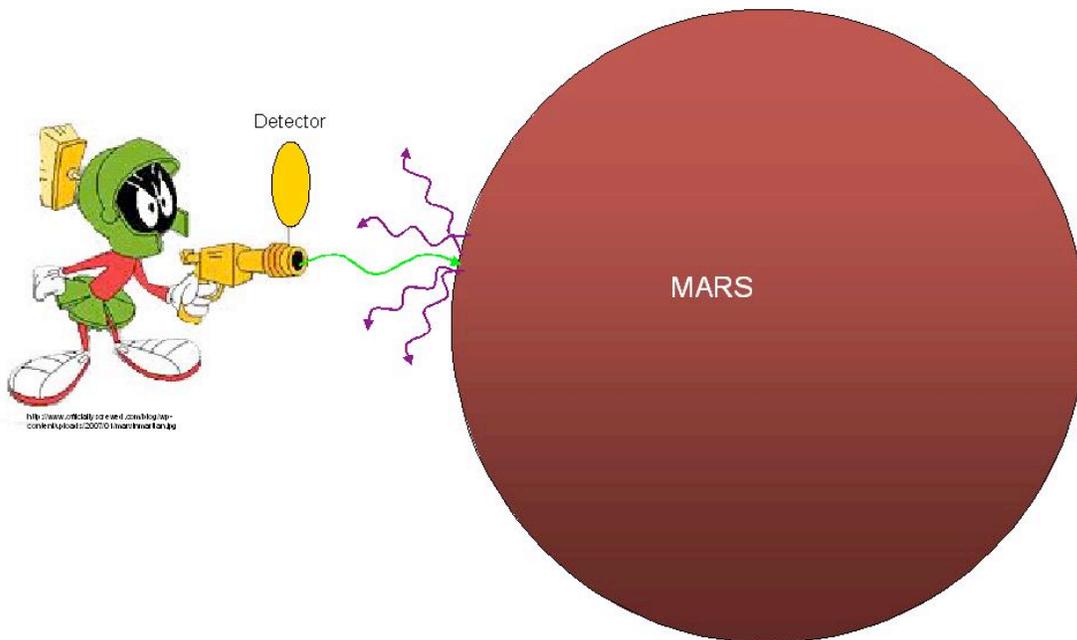
6.10 Worked out Example 1

These examples are fictional but they still require the same thought process.

Your supervisor put you in charge of monitoring the emissions from an old smoke stack that the plant just re-opened to keep up with product demand. It is important to measure the amount of smoke and oxygen leaving the stack. Choose the proper types of sensors to install and keep in mind that the contents of the stack are exiting at very high temperatures.

290

To monitor the smoke density one could install an opacity monitor that measures the absoprtion and scattering of light. Please refer to *Sensor Assmebly* on this website Opacity Monitor Assembly for a basic schematic diagram.

To monitor the oxygen concentration one could install either a polarographic sensor or a high temperature ceramic sensor because both types of sensors are specific to O_2 particle. The high temperature ceramic sensor is more suitable for this application since the emissions are exiting at high temperatures.

6.11 Worked out Example 2

Values of constants are ficitional but are being used to illustrate the thought process.

You are in charge of calibrating a new conductivity cell acquired by the company. The conductivity cell will be used to measure the concentration of NaCl in the beginning reagents. NaCl has a molar conductivity of 12.6 Sm^2 / $kmol$. The cell dimensions are as follows: area of the cells (A) is 100 m^2 and the distance between the cells (x) is 0.25 m. Calculate the concentration needed to produce a resistance of 10 ohms in the sensor.

Referring to the Conductivity Cell section, the equation can be arranged to find concentraton:

concentration = conductivity / molar conductivity

Unknown = conductivity

conductivity = $\left(\dfrac{A}{x}\right)\left(\dfrac{1}{R}\right)$

conductivity = $\left(\dfrac{100}{0.25}\right)\left(\dfrac{1}{10}\right)$

conductivity = 40

Plug back into initial equation

concentration = conductivity / molar conductivity

concentration = 40/12.6

concentration = 3.17 kmol

6.12 Sage's Corner

Some examples of conductivity sensors

```
http://video.google.com/googleplayer.swf?docId=-
7242947671477594432
```

For slides of this presentation:Media:conductivity.ppt

6.13 References

- Richardson, JF. Coulson & Richardson's Chemical Engineering Vol. 3, Oxford: Butterworth-Heinemann. ISBN 0080410030

- Perry, GH; Green, DW. Perry's Chemical Engineering Handbook 7th Edition, New York: Mcgraw-Hill ISBN 0070498415

- So, Peter; Dong, CY. *Fluorescence Spectrophotometry.* Encyclopedia of Life Sciences, London: McMillan

- OMEGA Engineering. *Turbidity Measurement.*

- Analytical Specialities. *Electrochemical Sensors*

- Shimadzu. *Quantification of Organic Peroxides Using FTIR and ATR.*

- Merriam-Webster's Dictionary.

- Wikipedia, Mass Spectrometry

- Berg, Jeremy M.; Tymoczko, John L.; Stryer, Lubert. *Biochemistry*, Sixth Edition. W.H. Freeman and Company.

- Wilks, Paul. *Process Monitoring: In-line, At-line, or Slip-stream?*.

- Thayer, Ann M. *KNOW THY PROCESS:Regulatory push for process analytics sets new goals for pharmaceutical manufacturing.*

- Randall, Willis C. *PAT Pending.*

- ABB Instruments. *Types of Industrial Composition Sensors available for purchase.*

- Topac Inc. *Conductivity Cells.*

- Wroblewski, Wojciech. *Ion Selective Electrodes.*

- Chemical Engineering Department at Case Western Reserve. *Electrochemistry Encyclopedia.*

- Basic Turbidimeter Design and Concepts.
  http://www.epa.gov/ogwdw/mdbp/pdf/turbidity/chap_11.pdf

Title: pH and Viscosity Sensors

Authors: Navin Raj Bora, Dallas Burkholder, Nina Mohan, Sarah Tschirhart

Date Presented: 14 September 2006 Date Revised:

- [First round reviews for this page](#)
- [Rebuttal for this page](#)

7.1 Introduction to pH

pH is a measure of concentration of H^+ ions ($[H^+]$) in solution, and can be defined as follows:

$$pH = -log([H^+])$$

Typically, pH values range between 0(strongly acidic) and 14 (strongly basic, or alkaline). However, that is not to say that 0 and 14 are the limits of pH values. A solution that has a H^+ concentration of 10 M, such as very strong HCl, would have a pH of -1. Pure water, which has a pH of 7, is said to be neutral because its H^+ concentration is equal to its OH^- concentration. An acid solution is said to have a greater H^+ concentration than its OH^- concentration. Conversely, a base, or alkaline solution, is said to have a greater OH^- concentration than its H^+ concentration. See for reference:
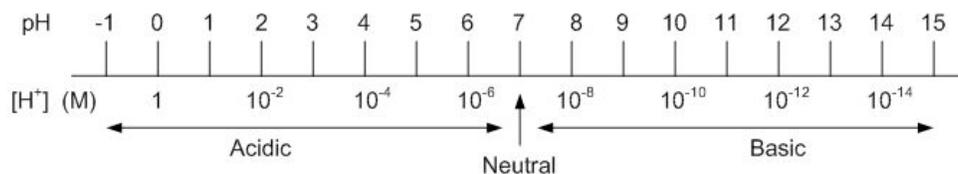
For weak acids and bases, the above equation is appropriate to use when calculating pH. However, many solutions are not weak acids and bases, which means that the other ions in solution play a large role in acidity and alkalinity, and thus a more accurate method is needed to calculate the pH. pH sensors are used to determine the acidity or alkalinity of a particular process solution. When dealing with a strong solutions, the sensors measure the effective concentration, or activity, of H^+ ions in solution, and do not measure the actual concentration. Hence, for the purposes of pH sensors, pH is defined as:

$$pH = -log(a_{H+})$$

294

$$a_{H^+} = (\gamma_{H^+})[H^+]$$
$$\gamma_{H^+} = \text{activity coefficient}$$

The activity coefficient accounts for the interactions between all the ions present. These interactions may prevent some of the H^+ ions from contributing to the acidity of the solution, and thus decrease the effective concentration of H^+ ions. This effective concentration is the concentration used when calculating pH.

7.2 Why is pH Relevant for Chemical Engineering?

pH sensors, while being relatively simple pieces of equipment, have numerous uses. Knowing the pH of a solution is valuable to an analytical chemist attempting to determine the contents of an unknown solution, as well as a farmer trying to determine the appropriate applications to his fields and many people in between. In industry pH sensors can be used as a simple way to monitor reactions involving acids or bases as either a reactant or product of a chemical process. They can also be used to monitor the conditions of a reactor to ensure optimal conditions. This is especially important in biological systems where very minor changes in pH can result in significantly lowered production levels or even a complete stoppage of production due to the death of the organisms on which the reaction is dependent.

An example of pH measurement's importance in industry is seen in food processing during the manufacture of fruit jelly. The gel formation brought about by pectin occurs over a small pH range, which is further complicated by the concentration of sugar. Too high of a pH value results in an unacceptably runny liquid. Too low of a pH causes the mixture to stiffen prematurely, resulting in a nearly-solid product. Continuous pH adjustment using edible acids such as citric acid yields the optimal gel consistency.

It is also important to keep in mind that application requirements should be carefully considered a pH sensor is chosen. In the case above, it is important that the pH sensor be made of a material that will not interfere with the quality of the product. For instance, the sensor should not be made of glass in order to avoid broken glass entering the product. Accurate pH measurement and the precise control that it allows for helps to optimize a process and can result in increased product quality and consistency. Continuous pH monitoring also controls chemical usage (such as the edible acids above), resulting in minimized system maintenance (i.e. less cleaning of hardened gel from containers, less corrosion of containers, etc.).

7.3 Typical pH Sensor Construction

Because it is difficult to directly measure the H+ concentration of a solution, pH is typically measured by comparing the potential of a solution with a known $[H^+]$ (this is known as a reference solution) with the potential of a solution with an unknown $[H^+]$. The potentials of these solutions are measured using a reference half-cell and a sensing half-cell, respectively.

7.3.1 Reference Half-Cell

The reference half-cell generally consists of a chamber with a conductor submerged in a reference electrolyte. The conductor, or external reference electrode wire, is typically silver coated with silver chloride (Ag/AgCl(s)) or mercury coated with mercurous chloride (Hg/Hg2Cl2(s)). The reference electrolyte is a standard solution such as KCl. The last component of the reference half-cell is a porous plug, which serves as a liquid-liquid interface between the standard solution and the process solution being analyzed. This plug allows standard solution to travel from the chamber out into the process solution, but does not allow process solution into the chamber. The purpose of this interface is to establish the electrical connection, which provides the reference potential.

7.3.2 Sensing Half-Cell

The sensing half-cell is of similar construction to the reference half-cell. A glass chamber contains an electrode (Ag/AgCl(s)), known as the internal reference electrode, which is submerged in a standard solution of constant pH. However, instead of a porous plug acting as the liquid-liquid interface between the standard and process solutions, there exists a glass membrane. This membrane is coated on the inside and outside with a hydrated gel. When the sensing half-cell is placed in solution, metal ions (Na+) in the gel on the outside of the membrane diffuse out into the sample being analyzed and are replaced by H+ ions from the sample. The change in energy associated with this substitution creates the change in electric potential that is detected by the electrode. The combination of the hydrated gel and the glass membrane are what make the sensor specific only to H+ ions. The solution of constant pH ensures that the surface of the glass membrane and the electrode are at the same pH before the cell is exposed to the process solution being analyzed.

The electrode wires from the reference and sensing half-cells are each connected to the voltage measuring device.

The diagram below shows the setup of a typical pH measurement system:

296
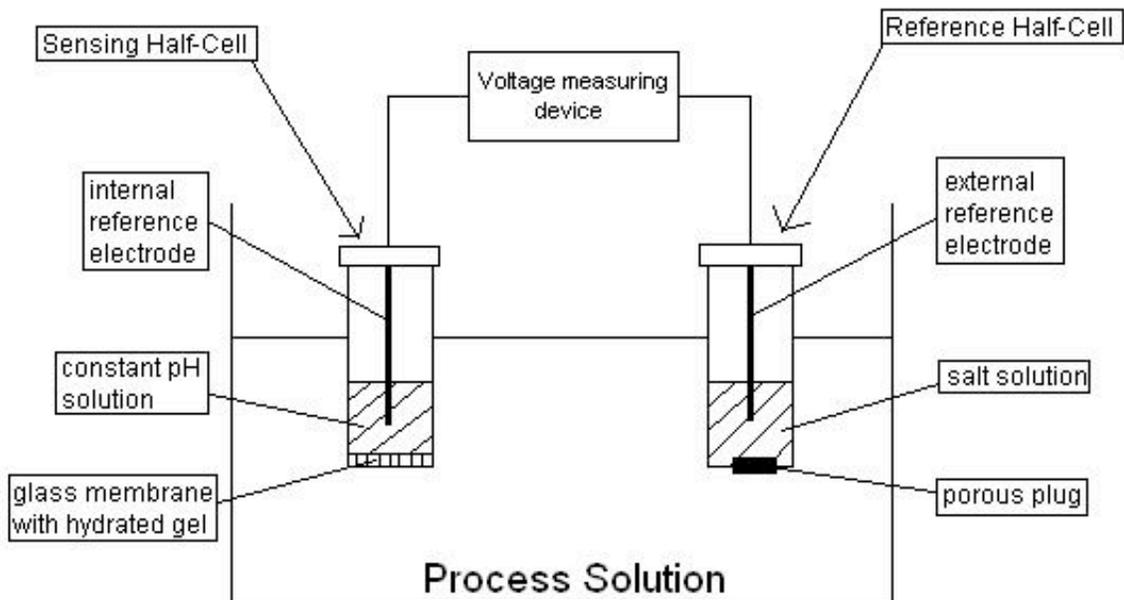
While this figure shows the essential components of a pH sensor, it is not an accurate depiction of what pH sensors actually look like. The reference and sensing cells in most pH sensors are generally encased together in a sturdy housing (usually made of some sort of plastic). This does not affect the operation of the sensor, but allows for ease of use, as there is only one major piece of equipment that the user needs to carry around.

pH sensors can be used both online and offline. If the sensor is being used to carefully control the pH of a system or to provide an accurate record of the pH of a solution it is often desirable to have an online pH sensor. In order to have an online pH sensor determine accurate readings it must be in a position to take a representative sample. This is most often done by diverting a side stream from the main process in which the pH sensor can be inserted. Often times continuous pH measurements are not necessary, but rather infrequent measurements to make sure that the system is running correctly are of interest. In these cases it is much more efficient to use an offline pH sensor.

7.4 How do pH sensors calculate pH?

pH is calculated using the potential drop between the reference and sensing electrodes. The potential across the overall pH measurement system can be expressed as:

$$E = E_{internal\,reference} + E_{membrane} + E_{porous\,plug} - E_{external\,reference}$$

The potential of the porous junction is usually negligible, and the potentials of the internal and external references can be lumped into a constant value, E_0. Thus, we get:

$$E = E_0 + E_{membrane}$$

Using the Nernst equation, the potential of the membrane in a half-cell reaction can be expressed as:

$$E = E_0 + \frac{RT}{nF} \ln(a)$$

$R =$ Universal Gas Constant
$F = 9.6490 * 10^7 coulombs/kmol$ (Faraday's Constant)
$n =$ net number of negative charges transferred in the reaction
$a =$ activity of the ions involved in the reaction

This equation is the standard equation used in electrochemistry to find the potential with respect to a reference electrode. It can be viewed as a calculation of the change in free energy at an electrode, due to changing ion concentrations at that electrode. Assuming that the electrode is completely selective to H^+ ions (a = 1) and that there is only one positive ion carrying one elementary charge in each half-cell reaction (n = 1) the pH of the unknown solution can be expressed as:

$$pH = \frac{F(E_0 - E)}{2.303RT}$$

$T =$ temperature of solution being analyzed

298

7.5 Temperature Compensation

As one can see from the Nernst equation, the pH as determined using the potential difference between the reference and sensing half-cells varies linearly with changes in temperature. That is, the voltage drop/pH unit across the cells increases with increasing temperature. Thus, pH measuring system needs to have some way to compensate for temperature changes in the process solution. Many pH control kits come equipped with automatic temperature compensation (ATC), which continuously measures and incorporates temperature changes into their pH calculations. Other older pH control kits require manual temperature compensation, meaning that the operator or user must manually input the process solution temperature into the control system.

7.6 Alternative Methods for Determining pH

While glass membrane pH sensors are an effective method of measuring pH they are not always the best choice. Certain properties of the system of interest make using a glass membrane pH sensor impractical. Glass membrane pH sensors can become clogged by viscous solutions or solutions containing suspended solids. Harsh conditions can also scratch or break the glass membrane. There are also cases where a glass membrane pH sensor would work but using one would be an unnecessary cost. In these situations there are several alternatives that can be used to determine pH. One method is using pH paper. This paper changes color when submerged in a solution. The color of the paper can be compared to a standard scale included with the paper to determine an approximate pH. pH paper is an excellent choice when all that is needed is a "ballpark" pH value. Another method of determining pH is the use of indicators. These are substances that change color depending on the pH of the solution. They can be very useful if the primary interest is determining if the pH of the system is above or below a particular point. Numerous indicators exist that change colors at different pH values.

7.7 Other Ion-Specific Sensors

The same concepts that allow pH to be measured using a reference and measuring electrode can also be applied to measure the concentration of other ions. The differences between the pH sensors and ion specific sensors are the composition of the measuring electrode the membrane used in the measuring electrode. For some ions of interest glass membranes can still be used, however for other ions solid state membranes or liquid ion exchange may be necessary. The choice of membrane depends on both the size and the charge of the ion of interest.

While ion specific electrodes can be fairly accurate a number of problems can compromise their accuracy. Ion specific electrodes can suffer from interference. This can be especially problematic in complex solutions, especially those containing many other ions similar in charge and size to the ion of interest. The measurements reported by ion specific sensors are also functions of the ionic strength of the solution. This can also result in significantly decreased accuracy of the sensor if the effect of the ionic strength of the solution is not accounted for. As with pH the property that is actually being

measured with an ion specific sensor is the activity of the ion, not its concentration. The activity can no longer be considered negligable in solutions with high ionic strength. In order to obtain proper readings the user must consider what effect the ionic strength of the solution has on the activity coefficient and adjust the measurement accordingly. Despite these limitations ion specific sensors can still provide accurate and valuable information, provided an appropriate measuring electrode membrane is used and the ionic strength of the solution is accounted for, and are widely used.

7.8 Problems with pH Sensors

Several problems can occur with pH sensors that can negatively affect the accuracy of these devices. The majority of these problems involve the reference electrode. One such problem is a partial clogging of the porous plug. Another error associated with the plug is due to a junction potential which can vary depending on the electrode setup. This junction potential adds error to the pH measurement when the ionic composition of the solution being tested is significantly different than the ionic composition of the standard buffer used to calibrate the electrode. In general, a junction potential develops at the interface of two solutions with different ionic compositions because the various ions in each solution have different mobilities, and as oppositely charged ions diffuse at different rates a charge separation develops at the interface between the solutions. A different type of error can occur in measurements of solutions with very low hydrogen ion concentration and very high sodium ion concentration; pH electrodes can "mistake" the sodium for hydrogen, leading to a lower pH reading than the true value.

Other problems include the contamination of the electrolyte in the reference electrode of a complete depletion of the electrolyte. Any of these problems will result in an incorrect potential in the reference electrode and ultimately inaccurate pH measurements.

Problems can also occur with the measuring electrode. Any damage to the glass membrane including cracks, etching, or abrasion, as well as coating of the glass membrane will result in poor pH measurements. In addition, if the electrode is dry it will not measure pH correctly; it must be soaked in some aqueous solution for a few hours. Problems with pH sensors can be fairly easily detected. Any problems such as difficulty calibrating the sensor, unsteady pH readings, or significant drift in the readings usually indicate damage to the pH sensor. It is now possible to test some forms of pH sensors online to quickly and easily determine which electrode is the cause of the problems and potentially the source of these problems. In some cases these problems can be fixed by a careful cleaning of the sensor, however if problems persist it may be necessary to replace the pH sensor.

7.9 pH Sensor Application Example

Would it be wise to use a pH electrode in the following situations? If not what method would you use to determine the pH of the system of interest?

A) You are an engineer responsible for a wastewater treatment facility. It is important that you know the pH of the water at several points in the treatment process to detect any potential problems.

B) You work for a pharmaceutical company who uses bacteria to produce a desired protein. The bacteria are very sensitive to changes in pH and production falls off significantly if the pH varies from optimal conditions.

C) You work in a plant that uses hydrochloric acid as a catalyst. After the reaction has gone to completion and the product is extracted it is desirable to separate the acid from the remaining solution to be reused in the plant and allow for easy disposal of the remaining solution. The EPA requires accurate data on the pH (among other things) of all discharged solution.

D) You work in a plant that produces acetic acid. You want to ensure that the solution exiting the reactor contains at least 75 mol% acetic acid.

--**SOLUTION**--

A) NO. A pH electrode would not hold up well under the harsh conditions of a wastewater treatment plant. It is also not necessary in this case to have constant measurements or exact pH values as the pH is being used simply as an early problem detection system. Simple pH paper would be sufficient in this situation.

B) YES. The sensitivity of the bacteria, as well as the likely high value of the proteins being produced calls for the accuracy and frequent measurements afforded the engineer by a glass membrane pH sensor.

C) YES. Here not only constant measurements are required, but records of these measurements must be kept to prove that environmental standards are being met. The most cost effective way to measure and record this data would be with a pH sensor connected to a computer.

D) NO. While pH could be used as an effective way to monitor the progress of the reaction a sensor is not necessary. Here you are only interested in ensuring that a certain benchmark, in this case a reaction completion percentage measured using pH, is reached. Choosing an appropriate dye indicator would be a much simpler and more cost effective way to make sure this standard is met.

7.10 Introduction to Viscosity

Viscosity quantifies the systematic deformation of the surface of a fluid upon an applied shear stress. Fluid, while in motion, typically travels at varying velocities due to the geometry of contact surfaces. It may be characterized either as a Newtonian fluid or a Non-Newtonian fluid. A Newtonian fluid flows continuously in a uniform manner

irrespective of the forces acting upon it (a common example, water). Conversely, the viscosity of a non-Newtonian fluid changes upon differing flow rates (common examples, table ketchup or mud).

Also, the effect of temperature on fluid viscosity is imperative. From a thermodynamic perspectice, temperature measures the random movement of molecules of a given substance, and as movement increases internal energy increases, and temperature also increases. For most traditional substances, having higher internal energy/temperature also implies a lower viscosity, as the substance will be less resistant to movement (i.e. will deform sooner to an applied shear stress). For example, consider honey: at room temperature it oozes out of the squeeze bottle, but after a minute in the microwave it may be easily poured out, similar to a syrup.

Nearly all modern viscometers have built in controls to address temperature issues, and a competent operator or engineer may input temperature settings to that the appropriate calculations are carried out correctly.

Viscosity may be further subdivided into two distinct forms: dynamic viscosity and kinematic viscosity. Dynamic viscosity (μ), a figure representing shear stress as proportional to the strain rate, has the SI unit Pa*s. Kinematic viscosity, ($v = \mu/\varrho$), describes shifts in momentum and has SI units m^2s^{-1}, but is also commonly reprented by the Stoke, cm^2s^{-1} (content adapted from http://scienceworld.wolfram.com/physics/DynamicViscosity.html).
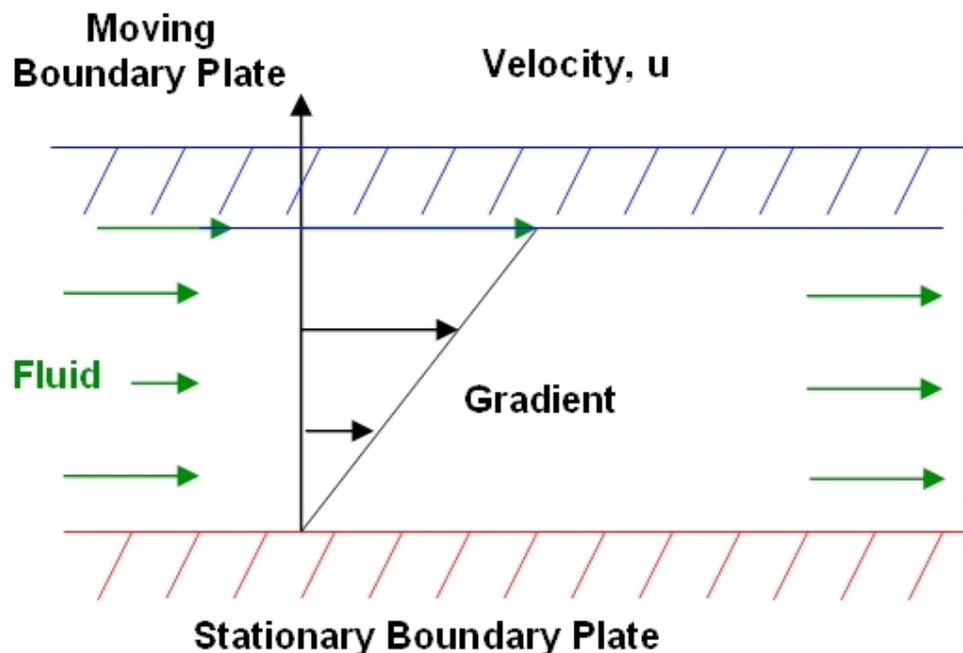
As noticed above, the typical Newtonian fluid experiences friction upon contacting a surface and resultantly develops a characteristic velocity profile. This profile may be

302

described by examining the known properties of the fluid and the surrounding structure through which it travels, commonly piping for plant chemical engineering.

In the above diagram we notice a non- cylindrical surface featuring two boundary plates, one stationary to the observer, and one mobile. Fluid contained within the two boundaries may behave according to the constraining conditions of the boundary plates, and subsequently characterized by fluid mechanics:

$$\tau = \mu * \frac{\partial u}{\partial y}$$

Assuming μ as the coefficient of viscosity, the above equation describes the relationship between the shear (parallel) stress and velocity of fluid travel per unit height above the stationary boundary plate. Fluids having a relatively high viscosity resultantly also require greater force to induce motion—resulting in a larger shear stress.

7.11 Why is Viscosity Relevant for Chemical Engineering?

For engineers a thorough knowledge of the physical and chemical properties of products is essential to the successful implementation of any design. Viscosity, similar to volatility, density, or any other physical property, offers significant insight into the potential behavior of a substance, whether classified formally as a solid, liquid or gas. By taking viscosity into account, engineers may correctly select and place instrumentation in an environmentally friendly, sustainable method.

Many professional assignments necessitate a genuine understanding of viscosity and its effects on process engineering. For a more detailed and insightful glance into the daily upkeep and activities associated with plant maintenance, please consult Michael Pearce at http://www.reynoldsmixers.com/whitepapers/030910_hotmelt.doc. He speaks at length regarding the cost- effective selection of specific types of blades used in mixing processes.

As an example, a key parameter in food process monitoring is viscosity. The viscosity is directly related to the flow characteristics of the product, which impact pumpability, pourability, and spreadability. In the food-processing environment there are a number of challenges to viscosity measurement, such as harsh process conditions and the complex rheological properties (relationships between deformations and stresses of materials). Offline measurements are often cumbersome, labor-intensive, and prone to operator error. Online viscometers must be able to deliver continuous measurement day after day with minimal maintenance. Therefore, considerations when choosing viscosity sensors for food processing should include ease of cleaning, minimal risk of fouling, and whether or not they meet sanitary requirements.

For the measurement of viscosity in controlled settings (typically laboratory) where the majority of variables are maintained as constant, engineers and scientists use off-line viscometers. Typical examples include: Capillary, Couette, Falling Ball, Cone and Plate and Oscillating Cylinder. It is very important to keep these instruments in a regulated environment with a stable temperature (such as in a water bath) because of viscosity's sensitivity to changes in temperature.

7.12.1 Capillary

The Ostwald U- Tube viscometer functions by measuing the amount of time a specified quantity of fluid takes to move from its initial position to a final position.

Typically, a suction device holds the bottom of the fluid meniscus at the start position (indicated by the solid horizontal red line), and upon its release the fluid drops an approximate distance of 5 mm (to reach the solid horizontal green line). The time for this drop is recorded so that the acquired data may then be used in Poiseuille's Law to determine the outcome of kinematic viscosity.

Both the dynamic and kinematic viscosities may be found by appling these formulae (where K_0 is the non-dimensionless viscometer constant):

Dynamic Viscosity: $\mu = K_0 \varrho t$

Kinematic Viscosity: $v = K_0 t$
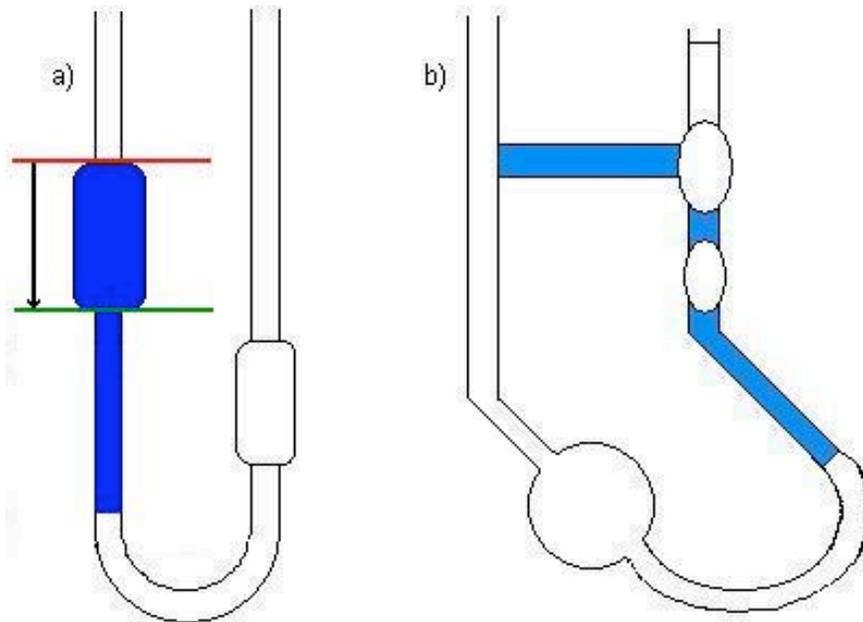
Examples of Capillary Viscometers:

a) Simple Ostwald Type
b) Cannon- Fenske Type

Content for Diagram Adapted from:

Coulson and Richardson's Chemical Enginering, Volume 3, Third Edition, Chemical, Biochemical Reactors, and Process Control; contributing editors: J.F. Richardson & D.G. Peacock,

7.12.2 Couette

This device may be used in both off-line and on-line applications with appropriate modifications in the output settings.

A couette type viscometer measures viscosity by spinning a cylinder encapsulated in fluid. This is accomplished through the synchronization (identical frequency, no phase difference) of a motor with magnetic coupling to rotate a magnet which in turns forces the inner cylinder to revolve within the fluid.

The torque reaction of the motor is resultant of the viscous drag on the rotating cylinder. This torque on the motor is effectively counteracted by the torsion bar (a thin rod connecting the control/ measuring bandto the linear variable displacement transformer).

The deflection of the torsion bar, a function of the fluid viscosity, is then subsequently converted into local signal available for laboratory analysis.

Engineers use viscometers of the couette type in in-line or in-tank applications. Also, the meters are appropriate for both Newtonian and non-Newtonian fluids. The acceptable range of viscosity spans from 10^{-3} to $5 * 10^{3} Ns / m^{2}$.

Information for this device has been adapted from Richardson's Chemical Engineering.
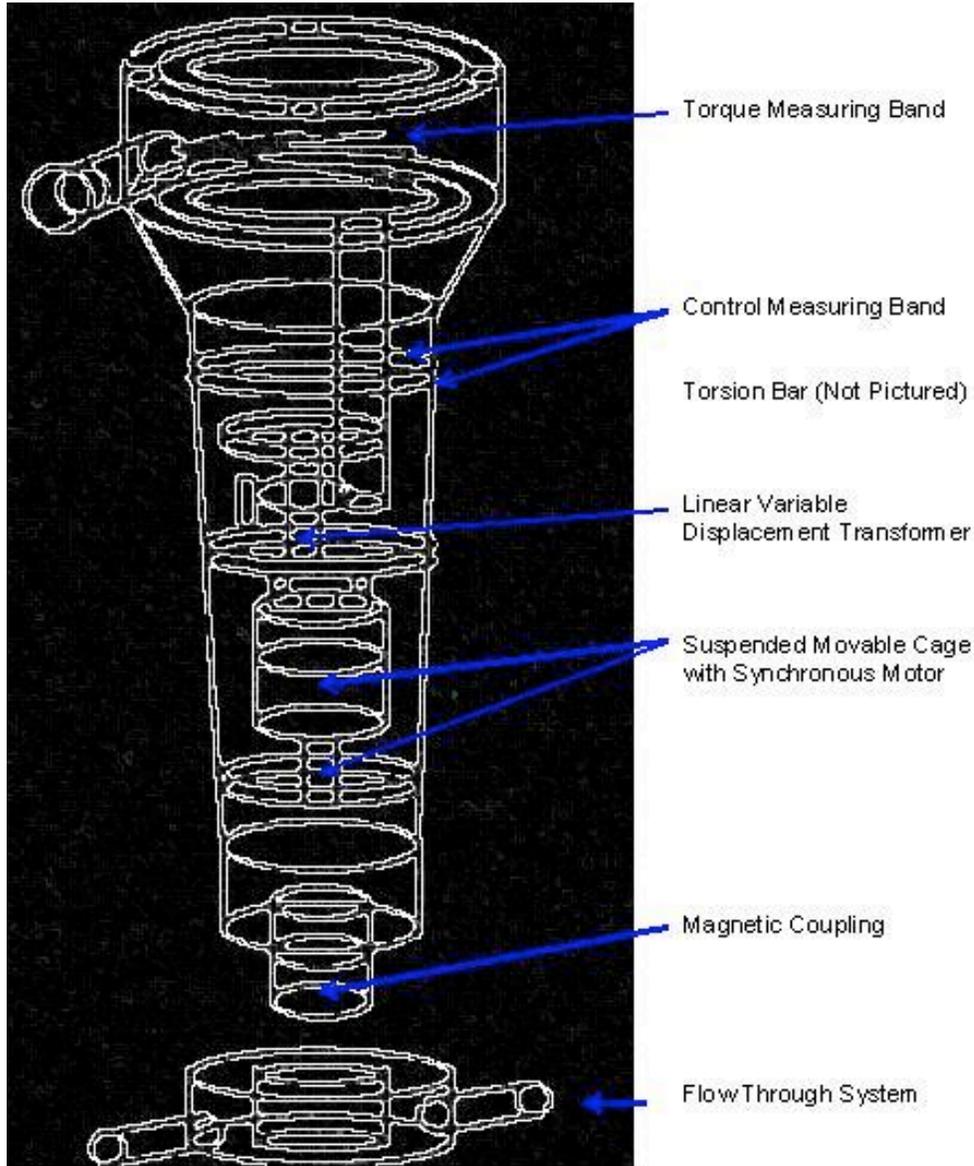
Couette- Type Viscometer for a Continuous Process

Content for Diagram Adapted from:

Coulson and Richardson's Chemical Enginering, Volume 3, Third Edition, Chemical, Biochemical Reactors, and Process Control; contributing editors: J.F. Richardson & D.G. Peacock,

7.12.3 Falling Ball

A substantially more direct approach to viscosity measurement, testers allow a ball to fall through a column containing liquid of unknown viscosity and then measure the amount of time necessary for the ball to reach the bottom of the column.

A commonly used method involves a stainless steel ball which sinks through the liquid under scrutiny; researchers measure the time necessary for the ball to drop from one preselected level to another. To attain the terminal velocity of the ball in the liquid, u_0, timers write the initial timing mark a minimum of six ball diameters below the ball release point.

Also important to note, the ratio of ball diameter to column tube diameter should not exceed 1:10 to account for the drag effects of the column tube walls on the terminal velocity of the falling steel ball.

A simple viscosity derivation from buoyancy principles is presented in Richardson's text, and is shown again here:

$$v = \frac{K_v(\rho_s - \rho)}{\rho}t$$

K_v = Dimensional viscometer constant
ρ_s = Density of the ball
ρ = Density of the liquid
t = Time required for the ball to sink the measured distance

7.12.4 Cone and Plate

Another popular method used employs a cone placed in a manner so that the tip touches the center of a stationary plate. Fluid encapsulates the cone and researchers take a measurement of the amount of torque required to keep the plate stationary.

This process allows the dynamic viscosity to be quantitatively described as:

$$v = \frac{3}{\pi \gamma r^3}T$$

T = Torque required to keep the top plate in place
γ = Rate of shear

The rate of shear is the ratio of the angular velocity of the cone to angle between the cone and the plate (note: this angle must be small enough so that the sin of the angle approximately equals the angle itself).

Important Additional Notes:

Edge effects of the cone or plate are neglected. The equation may also be used for Non-Newtonian fluids provided that γ is relatively constant.

Example of a Cone and Plate Viscometer
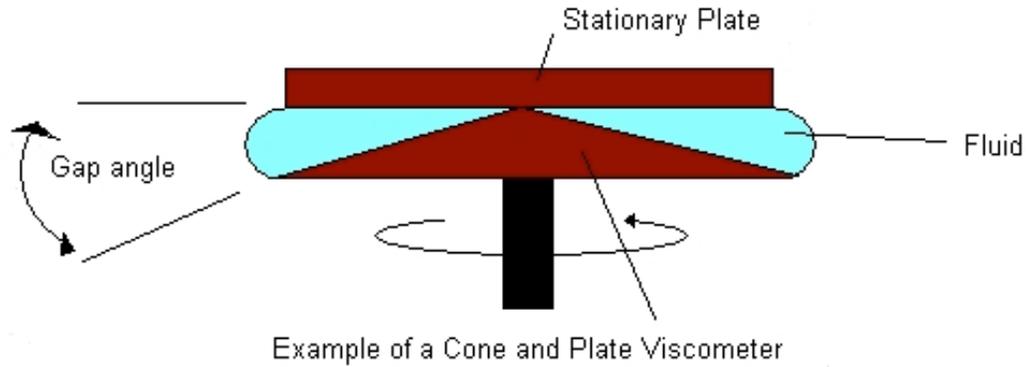
7.12.5 Parallel Plates

Parallel plates or discs viscometer is similar to the cone and plate method. The cone in the case is replaced with another plate. In this viscometer, the fluid flows in a gap between two parallel discs. One of the discs rotates with an angular velocity ω which creates the shear. Torque is applied to the other plate so it stays stationary. A normal force, F, is created by fluid elasticity and acts as to separate the two plates. No slip at boundaries is assumed.
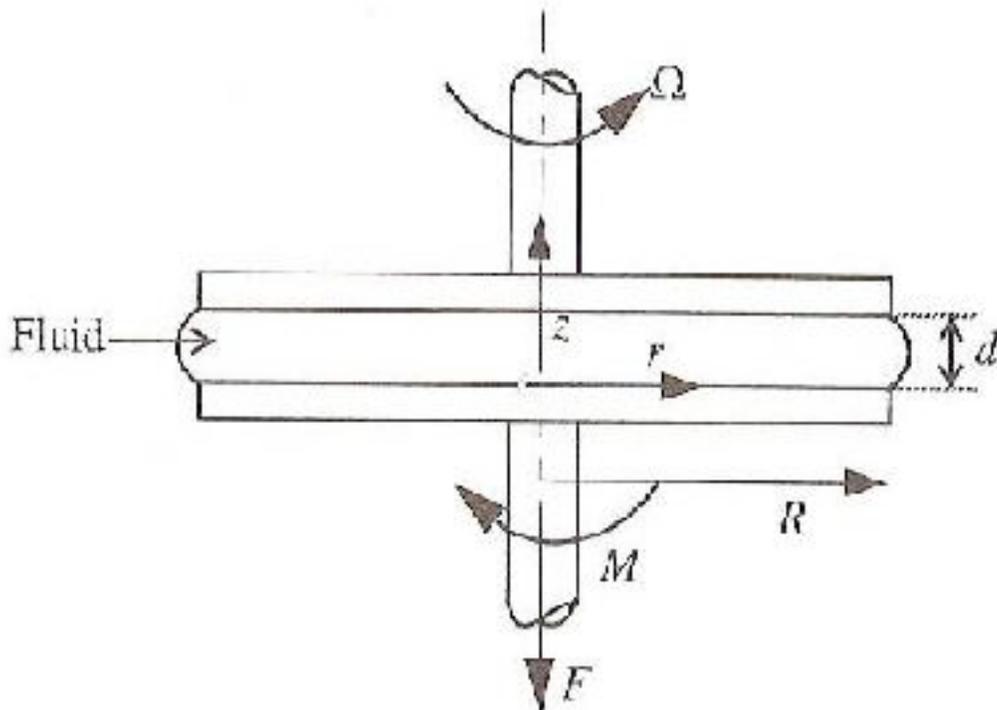
ω, T and F are all experimental parameters. Shear rate is given by:

$$\dot{\gamma} = \frac{\omega r}{d}$$

$$\dot{\gamma} = \text{Shear rate}$$

r = The radius of the disc

d = The distance between the two plates

Shear stress, τ, is given by :

$$\tau = \frac{T}{2\pi r^3}\left(3 + \frac{dlnT}{dln\dot{\gamma}}\right)$$

Viscosity is, therefore, the ratio of shear stress to shear rate. Some limitations associated with parallel plates include:

1. Sufficient data of T vs w and F vs w must be available

2. Uniform temperature at different points in the gap between the two plates is required

3. Error might be caused from edge fracture, wall slip, misalignment of the plates and viscous heating.

7.12.6 Oscillating Cylinder

This instrument involves an arrangement requiring a rotating cylinder placed in a quantity of a viscous fluid. The level of disturbance noticed in the fluid resulting from the rotation of the cylinder is then noted, and may be used if the following equation to determine viscosity:

$$v = \frac{(r_2^2 - r_1^2)}{4\pi\omega_v r_2^2 r_1^2}T'$$

r_2 = Outer cylinder radius

r_1 = Inner cylinder radius

ω_v = Angular velocity

The outer cylinder of radius r_2 rotates with an angular velocity of ω_v while the inner cylinder of r_1 remains stationary thus allowing for the torque (alternatively reffered to as 'viscous drag'), T', to be measured at predetermined values of ω_v.

7.13 On-line Instruments

On-line instruments are those that are capable of giving a precise measurement of viscosity under plant conditions. This means that they will be able to withstand the wide range of viscosities and the variable temperature, pressure, and flow rate that occur within a process plant while still obtaining the accuracy that is required. In all on-line instruments automatic temperature compensation is always required because of viscosity's temperature dependence.

7.13.1 Capillary

This type of viscometer is derived from Poiseuille's law:

$$\frac{dv}{dt} = v_s * \pi * r^2$$

v_s = Fluid velocity
r = Internal radius

The instrument has many temperature and pressure sensors in the intricate system because there must be a precisely controlled pressure drop and temperature in the stream. Response time is minimized by adding a fast loop sampling system. A capillary viscometer would be useful when using Newtonian fluids such as lubricating oils or fuel oils. It can endure viscosities ranging from $2 * 10^{-3}$ to $4\ Ns\ /\ m^2$.

7.13.2 Couette

Identical to the description provided earlier, except now with a change in the output data processing.

The deflection of the torsion bar, a function of the fluid viscosity, is now converted into a 4-20 mA output signal and is readily available for additional plant data mangement/processing.

Information for this device has been adapted from Richardson's Chemical Engineering.

7.13.3 Vibrating Rod or Cylinder

This type of viscometer can also be used in-line or in-tank. It is suitable for Newtonian, non-Newtonian fluids, and mineral slurries. The vibrating rod or cylinder works by using an electromechanical system as the driving force, or by a pulse of current through a specific alloy. The lowest viscosity that it can handle is $10^{-4} Ns/m^2$ and it can extend to $2 * 10^3 Ns/m^2$. Normally on lower viscosity ranges the cylinder is used, while with higher ranges the rod is used. Given a plants' natural vibration care must be taken when deciding on the location of an oscillation viscometer because of their sensitivity.

7.14 Table of Viscometer Uses

This table found in Richardson's text offers some guidance on the device selection process.

Common Viscometer Ranges (Content from Richardson's Chemical Engineering)

Viscometer Type	Lowest Viscosity (Ns/m^2)	Highest Viscosity (Ns/m^2)	Shear Rate Range (s^-1)
Offline (Laboratory Instruments)			
Capillary	2*10^-4	100	1- 1.5*10^4
Falling Ball	1	200	indeterminate
Couette	5*10^-4	4*10^6	10^-2- 10^4
Cone and Plate	10^-4	10^9	10^-3- 10^4
Oscillating Cylinder	10^-4	5*10^6	7*10^-3- 9*10^3
Online Instruments			Applications
Capillary	2*10^-3	4	Newtonian Fluids
Couette	10^-3	5*10^3	In-line or in-tank *
Vibrating Rod or Cylinder	10^-4	2*10^3	In-line or in-tank * #

* Also, Newtonian and Non Newtonian Fluids.
\# May be used also for mineral slurries

One may notice that the viscosity ranges of some of the viscometers overlap. In a situation where this occurs and it is necessary to choose between multiple viscometers, one needs to take into account process conditions such as changes in temperature, pressure, and solution flow rate. The appropriate viscometer should be able to handle these conditions without disrupting the process.

7.15 Problem Solving Schematic

The following schematic may be employed to select an appropriate viscometer. Additional considerations must also take place, as the usable viscometer ranges occasionally overlap. This requires the competent operator or engineer to base his or her selection from a host of other valuable criteria, including but not limited to: cost-effectiveness, difficulty of implementation, heat generation, availability of qualified local maintenance personnel, environmental impact, and long term sustainability.

Example Question:

How should an engineer correctly begin the viscometer selection process?

Solution:

Mentally develop a systematic decision schematic as indicated below:

Is the viscometer to be used for an on-line (plant implementation) or an off-line (lab use)?

Yes   No

What is the expected viscosity range of the test fluidl? (Ns/m^2)

| $2*10^{-4}$ to 100 | 1 to 200 | $5*10^{-4}$ to $4*10^9$ | 10^{-4} to 10^9 | 10^{-4} to $5*10^6$ |

Match to appropriate shear rate (s^{-1}) range:

| 1 to $1.5*10^4$ | indeterminate | 10^{-2} to 10^4 | 10^{-3} to 10^{-4} | $7*10^{-3}$ to $9*10^3$ |

Consider type of meter as an option, and proceed to rigorously check other parameters:

| Capillary | Falling Ball | Couette | Cone and Plate | Oscillating Cylinder |

What is the expected viscosity range of the test fluidl? (Ns/m^2)

| $2*10^{-3}$ to 4 | 10^{-3} to $5*10^3$ | 10^{-4} to $2*10^3$ |

What are the expected applications?

| Newtonian Fluids | In-Line or In-Tank. Used for both Newtonian and non-Newtonian fluids. | In-Line or In-Tank. Used for both Newtonian and non-Newtonian fluids. |

7.16 Viscosity Sensor Example

1) You are an engineer running a process that requires hot water at some times and cold water at others. To heat and cool your water, you run it through a heat exchanger. When you want to heat the water, you use 100°F glycol stream to heat the water (viscosity = 0.002 Ns/m^2). When you want to cool the water, you use 20°F glycol stream to cool the water (viscosity = 5 Ns/m^2). To ensure that the glycol does not clog the pumps in your process, you need to monitor its viscosity on-line. What type(s) of viscometers could you use on the hot glycol stream? On the cool glycol stream?

312

--**SOLUTION**--

1) After looking at the summary table given above, we can immediately narrow down our search to just the on-line instruments from the information given in the problem statement. For the hot glycol stream we find which instruments have a viscosity range which allows for a viscosity of 0.002 Ns/m^2. We find that the following viscometers would be acceptable: Capillary, vibrating rod or cylinder
The cool glycol stream involves the same process, but it must have a different range of viscosity which includes 5 Ns/m^2. Viscometers which are acceptable for this stream include: Couette, vibrating rod or cylinder

7.17 pH/Viscosity Sensor Key Learnings

As is apparent after reading this article, pH and viscosity measurements do not have much in common with each other. However, there is a common thread between these two types of sensors that needs to be acknowledged, and that is the method that needs to be taken when deciding what type of sensor to use. While there is only one single type of pH sensor available, before it is to be used the user needs to evaluate whether there is an easier and cheaper way (eg. using pH paper or a pH activated dye) to accomplish the task at hand. With viscometers, there are multiple types of viscometers to use over various ranges of viscosity. The user needs to evaluate process conditions such as temperature, pressure, and flow rate, in order to decide which type of viscometer is optimal for the application.

The math associated with both types of sensors is good to know for background purposes and to gain a fundamental understanding of how a sensor produces a measurement, however it is not information that greatly helps the user decided what type of pH indicator or viscometer to use. Hence, this article has tried to give the background theory behind the measurements, but a thorough analysis of this theory is not included and not necessary for the purposes of Process Control.

7.18 References

- Cole-Palmer Catalog, 2005/2006
- Liptak, Bela. *Instrument Engineers Handbook - Process Control and Optimization, Fourth Edition, Volume 2*. Taylor & Francis, New York, 2005.
- Perry, Richard. *Perry's Chemical Engineers' Handbook, Seventh Edition*. McGraw-Hill, 1997.
- Richardson, J.F. and Peacock, D.G. *Chemical Engineering, 3rd Edition, Volume 3*. Butterworth-Heinemann, 1994.
- Rodriguez, Ferdinand; et al. *Principles of Polymer Systems, 5th Edition*. Taylor & Francis, 2003. ISBN 1560329394
- Stephanopoulos, George. *Chemical Process Control - An Introduction to Theory and Practice*. Prentice Hall, New Jersey.
- Hydramotion. *Online Viscosity Control Increases Process Productivity and Efficiency*. ScientistLive 17 November 2008. 8 December 2008 http://www.scientistlive.com/European-Food-Scientist/Processing_Equipment/Online_viscosity_control_increases_process_productivity_and_efficiency/21292/

Section 8. Miscellaneous Sensors

Authors: Michael Kennett, Conor O'Bryan, Ryan Weigold

This page is a summary of different sensors not as commonly used as others in industry. This page details the methods of operation, industrial uses, and other relevant information about the sensors.

8.1 Humidity Sensors

"Humidity sensors are used to control the amount of water vapor present in many industrial processes. Textile, wood, and chemical processing is very sensitive to humidity" [1]

8.1.1 Psychrometers

Psychrometers use latent heat of vaporization to determine relative humidity in the system. This can be done by using a dry bulb thermometer with a wet bulb thermometer. The two temperatures recorded can be used with a psychrometric chart to obtain the relative humidity, water vapor pressure, heat content, and weight of water vapor in the air.

8.1.2 Hygrometers

Hygrometers are devices that sense the change in either their physical or electrical properties. Some materials such as hair or thin strips of wood change length depending on water content. The change in length is directly related to the humidity.

8.1.3 Dew Point Measuring Devices

These devices measure humidity by cooling the air until water starts to condense on the object. The amount that the air needed to be cooled before water started to form on the object can be used to determine the relative humidity.

8.1.4 Other Humidity Sensors

Microwave absorption by water vapor can be used to measure the humidity in a material

Infrared absorption can be used by hitting the object with infrared radiation and measuring the energy of the reflectance.

314

8.2 Sound Sensors

Sound sensors are important because they can be used in industrial applications such as detection of flaws in solids and location and linear distance measurements. Sound pressure waves can also cause vibration and failure.

8.2.1 Microphones

Microphones are pressure transducers, and are used to convert sound pressures into electrical signals. There are six types of microphones: electromagnetic, capacitance, ribbon, crystal, carbon, and piezoelectric.

8.3 Smoke Sensors

Smoke sensors are useful for not only safety for workers, but also environmental concerns and purity issues of processes.

8.3.1 Infrared Sensors

These sensors detect changes a signal received from an LED due to smoke or other objects in the light path from the LED.

8.3.2 Ionization Chambers

These devices can detect the difference in current between two plates that have a voltage between them. The difference is due to carbon particles from smoke that provide a conductive path between the two plates.

8.3.3 Taguchi-Type

Taguchi-type sensors are used to detect hydrocarbon gases, such as carbon monoxide and carbon dioxide. The sensor is coated with an oxidized element that when combined with a hydrocarbon creates a change in the electrical resistance of the sensor.

8.4 References

- [1] Dunn, William C. Introduction to Instrumentation, Sensors and Process Control. Boston: Artech House, 2006. Print.

315

Section 9. Valves: types, kinds, and selection

Note: Video lecture available for this section!

Authors: Daniel Katzman, Jessica Moreno, Jason Noelanders, and Mark Winston-Galant

Stewards: Jeff Byrd, Khek Ping Chia, John Cruz, Natalie Duchene, Samantha Lyu

Date Revised: September 5, 2007

- First round reviews for this page
- Rebuttal for this page

9.1 Introduction

Control valves are imperative elements in any system where fluid flow must be monitored and manipulated. Selection of the proper valve involves a thorough knowledge of the process for which it will be used. Involved in selecting the proper valve is not only which type of valve to use, but the material of which it is made and the size it must be to perform its designated task.

The basic valve is used to permit or restrain the flow of fluid and/or adjust the pressure in a system. A complete control valve is made of the valve itself, an actuator, and, if necessary, a valve control device. The actuator is what provides the required force to cause the closing part of the valve to move. Valve control devices keep the valves in the proper operating conditions; they can ensure appropriate position, interpret signals, and manipulate responses.

When implementing a valve into a process, one must consider the possible adverse occurrences in the system. This can include noise due to the movement of the valve, which can ultimately produce shock waves and damage the construction of the system. Cavitation and flashing, which involve the rapid expansion and collapse of vapor bubbles inside the pipe, can also damage the system and may corrode the valve material and reduce the fluid flow.

There are four general types of valves.

1. Electronic, or electrical valves. The movement of the ball or flap that controls flow is controlled electronically through circuits or digitally. These types of valves have very precise control but can also be very expensive.
2. Non-Return valves. These valves allow flow in only one direction, and are common in various industries. When pressure in the opposite direction is applied, the valve closes.

316

3. Electromechanical valves. These valves have electro magnets controlling whether the valve is open or closed. These valves can only be fully open or fully closed.
4. Mechanical Valves. These valves use mechanical energy in the process of opening and closing the actual valve. Larger valves can be opened and closed using mechanical processes such as levers and pulleys, whereas smaller mechanical valves can be opened or closed via a turning wheel or pulling a level by hand.

There are four major valve types: ball valves, butterfly valves, globe valves, and plug valves. There is also an array of many other types of valves specific to certain processes. Selecting which type of valve to use depends on what task the valve must carry out and in what space said valve can fit to carry out the task.

Some general features that one can take into consideration when selecting a valve are the following:

1. Pressure rating
2. Size and flow capacity
3. Desired flow condition
4. Temperature limits
5. Shutoff response to leakage
6. Equipments and pipes connected
7. Material compatibility and durability
8. Cost

9.2 Valve Types

There is a vast abundance of valve types available for implementation into systems. The valves most commonly used in processes are ball valves, butterfly valves, globe valves, and plug valves. A summary of these four valve types and their relevant applications is in the table below.

Valve Type	Application	Other information
Ball	Flow is on or off	Easy to clean
Butterfly	Good flow control at high capacities	Economical
Globe	Good flow control	Difficult to clean
Plug	Extreme on/off situations	More rugged, costly than ball valve

Following is a detailed description of the four main valve types.

9.2.1 Ball Valves

A ball valve is a valve with a spherical disc, the part of the valve which controls the flow through it. The sphere has a hole, or port, through the middle so that when the port is in line with both ends of the valve, flow will occur. When the valve is closed, the hole is perpendicular to the ends of the valve, and flow is blocked. There are four types of ball valves.

A full port ball valve has an over sized ball so that the hole in the ball is the same size as the pipeline resulting in lower friction loss. Flow is unrestricted, but the valve is larger. This is not required for general industrial applications as all types of valves used in industry like gate valves, plug valves, butterfly valves, etc have restrictions across the flow and does not permit full flow. This leads to excessive costs for full bore ball valves and is generally an unnecessary cost.

In reduced port ball valves, flow through the valve is one pipe size smaller than the valve's pipe size resulting in flow area becoming lesser than pipe. But the flow discharge remains constant as it is a multiplier factor of flow discharge (Q) is equal to area of flow (A) into velocity (V). A1V1 = A2V2; the velocity increases with reduced area of flow and decreases with increased area of flow.

A V port ball valve has either a 'v' shaped ball or a 'v' shaped seat. This allows the orifice to be opened and closed in a more controlled manner with a closer to linear flow characteristic. When the valve is in the closed position and opening is commenced the small end of the 'v' is opened first allowing stable flow control during this stage. This type of design requires a generally more robust construction due to higher velocities of the fluids, which would quickly damage a standard valve.

A trunnion ball valve has a mechanical means of anchoring the ball at the top and the bottom, this design is usually applied on larger and higher pressure valves (say, above 10 cm and 40 bars).

Ball valves are good for on/off situations. A common use for a ball valve is the emergency shut off for a sink.

9.2.2 Butterfly Valves

Butterfly valves consist of a disc attached to a shaft with bearings used to facilitate rotation. The characteristics of the flow can be controlled by changing the design of the disk being used. For example, there are designs that can be used in order to reduce the noise caused by a fluid as it flows through. Butterfly valves are good for situations with straight flow and where a small pressure drop is desired. There are also high performance butterfly valves. They have the added benefit of reduced torque issues, tight shutoff, and excellent throttling. It is necessary to consider the torque that will act on the valve. It will have water moving on both sides and when being used to throttle the flow through the valve it becomes a big factor. These valves are good in situations with high desired

pressure drops.They are desirable due to their small size, which makes them a low cost control instrument. Some kind of seal is necessary in order for the valve to provide a leak free seal. A common example would be the air intake on older model automobiles.

Top view, open configuration   Top view, closed configuration
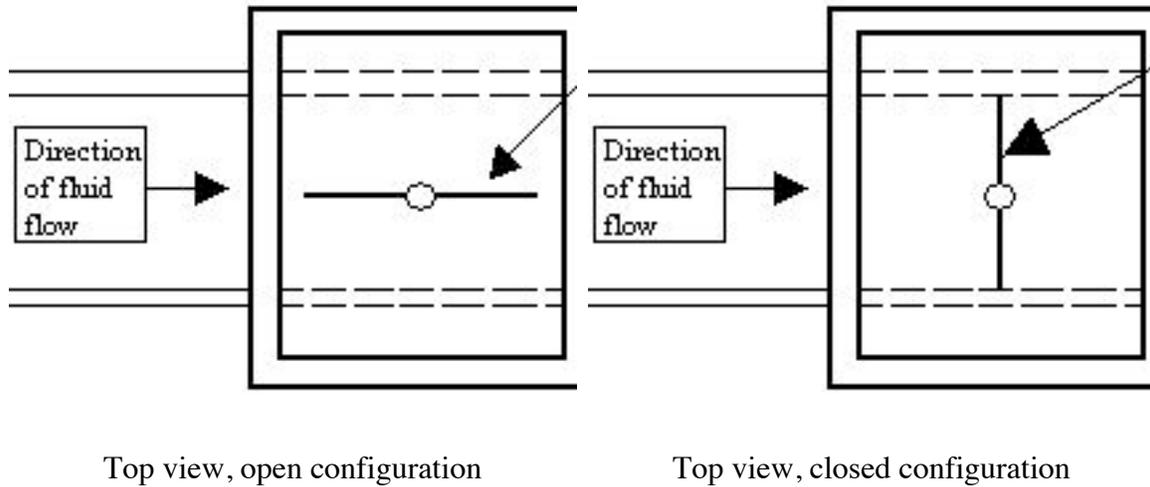
9.2.3 Globe Valves

A globe valve is a type of valve used for regulating flow in a pipeline, consisting of a movable disk-type element and a stationary ring seat in a generally spherical body. The valve can have a stem or a cage, similar to ball valves, that moves the plug into and out of the globe. The fluid's flow characteristics can be controlled by the design of the plug being used in the valve. A seal is used to stop leakage through the valve. Globe valves are designed to be easily maintained. They usually have a top that can be easily removed, exposing the plug and seal. Globe valves are good for on, off, and accurate throttling purposes but especially for situations when noise and caviatation are factors. A common example would be the valves that control the hot and cold water for a kitchen or bathroom sink.
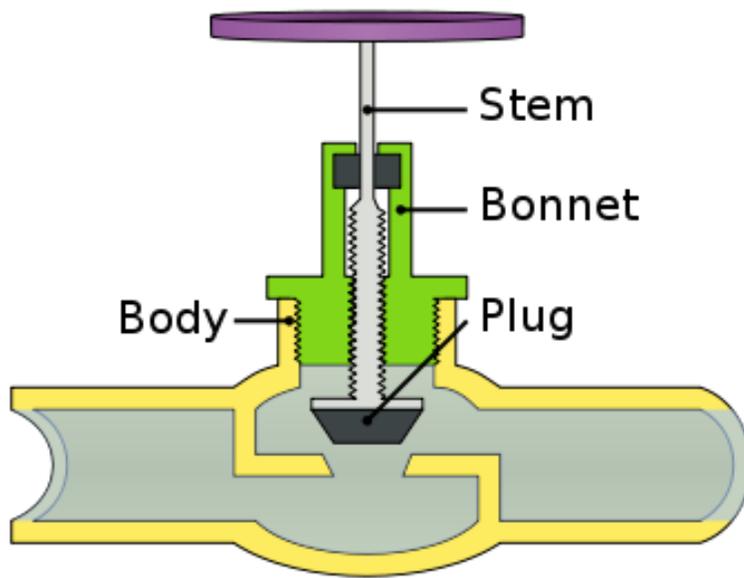
9.2.4 Plug Valves

Plug valves are valves with cylindrical or conically-tapered "plugs" which can be rotated inside the valve body to control flow through the valve. The plugs in plug valves have one or more hollow passageways going sideways through the plug, so that fluid can flow through the plug when the valve is open. Plug valves are simple and often economical. There are two types of plug valves. One has a port through a cylindrical plug that is perpendicular to the pipe and rotates to allow the fluid to proceed through the valve if in an open configuration. In the closed configuration, the cylinder rotates about its axis so that its port is no longer open to the flow of fluid. An advantage of these types of valves is that they are excellent for quick shutoff. The high friction resulting from the design, however, limits their use for accurate modulating/throttling. Schematics of this type of plug valve are below.

Top view, open
configuration

Top view, closed
configuration

Side view
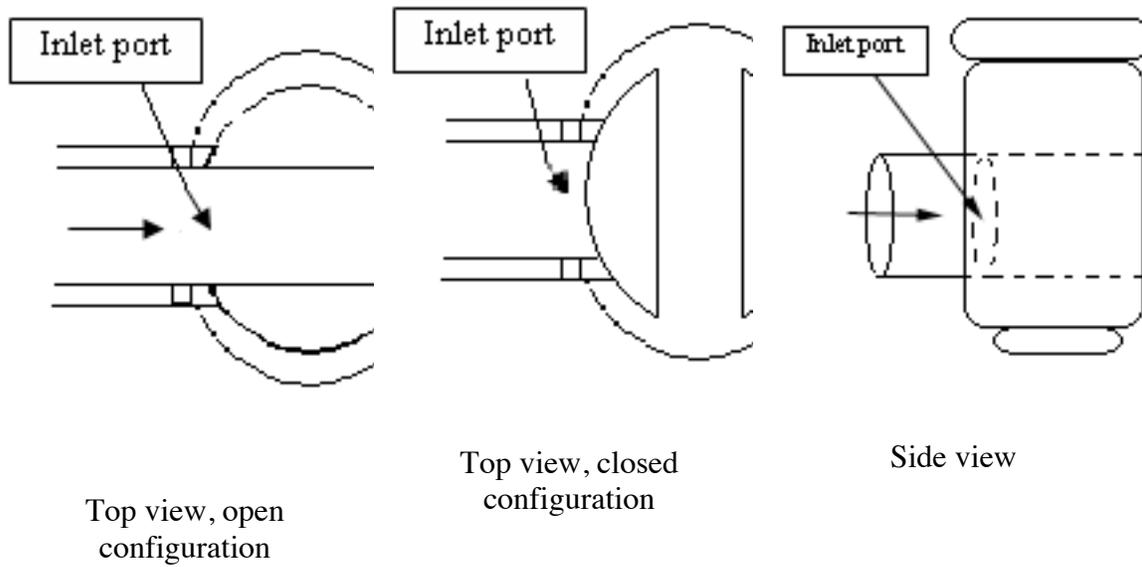
The other type of plug valve is the eccentric plug valve. In this design, the plug rotates about a shaft in a fashion similar to a ball valve. To permit fluid flow, the plug can rotate so that it is out of the way of the seat. To block fluid flow, it rotates to the closed position where it impedes fluid flow and rests in the seat. A schematic of this valve is below.
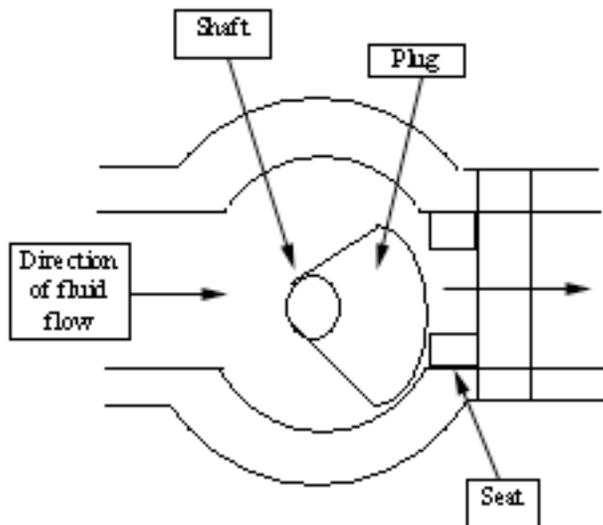
Top view, near-closed configuration

A common example would be a spray nozzle at the end of a garden hose.

9.3 Safety Valves

9.3.1 Pressure Relief Valves

Pressure relief valves are used as a safety device to protect equipment from over-pressure occurrences in any fluid process. Loss of heating and cooling, mechanical failure of valves, and poor draining and venting are some of the common causes of overpressure. The relieving system depends on the process at hand; pressure relief valves either bypass a fluid to an auxiliary passage or open a port to relieve the pressure to atmosphere. Some areas of common usage include reaction vessels and storage tanks. In the Petroleum Refining Industry, for example, the Fluidized Catalytic Cracker (FCC) reactor has several pressure relief valves to follow saftey codes and procedures on such a high pressure/high temperature process. Each of the pressure relief valves have different levels of pressure ratings to release different amounts of material to atmosphere in order to minimize environmental impact.

Here are three examples of pressure relief valves:

Conventional Spring Loaded Safety Valve

As the pressure rises, this causes a force to be put on the valve disc. This force opposes the spring force until at the set pressure the forces are balanced and the disc will start to lift. As the pressure continues to rise, the spring compresses more, further lifting the disc and alleviating the higher pressure. As the pressure inside the vessel decreases, the disc returns to its normal closed state.

Advantages:

-Most reliable type

-versatile

Disadvantages:

-Pressure relief is affected by back pressure

-Susceptible to chatter

322
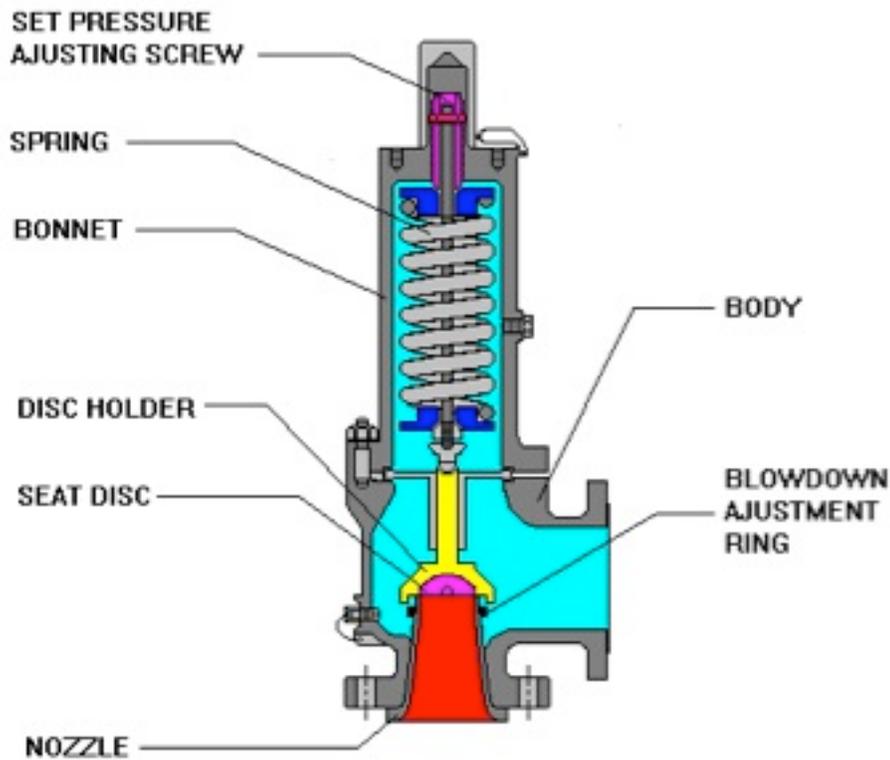
Bellows Spring Loaded Safety Relief Valve

The bellows spring loaded safety relief valve has the same principle as the conventional spring valve, with the exception of a vent located on the side of the valve. This vent lets releases the contents of the valve out to the surrounding environment.

Advantages:

-Pressure relief is not affected by back pressure

-Can handle higher built-up back pressure

-Spring is protected from corrosion

Disadvantages:

-Bellows can be susceptible to fatigue

-Not environmentally friendly (can release of toxics into atmosphere)

-Requires a venting system
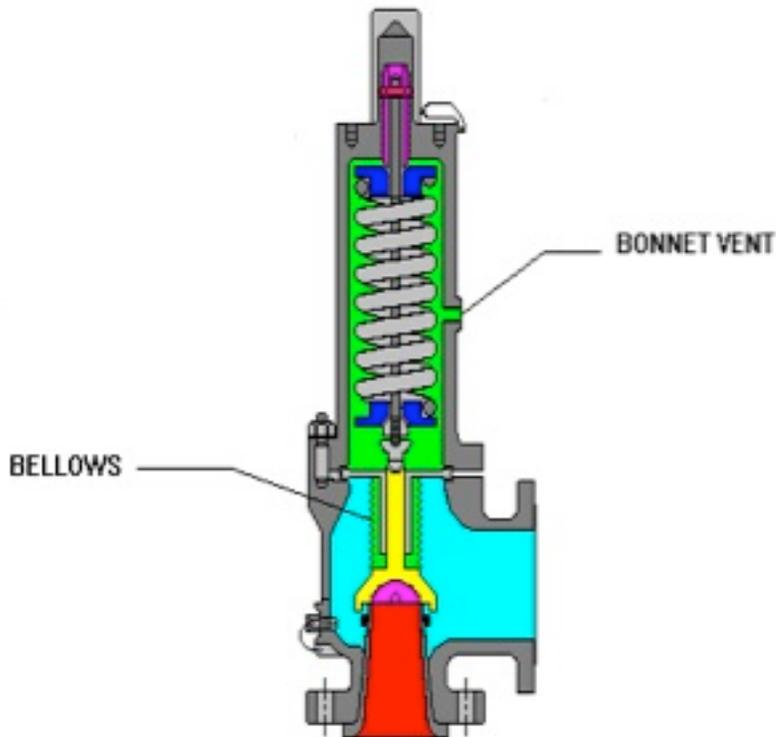
Pilot Assisted Safety Relief Valve

The pilot operated safety relief valve is also similar to the conventional safety relief valve except a pneumatic diaphragm or piston is attached to the top. This piston can apply forces on the valve when the valve is closed to balance the spring force and applies additional sealing pressure to prevent leakage.

Advantages:

-Pressure relief is not affected by back pressure

-Can operate at 98% of set pressure

-Less susceptible to chatter

Disadvantages:

-Pilot is susceptible to plugging

324

-Has limited chemical use

-Condensation can cause problems with the valve

-Potential for back flow
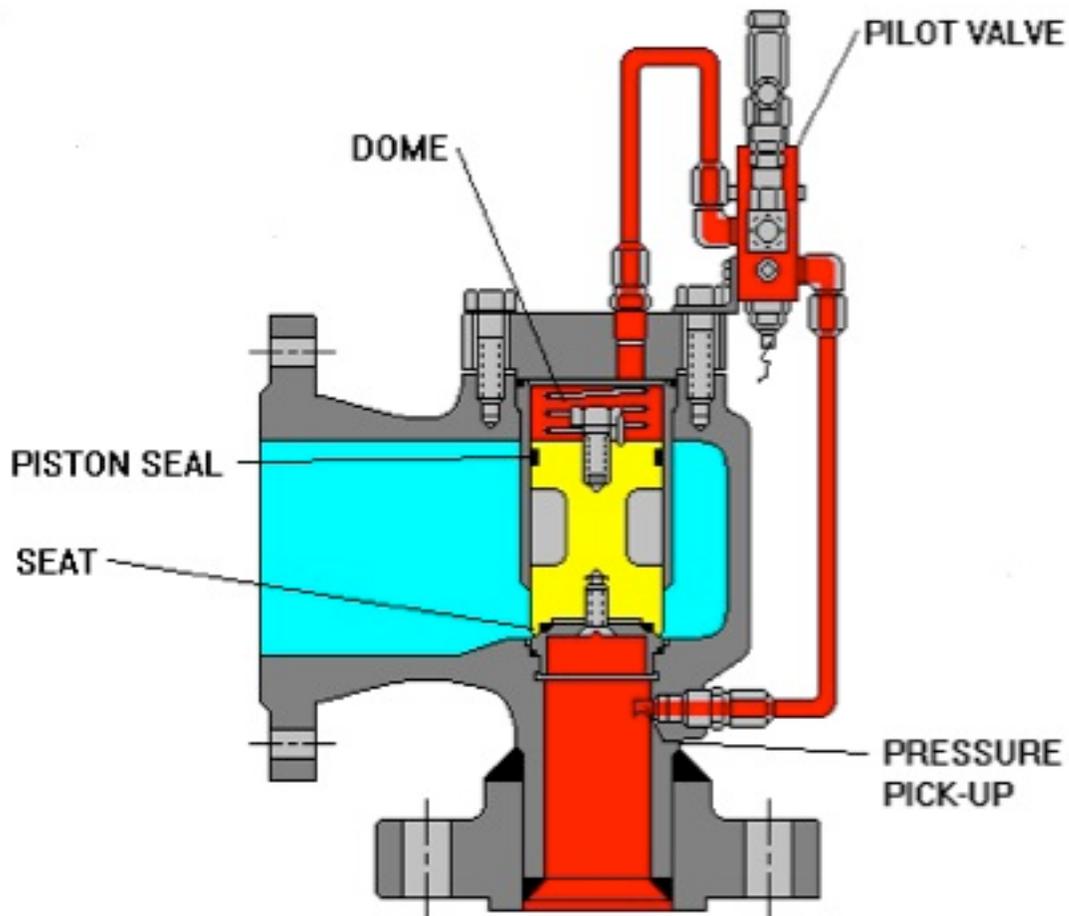
9.3.2 Steam Traps

Steam traps are devices that exist in low lying places within a pressurized steam line to release condensate and non-condensable gases from the system. Steam lines in industry are used to open/close control valves, heat trace pipelines to prevent freezing, etc. These steam traps are used in industry to save money on the prevention of corrosion and loss of steam. When these traps fail, it can mean a lot of money for the industry. There can be several hundred to several thousand in one process unit, therefore it is important to maintain and check the condition of each trap annually. The checks can be done by

visual, thermal, or acoustic techniques. Many suppliers have equipment to read the flow within the pipeline to see if it is: blocked, cold, leaking, working.

There are many types of steam traps that differ in the properties they operate on including Mechanical (density), Temperature (temperature), and Thermodynamic (pressure). For more information on the purpose and function of steam traps please view [Steam Traps].
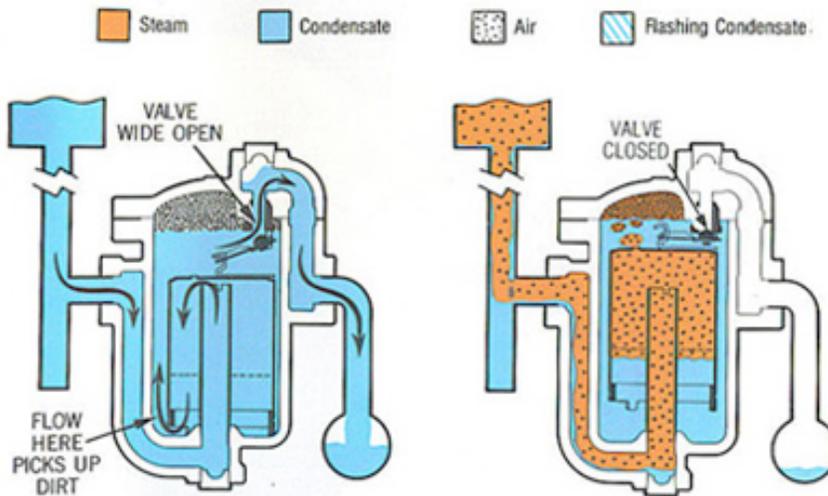
1. Steam trap is installed in drain line between steam heated unit and condensate return header. At this point, bucket is down and valve is wide open. As initial flood of condensate enters the trap and flows under bottom edge of bucket, it fills trap body and completely submerges bucket. Condensate then discharges through wide open valve to return header.

2. Steam also enters trap under bottom edge of bucket, where it rises and collects at top, imparting buoyancy. Bucket then rises and lifts valve toward its seat until valve is snapped tightly shut. Air and carbon dioxide continually pass through bucket vent and collect at top of trap. Any steam passing through vent is condensed by radiation from trap.

VALVE CLOSED

VALVE WIDE OPEN

SELF SCRUBBING FLOW
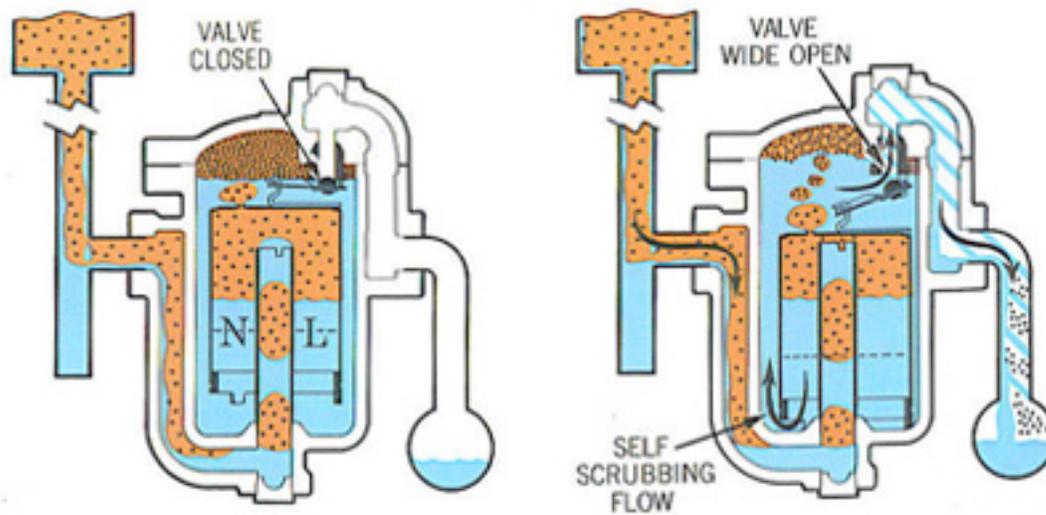
3. When entering condensate brings the condensate level slightly above the neutral line the bucket exerts a slight pull on the lever. The valve does not open, however, until the condensate level rises to the opening line for the existing pressure differential between the steam and the condensate return header.

4. When the condensate level reaches opening line the weight of the bucket, times leverage, exceeds the pressure holding valve to its seat. Bucket then sinks and opens trap valve. Any accumulated air is discharged first followed by condensate. Discharge continues until more steam floats bucket at which time cycle begins to repeat.

9.3.3 Other Safety Valves

- *Rupture Disc* - A rupture disc (also called a "bursting disc" or "safety disc") is a thin membrane of material (usually metal) that acts as a one-time use pressure relief device. At a critical pressure, the disc will fail and rupture allowing flow and the release of pressure. Often, rupture discs are used as a back-up to a conventional spring-controlled safety valve. Given primary safety valve failure (either no or inadequate pressure relief) the rupture disc will burst open and relieve pressure. Rupture discs are commonly used in chemical process plants and large air-cooled steam condensers.

- *Pressure Vacuum Valves* - Pressure vacuum (P-V) valves protect against both over-pressure and under-pressure conditions. They are commonly used in atmospheric storage tanks, to prevent the build up of excessive pressure or vacuum which can dangerously unbalance the system or damage the storage vessel.

9.4 Other Valves

In addition to the four main types of control valves, there are numerous other valves that may be necessary to manipulate fluid flow in a certain system. Some valves important to

chemical engineering processes are stated below with a brief description of their design and application.

- *Angle valves* include inlet and outlet ports that are oriented at a 90 degree angle. The fluid leaves at right angles to the direction in which it enters the valve.
- *Bleed valves* vent signal line pressure to atmosphere before removal of an instrument or to assist in calibration of control devices. Common bleed valves include ball and plug bleed valves.
- *Check valves* are one way valves. Check valves only allow fluid in one way and out the other. They are often placed on individual fluid streams when mixing fluids so as to prevent the mixture from flowing back into the original streams. Also, the speed with which the valve closes is significant to prevent reverse-flow velocity. There are different types of check valves which include lift check, swing check, tilting disk and diaphragm.
  - *Lift check valves* are quick closing. This valve should only be used for low viscosity fluids because they can be slowed by viscous fluids.

  -
  - *Swing check valve* has a disc like closing method from a hinge which may or may not be spring loaded.

  -
  - *Tilting disk check valve* is spring loaded for quick response. These are often more expensive and harder to fix when broken.

- *Diaphragm valves* have excellent shut-off characteristics and are used when there is particulate matter in the fluids. Not a good choice for controlling flow. A diaphragm valve has both a flexible and a rigid section. One advantage is that there are no crevices that affect the flow of the fluid when open. Used mostly in the pharmaceutical and food industry for their ability to be sanitary.

- *Piston valves* have a closure member shaped like a piston. When the valve opens, no flow is observed until the piston is fully moved from the seat bore, and when the valve closes, the piston removes any solids that might be deposited on the seat. Therefore, piston valves are used with fluids that have solid particles in suspension.

- *Gate valves* work by raising a gate or barrier out of the way of the flowing fluid. The water spigot outside a house works as a gate valve. Have the positive quality that, when fully open, the fluid flow is totally unobstructed. Two major types of gate valves are used: parallel and wedge gate valves. The wedge gate valve, in which the closure member has a wedge shape, provides the advantage of sealing against low pressure, as well as high pressure, flow. Gate valves have the ability to open very quickly. Advantages of gate valves

are that they have a high capacity, have good seals, relatively inexpensive, and don't have very much resistance to flow. Some disadvantages of gate valves are that they sometimes can have poor control, they can cavitate at lower pressures, and they cannot be used for throttling.

- *Needle valves* are similar to gate valves; however, they are usually applied to a much smaller orifice. Needle valves are excellent for precise control of fluid flow, typically at low flow rates.

- *Flush Bottom valves* are usually at the lowest point of a tank or reactor and used to drain out contents. Unique because it leaves no dead space in the valve when it is closed, this eliminates the problem of product buildup within the valve.

- *Pinch valves* are mainly used in order to regulate the flow of slurries in certain processes and systems. Pinch valves have flexible bodies that can be shut by pinching them. They are often used when it is necessary for the slurry to pass straight through when the valve is not pinched. They are often used for sanitary processes such as medical and pharmaceutical applications. Pinch valves can be controlled mechanically or with fluid pressure.
- *Knife valves* are used in systems that deal with slurries or powders. They are primarily used for on and off purposes; whether or not the slurry or powder flows or not. A knife gate valve can be used for fibrous material because it can cut through to close the valve.

- *Ballcock valves* are used in controlling levels in tanks. The valve is connected to a float in the tank using a lever. When the level in the tank rises, the float rises and forces the valve to be shut at the maximum level of the tank allowed. Ballcock valves are used mostly in water tanks and other low-risk areas within a certain process.
- *Solenoid valves* are used very frequently in industry. The valves have a loop of wire that is wrapped around a metal core. A current is shot through the valve creating a magnetic field, which in turn opens or closes plungers in pipelines controlling flow in that pipe. There are three types of solenoid valves:

1. Electromechanical solenoid valves use electrical energy
2. Pneumatic solenoid valves use compressed air
3. Hydraulic solenoid valves use energy from pressurized fluids

9.5 Flow Characteristics

For all valves, flow rates of fluid through the valve depend on the percentage of a full valve opening. In ball and butterfly valves, the valve opening is based on rotation. In the case for butterfly valves, an open valve is a result of a 90 degree rotation. When the valve is at the open position, the flow is parallel to the valve. The flow is uninterrupted, and therefore, no pressure is on the valve. When the valve is throttling, as is common for

globe valves, the flow rate is reduced or increased depending on the opening of the valve, and there is unequal pressure on the ends of the valve.

The ball valve flow stream involves at least two orifices; one for inlet and one for outlet flow. While traveling through an open ball valve, the fluid will continue in its flow straight through, with little pressure loss. When the ball valve is throttling, however, the fluid is subject to shearing and a change in flow rate in accordance with the percentage of the valve that is open. With high velocity liquids, the valve is susceptible to cavitation and erosion, and could produce noise. Also, at sonic velocities, the vena contracta will stop expanding, and create choking.

Globe valves are dependent on shape of the plug of the valve, in addition to the opening size, for flow variations. Lifting the globe valve will cause it to open. A flat plug is used for a quick opening; a conically shaped plug will create linear flow as the valve is raised; a rectangular shaped plug, where the bottom converges to one point directly in the middle of the plug, creates an equal percentage of flow through the valve.

9.6 Sizing

Once a specific valve type is chosen for a process, the next step is to figure out what size the valve will need to be. Topics to consider in valve sizing are the pressure-flowrate relationship and the flow characteristics to ensure an appropriate size for the desired valve.

9.6.1 Pressure-Flowrate Relationship

Control valves predominately regulate flow by varying an orifice size. As the fluid moves from the piping into the smaller diameter orifice of the valve, the velocity of the fluid increases in order to move a given mass flow through the valve. The energy needed to increase the velocity of the fluid comes at the expense of the pressure, so the point of highest velocity is also the point of lowest pressure (smallest cross section). This occurs some distance after leaving the smallest cross section of the valve itself, in a localized area known as the vena contracta.

Beyond the vena contracta, the fluid's velocity will decrease as the diameter of piping increases. This allows for some pressure recovery as the energy that was imparted as velocity is now partially converted back into pressure. There will be a net loss of pressure due to viscous losses and turbulence in the fluid.

9.6.2 Overview of Sizing Formulas

In order to determine the correct size of a valve for a specific system many factors must be considered. The most important factor is the capacity parameter, Cv, or the **flow coefficient**. The flow coefficient is a way of measuring how efficient a valve is at allowing fluid to flow through it and generally determined experimentally. Valve

manufacturers can provide you with Cv charts for the valves they sell. To determine the valve size needed for your system, you can estimate Cv with the following equations:

Liquid Flow

$$C_v = 11.7Q\sqrt{\frac{G_f}{\Delta P}}$$

Where:

- Q = flow rate (m^3 / h)

- ΔP = pressure drop through the valve (kPa)

- G_f = specific gravity of the fluid (dimensionless)

The next equation is also for calculating the flow coefficient. It should be used when you have a design flow rate, that will be the maximum flow through the valve, and a maximum pressure drop that will occur across the valve, which will be ΔP. This equation is also for turbulent flow and the units used in the calculation are important.

$$C_v = F_{max}\sqrt{\frac{G_t}{\Delta P}}$$

Where:

- F_{max} = maximum flow through the valve (gallons/minute)

- ΔP = pressure drop across the valve (psi)

- G_t = specific gravity of the liquid (unitless)

Another important piece of information about sizing and specifications is what diameter pipe may be used with a certain flow, seen in the following equation:

$$d = \sqrt{\frac{4F_{max}}{\pi v}}$$

Where:

- d = diameter of pipe (ft)

- F_{max} = maximum flow through the valve (ft^3 / s)

- v = velocity of flow (ft/s)

Note: The second and third equations are from Peter Woolf's Fall'08 lecture titled Modeling Case Study: Surge Tanks, Valves, Level sensors, and modeling.

Air and Gaseous Flow

(when $P_o < 0.53P_i$)

$$C_v = Q\frac{\sqrt{G_f(T + 460)}}{660P_i}$$

(when $P_o > 0.53P_i$)

$$C_v = Q\frac{\sqrt{G_f(T + 460)}}{1360\sqrt{\Delta P(P_o)}}$$

Where:

- Q = flow rate (ft^3 / h)
- G_f = specific gravity of the gaseous fluid (dimensionless)
- ΔP = absolute pressure drop through the valve (psia)
- P_o = outlet pressure (psia)
- P_i = inlet pressure (psia)
- T = gas temperature (degrees F)

The relationship between the inlet and outlet pressure is important, as seen above, in determining which equation to use for gaseous flow. When the outlet pressure is less than 53% the inlet pressure, this is said to be a **critical** pressure drop. The gas will behave differently when the pressure drop is critical, therefore it is necessary to use the correct equation depending on the extent of pressure drop. Once you have computed the desired value of C_v, you can choose the valve. The chosen valve must have a valve coefficient greater than or equal to the computed value.

Other sizing information

When sizing a valve it is important not to choose a valve that is too small. Take the following tips into consideration when choosing a valve.

- Valves should not be less than half the pipe size.

- Valves are easiest to control when they are open from 10 to 80% of maximum flow.

9.6.3 Using the Flow Coefficient to Determine Valve Sizes

The flow coefficient, C_v, is used as a standard sizing parameter. It is used to determine valve sizes and is also used in modeling programs for large piping systems. The C_v is an essential and practical variable that is essentially required when using pipes and valves. Try searching for a valve information and you will most likely find information requiring the C_v as well as the valve type and use.

Try looking at information for specific valves at the website of American Valve. [1]

You'll notice that every valve type and size has a listed C_v for correct sizing and function. Tables using this value are available for practically every valve imaginable and should be used to prevent over-use or under-use of valves in engineering processes.

9.6.4 Cv Charts

When sizing a valve, it is important to choose the correct Cv chart. Different Cv charts are used for different types of valves. Some different valve types are explained below. Note that *valve stroke* means the amount of flow through a pipe, and *valve travel* means the amount in which the valve is turned to achieve the valve stroke.

Equal Percentage Valves: Valve which produces equal valve stroke for equal increments in valve travel. This is the most common type of valve.

Linear Valves: Valves stroke is directly proportional to valve travel.

Quick Opening Valves: A small amount of valve travel produces a large valve stroke. Example: Gate Valves

Equal percentage and linear valve types offer the best throttling control. Examples are globe and butterfly valves.

Valves vary from manufacturer to manufacturer. Valve manufacturers can provide you with Cv charts to size each type of valve they manufacture. An examle CV chart is shown below. VT stands for valve travel.

Size (NPT)	Cv @ 10% VT	Cv @ 25% VT	Cv @ 50% VT	Cv @ 75% VT	Cv @ 100% VT
1/2"	0.03	0.075	0.3	0.5	1
1/2"	0.05	0.12	0.48	0.8	1.6
1/2"	0.08	0.19	0.75	1.25	2.5
1/2"	0.13	0.3	1.2	2	4
3/4"	0.21	0.47	1.89	3.15	6.3
1"	0.33	0.75	3	5	10
1 1/4"	0.53	1.2	4.8	8	16
1 1/2"	0.83	1.88	7.5	12.5	25
2"	1.33	3	12	20	40
2 1/2"	2.10	4.73	18.9	31.5	63
3"	3.33	7.5	30	50	100
4"	5.33	12	48	80	160
5"	8.33	18.75	75	125	250
6"	13.33	30	120	200	400

9.7 Materials

It is usually not enough to simply select the type of valve suited to given process parameters. Selecting compatible materials of construction helps ensure the lifespan of the valve, as well as the safety of the workforce, the environment and the public.

Selecting the most appropriate materials of construction of control valves is guided primarily by the service of the valve, then secondarily by cost; the least expensive material that is compatible with the service will be chosen to be used. Proper material selection promotes safety by avoiding materials of construction that may react with or be corroded by the process fluid.

The principle materials that need to be selected carefully are the wetted materials, that is, the materials that come into contact with the process fluid. These generally include the ball (for ball valves), the disk (for butterfly valves), and the plug (for plug and globe valves). Also included are the seats, which is the area where the plug or disk "sits" when closed to provide the actual shut off. The seals and the valve body are usually wetted as well.

There are many resources that contain what resources are compatible with a wide variety of process fluids, such as the Cole Parmer Chemical Resistance Database and the Cat Pumps' Chemical Compatibility Guide

334

There are design parameters inherent in the valve designs themselves that increase safety. For high service pressures (or in case of fire) some valves are designed with initial flexible seal rings that function as the primary seals. Behind these primary seals would be a backup seal of a more durable material such as 316 stainless, inconel or hastelloy. These backup seals assist in handling the additional pressure and heat.

In the highly specialized case where the process fluid is so dangerous or unsafe that any release of process fluid is unacceptable, the valve's packing can be slightly pressurized with a barrier fluid. As long as the pressure of the barrier fluid is higher than the process fluid, any leakage between the valve and the process will leak barrier fluid into the process, and not process fluid into the environment. Though as a side note, these applications usually require double containment piping and a whole host of other precautions beyond simply the safety of the valve. The most common barrier fluid is water or a water/antifreeze mix for freeze protection.

Some other considerations when selecting a material for valve are longevity/reliability of the valve and the temperature range of usage. If the valve is a control valve that gets constant use, it is important to select durable materials or to plan for replacement of the valve frequently. Service temperature is also important; materials need to be selected so the mechanical integrity of the valve is maintained throughout the entire service temperature.

9.8 Concerns

9.8.1 Noise

Noise in a system is created by the turbulence from a throttling valve. Throttling is when a valve is not used as a simple OPEN/CLOSE valve, but instead to control a flow rate.

It can be quite loud and damage people's hearing as well as the equipment. Two methods can be used to control noise. The first is to use a quiet trim that hinders system noise before it begins and the second is to treat the path (piping) where the noise already occurs. To do this, one can use thicker piping or insulation.

9.8.2 Cavitation

As previously mentioned, at the point where the fluid's velocity is at its highest, the pressure is at its lowest. Assuming the fluid is incompressible (liquid), if the pressure falls down to the vapor pressure of the fluid, localized areas of the fluid will vaporize forming bubbles that collapse into themselves as the pressure increases downstream. This leads to massive shock waves that are noisy and will certainly ruin process equipment. This is a process called cavitation.

For a given flow rate, a valve that has the smallest minimum flow area through the valve will have the most problems with cavitation (and choking, as mentioned in the next section).

9.8.3 Choking

If the fluid is compressible and the pressure drop is sufficient, the fluid could reach sonic velocity. At this point the static pressure downstream of the valve grows above design limits. This is a process known as choking, since it slows down, essentially "choking," the flow through the valve.

For a given flow rate, a valve that has the smallest minimum flow area through the valve will have the most problems with choking.

Choking is important for a gas-using process since the mass flow rate only depends on the upstream pressure and the upstream temperature. It becomes easier to control the flow rate since valves and orifice plates can be calibrated for this purpose. Choke flow occurs with gases when the ratio of the upstream pressure to the downstream pressure is greater than or equal to [(k + 1) / 2]^(k / (k - 1)). If the upstream pressure is lower than the downstream, it is not possible to achieve sonic flow.

When choking occurs in liquids, it is usually due to the Venturi effect. If the Venturi effect decreases the liquid pressure lower than that of the liquid vapor pressure at the same temperature, it will cause the bubbles in the liquid. These bubble burst can cause enough turbulence (known as cavitations) that can lead to physical damage to pipes, valves, controllers, gauges, and/or all other equipments linked to that section of the flow.

As mentioned above the limiting case of the **Venturi effect** is choked flow as the smaller space in a pipe will but a cap on the total flow rate. The Venturi effect is fluid pressure resulting from an incompressible fluid travelling through a tight or constricted section of a pipe. The Venturi effect can be explained by principles such as Bernoulli's principle and the continuity equation.

Bernoulli's equation :
$$\frac{v^2}{2} + gz + \frac{p}{\rho} = \text{constant}$$

Continuity equation:
$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0$$

for incompressible fluids: $\nabla \cdot \mathbf{u} = 0$

336

To satisfy the continuity equation, the velocity of the fluid must increase in the space but the pressure must decrease to agree with Bernoulli's equation, which is related to the law of conservation of energy through a pressure drop.

Venturi tubes are used to control the fluid speed. If an increase in speed is wanted, then the tube will decrease in diameter, or projected area, as the fluid moves away from the pump or energy source. If a decrease in fluid velocity is wanted, the tube will increase in diameter as it moves away from the energy source (which could be a pump). Using a venturi tube sometimes introduces air into the system which is mixed with the fluid in the pipe. This causes a pressure head at the end of the system.

Venturi tubes can be replaced by **orifice plates** in some cases. Orifice plates use the same principles as the venturi tubes but cause the system to suffer more permanent energy loss. Orifice plates, however, are less expensive to build then venturi tubes.

9.8.4 Upstream Pressure Increase

If a control valve is sized incorrectly and is too small, fluid velocity upstream of the valve will slow, causing an increase in pressure (much like when the end of a garden hose is partially obstructed). This increase in upstream pressure can be detrimental to certain processes, i.e. membrane filtration processes, where a large pressure difference across the membrane is desired. If the valve increases the pressure on the permeate side of the membrane, the driving force for separation will be reduced.

Decreasing the upstream static temperature will cause an increase in upstream static pressure therefore increasing the mass flow rate, so it is important to keep track of your temperature.

9.8.5 Hysteresis

Hysteresis, in general, is defined as the phenomenon where previous events can influence subsequent events in a certain process. This can create a relationship in a process going one direction, however, when carrying out the same process in the exact opposite direction, the same relationship does not occur. When processes exhibiting hysteresis are graphed there is an area between the two equally opposite curves that is used to describe the extent of hysteresis seen in the process.

The most commonly described process exhibiting hysteresis involves elastics and can be the loading and unloading of force (weight) to a stretchable rubber band. As you load more and more weight onto a rubber band, it will stretch in response to increased weight. You can subsequently remove the weight in similar increments and observe the ability of the rubber band to compress itself. Rubber bands, however, do not obey Hooke's Law perfectly therefore as you unload, the rubber band will be longer at the corresponding weights as you loaded. Below is a graphical representation of this phenomenon.

FORCE (y-axis)

Loading

Unloading

EXTENSION (x-axis)
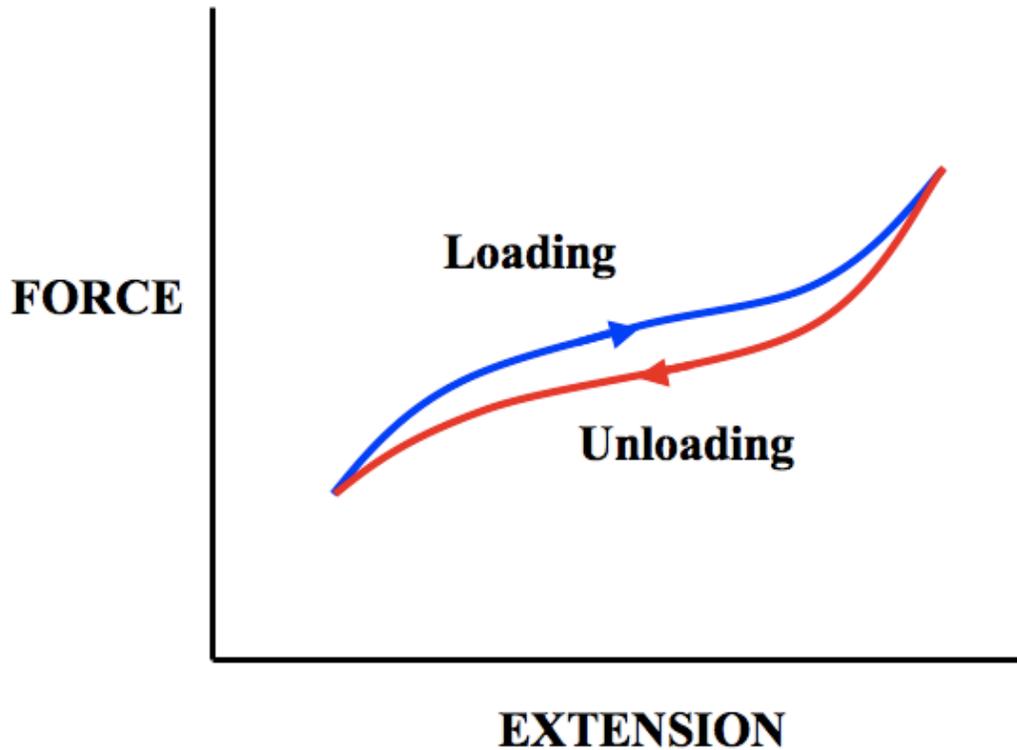
As seen in the above graph, for any given force applied to the rubber band, the extension of unloading is larger than the extension when loading. This is showing how the process of loading effected the outcome of the process of unloading. The area of the graph between the two curves is the area of hysteresis and, in this case, represents the energy dissipated as heat. The larger this area the less efficient this rubber band was at maintaining shape, and consequently the worse it obeys Hooke's Law.

Hysteresis, regarding control valves, is the principle that a control valve is dependent on the prior valve position. A common valve opening may correlate to different flow rates depending on if the valve was opened or closed to get to that position. This shows that the direction a valve is being changed may need to be accounted for in a control system to obtain a desired flow rate. If hysteresis becomes too large, it can cause the control architecture to force a system to oscillate around a desired point.

9.8.6 Chatter

Chatter is mainly a concern when selecting and sizing safety valves. Chatter is the rapid opening and closing of a safety valve or pressure relief device. Commonly, chatter is caused due to a large pressure drop between the vessel being relieved and the valve relieving the pressure.

If the pressure behind a closed safety valve builds up to some critical pressure, it will open. Once open, fluid will flow from the vessel being relieved out through the safety

338

valve in order to reduce pressure within the vessel. It is possible to have enough pipe length, fittings, or other energy-reducing obstacles so that once the high pressure flow reaches the safety valve, the pressure in the fluid is again below the critical pressure of the valve. Since there is not enough pressure in the fluid to keep the safety valve open, it closes and the vessel re-pressurizes, causing the safety valve to pop open again, locally depressurize and close again. This opening and closing is called chatter. The vibrations caused by chatter can be extremely damaging, causing unnecessary safety valve wear, possible seat-misalignment within the fittings, and even full failure in the valve or surrounding piping.

Chatter can be avoided by making sure that at critical release rates the pressure drop across the fittings to the safety valve is not large enough such that the valve will shut once fluid release is initiated.

9.9 Actuators

Actuators are the mechanical equipment that supply the force necessary to open or close a valve. Actuators are, essentially, the alternative to manual operation of a valve. The method of applying the opening/closing force to a valve is what differentiates the various types of actuators. When selecting the actuator, the most important feature to specify is whether you want fail-safe open or closed. This is determined entirely by a careful analysis of the process to decide which is safer. If all the power goes out or some other emergency occurs, the decision on the fail-safe mode of the valve is a huge factor in saving lives. For example, if a reaction is exothermic, the fail-safe mode of the cooling water in the jacket of the reactor should be fail-safe open. Pneumatic actuators have a delayed response which make them ideal for being resilient against small upsets in pressure changes of the source. Hydraulic actuators on the other hand use an incompressible fluid, so the response time is essentially instantaneous.

9.9.1 Pneumatic

Pneumatic actuators are the most popular type of actuators. The standard design of a pneumatic actuator consists of a pre-compressed spring that applies force against a disk on a sealed flexible chamber. The disk is usually attached to the stem of the valve it is intended to control. As the chamber is compressed with air, the chamber expands and compresses the spring further, allowing axial motion of the valve stem. Knowing the relationship between the air pressure in the chamber and the distance the stem moves allows one to accurately control flow through the valve.

The biggest advantage of the pneumatic actuators is their failsafe action. By design of the compressed spring, the engineer can determine if the valve will fail closed or open, depending on the safety of the process. Other advantages include reliability, ease of maintenance, and widespread use of such devices.

9.9.2 Motion conversion

Motion conversion actuators are generally used to adapt a common translational motion from the actuator's output to a rotary valve. The rod that moves axially from the translational motion actuator is connected to a disk and the connection is pivoted. The disk itself is also pivoted about its center. This system of pivots allows the translational motion to be converted into the rotation of the disk, which would open or close the rotary valve.

The main advantage of this setup is that an inexpensive translational motion actuator can be used with rotary valves. The key drawback is that the applications in which this can be used is very limited. Specifically, this setup is useless in the common case where the rotary motion required is greater than 90°.

9.9.3 Hydraulic

Hydraulic actuators work using essentially the same principal as pneumatic actuators, but the design is usually altered. Instead of a flexible chamber, there is a sealed sliding piston. Also, instead of using a spring as the opposing force, hydraulic fluid is contained on both sides of the piston. The differential pressure across the area of the piston head determines the net force.

Hydraulic actuators offer the advantages of being small and yet still providing immense force. Drawbacks of hydraulic actuators are primarily the large capital cost and difficulty maintaining them.

9.9.4 Electric

Electric actuators typically use standard motors, powered by either AC induction, DC, or capacitor-start split-phase induction. The motor is connected to a gear or thread that creates thrust to move the valve. As a failsafe, some motors are equipped with a lock in last position on its gear. This means that the gear cannot move from forces outside of the electric motor. This helps prevent overshoot on the motor as well as helps create better positioning for the gear.

Another type of motor that can be used is called a stepper motor. It uses increments on gear reduction to alleviate problems with positioning and overshoot. The increments are in a range of 5,000 to 10,000 increments in a 90 degree rotation.

A problem with using electric actuators is that a battery operated back-up system is needed or else the system is useless during power failure. Also, the actuator needs to be in an environment that is rendered safe, meaning a non-explosive environment.

9.9.5 Manual

Manual actuators are usually used for overrides of power actuators described above. This is an important safety measure in case the power actuator fails. Manual actuators typically consist of either a lever or a wheel (used for larger valves) connected to a screw or thread that turns the valve.

9.10 Summary Tables

Valve Type	Description	Advantages	Disadvantages
Ball	A ball that can be spun to allow or close flow	Operated in a 1/4 turn for quick on/off application	Size limitations
Butterfly	A disk that is rotated about it's diameter to allow, throttle, or close flow	Low pressure drop	Torque issues
Globe	A plug that moves into and out of a globe to allow, throttle, or close flow	Precise throttling	High pressure drop
Plug	A plug that rotates to allow, throttle, or close flow	Quick shutoff	Limited throttling ability

Actuator	Description	Advantages	Disadvantages
Pneumatic	Compressed air fills a chamber and pushes against a spring, operating the valve	Failsafe action, reliability, low cost, and ease of maintenance	Limited force applications
Motion Conversion	Consists of a disk connected to a rod. The rod moves axially and pivots the connection with the disk, causing the disk to also pivot.	Inexpensive and can be used for rotary valves	Limited to applications of 90° or less rotations.
Hydraulic	Similar to pneumatic, but a uses an incompressible fluid	Small size with large force	High cost
Electric	Electric motor with a rotating gear that moves the valve.	Low overshoot and accurate positioning	Needs a backup battery in case
Manual	Typically either a level or wheel connected to a screw	Good for an override in case of failure of	Not efficient as primary

The table below shows a list of typical icons for different valves found in industry.

341

Manual Valve · Angle Valve · Gate Valve · Globe Valve · Check Valve · Butterfly Valve · Ball Valve · Flush Bottom Valve · Diaphragm Valve · Plug Valve · Bleed Port Plug Valve · Needle Valve · Bleed Port Ball Valve
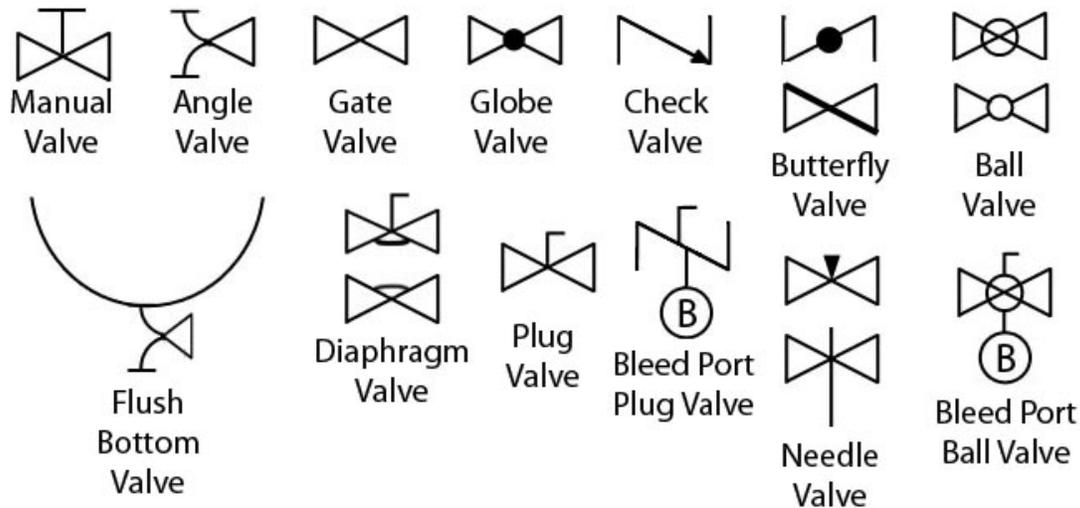
9.11 Valve and Actuator Selection Example

Note: This problem is completely fictionalized. Any relation to a real occurrence is completely coincidental.

A chemical engineer wants to use a valve at a start/stop flow of liquid water to a CSTR. The determination of the start/stop flow is a level sensor in the CSTR. Once the water reaches a certain level inside the CSTR, the flow of the water will stop. If the water level drops to a certain point, water will need to quickly flow into the CSTR. Your job is to list possible valves, and determine if each valve type can be used to start/stop the flow of water. Also, list all the potential actuators for the valves and determine which actuators *could* be used and how. Some issues to consider are choking, throttling, longevity, and reliability.

9.11.1 Solution:

Valves:

Ball valve: A ball valve can work for this example. A 2-way ball valve may not be the most ideal type of valve depending on the flow rate. If the water is moving at a high velocity, the 2-way ball valve will be susceptible to choking. If a ball valve is used, a cage valve would be the best choice, due to its durability and its on/off mechanics.

Globe valve: Just like the caged ball valve, a globe valve is a conceivable solution. The plug that would best help prevent water from flowing in the CSTR quickly would be the quick opening plug. This way, there will be no throttling when closing the valve, creating a quick on/off.

Butterfly valve: If a butterfly valve is selected, it would need to be a high performance valve. The benefits of this valve include a tight shutoff and improved torque over the non-high performance butterfly valves. This is important for the on/off capabilities for water flow the example calls for. Using a regular butterfly valve could lead to water leaking into the CSTR or torque issues after constant use.

Plug valve: A plug valve could work, but would encounter similar problems as a 2-way ball valve with choking. Also, the friction created when constantly rotating could cause a problem in the future with the valve not closing properly.

Actuators:

Pneumatic: A pneumatic actuator has a good potential to be used in this example. Its failsafe action is great in case there is a failure in the system. Its low maintenance and high reliability means that the pneumatic actuator is an ideal actuator for the valve.

Motion Conversion: Motion conversion actuators would be ideal for any rotating valve, like the high performance butterfly valve. Therefore, this actuator should be used with rotational valves for best reliability.

Hydraulic: Hydraulic actuators have similar advantages that the pneumatic actuators have. Hydraulic actuator, however, cost more than pneumatic actuators. Therefore, if another option exists, hydraulic actuators should not be used for this example.

Electric: An electric actuator will work best with rotational valves. So, if a high performance butterfly valve was chosen, then an electric actuator could possibly be used. As an effort to limit throttling issues, rotational models for ball valves would be less ideal. This is, of course, on the assumption that a back up battery would be included with the electric actuator

Manual: A manual actuator could be used in conjunction with any other actuator, but would it not be a good idea to use a manual actuator alone. The manual actuator would be a good backup method if any of the other actuators fail, but tedious if used as the primary actuator.

9.12 Selecting Control Valve Example

Note this example problem and the data contained within fictitious, with the exception of some of the Cv data, which is taken from Clark's Product Specification Site

An engineer in a chemical factory is asked to decide what type of control valve to use in throttling the line from a heat exchanger to a reactor. The chemical is part of a very sensitive reaction in which the precise addition of the chemical in question is crucial. If

an imprecise valve is utilized, the entire reactor's $5 million worth of product would need to go down the sewer. The process considerations are as follows:
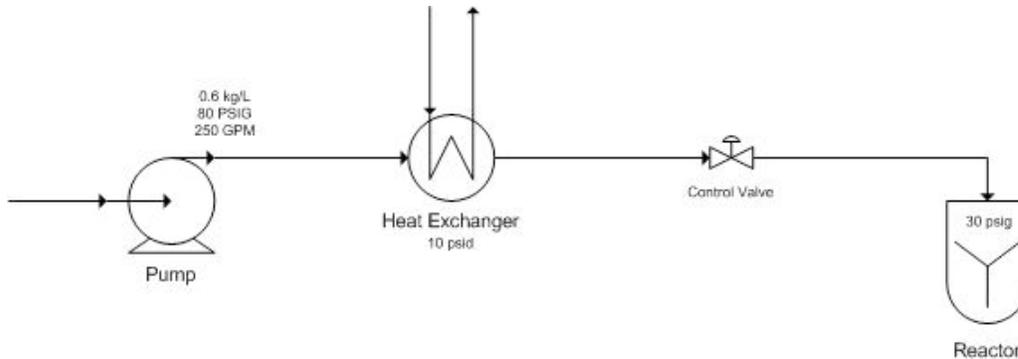
The process is designed to pump the liquid chemical at 80 psig and 250 GPM, though the flow occasionally needs to be varied from 125 to 400 GPM. The chemical has a density of 0.60 kg/L (S.G. = 0.6). The heat exchanger should take 10 psi of the pressure drop across it. Also, the reactor is operated at 30 psig Assume turbulent non-choked flow.

9.12.1 Solution:

First one must choose the type of valve. The four main valves of interest are ball, butterfly, plug and globe. Since the application is a throttling application a ball valve probably should not be our first choice. Since the chemical needs to be throttled at an extremely accurate rate, a plug valve would not be best suited here. Also, since there is significant variability of the desired flowrate, a butterfly valve probably is not the best decision either. A globe valve seems to work well here.

For reference, a list of Cv's by valve size (when fully opened) is listed below. The Cv's of each valve are listed also by % valve travel (as can be commonly acquired from a vendor), where 100% means the valve is fully open.

Size (NPT)	Cv @ 10% VT	Cv @ 25% VT	Cv @ 50% VT	Cv @ 75% VT	Cv @ 100% VT
1/2"	0.03	0.075	0.3	0.5	1
1/2"	0.05	0.12	0.48	0.8	1.6
1/2"	0.08	0.19	0.75	1.25	2.5
1/2"	0.13	0.3	1.2	2	4
3/4"	0.21	0.47	1.89	3.15	6.3
1"	0.33	0.75	3	5	10
1 1/4"	0.53	1.2	4.8	8	16
1 1/2"	0.83	1.88	7.5	12.5	25
2"	1.33	3	12	20	40
2 1/2"	2.10	4.73	18.9	31.5	63
3"	3.33	7.5	30	50	100
4"	5.33	12	48	80	160
5"	8.33	18.75	75	125	250
6"	13.33	30	120	200	400

Now one must size the globe valve. Recall the equation

$$C_v = Q\sqrt{\frac{G_f}{\Delta P}}$$

We can see that the pressure drop across the valve is the pressure at the pump minus the pressure at the reactor minus the pressure drop across the heat exchanger. Thus

$$\Delta P = 80 - 30 - 10 = 40_{\text{psi}}$$

So getting Cv,

$$C_v = 250\sqrt{\frac{0.6}{40}} = 30.62$$

Looking at the chart it would appear that the 2" Globe valve would work, but remember, the process fluid may sometimes need to be controlled up to 400 GPM. And the Cv at 400 GPM is:

$$C_v = 400\sqrt{\frac{0.6}{40}} = 49.0$$

Looking at this it is clear that the 2" globe valve would simply be fully open at 400 GPM. It appears that the 2 ½" globe valve would be a better choice since it has the capability to throttle up to the maximum design flowrate.

Checking the Cv at the minimum flow that will need to be throttled, 125 GPM:

$$C_v = 125\sqrt{\frac{0.6}{40}} = 15.3$$

We see that the 2 ½" globe valve will be between 25 and 50% open, so it will throttle well at these low flow rates at also. This reaffirms that the 2 ½" globe valve will work well for this application.

9.13 Multiple Choice Question 1

When designing a jacketed CSTR that will contain a highly exothermic reaction, what type of actuator should be used to operate the valve controlling the cooling water?

(a) pneumatic

(b) motion conversion

(c) hydraulic

(d) electric

Answer = (a)

9.14 Multiple Choice Question 2

Which of the following is a good application for a diaphragm valve?

(a) inlet to a CSTR where precise volumes of reactants are required

(b) food industry where ease of cleaning is important

346

(c) steam heating line for jacketed CSTR

(d) vent on a signal line to release pressure to atmosphere

Answer = (b)

9.15 Sage's Corner
```
http://video.google.com/googleplayer.swf?docId=1154505289288721815
```

Slideshow with no sound

9.16 References

- Edgar, Thomas F., Sale E. Seborg, and Duncan A. Mellichamp. *Process Dynamics and Control*. New York: John Wiley & Sons.
- Liptak, Bela G., *Process Control and Optimization*. Vol. II. New York: Taylor & Francis.
-
  Perry, Robert H.(1997), *Perry's Chemical Engineers' Handbook*, 7th Edition. McGraw-Hill. ISBN 0070498415
- "Valves", *Cole-Parmer*, 2005
- Considine, Douglas M. (1993), *Process/Industrial Instruments & Controls Handbook*, 4th Edition, McGRAW-HILL, INC. ISBN 0070124450
- Smith, Peter; Zappe, R.W. (2004),*Valve Selection Handbook*. Elsevier
- Matthews,Clifford . (2004),*A Quick Guide to Pressure Relief Valves (PRVs)*. John Wiley and Sons. ISBN 1860584578, 9781860584572
- Websites for steam traps:
  http://www.energysolutionscenter.org/BoilerBurner/Eff_Improve/Steam_Distribution/Steam_Traps.asp
  http://en.wikipedia.org/wiki/Steam_trap
  http://www.iklimnet.com/expert_hvac/steam_traps.html

Section 10. Valves: modeling dynamics

Note: ⊞Video lecture available for this section!

Authors: Erin Knight, Matthew Russell, Dipti Sawalka, Spencer Yendell

Date Presented: 9/19/06, Revised: 9/26/06

- First round reviews for this page
- Rebuttal for this page

10.1 Introduction

A valve acts as a control device in a larger system; it can be modeled to regulate the flow of material and energy within a process. There are several different kinds of valves (butterfly, ball, globe etc.), selection of which depends on the application and chemical process in consideration. The sizing of valves depends on the fluids processing unit (heat exchanger, pump etc.) which is in series with the valve. Sizing and selection of valves is discussed in the other wiki article on Valve Selection. Valves need to be modeled to perform effectively with respect to the process requirements. Important components for the modeling of control valves are:

1. Flow

2. Inherent Flow Characteristics

3. Valve Coefficient, C_v

4. Pressure Drop

5. Control Valve Gain

6. Rangeability

7. Installed Characteristics

Efficient modeling of the valves can optimize the performance and stability of a process as well as reduce development time and cost for valve manufacturers.

In the following sections we briefly define the various variables and equations involved in modeling valves.The purpose of the following sections is to give you an overview of the equations required to model the valves for a particular system. Example problems at the end of the article have been provided to aid in the qualitative and quantitative understanding of how valves are modeled for chemical engineering processes.

348

10.2 Flow through a Valve

The following equation is a general equation used to describe flow through a valve. This is the equation to start with when you want to model a valve and it can be modified for different situations. The unfamiliar components such as valve coefficient and flow characteristics will be explained further.

$$ F = C_v f(x) \sqrt{\frac{\Delta P_v}{sg}} $$

F = volumetric flow rate

Cv = valve coefficient, the flow in gpm (gallons per minute) that flows through a valve that has a pressure drop of 1psi across the valve.

ΔP_v = pressure drop across the valve

sg = specific gravity of fluid

x = fraction of valve opening or valve "lift" (x=1 for max flow)

f(x) = flow characteristic

10.2.1 Flow Characteristics

The inherent flow characteristic, f(x), is key to modeling the flow through a valve, and depends on the kind of valve you are using. A flow characteristic is defined as the relationship between valve capacity and fluid travel through the valve.

There are three flow characteristics to choose from:

1. f(x) = x for linear valve control
2. f(x) = $\sqrt{x}$ for quick opening valve control
3. f(x) = R^{x-1} for equal percentage valve control

   - R= valve design parameter (between 20 and 50)
   - note these are for a fixed pressure drop across the valve

Whereas a valve TYPE (gate, globe or ball) describes the geometry and mechanical characteristics of the valve, the valve CONTROL refers to how the flow relates to the "openness" of the valve or "x."

349

1. Linear: flow is directly proportional to the valve lift (used in steady state systems with constant pressure drops over the valve and in liquid level or flow loops)

2. Equal Percentage - equal increments of valve lift (x) produce an equal percentage in flow change (used in processes where large drops in pressure are expected and in temperature and pressure control loops)

3. Quick opening: large increase in flow with a small change in valve lift (used for valves that need to be turned either on or off frequently or where instant maximum flow is required, for example, safety systems)

For the types of valves discussed in the valve selection article, the following valve characteristics are best suited:

1. Gate Valves - quick opening 2. Globe Valves - linear and equal percentage 3. Ball Valves - quick opening and linear 4. Butterfly Valves - linear and equal percentage
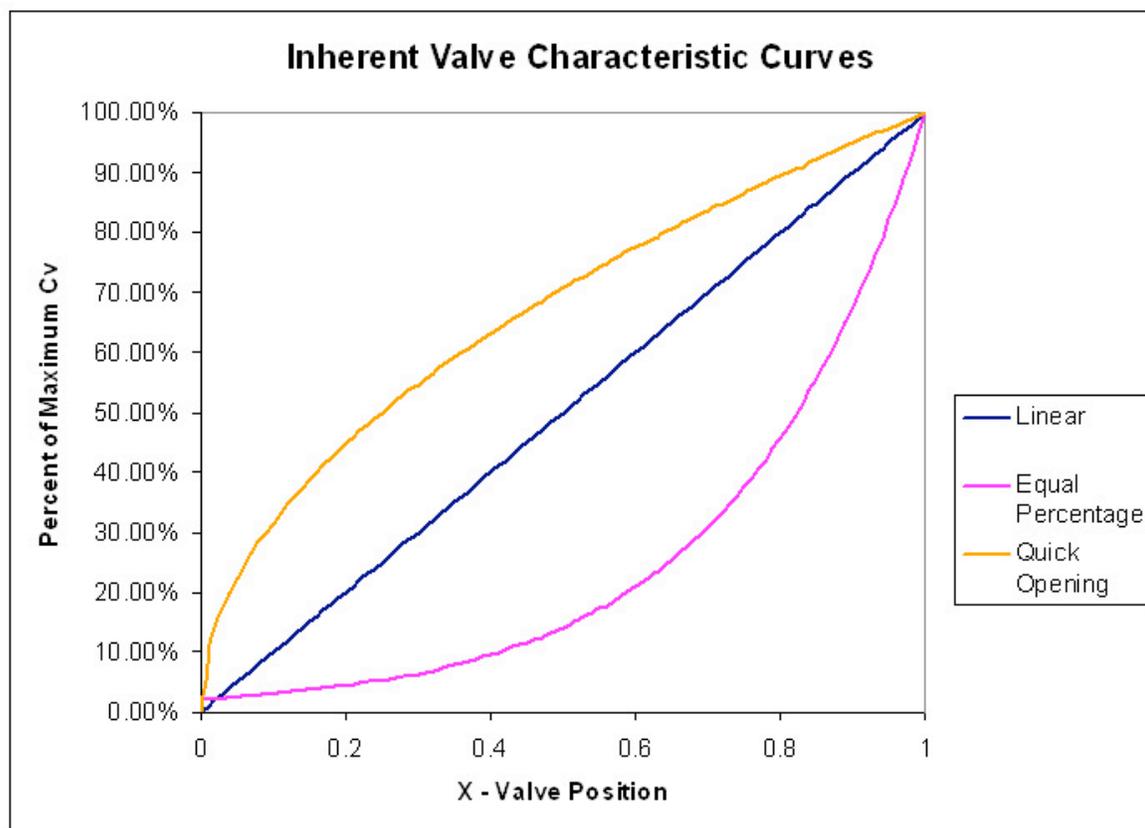
Cv_{max} depends on pipe characteristics and was chosen to be 110 gpm in this example. Constant pressure throughout the pipe line is assumed and the curves are accurate when the valve position is between 5% and 95% open.

Comparing the slopes of the graphs for the quick opening and equal percentage valves, we can see that a quick opening valve would experience greater change in flow with slight change in valve position in the lower flow range. The opposite is true for the higher range of flow. The equal percentage valve experiences slighter change in flow with respect to valve position in the lower range of flow.

When selecting the appropriate control valve, it is often the goal of the engineer to choose a valve that will exhibit a linear relationship between F and x over the normal operating position(s) of the valve. This linear relationship provides the most control for the operator. The flow characteristic observed through an installed valve, and all process factors considered (i.e. total pressure drop, etc.), is termed the installed flow characteristic. Therefore, it is not always the case that an inherently linear valve is desirable or even useful. An inherently linear valve is appropriate when there is a linear relationship between the valve position and the actual flow rate; however, consider the case where $\Delta P_L \neq 0$ and of significant value. In this case a valve with inherent equal percentage flow characteristic would be appropriate. The inherently non-linear valve would compensate for ΔP_L and result in an installed linear flow characteristic.

10.2.2 Valve Coefficient, C_v

The valve coefficient, C_v, is defined as the flow in gpm that flows through a valve with a pressure drop of 1psi across the valve ($\Delta P_v = 1psi$). Cv is an important parameter that comes up in other modeling equations. It is specific to the valve you are using.

$$C_v = \frac{29.9d^2}{\sqrt{K}}$$

d = internal diameter of the pipe in inches

K = resistance coefficient

K is specific to the pipe shape, diameter and material. Table of typical K values

10.3 Pressure Drop

The pressure drop in the pipe line (pressure drop due to the pipe line and any other equipment in series with the valve), ΔP_L, is defined as:

$$\Delta P_L = k_L \times sg \times f^2$$

f = flow through the pipe in gallons per minute [gpm]

351

$k_L = [\dfrac{psi}{gpm^2}]$ = constant friction coefficient for the pipe and any equipment in series with the valve

sg = specific gravity of the liquid

The pressure drop across the valve is defined as:

$$\Delta P_v = sg \dfrac{f^2}{(C_v)^2}$$

So, the total pressure drop is described by the equation:

$$\Delta P_o = \Delta P_v + \Delta P_L = (\dfrac{1}{C_v^2} + k_L)sgf^2$$

If the line pressure drop is negligble (constant pressure in the pipe line) then $\Delta P_L = 0$ and $\Delta P_o = \Delta P_v$. When $\Delta P_L = 0$ a valve with a linear flow characteristic will be desirable. When $\Delta P_L \neq 0$ and of significant value, a valve with flow characteristics closer to an equal percentage or quick opening valve will be more desirable.

10.4 Control Valve Gain

The gain of a control valve (K_L) is defined as the steady-state change in output (flow through a valve, f) divided by the change in input (controller signal, m). The flow through a valve, f, can have units of gallons per minute (gpm), pounds per hour (lb/hr) or standard cubic feet per hour (scfh). The controller signal, m, usually has units of percent of controller output (%CO). The basic relationship for control valve gain is shown below.

$$K_v = \dfrac{df}{dm}$$

One objective when choosing a valve is to achieve "constant valve gain". The gain is a product of the dependence of valve position on controller output, the dependence of the flow on Cv, and the dependence of Cv on the valve position. The change in valve coefficient, Cv, with respect to valve position depends on the valve characteristics f(x).

352

For linear characteristics $\dfrac{dC_v}{dx} = Cv_{max}$

For equal percentage $\dfrac{dC_v}{dx} = (lnR)\, C_v$

10.4.1 Constant Pressure Drop

The dependence of flow on the Cv depends on the pressure drop, so the equation for gain is different when there is a constant pressure drop or a variable pressure drop. If the inlet and outlet pressures do not vary with flow, the gain for either liquid or gas flow in mass units is:

$$K_v = \pm \frac{lnR}{100} W \frac{lb/hr}{\%CO}$$

%CO = percent controller output

W = mass flow rate

R = valve design parameter (usually between 20 and 50)

Note: the sign is positive if the valve fails closed (air-to-open) and negative if the valve fails open (air-to-close)

10.4.2 Variable Pressure Drop

The valve gain for variable pressure drop is more complicated. As an example, the gain for an equal percentage is

$$K_v = \pm \frac{ln\alpha}{100} \frac{flow}{(1 + k_L C_v^2)} \frac{gpm}{\%CO}$$

k_L = constant friction coefficient for line, fittings, equipment, etc.

The flow term cancels some of the effect of the Cv term until the valve is fully opened, so this gain is less variable with valve opening. Therefore the installed characteristics are much more linear when compared to the inherent characteristics of an equal percentage valve.

353

10.5 Rangeability

Valve rangeability is defined as the ratio of the maximum to minimum controlable flow through the valve. Mathematically the maximum and minimum flows are taken to be the values when 95% (max) and 5% (min) of the valve is open.

$$\text{Rangeability} = \frac{\text{Flow at 95\% valve position}}{\text{Flow at 5\% valve position}}$$

A smaller rangeablilty correlates to a valve that has a small range of controllable flowrates. Valves that exhibit quick opening characteristics have low rangeablilty values. Larger rangeability values correlate to valves that have a wider range of controllable flows. Linear and equal percentage valves fall into this category.

Another case to consider is when the pressure drop across the valve is independent of the flow through the valve. If this is true then the flow is proportional to C_V and the rangeability can be calculated from the valve's flow characteristics equation.

10.6 Modeling Installed Valve Characteristics

When a valve is installed in series with other pieces of equipment that produce a large pressure drop in the line compared to the pressure drop across the valve, the actual valve characteristics deviate from the inherent characteristics. At large in-line pressure drops the pressure drop, and consequently the valve coefficient, varies with flow through the valve. These changes can cause changes in the rangeabilty and distorts inherent valve characteristics.

In the following Microsoft® Excel model, the variation from inherent valve characteristics is illustrated. A number of parameters can be changed to match the flow conditions through a valve. This model simulates both linear and equal percentage valve characteristics. To more clearly demonstrate the deviation from inherent characteristics; simply change the C_{Vmax} value of the valve. Notice how the installed valve characteristics and valve rangeability change drastically.

Installed Valve Characteristcs Model

10.7 Special Considerations for the Equation describing Flow Through a Valve

Compressible Fluids

- Manufacturers such as Honeywell, DeZurik, Masoneilan and Fischer Controls have modified the flow equation to model compressible flows. The equations are derived from the equation for flow through a valve but include unit conversion factors and corrections for temperature and pressure, which

354

affect the density of the gas. It is important to remember to account for these factors if you are working with a compressible fluid such as steam.

Accuracy

- This equation, and its modified forms, is most accurate for water, air or steam using conventional valves installed in straight pipes. If you are dealing with non-Newtonian, viscous or two phase systems the calculations will be less accurate.

10.8 Example 1: Verbal Model of a Control Valve

10.8.1 Problem Statement:

Verbally model a fail open control valve positioned as a safety measure on a reactor processing an exothermic reaction.

10.8.2 Solution

1.)**Describe the Process:** In the fail-open control valve a quick opening valve opens with a failure signal. Open is its default position once the signal goes off.

2.) **Identify Process Objectives and Constraints:** A fail-open control valve is a safety measure. For example, if your cooling heat exchanger fails and the reactor starts to heat up and produce excess gases, the fail-open control valve would release excess gasses before pressure builds up to produce an explosion. The size of the valve is a constraint because it limits how much fluid can escape. The valve size determines the maximum flow rate. The shape and angles of the valve are modeling constraints. Sudden and gradual contraction or enlargement of the pipe diameter inside the valve and the connecting piping, will change the resistance coefficient and therefore the maximum velocity.

3.) **Identify Significant Disturbances:** Significant internal disturbances include the escalating pressure and temperature as the exothermic reaction gets out of control.

4.) **Determine the Type and Location of Sensors:** A pressure sensor would be located in the tank with the control valve that would provide the signal to the fail-open control valve. To achieve redundancy, a temperature sensor would be located on the heat exchanger to signal failure of the cooling apparatus.

5.) **Determine the Location of Control Valves:** A fail-open control valve (or multiple valves) would be placed on the top of the tank to allow exit of the gasses in the processing unit.

6.) **Apply a Degree-of-Freedom Analysis:** The only manipulated variable is the valve coefficient. This depends on the valve's diameter and resistance coefficient K. The

control objective is the maximum flow rate. The pressure drop will vary according to the failure. Therefore there is one degree of freedom.

7.) **Implement Energy Management:** This doesn't apply to our confined example, but in a larger system we could install a backup cooler if this reaction were truly dangerous.

8.) **Control Process Production Rate and Other Operating Parameters:** The exit flow rate can not exceed the maximum flow rate through the control valve.

9.) **Handle Disturbances and Process Constraints:** If our first control valve fails to sufficiently lower the tank pressure, a signal would be sent to a second valve and depending on the reaction, a backup cooling system. A secondary cooling system would be too expensive for many cases, but if you were dealing with a nuclear reactor or something highly explosive it may be worth the investment.

10.) **Check Component Balances:** Does not apply. Preventing accumulation is the point of this control valve.

11.) **Apply Process Optimization:** Our manipulatable variable is choosing a valve with a specific C_v. The valve should be able to withstand extreme temperatures and high pressures. It would be a gate valve, which opens completely upon failure. For other sizing concerns refer to "Valve Sizing."

10.9 Example 2: Modeling Installed Valve Characteristics

10.9.1 Problem Statement

A new valve is being installed downstream from a water pump. The friction coefficient of the pump and associated piping that will be in series with the new valve is

$$k_L = 1.4 * 10^{-4} \left(\frac{psi}{gpm^2} \right)$$

. The flow through the line from the pump is 300 gpm. The desired pressure drop across the valve is 4 psi. A high level of control is desired for the flow through the new valve. Two valves are being considered, one has an inherent linear characteristic, the other is equal percentage ($\alpha=50$). From the manufacturer's literature, both have a C_{Vmax} value of 200. Use the [Installed Valve Characteristcs Model](#) to determine which valve has a higher range of controllable flows values.

10.9.2 Solution

To view the valve model spreadsheet with the correct inputs, [click here](#).

Note that the pressure drop across the pipe is 13.5psi, which is significantly larger than

356

the pressure drop across the valve (4 psi). These conditions indicate that the characteristic flow through the valves may not match the inherent characteristics. This is verified by the plots and also by the calculated rangeability values shown in the valve model spreadsheet. The equal percentage valve has a higher rangeabilty value, corresponding to a higher range of controllable flows.

10.10 References

- Bequette, B. Wayne. *Process Control Modeling, Design, and Simulation*, Upper Saddle River, New Jersey: Prentice Hall.
- Crane Co. *Flow of Fluids Through Valves, Fittings, and Pipe*, Joliet, IL: CRANE.
- "Friction Losses in Pipe Fittings"(PDF), Western Dynamics, LLC., retrieved September 11, 2006.
- Perry, R. H., and D. Green (ed). *Perry's Chemical Engineering Handbook*, 7th ed. New York: McGraw-Hill.
- Seborg, Dale E., Thomas F. Edgar, Duncan A Mellichamp. *Process Dynamics and Control*, New York: John Wiley & Sons.
- Smith, Carlos A., Armando B. Corripio. *Principles and Practice of Automatic Process Control*, 3rd ed. New York: John Wiley & Sons.
- "Valve Sizing and Selection." The Chemical Engineers' Resource Page. 1442 Goswick Ridge Road, Midlothian, VA 23114. retrieved Sept 24, 2006.

Chapter 4. Piping and Instrumentation Diagrams

Section 1. P&ID General Information

Piping and Instrumentation Diagrams (P&IDs) use specific symbols to show the connectivity of equipment, sensors, and valves in a control system. The following sections will outline general information about P&IDs that is necessary to to know before trying to draw one.

1.1 P&ID vs. PFD

P&IDs may often be confused with PFDs, or process flow diagram. P&IDs and PFDs generally utilize the same notation for equipment. However, they serve different purposes and provide different information. The purpose of a PFD is to show exactly what a process does during operation, and a P&ID shows all controllers, valve types and the materials that are used in construction. A PFD shows the connectivity and relationships between the major equipment and materials in a process. However, it also includes tabulated design values such as normal, minimum, and maximum operating conditions that a P&ID does not contain. A PFD does not include minor piping systems or other minor components that a P&ID typically includes. The difference between P&IDs and PFDs is that P&IDs typically include more information regarding piping and safety relief valves than process flow diagrams. P&IDs do not contain operating specifications that PFDs contain, such as stream flows and compositions. It is important to note that differences between PFDs and P&IDs will vary between institutions. Most corporations maintain designated standards to create and modify the documents. Both PFDs and P&IDs are controlled documents and need to be maintained with a document control procedure. To see an industrial example of PFD vs P&ID refer to, [PFD/PID industry example Wiki Page](#).

1.2 Information Incorporated in P&IDs

The following information is given on a P&ID that is not explicit on a PFD:

- ALL valves and valve types
- Controllers present
- Controller architectures
- Pipe diameters, materials of construction, and insulation properties (including minor piping systems)
- Equipment materials of construction

1.3 Uses of P&IDs

- Develop operational methodology
- Develop safety philosophy and safeguards
- Develop control philosophy
- Serve as a basis for control programming
- Serve as a communication document for how the process works
- Serve as a basis for equipment design, pipe design, estimating cost, purchasing
- Use for evaluation of construction process
- Train employees
- Serve as a conceptual layout of a chemical plant
- Provide a common language for discussing plant operations

1.4 Characteristics of P&IDs

- Grouped by specific section of plant
- Show the connections between all the sensors and actuators
- A general schematic - NOT a layout and NOT to scale. It should be noted that P&IDs do not specifically imply the following: same elevation of equipment, relative sizes, where valves are located on the equipment, how close equipment is to each other, and impeller types/location. They are also not the same as control or incidence diagrams. This type of information can be seen in either a plant layout drawing (can be either satellite view, showing distance between units, or a slice of building, showing height of units) or construction drawings, such as plant blueprints.
- Must be clear and uncluttered
- Must be systematic and uniform. P&IDs are used extensively in industry to document plant information. These documents need to be easily read by anyone working within the company, and easily explained to anyone else. OSHA audits can occur anytime and it is imperative that operational information can be provided to the auditor when requested. Without standard notation, it would be very difficult to go from plant to plant within your company and understand the P&IDs.
- Are generally highly confidential and have restricted access, as they can be used to replicate a process

1.5 What A P&ID Is Not

- Not an architectural diagram of a process (should show the flow of material across the plant floor between sensors and actuators, not necessarily corresponding to a 3D location)
- Does not need to be drawn perfectly to scale
- Does not imply any relative elevations
- Do not need manual switches

- No actual temperature, pressure, or flow data
- Leave out any extensive explanations

1.6 What A P&ID should include

- Instrumentation and designations
- Mechanical equipment with names and numbers, and their specifications such as material, insulation, maximum flow rate, working pressure and temperature, maximum power etc.
- All valves and their identifications
- Process piping, sizes and identification
- Miscellaneous - vents, drains, special fittings, sampling lines, reducers, increasers and swagers
- Permanent start-up and flush lines
- Flow directions
- Interconnections references
- Control inputs and outputs
- Interfaces for class changes
- Vendor and contractor interfaces
- Identification of components and subsystems delivered by others
- Intended physical sequence of the equipment

1.7 P&ID Revisions

- Revisions should be clearly identified
- Regularly issued to all related employees at each significant change to the process, as well as at benchmark points
- If small changes are made that don't warrant a completely new revision, "red pencil" additions are generally accepted between issues
- 15-20 revisions are typical during process design
- All revisions need to be communicated to EVERYONE so that only the latest revision is used. This is critical in order to avoid serious (not to mention expensive) construction mistakes. (Typically outdated P&ID is discarded to avoid confusion)

1.8 How do you generate a P&ID?

P&IDs can be created by hand or computer. Common programs, for both PC and Mac, that create P&IDs include Microsoft Visio (PC) and OmniGraffle (Mac). For creating a P&ID by hand or on a computer, please refer to the P&ID Standard Notation Wiki Page for standard equipment notation as well as computer templates.

Section 2. P&ID standard notation

Title: Piping and Instrumentation Diagram Standard Notation
Authors: Halley Crast, Andrew Laskowski, Maurice Telesford, Emily Yatch
Date Presented: September 21, 2006
Date Revised: October 3, 2006
Stewards: Ji Sun Sunny Choi, Sang Lee, Jennifer Gehle, Brian Murray, Razili Stanke-Koch

Last Updated: December 14, 2009

By: Stephanus Oscar, Kaitlin Harrington , Suhendra Lie

- First round reviews for this page
- Rebuttal for this page

2.1 Introduction

Piping and Instrumentation Diagrams (P&IDs) use specific symbols to show the connectivity of equipment, sensors, and valves in a control system. These symbols can represent actuators, sensors, and controllers and may be apparent in most, if not all, system diagrams. P&IDs provide more detail than a process flow diagram with the exception of the parameters, i.e. temperature, pressure, and flow values. "Process equipment, valves, instruments and pipe lines are tagged with unique identification codes, set up according to their size, material fluid contents, method of connection (screwed, flanged, etc.) and the status (Valves - Normally Closed, Normally Open)."[1] These two diagrams can be used to connect the parameters with the control system to develop a complete working process. The standard notation, varying from letters to figures, is important for engineers to understand because it a common language used for discussing plants in the industrial world.

P&IDs can be created by hand or computer. Common programs, for both PC and Mac, that create P&IDs include Microsoft Visio (PC) and OmniGraffle (Mac). As with other P&IDs, these programs do not show the actual size and position of the equipment, sensors and valves, but rather provide a relative positions. These programs are beneficial to produce clean and neat P&IDs that can be stored and viewed electronically. See below for P&ID templates for these programs.

This section covers four main types of nomenclature. The first section describes the use of lines to describe process connectivity. The second section describes letters used to identify control devices in a process. The third section describes actuators, which are devices that directly control the process. The final section describes the sensors/transmitters that measure parameters in a system.

Line symbols are used to describe connectivity between different units in a controlled system. The table describes the most common lines.

Table 1: Line Symbols
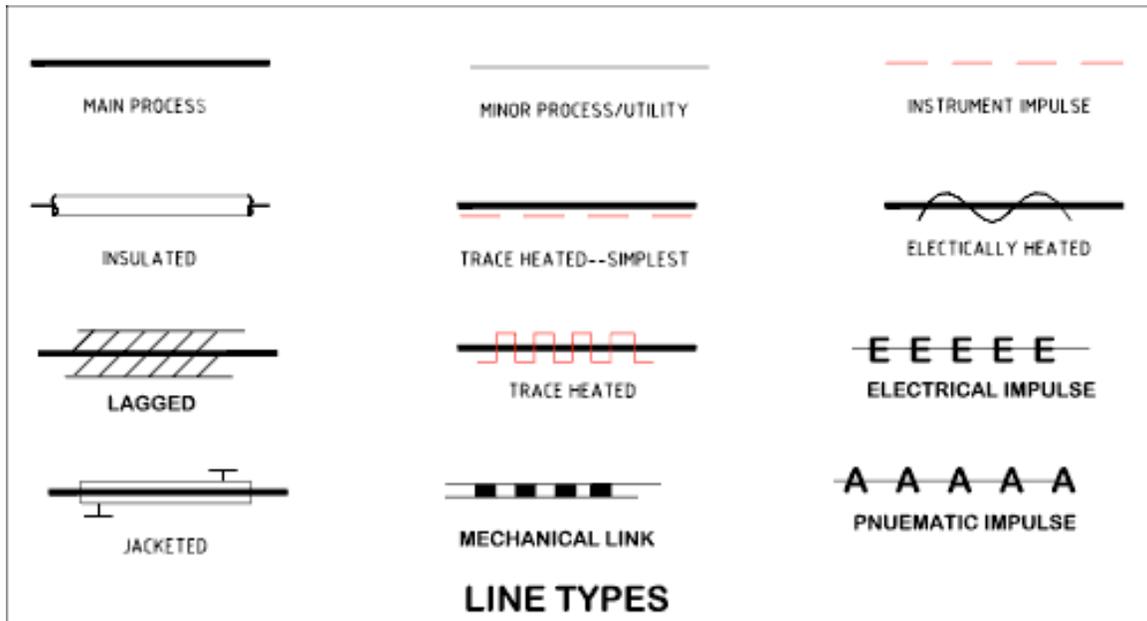
In Table 1, the "main process" refers to a pipe carrying a chemical. "Insulated" is straightforward, showing that the pipe has insulation. "Trace heated" shows that the pipe has wiring wrapped around it to keep the contents heated. "Lagged" indicates on a P&ID that the pipe is wrapped in a cloth or fiberglass wrap as an alternative to painting to improve the appearance of the pipe see here for more information. The last column in Table 1 shows pipes that are controlled by a controller. "Electrical impulse" shows that the manner in which information is sent from the controller to the the pipe is by an electrical signal, whereas "pneumatic impulse" indicates information sent by a gas.

In addition to line symbols, there are also line labels that are short codes that convey further properties of that line. These short codes consist of: diameter of pipe, service, material, and insulation. The diameter of the pipe is presented in inches. The service is what is being carried in the pipe, and is usually the major component in the stream. The material tells you what the that section of pipe is made out of. Examples are CS for carbon steel or SS for stainless steel. Finally a 'Y' designates a line with insulation and an 'N' designates one without it. Examples of line short codes on a P&ID are found below in Figure A.
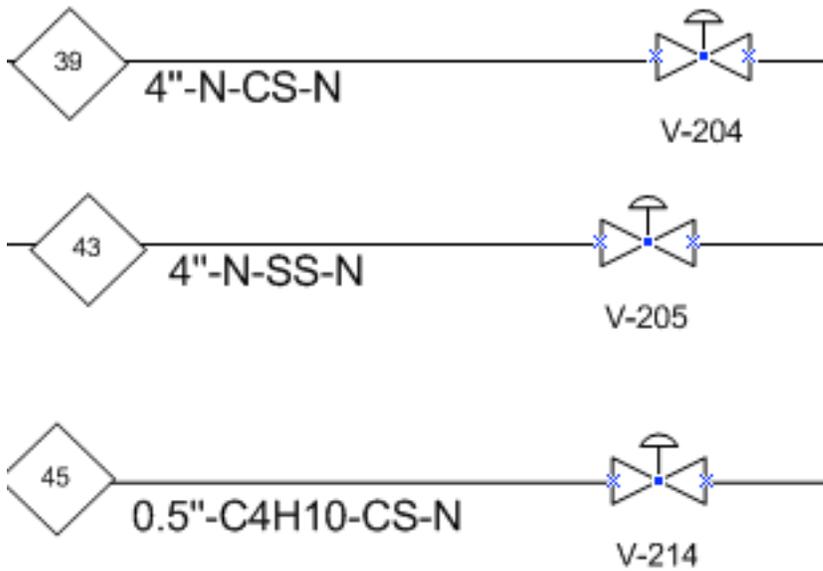
Figure A: Line Labels

This is useful for providing you more practical information on a given pipe segment.

For example in stream 39 in Figure A, the pipe has a 4" diameter, services/carries the chemical denoted 'N', is made of carbon steel, and has no insulation.

2.3 Identification Letters

The following letters are used to describe the control devices involved in a process. Each device is labeled with two letters. The first letter describes the parameter the device is intended to control. The second letter describes the type of control device.

Table 2: First Identification Letter   Table 3: Second Identification Letter

First Letter	Parameter Controlled
A	Analysis
C	Conductivity
D	Density
E	Voltage
F	Flow rate
I	Current
L	Level
M	Moisture (Humidity)
P	Pressure or Vacuum
T	Temperature
V	Viscosity

For example, the symbol "PI," is a "pressure indicator."

2.4 Valve Symbols

The following symbols are used to represent valves and valve actuators in a chemical engineering process. Actuators are the mechanisms that activate process control equipment.

Table 4: Valve Symbols
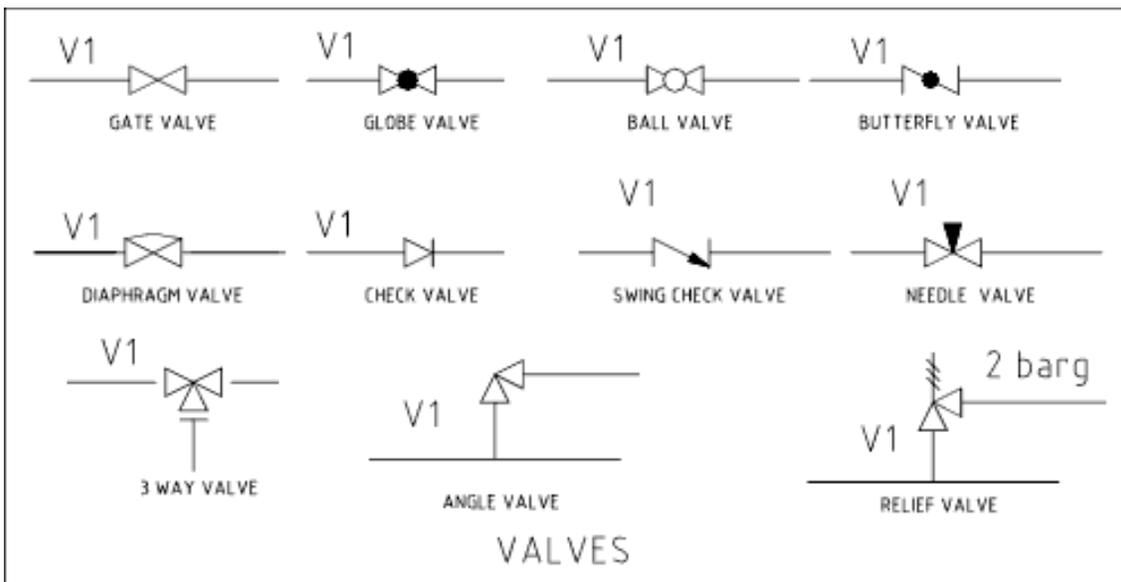
Table 5: Valve Actuator Symbols

VALVES ACTUATORS
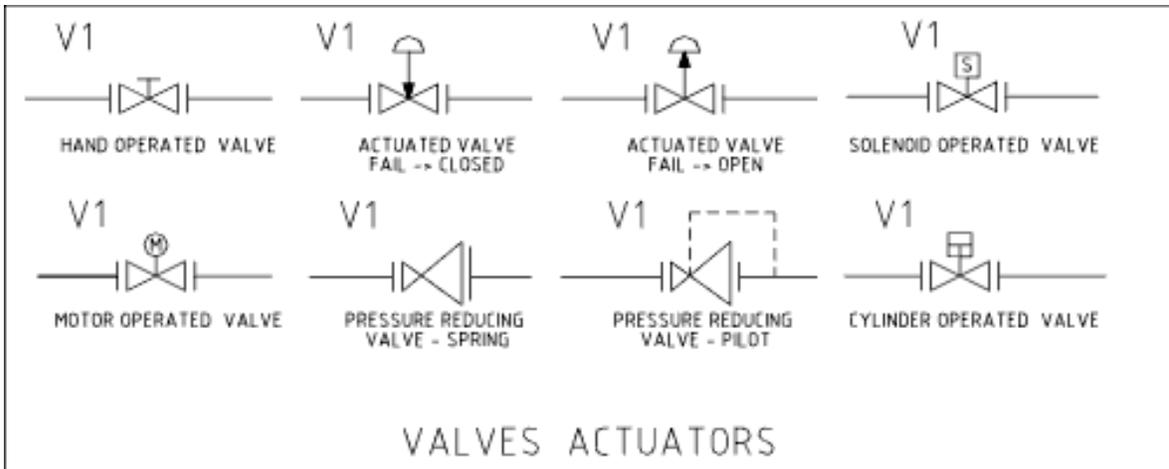
2.5 General Instrument or Function Symbols

Instruments can have various locations, accessibilities, and functionalities in the field for certain processes. It is important to describe this clearly in a P&ID. Below is a table of these symbols commonly used in P&IDs.

General instrument or function symbols			
	Primary location accessible to operator	Field mounted	Auxiliary location accessible to operator
Discrete instruments	1 ⊖	2 ○	3 ⊖
Shared display, shared control	4	5	6
Computer function	7	8	9
Programmable logic control	10	11	12

1. Symbol size may vary according to the user's needs and the type of document.
2. Abbreviations of the user's choice may be used when necessary to specify location.
3. Inaccessible (behind the panel) devices may be depicted using the same symbol but with a dashed horizontal bar.

Source: Control Engineering with data from ISA S5.1 standard

Discrete instruments are instruments separate or detached from other instruments in a process. Shared display, shared control instruments share functions with other instruments. Instruments that are controlled by computers are under the "computer function" category. Instruments that compute, relay, or convert information from data gathered from other instruments are under the "Programmable logic control" section.

For example, a discrete instrument for a certain process measures the flow through a pipe. The discrete instrument, a flow transmitter, transmits the flow to a shared display shared control instrument that indicates the flow to the operator. A computer function instrument would tell the valve to close or open depending on the flow. An instrument under the "Programmable logic control" category would control the valve in the field if it was pneumatically controlled, for instance. The instrument would gather information from discrete instruments measuring the position of the actuator on the valve, and would then adjust the valve accordingly.

In the chart above, it is necessary to know where the instrument is located and its function in order to draw it correctly on a P&ID. A primary instrument is an instrument

366

that functions by itself and doesn't depend on another instrument. A field mounted instrument is an instrument that is physically in the field, or the plant. Field mounted instruments are not accessible to an operator in a control room. An auxiliary instrument is an instrument that aids another primary or auxiliary instrument. Primary and auxiliary instruments are accessible to operators in a control room.

2.6 Transmitter Symbols

Transmitters play an important role in P&IDs by allowing the control objectives to be accomplished in a process. The following are commonly used symbols to represent transmitters.

Below are three examples of flow transmitters. The first is using an orifice meter, the second is using a turbine meter, and the third is using an undefined type of meter.

Table 6: Transmitter Symbols
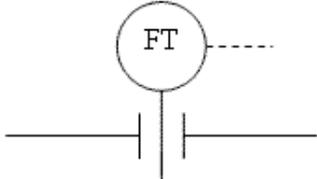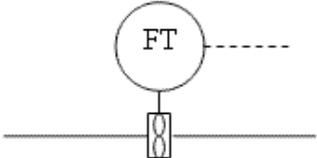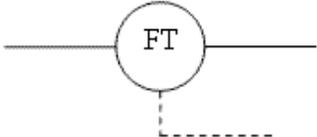
Flow Transmitters		
Orifice Meter	Turbine Meter	Undefined Meter

The location of the transmitter depends on the application. The level transmitter in a storage tank is a good example. For instance, if a company is interested in when a tank is full, it would be important for the level transmitter to be placed at the top of the tank rather than the middle. If the transmitter was misplaced in the middle because a P&ID was misinterpreted then the tank would not be properly filled. If it is necessary for the transmitter to be in a specific location, then it will be clearly labeled.

2.7 Miscellaneous Symbols

The following symbols are used to represent other miscellaneous pieces of process and piping equipment.

Table 7: Process Equipment

VESSEL/TANK

PUMP
FLOW SHEET

PUMP
LINE DIAGRAM

CW

HEAT EXCHANGER
FLOW SHEET

HEAT EXCHANGER
LINE DIAGRAM

FAN

EQUIPMENT PROCESS PLANT -1

BREAK VESSEL
WITH DEMISTER RING

TRAY COLUMN
BASIC

FLUID CONTACTING COLUMN
BASIC   DETAILED

REACTION OR ABSORBING VESSEL
BASIC   DETAILED

COOLING TOWER

FUME EXTRACTION HOOD

OPEN TANK

OPEN TANK + COVER

CLOSED TANK

SEALED TANK

EQUIPMENT- PROCESS PLANT-2

368

EQUIPMENT PROCESS PLANT -3

Table 8: Line Fittings

LINE FITTINGS

Table 9: Pipe Supports
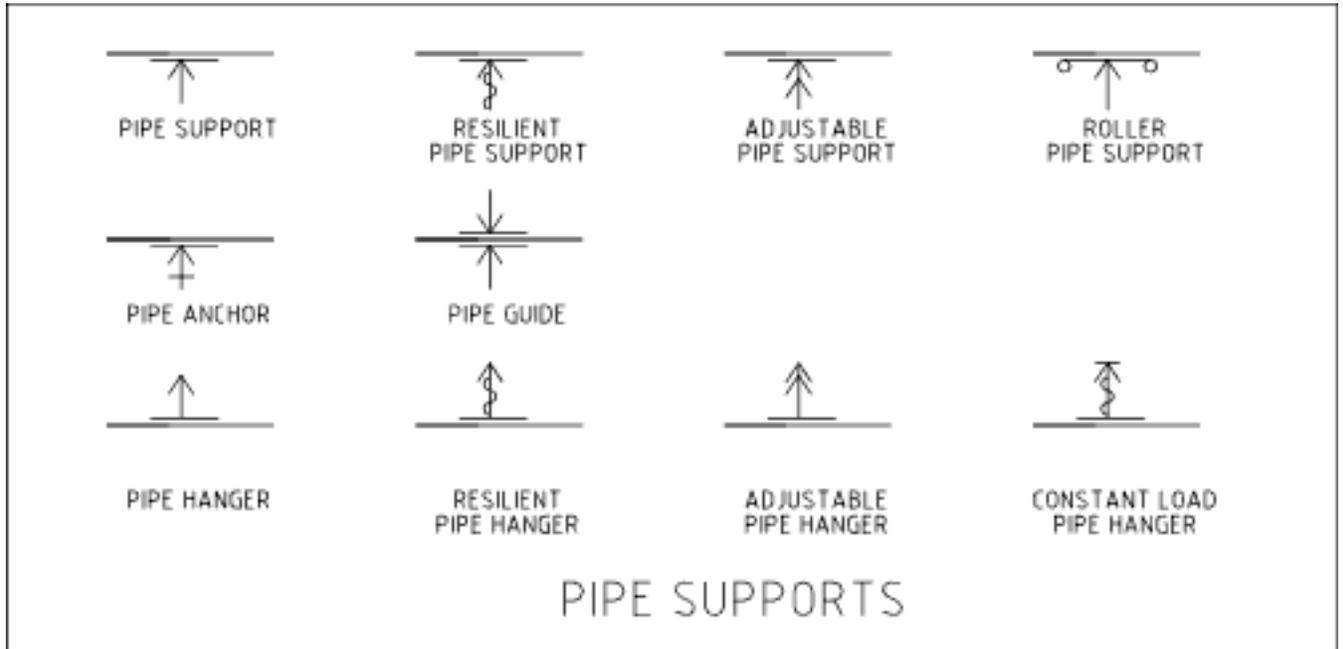
2.8 Crafting a P&ID

In order to greatly simplify P&ID diagrams for the purposes of this class, a standard convention must be employed. This convention simplifies the many control devices that need to be used. For the sake of brevity, sensors, transmitters, indicators, and controllers will all be labeled on a P&ID as a controller. The type of controller specified (i.e. temperature or level) will depend on the variable one wished to control and not on the action needed to control it.

For instance, consider if one must control the temperature of fluid leaving a heat exchanger by changing the flow rate of cooling water. The actual variable to be controlled in this case is temperature, and the action taken to control this variable is changing a flow rate. In this case, a temperature controller will be represented schematically on the P&ID, not a flow controller. Adding this temperature controller to the P&ID also assumes that there is a temperature sensor, transmitter, and indicator also included in the process.
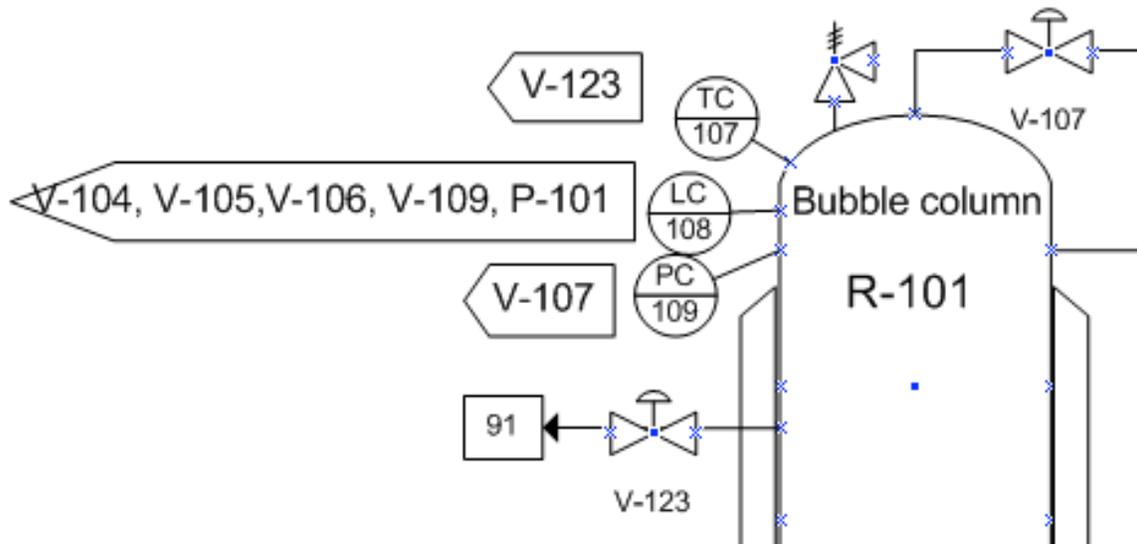
As you can see on the P&ID above, these controllers are represented as circles. Furthermore, each controller is defined by what it controls, which is listed within arrow boxes next to each controller. This simplifies the P&ID by allowing everyone the ability to interpret what each controller affects. Such P&IDs can be constructed in Microsoft Office Visio.

2.9 Sample Diagram

Below is a sample P&ID Diagram that is actually used in an industrial application. It is clearly more complicated than what has been detailed above, however, the symbols used throughout remain the same.

371

Table 10: Sample P&ID Diagram
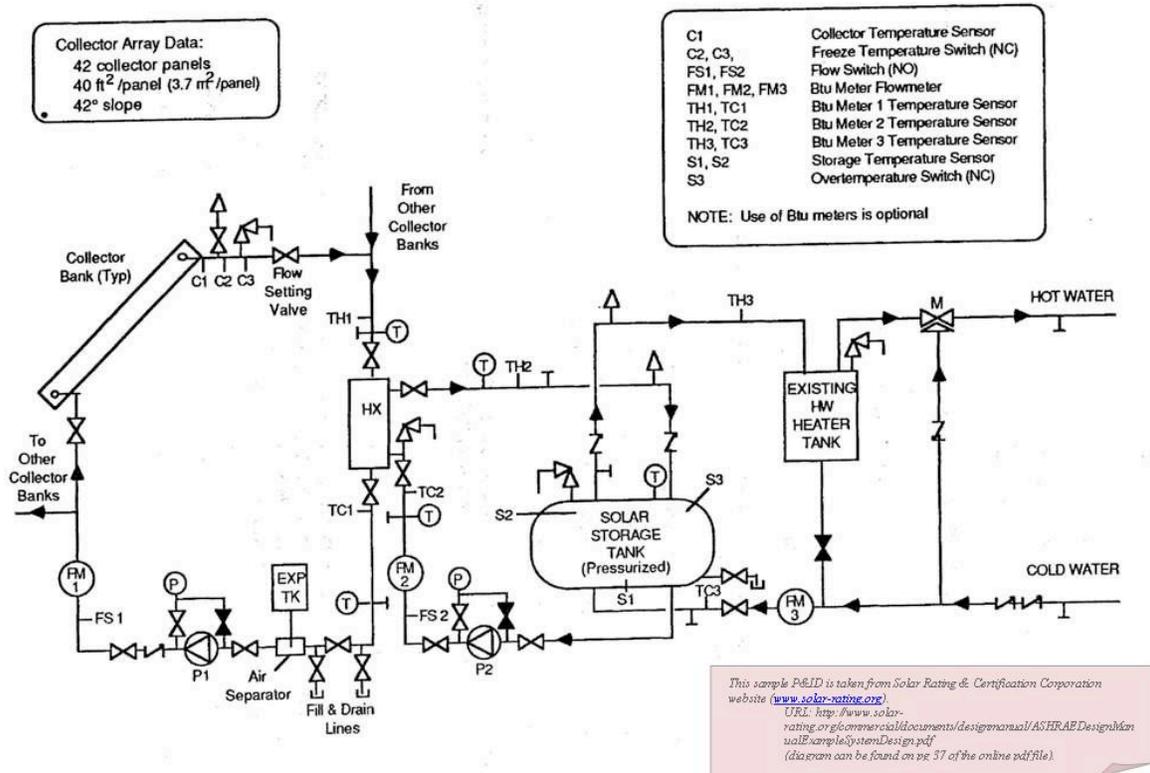
Figure B. Example System P&ID.

2.10 Example 1

Describe the following controlled process in words:
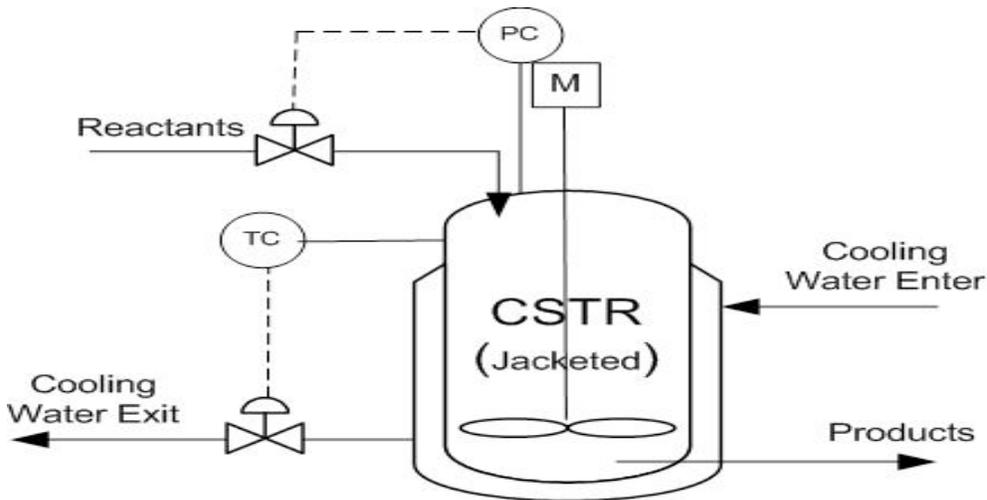
Answer: Reactants enter a jacketed CSTR where a reaction takes place and the products exit. The reactor is cooled via a coolant water stream. The temperature inside the reactor vessel is monitored with a temperature controller (also contained in the controller is a sensor, indicator, and transmitter), which electrically controls a valve. The valve can alter the flowrate of the coolant water stream, thereby controlling the temperature inside the reactor. A pressure controller is also present which feeds back to an inlet valve. Therefore, we can deduce that this reaction is most likely gas phase and if the CSTR becomes too full (high pressure) the inlet valve will close.

2.11 Example 2

Draw a proper P&ID diagram of the following process:

A storage tank is filled with condensed products formed via the CSTR in Example 1. The tank contains a level controller at a set point on the top of the tank. If this tank were to fill, materials would get clogged up in the reactor. Therefore, if the tank reaches 90% of its total capacity, the level controller will send an electric signal, which opens an emergency drainage line located at the bottom of the tank. The level controller will also activate an alarm alerting plant engineers that there is a problem with the storage tank. Finally, the level controller will also close the inlet valve to the storage tank.
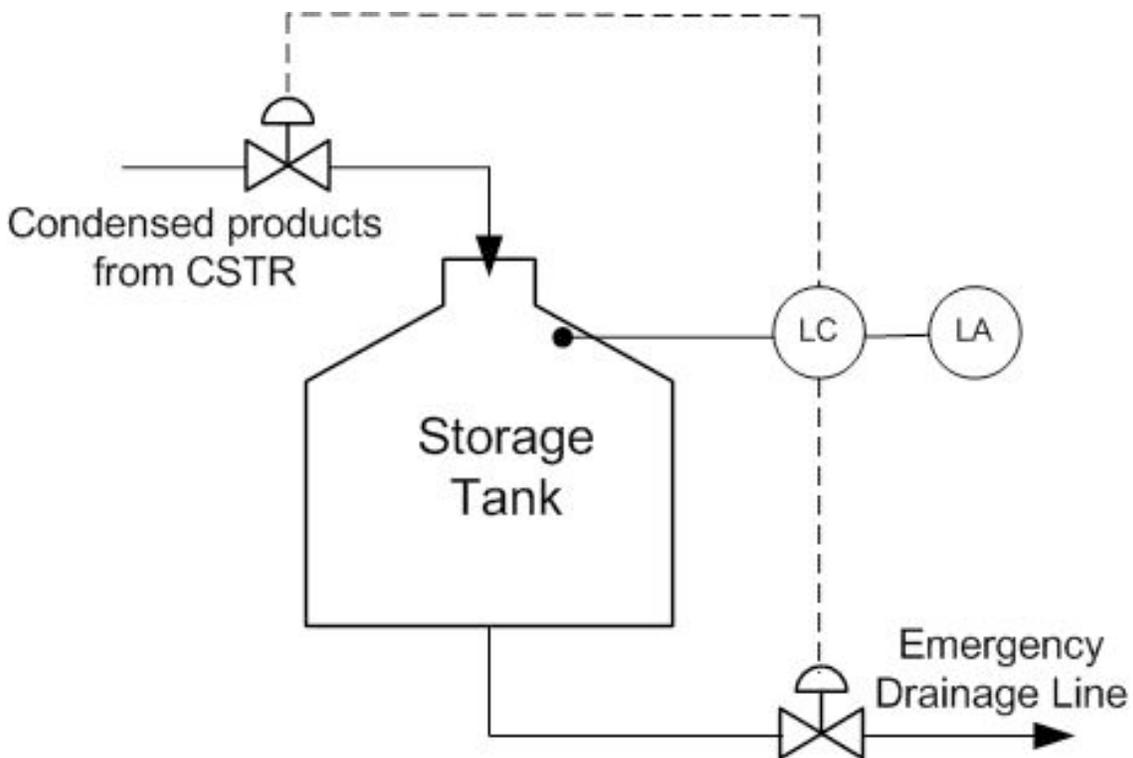
2.12 Example 3

Below is a P&ID diagram of the transesterification process to produce biodiesel. Soybean oil, methanol, and the sodium methoxide catalyst are pumped in to the reactor. The temperature of the reactor is regulated by the circulation water. The resulting biodiesel is then pumped out of the reactor and goes on to other processes so that it can be sold. Below is a P&ID of the process that is missing the valves, pumps, and sensors. Add the pumps, sensors, and valves that are needed to successfully control the process.
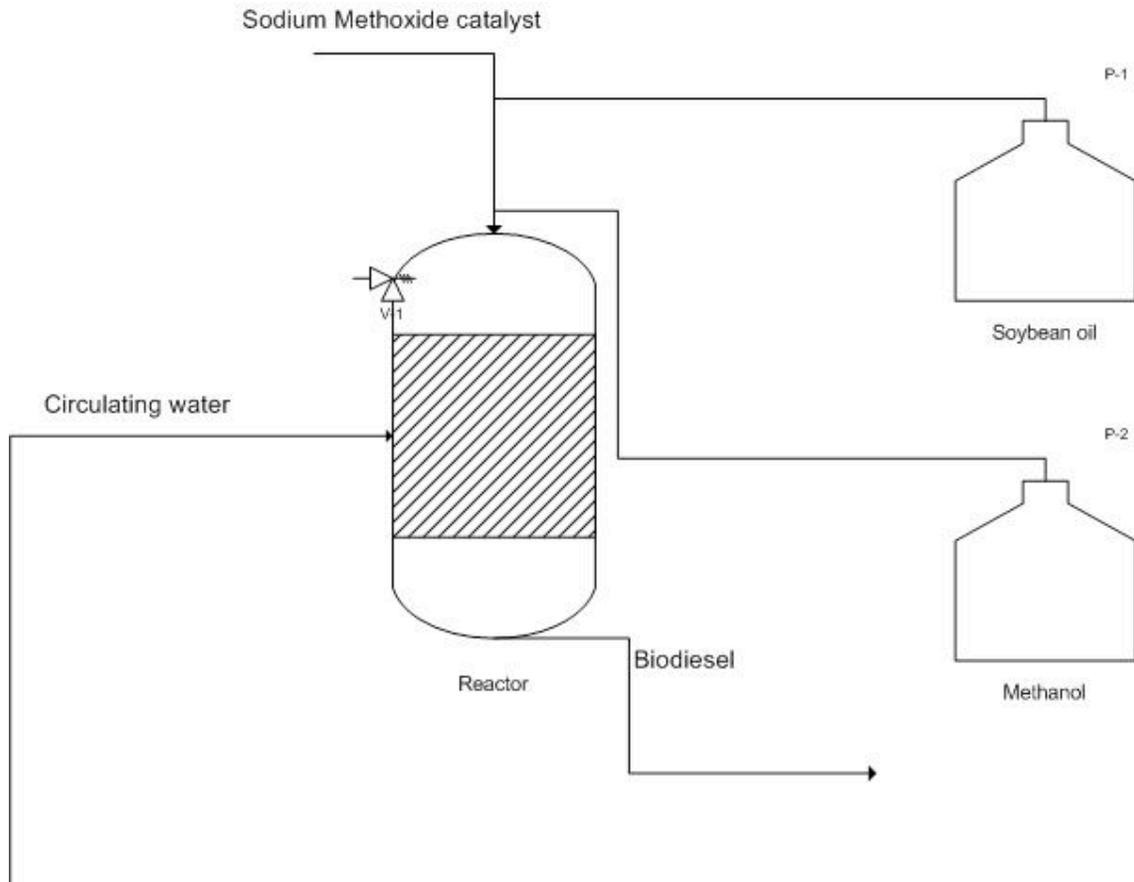
Solution:
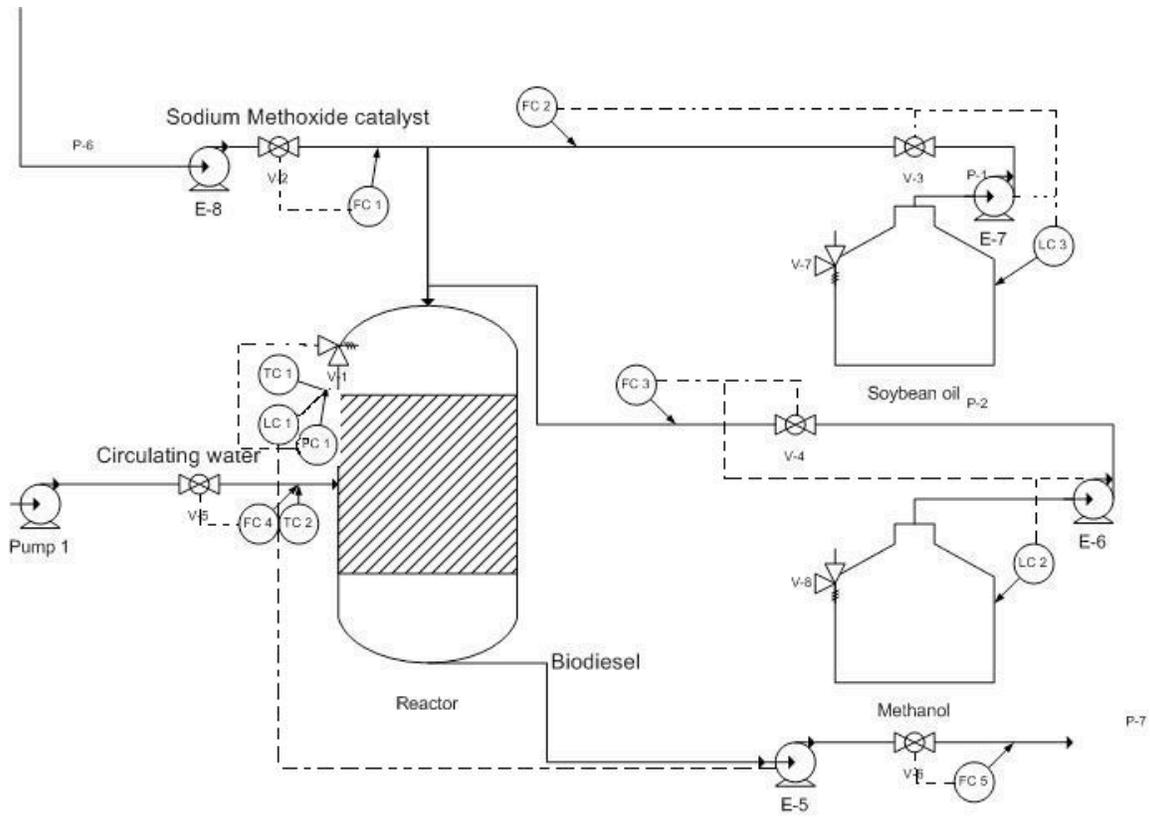
2.13 Example 4

Below is a example problem of a typical P&ID problem. A is a liquid at Tamp but boils at Trx. B and P are high boiling point liquids and C is a solid. The reaction for the process is 2A+B+C-->P at Trx. Ais fed in excess.

ChE 466
F09
HW2
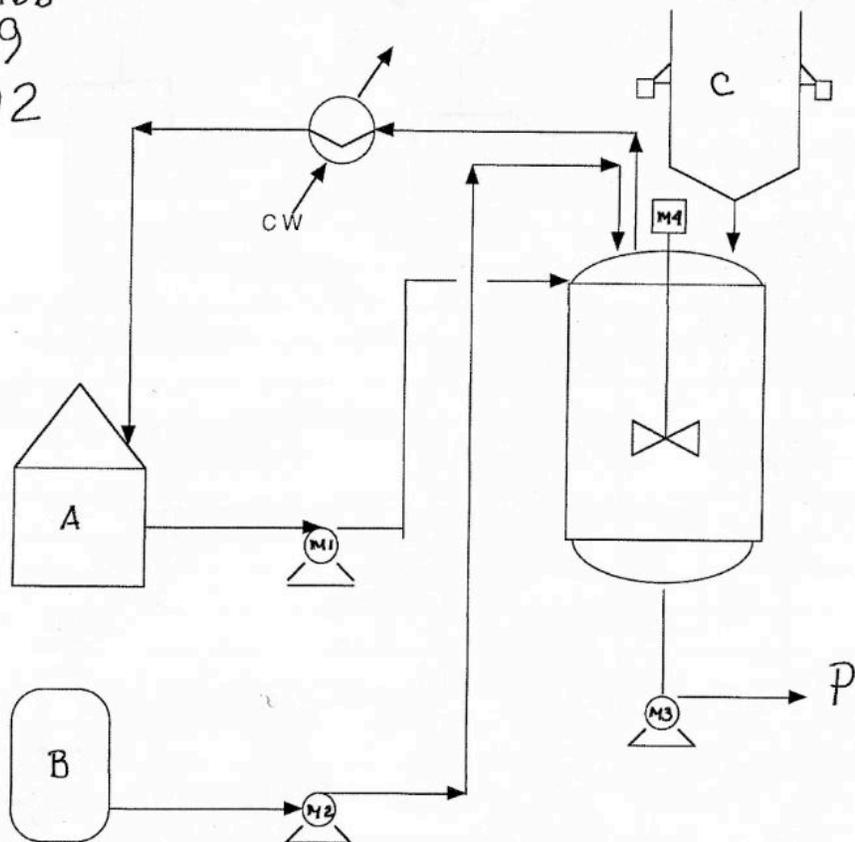
A is a liquid at T_{amp} but boils at T_{Rx}
B and P are high boiling liquids
C is a solid

$$2A + B + C \rightarrow P \ @ \ T_{Rx}$$
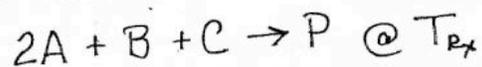
A is fed in excess

1) Instrument the process ie add ◯ as needed
2) Show control devices ie ⋈ or [M]
3) Show "what" controllers control ie ⟨M_1, M_2⟩

Below is the solution to the problem above.

376
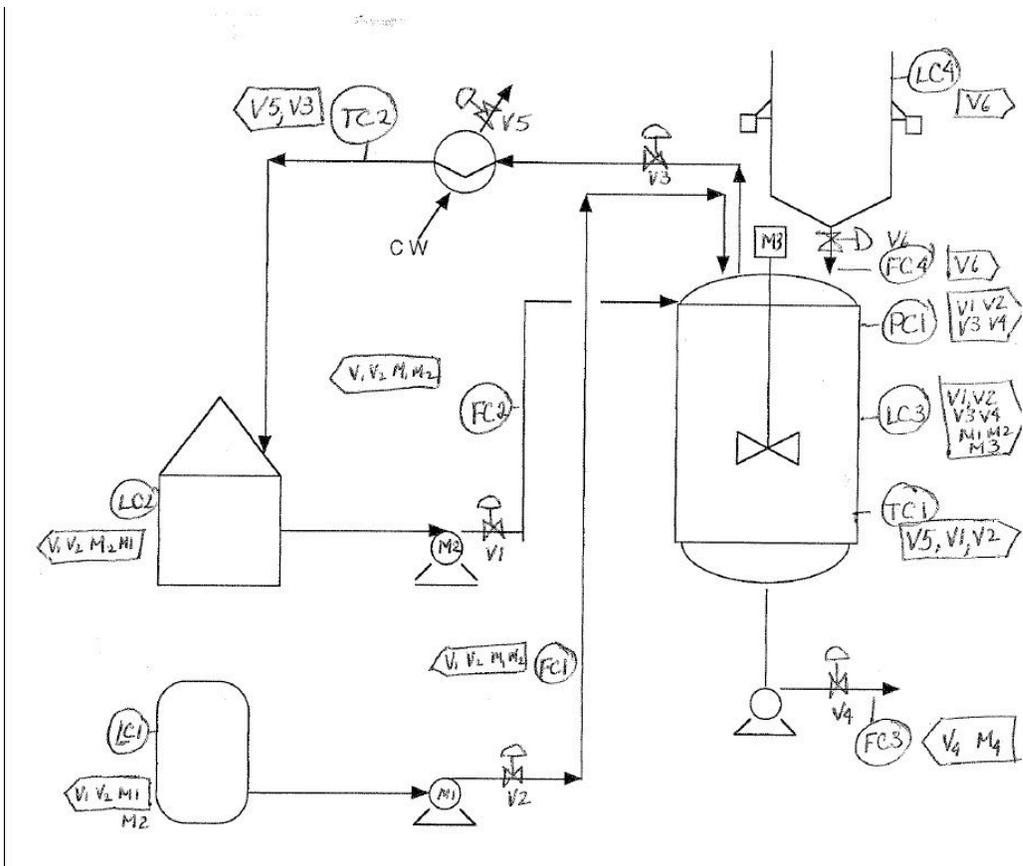
ChE 466   HW2

2.14 Sage's Corner

A Brief Guide To Piping and Instrumentation Diagram Standard Notation

```
http://video.google.com/googleplayer.swf?docId=6205940157432698049
```

slides for this talk

2.15 Additional Reading

1. Wikipedia ChemE P&ID symbols entry
2. Wikipedia P&ID entry
3. Control Engineering Online Article: how to read P&IDs

Here are templates for creating P&IDs in Visio (PCs) and OmniGraffle (Mac).

2.16 References

- Mineral Engineering Technical Services; "Process Engineering.". Last Updated 05/09/2007.
- Solar Rating and Certification Corporation; "Example System Design.". Last Updated 05/09/2007.
- Karim, Nazmul M.; Riggs, James B. "Chemical and Bio-Process Control." 3rd Edition. Ferret Publishing.
- Ogunnaike, Babatunde A.; Ray, W. Harmon. "Process Dynamics, Modeling, and Control." New York Oxford: Oxford University Press, 1994.
- Beardmore, Roy; "Flow Sheet Symbols.". Last Updated 12/02/2006.

Section 3. P&ID standard structures, location of features

Title: Piping and Instrumentation Diagrams: Location of Controls and Standard Control Structures
Note: Video lecture available for this section!
Authors: Christopher Bauman, Jennifer DeHeck, Merrick Miranda, Evan Leonard
Stewards: Robert Chockley, Chloe Sweet, Irene Brockman, Scott Dombrowski
Anchors: Megan Boekeloo, Stacy Young

Date Presented: September 21, 2006 / **Date Revised:** September 7, 2007

- First round reviews for this page
- Rebuttal for this page

3.1 Introduction

A Piping & Instrumentation Diagram (P&ID) is a schematic layout of a plant that displays the units to be used, the pipes connecting these units, and the sensors and control valves. Standard structures located on a P&ID include storage tanks, surge tanks, pumps, heat exchangers, reactors, and distillation columns. The latter three comprise most of the chemical process industry single-unit control problems.

P&IDs have a number of important uses in the design and successful operation of chemical process plants. Once a process flow diagram is made, P&IDs help engineers develop control strategies that ensure production targets are met while meeting all safety and environmental standards. These diagrams are also very important for locating valves and process components during maintenance and troubleshooting.

Each page of a P&ID should be easy to read and correspond to a specific action of the plant. Also, symbols used in the P&ID are uniform throughout. Diagrams for heat exchangers, continuously stirred tank reactors (CSTRs), and distillation columns shown in subsequent sections of this wiki represent a small sample of standard symbols for these structures. For a list and explanation of standard controls-related symbols, please refer to PID Standard Notation.

3.2 General Strategies for Implementation of Good Control Systems

There are several guidelines to follow when designing P&IDs for a plant. The first consideration is stable operation of the plant so that all safety and environmental standards are met. Maintaining product quality should also be a primary design objective. Additionally, systems should be designed to respond quickly to rapid changes in rate and product quality. Usually control systems should be run in an automatic mode so the system will correct itself, as opposed to a manual mode, which requires operator supervision.

379

When setting up a control system, it is useful to first focus on mass balance control, which can be monitored by level and pressure control loops that use gas or liquid flow rates. Additionally, a product control structure should be set up to ensure efficient process operation. To prevent controls from conflicting with one another, a final examination of the controlled streams should be performed. Further instrumentation can then be installed to prevent conflicts between controllers. Finally, appropriate tolerances should be established for controllers that directly affect the action of other controllers. Determination of the optimal control placement is essential for successful plant operation.

3.3 Standard Structures and Location of Control Features

As mentioned above, control of heat exchangers, reactors, and distillation columns represent the majority of single-unit control problems. The table below summarizes the control schemes derived in the following examples. The most common control schemes are listed in the table below.

Table 1: Summary of common P&ID structures

Structure	Control Variable	Manipulated Variable	Location of Control Valve	Comments	Typical Disturbances
Steam-heated Heat Exchanger	Temperature of outgoing stream	Steam pressure	Steam line or condensate stream	Placement of a control valve on the steam line gives tighter control, but this type of valve is bigger and more expensive than one placed on a condensate line	Fluctuation of the inlet flow rate
Liquid/Liquid Heat exchanger	Temperature of outgoing stream	Bypass flow rate	Bypass stream	This allows high coolant flow rates and decreases fouling. Changes in bypass stream's flow-rate result in temperature control loop disturbances that are easily regulated.	Fluctuations in inlet flow rates
Jacketed CSTR w/ Endothermic RXN	Product stream temperature	Steam flow rate or steam pressure in the jacket	Steam line	Changing steam flow rates allows direct measure of the amount of heat entering the system and therefore a measure of conversion. It also makes a reactor sensitive to heat load changes and enthalpy changes. Manipulating jacket pressure is better at absorbing reactor heat duty and steam enthalpy upsets but it does not directly measure heat load on the reactor.	Fluctuations in feed flow rate, temperature, and enthalpy of heating medium
Jacketed CSTR w/ Exothermic RXN	Inlet temperature of coolant	Coolant flow rate	Coolant feed stream	There is a slow response to fouling on heat transfer surfaces, but there is a fast response to changes in inlet coolant temperature and coolant supply pressure.	Changes in feed flow rate, temperature; Fluctuation of coolant enthalpy

For most equipment, a degrees of freedom analysis is first performed and then control scheme is designed based on the degrees of freedom.

3.3.1 Heat Exchanger

To monitor the performance of a heat exchanger, the product stream is important. Usually, the product stream must be within some temperature range before it continues to

downstream process units. The outlet temperature of this stream can be used to calculate the heat transfer. The steam is controlled in order to obtain the desired product stream temperature. One way to influence the product temperature is by controlling the flow of the heated steam. This flow-based control may take some time to implement and therefore cause fluctuations in the process. Depending on the process, these fluctuations may or may not be acceptable. Instead of monitoring the flow-rate, the steam pressure may be monitored, achieving tighter control of temperature. A change in pressure is much easier to monitor and correlates directly with a change in temperature of the steam. This offers an effective way to control process temperatures. Pressure control also enables the physical condition of the piping to be monitored, since pressure changes occur as fouling progresses. The pressure, combined with the flow-rate and temperature of the heated steam, can be used to calculate this fouling occurring inside a piece of equipment. It is important to remember that the steam should always be controlled on the inlet side of the heat exchanger for better pressure control and safety reasons, as shown in figure 1 below.
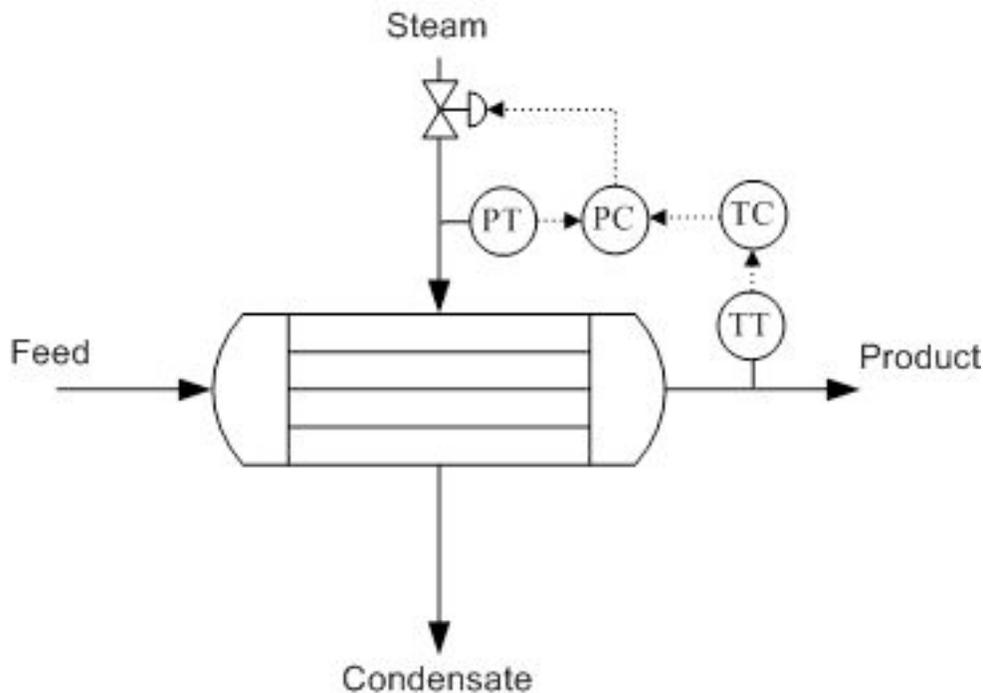
Figure 1. Heat Exchanger with pressure control on steam inlet and temperature control on the product stream.

3.3.2 Distillation Column

Because the economic viability of an overall process is based significantly on product purity, it is important that distillation columns maintain stable operation. Changes in composition and flow-rate of the feed stream are common disturbances in distillation column operation. Improper functioning of controllers can undermine the effectiveness of the product composition. A degrees of freedom analysis can help place sensors and

381

actuators in appropriate places, while not including too many sensors and actuators, in order to obtain an efficient control system.

Degrees of Freedom Analysis

A simple degrees of freedom analysis can make design of a control scheme easier as well as improve the control scheme overall. A process control approach to degrees of freedom, adapted from ECOSSE Module 3.1, requires tabulation of streams and extra phases.

DOF = Streams − ExtraPhases + 1

A typical distillation column contains the following streams; feed (1), bottoms (2), distillate (3), reflux (4), product (5), vapor-liquid mix (6), cooling water (7), and steam (8). There are also three locations where there are two phases, vapor and liquid, present in equilibrium. These are denoted as one "extra phase", since temperature and pressure are not independent in a two-phase system.

DOF = 8 − 3 + 1 = 6
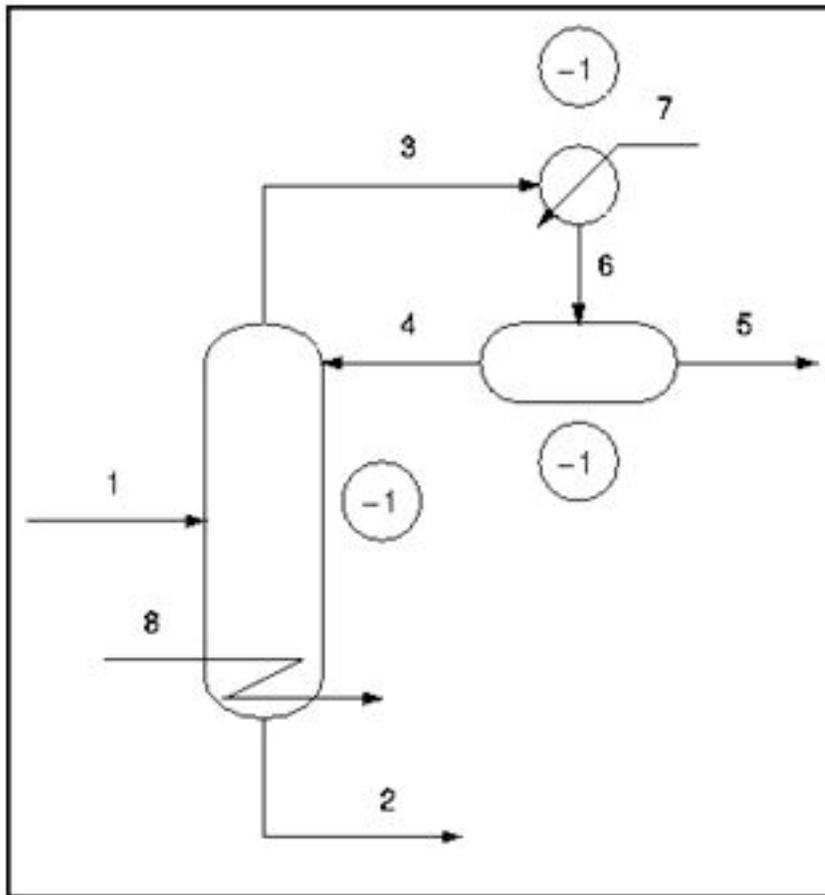
Figure 2. Degrees of freedom analysis on a typical distillation column (adapted from ECOSSE).

In typical practice, the condenser and reboiler pressure are specified as atmospheric pressure. Once the pressure is specified in the condenser and reboiler (both two-phase systems), the temperature is specified and therefore the degrees of freedom are reduced by two. The total degrees of freedom, in practice, is six.

$$DOF = 8 - 5 + 1 = 4$$

A simple degrees of freedom analysis, in any system, can help define the variables and reveal where the critical sensors for process control should be located. This type of analysis also minimizes the design of too many sensors, actuators, and valves, which may in theory control and regulate the same variable.

Sensor and Actuator Options

Product streams, reboiler steam, and the reflux stream should have flow sensor/transmitters so that each flow can be adjusted to meet the column's control objectives. A flow sensor should also be placed on the feed stream because disturbance to the feed flow rate are common. Each flow sensor should be connected to a flow valve to control the corresponding stream.

Differential pressure level sensor/transmitters should also be used for the accumulator (the vessel that collects condensed distillate) and bottom of the column because maintaining these levels is essential for reliable operation of the column. If flooding is an issue, a pressure differential across the column should indicate the onset of flooding. A pressure indicator should be installed at the top of the accumulator to monitor column pressure. Temperature in a distillation column is typically controlled by manipulating steam flow to the reboiler.

Control Schemes

There are a number of common control schemes for distillation columns. Optimally, a distillation column should be run with *dual composition* control because it saves energy. In dual compostion control, the temperature of both chemicals in a binary distillation is controlled. The system is more complex to setup and measurements required for control may be difficult. As a result, many distillation columns use single composition control instead. Common control configurations for distillation columns include reflux-boilup and distillate-boilup.

In the reflux-boilup configuration, the distillate composition and bottoms composition are the control variables. The reflux flow and the heat input control (vapor boilup) are the manipulated variables, which allow control of the liquid and vapor flow-rates in the column. With this control system, a quick response to changes in the feed composition is possible. In the distillate-boilup control configuration, the distillate flow and the vapor

boilup are used to control composition. This configuration is a better choice for columns where the reflux ratio is high.
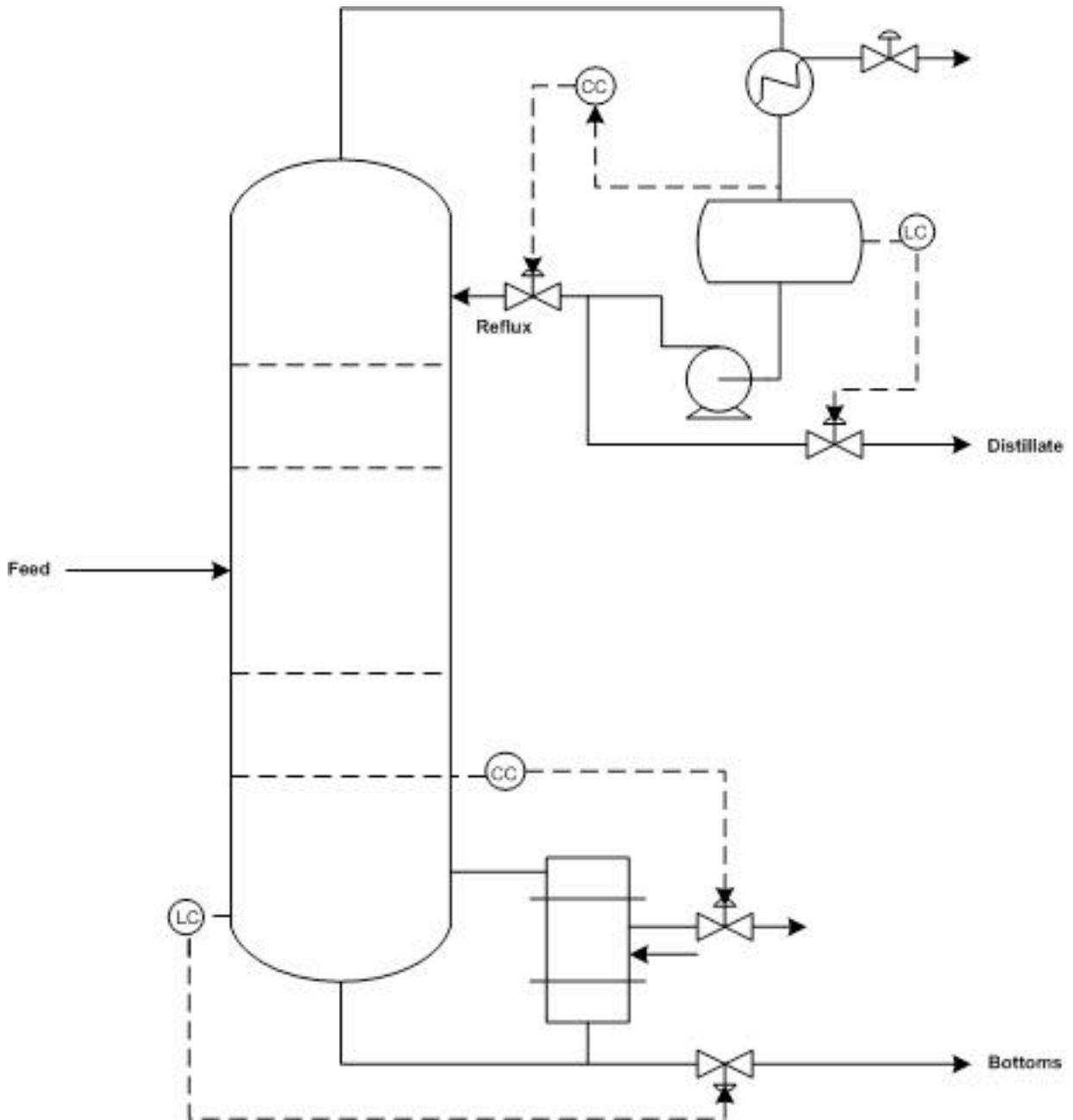
Figure 3. Distillation column with reflux-boilup control scheme (adapted from Luyben).

3.3.3 Reactor (Exothermic Reaction in CSTR)

In implementing controls in reactive systems, temperature is a good indicator for unit performance. Temperature is often related to reaction rate and varies with time in most

reactors. However, a reactor exit stream with a constant temperature is often desired downstream in the process.

In an exothermic reaction in a CSTR, lower initial temperatures result in lower reaction rates and low heat generation. As the reaction progresses, heat generation increases rapidly due to higher reaction rates and high concentrations of reactants. As reactant concentration decreases, the heat generation once again becomes low. Exothermic reactor temperatures must be controlled to assure stable reactor operation.

To remove heat from an exothermic reaction, basic heat transfer principles are employed. A coolant is pumped through a shell outside of the reactor. Since the heat removal is linear, the temperature of the coolant should be controlled. In doing this, it is possible to increase the driving force for heat transfer to slow a reaction, or conversely, to allow the reaction to further progress by decreasing the driving force for heat transfer.

Degrees of Freedom Analysis

In order to determine where to place controls, sensors, and valves in a exothermic reaction operation, a degrees of freedom analysis, similar to that carried out in the distillation section 3.2, can be helpful.

A typical CSTR contains the following streams; reactant A (1), reactant B (2), product (3), and coolant to the jacket (4). There are no locations where there are two phases, vapor and liquid, present in equilibrium - assuming the reaction is liquid phase with no simultaneous evaporation or sublimation.

$$DOF = Streams - ExtraPhases + 1$$

$$DOF = 4 - 0 + 1 = 5$$
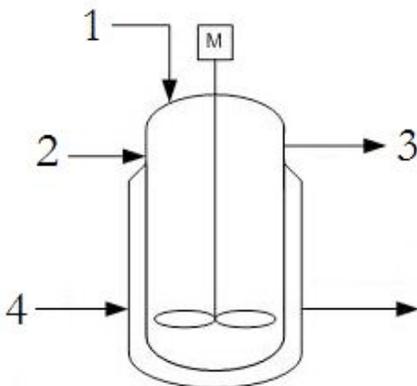
Figure 4. Degrees of freedom analysis on a typical exothermic reaction in a CSTR.

Since in most practical applications the reactants (1 and 2) as well as the product (3) flow rates are defined by demand for the product, there are only 2 degrees of freedom which allow for placement of a 2 valves or controllers on the coolant stream (4) as well as one of the reactant streams (1 or 2). Only one of the reactant streams need to be specified as the other can be determined by a ratio controller using stoichiometric coefficients.

Control Schemes

As discussed previously, a valve can be placed on the coolant inlet stream to ensure proper temperature control of the reactor. By controlling the coolant stream based on inlet conditions of the reactant streams, the control can respond quickly. But, if the coolant is controlled based on outlet conditions of the product stream, there is a lag in response, but it is easy to monitor heat transfer performance.
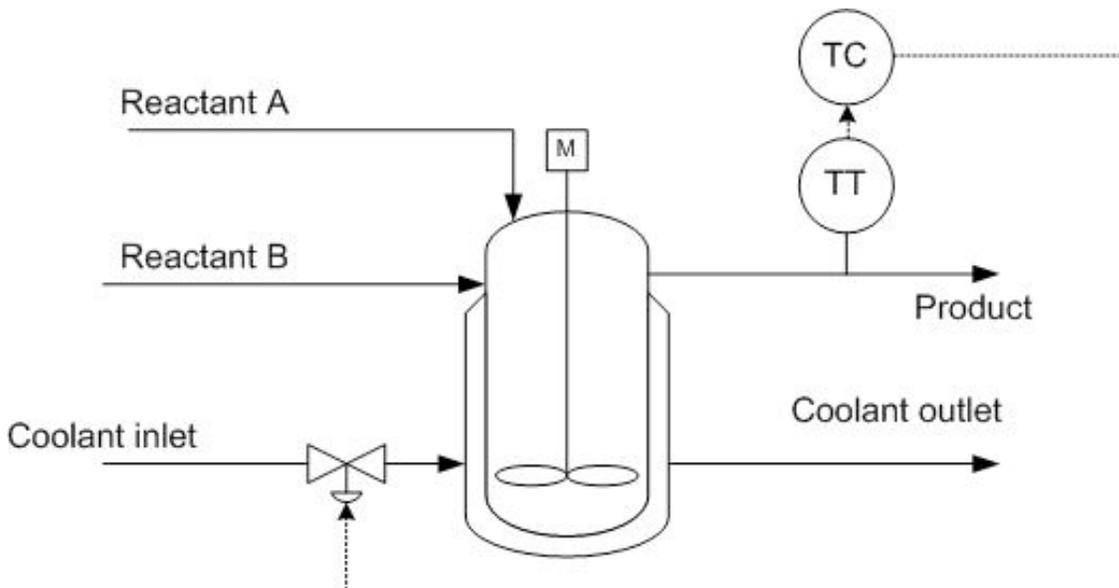
Figure 5. Jacketed CSTR with coolant control based on the outlet temperature conditions.

To translate this example to an endothermic reaction, the coolant flow would simply be translated to steam flow in order to provide heat to the reaction through the CSTR jacket.

3.3.4 Other Common Process Equipment

Beyond the heat exchangers, reactors, and distillation columns, many other pieces of process equipment, including furnaces, compressors, decanters, refrigerators, liquid-liquid extractors (LLEs), and evaporators, are subject to disturbances and require careful control.

Furnaces

For example, furnaces may be subject to frequent load changes as a process or customer requires more energy. To cope with these demands, the temperature of the outlet stream must be monitored and manipulated. Information from a temperature controller at the outlet stream can be used to effect changes in valves that control the flow-rate of air and fuel to the furnace. At this point, the best setup of the control system must be considered in light of factors such as safety, lag time for changes to occur, and equipment wear. For a furnace, the controls should generally be set up so that an excess of air is more likely than an excess of fuel. This reduces the risk of explosion due to the furnace filling with uncombusted fuel.

Liquid-Liquid Extractors

In liquid-liquid extractors, the interface level and pressure are the controlled variables. Disturbances in flow rate of the entering stream can affect interface level and prevent complete separation of the heavy and light components. From this, it is obvious that there should be valve controls on both exit streams. The best control scheme depends on the operation of the process. When the heavy phase is continuous (light phase flows upward through heavy phase), changes in interface level should be controlled by adjusting the flow-rate of the light product, while the pressure is controlled by adjusting the flow-rate of the heavy product out of the column. Figure A is a representation of what occurs in a single stage extractor. Generally, single stage extractors are using in chemical labs, where as multistage extractors are used in industry. A multistage extractor uses the immiscible liquid stream from the previous stage as the feed in the following stage. Figure 6 depicts the control scheme previously described.
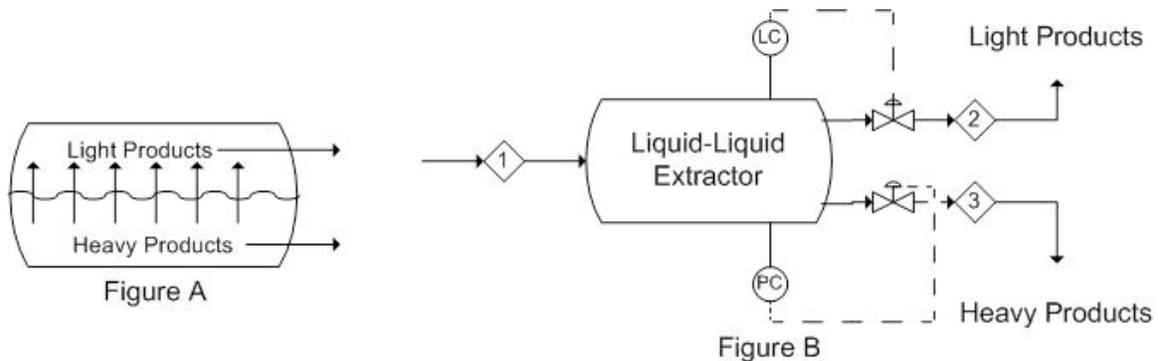
Figure 6. Liquid-liquid extractor control scheme with interface level and pressure control.

When the light phase is continuous, the control system must be set up in the opposite manner. Figure C is again a depiction of what occurs in a single stage extractor. Figure D is a representation of a control scheme that could be implemented. This is the reverse of the control scheme in Figure 6.
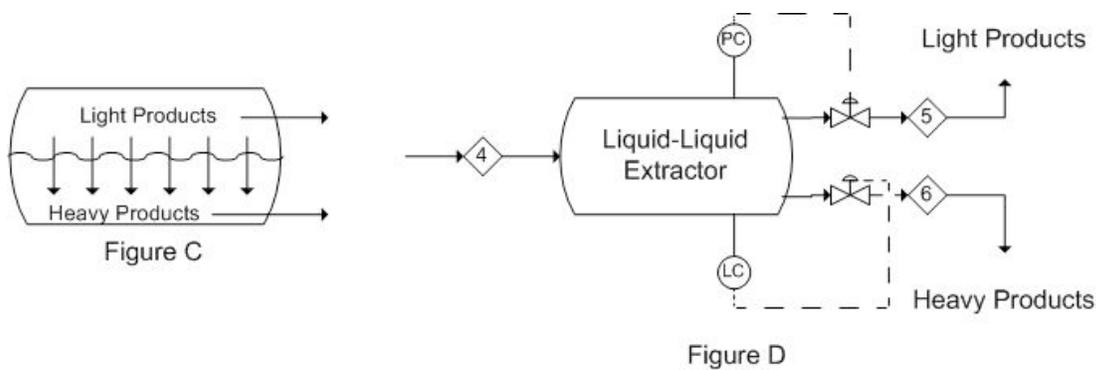
Figure C

Figure D

Figure 7. Liquid-liquid extractor control scheme with light-phase control.

These representations are only two possibilities for basic control schemes that can be implemented in a process as there are several controllers and aspects of the specific processes that may need to be controlled or monitored.

Compressors

Compressors are another valuable component in process design. They allow for the reduction of the volume of an incoming stream by increasing the pressure at which the stream is maintained. They can also be used to pump liquids through the process, as liquids are highly incompressible compressors cannot be used for volume reduction. For this, there must be a specific control system as to prevent adverse effects due to extremely high or low pressures. There are several types of compressors, including: dynamic, axial, and rotary [1] to name a few. Because the increase in pressure is governed by the ideal gas law, there is most often and increase in temperature as well. This can be left as is, or sent to a heat exchanger for temperature reduction. Heat exchangers were discussed above.

One such example is the use of a centrifugal compressor to reduce the volume of a fuel stream for storage. Using a compressor will allow for volume reduction as gasses are easily compressed, this is also economically friendly as it reduces the size of tank necessary to store the fuel stream. The tank should also be equipped with a pressure reducing valve, to bring the stream back to a desired pressure, depending on the process. A diagram of this scheme is as follows:

388
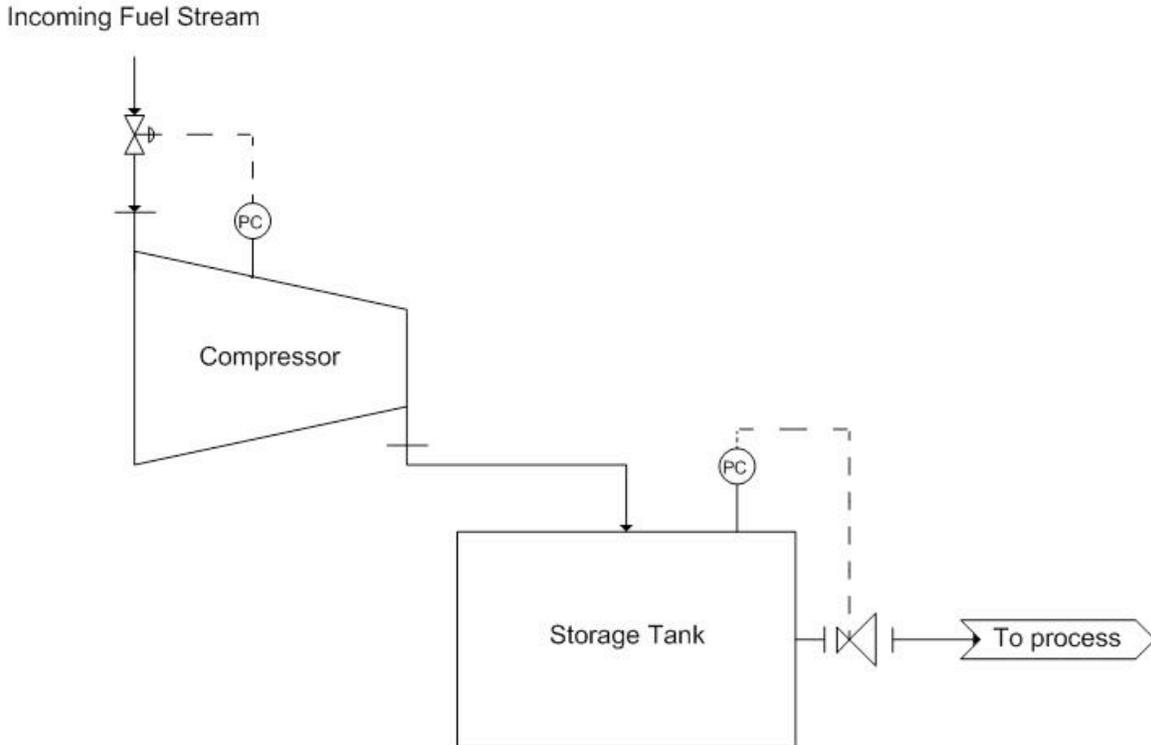
Figure 8. Compressor control scheme with pressure control.

The pressure controller on the compressor controls the valve on the incoming fuel stream. This ensures that if there is a build up in pressure, the flow into the system will be stopped in time. Also, a pressure controller should be placed on the storage tank. This is controlled by the pressure reducing valve mentioned earlier.

Decanters

Decanters, much like Liquid-liquid extractors, use solubility as their principle of separation. Unlike Liquid-liquid extraction, these require some time for the separation to occur. Generally, the separation is a liquid-liquid or liquid-solid separation. Decanters are widely used in the wine industry to separate out sediment from the wine. Utilizing the wine separation example, a possible control scheme is as follows. Here, there is only a level sensor on the decanter as it is a liquid-solid separation. Note that there is a release stream, used to remove collected sediment. An analytical or pH sensor could also be used here to maintain the correct chemistry or pH, as wine is the final product. It is also important to note that the exact placement of the level sensor would vary depending on the desired level in the tank. Here, the level sensor is not shown in its exact placement.
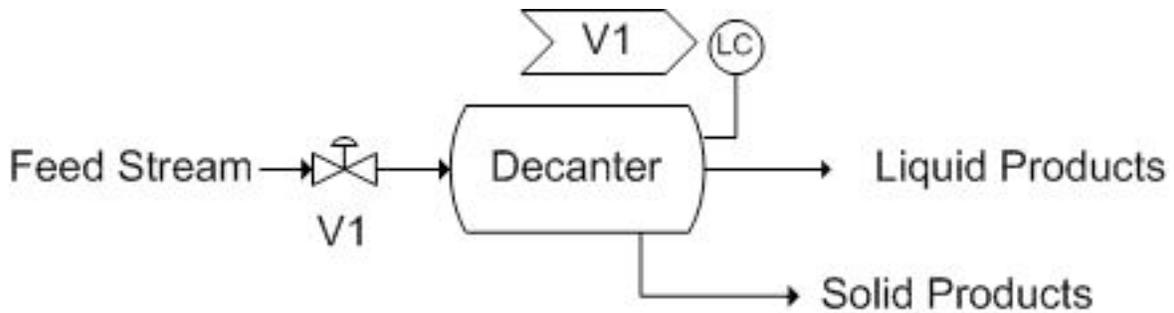
Figure 9. Decanter control scheme with level control and pH control.

Another vessel that is very similar to decanters are knockout drums. These vessels are generally located after heat exchangers or other pieces of equipment that result in a multiphase system. These vessels are used to separate the two phases, generally gas-and liquid separation. A possible control scheme is depicted below. The incoming stream is a liquid-gas mixture coming from a heat exchanger. Thus, there is a pressure sensor on the knockout drum. A level controller could be used, but this is effectively measuring the same thing, so it has been omitted. Also, because it is coming from a heat exchanger, a temperature controller has been included to control the amount of cooling taking place in the heat exchanger.

The pressure controller (PC) controls V2, the stream into the tank. The temperature controller (TC) controls V1, the valve on the coolant stream to the heat exchanger.
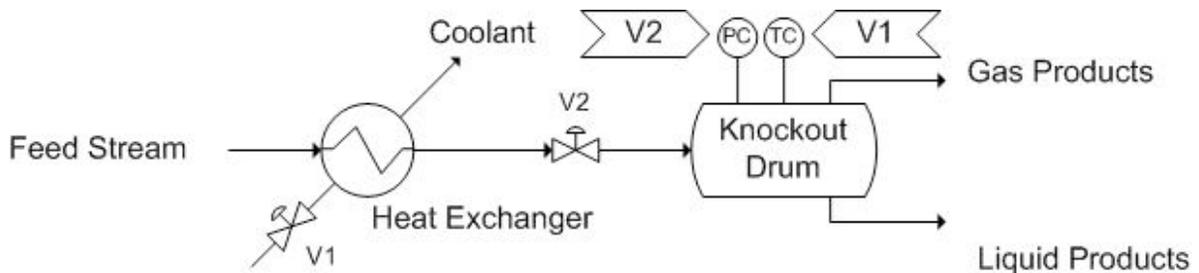
Figure 10. Knock-out drum control scheme with pressure and temperature control on different streams.

These examples illustrate the typical method of locating control systems for process equipment. After choosing the location for valves based on process constraints, there still remain a number of possibilities for the actual manner in which to control the valves. The lag-time for changes may be longer for certain configurations, while the danger of damaging equipment is smaller. The controls configuration will depend strongly on the safety concerns associated with a specific process.

390

The following steps should be followed when setting up controls for multi-unit processes. For more in-depth descriptions, refer to the [Step-by-Step Method For Describing Controls and Their Purpose](#).

1. Determine process objectives, taking into consideration product specifications, economic constraints, environmental and safety regulations, etc.

2. Identify boundaries for normal operation. These can be based on equipment limitations, safety concerns, environmental regulations, and the economic objectives of the processes.

3. Identify units and streams in the process that are susceptible to significant disturbances. These disturbances commonly occur in feed streams, product streams, and reactor vessels, but can be present anywhere that temperature or pressure or other variables are changing.

4. Select the types and locations of sensors in order to properly measure and monitor critical process variables.

5. Determine the appropriate types and locations for control valves in order to appropriately adjust process variables so that they remain within the normal operating boundaries. Controls should be set up to minimize response time between sensing a change and taking corrective actions. The ideal location for any given control depends on the process unit or units that it affects.

6. Perform a degree of freedom analysis.

7. Energy Considerations. An energy balance should be performed for the process. This step involves transporting energy to and from process units. This may include removing heat generated by a reactor and using it elsewhere in the process. Control valves will help regulate the flow of such streams.

8. Control Process Production Rate and Other Operating Parameters. Adjusting process inputs, such as reactant feed rates, can alter other variables in the process. Process controls must be able to respond to these adjustments to keep the system within operating boundaries.

9. Set up control system to handle disturbances and minimize their effects. (See Chemical Process Controls > PID Control)

10. Monitor Component Balances. Accumulation of materials within a system is not desirable and can lead to inefficiency in the process or catastrophic failure.

11. Control individual unit operations. Each unit of a multi-unit process needs to be individually controlled in order for control of the entire system to be possible.

12. Optimize the process. If the system has degrees of freedom, process variables can be manipulated in order to more efficiently or economically create product.

3.5 Example 1: Completing a Control System

Consider the following reaction in a semi-batch reactor: (CH3CO)2O + H2O -> 2CH3COOH
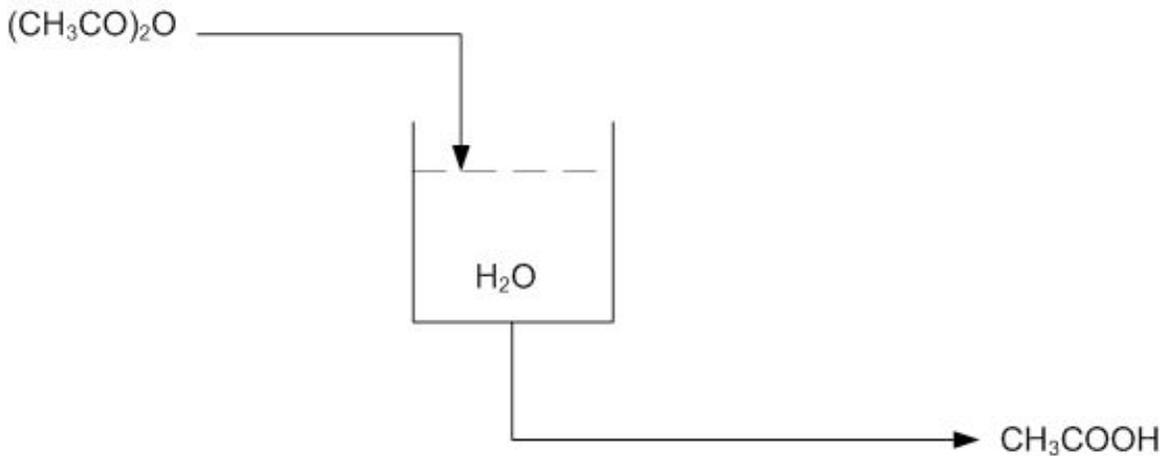
Figure E-1. Basic Flow Diagram of reaction of Acetic Anhydride with water to produce Acetic Acid.

The objective is to produce CH3COOH at a given concentration, i.e. 15-20%. The tank should also be drained once the given concentration is achieved and the reaction has gone to completion. The tank can only hold a specified volume and 80% is filled with water. The reaction should be run in excess water. Once a given amount of acetic anhydride has entered the reactor, it should be left in the tank until the reaction has gone to completion. The completion of the reaction can be determined by temperature, as the reaction is exothermic.

Using the multi-unit controller design from above, determine the location of needed sensors and valves on the flow diagram.
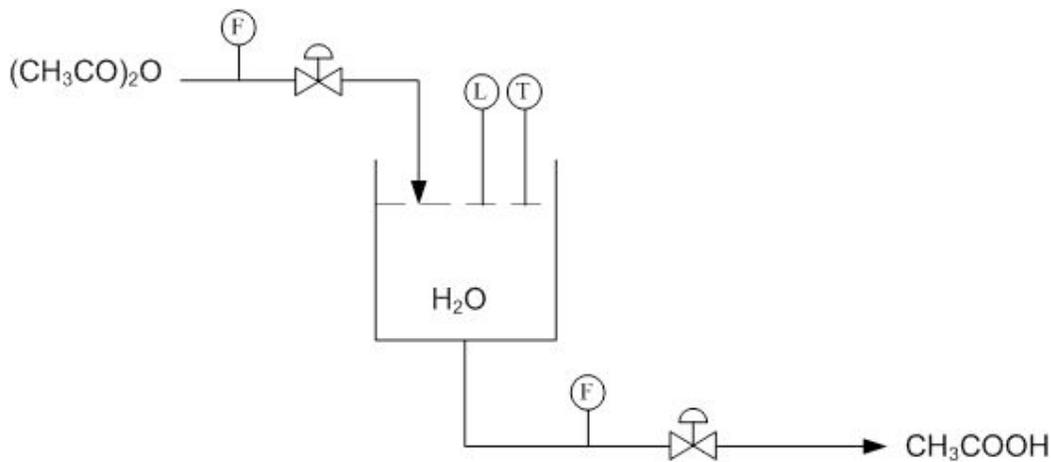
Figure E-2. Location of sensors and valves.

Using Figure E-2 determine how to use the level and temperature sensors to limit the amount of acetic anhydride entering the reactor and also how to drain the system once the reaction has gone to completion.
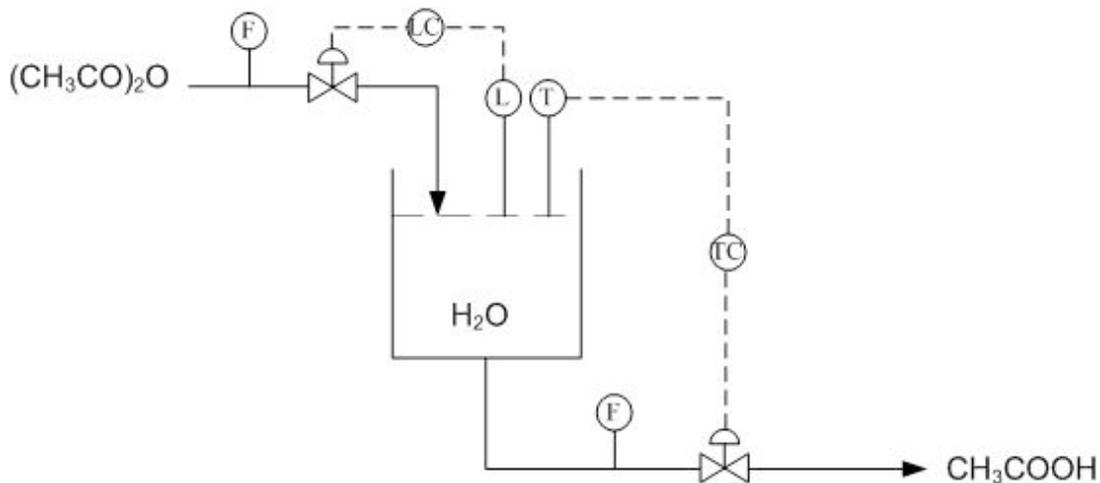
Figure E-3. Control system for production of acetic acid.

Using the above system, when the right amount of acetic anhydride has entered the reactor the level sensor will stop the flow into the reactor. Once the reaction has gone to completion and the temperature sensor has been triggered, due to a peak in temperatures from the exothermic reaction, the tank will drain.

3.6 Example 2: Control of a Heat Exchanger

Examine the figure below, which shows two different control schemes for a heat exchanger. The outlet stream temperature must remain constant. How is this goal

accomplished using the control systems shown below? Which stream is the manipulated stream? What are the advantages and disadvantages of each setup?

(a)
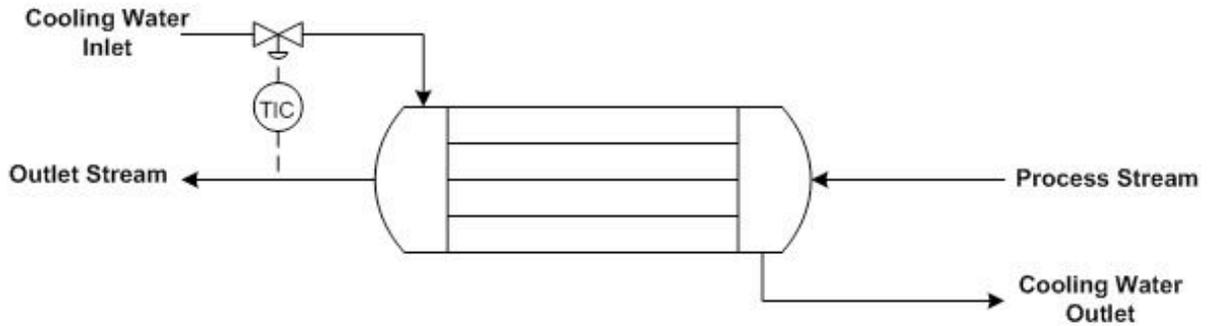
(b)
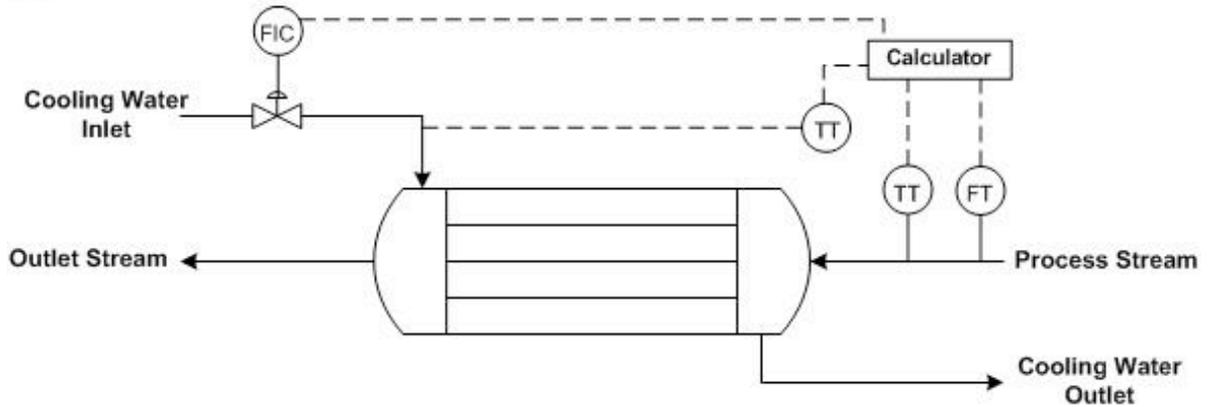
Figure E-4. Two different control strategies for a heat exchanger (adapted from Turton)

In both setups, the cooling water flow is manipulated to control the temperature of the outlet stream. The control strategy for each heat exchanger is different.

Figure E-4 (a) shows a heat exchanger where the flow of cooling water is adjusted based on the temperature of the outlet stream. The advantage to this setup is that it is quite straightforward and the cause of the change in the temperature of the outlet stream does not have to be identified. The disadvantage is that a temperature change has already occurred before action is taken, and if the lag-time for the additional cooling water to cool the process stream is long, the outlet stream temperature may be inconsistent.

Figure E-4 (b) shows a heat exchanger were the flow of cooling water is adjusted based on the flow-rate and temperature of the inlet stream and the temperature of the coolant stream. Based on a set of calculations using the heat transfer coefficient of the heat exchanger, the outlet stream temperature can be predicted. The advantage of this system is that action is taken before changes occur in the outlet stream temperature. The

394

disadvantage is that if the actual heat exchange capacity differs from the expected value due to fouling or mechanical problems, the outlet temperature will not be correct.

3.7 Example 3: An alternate notation, with explanation of controller location and function

This presents an alternate notation of controls. The arrows above the controller indicate what parts of the process that the controller controls.

3.7.1 The Problem

A + B −> AB

1. B is added slowly into a full charge of A
2. The temperature is maintained at TR (RXN temperature)
3. Upon completion of reaction, AB is cooled to TP
4. C is added to AB to make final product, which leaves the tank
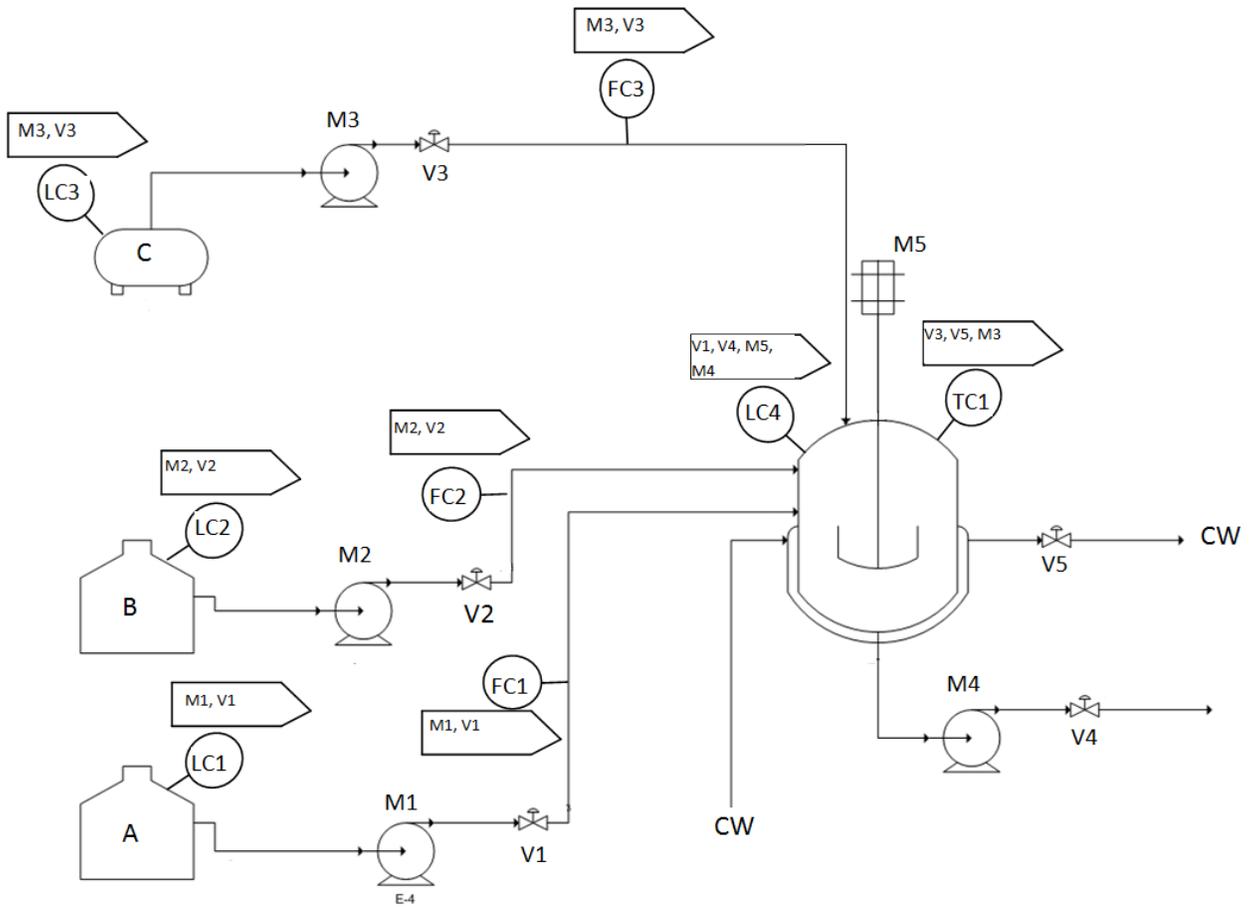
3.7.2 The Solution

- **LC1**: This controls M1 and V1.

  This controller is used to keep the level of A in the A tank above a preset minimum.

- **LC2**: This controls M2 and V2.

  This controller is used to keep the level of B in the B tank above a preset minimum.

- **LC3**: This controls M3 and V3.

  This controller is used to keep the level of C in the C tank above a preset minimum.

- **FC1**: This also controls M1 and V1.

  This controller totalizes the amount of A that has flowed out of the A tank, and turns off flow when the flow amount has reached a preset value.

- **FC2**: This also controls M2 and V2.

  This controller totalizes the amount of B that has flowed out of the B tank, and turns off flow when the flow amount has reached a preset value.

- **FC3**: This also controls M3 and V3.

  This controller totalizes the amount of C that has flowed out of the C tank, and turns off flow when the flow amount has reached a preset value.

- **LC4**: This controls V1, V4, M4, and M5.

  V1 - If level of liquid in the reaction tank goes over a preset maximum, V1 is closed.
  V4 - If level of liquid in the reaction tank goes under a preset minimum, V4 is closed.
  M4 - If level of liquid in the reaction tank goes under a preset minimum, M4 is turned off.
  M5 - If level level of liquid in the reaction tank goes under the level of the agitator, M5 is turned off.

- **TC1**: This controls V3, V5, and M3.

M3, V3 - If the temperature in the reaction tank is at TP, M3 is turned on and V3 is opened.
V5 - To maintain the temperature in the reaction tank at TP, V5 is opened and closed accordingly.

3.8 Sage's Corner

Proper Controls Placement Case Study

http://www.youtube.com/v/B3S_Otjht_s

For just the slides:
http://controls.engin.umich.edu/wiki/index.php/Image:PID_Controls_Slides.ppt

3.9 References

- Bequette, W.B. *Process Control Modeling Design and Simultion.*, New Jersey: Prentice Hall, pp 466-477
- Luyben, W.L. *Plantwide Process Control.*, New York: McGraw-Hill, pp 53-69, pp 201-203, pp 235-247
- *Module 3.1: Control of Distillation Columns.* The ECOSSE IGDS Controls HyperCourse. University of Edinburgh, Scotland.
- Ogunnaike, Babatunde A. *Process dynamics, modeling, and control*, New York: Oxford University Press 1994, pp 1149-1153
- Riggs, James B. *Chemical and Bio Process Control.*, Ferret Publishing, pp 491-504
- Turton, Richard. *Analysis, Synthesis, and Design of a Chemical Process*, New Jersey: Prentice Hall, pp 408-409
- U.S. Chemical Safety and Hazard Investigation Board; "Morton International Explosion." http://www.csb.gov/completed_investigations/docs/mortondigest.pdf. Last Updated 07/17/2003.
- Tompkins, Bill. "Reaction and action." http://www.ehponline.org/members/2004/112-6/focus.html. Last Updated 09/09/2007
- Wikipedia. "Gas Compressors" http://en.wikipedia.org/wiki/Gas_compressor. Last Updated 11/27/2008
- Wikipedia. "Liquid-Liquid Extraction" http://en.wikipedia.org/wiki/Liquid-liquid_extraction. Last Updated 11/20/2008

Section 4. P&ID Standard Pitfalls

Title: P&ID Standard Pitfalls
Note: ⬛Video lecture available for this section!
Authors: Angela Yeung, Curt Longcore, Ben Van Kuiken, Jeffrey Carey
Date Presented: 5 October 2006 /Date Revised: 19 October 2006
Stewards: Eric Black, Megan Boekeloo, Daniel Carter, Stacy Young

Date Updated: 16 September 2007

- First round reviews for this page
- Rebuttal for this page

4.1 Introduction

Piping and Instrumentation Diagrams (P&ID) are standardized in many ways, and there are some fundamental safety features that are absolute requirements for all P&IDs. Unfortunately, many people forget these features in their designs unintentionally. Lacking these safety features could lead to serious engineering problems. It is important to eliminate these pitfalls when designing a P&ID.

In the following sections, different pitfalls of P&IDs will be discussed. The equipment design section will concentrate on how equipment might be accidentally neglected and misplaced while designing a safe and functional process. The design process section will describe how a lack of consideration of process conditions would lead to serious design pitfalls. Overspecification and underspecification of equipment and design will be discussed. There are also some miscellanous pitfalls associated with interpretation and inherent problems of the P&ID itself. Examples are introduced at the end to illustrate common errors and pitfalls of P&IDs.

4.2 P&ID Production Pitfalls

Most of the common pitfalls of P&ID production result from the engineer forgetting that the design on paper represents a real physical process and that there are practical and physical limitations that need to be considered in the process design. The following are some errors to be particularly careful of when drawing up a P&ID:

4.2.1 Equipment Design
Safety valves
Safety valves are part of the essential valves system for P&IDs. Together with isolation valves, they are an absolute requirement for instrument design. Safety valves are required to install for all gas, steam, air and liquid tanks regardless of the tank's function for pressure relief purposes. Engineers should be aware of their system's set pressure, relief pressure, percent overpressure, maximum allowable

398

pressures etc. when selecting a safety valve. The US law requires all tanks of pressure greater than 3 psig to have safety valves installed.

Different pressure tanks require different safety valves to best fit their safety design. Therefore, engineers must be very careful in selecting the right safety valves for their systems. However, when constructing a P&ID, engineers sometimes forget adding safety valves to their design, and this could cause serious problems. For example, if one forgets to add a pressure relief valve or safety valve on a reaction tank of gas and liquid, the extra pressure accumulating would exceed the preset pressure limits for safety design. This could lead to a serious explosion!

A pressure relief valve is symbolized by two triangles orientated at 90 degrees to each other, as shown below.

Pressure Relief Valve
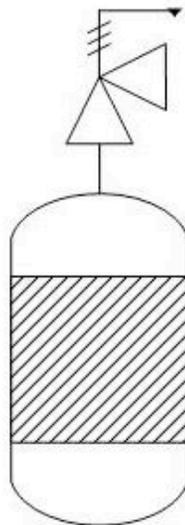
This eliminated excess pressure that might build up inside a reactor.

Isolation valves
The isolation valve is used to isolate a portion from the system when inspection, repair or maintenance is required. Isolation valves are placed around the junctions in the distribution system. They are also part of the absolute requirement for P&ID construction. Engineers should be aware of their system's pressure, voltage, process medium, pipe sizing and flow rates when selecting the right isolation valves that work under the conditions for their system.

If engineers forget to add isolation valves on their P&ID design, serious problems would occur when the system needs to be partially shut down for maintance or other reasons. There would be no way to control the unit's operation other than existing flow valves.

Some upstream production problems could affect downstream production since appropriate isolation valves are not installed for safety and production purposes.
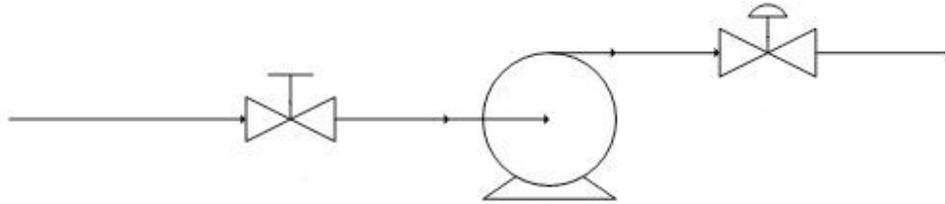
Isolation Valve Placement

Notice that the automatic valve after the pump is a flow valve, while the manual valve before the pump is an isolation valve. One example of this use is for cleaning. If some parts of the pump cannot be cleaned in place, the pump will need to be taken apart. The isolation valve can cut off the flow to the pump, allowing it to be safely taken apart.

Valves and pumps
Some of the most commonly used pieces of process equipment that show up on P&IDs are valves. Valves control the flow of fluid through pipes by opening to allow flow or closing to stop flow. One of the problems associated with valves on P&IDs is the sheer number of them and deciding where to place them in the process.

A common mistake with valve placement has to do with the interaction between valves and pumps. It is important that valves be placed **after** pumps in a given pipe. When a valve is placed before a pump, should the valve close and the pump has not been shut off, there will not be a constant supply of fluid to pump. This is known as *starving* the pump. Starving the pump can create a disturbance known as cavitation, and it is very hard on the equipment and can cause the pump to break. Placing valves after the pump ensure that even if the pump does not shut off, it is still filled with the proper fluid.
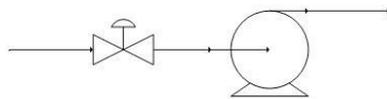
Improper Valve Placement   Proper Valve Placement
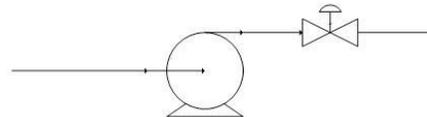
These same principles apply to valve placement with respect to compressors. Placing a control valve upstream of a compressor eliminates your ability to control pressure in the pipeline downstream of the compressor and introduces a risk of starving the compressor.

Agitators
A point that is very easy to miss, and very expensive if it is missed, is that if a vessel

400

is equipped with an agitator, the vessel **must** be filled enough to submerge the agitator before the motor is turned on. Agitators are designed to operate against the resistance of fluid. Without this resistance, the agitator will rotate much faster than the equipment is designed for. This rotation may result in hitting the harmonic frequency of the structure, compromising its integrity and causing the agitator to rip from its foundation. This creates not only a fiscal predicament (not only ruining the agitator but also the vessel), but a safety nightmare. When designing a process, one must make sure he or she knows and accounts for how much fluid must be in an agitated vessel to avoid this situation. This can easily be solved by adding a level sensor to the tank that the agitator is being used in.
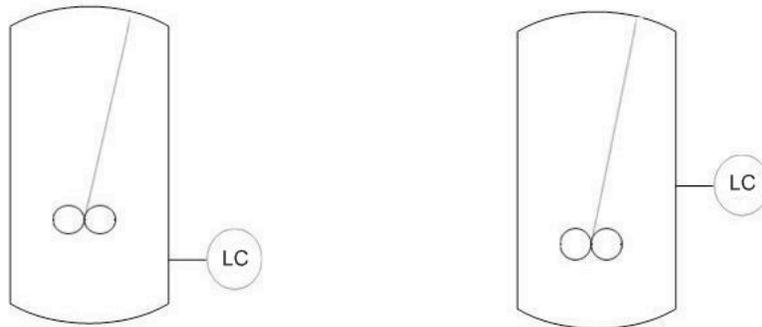
Improper Placement   Proper Placement

When placing the level sensor on the tank, make sure to place the sensor above the level of the agitator. This will ensure that the agitator is submerged in the fluid. It would be incorrect to place the level sensor below the agitator.

Instrument selection and placement

Instruments are designed to operate properly under specific conditions. Every instrument has a range over which it functions properly, and instruments must be selected that are appropriate for their applications. For example, a pressure gauge might have a working range of 5 psig - 50 psig. You would not want to use this gauge for sensitive measurements in the range 3 - 6 psig. Instrument material must also be considered during the selection process. If the substance being monitored is corrosive, for example, the instrument must be made of a corrosion-resistant material.

Once an appropriate instrument has been selected, it must be appropriately placed. For example, a level control is not useful in a pipe because there is no need to measure any water level inside of a pipe, much like a flow controller is not useful in a storage tank because there is no flow. Similarly, a flow controller should not be placed on a valve, but instead downsteam from the valve. However, level controls **are** useful in storage tanks and reactors, while flow controllers are useful in pipelines. Instruments must be selected and placed to reliably provide useful information and to accurately control the process.

Equipment Selection

When creating a P&ID, the equipment that is selected to be used is very important, not only to maintain a smooth process but also for safety purposes. Each and every piece of equipment from 100,000 liter storage tanks to temperature sensors has *Operational Limitations*. These are the conditions under which a given piece of equipment operates as expected, with safe, consistent, and reproducible results. For example, when storing a highly pressurized gas at 2,000 psig, one wouldn't want to use a storage tank that has been pressure tested up to 3 psig. The process conditions are way outside the operational limitations and would pose a serious safety hazard. The same goes for sensors and gauges of all types. On a pressurized vessel at 2,000 psig, it would be no good to use a pressure control system that has a sensor that is meant to measure up to 100 psig, the results would not be accurate.

Operational limitations can usually be found in the equipment manual sent by the manufacturer or possibly on the manufacturer's website. If it is not found in either place, an engineer is obligated to contact the manufacturer and find the operational limitations before using a piece of equipment in a process. More can be found on equipment selection for some (but by no means all) equipment below:

Temperature Sensors

Pressure Sensors

Level Sensors

Flow Sensors

Composition Sensors

pH and Viscosity Sensors

Biosensors

Valves

4.2.2 Process Design
Unit operation input/output
Providing clearly specified inputs and outputs to and from process units is vital. For the safety of the system, proper control and the prevention of disaster, it is important to show where each substance came from and where it is going. The P&ID must show all material streams to and from separation units, heat exchangers, and reactors.

For example, if the P&ID is not clearly denoted in a reaction to create construction explosive materials, then a large and perhaps fatal calamity can occur at the chemical

plant. The exiting streams of a plant may pour into a river which may violate environmental regulations. Therefore, specifying inputs and outputs is imperative in a P&ID.

Pressure and flow

The movement of fluid is essential in many production lines. Transporting material from a tank to a reactor, a reactor to a distillation column or from a column into a tanker truck all involve the movement of fluids. When designing P&IDs, a process engineer must decide how they are going to attack this problem, namely whether or not a pump is needed to move the fluid.

The main issue here is pressure. Fluid, both gaseous and liquid, moves down the pressure gradient from high pressure to low pressure. The rule of thumb is that if the source of the fluid has a much higher pressure than the destination of the fluid, a pump is not needed. Even if the source has a pressure only a fraction higher than that of the destination, a pump may yet be needed for the fluid to flow through the pipes, from the source to the destination. However, the flow and pressure from the source must also be compared to the needs of the process. For example, if a inlet flow has highly fluctuating pressure, a pump or valve should be used to regulate the pressure for the process. Also, if the diameter of the pipe does not remain constant and is fluctuating as well, perhaps due to outside pressure or force, a pump should be used to control the pressure inside the system.

A common mistake young process engineers make is when charging a pressurized vessel, they do not use a pump powerful enough to overcome that pressure. This causes *backflow* and can ruin process equipment. This problem is especially prevalent in recycle streams. In some situations it may be appropriate to use a check valve (a valve allowing fluid to flow through into it from only one direction), to add an additional barrier against backflow. An example of such a situation would be when a fuel is to be mixed with air for combustion: if any air were to backflow to the fuel source, a dangerous, explosive situation would arise. In this case the use of a check valve would be appropriate.

An example of when a pump is never needed is when liquid from one vessel at ambient pressure is being transported to another vessel at ambient pressure that is at a lower elevation. This is known as a *gravity feed* and utilizing gravity feeds where possible can significantly decrease the cost of a process. As long as the pressure at the exit of the pipe is lower than the pressure created by gravity, a gravity feed can be used. An example of such a gravity feed is the flushing of a toilet, by which the water from the cistern at a higher elevation falls to the water closet at the lower elevation.

4.2.3 Underspecification

For safety and control purposes, **redundancy** is desirable in control systems. A process lacking in suitable redundancy would be **underspecified**. For example, a reactor containing an exothermic reaction would be underspecified if it only contained a temperature control. The potential danger posed by an exothermic reaction warrents a

high degree of safety: if one controller were to malfunction, a backup must be in place. A pressure control should also be included, and perhaps a level control as well depending on the specific process.

4.2.4 Overspecification

On the flipside of underspecification is **overspecification**. Adding too many controllers and valves on vessels and lines is unnecessary and costly. Since process control is highly dependent upon the situation, it is difficult to set specific limits on the necessary amount of specification. However, a controller should not be used unless there is a specific need for it. Misused redundancy is a common example of overspecification. It is unlikely that a water storage tank needs level, temperature, and pressure controllers. A degree of freedom analysis, as shown in a distillation column example, can be helpful in determining the exact number of controllers, sensors, actuators, and valves that will be necessary to adequately control a process.

4.3 Problems When Using a P&ID

Other than equipment and production pitfalls, there are also some general P&ID interpration and inherent problems.

4.3.1 Interpretation of P&IDs

Although it is essential for P&IDs to represent the right instruments with the right references, many P&IDs do not support scaling or do not require scaling as part of the system. P&IDs are drawn in a way that equipment and piping are displayed for ease of interpretation. They do not show the physical placement and location of different systems or the actual sizes and length of equipment and pipelines.

Experience tells us that many engineers have overlooked certain pieces of equipment or over-estimated the size of equipment and piping, leading to calculation and construction errors. For example, an engineer could be designing new equipment on the existing P&ID and not realize that existing equipment is so densely populated that extra equipment will not fit. It is important to check the physical space of an area before adding equipment to a process.

4.3.2 Inherent Problems of P&IDs

There are a few problems with P&IDs that are only solved by being conscious of them. One problem is that P&IDs are constantly being updated, revised, changed, and added to. When reading a P&ID it is always important to check the date it was last revised and if there is a later revision available. Making adjustments to equipment that isn't there anymore not only causes confusion and frustration, but is a waste of everyone's time. Moreover, making sure you are using the most recent edition to the P&ID will ensure that you do not purchase equipment to install that has already been installed.

Usually, companies will have a computer database with the most current P&IDs. Before modifying or working with an old paper copy of a P&ID, check to make sure it is up to date with the most current revisions.

Another problem with P&IDs is that even if they are well made and technically perfect, complex processes often appear cluttered. There is an inordinate amount of information contained in a P&ID, and as such P&IDs are next to impossible to take in at a glance. It is important to carefully study each document to fully understand the process. The mistake of assuming you gained all the information off of a P&ID from simply scanning it can leave you with mental holes that you don't even realize you have. This can be avoided by reading the P&ID carefully on the first pass and constantly referring back to it when questions arise.

4.4 Worked out Example 1

A pipe connects a water tank (open to the atmosphere) and a vessel pressurized to 28psi. The open tank has 10 feet of water in it. A level control system ensures that the 10 foot level is maintained in the open tank. The bottom of the pressurized vessel is 20 feet below the bottom of the open tank and starts with no water in it. The goal is to fill the pressurized vessel up to 5 feet. Due to a poor design by the project engineer, the water is fed into the bottom of the pressurized vessel. Given that the density of water is 62.4 lbm/ft^3 and the gravitational constant is 32.2 ft/s^2, is a pump needed? If so, where should it be placed? Assume that there is no pressure drop due to friction against the pipe and that the air pressure of the pressurized tank remains at a constant 15psi. The figure below may be helpful in visualizing the process. (Hint: Remember in English units, a "fudge factor" is needed when using $P = \rho*g*h$)
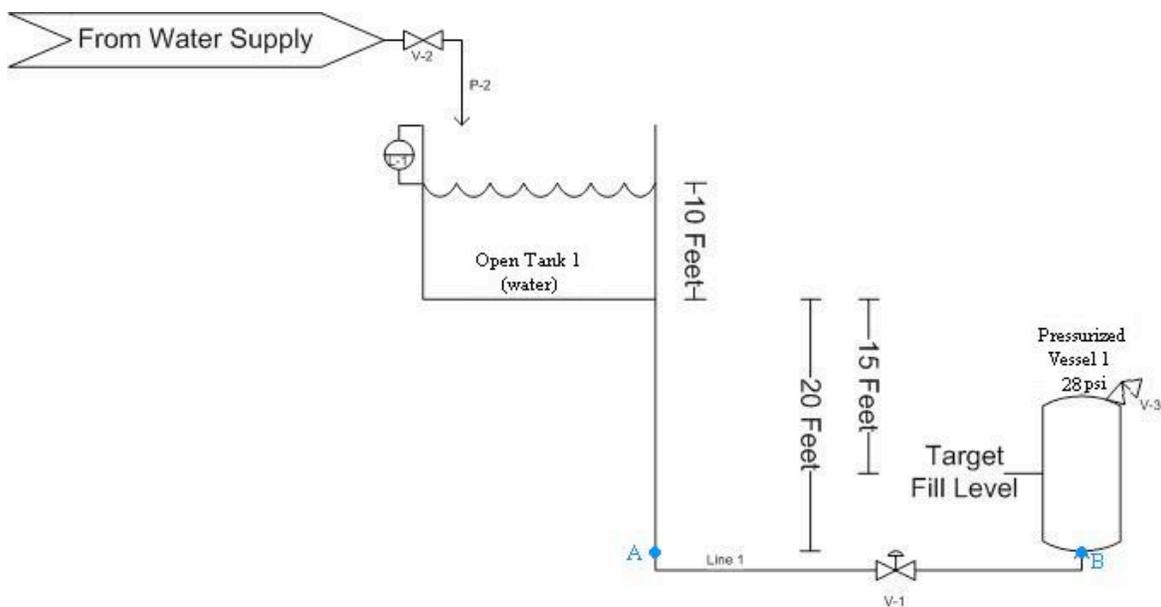
Answer: First, we must determine the pressure at the stream exit (point B). When the vessel is full, we know that the pressure created by the water is:

$$P = \frac{\rho g h}{32.2}$$

Plugging in 62.4 for the density, 32.2 for g and 5 for h, we find that the pressure caused by the water is 312 lbf/ft^2. Dividing by 144 to convert to psi, we see this is equivalent to 2.2 psi. When we add the 28psi from the air pressure in the vessel, we get a grand total of 30.2 psi at the pipe exit.

The pressure in the feed pipe at the height of the exit to the pressurized vessel (point A) is again given by:

$$P = \frac{\rho g h}{32.2}$$

Only this time, the top of the water is 30 feet above the pipe exit instead of 5 feet. Plugging in this value, we find that the pressure is 1872 lbf/ft2. After converting, we find this is equivalent to 13psi. Since the water tank is open, we must also add 14.7psi for atmospheric pressure. This gives us a total pressure of 27.7psi.

The pressure of the feed at point A in the diagram is LESS than the pressure at point B, so a pump is needed. The pump should be placed BEFORE the valve labelled V-1 on the pipe as shown below. This is to prevent starving the pump when the valve closes.
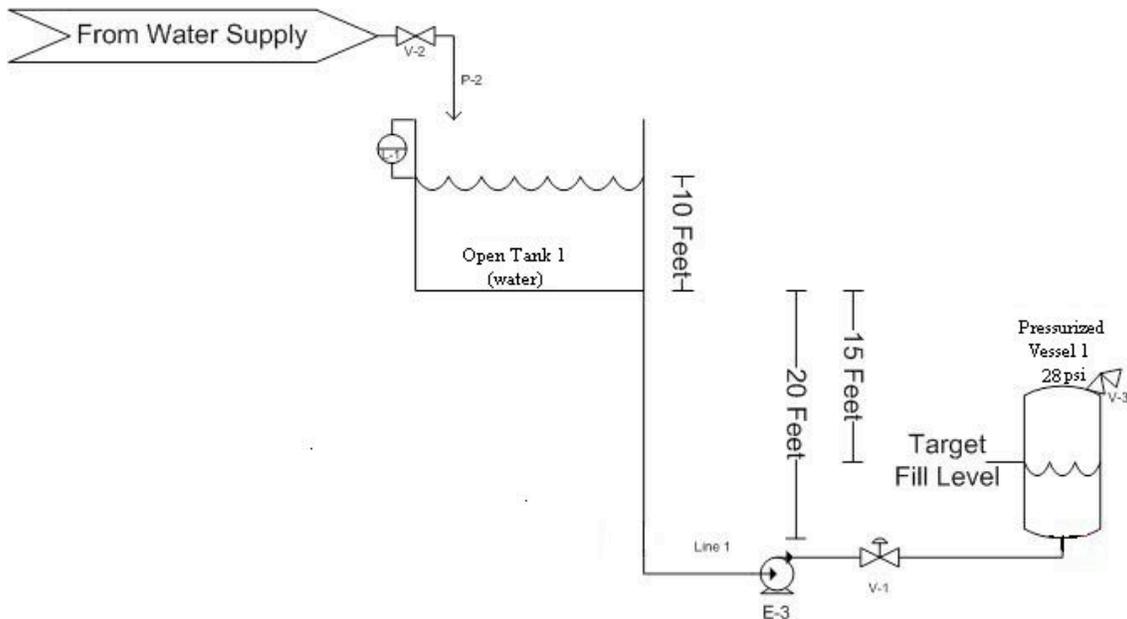
There are several errors in the following P&ID, which shows a vessel where a liquid is vaporized using heat transfered from steam passing through a surrounding jacket. Identify the errors and explain why they are mistakes.
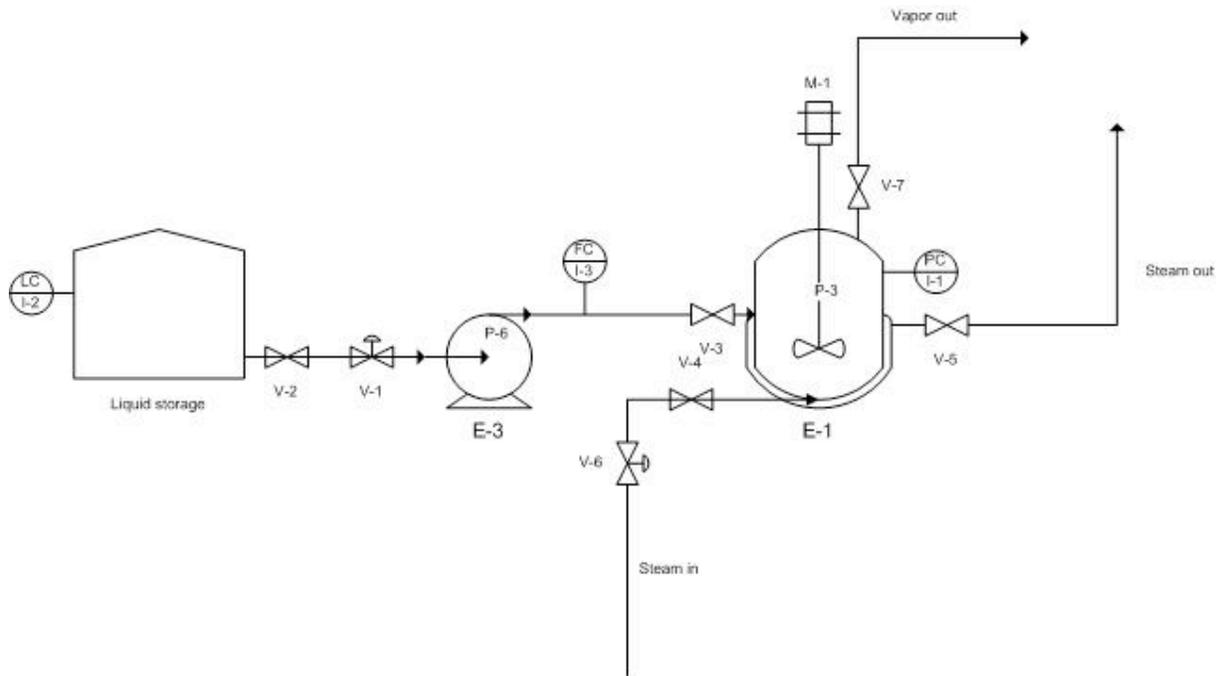
Answer:

- 1. Control valve V-1 should be after the pump, not before. Placing a pump after a control valve takes away your control of the flow rate and can also damage the pump if the valve is shut off while the pump is running.
- 2. Vessel E-1 is only being monitored with a pressure control. This is not adequate redundancy for process control. There should also be a temperature control (to make sure the vessel is at an appropriate temperature for vaporization) and a level control (to make sure the agitator is always submerged in liquid).

- 3. Although the vaporization takes place at ambient pressure, there should be a safety valve on the vessel. What if there is a malfunction and too much vapor is produced?
- 4. The input and output streams for vessel E-1 are not sufficiently labeled. The source of input streams and destination of output streams should be specified to ensure adequate control of a process.
- 5. There should be a flow meter and control valve on the vapor stream exiting the vaporization vessel so that you can monitor and control the amount of vapor exiting.

4.6 Sage's Corner

P&ID Standard Pitfalls Presentation #1
http://video.google.com/googleplayer.swf?docId=-27175650286417763

P&ID Standard Pitfalls Presentation #2
http://video.google.com/googleplayer.swf?docId=-2067825845180670255

[Powerpoint Slides without narration](#)

[Powerpoint Slides without narration](#)

[PDF of slides from Presentation 1](#)

[PDF of slides from Presentation 2](#)

4.7 References

- Center for Chemical Process Safety. (March 2006). *Process Safety Beacon: Pressure Relief Systems - Do You See Any Hazards?* CEP
- Melhem, G.A. and Howell, P. (September 2005). *Designing Emergency Relief Systems for Runaway Reactions.* CEP
- Ulrich, G.D. and Vasudevan, P.T. (July 2006). *Predesign With Safety in Mind.* CEP
- Williams, J.P. (September 2004). *Reliability for Safety Instrumented Systems.* CEP

Section 5. Safety features in P&ID

Title:Saftey Features in Chemical Processes
Note: Video lecture available for this section!
Authors:Bradley Anderson, Aaron Bennick, Michael Salciccioli Stewards: Jocelyn Anleitner, Stephanie Combs, Diane Feldkamp, Heeral Sheth

Date Presented:8/3/06 Date Revised:

- First round reviews for this page
- Rebuttal for this page

5.1 Introduction

Safety has become integral to the manufacturing world. The implementation of proper safety techniques and accident prevention can not only save time and money, but prevent personal injury as well. P&IDs, when properly utilized, are a powerful resource to identify safety hazards within the plant operations. The following sections provide an overview for the safety hazards that exist within a process, and illustrate the importance of P&IDs in a chemical plant.

5.2 Safety in Design

During the early stages of plant design it is critical to determine important safety features that remove potential hazards from effecting the facility environment. Regulations require that plant designers play a major role in minimizing the risks associated with these hazards. However, in order to do so, designers need to be aware of the hazards that exist during plant activity. The facility design team must develop a detailed drawing (P&IDs) including specifications of the plant process and environment to ensure that every aspect with regards to safety is covered.

5.2.1 Hazard vs, Risk

When discussing safety, the terms hazard and risk are often used interchangeably. However, the difference in definition between the two terms is critical in utilizing the information they provide in increasing the safety within a plant. Hazard is defined as a potential source of danger and risk is defined as the level of threat associated with the hazard. A risk of a hazard occurring can be represented mathematically by the following equation:

$$Risk = Frequency \times Consequences \quad \text{(1)}$$

Frequency represents the probability that a hazard will occur, and consequence represents the impact of that hazard. Values for each parameter of the equation above are assigned by using either experimental information or educated judgment based on engineering models and experience. Each plant design process will have specific safety hazards and risks associated with it. Therefore, there is no predetermined value that can be assigned to each variable. For example, in one situation, a water tower may be well supported inside the plant facility, and in another situation a water tower of similar structure may be hoisted against a rusted frame work outside of the plant facility. Both of these situations have different levels of risk associated with them. Needless to say, the risks associated with a plant setup can be reduced substantially by minimizing the probability that a hazard will occur.

5.2.2 Hazard Locations and Risk Hotspots

Common hazard locations exist in any place containing large amounts of energy. The degree of danger is proportional to the amount of energy stored at that location. Risk can be directly linked to kinetic energy, potential energy, work, heat, enthalpy and internal energy sources. Kinetic energy, otherwise known as energy in motion, is present in any moving component. The component may be vibrating, rotating, or translating, and this motion causes the kinetic energy of that part to greatly increase. Within industry, personal injuries and fatalities are common hazards associated with moving parts. Reactors and cooling towers placed in high locations contain potential energy. If a structural failure were to occur within a plant, these structural units and their contents could fall from a large elevation onto another processing system or a human being, releasing all the chemical contents stored inside. Similarly, the stored work in springs and other devices can cause fatigue and wear on the mechanical system over time, and result in eventual machine failure. Heat released from a reaction within a chemical reactor can be rapid and fatal if not accounted for. The buildup of heat can cause serious consequences with runaway reactions and boiler explosions. The enthalpy and internal energy of a reaction typically are the cause of most runaway reactions and destructive fires in a plant.

Risk Hotspots

Uncontrolled chemical hotspots created within a chemical plant are a common source of hazard, besides the hazard locations pertaining to energy sources such as kinetic and potential energy. Risk hotspots mostly occurs in the piping system and associated valves of the system, joints, traps, and various other piping elements. Possible malfunction of the system due to structural corrosion can be triggered by the failure to maintain the piping systems efficiently and periodically. Even if one of the valves in the system has corroded and is unable to function properly, the fluid flowing through the pipe might get trapped, and the resultant buildup of pressure in the pipe may cause major safety hazards including fatal ones such as an explosion. If the system is not shut off before the pressure gets out of control, the pressure buildup in the pipe will cause it to burst, releasing all of its internal contents to the surrounding environment.

Storage vessels are other pieces of equipment that must be cared for properly. Since the plant operators and engineers do not usually interact with the storage vessels, as compared to other parts of the plant process such as the piping system, they are considered to be of secondary importance and commonly overlooked. Storage vessels have much more content inside them than pipes, so a leak or a burst would take longer to get back under control than pipes, which can by plugged more easily. Problems which may arise from storage vessels are not associated with their design, but in fact from not thoroughly and periodically maintaining them. Possible complications that can arise from neglecting the storage tanks are over-pressurization, overfilling, heating element malfunction, or simply equipment malfunction.

Chemical reactors are another common location where risk hotspots occur. The nature and design of commercial chemical reactors is to handle a controlled explosion. However, if the control element is removed, disaster is bound to occur, just as in any other part of the chemical plant. The most common type of hazard is a runaway reaction inside a batch reactor. When the plant facility looses electricity or cooling water, a runaway reaction will strike inside the reactor. Once a runaway reaction has spun up inside the reactor, many other hazards may follow, such as flow reversal in the pipes, incorrect reagent charging, heat exchanger failures, and external fires. Other hazards may perhaps be even more serious, such as engineering errors that could potentially cause a runaway reaction to occur, including inappropriate material selection, inadequate equipment inspection or failure to fully understand the chemistry or exothermic nature of a reaction. A runaway reaction originating inside the chemical reactor can easily cause a chain reaction across the rest of the equipment at the facility, and can result in the entire system malfunctioning.

Other process equipment that may be hazardous and where risk hotspots commonly arise, are vacuum operators, furnaces, pumps, gas movers, compressors, and heat exchangers. The location and type of specific piping and unit operations are available on the process P&ID. A responsible process engineer should use the P&ID to identify all risk hotspots, and act accordingly to monitor and maintain a safe working environment. In addition, a standardized plan should be constructed so that in the event of a malfunction, the correct steps can be taken to bring the faulty part back under control. Supplementary precautions should be taken to prevent a comparatively minor malfunction in the system from becoming a disaster which may violate environmental regulations and even endanger human lives.

5.2.3 Safe Design Principles

The ISD or Inherently Safer Design movement was a doctrine striving for safer chemical processing procedures. This movement was pioneered by Trevor Kletz in 1976, and promotes the design of processes so safe, that no catastrophic failure can occur within the plant. The following principles apply to initial process design:

1. Use the fewest number of hazardous substances in the smallest quantities and still maintaining plant productivity

2. When possible, substitute hazardous chemicals with chemicals that are less dangerous

3. Practice moderate operating conditions in the plant

4. Use the simplest plant design possible

5. Design equipment in the plant to minimize the effects of a hazardous incident

The specifications determined by the process designers are communicated through the P&ID.

5.2.4 Hazards in Construction

In order to eliminate hazards, a operations personnel must be able to identify that a hazard exists. Hazards that may be encountered on plant sites may be categorized into three main types:

A) Hazards harmful to health

B) Hazards likely to cause personal injury

C) Hazards likely to lead to catastrophic events

Hazards harmful to health:

When workers are exposed to or come in contact with asbestos, corrosives, irritants, toxins, or noxious gases try to avoid by specifying the processes, which lead to this exposure

Hazards likely to cause personal injury:

Hazard awareness is increased when people have to work in situations likely to expose them to the risk of personal injury, including moving plant machinery or working in areas where objects are likely to fall. Situations where there are live electrical circuits overhead, buried power lines, and confined working conditions are likely to cause personal injury.

Hazards likely to lead to a catastrophic event:

These hazards have consequences beyond the site boundary. They include fire outbreaks, explosions, flooding, or premature collapse of structures, cranes, tunnels and excavations.

5.2.5 Fail-Safe Design

The fail-safe design of a unit operation (such as a reaction vessel) requires a complete understanding of the operation at hand, and the knowledge of all the worst-case

conditions. A fail-safe system is a unit operation such that, if any or all of the worst-case conditions were to occur, the operation would shut itself down automatically and in a safe fashion. In the case of a run-away reaction, if reagent feed limits, interlocking controls, and integrated heat balances are all properly maintained, the reaction cannot "run away." Other precautions such as purges, vents, dump tanks and quenches are available for reaction vessels, and should be visible on the P&ID.

5.2.6 Inherent Safer Predesign

The table below provides a guideline for identifying and minimizing hazards partly based upon Kletz's rules for ISD in the Safe Design Principles section above. The step/rule column describes the action taken. The tools column describes the mechanism by which the actions are taken. The experimental and analytical resources column describes the knowledge by which the mechanisms are created. The Literature References column describes where the knowledge can be found.

Step/Rule	Tools	Experimental and Analytical Resources	Literature References
1. Identify	Fire triangle, Flammability-limit chart, Chemical reactivity chart, Safety compatibility chart, Safety stream chart, Toxicity ratings	Thermodynamic calculations, Reactor design equations, calorimetry, flammability charts	Perry's Handbook (10) Table 26-10 and associated pages
2. Eradicate (ISD Rule 2)	Inert-gas blanketing, Failsafe design	see above	see above
3. Minimize, Simplify, Moderate, Attenuate (ISD Rules 1, 3, 5)	Moderate ignition sources, Keep reactive volumes small	Logical chemical engineering analytical and computational skills	Merk Index, MSDSs, Chemical Engineering textbooks and monographs
4. Isolate (ISD Rule 4)	Separate hazardous operations - Surround in an impenetrable structure	Logical chemical engineering judgment	

5.3 Relief Systems

Organizations such as the American Society of Mechanical Engineers, American Petroleum Institute and National Fire Protection Association layout recommendations and design standards so that most engineers with proper training can setup proper emergency relief systems for single phase flow. Unfortunately, this is often not good

enough for reactive systems. These systems are much more complex and include multiple phase flow, runaway reaction potential and self reactant material. When designing an emergency relief system (ERS) it is necessary to understand all aspects of the chemicals and processes that will be in play. This includes but is not limited to: kinetics of the possible reactions, contamination, interactions with air, rust, piping, or water, phase changes and runaway reactions. The follwing topics deal with the hazards that require designing relief systems and the prevention of runaway reactions.

5.3.1 Planning and Design

When designing an ERS, it is important to consider the worst-case scenario. This is based upon a thorough knowledge of the reactions, materials, and environment of the process. Some of the most critical scenarios are over-pressurization in a reactor.

All disaster scenarios can be analyzed using a hazard and operability (HAZOP) study. This HAZOP study will analyze a process based upon human, equipment, and environmental factors.

The HAZOP team must take the following steps to ensure that all potential scenarios are taken into account:

• Review the potential hazards of all chemicals. This includes non-operational conditions and interactions with contaminants

• Study the chemical process including all possible reactions, rearrangements, decompositions, etc.

• Review the P&ID's for the process

• Study the specific reactor and storage vessels for material composition compatibility, size, surface area, instrument ranges, and set points

With this information, the HAZOP team can determine most of the potential disaster scenarios. Most likely the worst case scenario will involve fire induced runaway reactions.

5.3.2 Design Strategies

There are various techniques often used to prevent failure from over-pressurization, fire, runaway reaction or other disaster scenarios. The following are some design suggestions that will decrease this risk:

-The use of insulation in case of fire. Unfortunately, insulation will also minimize heat loss from the reactor during a runaway reaction. It is important to consider the ability of a vessel to drain when using insulation. If the contents can be drained and cooled before the reaction starts to runaway, catastrophe can be prevented.

- Design to avoid fire damage on sensitive equipment. The possibility of fire damage on electrical equipment or sensors will cause control difficulties when trying to slow down a runaway reaction, or monitor another emergency situation.

- Consider the structural integrity of the entire system due to fire damage.

- Use multiple purge streams and valves to separate materials. Separating reaction components into a storage vessel with a large surface area will allow for faster cooling.

- Install two separate relief devices in case one has been compromised due to fouling or solid particle blockage.

Overall safety relief plans must be made for all possible scenarios.

5.3.3 Reaction Kinetics

Reaction kinetics usually determine the potential for safety disaster. For the average exothermic reaction, the reaction rate doubles with every 10°C increase. This will lead to an exponential increase in energy which will force both the temperature and pressure in a system to uncontrollable levels. Since reaction rates are sensitive not only to temperature, but also pressure, contaminants, concentrations and phases, all possibilities must be adequately analyzed for plant operation to be deemed safe.

Relief devices should be designed to handle chemicals and/or mixtures in any phase. For example, if a reactor temperature increases suddenly and a safety relief valve is activated the material must be able to flow through the valve quickly and safely. If the chemical has changed phases, or has a higher pressure that the valve cannot accommodate, the chemical will not be able to escape and build up pressure in a reactor. This would greatly increase the risk of a disaster. Because of the extent of possibilities when considering reaction rates and kinetics, it is usually necessary to consult some sort of computer simulation or dynamic simulation tool to plan for every possible scenario.

5.3.4 Runaway Reactions

Runaway reactions are caused when exothermic reactions are fed more energy due to malfunctioning cooling systems. This causes an exponential increase in temperature, which in turn causes an increase in pressure, and finally damage to the reactor and/or plant. The possibility of malfunctioning cooling systems must always be considered for flammable materials. If materials are being stored at temperatures above their flash point,

fire is always a possibility. Fire will cause a reaction to reach runaway conditions with very little reactant. Relief systems for fire induced runaway reactions must be larger than conventional runaway reactions.

While relief systems are often designed for over-pressurization, this might not be enough when considering flammable materials. Constant or prolonged exposure to flame will cause most normal reactant or storage vessels to fail causing chemical leaks or plant fires. Fire proof insulation must be used together with normal pressure relief systems to prevent system failure. An important design variable when considering fire induced runaway reactions is fire flux. This variable considers heat impact on a reactor due to fire. Formulas for calculating fire flux can be obtained from the National Fire Protection Association. The elevation of flames must also be considered when designing reactor vessels and safety relief systems. Pool fires can produce flames that are hundreds of feet high; using the P&ID will be important in ensuring that process components sensitive to fire are sufficiently protected from something like pool fires.

5.3.5 Two Phase Flow

For systems in which multiple phase flow is possible, all phases must be planned for. For vapor hybrid systems, all-vapor flow should be considered. Likewise, when foam flow is possible, all-foam flow should be planned for. This is all part of the mindset of planning for the worst case scenario. The most conservative design should always be used when faced with uncertainties of phase flow.

Often times, a runaway reaction will cause high-viscosity two phase flow. Relief valves and safety features must be ready for viscous flow. Many resources suggest averaging the viscosities of the two phases to plan for two phase flow. It is important to keep in mind however, that two phase flow discharge will separate in the discharge line. This will lead to higher pressure drop. Piping sizes are often underestimated due to this unplanned scenario. Undersized piping will lead to valve rupture and back pressure and could cause venting disturbances during the emergencies where venting is critical to the safe shutdown of a process.

5.3.6 Regulations

There are often conflicts and contradictions in federal regulations and recommended practices for safety design guidelines. When faced with such confusion the designer should at least design to the level of the Occupational Safety and Health Administration's (OSHA) requirements, as well as meeting regulations set by the Environmental Protection Agency (EPA), the Process Safety Management(PSM), and the Risk Management Program (RMP). While these guidelines and regulations produce a minimum standard to follow, when dealing with reactive systems, this does not always meet process needs. For more details on the various regulatory agencies that govern chemical processes, see Regulatory Agencies and Compliance.

416

5.4 Additional Safety Hazards in Chemical Plants

This section outlines potential safety hazards commonly found in places where chemicals are stored or chemical processes are taking place. While these hazards are rather easily prevented through attention to detail and general awareness, neglecting them can have catastrophic consequences.

5.4.1 Time Sensitive Chemicals

Some chemicals have a "shelf life," or an expiration date provided by the manufacturer. The chemical must be used by this date or properly discarded. These are typically reactive chemicals, which can become unstable after a certain period of time, possibly rupturing the vessel in which they are stored. An example of this is a monomer that begins to polymerize unless an inhibitor is present. This inhibitor is completely consumed after a certain period of time, allowing polymerization to occur, and therefore must be used or discarded by that time. Another example deals with the formation of peroxide, which can be a severe fire and explosion hazard. It may also be a health concern, causing severe mucous membrane, respiratory tract, skin, and/or eye burns. Peroxide-forming materials should be stored carefully, labeled with the date received and the date first opened on the container. Chemicals should be disposed of or checked for peroxide formation after six months; do not open any container with solid formation around the lid. There are several ways to prevent this from happening. Material Safety and Data Sheets (MSDS) should be available for any chemical the plant uses. These can inform you if a particular chemical becomes unstable after a certain period of time. It is also important to make sure there are procedures in place for handling time-sensitive materials. Lastly, investigating near-miss occurrences can help to ensure that future incidents do not occur.

5.4.2 Pressure Relief Systems

Any open pipe in a chemical plant is a potential discharge site. While operator convenience and maintenance remains a concern, safety takes precedence when dealing with relief systems. It is essential to operator safety that relief valve discharge sites are located in areas that pose a low risk of exposing personnel to chemical hazards and are directed away from all access platforms. It is common in industry to tie multiple pressure relief devices to an emergency vent header that releases on the roof of the plant. Long stretches of unsupported pipes also pose a potential threat. Force generated by material flow could bend or break the pipe impeding on plant operations, and more importantly, injuring personnel.

5.4.3 Dust Explosion Hazards

Most flammable solids can form an explosive dust cloud if the particles are small enough. Materials such as wood, grain, sugar, plastics, and many metals can all form these explosive dust clouds. Dust explosions occur when a combustible material accumulates in a confined area and is exposed to an ignition source. High risks areas are usually those

that are neglected such as tops of vessels and tanks, on pipes, storage bins, bucket elevators, and dust collectors.

To prevent dust explosions, it is essential to implement good housekeeping practices. Care must be taken so that the cleaning process, such as sweeping, doesn't turn a dust layer into a more dangerous cloud. Operations and process engineers should be aware of all the mechanical and safety control equipment associated with preventing fires. Equipment like electric vacuums must be appropriate for use in an area where an explosive dust cloud could form. This entails no holes or cracks in the cord, sufficient grounding, and receiving site approval for use of the equipment.

5.4.4 Overfilling Tanks

The overfilling of vessels has long been a leading cause of serious incidents in chemical and petroleum industries. When a level sensor or high-level alarm fails, reactive material can spill over the tank and accumulate. If this material is exposed to an ignition source, there is the potential for an explosion, leading to serious property damage, environmental issues, and injury to operations personnel. When filling or draining a tank with material, operators should be aware of all relevant level, pressure and temperature controls in place, and watch for abnormal trends. Also, all safety critical alarms surrounding a vessel should be tested regularly at frequencies recommended in plant process-safety-management procedures. Conducting regular process maintenance on safety critical alarms have numerous benefits, which include reducing the risk of operating a plant with faulty equipment, and increasing operator awareness with the location and function of critical safety devices.

5.4.5 Containing Storage Tank Spills

Engineering controls are implemented into plant designs to account for potential disasters, such as a spill, leak, or complete emptying of a storage tank. Industrial-sized chemical plants store large amounts of raw materials, products, and byproducts on-site. The amount of each varies, but is typically between three to thirty (or more) days of the required supply or amount produced. The stored chemicals can be hazardous, flammable, explosive, and/or reactive with each other. In the event of a tank spilling, dikes are built around tanks to contain the spill and protect the surrounding community from the spill. The regulations of the dikes include the following: dike volume must be 1.5X the largest storage tank contained by the dike, reactive materials cannot be stored in the same dike, and scuba gear must be present on-site if any dike is deeper than four feet. Since the dikes cannot contain any reactive materials, the implementation of dikes affects the overall plant layout. Typically, dikes are designed to be like speed bumps and have a height less than one foot so fork-lifts and tankers can easily maneuver through the plant. Dikes are an effective engineering control that greatly improve the safety of a chemical plant.

418

5.4.6 Temperature and Pressure Ratings

Before a vessel is put into plant operation, it is rated and stamped by the manufacturer with temperature and pressure limits. Problems arise when personnel overlook the inverse relationship between temperature and pressure for gases. Be aware that equipment rated for a specific temperature and pressure, cannot be operated at the same pressure if the temperature is increased. While it may seem extremely intuitive (think ideal gas law), this relationship is too often overlooked, usually with serious consequences. When operating any process, pay attention to the temperature and pressure ratings. If they aren't readily known, review equipment files or contact the manufacturer before making any changes to the process. Also, operation and maintenance should always be performed according to strict standards laid out by a plant standard operating procedure (SOP). Any changes must be reviewed and approved by a cross functional team.

The above hazards and preventive measures have been outlined to illustrate how important safety is in a chemical plant, and the importance of being aware of your surroundings and all possible safety hazards. Many of the aforementioned topics may seem like common sense, but it is very easy to overlook small details in the scope of a large-scale chemical process. Keeping safety in mind at all times as the paramount of any process can ensure that people leave their shift the same way they arrived.

5.5 Alarms in Processes

Alarm configuration and specification is an important part in the design and operation of any chemical process. Alarms are implemented in a process design to aid in the control of the process. Federal and industrial documents only specifically reference alarms in the context of processes exceeding regulatory compliance limits. In this sense, alarms are used to control safety and environmental hazards. Other important uses of alarms are to control product yield, product quality, and operational limits of process equipment. This section will discuss the steps taken to implement alarms in processes, common different levels of alarms, and common instances in which alarms are useful to comply with regulations.

5.5.1 Alarm Lifecycle

When it has been determined that an alarm is needed to aid in the control of a process, the alarm must be specified. The following are basic steps in implementing an alarm.

1. First, the process designer needs to know what category the alarm fits into. This is important because responses from different categories of alarms are usually managed differently. They may be prioritized in case multiple alarms occur at once, so that the proper follow-up reports and procedures may be taken accordingly. Common categories include product quality parameters, safety, environmental considerations, and equipment protection.

2. After determining the specific use of the alarm, the limits must be set. If the process variable exceeds the set limits, the alarm will be triggered. (More information on limits will be discussed in Alarm Levels.)

3. Next, the computer system for the alarm must be configured. The computer system may contain logic loops that automatically change control parameters to offset the problem and merely inform the operator that it has done so. Alternatively, an alarm may trigger horns, flashing lights, or send a page to an operator, alerting them that there is a problem that needs attention.

4. The correct user response and interaction must then be defined for the alarm. This includes providing proper training for the particular process, procedures and operator manuals describing how the event should be investigated, guidelines on when action needs to be taken, and guidelines on when to escalate the situation to a more serious event.

Proper communication from the automated alarm system is critical. Care needs to be taken that alarms only signify abnormal conditions that require a response. For example, the successful completion of a batch operation is an important piece of information and should generate a computer message so that the operator knows. However, this does not represent an abnormal situation and should therefore not show up on the computer as an alarm. While this seems quite obvious, most industrial plants struggle with maintaining alarm systems free of "nuisance alarms." It is also important that alarms have proper descriptions so that the reason an alarm appears is clear. For example, an alarm could appear in a large plant that says, "LI-501 exceeds limits" by default. Although this might be useful to the engineer designing the alarm system as an indication that tank 501 has a high level of material, an operator that sees the alarm or other engineers working on the system may not know what the alarm is communicating. A more universally meaningful alarm indicator might say, "Material level in tank 501 is high." With this simple change in the computer system, the alarm would be more effective for personnel to locate the problem quickly and act accordingly.

5.5.2 Alarm Levels

Alarms are available in a wide variety of types, with multiple levels of alarm. In all processes, disturbances occur that can shift a plant's operation away from normal. When this happens, measures are usually taken by computers, such as with the use of P&ID control loops, to keep the process under control. With these control systems, processes are designed to fall within a range of acceptable normal operating limits. When a process deviates beyond these normal limits, an alarm should be triggered.

For most processes, the minimum for safe operation is two levels of alarms: warning and critical. The warning alarm tells plant operators that the process has deviated beyond the acceptable limits and provides them with the time and ability to take corrective action so that the product quality is not affected and environmental and safety regulations are not exceeded. If the right actions are not taken or are not taken quickly enough to correct the

problem, a critical alarm may then be triggered. The critical alarm tells the plant operators that conditions are dangerously close to breaching what is allowed. In many cases, the critical alarm will call for a systematic shut-down of the operation until the problems can be addressed.

The conditions at which warning and critical alarms are triggered are those conditions that exceed the limits determined for the process. Measurement uncertainty must always be considered because all devices in the control system will be subject to some possible error, even if it is small as +/- 1%. Setting an alarm at exactly the proven acceptable range for the process could allow a measured value to fall within this range, even though the actual value lies outside. This is called a "false acceptance." By performing error analysis and statistical distribution theory, the alarm limits can be adjusted as needed. This is a process called "guard banding," and it prevents real disturbances in the process from being ignored by the alarm system. Information about alarm limits should be well documented so if changes to the system are proposed, designers know how the limits were originally determined.

The figure below provides a visual representation of alarm ranges.
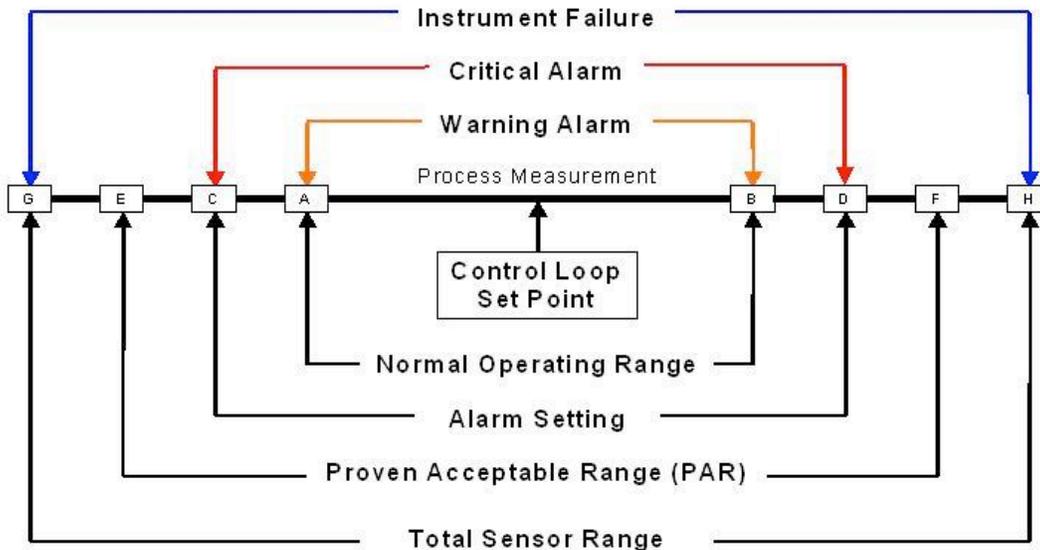
As seen in the center of the above figure, the control loop set point is the optimum point of control for the process (e.g. the optimum temperature and concentrations of reactants for a reaction). It is impossible to maintain the process at exactly this point, so there is a range of "normal" operation, inside which the process is still considered to be running in an acceptable way. The warning alarm would be triggered when the process goes outside of the limits of this range (lower than A or higher than B), allowing time for the process to be brought back under control. The critical alarm would then go off if the process goes beyond the alarm setting (lower than C or higher than D). This setting is determined by guard banding the process acceptable range (PAR) for uncertainty, seen in the figure by the fact that the alarm setting lies well inside the PAR. The spaces between E and C and

between D and F are determined by the uncertainty. Lastly, this PAR must be inside the total range of the sensor that determines the instrument failure.

Alarms must be analyzed based on their priority:

- HIGH	= Hazard to people or equip., action required
- MEDIUM	= Loss of $$, close monitoring required
- LOW	= investigate when time available

SAFETY STEPS TO HIGH RELIABILITY

A safe system involves many layers of reponses when an incident occurs.
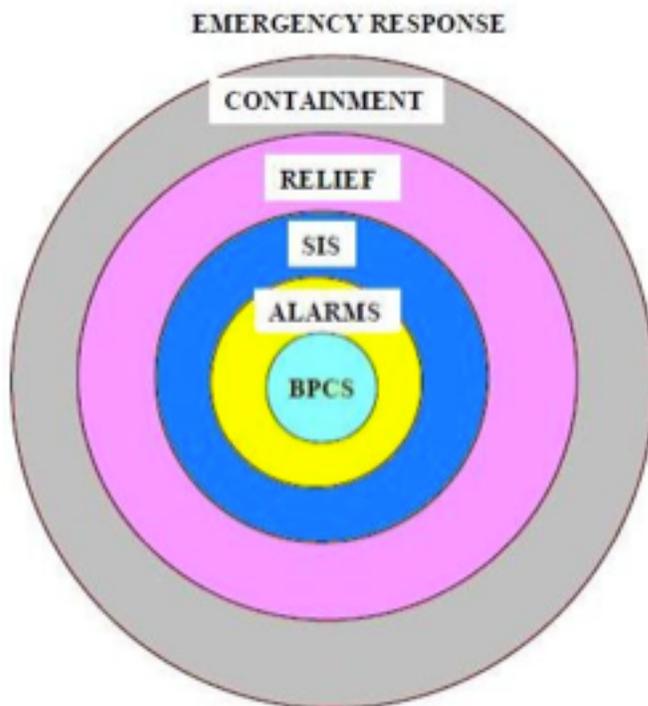
The center of the ring is the basic process control system.

The first layer of response is the alarm system which draws attention.

The second layer is the Safety Interlock System which can stop/start the equipment.

The third layer is the Relief system which leases pressure build-up in the sytem.

The fourth layer is containment which prevents material from reaching workers, community, or the environment.

The last layer to the ring is the emergency response system which involves evacuation, fire fighting, etc.

5.6 Worked out Example 1

A reagent recovery unit for a chemical process plant is being designed. The goal is to recover tin from a tin-plating waste stream through binary extraction with carbon tetrachloride. The equipment used in the process and P&ID are as follows:

- Carbon tetrachloride storage tank
- Mixing Vessel
- Pumps
- Heat Exchanger
- Associated piping
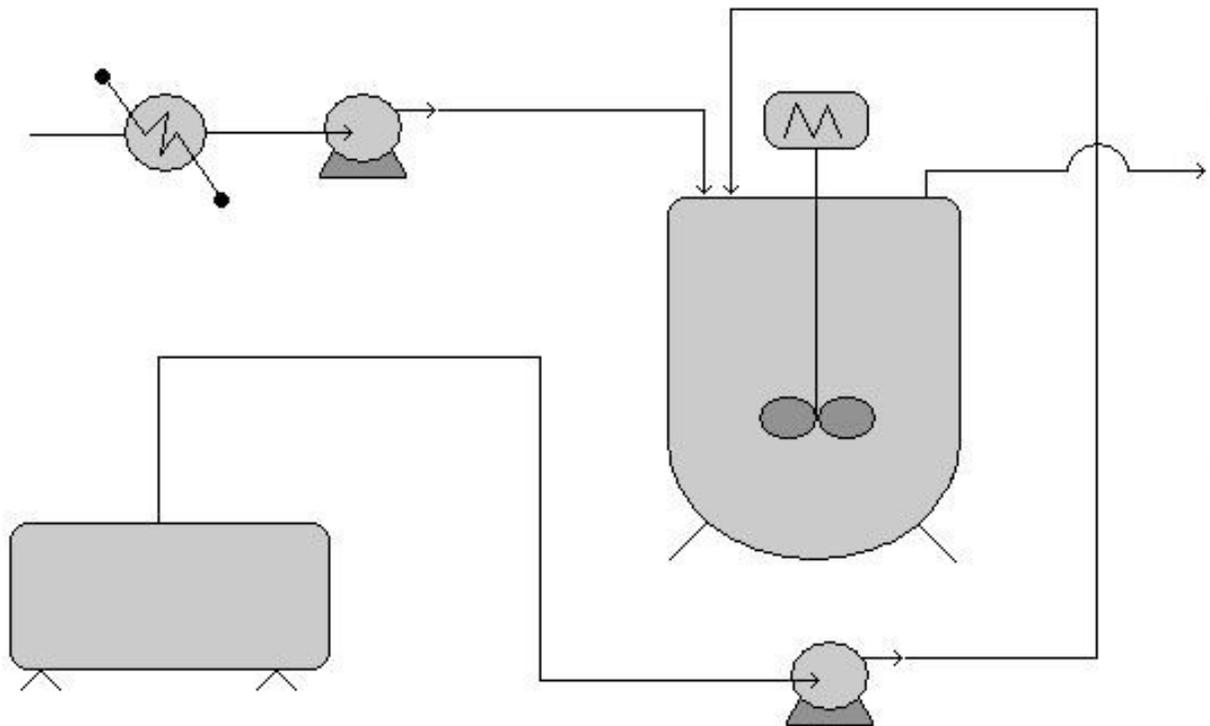
With knowledge of hazardous locations and risk hotspots, and safe design principles,

analyze the flow diagram. Identify areas of risk and specific improvements that should be made to the process design before implementation begins. (note: no piping, valves, or controllers exist on the P&ID for the sake of clarity. Ignore these in your analysis) .

Areas of risk

Any piece of equipment could potentially become dangerous if the right situation were to arise. Each piece of equipment used in this process is no exception.

- Storage tank - These process units that are of secondary importance to the process goal, don't draw as much attention from plant operators and engineers, and tend to receive less maintenance.
- Mixing Vessel - The potential for heat buildup due to heat of mixing and the kinetic energy of the spinning motor makes this vessel a risk hotspot. Improper control of temperature within the vessel or motor speed (especially during periods when the tank may be empty) can lead to equipment malfunction or even explosions
- Pumps and Heat exchangers - The energy stored through pressurization by the pumps and potentially high temperature solutions handled by the heat exchanger present the possibility for danger. Wherever energy is stored, danger is associated with its potential release.
- Associated piping - Corrosion or failure to maintain pipes and associated elements are the main cause of the malfunction leading to danger. Leaks or total failures can release reactive materials on other equipment causing corrosion or malfunction

Specific Improvements

This process could be improved and made safer in the following ways.

- Simplification of the piping used to connect the storage tank to the mixing vessel will reduce the risk of leaks and malfunctions due to piping complications.
- The addition of a emergency relief valve and temperature controlling/insulating element to the mixing vessel will greatly reduce the risk of equipment failure due to overheating or overpressureization due to heats of mixing or outside heating influences.
- Use a reactor with a rounded top, as this type can withstand much higher pressures.
- The substitution of carbon tetrachloride with a less hazardous solvent, if the specific solubility required to extract the tin still exists, like cyclohexane, isopropyl alcohol, or 1,1,1-trichloroethane reduces the risk of health complications due to exposure, and possible explosions.

5.7 Example 2: Alarms in P&ID

A P&ID appears below for the production of a solution containing a pharmaceutically active compound. The reaction taking place in the CSTR is highly exothermic. After examining the P&ID for this part of the process, describe a possible alarm system.
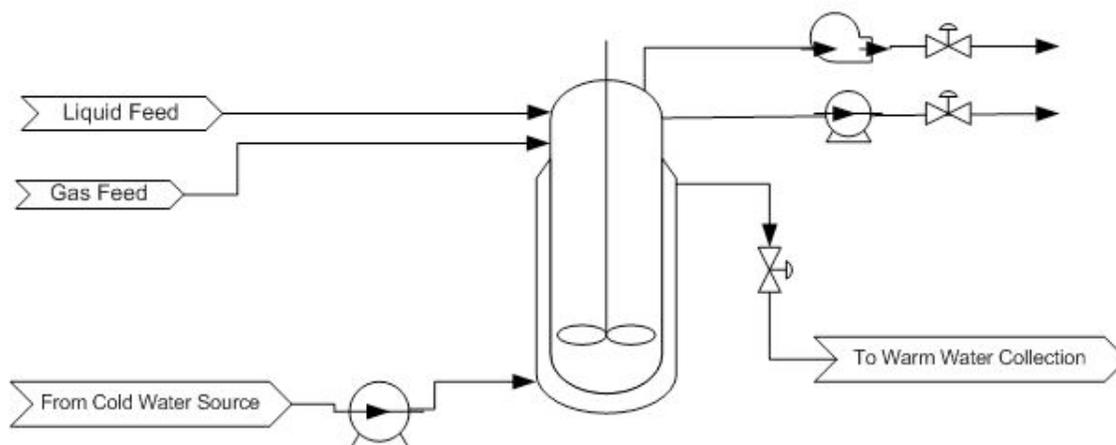
Answer: The CSTR for the exothermic reaction is jacketed with a cooling water stream. An alarm should be in place to monitor the reactor temperature. A warning alarm can notify the operator that the temperature is too high and corrective action needs to be taken. A critical alarm should be in place to warn that the reactor is nearing runaway conditions and an immediate response is needed. If the necessary action is not taken, systematic shutdown of the reactor could occur. This would involve closing the valves, flooding the jacket with cooling water, and having the impeller on. Another possibility for an alarm, although we do not know how the products are being used specifically, is in a composition measurement of the product containing the pharmaceutically active compound. Depending on where this stream is going and how it is being used, too high a concentration could be dangerous if no other concentration-altering steps occur before the finished product goes out to consumers.

5.8 Actual Case

The bulletin below is taken for the US Chemical Safety and Hazard Invesitgation Board

Washington, DC, July 15, 2004 - The U.S. Chemical Safety and Hazard Investigation Board (CSB) today released a Safety Bulletin calling on chemical plant and refinery operators to exercise appropriate caution when performing work on piping and equipment that could contain hazardous materials.

The bulletin, set to be considered by the Board at a public meeting today, results from the CSB's investigation into a January 13, 2004, explosion and fire at the Huntsman

Petrochemical facility in Port Neches, Texas, where two employees were seriously burned and significant damage occurred to nearby equipment. The explosion and fire occurred as workers attempted to purge a thousand-foot-long chemical process pipe in preparation for a cutting and welding operation.

Huntsman managers and workers were aware of the importance of completely removing hazardous material before cutting into the piping, and relevant warnings were contained in the company's written operating procedures. In this incident, workers first purged the piping with nitrogen to force out residual chemicals, including a hazardous mixture of peroxide and alcohol that reacts violently when heated. But unknown to the workers, the piping included a 300-foot-long section that was three feet lower than the rest of the piping, and despite the nitrogen purge, a significant amount of the hazardous mixture remained trapped. The next step in the operation was to use high-temperature steam to purge the piping of what workers believed would be a small amount of residual flammable hydrocarbon vapor. But the steam heated the peroxide that was trapped in the low section of piping. The peroxide then began to decompose, releasing heat and creating intense pressure. The pressure blew out a valve gasket and violently ruptured the pipe. Flammable vapors shot out of the openings and ignited into a large fireball, injuring plant workers.

After the accident, Huntsman found two drains in the low section of the pipe, which could have been used to remove the trapped liquid. Had Huntsman's procedures called for reviewing plant pipe drawings and physically walking the entire line within the work boundaries, the accident would likely have been avoided, CSB said.

Lessons Learned for this incident:

- *Physically examine all piping and components between isolation devices such as valves, and be sure piping drawings are current*
- *Use the drawings to identify key components, such as low-point drains that can be used to remove dangerous chemicals*
- *Prepare a specific written procedure for removing hazardous material and consider the consequences of working on piping that is not completely purged."*

5.9 Sage's Corner

Case Study: Texas City Refinery Disaster	Got Explosion?
http://www.youtube.com/v/jGLG1 2ymQEM	http://video.google.com/googleplayer.swf ?docId=-5627107926354267537

| Unnarrated Version: PID_Safety_No_Narr.ppt | Click the link to watch a movie about plant safety! |

5.10 References

- AIChE (January 2006). Process Safety Beacon, CEP Magazine
- AIChE (March 2006). Process Safety Beacon, CEP Magazine
- AIChE (May 2006). Process Safety Beacon, CEP Magazine
- AIChE (November 2003). Process Safety Beacon, CEP Magazine
- AIChE (September 2006). Process Safety Beacon, CEP Magazine
- Gael D. Ulrich, Palligarnai T. Vasudevan (2006). Predesign With Safety in Mind, Univ. Of New Hampshire
- Georges A. Melhem, Peter Howell (2005). Designing Emergency Relief Systems for Runaway Reactions, CEP Magazine
- Joseph S. Alford, John Kindervater, Robert Stankovich (April 2005). Alarm Management for Regulated Industries, CEP Magazine
- U.S. Chemical Safety and Hazard Investigation Board, 15 July 2004. CSB News Release
-
  http://www.csb.gov/index.cfm?folder=news_releases&page=news&NEWS_ID=171

Section 6. Regulatory Agencies and Compliance

Title: Piping and Instrumentation Diagrams: Regulatory Agencies and Compliance
Authors: Jayshree Mahajan, Bradley Anderson, Aaron Bennick, Michael Salciccioli
Stewards: Jocelyn Anleitner, Stephanie Combs, Diane Feldkamp, Heeral Sheth

6.1 Introduction

Regulatory agencies govern various sections of chemical plants from emissions and water treatment to personnel behaviors and company policy. Understanding of the types of organizations that demand high performance from chemical plants on a national and local level are extremely important for the safety of the employees, community and environment.

6.2 Regulatory Compliance

Compliance is an integral part of ensuring the safety of all that work in close contact with the chemical plant as well as minimizing fines and fees that come with violating regulations. Compliance can be organized into two main categories; plant safety and environmental safety. Plant safety ensures employees and the surrounding community are adequately protected during full-operation of the plant. Environmental safety comprises protecting and treating the environmental carefully when it comes to various contaminated streams in the plant.

6.2.1 Plant Safety for Employees and the Community

Safety is paramount in any chemical process, and alarms are essential for compliance with safety constraints. OSHA has established guidelines that must be followed in any plant when dealing with chemicals defined as highly hazardous. Limits for temperature, pressure, flow rates, and compositions need to be regulated. Alarms should be in place to warn operators when a limit is near so that steps can be taken to ensure the safety of people in the plant. An example of this is a safe temperature limit for a CSTR. The alarm signals that the temperature is too high and action is needed to prevent a runaway scenario. If the corrective action is not taken, or not taken quickly enough, a critical alarm can signal a computer program to automatically shut down the entire process or specific unit operation.

Some industries may expect product quality to be closely regulated by the FDA or other government agencies. Typically, this will be in a process where the final product is directly used by people and the margins for error are small. These processes include food processing and manufacturing consumer products, especially pharmaceuticals. These industries usually require systems in place that frequently validate alarms, as well as documentation for all critical alarm events. Measurements such as the weight percent of a pharmaceutically active compound in a solution must be carefully monitored, with recorded uncertainty analysis.

428

6.2.2 Environmental Safety

Emissions of solids, liquids, and gases in a plant are heavily regulated by government agencies. Regulations apply for processes that emit chemicals to the atmosphere (either directly or following a scrubber), processes that discharge material into a body of water, or processes that require containment control devices like check valves and rupture disks. Alarms are frequently used to comply with these regulations by measuring things such as pH and organic solvent concentration. Typically, a warning alarm will alert personnel that a threshold may be breached if action is not taken, allowing time to avoid an incident requiring a formal report. Critical alarms can alert operators that a threshold has been passed and automatically trigger the appropriate action, such as a systematic shutdown of the process.

6.3 Regulatory Agencies

Federal and national agencies maintain smaller state subsections of various programs and administrations. In most cases, the state level requires stricter compliance and lower limits. Solely state-controlled programs usually handle the air and water quality since any regulation violation results in the consequences for the immediate community.

6.3.1 Federal and National Agencies or Programs

Plant safety and environmental safety programs regulated on the national or federal level are monitored by three main agencies; the Environmental Protection Agency (EPA), the U.S. Department of Labor, the Food and Drug Administration, and the Department of Homeland Security. These three governing bodies have created numerous acts, committees, administrations and policies that protect the welfare of the employee, community, and environment.

The EPA at the federal level, provides acts, laws, and regulations, that help maintain and improve the air and water quality. The risk management program (RMP) is a mandatory program that "*require facilities that produce, handle, process, distribute, or store certain chemicals to develop a Risk Management Program*". A Risk Management Plan (RMP) must be submitted to the EPA for approval. Overall, risk management is a large part of process control as control systems must adequately function and maintain compliance of an entire facility. Failure logic for instrumentation, redundant sensors, and critical alarms are essential in maintaining compliance, but a RMP is crucial for handling low-likelihood emergency situations.

The U.S. Department of Labor maintains the Occupational Safety and Health Administration (OSHA) which provides rules and regulations for employers and employees on safe workplace practices. Although individual states may maintain their own occupational health and safety plans, OSHA is the governing body and authority on those programs. Inspections are performed to ensure that all employees have a clean and safe working environment that is hazard-free and risk-mediated. OSHA also maintains the Process Safety Management Program (PSM) which regulates requirements for

facilities that handle highly hazardous chemicals. A [list of chemicals that qualify](#) include chlorine, formaldehyde, and hydroflouric acid. Requirements of PSM include frequent process hazard analysis (PHA), pre-startup safety review (PSSR), and incident investigations. A cooperative program that OSHA maintains is The Voluntary Protection Programs (VPP). These programs aim to bring together management and labor to provide comprehensive safety and health guidelines and regulations that keep all employees safe while on the job. Each facility must submit application for entry into the program and upon acceptance will follow a set of standards to ensure continued safety. The following is a detailed description of why PSM should be important to chemical engineers, and also some highlights of the main aspects of PSM as required by OSHA.

6.3.2 Process Safety Management (PSM) -- Why is this important to chemical engineers?

See the following links to videos provided by the United States Chemical Safety Board describing a few catastrophic events that have occurred in the chemical engineering industry:

- **[2005 Texas City Refinery Explosion](#)**
- **[Bhopal Disaster](#)**

Besides the catastrophic nature of events that can occur from neglecting Process Safety, large chemical facilities are granted a privilege and license to operate by the different federal regulatory agencies. If these regulatory agencies perform an audit on a specific facility, and find that their regulations are not being followed, then extremely large fines can be levied to the company, even to the extent of shutting the facility down permanently by removing that facility's privilege to operate. In the case of PSM, it is OSHA who deals out these fines. For example, in 2009 OSHA attempted to levy a record *87 million* dollar fine to an integrated oil company, which has not been finalized in the legal system yet, but gives a good example of how important it is for companies operating in the U.S., if they want to continue to operate safely and economically, to follow all government regulations as closely as possible.

Unexpected releases of toxic, reactive, or flammable liquids and gases in processes involving highly hazardous chemicals have been reported for many years in various industries that use chemicals with such properties. Regardless of the industry that uses these highly hazardous chemicals, there is a potential for an accidental release any time they are not properly controlled, creating the possibility of disaster. As a result of catastrophic incidents in the past, and to help ensure safe and healthful workplaces, OSHA has issued the Process Safety Management of Highly Hazardous Chemicals standard (29 CFR 1910.119), which contains requirements for the management of hazards associated with processes using highly hazardous chemicals. OSHA's standard 29CFR 1910.119 emphasizes the management of hazards associated with highly hazardous chemicals and establishes a comprehensive management program that integrates technologies, procedures, and management practices. A detailed list of these standards can be found on the United States Department of Labor website: http://www.osha.gov/SLTC/processsafetymanagement/standards.html An effective

process safety management program requires a systematic approach to evaluating the whole process. Using this approach the process design, process technology, operational and maintenance activities and procedures, training programs, and other elements which impact the process are all considered in the evaluation. Process safety management is the proactive identification, evaluation and mitigation or prevention of chemical releases that could occur as a result of failures in process, procedures or equipment. OSHA prescribes essential tools to the success of process safety management including:

- **Process Safety Information**
- **Process Hazard Analysis**
- **Operating Procedures and Practices**
- **Employee Training**
- **Pre-Startup Safety Review**
- **Mechanical Integrity**
- **Management of Change**
- **Incident Investigation**
- **Emergency Preparedness**
- **Compliance Audits**

The thought is, with the simultaneous implementation of all of these things at a facility dealing with large amounts of highly hazardous chemicals, the risk of a catastrophic incident resulting from an unplanned release will be minimized. Following is a detailed discussion of each of these tools prescribed by OSHA.

Process Safety Information (PSI)

Complete, accurate, and up-to-date written information concerning process chemicals, process technology, and process equipment is essential to an effective process safety management program. The compiled information will be a necessary resource to a variety of users including the team that will perform the process hazards analysis, those developing the training programs and operating procedures, contractors whose employees will be working with the process, those conducting the pre-startup safety reviews, local emergency preparedness planners, and insurance and enforcement officials. PSI includes, but is not limited to:

- Material and safety data sheets (MSDS)
- A block flow diagram showing the major process equipment and interconnecting process flow lines
- Process Flow Diagrams (PFDs)
- Piping and Instrument Diagrams (P&IDs)
- Process design information, including the codes and standards relied on to establish good engineering design

Process Hazards Analysis (PHA)

431

A process hazards analysis (PHA) is one of the most important elements of the process safety management program. A PHA is an organized and systematic effort to identify and analyze the significance of potential hazards associated with the processing and handling of highly hazardous chemicals. A PHA is directed toward analyzing potential causes and consequences of fires, explosions, releases of toxic or flammable chemicals, and major spills of hazardous chemicals. The PHA focuses on equipment, instrumentation, utilities, human actions, and external factors that might impact the process. These considerations assist in determining the hazards and potential failure points or failure modes in a process.

A team from each process unit in the facility will be tasked with conducting a PHA for their process unit at regularly scheduled intervals as defined by OSHA. One example is in an oil refinery, where a PHA has to be conducted and documented for each process unit every five calendar years. The competence of the team conducting the PHA is very important to its success. A PHA team can vary in size from two people to a number of people with varied operational and technical backgrounds. The team leader needs to be fully knowledgeable in the proper implementation of the PHA methodology that is to be used and should be impartial in the evaluation. The other full or part time team members need to provide the team with expertise in areas such as process technology, process design, operating procedures and practices, alarms, emergency procedures, instrumentation, maintenance procedures, safety and health, and any other relevant subject as the need dictates. The ideal team will have an intimate knowledge of the standards, codes, specifications and regulations applicable to the process being studied.

There are various methodologies for conducting a PHA. Choosing which one is right for each individual facility will be influenced by many factors, including the amount of existing knowledge about the process. For more information on the different methodologies for conducting a PHA, see Center for Chemical Process Safety of the American Institute of Chemical Engineers

Operating Procedures

Operating procedures provide specific instructions or details on what steps are to be taken or followed in carrying out the task at hand. The specific instructions should include the applicable safety precautions and appropriate information on safety implications. For example, the operating procedures addressing operating parameters will contain operating instructions about pressure limits, temperature ranges, flow rates, what to do when an upset condition occurs, what alarms and instruments are pertinent if an upset condition occurs, and other subjects. Another example of using operating instructions to properly implement operating procedures is in starting up or shutting down the process.

Operating procedures and instructions are important for training operating personnel. The operating procedures are often viewed as the standard operating practices (SOPs) for operations. Control room personnel and operating staff, in general, need to have a full

understanding of operating procedures. In addition, operating procedures need to be changed when there is a change in the process. The consequences of operating procedure changes need to be fully evaluated and the information conveyed to the personnel. For example, mechanical changes to the process made by the maintenance department (like changing a valve from steel to brass or other subtle changes) need to be evaluated to determine whether operating procedures and practices also need to be changed. All management of change actions must be coordinated and integrated with current operating procedures, and operating personnel must be alerted to the changes in procedures before the change is made. When the process is shut down to make a change, the operating procedures must be updated before re-starting the process.

Employee Training

All employees, including maintenance and contractor employees involved with highly hazardous chemicals, need to fully understand the safety and health hazards of the chemicals and processes they work with so they can protect themselves, their fellow employees, and the citizens of nearby communities. Training conducted in compliance with the OSHA Hazard Communication standard (Title 29 Code of Federal Regulations (CFR) Part 1910.1200) will inform employees about the chemicals they work with and familiarize them with reading and understanding MSDSs. However, additional training in subjects such as operating procedures and safe work practices, emergency evacuation and response, safety procedures, routine and non-routine work authorization activities, and other areas pertinent to process safety and health need to be covered by the employer's training program.

In establishing their training programs, employers must clearly identify the employees to be trained, the subjects to be covered, and the goals and objectives they wish to achieve. The learning goals or objectives should be written in clear measurable terms before the training begins. These goals and objectives need to be tailored to each of the specific training modules or segments. Employers should describe the important actions and conditions under which the employee will demonstrate competence or knowledge as well as what is acceptable performance.

Careful consideration must be given to ensure that employees, including maintenance and contract employees, receive current and updated training. For example, if changes are made to a process, affected employees must be trained in the changes and understand the effects of the changes on their job tasks. Additionally, as already discussed, the evaluation of the employee's absorption of training will certainly determine the need for further training.

Pre-Startup Safety Review

For new processes, the employer will find a PHA helpful in improving the design and construction of the process from a reliability and quality point of view. The safe operation of the new process is enhanced by making use of the PHA recommendations before final installations are completed. P&IDs should be completed, the operating procedures put in place, and the operating staff trained to run the process, before startup. The initial startup procedures and normal operating procedures must be fully evaluated as part of the pre-startup review to ensure a safe transfer into the normal operating mode.

For existing processes that have been shut down for turnaround or modification, the employer must ensure that any changes other than "replacement in kind" made to the process during shutdown go through the management of change procedures. P&IDs will need to be updated, as necessary, as well as operating procedures and instructions. If the changes made to the process during shutdown are significant and affect the training program, then operating personnel as well as employees engaged in routine and non-routine work in the process area may need some refresher or additional training. Any incident investigation recommendations, compliance audits, or PHA recommendations need to be reviewed to see what affect they may have on the process before beginning the startup.

Mechanical Integrity

Employers must review their maintenance programs and schedules to see if there are areas where "breakdown" is used rather than the more preferable on-going mechanical integrity program. Equipment used to process, store, or handle highly hazardous chemicals has to be designed, constructed, installed, and maintained to minimize the risk of releases of such chemicals. This requires that a mechanical integrity program be in place to ensure the continued integrity of process equipment.

Elements of a mechanical integrity program include identifying and categorizing equipment and instrumentation, inspections and tests and their frequency; maintenance procedures; training of maintenance personnel; criteria for acceptable test results; documentation of test and inspection results; and documentation of manufacturer recommendations for equipment and instrumentation.

Management of Change

To properly manage changes to process chemicals, technology, equipment and facilities, one must define what is meant by change. In the process safety management standard, change includes all modifications to equipment, procedures, raw materials, and processing conditions other than "replacement in kind." These changes must be properly managed by identifying and reviewing them prior to implementing them. For example, the operating procedures contain the operating parameters (pressure limits, temperature

434

ranges, flow rates, etc.) and the importance of operating within these limits. While the operator must have the flexibility to maintain safe operation within the established parameters, any operation outside of these parameters requires review and approval by a written management of change procedure. Management of change also covers changes in process technology and changes to equipment and instrumentation. Changes in process technology can result from changes in production rates, raw materials, experimentation, equipment unavailability, new equipment, new product development, change in catalysts, and changes in operating conditions to improve yield or quality. Equipment changes can be in materials of construction, equipment specifications, piping pre-arrangements, experimental equipment, computer program revisions, and alarms and interlocks. Employers must establish means and methods to detect both technical and mechanical changes.

Temporary changes have caused a number of catastrophes over the years, and employers must establish ways to detect both temporary and permanent changes. It is important that a time limit for temporary changes be established and monitored since otherwise, without control, these changes may tend to become permanent. Temporary changes are subject to the management of change provisions. In addition, the management of change procedures are used to ensure that the equipment and procedures are returned to their original or designed conditions at the end of the temporary change. Proper documentation and review of these changes are invaluable in ensuring that safety and health considerations are incorporated into operating procedures and processes. Employers may wish to develop a form or clearance sheet to facilitate the processing of changes through the management of change procedures. A typical change form may include a description and the purpose of the change, the technical basis for the change, safety and health considerations, documentation of changes for the operating procedures, maintenance procedures, inspection and testing, P&IDs, electrical classification, training and communications, pre-startup inspection, duration (if a temporary change), approvals, and authorization. Where the impact of the change is minor and well understood, a check list reviewed by an authorized person, with proper communication to others who are affected, may suffice. For a more complex or significant design change, however, a hazard evaluation procedure with approvals by operations, maintenance, and safety departments may be appropriate. Changes in documents such as P&IDs, raw materials, operating procedures, mechanical integrity programs, and electrical classifications should be noted so that these revisions can be made permanent when the drawings and procedure manuals are updated. Copies of process changes must be kept in an accessible location to ensure that design changes are available to operating personnel as well as to PHA team members when a PHA is being prepared or being updated.

Incident Investigation

Incident investigation is the process of identifying the underlying causes of incidents and implementing steps to prevent similar events from occurring. The intent of an incident investigation is for employers to learn from past experiences and thus avoid repeating

past mistakes. The incidents OSHA expects employers to recognize and to investigate are the types of events that resulted in or could reasonably have resulted in a catastrophic release. These events are sometimes referred to as "near misses," meaning that a serious consequence did not occur, but could have.

Employers must develop in-house capability to investigate incidents that occur in their facilities. A team should be assembled by the employer and trained in the techniques of investigation including how to conduct interviews of witnesses, assemble needed documentation, and write reports. A multi-disciplinary team is better able to gather the facts of the event and to analyze them and develop plausible scenarios as to what happened, and why. Team members should be selected on the basis of their training, knowledge and ability to contribute to a team effort to fully investigate the incident.

Emergency Preparedness

Each employer must address what actions employees are to take when there is an unwanted release of highly hazardous chemicals. Emergency preparedness is the employer's third line of defense that will be relied on along with the second line of defense, which is to control the release of chemical. Control releases and emergency preparedness will take place when the first line of defense to operate and maintain the process and contain the chemicals fails to stop the release.

Employers will need to select how many different emergency preparedness or third lines of defense they plan to have, develop the necessary emergency plans and procedures, appropriately train employees in their emergency duties and responsibilities, and then implement these lines of defense. Employers, at a minimum, must have an emergency action plan that will facilitate the prompt evacuation of employees when there is an unwanted release of a highly hazardous chemical. This means that the employer's plan will be activated by an alarm system to alert employees when to evacuate, and that employees who are physically impaired will have the necessary support and assistance to get them to a safe zone. The intent of these requirements is to alert and move employees quickly to a safe zone. Delaying alarms or confusing alarms are to be avoided. The use of process control centers or buildings as safe areas is discouraged. Recent catastrophes indicate that lives are lost in these structures because of their location and because they are not necessarily designed to withstand overpressures from shock waves resulting from explosions in the process area.

Preplanning for more serious releases is an important element in the employer's line of defense. When a serious release of a highly hazardous chemical occurs, the employer, through preplanning, will have determined in advance what actions employees are to take. The evacuation of the immediate release area and other areas, as necessary, would be accomplished under the emergency action plan. If the employer wishes to use plant personnel-such as a fire brigade, spill control team, a hazardous materials team-or employees to render aid to those in the immediate release area and to control or mitigate

the incident, refer to OSHA's Hazardous Waste Operations and Emergency Response (HAZWOPER) standard (Title 79 CFR Part 1910.1 20). If outside assistance is necessary, such as through mutual aid agreements between employers and local government emergency response organizations, these emergency responders are also covered by HAZWOPER. The safety and health protection required for emergency responders is the responsibility of their employers and of the on-scene incident commander.

Compliance Audits

An audit is a technique used to gather sufficient facts and information, including statistical information, to verify compliance with standards. Employers must select a trained individual or assemble a trained team to audit the process safety management system and program. A small process or plant may need only one knowledgeable person to conduct an audit. The audit includes an evaluation of the design and effectiveness of the process safety management system and a field inspection of the safety and health conditions and practices to verify that the employer's systems are effectively implemented. The audit should be conducted or led by a person knowledgeable in audit techniques who is impartial towards the facility or area being audited. The essential elements of an audit program include planning, staffing, conducting the audit, evaluating hazards and deficiencies and taking corrective action, performing a follow-up review, and documenting actions taken.

6.3.3 Other Federal Entities

The Food and Drug Administration (FDA) is an agency of the United States Department of Health and Human Services. The FDA regulates food, drugs, cosmetics, biologics, medical devices, radiation-emitting devices and vetenary products manufactured in the United States. The main goal of the FDA is to maitain that the products they regulate are safe, effective, and secure. The FDA is also responsible that the products are accurately represented to the public. State and local governments also help regulate these products in cooperation with the FDA. The FDA does not regulate alcohol, illegal drugs, and meat and poultry.

The Department of Homeland Security has recently taken a role in regulating chemical plants because of 9/11. Chemical plants are seen by the government as targets for terrorists and security in and around the plant is a major concern. Although the laws are typically state run, the Department of Homeland Security has required mandatory national security standard to chemical plants throughout the nation. Although the mandates were fought by legislative for years, the Department of Homeland Security has influence in the security in chemical plants. The law requires the plant to prepare a

vulnerability test and submit a site security plan. In order to validate these activities, audits and site visits are/will be performed by government officials.

6.3.4 State Regulations

State regulations vary greatly from state-to-state depending on the main concern. For instance, beach quality is more important in California than Nebraska due to the geographical location. Similarly state legislative acts and administrations can expect high performance from the chemical industry as it has the capability to affect all parts of a community for many generations.

6.4 References

- U.S. National Profile on the Management of Chemicals. http://www.epa.gov/oppfead1/profile/
- U.S. Department of Labor. Occupational Safety & Health Administration. Process Safety Management. http://www.osha.gov/SLTC/processsafetymanagement/index.html
- U.S. Department of Labor. Occupational Safety & Health Administration. State Occupational Health and Safety Plans. http://www.osha.gov/dcsp/osp/index.html
- U.S. Department of Labor. Voluntary Protection Programs. http://www.osha.gov/dcsp/vpp/index.html
- Environmental Protection Agency. Regulatory Information by Business Sector. http://www.epa.gov/lawsregs/bizsector/index.html
- Environmental Protection Agency. Regulatory Information by Environmental Topic. http://www.epa.gov/lawsregs/envtopics/index.html#air
- Environmental Protection Agency. State Specific Regulatory Information Search Engine. http://www.epa.gov/lawsregs/where/index.html
- Environmental Protection Agency. Risk Management Plan. http://www.epa.gov/oem/content/rmp/index.htm
- United States Chemical Safety and Hazard Regulation Board. Homepage. http://www.chemsafety.gov/
- U.S. Department of Health & Human Services. Food and Drug Administration. http://www.fda.gov/
- United States Chemical Safety Board (CSB). http://www.csb.gov
- Center for Chemical Process Safety (CCPS). http://www.aiche.org/ccps

Chapter 5. Logical Modeling

Section 1. Boolean models: truth tables and state transition diagrams

Title: Boolean Models

Authors: Joseph Casler, Andry Haryanto, Seth Kahle and Weiyin Xu
Date Presented: Thursday, September 26, 2006
Date Revised: Thursday, September 21, 2006
Stewards: (September 12, 2007) Adhi Paisoseputra, Andrew Kim, Hillary Kast, Stephanie Cleto

- First round reviews for this page
- Rebuttal for this page

1.1 Introduction to Boolean Models

A Boolean is a variable that can only attain two values: True or False. In most applications, it is convenient to represent a True by the number 1, and a False by the number 0. A **Boolean model**, or **Boolean network**, is a collection of Boolean variables that are related by logical switching rules, or **Boolean functions**, that follow an If-Then format. This type of Boolean model is known as an autonomous model and will be the primary type of model discussed in this article.

In chemical engineering, Boolean models can be used to model simple control systems. Boolean functions can be used to model switches on pumps and valves that react to readings from sensors that help keep that system operating smoothly and safely.

A simple application for level control of a CSTR is included in worked-out example 1. In this example, Boolean function is used to close the inlet stream and open the outlet stream when the level is higher than a specified point.

1.1.1 Boolean Functions

Boolean functions are logical operators that relate two or more Boolean variables within a system and return a true or false. A Boolean expression is a group of Boolean functions, which will be described individually below. When computing the value of a Boolean expression, Parentheses are used to indicate priority (working from inside out as in algebra). After that, LOGICAL INVERSION will always be first and LOGICAL EQUIVALENCE will be last, but the order of operation for the AND, OR, and EXCLUSIVE OR functions are specified with parenthesis.
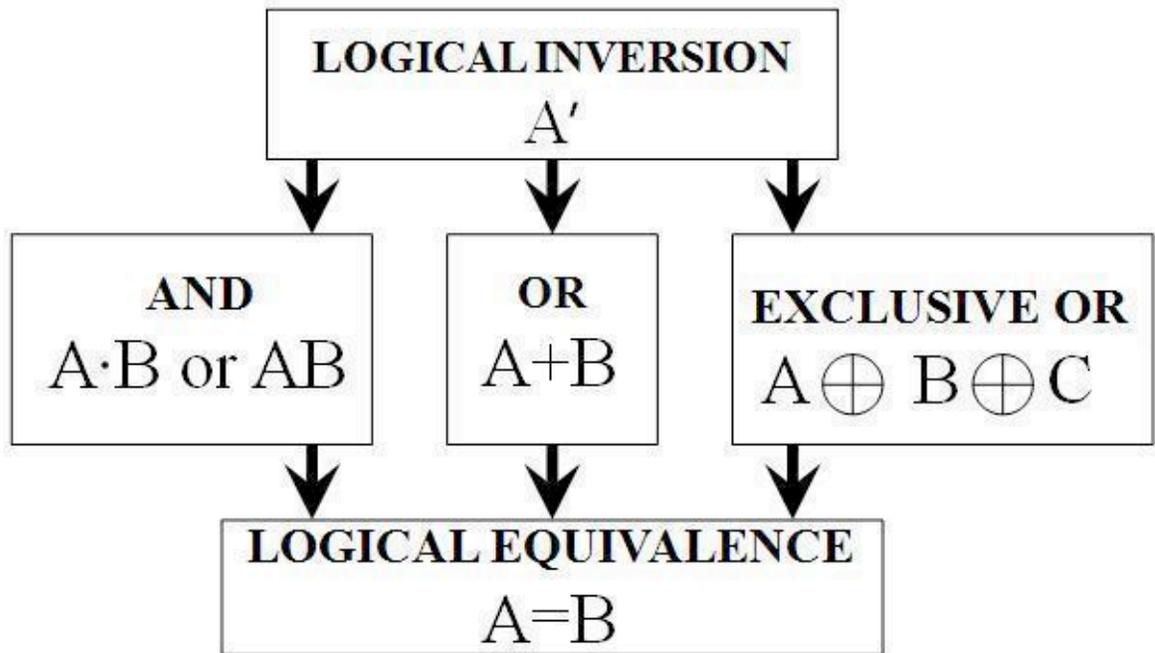
Descriptions and examples of these functions are given below. A quick reference of each of the functions can be found after the examples.

LOGICAL INVERSION

LOGICAL INVERSION is a function that returns the opposite value of a variable. The function is denoted as a prime on the variable (e.g. A' or B') For example, if we say that A is true (A=1), then the function A' will return a false (A'=0). Similarly, if we say that A is false (A=0) then the function A' will return true (A'=1).

Example:
A=1, B=A' then B=0

AND

The AND function relates two or more Boolean variables and returns a true if-and-only-if both variables are true. A dot is used to denote the AND function, or it is simply omitted. For example "A and B" can be written as "A•B" or as "AB." In this example, the AND function will only return a true if-and-only-if both Boolean variables A and B have a value of 1.

Examples:

Variables	Results

440

A=1, B=1	AB = 1
A=1, B=0	AB = 0
A=1, B=1, C=1	ABC = 1
A=1, B=0, C=1	ABC = 0

OR

The OR function relates two or more Boolean variables and returns a true if any referenced variables are true. A plus is used to denote the OR function. For example "A or B" can be written as "A+B." In this example, the OR function will return true if either Boolean variable, A or B, has a value of 1.

Examples:

Variables	Results
A=1, B=1	A+B = 1
A=1, B=0	A+B = 1
A=0, B=0	A+B = 0
A=0, B=0, C=1	A+B+C = 1
A=0, B=0, C=0	A+B+C = 0

EXCLUSIVE OR

The EXCLUSIVE OR function relates two or more Boolean variables and returns **true** only when **one** of the variables is true and **all** other variables are false. It returns **false** when **more than one** of the variables are true, or **all** the variables are false. A circumscribed plus is used to denote the EXCLUSIVE OR function. For example "A EXCLUSIVE OR B" can be written as "A $\oplus$ B."

Examples:

Variables	Results
A=1, B=1	A $\oplus$ B = 0
A=1, B=0	A $\oplus$ B = 1
A=0, B=1	A $\oplus$ B = 1

A=0, B=0	$A \oplus B = 0$
A=0, B=0, C=0	$A \oplus B \oplus C = 0$
A=1, B=0, C=0	$A \oplus B \oplus C = 1$
A=1, B=0, C=1	$A \oplus B \oplus C = 0$
A=1, B=1, C=1	$A \oplus B \oplus C = 0$

LOGICAL EQUIVALENCE

The LOGICAL EQUIVALENCE function equates two Boolean variables or expressions. The LOGICAL EQUIVALENCE function, denoted as =, assigns a Boolean variable a true or false depending on the value of the variable or expression that it is being equated with. For example, A LOGICAL EQUIVALENCE B can be written as A = B. In this example, the value of A will be assigned the value of B.

1.1.2 Boolean Networks

As stated in the introduction, a Boolean network is a system of boolean equations. In chemical engineering, Boolean networks are likely to be dependant on external inputs as a means of controlling a physical system. However, the following sections pertain mostly to synchronous autonomous systems. An autonomous system is one that is completely independent of external inputs. Every Boolean variable is dependent on the state of other Boolean variables in the system and no variable is controlled by an external input. A synchronous system is one that logical switching (the changing of Boolean variables) occurs simultaneously for all variables based on the values prior to the incidence of change.

Here is an example of an autonomous boolean network:

Boolean Functions
A = B+C'
B = AC
C = A'

1.2 Truth Tables

A truth table is a tabulation of all the possible states of a Boolean Model at different time frames. A simple truth table shows the potential initial states at time, T_i, and the

corresponding subsequent states at time T_{i+1}, of a Boolean network. Truth tables can provide one with a clearer picture of how the rules apply and how they affect each situation. Hence, they help to ensure that each output only has one control statement so that the Boolean rules do not conflict with each other.

1.2.1 Constructing Truth Tables

1) Draw up a table with the appropriate number of columns for each variable; one for each input and output.
2) The left side of the column should contain all possible permutations of the input variables at time T_i. One method to accomplish this might be to list all possible combinations in ascending binary order.
3) The right side of the column should contain the corresponding outcome of the output variables at the subsequent time T_{i+1}. A generic example of this with 2 variables can be seen below:

Inputs		Outputs	
T_i		T_{i+1}	
variable1	variable2	variable1	variable2

A quick way to check that you have all of the possible permutations is that there should be 2^x possible permutations for X input variables.

1.2.2 Example of a Truth Table

The sample system we will be using is based on hydrogen fuel cell technology. The equation for the operation of hydrogen fuel cells is $H_2 + O_2 \rightarrow H_2O$. One aspect of Proton Exchange Membrane (PEM) fuel cells (a type of fuel cell) is that the performance of the fuel cell is highly dependent on the relative humidity of the system (if humidity rises too high, the fuel cell will flood and H_2 and O_2 will be unable to reach the cell. If humidity falls too low, the fuel cell will dry up and the performance will drop.) The task is to create a Boolean model for this simplified water management system.

PEM Fuel Cell System
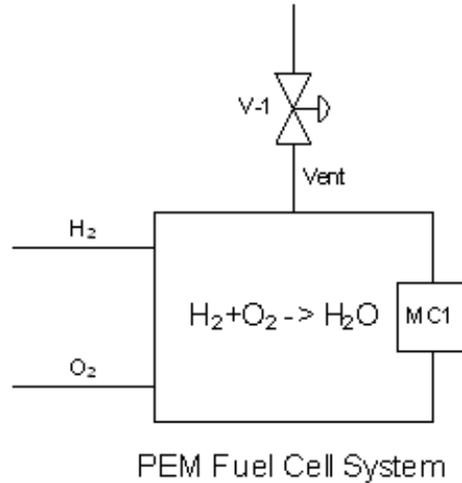
The system produces steam within the system, and there is a vent to release steam if the system becomes too saturated. In our system, we will assume that the inputs are stoichimetric and react completely. Also we will assume that pressure buildup from steam is negligible compared to the change in relative humidity. The only variable in question is the %relative humidity in the system.

- note: this is not how water management actually works in a fuel cell system, but it is a simple example.

A will represent the moisture controller response (0 indicates relative humidity or %RH < 80%, 1 indicates %RH >80%)
B will represent the valve status (0 is closed, 1 is open)

The corresponding Boolean functions for this model are given below (normally you would have to design these yourself to meet the criteria you desire):

A = A
B = A

For this example with 2 input variables, there are $2^2 = 4$ possible permutations and $2^2 = 4$ rows. The resultant permutations for the outputs are: For A where Y=1, the number of 0s and 1s are $2^{(Y-1)}=2^{(1-1)}=1$. For B where Y=2, the number of 0s and 1s are $2^{(Y-1)}=2^{(2-1)}=2$.

444

The resultant truth table is below:

Inputs		Outputs	
T_i		T_{i+1}	
A	B	A	B
0	0	0	0
1	0	1	1
0	1	0	0
1	1	1	1

1.3 State Transition Diagrams

A state transition diagram is a graphical way of viewing truth tables. This is accomplished by looking at each individual initial state and its resultant state. The transition from one state to another is represented by an arrow. Then they are pieced together like a jigsaw puzzle until they fit in place. When one state leads to itself it simply points to itself. The following example is based on the truth table in the previous section.
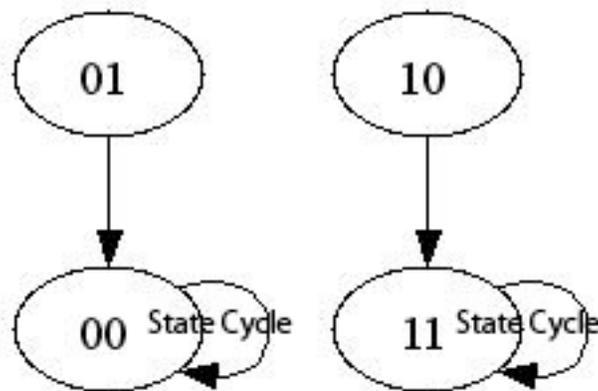
In this example, there are two state cycles. A state cycle is a combination of states around which the system continually enters and reenters. For a finite number of states, there will always exist at least one state cycle. A state cycle is also a pathway or a flowchart that shows the "decision making process" of a Boolean network. This feature is a direct result from two attributes of Boolean networks:
1. Finite number of states
2. Deterministic (there is a certain set of rules that determines the next state that will be entered)

In the example presented in the previous section, there were two state cycles. One advantage of state cycles is it easily allows you to see where your model will end up

cycling and if there are any states that are not accounted for properly by your model. In the previous diagram, if the moisture controller indicated the humidity was below the set value, it would close the valve or hold the valve closed. If the moisture controller indicated that the humidity was above the set value, it would either open the valve or hold it open.

Consider this alternate system.
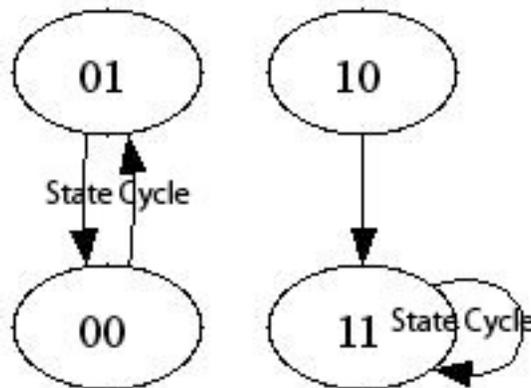
In this example, the state cycle says that if the meter says that the humidity is below the set point it would cycle the vent valve open and closed. This would hurt the system and is not a desired outcome of the model.

For safety and functionality issues, a process control engineer would want to consider all possiblities in the design of any Boolean network modeling a real system.

1.4 Limitations of Boolean Networks

1.4.1 Advantages of Boolean Networks

- Unlike ordinary differential equations and most other models, Boolean networks do not require an input of parameters.
- Boolean models are quick and easy to compute using computers.
- Boolean networks can be used to model a wide variety of activities and events.
- Boolean networks can be used to approximate ordinary differential equations when there are an infinite number of states.

1.4.2 Disadvantages of Boolean Networks

- Boolean networks are restrained to computing very simple math. They cannot be used for calculus and to calculate large quantities.
- Boolean models have relatively low resolution compared to other models.
- It is very time consuming and complicated to build Boolean networks by hand.

446

1.5.1 Worked out Example 1

A hypothetical CSTR needs to have its liquid level maintained below a safety mark by means of a sensor, L1, on the corresponding mark and a control valve placed on the inlet and outlet streams – V1 and V2 respectively. A typical application of the afore-mentioned system could involve heterogeneously catalyzed liquid reaction(s) with liquid product(s).
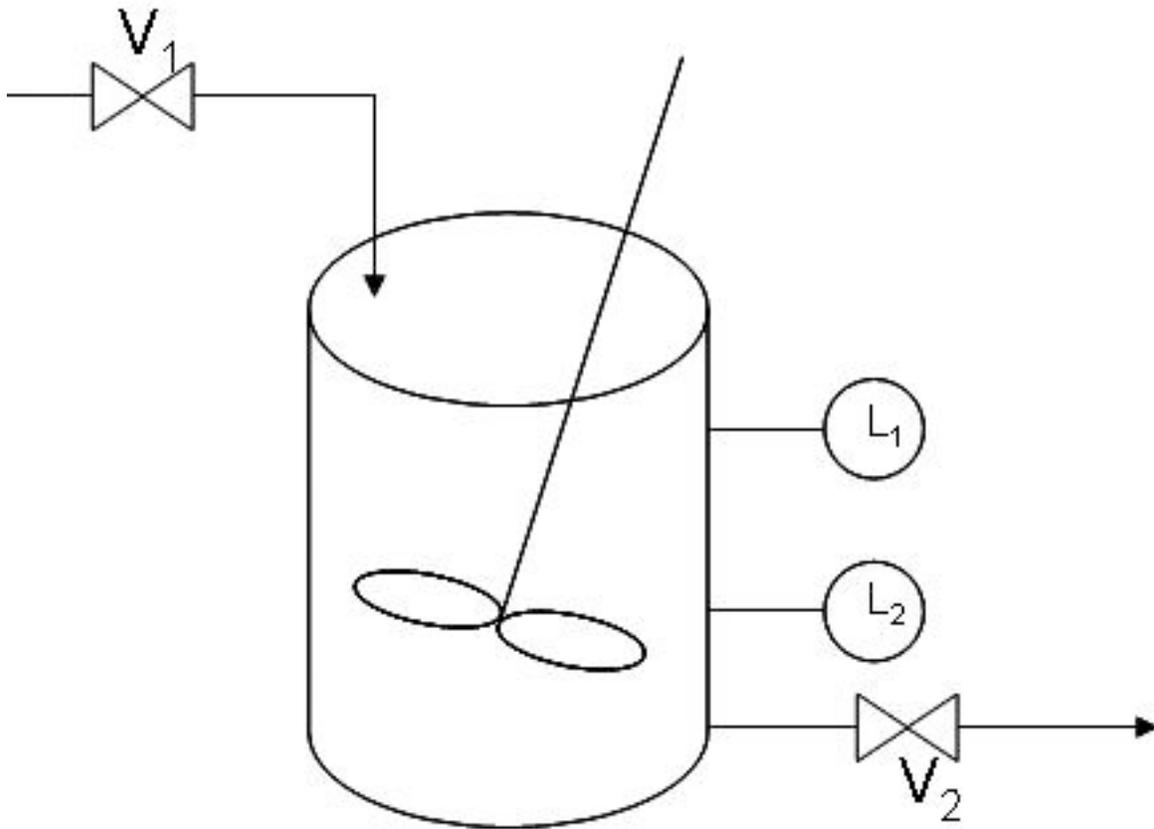
Solution to Example 1

Conventions

Water level sensor

L1	0	1

water level	desirable	too high

Valve

V	0	1
position	closed	open

Initial State

Assume that the CSTR is empty and being filled up. CSTR, being empty, sets the value of L1 to zero. Filling up the CSTR could be done by opening valve 1 - V1 assuming a value of one - and closing valve 2 - V2 assuming a value of zero.

In coordinate form, the initial state is as such: $(L1, V1, V2) = (0, 1, 0)$

Problem Interpretation

Let h be the water level and WL1 be the safety mark defined in the CSTR. The system could assume one of the following states at any one time:

1) $h < WL1$: desirable water level

```
   Maximizing production of the chemical prompts the system
to remain in its current state - that is,   its initial
state.    (L1, V1, V2)final = (0, 1, 0)        final state
```

2) $h > WL1$: water level too high

```
   Prevention of flooding requires that the tank be emptied.
As such, valve 1 (V1) should be closed   to stop the input
while valve 2 (V2) should be open to empty the extra water
above the safety   water mark.     (L1, V1, V2)' = (1, 1, 0)
trigger to valve   (L1, V1, V2)final = (1, 0, 1)
final state
```

State Cycle
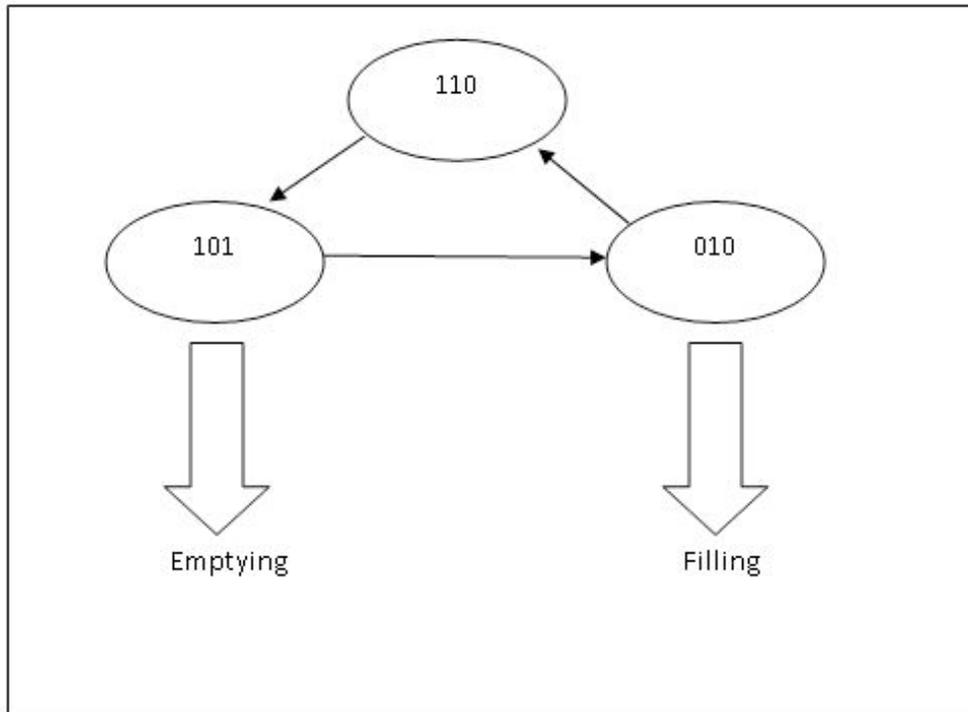
Physical Significance
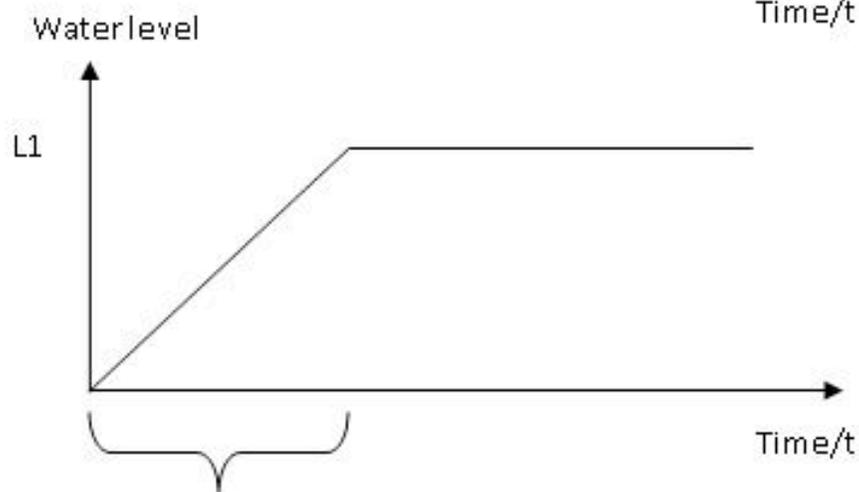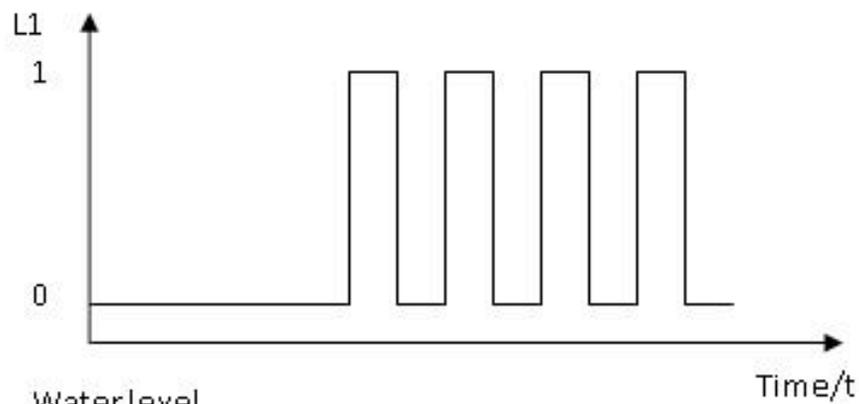
At, time taken to fill the tank up to L1

An illustration of how Boolean system controls the water level inside CSTR

450

Boolean Gameshow:

```
http://video.google.com/googleplaye
r.swf?docId=5205889672082992278
```
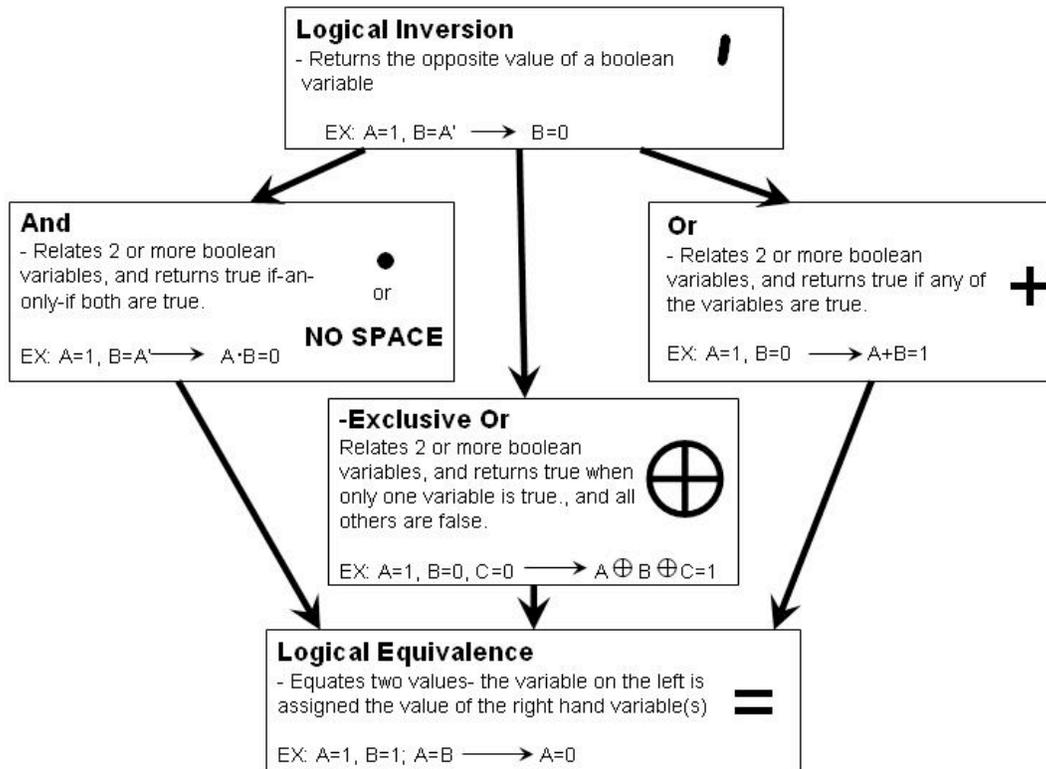
Boolean Models: A mechanism for constructing truth tables

```
http://video.google.com/googleplaye
r.swf?docId=3769987563796006429
```

Slides without narration

Slides without narration Truth Tables

1.8 References

- James E. Palmer and David E. Perlman (1993). *Schaum's Outline of Theory and Problems of Introduction to Digital Systems,* McGraw-Hill Professional. ISBN 0070484392
- Stuart A. Kauffman (1993). *The Origins of Order Self-Organization and Selection in Evolution,* Oxford University Press. ISBN 0195079515

Section 2. Logical control programs: IF... THEN... WHILE...

Title: Logical control programs: IF, THEN, WHILE...
Note: ⊞Video lecture available for this section!
Authors: Stephanie Fraley, Michael Oom, Benjamin Terrien, John Zalewski
Stewards: Ross Bredeweg, Jessica Morga, Ryan Sekol, Ryan Wong

Date Presented:September 26, 2006 /Date Revised: September 13, 2007

- First round reviews for this page
- Rebuttal for this page

2.1 Introduction

A logical control program is a set of conditional statements describing the response of a controller to different inputs. A controller is a computer used to automate industrial processes (See Wikipedia). Process engineers use control logic to tell the controller in a process how to react to all inputs from sensors with an appropriate response to maintain normal functioning of the process. Control logic (sometimes called process logic) is based on simple logic principles governed by statements such as IF X, THEN Y, ELSE Z yet can be used to describe a wide range of complex relationships in a process. Although different controllers and processes use different programming languages, the concepts of control logic apply and the conditions expressed in a logical control program can be adapted to any language.

The concepts behind logical control programs are not only found in chemical processes; in fact, control logic is used in everyday life. For example, a person may regulate his/her own body temperature and comfort level using the following conditional logic statements: IF the temperature is slightly too warm, THEN turn on a fan; IF the temperature is way too warm, THEN turn on the air conditioning; IF the temperature is slightly too cold, THEN put on a sweatshirt; IF the temperature is way too cold, THEN turn on the fireplace. The person takes an input from the environment (temperature) and if it meets a certain prescribed condition, she executes an action to keep herself comfortable. Similarly, chemical processes evaluate input values from the process against set values to determine the necessary actions to keep the process running smoothly and safely (aforementioned example illustrated below).

452

Environmental Signal ("Input")

Way too cold   Slightly too cold   "Set Value" for Optimal Comfort   Slightly too warm   Way too warm
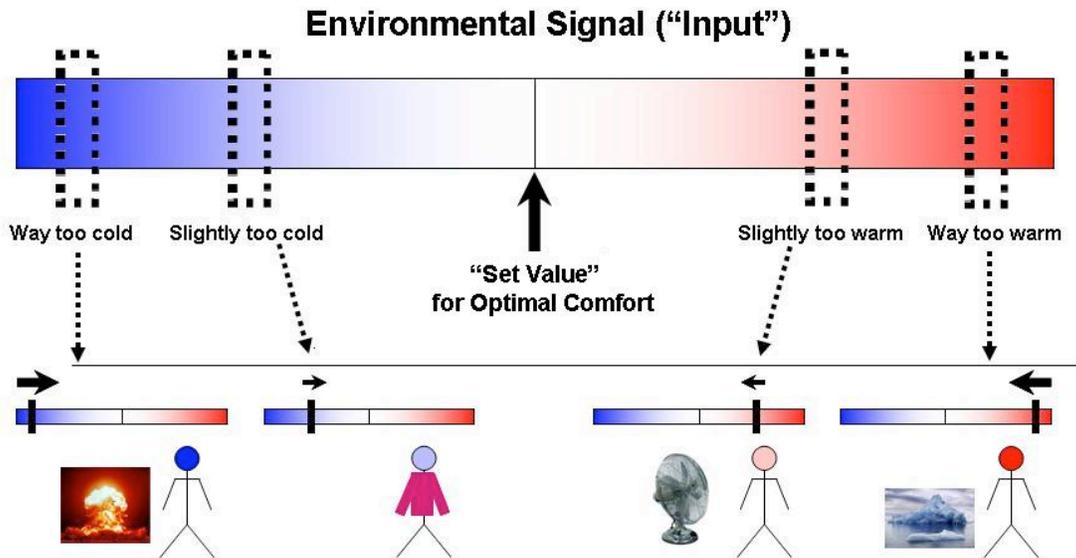
The following sections elaborate on the construction of conditional logic statements and give examples of developing logical control programs for chemical processes.

2.2 Logic Controls

Logical controls (IF, THEN, ELSE, and WHILE) compare a value from a sensor to a set standard for the value to evaluate the variable as True/False in order to dictate an appropriate response for the physical system. The control program for a chemical process contains many statements describing the responses of valves, pumps, and other equipment to sensors such as flow and temperature sensors. The responses described by the system can be discrete, such as an on/off switch, or can be continuous, such as opening a valve between 0 and 100%. The goal of a control program is to maintain the values monitored by the sensors at an acceptable level for process operation considering factors like product quality, safety, and physical limitations of the equipment. In addition to describing the normal activity of the process, a control program also describes how the process will initialize at the start of each day and how the controller will respond to an emergency outside of the normal operating conditions of the system. Unlike a linear computer program, logic programs are continuously monitoring and responding without a specific order. Before constructing logical control programs, it is important to understand the conditional statements, such as IF-THEN and WHIlE statements, that govern process logic.

The following logic controls found below are written in pseudocode. Pseudocode is a compact and informal way of writing computer program algorithms. It is intended for human reading instead of machine reading, and does not require stringent syntax for people to understand. Pseudocode is typically used for planning computer program development and to outline a program before actual coding occurs.

2.2.1 IF-THEN statements

IF-THEN statements compare a value from a sensor to a set value and describe what should happen if the relationship holds. The IF-THEN statement takes the form IF X, THEN Y where X and Y can be single variables or combinations of variables. For example, consider the following statements 1 and 2. In statement 1, both X and Y are single variables whereas in statement 2, X is a combination of two variables. The ability in conditional logic to combine different conditions makes it more flexible than incidence graphs, which can only describe monotonic relationships between two variables (See incidence graphs). A monotonic relationship is one where if X is increasing, Y is always decreasing or if X is increasing, Y is always increasing. For complex processes, it is important to be able to express non-monotonic relationships.

1. IF T>200 C, THEN open V1
2. IF T> 200 C and P> 200 psi, THEN open V1.
Where T is Temperature, P is pressure, and V represents a valve.

In statement 1, if the temperature happens to be above 200 C, valve 1 will be opened. In statement 2, if the temperature is above 200 C and the pressure is above 200 psi, then the valve will be opened.

Otherwise, no action will be taken on valve 1

If the conditions in the IF statement are met, the THEN statement is executed, and depending on the command, the physical system is acted upon. Otherwise, no action is taken in response to the sensor input. In order to describe an alternate action if the IF condition does not hold true, ELSE statements are necessary.

2.2.2 ELSE statements

The simple form of an IF-THEN-ELSE statement is IF X, THEN Y, ELSE Z where again X, Y, and Z can be single variables or combinations of variables (as explained in the IF-THEN section above). The variable(s) in the ELSE statement are executed if the conditions in the IF statement are not true. This statement works similar to the IF-THEN statements, in that the statements are processed in order. The ELSE statement is referred to last, and is a condition that is often specified to keep the program running. An example is the following:

IF P>200 psi, THEN close V1
ELSE open V4

In this statement, if the pressure happens to be 200psi or less, the THEN statement will be skipped and the ELSE statement will be executed, opening valve 4.

Sometimes, if X, Y or Z represent many variables and several AND or OR statements are used, a WHILE statement may be employed.

454

CASE statement is an alternative syntax that can be cleaner than many IF..THEN and ELSE statements. The example shown in the table below shows its importance.

CASE	IF..THEN and ELSE

T>Tset+2:
 	
     v2=v2+0.1
 	
 T>Tset+1:
 	
     v2=v2+0.05
 	
 T<Tset-1:
 	
     v2=v2-0.05
 	
 T<Tset-2:
 	
     v2=v2-0.1
	IF T>Tset+1:
	     IF T>Tset+2:
	     v2=v2+0.1
	     ELSE:
	     v2=v2+0.05
	
	 ELSE IF T<Tset-1:
	
	     IF T<Tset-2:
	     v2=v2-0.1
	     ELSE:
	     v2=v2-0.05

Thus CASE statements make the code easier to read for debugging.

2.2.4 WHILE statements

The WHILE condition is used to compare a variable to a range of values. The WHILE statement is used in place of a statement of the form (IF A>B AND IF A<C). WHILE statements simplify the control program by eliminating several IF-AND statements. It is often useful when modeling systems that must operate within a certain range of temperatures or pressures. Using a WHILE statement can allow you to incorporate an alarm or a shut down signal should the process reach unstable conditions, such as the limits of the range that the WHILE statement operates under. A simple example illustrating the use of the WHILE statement is shown below.

Example:
A tank that is initially empty needs to be filled with 1000 gallons of water 500 seconds after the process has been started-up. The water flow rate is exactly 1 gallon/second if V1 is completely open and V1 controls the flow of water into the tank.

Using a IF...THEN statement the program could be written as follows:
IF t > 500 and t < 1501 THEN set V1 to open
ELSE set V1 to close.

The WHILE statement used to describe this relationship is as follows:

WHILE 500 < t < 1501 set V1 to open
ELSE set V1 to close.

It may not seem like much of a change between the two forms of the code. However, this is a very simple model. If modeling a process with multiple variables you could need many IF...THEN statements to write the code when a single WHILE condition could replace it.

Example:
V1 controls reactants entering a reactor that can only run safely if the temperature is under 500K.

The WHILE can be used to control the process as follows:
WHILE T < 500 set V1 to open
ELSE set V1 to close. ALARM.

This example shows how a WHILE condition can be used as a safety measure to prevent a process from becoming unstable or unsafe.

In addition to lists of IF-THEN-ELSE-WHILE statements, control logic can be alternately represented by truth tables and state transition diagrams. Truth tables show all the possible states of a model governed by conditional statements and state transition diagrams represent truth tables graphically. Oftentimes, these are used in conjunction with booleans, variables that can only have two values, TRUE OR FALSE. A Boolean model or Boolean function follows the format of IF-THEN statements described here. A description of Boolean models, truth tables, and state transition diagrams is given here.

2.2.5 GO TO statements

The GO TO statement helps to break out of current run to go to a different configuration. It can be an important operator in logical programming because a lot of common functions are accessed using GO TO. However, many programmers feel that GO TO statements should not be used in programming since it adds an extra and often unnecessary level of complex that can make the code unreadable and hard to analyze. Even though the GO TO operator has its downsides, it is still an important operator since it can make help to simplify basic functions. It can simplify code by allowing for a function, such as a fail safe, be referenced multiple times with out having to rewrite the function every time it is called. The GO TO operator is also important because even advanced languages that do not have a GO TO function often have a different operator that functions in a similar manner but with limitations. For example, the C, C++ and java languages each have functions `break` and `continue` which are similar to the GO TO operator. `Break` is a function that allows the program to exit a loop before it reaches completion, while the `continue` function returns control to the loop without executing

456

the code after the `continue` command. A function is a part of code within a larger program, which performs a specific task and is relatively independent of the remaining code. Some examples of functions are as follows:

FUNCTION INITIALIZE: It runs at the beginning of a process to make sure all the valves and motor are in correct position. The operation of this function could be to close all valves, reset counters and timers, turn of motors and turn off heaters.

FUNCTION PROGRAM: It is the main run of the process.

FUNCTION FAIL SAFE: It runs only when an emergency situation arises. The operation of this function could be to open or close valves to stop the system, quench reactions via cooling, dilution, mixing or other method.

FUNCTION SHUTDOWN: It is run at the end of the process in order to shutdown.

FUNCTION IDLE: It is run to power down process.

All the functions mentioned above except FUNCTION IDLE are used in all chemical processes.

2.2.6 ALARM statements

The ALARM statement is used to caution the operators in case a problem arises in the process. Alarms may not be sufficient danger to shut down the process, but requires outside attention. Some example when ALARM statements are used in the process are as follows:

-If the storage tank of a reactant is low, then ALARM.

-If the pressure of a reactor is low, then ALARM.

-If no flow is detected even when the valve is open, then ALARM.

-If the temperature of the reactor is low even after heating, then ALARM.

-If redundant sensors disagree, then ALARM.

In conclusion ALARM functions are very important in order to run a process safely.

2.3 Control Language in Industry

As stated before, these commands are all in pseudocode, and not specific to any programming language. Once the general structure of the controller is determined, the pseudocode can be coded into a specific programming language. Although there are many proprietary languages in industry, some popular ones are:

- Visual Basic

- C++

- Database programming (ex. Structured Query Language/SQL)

- Pascal

- Fortran

Pascal and Fortran are older languages that many newer languages are based on, but they are still used with some controllers, especially in older plants. Any experience with different computer languages is a definite plus in industry, and some chemical engineers make the transition into advanced controls designing, writing, and implementing code to make sure a plant keeps running smoothly.

2.4 Logical Functions in Microsoft Excel

Microsoft Excel has basic logical tools to help in constructing simple logical statements and if needed more complex logical systems. A list of the functions is shown below.

TRUE()...Returns the logical value, TRUE.

FALSE()...Returns the logical value, FALSE.

AND(logical_expression_A,B,C).........Returns TRUE if all the expressions are true.

OR(logical_expression_A,B,C)...........Returns TRUE if one of the expressions are true.

NOT(logical_expression)....................Returns the opposite of the expected logical value. If the expression is TRUE, it will return FALSE

IFERROR(value,value_if_error)............Returns the value unless there is an error in which it will return the value_if_error.

IF(logical_expression,value_if_true,value_if_false)..............Checks the validity of the expression and returns TRUE or FALSE likewise.

The IF() function will be most useful in logical programing. It is essentially an IF THEN or IF ELSE function returning one of two values based on the logical expression it is testing. Excel also allows logical functions within functions. This allows for logical expressions involving more than one IF statement within itself for example. Use of these tools will be practical in quickly setting up control programs and other systems in Microsoft Excel.

458

2.5 Constructing a Logical Control Program

Understanding the conditional statements used in control logic is the first step in constructing a logical control program. The second step is developing a thorough understanding of the process to be controlled. Knowledge of the equipment, piping, and instrumentation (contained on a [P&ID diagram](#)), operating conditions, chemical compounds used, and safety concerns is necessary. Particularly it is important to know the measured and controlled variables. For example, the pressure limits of a tank must be known in order to develop a control plan to ensure safety; ignoring this constraint could lead to explosion and injury. Once the necessary controls are known, one can develop a plan using the logical statements described previously. The third step is constructing a logical control program is understanding that there is not always a right answer, meaning there are many different ways to ensure the same desired outcome.

Worked out Examples 1, 2, and 3 demonstrate the construction of simple logical control programs. The more complex the situation, the longer the logical control plan becomes yet the process is still the same. An example of a more complex logical control program is given [here](#).This example is from the 2005 ChE 466 class at the University of Michigan and describes an entire chemical process from the delivery of raw materials to the output of the final product.

2.6 Determining Fail Safe Conditions

Fail safe is the practice of designing a system to default to safe conditions if anything or everything goes wrong. The goals of fail safe conditions are to:

- Protect plant personal
- Protect the local community around the plant
- Protect the environment
- Protect plant equipment

In order to establish safe conditions, fail safe programs must specify the desired positions of all valves and status of all motors and controlled equipment. For example, in an exothermic reactor, fail safe conditions would specify opening all cooling water valves, closing all feed valves, shutting off feed pump motors, turning on agitator motor, and open all vent valves.

Control programs frequently define fail safe conditions at the beginning of the program. These conditions are then activated using a GO TO command when process conditions exceed the maximum or fall below the minimum allowable values.

All processes must be evaluated for conditions that could cause hazards and fail safe procedure must be designed counteract the effects.

2.7 Worked Out Example 1: Reboiler

Disclaimer: This is a fictional example of an isolated unit operation, neglecting possible outcomes on the rest of the plant.

Reboilers are used in industry to cool down process streams by creating steam from water. This chemical process involves a phase change from liquid (water) to gas (steam) and it is important to monitor flowrates, temperatures, and pressures. Below is a diagram of a reboiler. The controlled variables are F1, F2, and F3; these are controlled by manipulating the corresponding valves. The measured variables are P1, T1, and T2. Considering the operating conditions and constraints given below, write a logical control program for the reboiler.
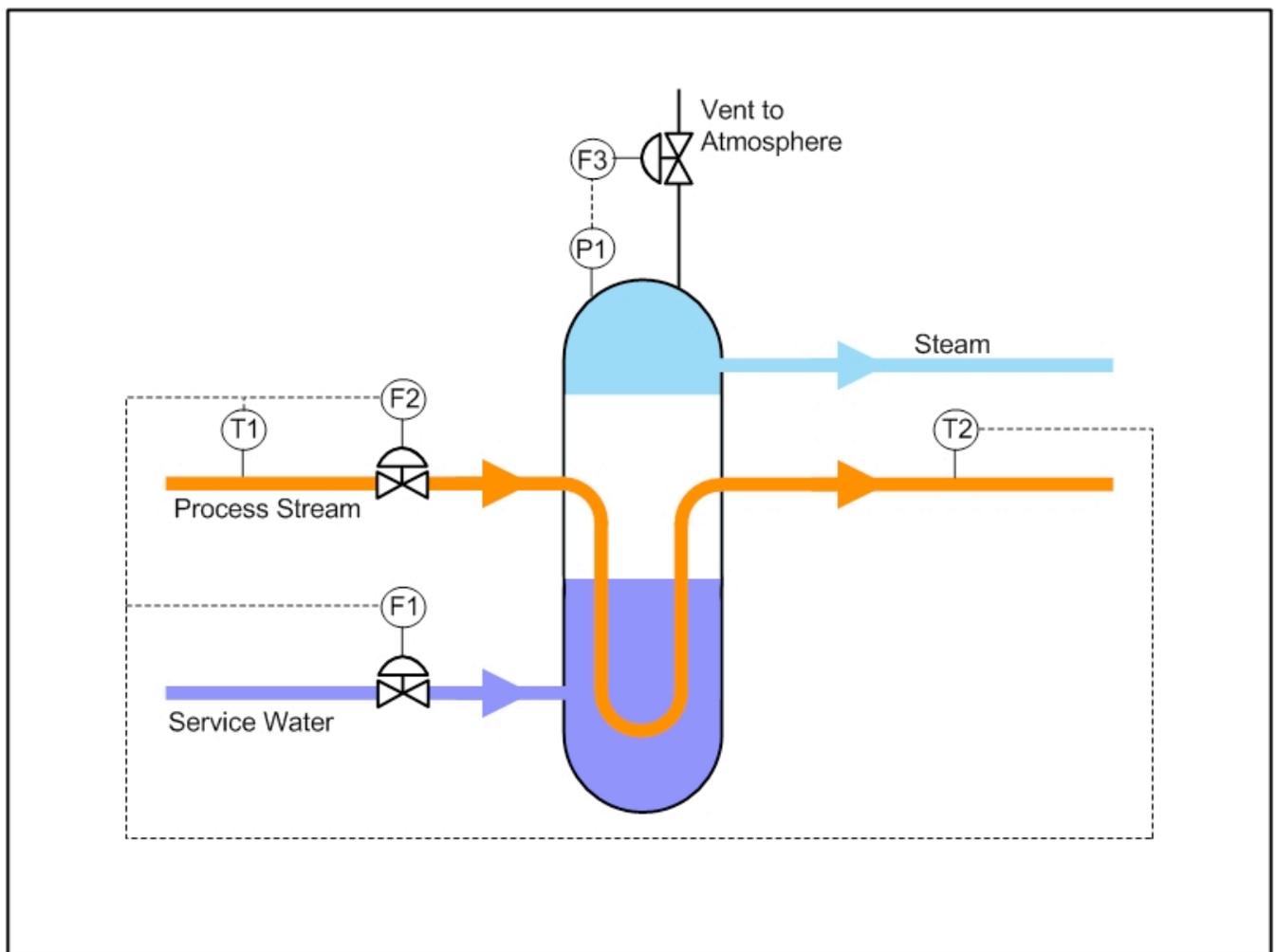
The operating constraints are as follows:

- T1 must not exceed 350 ℃
- T2 must be between 100 and 200 ℃
- P1 cannot exceed 150 psi

The normal operating conditions for the controlled variables are as follows:

- F1 is 20 gal/min
- F2 is 10 gal/min
- F3 is closed

Solution
The first step is understanding the system to be controlled. Looking at the diagram, one can see that F1 controls the flow rate of the service water, F2 controls the flow rate of the process stream, and F3 controls the vent to the atmosphere. T1 is the temperature of the process stream entering the reboiler and T2 is the temperature of the process stream exiting the reboiler. P1 is the pressure inside the reboiler.
Below is a possible control program to ensure the operating constraints are met; there may be other solutions to achieve the same objective:

IF T2 > 200 °C, THEN F1 = 30 gal/min

IF T2< 100 °C, THEN F1 = 10 gal/min

IF P1 > 150 psi, THEN open F3

IF T1 > 350 °C, THEN F2 = 2 gal/min

In order to control the exit temperature of the process stream, one can increase or decrease the service water from the normal flow rate. Opening the vent to the atmosphere reduces the pressure if it reaches an unsafe value. If the entering temperature is too high for the process stream, reducing the entering flow rate will ensure the exit temperature is not too high for the remainder of the process. Note: reducing the process stream may negatively impact the rest of the plant.

2.8 Worked out Example 2: Thermostat

Chemical engineers are increasingly involved in biological applications. In many biological processes, temperature control can be very critical to the process. Take, for example, a simulated cell on a computer chip. Let's say that it is imperative for the chip to remain at 97°F ± 1°F, very similar to the temperature of the human body. Some of the reactions and processes on the chip will not function at temperatures outside this range and can become irreversibly damaged unless the power supply to the chip is turned off. The power is turned on and off by a switch S1. There are flows of cool water with an automatic valve V1 and warm water with an automatic valve V2 for cooling and heating purposes. The chip is attached to a thermostat to sense and control the temperature, T. Write out a controlled logic scheme to maintain the chip temperature and to prevent damage.

Hint: Heat or cool the chip before reaching the limits (about half a degree)

<u>Solution</u>

To control the temperature, the chip must be heated or cooled depending on the environment

IF (T<96.5) THEN V2 is open

ELSE V2 is closed

IF (T>97.5) THEN V1 is open

ELSE V1 is closed

The control is executed before the temperature limit is reached to allow for lag time in the heating or cooling water flow.

Also, to ensure that the chip is not damaged, the power needs to be shut off if the temperature goes above 98°F or below 96°F

WHILE (96<T<98) THEN S1 is on

ELSE S1 is off

2.9 Worked out Example 3: Chemical Reactor

There is an exothermic chemical reaction occurring in a CSTR (Continuous Stirred Tank Reactor) that involves two reactants being fed at a 1:1 ratio. All valves are set to be 50% open normally. Write a control program that keeps the level in the CSTR vessel less than 8 meters (the tank is 10 meters tall) and the reactor temperature below 450 degrees Celsius.
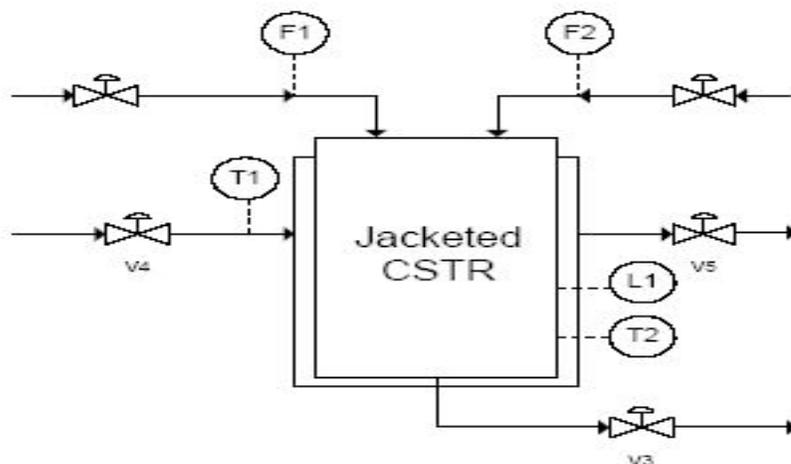
Solution

While L1 > 8 set V3 to 100% open and close V1 and V2
Else set V1 and V2 and V3 to 50% open

If T2 > 450 THEN set V5 and V4 to 100% open

Else set V5 and V4 to 50% open

This solution gives an example of using AND statements to control multiple valves with just one condition.

2.10 Worked out Example 4: Programming and Alarms

There is a process that is run by the P&ID shown below.
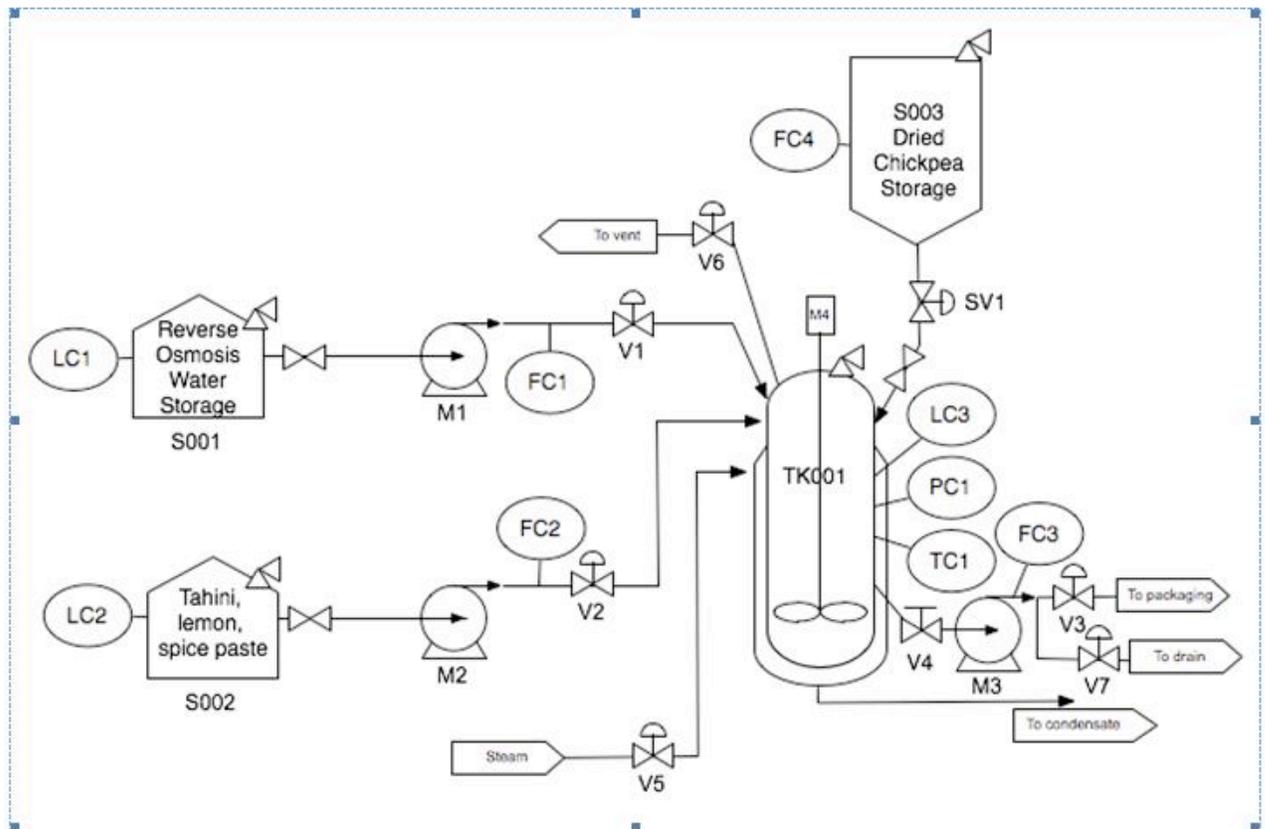
Based on this process and the steps listed below, write out a detailed control program for the process. Use comments (denoted with #) if necessary to explain the logic behind the steps.

1) Measure Qw units of water into a tank

2) Add Qc units of dried chickpeas

3) Let dried chickpeas soak for 20 hours without mixing

4) Drain off soaking water to waste (assume the filter in the tank will not allow whole chickpeas through the pump) 5) Add Qw units of fresh water to the tank.

6) Heat the tank to Tcook and maintain the pressure at 4 atm. Note that your tank is rated to withstand pressures between 0.5 and 6 atm, while outside of that range the tank may implode or explode.

7) Cook chickpeas for 20 minutes.

8) After cooking, turn off heat and allow the system to return to ambient temperature (Tamb) and ambient pressure. Beware of a strong vacuum forming in the tank as the water vapor condenses!

9) Drain cooking water to drain.

10) Pump in Qs units of the tahini spice mix

11) Blend the mixture for 10 minutes to produce a smooth hummus paste.

12) Pump out product to packaging.

13) Fill tank with clean water and agitate to clean the reactor.

14) Pump wash water to drain.

Solution

FUNCTION INITIALIZE

Turn off M1, M2, M3, M4
Close V1, V2, V3, V5, V6, V7, SV1
Set all timers to zero
Set all totalizers to zero

FUNCTION FAILSAFE

464

Turn off M1, M2, M3, M4
Close V1, V2, V3, V5, V7, SV1
Open V6

FUNCTION PROGRAM

#Step 1 – Measure Qw unites of water into a tank
Turn on M1
Open V1
WHILE FC1tot < Qw:
Adjust V1 to FC1set

IF LC1 < LC1min:

ALARM

Close V1

Turn off M1
#FC1tot is the total amount of fluid that has gone through the flow meter
#FC1set is the set point (amount the valve is open) for V1 that FC1 has already
programmed into it
#LC1min is the minimum acceptable level of fluid in S001

#Step 2 – Add Qc units of dried chickpeas
Open SV1
WHILE FC4 < Qc:
Adjust SV1 to FC4set

IF LC4 < LC4min:

ALARM
Close SV1
#FC4set is the set point for SV1 that FC4 has already programmed into it
#LC4min is the minimum acceptable level of fluid in S003. LC4 is not on the P&ID,
however, it makes sense to have one on it so the level on the tank can be properly
monitored

#Step 3 – Let dried chickpeas soak for 20 hours without mixing
WAIT 20 hours

#Step 4 – Drain off soaking water to waste
Open V7
Turn on M3
WHILE FC3tot < Qw:
Adjust V7 to FC3set2

Turn off M3

Close V7
#FC3tot is the total amount of fluid that has gone through the flow meter
#FC3set2 is the set point for V7 that FC3 has already programmed into it

#Step 5 – Add Qw units of fresh water to the tank
Clear FC1tot
Turn on M1
Open V1
WHILE FC1tot < Qw:
Adjust V1 to FC1set

IF LC1 < LC1min:

ALARM

Close V1

Turn off M1

#Step 6 – Heat the tank to Tcook and maintain the pressure at 4 atm.
WHILE TC1 < Tcook:
Adjust v5 to Tcook

IF OR (PC1 < 0.5, PC1 > 6):

GO TO FAILSAFE

IF PC1 < PC1set:

Close V6

ELSE:

466

Adjust V6 to PC1set

IF LC3 > LC3max:

GO TO FAILSAFE
#PC1set is the setting that V6 must be set to for the tank to have 4 atm of pressure
in it
#LC3max is the maximum level that the contents of the tank are allowed to get to.
Anything higher indicates a problem with one of the flow meters.

#Step 7 – Cook chickpeas for 20 minutes#
WAIT 20 minutes

#Step 8 –After cooking, turn off heat and allow the system to return to ambient
temperature and pressure
IF PC1 > PC1amb:
Adjust V6 to PC1set2

IF OR (PC1 < 0.5, PC1 > 6):

GO TO FAILSAFE

WHILE TC1 > Tamb:

Close V5
#PC1amb is the ambient pressure of 1 atm that the system needs to get to
#PC1set2 is the second setting on PC1 which effects how open V6 is

#Step 9 – Drain cooking water to drain
Clear FC3tot
Open V7
Turn on M3
WHILE FC3tot < Qw:
Adjust V7 to FC3set

Turn off M3

Close V7

#Step 10 – Pump in Qs units of the Tahini spice mix
Clear FC2tot
Turn on M2
Open V2
WHILE FC2tot < Qs:
Adjust V2 to FC2set

IF LC2 < LC2min:

ALARM

Close V2

Turn off M2
#FC2tot is the total amount of fluid that has gone through the flow meter
#FC2set is the set point for V2 that FC2 has already programmed into it

#Step 11 – Blend the mixture for 10 minutes to produce a smooth hummus paste
Turn on M4
WAIT 10 minutes
Turn off M4

#Step 12 – Pump out product to packaging
Open V3
Turn on M3
WHILE LC3 > 0:
Adjust V3 to FC3set

Turn off M3

Close V3
#FC3set is the set point for V3 that FC3 has already programmed into it

#Step 13 – Fill tank with clean water and agitate to clean the reactor
Turn on M1
Open V1
WHILE LC3 < LC3max:
Adjust V1 to FC1set

IF LC1 < LC1min:

ALARM

Close V1
Turn off M3
Turn on M4
WAIT 10 minutes

Turn off M4

#Step 14 – Pump was water to drain
Open V7
Turn on M3
WHILE LC3 > 0:
Adjust V7 to FC3set2

Turn off M3

Close V7

2.11 Worked Out Example 5: Another Chemical Reactor (taken from Prof. Barkel's 9/29/09 lecture)
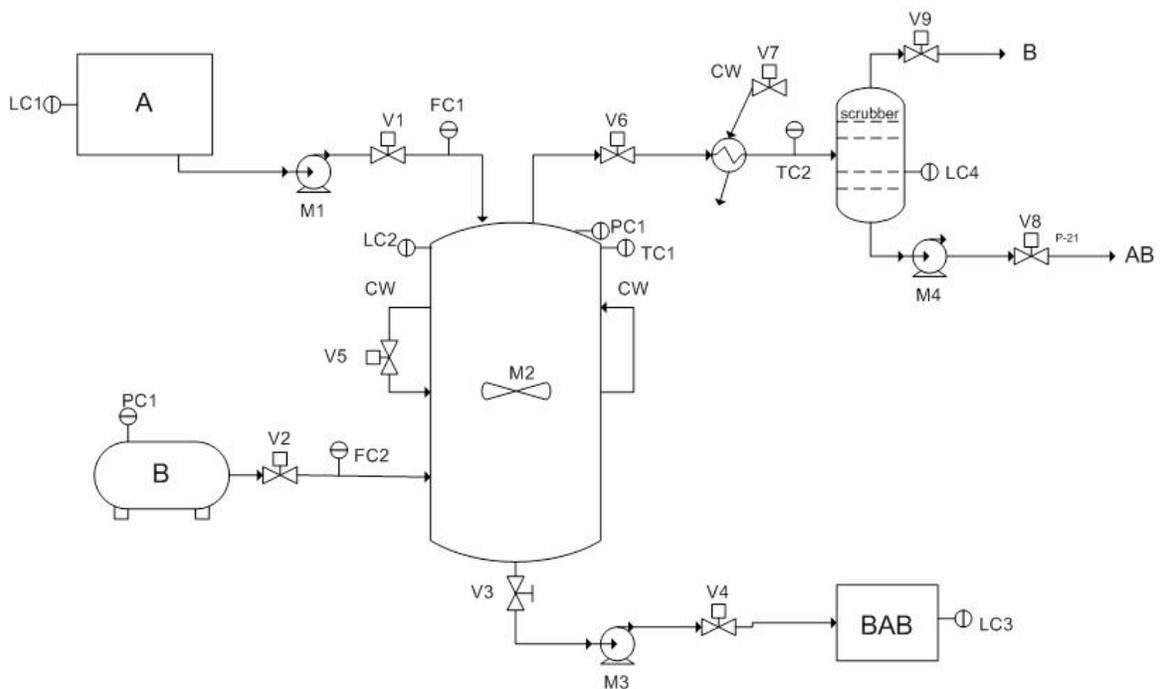
WRITING THE CONTROL PROGRAM

Nomenclature:
x_A = input amount of A
y_B = input amount of B

Readings:
L_L = low level
L_H = high level
Lmax = maximum high level
Lmax = maximum low level

Things to note:
Be sure to use If, Then statements.
Lines in *italics* are comments used to organize the program.

CONTROL PROGRAM

Equal means within +/- 0.05 % of the value being held equal.

Initialize *This is used to set all control devices to the position you want them in.*
1. Turn off all valves
2. Shut off all motors

Fail Safe *This is used to shut down the process in case something goes wrong and must be terminated immediately.*
1. Turn off all valves except V5, V7, V9, V6
2. Shut off all motors except M2

Universal Statements
If T1>Tmax, then go Fail Safe.

1. Initialize
2. Open V6, V9, V1. *[so that systems doesn't build up pressure]*
3. If L1>L_L, then turn on M1. *[fill with A]*
4. If L2>=*Lag*, turn on M2. *[start stirrer]*
5. If F1>=x_A, turn off V1, shut off M1. *[stop flow of A]*
6. Open V5, V7. *[allow for cooling]*
7. If P1>=P_L, then open V2. *[fill with B]*
8. If L4>=L4min, open V8, turn on M4. *[release product AB]*
9. If L2<*Lag*, then shut off M2. *[so motor doesn't run dry]*
10. If L4<$L4_L$, close V8. *[stop outflow of product AB]*
11. If T1><T_H, then open V5. *[cool reactor temperature]*
12. If T2><T_H, then open V5. *[cool upper product stream]*
13. If T1<<T_L, then close V5. *[stop cooling of reactor]*
14. If F2>=<y_B, then turn off V2. *[stop flow of B]*

470

15. Close V8, M4, leave V6, V9 open. *[shut down process but allow it to vent]*
16. Pump out BAB.

2.12 Worked Out Example 6
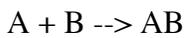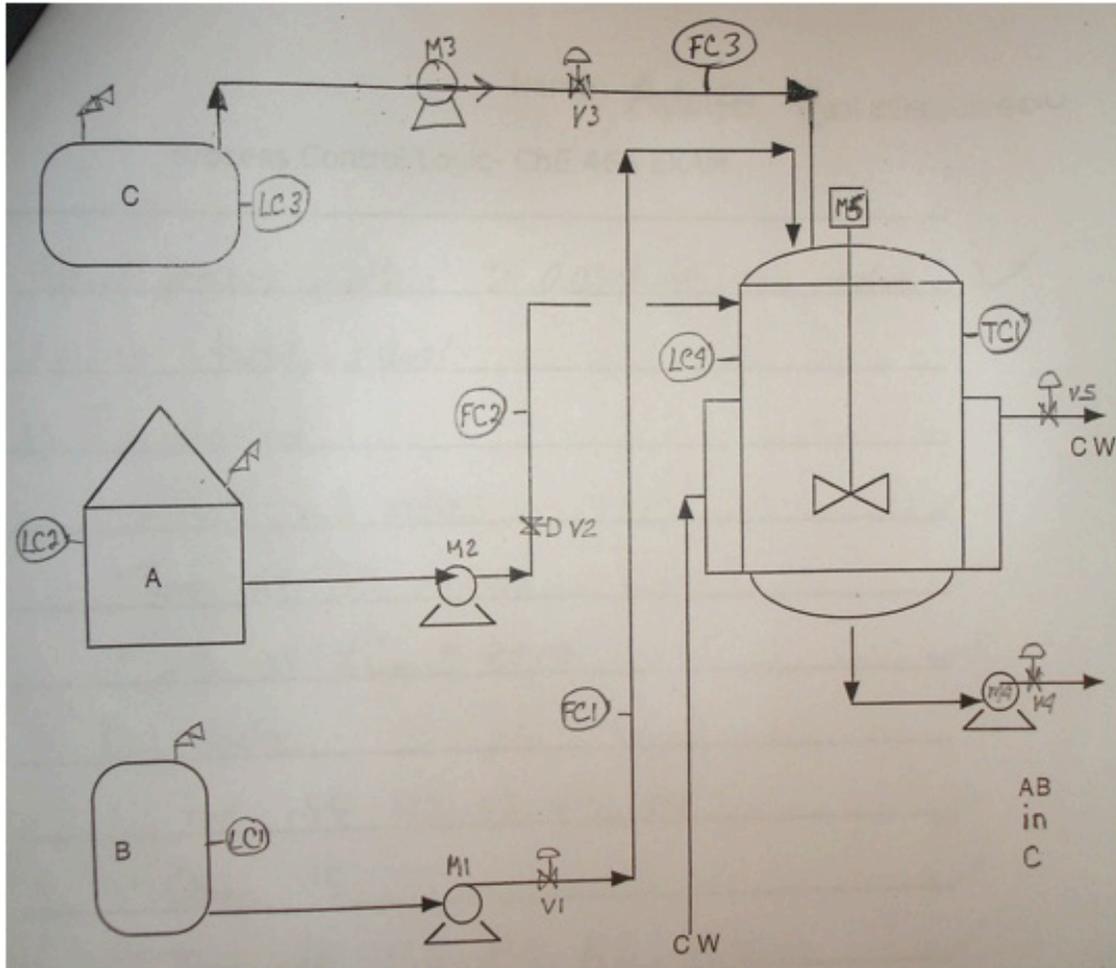
A + B --> AB

The reaction is exothermic
A, B, AB are all liquids
It goes to 100% completion
X_A = total amount of A used
Y_B = total amount of B used

B is added slowly into a full charge of A
The temperature is maintained at T_R
Upon completion of the reaction, the AB is cooled to T_P
Z_C, an amount of solvent C is added to the AB to make the final product

Using only the equipment and instrumentation shown, write the control logic for this

BATCH reaction. Use **If**, **Then** logic.

Use of subscripts:
Ag for agitator
L for low-bottom of control range
H for high-top of control range
Min for minimum - lowest level allowed
Max for maximum - highest level allowed
tot for totalizer

Process Control Logic

- Equal means within +/- 0.05% of the value being held equal.

1) Initialize

```
Shut all valves
 Turn off all motors
 Set all FC_tot = 0
```

2) Fail safe

```
Shut off V1, V2, V3, V4
 Open V5
 Turn off M1, M2, M3
 Turn off M4, Open V4
 Turn on M5 if LC4>Lag, else turn off M5
```

3) If T1>T_{max}, then go Fail Safe
4) If L2>$L2_{min}$, then turn on M2, open V2
5) If L2<$L2_{min}$, then turn off M2, shut off V2
6) If $FC2_{tot} = X_A$, then turn off M2, shut off V2
7) If L4>Lag, then turn on M5
8) If L1>$L1_{max}$, then go Fail Safe
9) If L1>$L1_{min}$, then turn on M1, open V1
10) If L1<$L1_{min}$, then turn off M1, shut off V1
11) If $FC1_{tot} = Y_B$, then turn off M1, shut off V1
12) If T1>$T1_H$, then open V5
13) If T1<$T1_L$, then close V5
14) If L4>$L4_H$, then close V1, turn off M1
15) If T1>T_P, then open V5
16) If T1=T_P, then close V5
17) If L3>$L3_{min}$, then turn on M3, open V3
18) If L3<$L3_{min}$, then turn off M3, shut off V3
19) If $FC3_{tot} = Z_C$, then turn off M3, shut off V3

20) If LC4<Lag, then turn off M5
21) Open V4
22) Turn on M4
23) If L4<$L4_{min}$, then turn off M4, shut off V4

2.13 Sage's Corner

Narrated Presentation of Example 1

```
http://video.google.com/googleplayer.swf?
docId=5141188046317469856
```

Narrated Presentation of Example 2

```
http://video.google.com/googl
eplayer.swf?docId=-
2693631730278014126
```

Media:LogicalControlsProgram9122007.ppt

Media:ProcessLogicN.ppt

2.14 References

- Savitch, Walter. "Problem Solving with C++ The Object of Programing". Boston: Pearson Education, Inc. 2005.
- Stanford Encyclopedia of Philosophy
- Woolf, Peter. "Possible Useful Notes", Notes for Boolean Logic, September 22, 2005.
- Woolf, Peter. "Project 1 Sample Solution", October 6, 2005.
- Woolf, Peter. "Useful Definitions", Terms Review, October 11, 2005.

Chapter 6. Modeling Case Studies

More information on chemical process modeling in general at ECOSSE example 1
and ECOSSE example 2

Section 1. Surge tank model

Note: Video lecture available for this section!

Authors: (September 6, 2007) Cara Canady, David Carpenter, Che Martinez, Jeremy Minty, Bradley
Novak

1.1 Introduction

Used to regulate fluid levels in systems, surge tanks act as standpipe or storage reservoirs
that store and supply excess fluid. In a system that has experienced a surge of fluid, surge
tanks can modify fluctuations in flow rate, composition, temperature, or pressure.
Typically, these tanks (or "surge drums") are located downstream from closed aqueducts
or feeders for water wheels. Depending upon its placement, a surge tank can reduce the
pressure and volume of liquid, thereby reducing velocity. Therefore, a surge tank acts as
a level and pressure control within the entire system.

Since the flow to the surge tank is unregulated and the fluid that is output from the surge
tank is pumped out, the system can be labeled as unsteady-state [MIT], but the approach
to an approximate solution (below) utilizes techniques commonly adhered to when
solving similar steady-state problems.

The technology behind surge tanks has been used for decades, but researchers have had
difficulty fully finding a solution due to the non-linear nature of the governing equations.
Early approximations involving surge tanks used graphical and arithmetical means to
propose a solution, but with the evolution of computerized solving techniques, complete
solutions can be obtained. [Wiley InterScience].

1.2 Derivation of Ordinary Differential Equation

To accurately model a surge tank, mass and energy balances need to be considered across
the tank. From these balances, we will be able to develop relationships for various
characteristics of the surge tank

474

1.2.1 Diagram of Surge Tank System
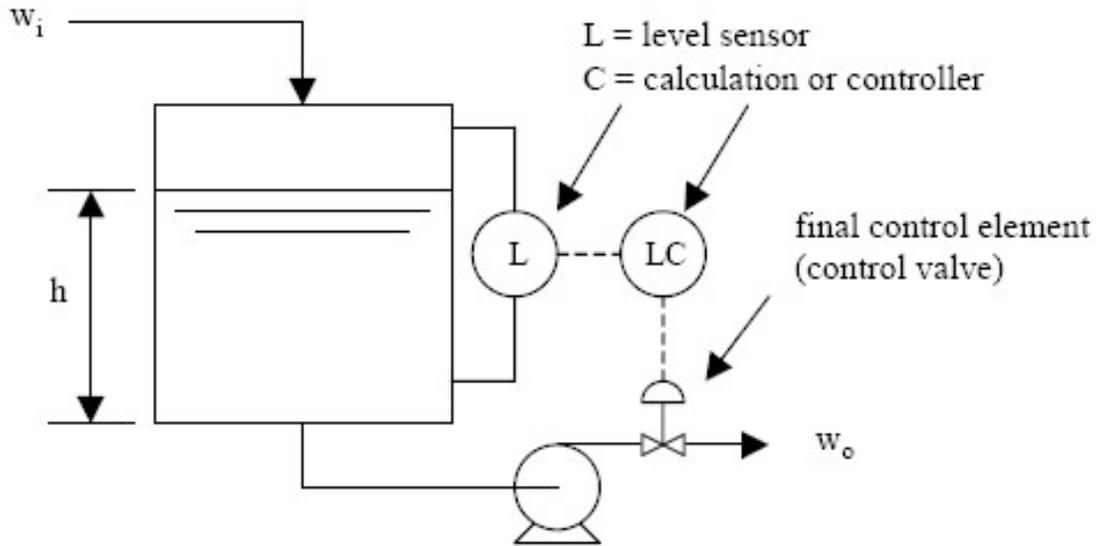
A surge tank relies on the level sensor to determine whether or not fluid stored in the tank should be removed. This regulated outflow is pumped out in response to calculations made by a controller that ultimately opens and closes the control valve that releases the fluid from the tank.

1.2.2 Governing Equations of Surge Tank Model

A surge tank's components must be divided up and evaluated individually at first, then as a whole. First, the expression for the inlet stream must be obtained. The simplistic sine function will be used as the basis for the expression of a stream because it typically describes the tidal surge pattern of a low-viscous fluid, like water. The flowrate, w, will be given in units of kg h^{-1}.

$$w_i = a + bsin(c\pi t)$$

where *a* and *b* are constants determined by the specific problem circumstance.

A mass balance must now be performed on the tank as a system. Using the concept that *mass in = mass out*, and assuming that the tank is a perfect cylinder devoid of diversions, an expression can be derived:

rate mass in - rate mass out = accumulated fluid in tank

Rewriting,

$$w_i - w_o = \rho \frac{dV}{dt}$$

$$\frac{dh}{dt} = \frac{w_i - w_o}{\rho A}$$

$$\int_0^h dh = \int_0^t \frac{w_i - w_o}{\rho A} dt$$

$$h(t) - h_o = \frac{1}{\rho A} \int_0^t w_i - w_o dt$$

$$h(t) = h_o + \frac{1}{\rho A} \int_0^t w_i - w_o dt$$

Where, at time $t=0$, the amount of fluid in the surge tank is constant; thus $h(0) = h_0$.

Substituting the original equation for the inlet stream, w_i, into the expression for the height of the tank, $h(t)$, the governing equation for the height of the tank is obtained:

$$h(t) = h_o + \frac{1}{\rho A} \int_0^t [(a + bsin(c\pi t) - w_0] dt$$

Integrating by parts,

$$h(t) = h_o + \frac{x}{\rho A}(1 - cos(c\pi t))$$

where x is formed from the constants a and b during integration.

1.3 Secondary uses of Surge Tanks

Surge tanks are most commonly used to protect systems from changes in fluid levels; they act as a reservoir that stores and supply excess fluid. In addition, the tanks shield the systems from dramatic changes in pressure, temperature, and concentration. They can

476

also allow one unit to be shut down for maintenance without shutting down the entire plant.

1.3.1 Pressure

There can be moments of high pressure, called hammer shock, in a system when the liquid (incompressible) flow is stopped and started. The energy that liquids possess while traveling through pipes can be categorized as potential or kinetic. If the liquid is stopped quickly, the momentum that the liquid carries must be redirected elsewhere. As a result, the pipes vibrate from the reactive force and weight of the shock waves within the liquid. In extreme cases, pipes can burst, joints can develop leaks, and valves and meters can be damaged.

The extreme amounts of pressure are dampened when the fluid enters the surge tanks. The surge tank acts as a buffer to the system, dispersing the pressure across a greater area. These tanks make it possible for a system to more safely execute their tasks.

1.3.2 Temperature

The temperature of a fluid can either be controlled or changed through the use of a surge tank. The surge tank allows for a rapid change in fluid temperature. This is exemplified by the process of pasteurization; the milk needs to be at a high temperature for just a short period of time, so it is exposed to the high temperature and then moved to the surge tank where it can be stored and cooled (see heated surge tank).

A substance can enter the surge tank at room temperature, and it will instantaneously mix with the rest of the tank. Substances entering the tank will also subsequently rise to meet the high temperature and then exit the surge tank quickly thereafter.

1.3.3 Concentration

Concentration inside the surge tank is kept relatively constant, thus the fluid exiting the surge tank is the same as the fluid in the tank. This is favorable when there is a concentration gradient in the incoming fluid to the surge tank. The tank homogenizes the entering fluid, keeping the concentration of the reactants the same throughout the system, therefore eliminating any concentration gradient.

1.4 Modeling Surge Tank: Example 1

Suppose we are to design a surge tank that deals with flow swings of +/- 40% over a 10 minute period modeled by the following equation:

$$w_1 = 500 + 200 sin\frac{\pi t}{10}$$

(1)

where flow is m³h⁻¹ and time in hours. The outlet flow rate, w_o, is desired to be 500 m³h⁻¹.

The surge tank is 30m tall and has a cross sectional area of 5m². The initial height of the fluid ($\varrho = kg/m^3$) is 10m.

[A.] Model this hypothetical example.

[B.] The surge tank has upper and lower limit controls that open or close the valve if the level gets too high or low. What is the highest maximum set point that can be used without the surge tank overflowing?

1.4.1 Solution

[A.] First we should write out the Material Balance:

$$\rho A \frac{dh}{dt} = w_i - w_o, \; h(0) = h_o$$

(2)

$$h(t) = h_o + \frac{1}{\rho A} \int_0^t (w_i(t) - w_o(t)) dt$$

Substituting (1) into (2), we get

$$h(t) = h_o + \frac{2000}{\rho \pi A} \left(1 - \cos\frac{2\pi t}{20}\right)$$

(3)

To complete the design, we must have the cross-sectional area A of the surge tank, this would be given.

If we were to apply the following condition,

$$w_i - w_o = C$$

(4)

and were to substitute (4) into (2), we discover

$$h(t) = h_o + \frac{C}{\rho A} t$$

(5)

478

[B.] Using the Microsoft Excel model provided, the highest possible maximum set point is 23m. Media:Model Surge Tank.xls
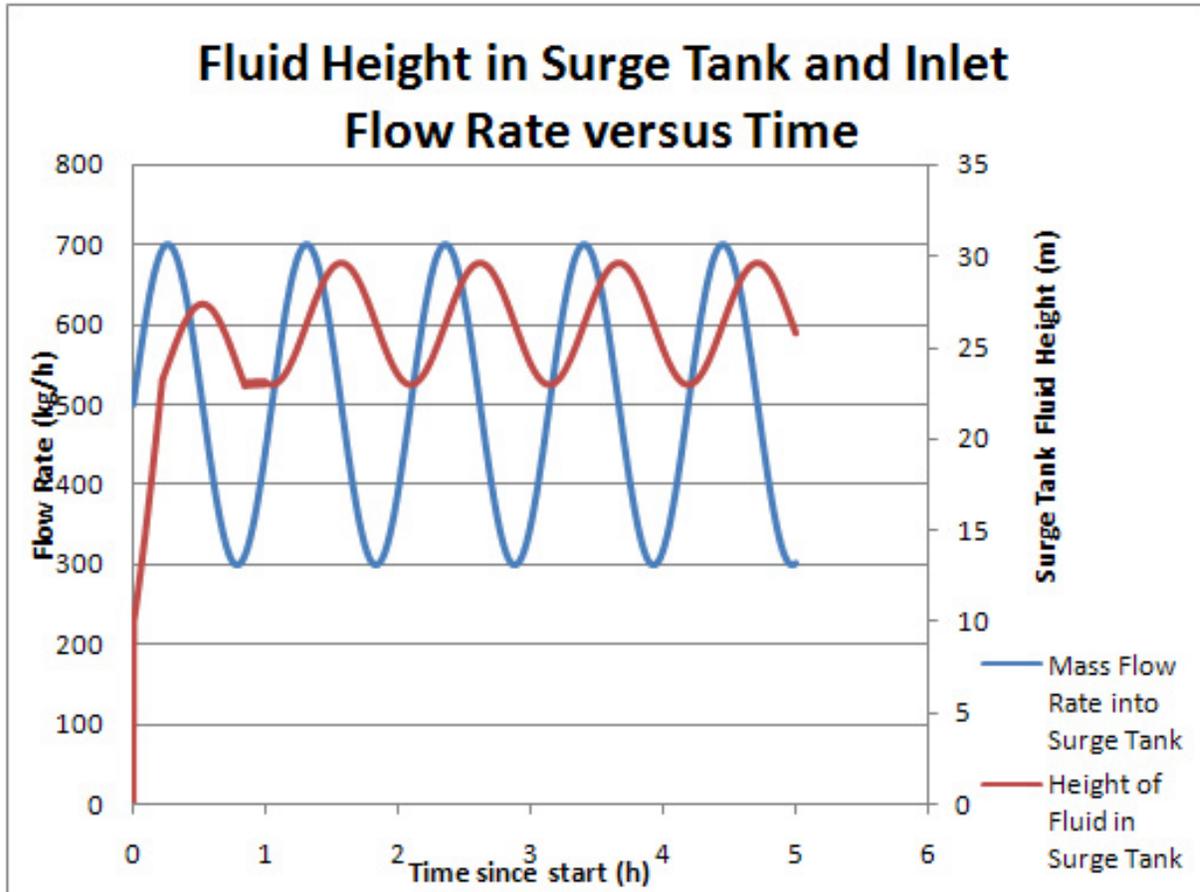
1.5 Modeling Surge Tanks: Example 2

In this example, we simulate random fluctuations that might be observed in a real world process stream. A random number generating function (in MS Excel) is used to generate pseudo-random fluctuations of $\pm$ 200 kg/h about an average value of 1000 kg/h for the inlet stream, $w_i(t)$ over a period of 5 hours. The result is as follows:

479
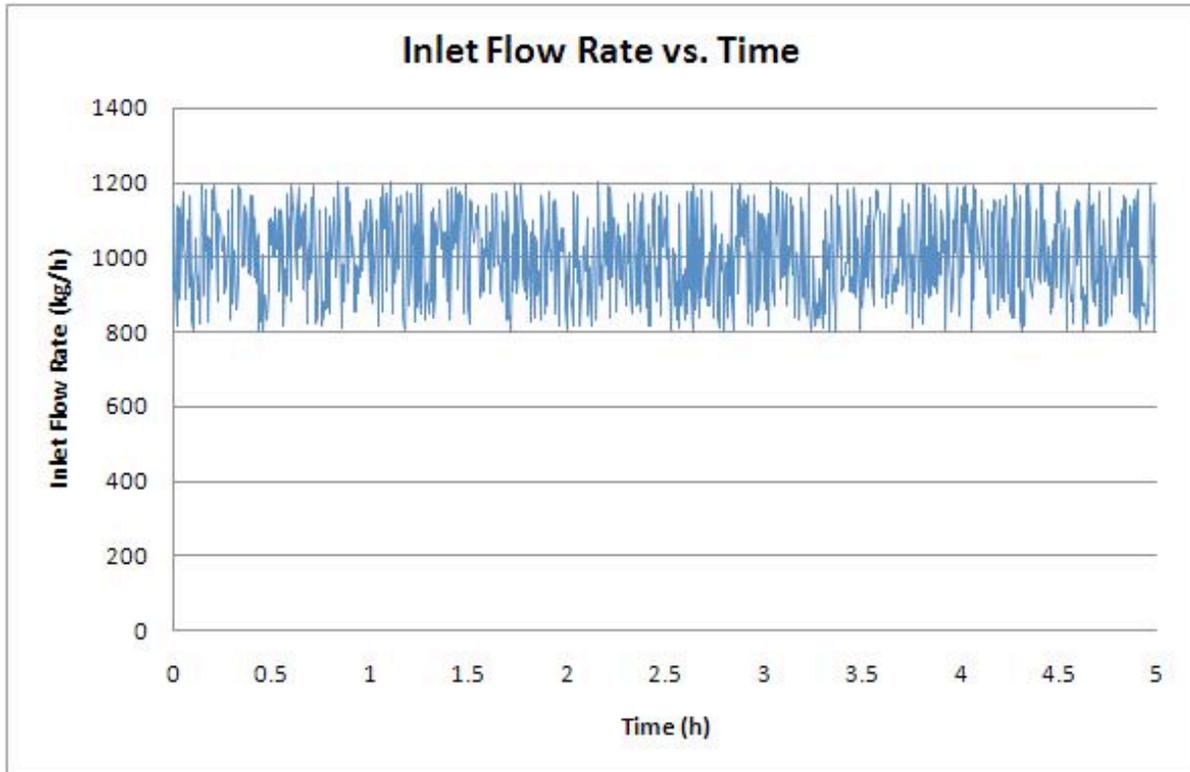
Examine the effect that these fluctuations have on the fluid level in surge tanks of various volumes. To vary the volume, assume tanks of constant height h_{max}=20m, and vary the cross sectional area from A=1m^2 to A=5m^2. Use the following parameters for the surge tank: Initial fluid level h_0=10m, ρ=1 kg / m^3, w_0=1000 kg / m^3, t_0=0 h and t_f=5h.

1.5.1 Solution

In the solution to the first example problem, we used a trigonometric function to simulate fluctuations in the inlet stream, and we obtained an analytical solution to the differential equation. Since we are using a pseudo-random fluctuating inlet stream in this example, we will solve this problem via numerical methods. Using the pseudo-random data for $w_i(t)$, we perform a numerical integration on the previously derived expression:

$$h(t) = h_o + \frac{1}{\rho A} \int_0^t (w_i(t) - w_o(t))dt$$

(2)

This integration was performed with the trapezoid rule (in MS Excel, using a slight modification of the posted Excel model), using the specified surge tank parameters, with A=1 m^2, 2 m^2, and 5 m^2 1. The following results were obtained:

480

Height of Fluid in Surge Tank vs. Time for A=1 m^2

Height of Fluid in Surge Tank vs. Time for A=2 m^2

481

Height of Fluid in Surge Tank vs. Time for A=5 m^2
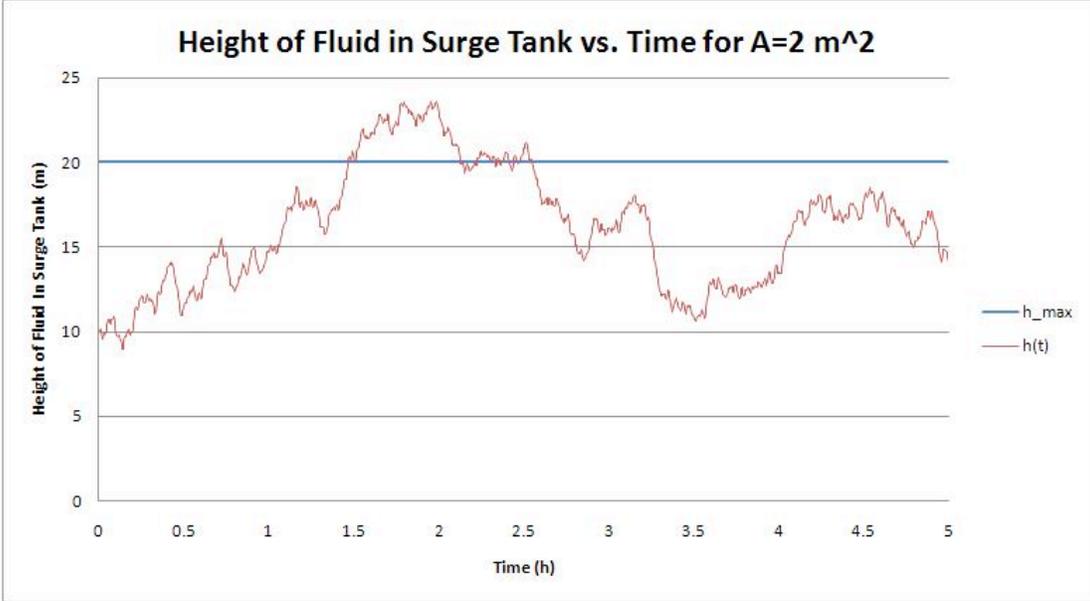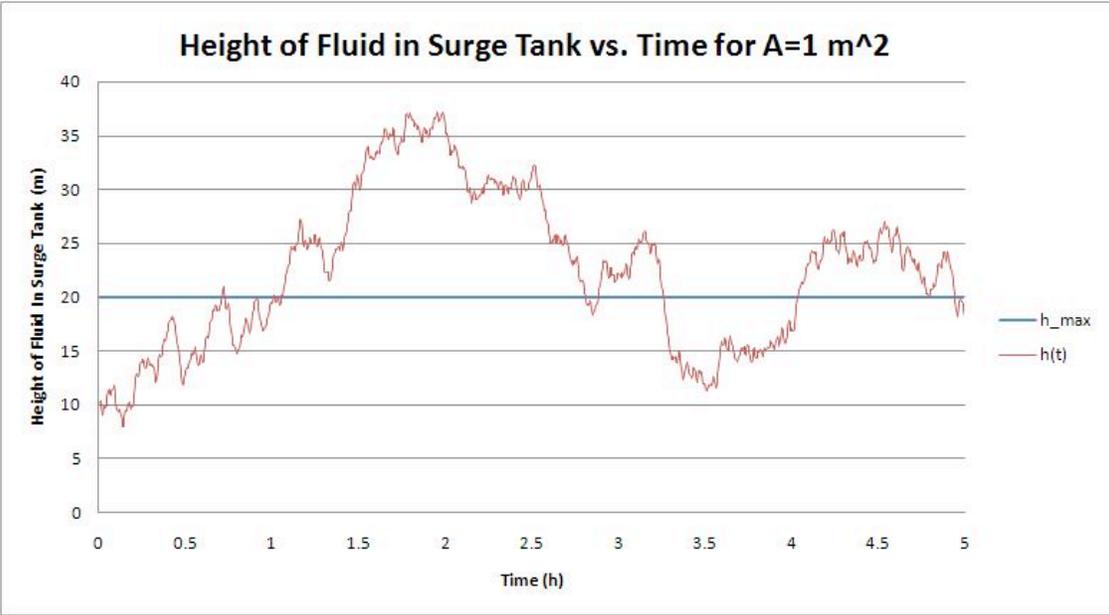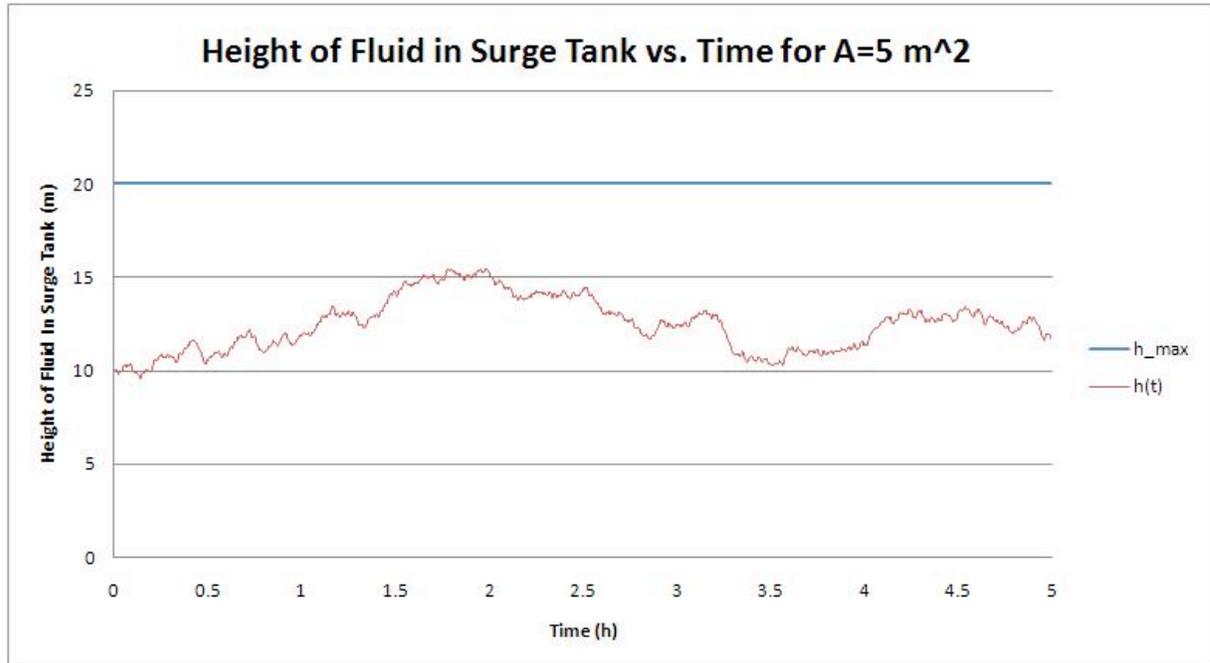
We see that increasing the volume of the surge tank by increasing the cross sectional area A reduces the magnitude of the fluctuations of the fluid level h in the surge tank. For A=1 m^2 and $2m^2$, the capacity of the surge tank is exceeded and it overflows.

1.6 Modeling Surge Tanks: Example 3

An operator quickly adds 50 gallons from a drum of water into a cylindrical surge tank with a diameter of 4 feet. The initial volume of liquid in the tank is 40 cubic feet and the total height of the tank is 5 feet. The inflow and exit flow rates are initally 6 cubic feet per minute. The resistance has a linear relationship with liquid height.

(a) Derive a first order linear model for the tank height and determine the value of R for the outflow

(b) Will the tank overflow when the drum is added?

(c) Show the height h(t) after the drum of water is added; treat the change in h(t) as instantaneous

(d) Does the tank height return to steady state? What is it?

482

V = 50 gal

q_{in}

q_{out}

$h_m = 5$ ft

h
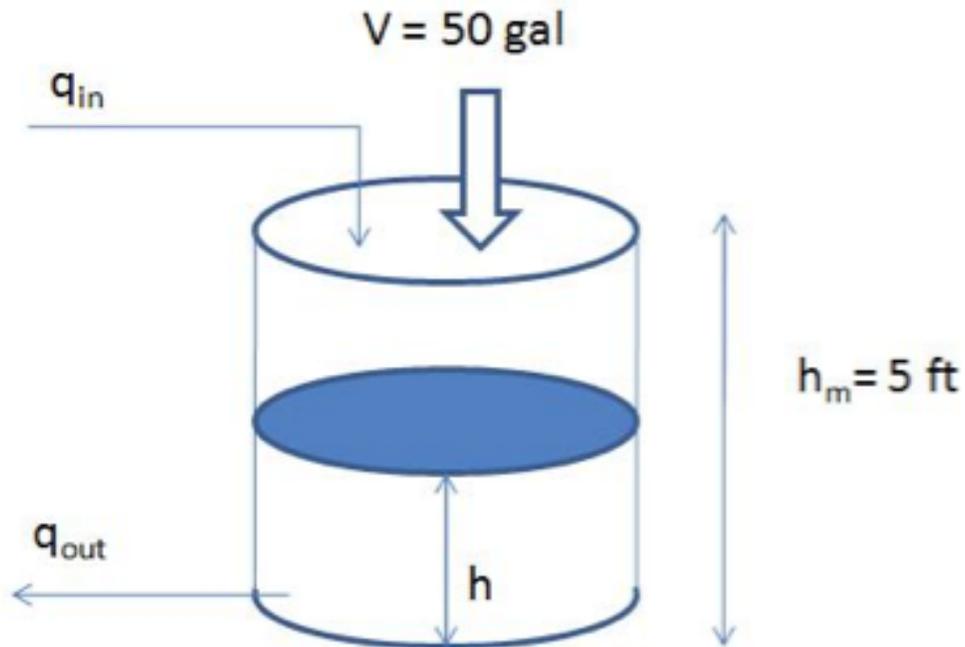
(a)

Assume the operator adds the water at t = 0, so when t < 0, the tank is under steady state which can be described as: $q_{IN} - q_{OUT} = 0$ (1)

$$q_{OUT} = \left(\frac{h}{R}\right)$$

Assume: where R is the resistance and h is the height of the liquid. Initial volume of the tank is:

$$V_{SS} = \pi \left(\frac{d^2}{4}\right) h_{SS}$$

$$40 = \pi \left(\frac{4^2}{4} \right) h_{SS}$$

$$h_{SS} = \left(\frac{10}{\pi} \right) = 3.183 ft$$

Thus:

$$q_{IN} = \left(\frac{6 ft^3}{min} \right) \quad \text{and} \quad q_{OUT} = \left(\frac{h_{SS}}{R} \right)$$

Substitute above: into equation 1

$$0 = 6 - h_{SS} / R$$

$$R = \left(\frac{h_{SS}}{6} \right) = \left(\frac{0.531 min}{ft^2} \right)$$

At time $t > 0$, the 50 gallons (6.685 cubic feet) of water is added to the system which disrupts the steady state and the corresponding initial condition of the system is:

$$V_0 = V_{SS} + V_{operator} = 40 + 6.685 = 46.685 ft^3$$

$$h_0 = \left(\frac{46.685}{4\pi} \right) = 3.7151 ft$$

The corresponding dynamic linear model:

$$A \left(\frac{dh}{dt} \right) = q_{IN} - q_{OUT} = q_{IN} - \left(\frac{h}{R} \right) \quad (2)$$

with intitial condtion: $h_0 = 3.7151 ft$

(b)

After adding the drum, $h_0 = 3.7151$ which is lower than 5 feet. The system will not overflow at t_0. Because the liquid level is higher than the steady state level and

484

$$q_{out} = \left(\frac{h}{R}\right)$$

, the out flow rate will be larger than q_{IN} which pushes the liquid level drop. As a result, water will not over flow.

(c)

Define deviation variables. $h' = h - h_{ss}$; $q_{IN}' = q_{IN} - q_{INSS} = 0$ and substitute into equation (2)

$$A\left(\frac{dh'}{dt}\right) = q_{IN}' - \left(\frac{h'}{R}\right) = -\left(\frac{h'}{R}\right) \quad (3)$$

Solve the linear model above (Eq. 3): (Either by direct integration or laplace transform and inverse lapace transform)

$$h' = h_0' exp\left(-\left(\frac{t}{AR}\right)\right) \quad (4)$$

Put $h' = h - h_{ss}$ into equation 4:

$$h - h_{ss} = (h_0 - h_{ss})exp\left(-\left(\frac{t}{AR}\right)\right)$$

Then:

$$h = h_{ss} + (h_0 - h_{ss})exp\left(-\left(\frac{t}{AR}\right)\right)$$

$$h = 3.183 + (0.5321)exp\left(-\left(\frac{t}{AR}\right)\right)$$

(d)

$$\lim_{t \to \infty} h = 3.183 + 0.5321 exp(-0.1499t) = 3.183$$

Yes, as

ft which was the original steady state

1.8 Sage's Corner

If you are a sage for this page, please link your narrated powerpoint presentation here.

1.9 References

- Kundur, Prabha. *Power System Stability and Control*, McGraw-Hill 1994.
- Cheng-Nan Lin and John S. Gladwell. *Non-dimensional Surge Tank Analysis,* State of Washington Water Research Center, Washington State University 1970.
- Slocum, Stephen Elmer. *Elements of Hydraulics*, McGraw-Hill 1915.
- MIT OpenCourseWare. http://www.ocw.cn/OcwWeb/Chemical-Engineering/10-450Process-Dynamics--Operations--and-ControlSpring2003/LectureNotes/index.htm. Design of a surge tank to smooth out fluctuations in flow. Definition of important process control terms
- Use the Harvard Referencing style for references in the document.
- For more information on when to reference see the following wikipedia entry.

486

Section 2. Heated surge tank
see also ECOSSE

Authors: *Angela Antosiewicz, Christopher Kline, Peter Heisler, Paul Niezguski*
Stewards: *Karen Staubach, Soo Kim, Kerry Braxton-Andrew, Joshua Katzenstein*
Date Released: September 6, 2007
Date Revised: September 9, 2007

2.1 Introduction

A surge tank is an additional safety or storage tank that provides additional product or material storage in case it becomes needed. Heat exchange can be added to surge tanks, which provides temperature control for the tank. Within a system these tanks can appear as distillation columns, reboilers, heated CSTR's, and heated storage. They can increase production rates by allowing a batch of product to finish reacting while the initial tank is reloaded, provide constant system parameters during start up and shut down, or create additional storage space for product overflow or backup material.

Uses for Heated Surge Tanks:

- Fuel surges caused by motion of a vehicle: If fuel cannot be drawn from the primary tank, the engine resorts to a surge tank. The heat maintains the fuel's temperature.

- Caramelization: During the formation of caramel, the mixture must be maintained at a specific temperature for a predetermined amount of time. Once the ingredients are thoroughly dissolved, the mixture is transferred to a heated surge tank and maintained until the caramel has thickened and is ready to be drawn out.

- Mixing of gases: Bulk gas lines can be connected to a heated surge tank with a pressure sensor. The pressure sensor would control the temperature. By heating the gas when it first enters the tank, there is no risk of explosion later due to expansion.

- Heated pools: Surge tanks are used to catch and store displaced water from a pool. If the pool is heated, a heated surge tank should be used to maintain the temperature of the water.

- De-aeration: Heated surge tanks are often used with de-aerators. They heat the component that will enter the de-aerator, because if the component is not

487

preheated, the de-aerator must wait until the component reaches the correct temperature. This could waste a lot of time and energy.

- Chemical Baths: Often in industry, things need to be treated with a chemical bath. The chemicals usually need to be at a certain temperature so that it will adhere to the object. A heated surge tank is perfect for this application.

- Reboilers: Liquids coming off of a distillation column can be reheated to enter the column again at a higher temperature to drive the separation process. Many industries use this tool to obtain a more efficient separation and produce a higher net profit.

- Product or Material Backup: Heated surge tanks can also be used as simple storage in two ways. First, a surge tank can be used excess product not yet sold or otherwise moved out of the production system. Second, heated surge tanks can serve as backup for chemical or fuel supplies to a production plant, such as outdoor gasoline tanks for a backup generator in case of power failure.

2.2 Basic Design for Heated Surge Tanks
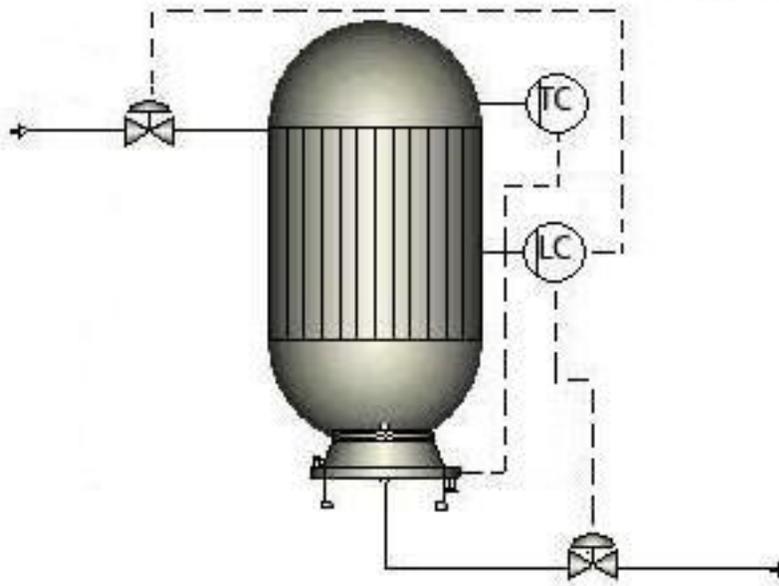
Above is a basic example of a heated surge tank. While surge tanks can have multiple inputs and outputs, for simplicity we have only included one of each here.

488

Connected to the tank is a temperature control, which controls the heater. Depending on the temperature of the fluid, this control will increase or decrease the heating to the tank. This will keep the fluid at the necessary temperature to meet the process requirements.

There is also a level control connected to the tank to indicate when the tank has neared maximum capacity. When this happens, the control will open the valve at the bottom of the tank, allowing the product to flow further down the process. The control can also slow or stop the flow coming through the input valve. These mechanisms will prevent the tank from overfilling. The position of the level control depends on the type of material in the process, the phase of the material, the type of level control, and the requirements of the system.

2.3 Useful Equations

The basic equations that govern heated surge tanks are shown below. First, a simple mass balance is done on the system. Second, the energy balance was simplified using the assumptions listed below. Most problems involving this type of tank can be described by these equations. Additional considerations may require additional variables and equations.

Assumptions:

1) The substance coming into the tank is uniform.

2) No reaction is taking place.

3) The tank is well mixed, which means the temperature profile is constant throughout the tank.

2.3.1 Mass Balance

Since there is no generation from reactions inside the heated surge tank, we obtain the rate of accumulation or level inside the tank by subtracting what is coming out from what is coming in.

$$(Rate\ of\ Accumulation) = (Flow\ In) - (Flow\ Out)$$

2.3.2 Energy Balance

The temperature change with respect to time is essential for the purpose of configuring a system to reach steady state. When turning a system on or off, there is a time period in which the system is in unsteady state. During this time, the system is difficult to model.

In steady state, the system is easier to model because once steady state is reached the left hand term will become zero.

$$\frac{dT}{dt} = \frac{v\rho \, C_p(T_0 - T) + UA(T_C - T)}{V\rho \, C_p}$$

2.4 Case Study - Water Purification at IBM
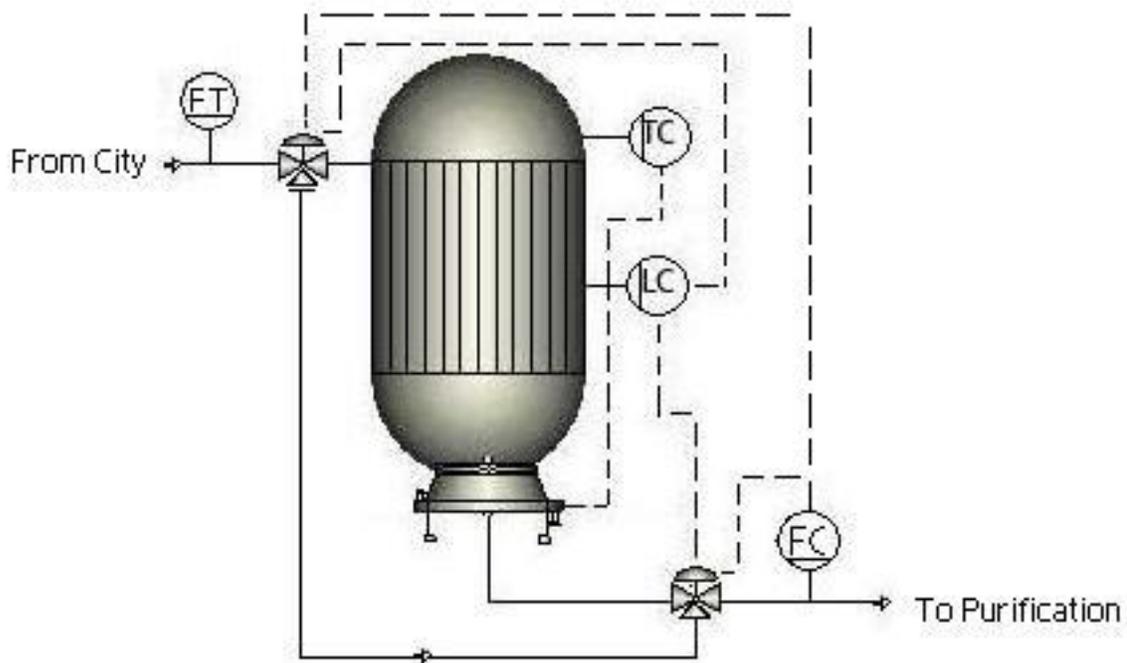
At IBM's manufacturing facility outside Burlington, Vermont, a heated surge tank is used in the de-ionized water system. In order to wash semi-conductor wafers in manufacturing, the water has to be about 1,000,000 times cleaner than the incoming city water. All of this purification is done on site.

The water comes in from the municipal water source at a constant flow rate, but manufacturing demand is not constant. In order to compensate for this, when the demand in manufacturing is low, a surge tank is used to store extra water for high demand periods. Because the large tank is located outside and the winter in Vermont is very cold, the tank is heated to prevent the water inside from freezing.

During normal operation of the system, the surge tank is bypassed. When a flow controller downstream has low demand, the inlet valve opens, letting water into the surge

tank. A level controller monitors the tank to make sure it doesn't overfill and can shut off the inlet valve and let water out. A temperature controller controls the heater jacket to maintain the water around 50°C. When the demand for water increases, the flow controller near the outlet can shut off the inlet valve to the tank, and/or further open the outlet valve to access the extra water supply in the tank.

2.5 Worked out Example 1 - Determining Temperature of Heating Fluid for Tank

A heated surge tank is being designed to hold paraffin wax coming from a distillation column at an oil refinery. High pressure steam will be used as a heating fluid in the heat exchanger to heat up and maintain the paraffin at 51°C (to maintain high viscosity and prevent solidification). The physical parameters of the tank (volume of 5 m^3) and heat exchanger within it are given. The tank is originally filled with paraffin at room temperature. At what temperature must the high pressure steam be to sufficiently heat the paraffin; will a proposed temperature of 130°C be adequate?

The paraffin comes into the tank at 37°C at a volumetric flow rate of 0.0005 m^3/s. The heat exchange coefficient is equal to 50 W/m^2/K and the area of the heat exchanger is 2.0 ^{2}. The heat capacity of the paraffin is 2900 J/Kg/K and the density is 800 Kg/m3.

Worked out Answer to Example 1

POLYMATH CODE:

$d(T)/d(t) = (-v_{in}*rho*C_p*(T-Tin)+(U*A*(T_a-T)))/(C_p*V*rho)$

vin= 0.0005 m^3/s

U= 50 W/m^2/K

A= 2 m^2

Ta= 130°C= 403K

Cp= 2900 J/Kg/K

rho= 800 Kg/m3

Tin= 37°C= 310K

V= 5 m^3

t(0)= 0 s

T(0)= 298K

t(f)= 36000 s
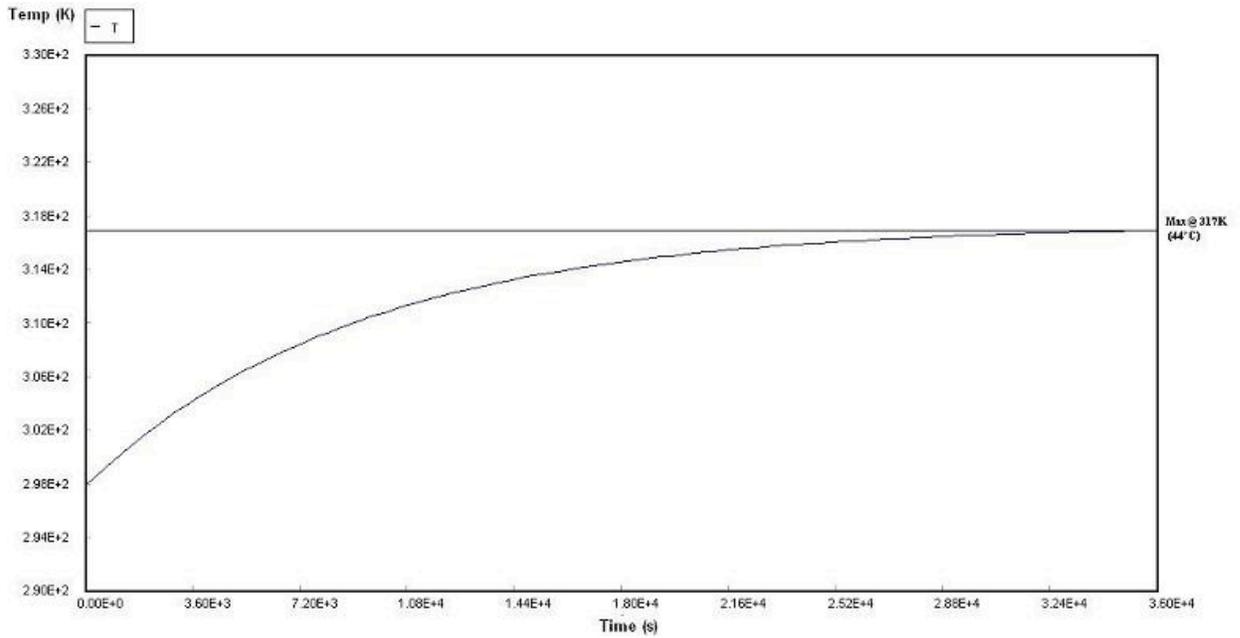
With a heating fluid at a temperature of 130°C, the fluid only reaches a temperature of 44°C (317K). A higher capacity heating fluid must be used.

Trying a heating fluid at 277°C, we generate the plot below.
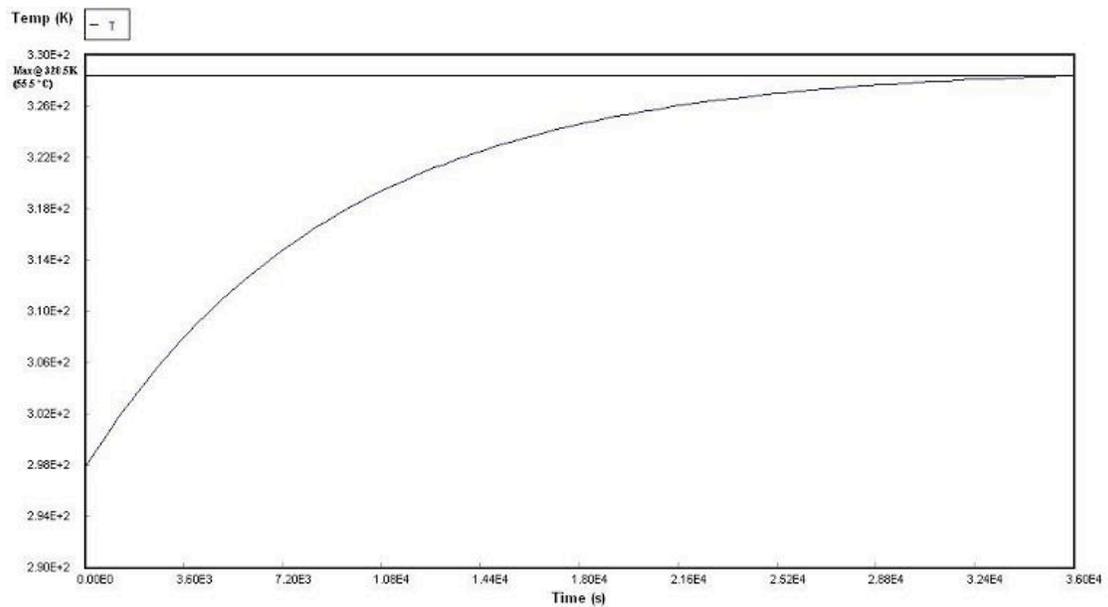
The plot shows that the tank will reach a temperature of about 55.5°C (328.5K) with a heating fluid at 277°C. This will be sufficient to maintain the paraffin in the liquid phase.

2.6 Worked out Example 2 - Time to Reach Steady State after Heating Failure

For the same surge tank from problem 1, if the heater fails for 2 hours after 10 hours of operation and is then restarted, how long will it take after it is restarted to reach steady state?

Answer: Approximately 20 hours

The first manipulation that must be done to your Polymath code is to create an "if-then-else" statement for the dT/dt line for the case before hour 10, the case between hour 10 and hour 12, and the case after hour 12. For the time ranging between 10 and 12 hours, because the heating element has failed, the differential equation over this period must reflect that. Thus, the section (U*A*(Ta-T)) from the differential equation is dropped and the equation appears as follows: d(T)/d(t)= (-vin*rho*Cp*(T-Tin))/(Cp*V*rho). For any other time in this simulation, the normal dT/dt equation is used. In order to determine where the surge tank reaches steady state again, the final time is increased until dT/dt appoximately reaches zero. The graph generated in Polymath will look like the following, using the code below as a parameter:

d(T)/d(t)= if(36000<t and t<42500) then(-vin*rho*Cp*(T-Tin)/(Cp*V*rho)) else((-vin*rho*Cp*(T-Tin)+(U*A*(Ta-T)))/(Cp*V*rho))

2.7 Sage's Corner
http://video.google.com/googleplayer.swf?docId=118567754154073800
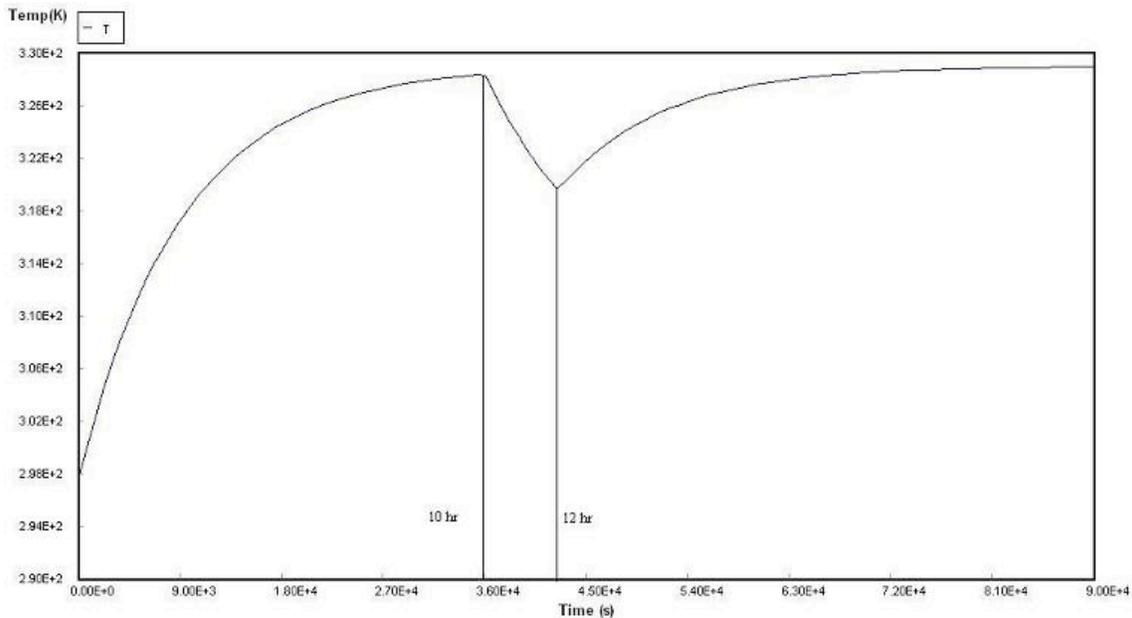
Media:HeatedSurgeTankSlides.ppt

2.8 References

- Use the Harvard Referencing style for references in the document.
- For more information on when to reference see the following wikipedia entry.

Section 3. Bacterial chemostat

Authors: Shoko Asei, Brian Byers, Alexander Eng, Nicholas James, Jeffrey Leto
Date Released: September 18, 2007
Stewards: Jeffrey Falta, Taylor Lebeis, Shawn Mayfield, Marc Stewart, Thomas Welch
Date Revised: September 25, 2007
Stewards: Sarah Hebert, Valerie Lee, Matthew Morabito, Jamie Polan
Date Revised: September 27, 2007

3.1 Introduction

Bioreactors are used to grow, harvest, and maintain desired cells in a controlled manner. These cells grow and replicate in the presence of a suitable environment with media supplying the essential nutrients for growth. Cells grown in these bioreactors are collected in order to enzymatically catalyze the synthesis of valuable products or alter the existing structure of a substrate rendering it useful. Other bioreactors are used to grow and maintain various types of tissue cultures. Process control systems must be used to optimize the product output while sustaining the delicate conditions required for life. These include, but are not limited to, temperature, oxygen levels (for aerobic processes), pH, substrate flowrate, and pressure. A bacterial chemostat is a specific type of bioreactor. One of the main benefits of a chemostat is that it is a continuous process (a CSTR), therefore the rate of bacterial growth can be maintained at steady state by controlling the volumetric feed rate. Bacterial chemostats have many applications, a few of which are listed below.

Applications:

Pharmaceuticals: Used to study a number of different bacteria, a specific example being analyzing how bacteria respond to different antibiotics. Bacteria are also used in the production of therapeutic proteins such as insulin for diabetics.

Manufacturing: Used to produce ethanol, the fermentation of sugar by bacteria takes place in a series of chemostats. Also, many different antibiotics are produced in chemostats.

Food Industry: Used in the production of fermented foods such as cheese.

Research: Used to collect data to be used in the creation of a mathematical model of growth for specific cells or organisms.

The following sections cover the information that is needed to evaluate bacterial chemostats.

3.2 Bacterial Chemostat Design

The bacterial chemostat is a continuous stirred-tank reactor (CSTR) used for the continuous production of microbial biomass.

3.2.1 Chemostat Setup
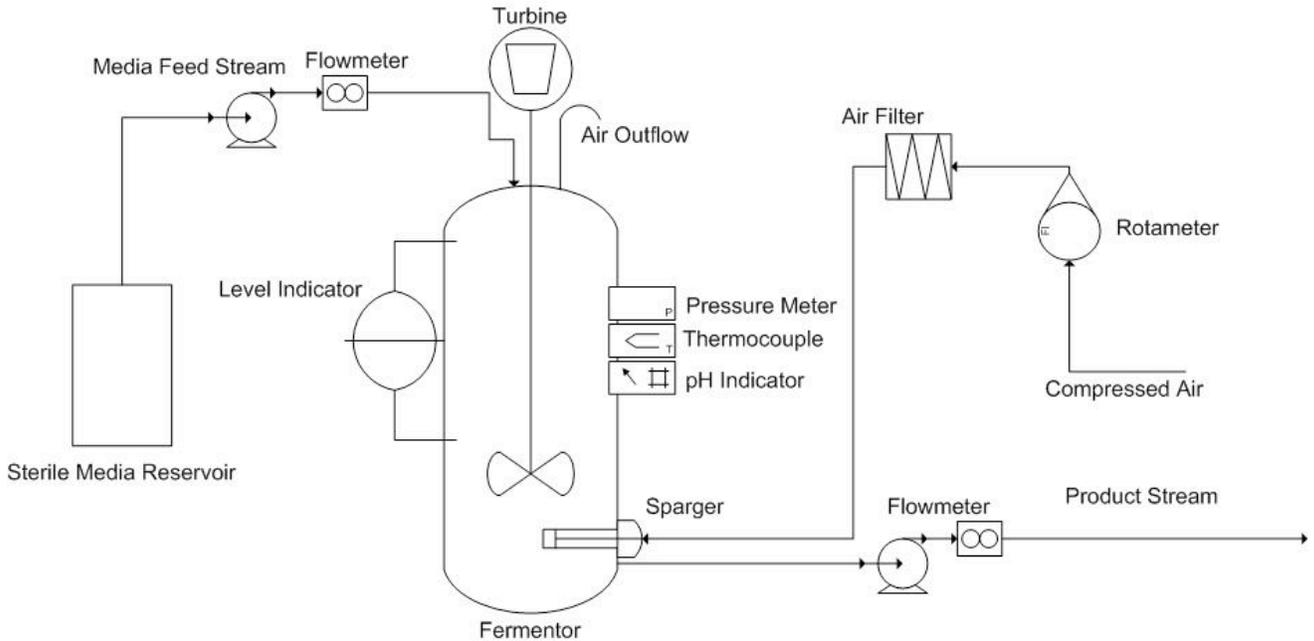
The chemostat setup consists of a sterile fresh nutrient reservoir connected to a growth chamber or reactor. Fresh medium containing nutrients essential for cell growth is pumped continuously to the chamber from the medium reservoir. The medium contains a specific concentration of growth-limiting nutrient (C_s), which allows for a maximum concentration of cells within the growth chamber. Varying the concentration of this growth-limiting nutrient will, in turn, change the steady state concentration of cells (C_c). Another means of controlling the steady state cell concentration is manipulating the rate at which the medium flows into the growth chamber. The medium drips into culture through the air break to prevent bacteria from traveling upstream and contaminating the sterile medium reservoir.

The well-mixed contents of the vessel, consisting of unused nutrients, metabolic wastes, and bacteria, are removed from the vessel and monitored by a level indicator, in order to maintain a constant volume of fluid in the chemostat. This effluent flow can be controlled by either a pump or a port in the side of the reactor that allows for removal of the excess reaction liquid. In either case, the effluent stream needs to be capable of removing excess liquid faster than the feed stream can supply new medium in order to prevent the reactor from overflowing.

Temperature and pressure must also be controlled within the chemostat in order to maintain optimum conditions for cell growth. Using a jacketed CSTR for the growth

496

chamber allows for easy temperature control. Some processes such as biological fermentation are quite exothermic, so cooling water is used to keep the temperature at its optimum level. As for the reactor pressure, it is controlled by an exit air stream that allows for the removal of excess gas.

For aerobic cultures, purified air is bubbled throughout the vessel's contents by a sparger. This ensures enough oxygen can dissolve into the reaction medium. For anaerobic processes, there generally is not a need for an air inlet, but there must be a gas outlet in order to prevent a build up in pressure within the reactor.

In order to prevent the reaction mixture from becoming too acidic (cell respiration causes the medium to become acidic) or too basic, which could hinder cell growth, a pH controller is needed in order to bring pH balance to the system.

The stirrer ensures that the contents of the vessel are well mixed. If the stirring speed is too high, it could damage the cells in culture, but if it is too low, gradients could build up in the system. Significant gradients of any kind (temperature, pH, concentration, etc.) can be a detriment to cell production, and can prevent the reactor from reaching steady state operation.

Another concern in reactor design is fouling. Fouling is generally defined as the deposition and accumulation of unwanted materials on the submerged surfaces or surfaces in contact with fluid flow. When the deposited material is biological in nature, it is called biofouling. The fouling or biofouling in a system like this can cause a decrease in the efficiency of heat exchangers or decreased cross-sectional area in pipes. Fouling on heat exchanger surfaces leads to the system not performing optimally, being outside the target range of temperature, or spending excess energy to maintain optimum temperature. Fouling in pipes leads to an increase in pressure drop, which can cause complications down the line. To minimize these effects, industrial chemostat reactors are commonly cylindrical, containing volumes of up to 1300 cubic meters, and are often constructed from stainless steel. The cylindrical shape and smooth stainless steel surface allow for easy cleaning.

3.2.2 Design Equations

The design equations for contiuous stirred-tank reactors (CSTRs) are applicable to chemostats. Balances have to be made on both the cells in culture and the medium (substrate).

Mass Balance

The mass balance on the microorganisms in a CSTR of constant volume is:

[Rate of accumulation of cells, g/s] = [Rate of cells entering, g/s] – [Rate of cells leaving, g/s] + [Net rate of generation of live cells, g/s]

The mass balance on the substrate in a CSTR of constant volume is:

[Rate of accumulation of substrate, g/s] = [Rate of substrate entering, g/s] – [Rate of substrate leaving, g/s] + [Net rate of consumption of substrate, g/s]

Assuming no cells are entering the reactor from the feed stream, the cell mass balance can be reworked in the following manner:

$$(Rate\ Accumulation\ Cells) = V\frac{dC_C}{dt}$$

(1)

$$(Flow\ Entering) - (Flow\ Leaving) = 0 - \nu_0 C_C$$

(2)

$$(Rate\ Cell\ Generation) = V(r_g - r_d)$$

(3)

Similarly, the substrate mass balance may be reworked in the following manner:

$$(Rate\ Accumulation\ Substrate) = V\frac{dC_S}{dt}$$

(4)

$$(Flow\ Entering) - (Flow\ Leaving) = \nu_0 C_{S0} - \nu_0 C_S$$

(5)

$$(Rate\ Substrate\ Consumption) = V r_S$$

(6)

Putting equations 1, 2, and 3 together gives the design equation for cells in a chemostat:

$$V\frac{dC_C}{dt} = 0 - \nu_0 C_C + V(r_g - r_d)$$

(7)

498

Similarly, equations 4, 5, and 6 together gives the design equation for substrate in a chemostat:

$$V\frac{dC_S}{dt} = v_0 C_{S0} - v_0 C_S + V(r_g - r_d)$$

(8)

Assumptions made about the CSTR include perfect mixing, constant density of the contents of the reactor, isothermal conditions, and a single, irreversible reaction.

Rate Laws

Many laws exist for the rate of new cell growth.

Monod Equation

The Monod equation is the most commonly used model for the growth rate response curve of bacteria.

$$r_g = \mu C_c$$

(9)

where r_g = cell growth rate

C_c = cell cencentration

μ = specific growth rate

The specific cell growth rate, μ, can be expressed as

$$\mu = \mu_{max}\frac{C_s}{K_s + C_s}$$

(10)

where μ_{max} = a maximum specific growth reaction rate

K_s = the Monod constant

C_s = substrate concentration

Tessier Equation and Moser Equation

Two additional equations are commonly used to describe cell growth rate. They are the Tessier and Moser Equations. These growth laws would be used when they are found to better fit experimental data, specifically at the beginning or end of fermentation.

Tessier Equation:

$$r_g = \mu_{max}\left[1 - exp(-\frac{C_s}{k})\right]C_c$$

(11)

Moser Equation:

$$r_g = \frac{\mu_{max}C_s}{1 + kC_s^{-\lambda}}$$

(12)

where λ and k are empirical constants determined by measured data.

Death Rate

The death rate of cells, r_d, takes into account natural death, k_d, and death from toxic by-product, k_t, where C_t is the concentration of toxic by-product.

$$r_d = (k_d + k_t C_t)C_c$$

(13)

Death Phase The death phase of bacteria cell growth is where a decrease in live cell concentration occurs. This decline could be a result of a toxic by-product, harsh environments, or depletion of nutrients.

Stoichiometry

In order to model the amount of substrate and product being consumed/produced in following equations, yield coefficients are utilized. Y_{sc} and Y_{pc} are the yield coefficients for substrate-to-cells and product-to-cells, respectively. Yield cofficients have the units of g variable/g cells. Equation (14) represents the depletion rate of substrate:

$$-r_s = Y_{sc}r_g + mC_c$$

500

Equation (15) represents the rate of product formation:

$$r_p = Y_{pc} r_g$$

3.3 Control Factors

The growth and survival of bacteria depend on the close monitoring and control of many conditions within the chemostat such as the pH level, temperature, dissolved oxygen level, dilution rate, and agitation speed.

As expected with CSTRs, the pumps delivering the fresh medium and removing the effluent are controlled such that the fluid volume in the vessel remains constant.

pH level

Different cells favor different pH environments. The operators need to determine an optimal pH and maintain the CSTR at it for efficient operation. Controlling the pH at a desired value during the process is extremely important because there is a tendency towards a lower pH associated with cell growth due to cell respiration (carbon dioxide is produced when cells respire and it forms carbonic acid which in turn causes a lower pH). Under extreme pH conditions, cells cannot grow properly, therefore appropriate action needs to be taken to restore the original pH (i.e. adding acid or base).

Temperature

Controlling the temperature is also crucial because cell growth can be significantly affected by environmental conditions. Choosing the appropriate temperature can maximize the cell growth rate as many of the enzymatic activates function the best at its optimal temperature due to the protein nature of enzymes.

Dilution rate

One of the important features of the chemostat is that it allows the operator to control the cell growth rate. The most common way is controlling the dilution rate, although other methods such as controlling temperature, pH or oxygen transfer rate can be used. Dilution rate is simply defined as the volumetric flow rate of nutrient supplied to the reactor divided by the volume of the culture (unit: time-1). While using a chemostat, it is useful to keep in mind that the specific growth rate of bacteria equals the dilution rate at steady state. At this steady state, the temperature, pH, flow rate, and feed substrate

concentration will all remain stable. Similarly, the number of cells in the reactor, as well as the concentration of reactant and product in the effluent stream will remain constant.

Negative consequences can occur if the dilution rate exceeds the specific growth rate. As can be seen in Equation (16) below, when the dilution rate is greater than the specific growth rate ($D > \mu$), the dC_C/dt term becomes negative.

$$\frac{dC_C}{dt} = (\mu - D)C_C$$

(16)

This shows that the concentration of cells in the reactor will decrease and eventually become zero. This is called wash-out, where cells can no longer maintain themselves in the reactor. Equation (17) represents the dilution rate at which wash-out will occur.

$$D_{max} = \frac{\mu_{max}C_{s0}}{K_s + C_{s0}}$$

(17)

In general, increasing the dilution rate will increase the growth of cells. However, the dilution rate still needs to be controlled relative to the specific growth rate to prevent wash-out. The dilution rate should be regulated so as to maximize the cell production rate. Figure 1 below shows how the dilution rate affects cell production rate(DC_C), cell concentration (C_C), and substrate concentration (C_S).

502
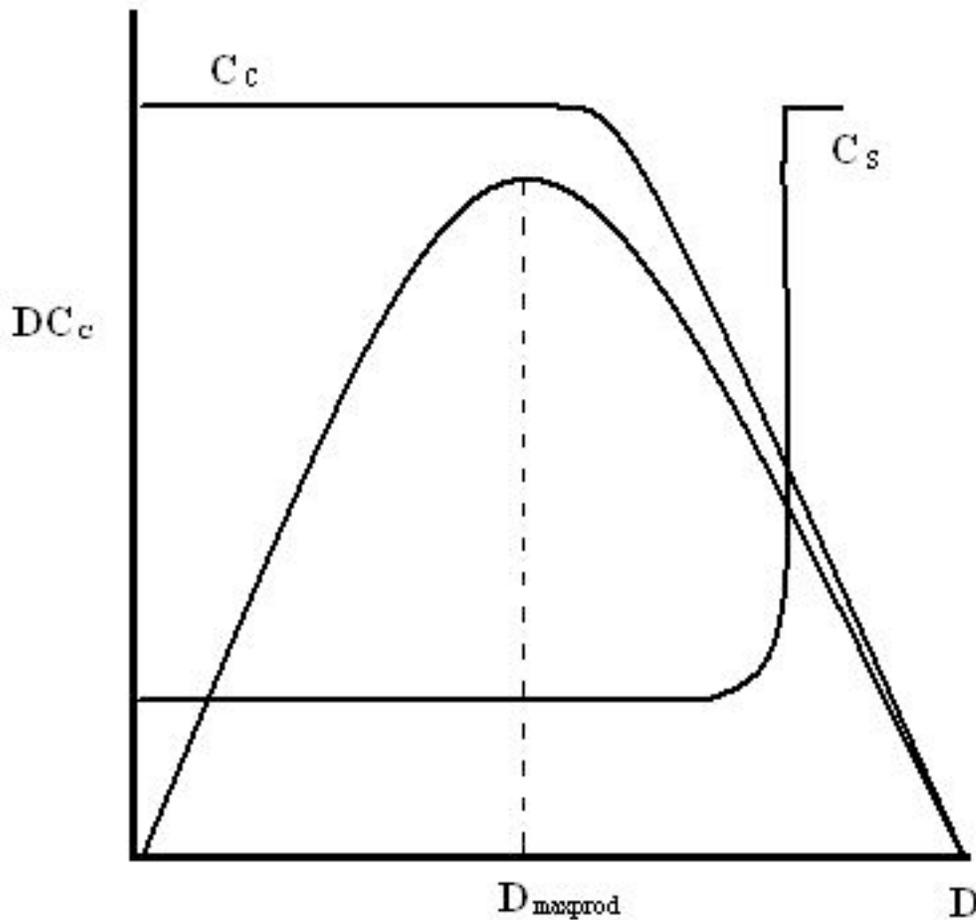
Figure 1: Cell concentration, cell production, and substrate concentration as a function of dilution rate

Initially, the rate of cell production increases as dilution rate increases. When $D_{maxprod}$ is reached, the rate of cell production is at a maximum. This is the point where cells will not grow any faster. $D = \mu$ (dilution rate = specific growth rate) is also established at this point, where the steady-state equilibrium is reached. The concentration of cells (C_C) starts to decrease once the dilution rate exceeds the $D_{maxprod}$. The cell concentration will continue to decrease until it reaches a point where all cells are washed out. At this stage, there will be a steep increase in substrate concentration because fewer and fewer cells are present to consume the substrate.

Oxygen transfer rate

503

Since oxygen is an essential nutrient for all aerobic growth, maintaining an adequate supply of oxygen during aerobic processes is crucial. Therefore, in order to maximize the cell growth, optimization of oxygen transfer between the air bubbles and the cells becomes extremely important. The oxygen transfer rate (OTR) tells us how much oxygen is consumed per unit time when given concentrations of cells are cultured in the bioreactor. This relationship is expressed in Equation (18) below.

$$\text{Oxygen Transfer Rate (OTR)} = Q_{O2}C_C$$

(18)

Where C_C is simply the concentration of cell in the reactor and Q_{O2} is the microbial respiration rate or specific oxygen uptake rate. The chemostat is a very convenient tool to study the growth of specific cells because it allows the operators to control the amount of oxygen supplied to the reactor. Therefore it is essential that the oxygen level be maintained at an appropriate level because the cell growth can be seriously limited if inadequate oxygen is supplied.

Agitation speed

A stirrer, usually automated and powered with a motor, mixes the contents of the chemostat to provide a homogeneous suspension. This enables individual cells in the culture to come into contact with the growth-limiting nutrient and to achieve optimal distribution of oxygen when aerobic cultures are present. Faster, more rigorous stirring expedites cell growth. Stirring may also be required to break agglutinations of bacterial cells that may form.

3.4 Q&A

Q1: Why is a chemostat called a chemostat?

A1: Because the chemical environment is static, or at steady state. The fluid volume, concentration of nutrients, pH, cell density, and other parameters all are assumed to remain constant throughout the operation of the vessel.

Q2: What are some concerns regarding chemostats?

A2: a) Foaming results in overflow so the liquid volume will not be constant. b) Changing pumping rate by turning the pump on/off over short time periods may not work. Cells respond to these changes by altering rates. A very short interval is needed for it to respond correctly. c) Fragile and vulnerable cells can be damaged/ruptured when they are caught between magnetic stirring bar and vessel glass. d)Bacteria contamination

occurs because bacteria travel upstream easily and contaminate the sterile medium. This can be solved by interrupting the liquid path with an air break.

Q3: The Monod equation uses a Michaelis-Menten relationship which is based on a quasi-state assumption. (T/F)

A3: T

Q4: An important feature of chemostat is the dilution rate. Define dilution rate.

A4: Dilution Rate = volume of nutrient medium supplied per hour divided by the volume of the culture.

Q5: What are the advantages/disadvantages over choosing a chemostat instead of a batch reactor for bioreactions?

A5: Advantages: 1. A chemostat has better productivity than a batch reactor. There is a higher rate of product per time per volume. A batch process wastes time. 2. A chemostat is operated at steady state, therefore it has better control maintaining the same conditions for all product produced.

Disadvantages: 1. A chemostat is less flexible than a batch reactor. A batch reactor can be used to make more than one product. 2. It is harder to maintain a sterile system in a chemostat. A batch reactor is easier to clean.

Q6: What is the physical meaning of the Monod constant?

A6: The Monod constant is a substrate concentration at which the growth rate of the biomass of microbial cells participating in the reaction is half the maximum growth rate.

3.5 Worked out Example 1

Note: The context and values given in this problem are not factual.

Researchers at the University of Michigan are using a bacterial chemostat to model the intestinal tract of a pig in order to study the metabolism of E. Coli bacteria in that particular environment. The growth chamber of the chemostat has a volume of 500 dm^3.

The initial concentration of E. Coli bacteria inoculated in the chemostat growth chamber is 1 g/dm^3. A 100g/dm^3 substrate feed is fed to the chemostat at a volumetric flow rate of 20 dm^3/hr. How much time is required for this biochemical process to reach steady rate

from the point of startup? Assume the growth rate is the Monod equation for bacteria bacterial cell growth, shown above.

Additional data pertaining to the problem is given: $\mu_{max} = 0.8$; $K_s = 1.7$ g/dm³; $Y_{s/c} = 8$; $Y_{p/c} = 5$; m = 0; $r_d = 0$;

<center>Schematic Drawing</center>

Cs0 = 100 g/dm^3
v0 = 20 dm^3/hr

V = 500 dm^3
Cc0 = 1 g/dm^3

Cc
Cs
Cp
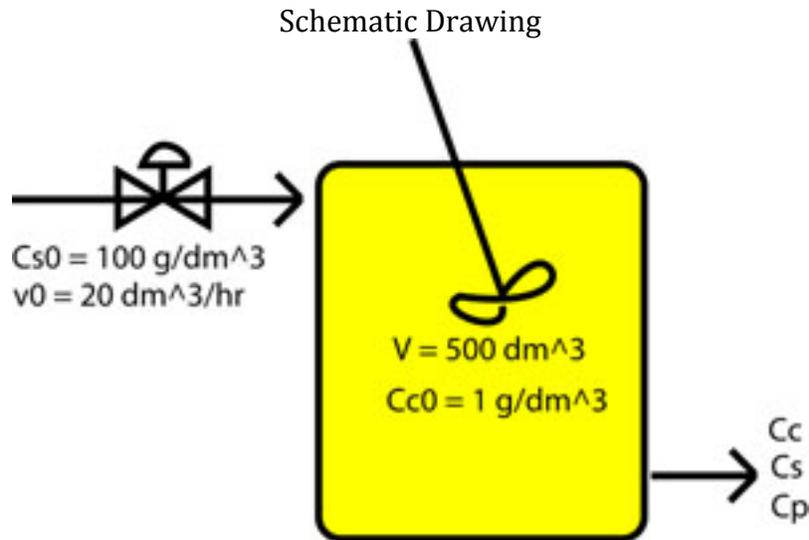
Answer = 3.7 hours

Solution:

The Chemostat was modeled in Excel using the design equations above and Euler's Method. A graph of Cell Concentration (g/dm³) vs Time(hr) was then plotted. When the Cell Concentration become stable, steady state has been reached and the time can be read off the graph. Below is a screen shot of the model and the graph created.

Excel Model Screen Shot

Given Information

Symbols	Values	Units	Description
Cs0	100	g/dm^3	Initial Feed Concentration of Substrate
mumax	0.8	h-1	Max specific growth rate
Ks	1.7	g/dm^3	Monod Constant
V	500	dm^3	Volume of Chemostat
vo	20	dm^3/hr	Initial Volumetric Flowrate
Ys/c	8	g Substrate/g Cells	Yield Coefficient Substrate/Cells
Yp/c	5	g Product/g Cells	Yield Coefficient Product/Cells
Cc0	1	g/dm^3	Initial Concentration of Cells
m	0	g Substrate/(g Cells·	Maintenance Constant for Cells
kvalve	0	dm^3/hr^2	Increase Flowrate Constant
deltat	0.001	hr	Step size

Calculation

t (s)	Cs	dCs/dt	Cc	dCc/dt	Cp	dCp/dt	v	dv/dt	rg
0	100		1		0		20	0	0.786627
0.001	100	-6.293018682	1	0.746627	0	3.933137	20	0	0.786627
0.002	99.9937	-6.293018682	1.000746627	0.746627	0.003933	3.933137	20	0	0.787214
0.003	99.9874	-6.297458876	1.001493255	0.747184	0.007866	3.936069	20	0	0.7878
0.004	99.9811	-6.301899059	1.002240439	0.747741	0.011802	3.939002	20	0	0.788387
0.005	99.9748	-6.306342552	1.002988179	0.748298	0.015741	3.941936	20	0	0.788975
0.006	99.9685	-6.310789355	1.003736477	0.748855	0.019683	3.944873	20	0	0.789562
0.007	99.9622	-6.31523947	1.004485332	0.749413	0.023628	3.947812	20	0	0.790151
0.008	99.9559	-6.3196929	1.005234745	0.749971	0.027576	3.950753	20	0	0.790739

Excel Graph

E. Coli Bacterial Chemostat
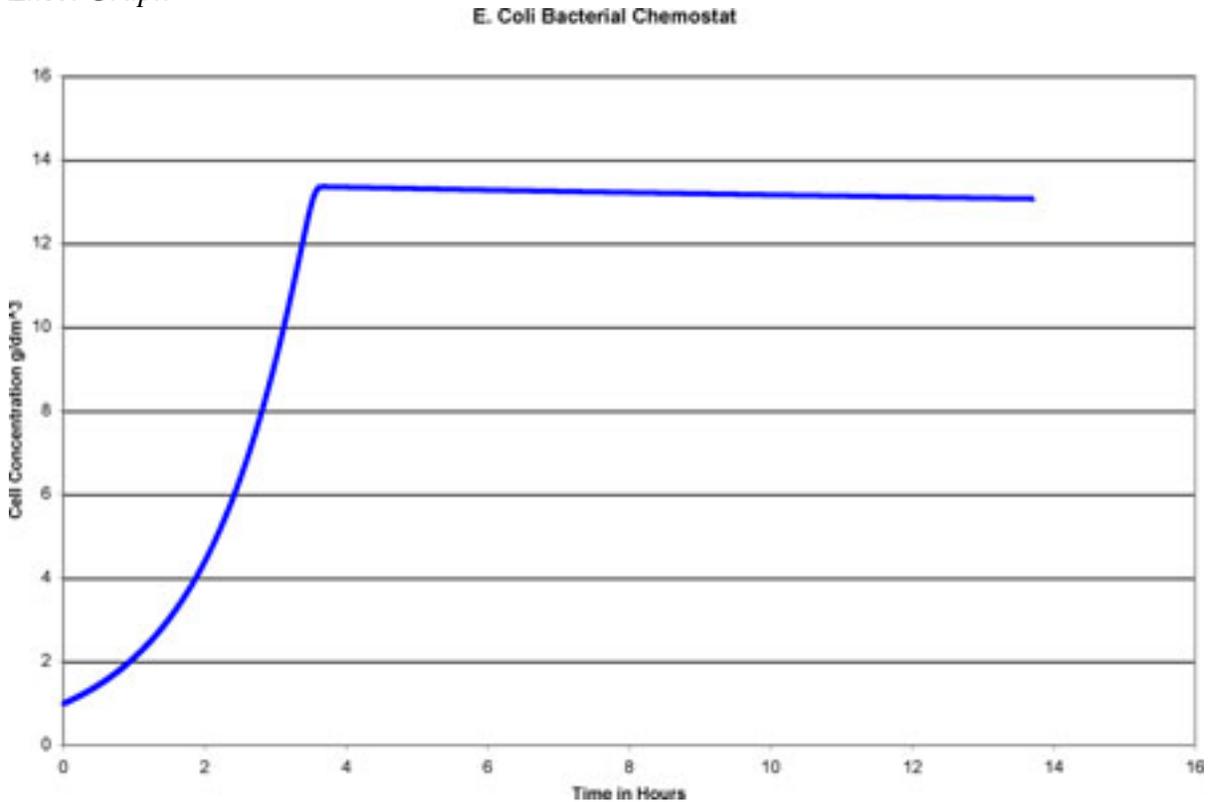
This graph clearly shows that steady state is reached 3.7 hours after start up.

3.6 Worked out Example 2

Note: The context and values given in this problem are not factual.

After calculating the time required to reach steady state, the researchers decide to start up the chemostat. While do so, the control valve for the inlet substrate feed malfunctions. The flow rate of substrate into the chemostat is accelerating at 40 dm^3/hr^2. Determine how long they have to correct the problem before wash-out occurs and all of the bacteria in the chemostat is lost.

Schematic Drawing

Cs0 = 100 g/dm^3
v0 = 20 dm^3/hr
d(v)/dt = 40 dm^3/hr^2

V = 500 dm^3
Cc0 = 1 g/dm^3

Cc
Cs
Cp
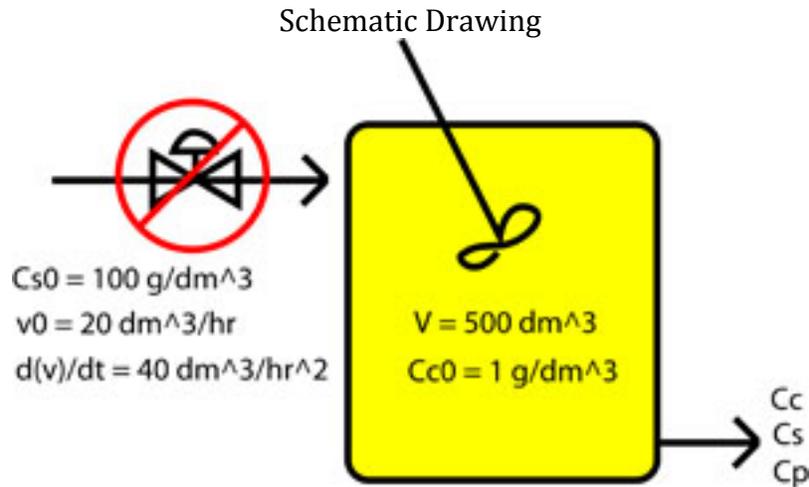
Modeling the Malfunction

$$\frac{dv_0}{dt} = kvalve = 40$$

Answer = 20 hours

Solution:

The Chemostat was modeled in Excel using the design equations above and Euler's Method. A graph of Cell Concentration (g/dm^3) vs Time(hr) was then plotted. When the Cell Concentration becomes zero wash-out of the bacteria took place. Below is a screen shot of the model and the graph created.

Excel Model Screen Shot

Given Information

Symbols	Values	Units	Description
Cs0	100	g/dm^3	Initial Feed Concentration of Substrate
mumax	0.8	h-1	Max specific growth rate
Ks	1.7	g/dm^3	Monod Constant
V	500	dm^3	Volume of Chemostat
vo	20	dm^3/hr	Initial Volumetric Flowrate
Ys/c	8	g Substrate/g Cells	Yield Coefficient Substrate/Cells
Yp/c	5	g Product/g Cells	Yield Coefficient Product/Cells
Cc0	1	g/dm^3	Initial Concentration of Cells
m	0	g Substrate/(g Cells*	Maintenance Constant for Cells
kvalve	40	dm^3/hr^2	Increase Flowrate Constant
deltat	0.001	hr	Step size

Calculation

t (s)	Cs	dCs/dt	Cc	dCc/dt	Cp	dCp/dt	v	dv/dt	rg
0	100		1		0		20	40	0.786627
0.001	100	-6.293018682	1	0.746627	0	3.933137	20.04	40	0.786627
0.002	99.9937	-6.293018682	1.000746627	0.746547	0.003933	3.933137	20.08	40	0.787214
0.003	99.9874	-6.297458876	1.001493175	0.747024	0.007866	3.936069	20.12	40	0.7878
0.004	99.9811	-6.301898556	1.002240199	0.7475	0.011802	3.939001	20.16	40	0.788387
0.005	99.9748	-6.306341041	1.002987699	0.747977	0.015741	3.941935	20.2	40	0.788974
0.006	99.9685	-6.310786331	1.003735675	0.748454	0.019683	3.944871	20.24	40	0.789562
0.007	99.9622	-6.315234427	1.004484129	0.748931	0.023628	3.947809	20.28	40	0.79015
0.008	99.9559	-6.31968533	1.005233059	0.749408	0.027576	3.950748	20.32	40	0.790738

Excel Graph

E. Coli Bacterial Chemostat Malfunction Valve
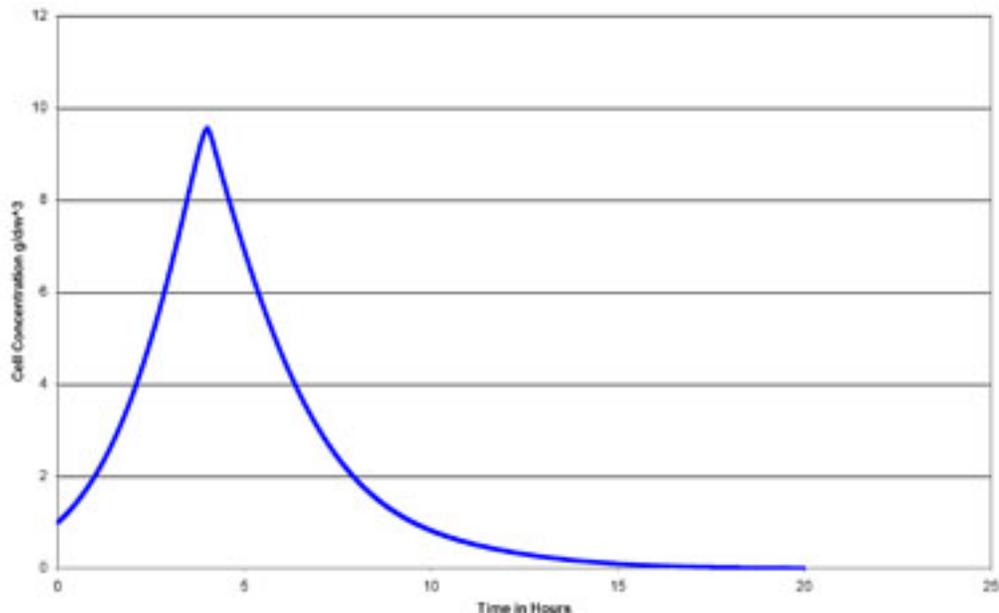
This graph clearly shows wash-out occurs 20 hours after start up. We can see in example that process controls are extremely important for Bacterial Chemostats.

The template model used for both Worked Out Example 1 and 2 can be downloaded here
Media: Bacterial Chemostat Template.xls

3.7 Sage's Corner
Bacterial Chemostats
http://www.youtube.com/v/qVhS0tdJuPQ

3.8 References

- "Chemostat." *McGraw-Hill Dictionary of Scientific and Technical Terms.* McGraw-Hill Companies, Inc., 2003. Accessed September 16, 2007. Available http://www.answers.com/topic/chemostat
- Fogler, H. Scott (2006). *Elements of Chemical Reaction Engineering.* New Jersey: Prentice Hall PTR. ISBN 0-13-047394-4
- Kelly, C. "Ch 9 - Reactors" 17 February, 2004. Retrieved 2007-09-24. Available http://www.ecs.syr.edu/faculty/kelly/handouts/biochemical%20engineering/chapter%209%20reactors.ppt
- Smith, H.L. (2006). "Bacterial Growth". Retrieved on 2007-09-15.
- Strandberg, Per Erik (2004). "Mathematical models of bacteria population growth in bioreactors: formulation, phase space pictures, optimisation and control". Retrieved on 2007-09-16.
- Strandberg, Per Erik (2003). "The Chemostat". Retrieved on 2007-09-15.
- "What is a Bioreactor?" Transgalactic Ltd., 25 May, 2005. Retrieved 2007-09-24. Available http://www.bionewsonline.com/o/what_is_bioreactor.htm

Section 4. ODE & Excel CSTR model w/ heat exchange

Title: ODE & Excel CSTR Modeling with Heat Exchange
Authors: Jason Bourgeois, Michael Kravchenko, Nicholas Parsons, Andrew Wang
Date Presented: 10/3/06 Date Revised: 10/19/06

- First round reviews for this page
- Rebuttal for this page

4.1 Introduction

CSTR with Heat Exchange

A CSTR (Continuous Stirred-Tank Reactor) is a chemical reaction vessel in which an impeller continuously stirs the contents ensuring proper mixing of the reagents to achieve a specific output. Useful in most all chemical processes, it is a cornerstone to the Chemical Engineering toolkit. Proper knowledge of how to manipulate the equations for control of the CSTR are tantamount to the successful operation and production of desired products.

The purpose of the wiki is to model *dynamic conditions* within a CSTR for different process conditions. Simplicity within the model is used as the focus is to understand the dynamic control process.

4.2 Assumptions

For the purposes of this Wiki, we have made the following assumptions to explain CSTR with heat exchange modeling.

Perfect mixing

- The agitator within the CSTR will create an environment of perfect mixing within the vessel. Exit stream will have same concentration and temperature as reactor fluid.

Single, 1st order reaction

- To avoid confusion, complex kinetics are not considered in the following modeling.

$$-r_a = kC_A$$

Parameters specified

- We assume that the necessary parameters to solve the problem have been specified.

Volume specified

- In a control environment, the size of the vessel is usually already specified.

Constant Properties

- For this model, we have made the assumption that the properties of the species we are trying to model will remain constant over the temperature range at which we will be looking at. It is important to make this assumption, as otherwise we will be dealing with a much more complex problem that has a varying heat capacity, heat of reaction, etc.

4.3 ODE Modeling in Excel

To setup the model, the mass and energy balances need to be considered across the reactor. From these energy balances, we will be able to develop relationships for the temperature of the reactor and the concentration of the limiting reactant inside of it.

4.3.1 Variable Definitions

The following table gives a summary of all of the variables that we used in our mathematical formulas.

Symbol Explanations			
Symbol	**Meaning**	**Symbol**	**Meaning**
V	**Volume of Reactor**	ρ	**Density of Stream**
m	**Mass Flow Rate**	C_{A0}	**Original Concentration**
C_A	**Current Concentration**	T_0	**Original Temperature**
T	**Current Temperature**	T_C	**Coolant Temperature**
ΔH_{rxn}	**Heat of Reaction**	ΔC_p	**Overall change in Heat Capacity**
k_0	**Rate Law Constant**	E	**Activation Energy**

R	Ideal Gas Constant	UA	Overall Heat Transfer Coefficient

4.3.2 Mass Balance

From our energy and material balances coursework, we know that the general equation for a mass balance in any system is as follows:

$$(Rate\ Accumulation) = (Flow\ In) - (Flow\ Out) - (Rate\ Generation)$$

$$(Rate\ Accumulation) = V\frac{dC_A}{dt}$$

$$(Flow\ In) - (Flow\ Out) = \frac{m}{\rho}(C_{A0} - C_A)$$

$$(Rate\ Generation) = V(-r_a) = VC_A k_0 e^{-E/RT}$$

In the case of a CSTR, we know that the rate of accumulation will be equal to $V\frac{dC_A}{dt}$. This comes from the fact that the overall number of moles in the CSTR is VC_A, so the accumulation of moles will just be the differential of this. Since V is a constant, this can be pulled out of the differential, and we are left with our earlier result. We also know that the flow of moles in versus the flow of moles out is equal to $\frac{m}{\rho}(C_{A0} - C_A)$, which is the mass flow rate, divided by the density of the flow, and then multiplied by the difference in the concentration of moles in the feed stream and the product stream. Finally, we can determine what the rate of generation of moles in the system is by using the Arrhenius Equation. This will give us the rate of generation equal to $VC_A k_0 e^{-E/RT}$.

Combining all of these equations and then solving for $\frac{dC_A}{dt}$, we get that:

513

$$\frac{dC_A}{dt} = \frac{m}{\rho V}(C_{A0} - C_A) - k_0 C_A e^{-E/RT}$$

4.3.3 Energy Balance

From our thermodynamics coursework, we know that the general equation for an energy balance in any system is as follows:

$$(Rate\ Energy\ Accumulation) = (Heat\ In) - (Heat\ Out) + (Rate\ Heat\ Generation) + (Heat\ Transfer)$$

$$(Rate\ Energy\ Accumulation) = V\rho C_p \frac{dT}{dt}$$

$$(Heat\ In) - (Heat\ Out) = m\Delta C_p(T_0 - T)$$

$$(Rate\ Heat\ Generation) = -V\Delta H_{rxn}(-r_a) = -V\Delta H_{rxn}k_0 C_A e^{-E/RT}$$

$$(Heat\ Transfer) = UA(T_C - T)$$

In the case of a CSTR, we know that the rate of energy accumulation within the reactor will be equal to $V\rho C_p \dfrac{dT}{dt}$. This equation is basically the total number of moles (mass actually) in the reactor multiplied by the heat capacity and the change in temperature. We also know that the heat generated by this reaction is $-V\Delta H_{rxn}k_0 C_A e^{-E/RT}$, which is the rate of mass generation ($-Vr_a$) times the specific heat of reaction (ΔH_{rxn}). The overall rate of heat transfer into and out of the system is given by $m\Delta C_p(T_0 - T)$. This equation is the flow rate multiplied by the heat capacity and the temperature difference, which gives us the total amount of heat flow for the system. Finally, the amount of heat transferred into the system is given by $UA(T_C - T)$. Combining all of these equations and solving the energy balance for $\dfrac{dT}{dt}$, we get that:

514

$$\frac{dT}{dt} = \frac{m\Delta C_p(T_0 - T) - V\Delta H_{rxn}C_A k_0 e^{-E/RT} + UA(T_C - T)}{V\rho\Delta C_p}$$

In a realistic situation in which many chemical processed deal with multiple reactions and heat effects slight changes to the modeled equation must be done. The diagram below evaluates the heat exchanger under heat effects in which there is an inlet and outlet temperature that is accounted for in the enthalpy term in the newly modeled equation.
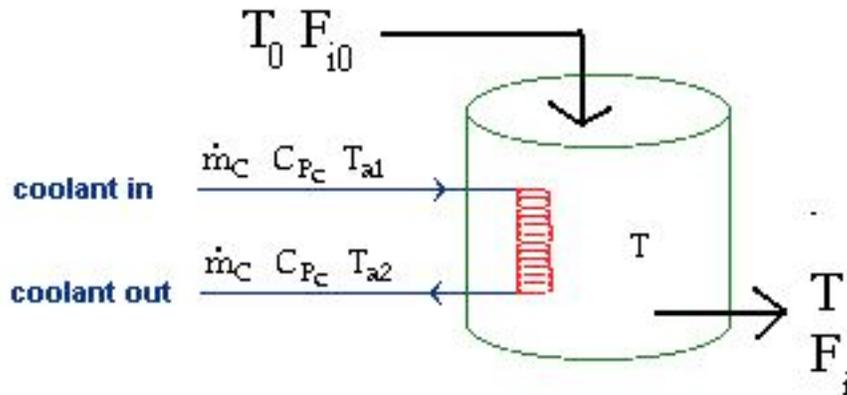
To model a heat exchanger that accounts for multiple reactions simply take the deltaHrxn and deltaCp term and add the greek letter sigma for summation in front of the terms. When considering a case with multiple reactions and heat effects, the enthalpy and heat capacity of each reaction must be implemented in the energy balance, hence i and j represents the individual reaction species. The equation now looks something like this:

$$\frac{dT}{dt} = \frac{m\sum\Delta C_{pij}(T_0 - T) - V\sum\Delta H_{rxnij}C_A k_0 e^{-E/RT} + UA(T_C - T)}{V\rho\sum\Delta C_{pij}}$$

4.3.4 Euler's Method

In order to model an ODE in Excel, we will be using Euler's Method. The reason why we chose Euler's Method over other methods (such as Runge-Kutta) was due to the simplicity associated with Euler's. Trying to model with methods such as Runge-Kutta are much more difficult in Excel, while Euler's method is quite simple. We also checked the accuracy of our results using Euler's Method by comparing our answers with that of Polymath, a highly accurate ODE modeling program. We found that our answers were the same as that of Polymath, so we are quite comfortable using Euler's Method for our model. Here are the full details on Euler's Method. For this model, we will primarily be interested in the change in concentration of a reactant and the temperature of the reactor. These are the two differential equations that we were able to obtain from the mass and energy balances in the previous section. The following is an application of Euler's method for the CSTR on the change in concentration:

$$C_{Ai} = C_{A(i-1)} + \frac{dC_A}{dt}\Delta t$$

This same application can be made to the change in temperature with respect to time:

$$T_i = T_{i-1} + \frac{dT}{dt}\Delta t$$

Assuming that all values in the ODE's remain constant except for T and C_A , the new value is then found by taking the pervious value and adding the differential change multiplied by the time step.

4.3.5 List of Equations

The following are a summary list of all of the equations to be used when modeling CSTR with a heat exchange.

$$k = k_0 e^{-E/RT}$$

$$\frac{dC_A}{dt} = \frac{m}{\rho V}(C_{A0} - C_A) - k_0 C_A e^{-E/RT}$$

$$\frac{dT}{dt} = \frac{m\Delta C_p(T_0 - T) - V\Delta H_{rxn}C_A k_0 e^{-E/RT} + UA(T_C - T)}{V\rho C_p}$$

$$C_{Ai} = C_{A(i-1)} + \frac{dC_A}{dt}\Delta t$$

$$T_i = T_{i-1} + \frac{dT}{dt}\Delta t$$

4.3.6 Combining Using Excel

To model the CSTR using Excel you use Euler's Method, Energy Balance and Mass Balance together to solve for the concentration at time t.

516

	A	B	C	D	E	F	G	H
10	m	4 kg/s		Mass feed rate and product rate				
11	k_0	1.97E+20 s^{-1}		Rate constant				
12	UA	7000 J/s		Heat Transfer Coefficient and Area				
13	T_0	400 K		Feed temperature				
14	$T_{0,2}$	400 K		Change Initial Feed Temperature (leave the same if this				
15	t for $T_{0,2}$	0 s		Time when Initial Feed Temperature Changes				
16	T_c	350 K		Coolant temperature				
17	$T_{c,2}$	350 K		Change Initial Coolant Concentration (leave the same if				
18	t for $T_{c,2}$	0 s		Time when Initial Coolant Temperature Changes				
19	V	3000 L		Reactor volume				
20	ρ	1 kg/L		Density of reactor feed and product (assumed constant)				
21	ΔH_{Rx}	-200 J/mol		Heat of reaction				
22								
23								
24	t (s)	T_0	T (K)	T_c	C_{A0}	C_A	dC_A/dt	dT/dt
25	0	400	400	350	9	9	0.00E+00	0
26	1	400	400	350	9	9	-3.72E-01	-0.04227
27	2	400	399.9577	350	9	8.627994	-3.54E-01	-0.04556
28	3	400	399.9122	350	9	8.273736	-3.37E-01	-0.04869
29	4	400	399.8635	350	9	7.936446	-3.21E-01	-0.05167
30	5	400	399.8118	350	9	7.615362	-3.06E-01	-0.05448

Screenshot of Excel Model for CSTR with Heat Exchange

4.4 How to Use Our Model

In order to help facilitate understanding of this process, we have developed an Excel spreadsheet specifically for looking at the changes in concentration and temperature given some change in the input to the CSTR system. An example of a change to the system could be that the temperature of the feed stream has dropped by a given number of degrees, or that the rate at which the feed stream is being delivered has changed by some amount. By using our spreadsheet, you will be able to easily plug in your given parameters, and look at the trend of the concentration and temperature over a wide time interval.

The way in which this spreadsheet works is quite simple. Boxes are provided for you to input all of the given information for your CSTR problem. Various columns containing values for the temperature, concentration, etc., with respect to time have also been provided. There are then more columns that contain the values for the various differential equations from above. With the time derivative in hand, we are then able to predict the value of the temperature or concentration at the next given time interval.

Our easy-to-use Excel model is given here: CSTR Modeling Template. In our model, you will find a column of unknowns that must be specified in order to solve for the optimal conversion temperature and optimal concentration of A. There are then two cells that will display the optimal temperature and concentration. Graphs are also provided to look at the change in temperature and concentration over time. Most of the variables in the model are self-explanatory. One important feature of our model is the option to have a change in the temperature of the feed stream or the concentration of A after a given time t. You do not need to input a value for these cells if there is no change in the feed; it just provides a convenient way to look at the change of temperature and concentration of A.

You are also provided with a cell for the time step, Δt. Depending on what size time step you choose, you may need to choose a larger value if your graphs do not reach steady state. If this is the case, the output cells will tell you to increase the time step.

4.5 Applications to Process Control

The model created can account for many different situations. However, in process control, it is very important for the following to be considered.

Exothermic reaction - In industry, the exothermic reaction is typically of more importance to controls as there is a safety factor involved. (i.e. explosive reactor conditions)

Volume constant - For CSTRs, liquids are mostly used.

The model developed can assist in controlling an actual process in many ways including the following, but not limited to:

- Maintaining product concentration despite changes operating conditions
- Predicting and preventing explosive reactor conditions
- Flow rate optimization
- Input concentration optimization
- Operating temperature optimization
- Coolant temperature optimization

This model is very useful for many simple reactor situations and will aid in the understanding of a process and how dynamic situations will affect operation.

In real-life applications, the reaction may not be first order or irreversible or involve multiple reactions. In all these cases, we modify the Energy and Mass Balance with additional rate law considerations.

518

You are contracted by WOW Chemical to control the operation of their 3000 L jacketed CSTR. They desire to create chemical B from chemical A at an optimal conversion. What is the temperature at which the optimal conversion is achieved?

A → B is a first order, irreversible reaction.

Some information about the process:

Feed stream temperature = 400 K

Coolant temperature = 350 K

Heat of reaction = -200 J/mol

Inlet concentration of A = 9 mol/L

Inlet flow rate = 4 kg/s

Density of A = 1000 g/L

UA of the heat exchanger = 7 kcal/s

Rate constant = 1.97×10^{20} s^{-1}

Activation energy = 166 kJ/mol

Overall change in Heat capacity = 1 kcal/kg-K.

$T_0 = 400$ K
$m = 4$ kg/s
$\rho = 1000$ g/L
$C_{A0} = 9$ mol/L

$T_{coolant} = 350$ K

© CSTR Heat Exchange Model
for ChE 466, Fall 2006

UA = 7 kcal/s

$k_0 = 1.97 \times 10^{20}$ s^{-1}
$E = 166$ kJ/mol

$\Delta C_p = 1$ kcal/kg-K
$\Delta H_{rxn} = -200$ J/mol
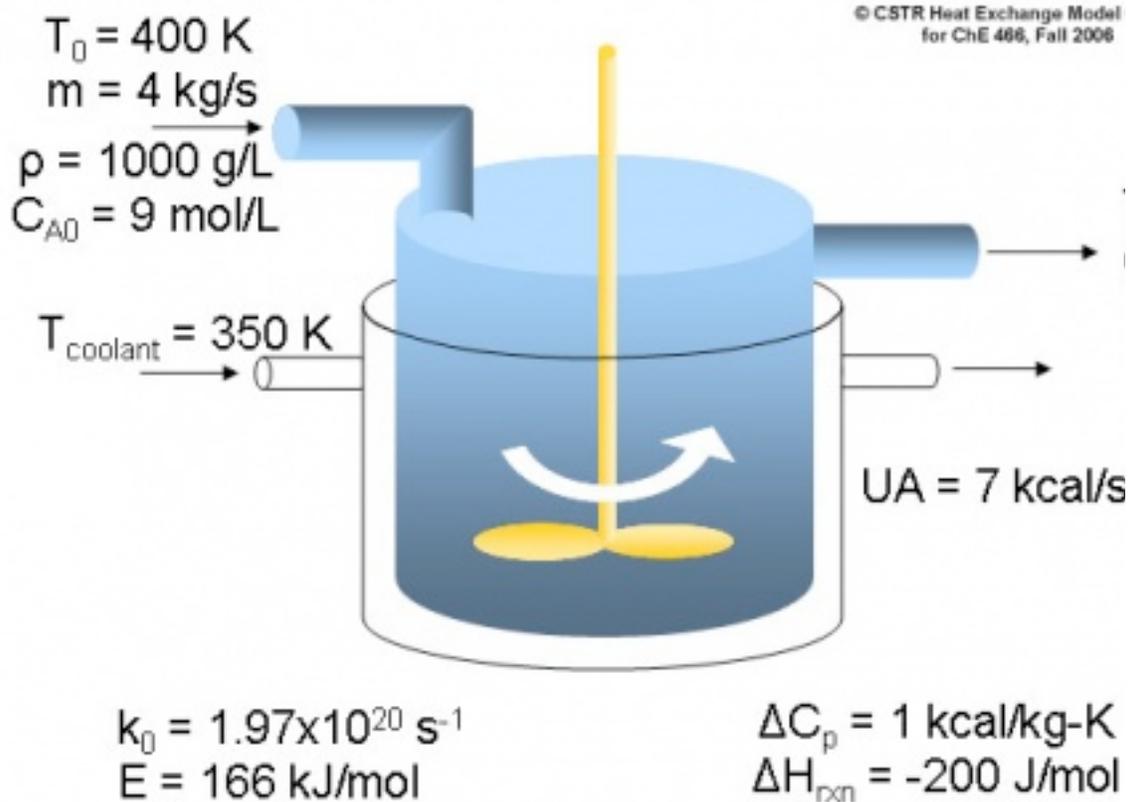
4.6.1 Answer to Example Problem #1

From using our handy-dandy worksheet, we see that the optimal conversion occurs at 368.4 K. This comes from inputting all of the given information from the problem statement into the Excel sheet, and then reading off the value for the optimal temperature. Also, make sure you are converting to proper units, such as from kJ to J, for use of our model (just to see if you were paying attention to units).

Example 1 Solution

520

	A	B	C	D
1	Enter Given Information HERE			Description
2	C_{A0}	9	mol/L	Initial Feed Concentration
3	$C_{A0,\,2}$	9	mol/L	Change Initial Feed Concentration (leave the same if this does not occur)
4	t for $C_{A0,\,2}$		s	Time when Initial Feed Conc changes
5	Δt	1	s	Time Step
6	ΔC_p	1000	J/kg-K	heat capacity of the reactor feed and product
7	E	166000	J/mol	Activation energy
8	m	4	kg/s	Mass feed rate and product rate
9	k_0	1.97E+20	s^{-1}	Rate constant
10	UA	7000	J/s	Heat Transfer Coefficient and Area
11	T_0	400	K	Feed temperature
12	$T_{0,\,2}$	400	K	Change Initial Feed Temperature (leave the same if this does not occur)
13	t for $T_{0,\,2}$		s	Time when Initial Feed Temperature Changes
14	T_c	350	K	Coolant temperature
15	$T_{c,\,2}$	350	K	Change Initial Coolant Concentration (leave the same if this does not occur)
16	t for $T_{c,\,2}$		s	Time when Initial Coolant Temperature Changes
17	V	3000	L	Reactor volume
18	ρ	1	kg/L	Density of reactor feed and product (assumed constant)
19	ΔH_{Rx}	-200	J/mol	Heat of reaction
20				
21	Optimal Conversion Values			
22	T	368.377		K
23	C_A	6.301		mol/L

4.7 Worked out Example 2

1. In the previous example, the plant was operating in the daytime. At night, there is a drop of 5 K in feed stream temperature. In order to maintain the same concentration do we need to...

A. Decrease the temperature of the cooling water in the heat exchanger.

B. Increase the temperature of the cooling water in the heat exchanger.

C. Maintain the same temperature of the cooling water in the heat exchanger.

2. At night, the inlet flow rate of the system also decreases due to the plant not operating at full capacity. What happens to the temperature of the reaction if concentration of A in the feed stream stays the same?

A. Decrease

B. Increase

C. Stay the same

3. Finally, if the rate constant doubled, what effect will this have on the overall concentration if temperature is held constant?

A. Increase

B. Decrease

C. Stay the same

4.7.1 Answers to Example Problem #2

1: B. Since the temperature of the feed stream has decreased, we will need to increase the temperature of the cooling stream in order to maintain the same temperature at which the CSTR was operating at before. Using the Excel spreadsheet from Example 1 and lowering the feed stream temperature by 5 K, we can first see that the temperature of the CSTR has fallen to 366.5 K.

	A	B	C	D
1	Enter Given Information HERE			Description
2	C_{A0}	9	mol/L	Initial Feed Concentration
3	$C_{A0, 2}$	9	mol/L	Change Initial Feed Concentration (leave the same if this does not occur)
4	t for $C_{A0, 2}$		s	Time when Initial Feed Conc changes
5	Δt	1	s	Time Step
6	ΔC_p	1000	J/kg-K	heat capacity of the reactor feed and product
7	E	166000	J/mol	Activation energy
8	m	4	kg/s	Mass feed rate and product rate
9	k_0	1.97E+20	s^{-1}	Rate constant
10	UA	7000	J/s	Heat Transfer Coefficient and Area
11	T_0	395	K	Feed temperature
12	$T_{0, 2}$	395	K	Change Initial Feed Temperature (leave the same if this does not occur)
13	t for $T_{0, 2}$		s	Time when Initial Feed Temperature Changes
14	T_c	350	K	Coolant temperature
15	$T_{c, 2}$	350	K	Change Initial Coolant Concentration (leave the same if this does not occur)
16	t for $T_{c, 2}$		s	Time when Initial Coolant Temperature Changes
17	V	3000	L	Reactor volume
18	ρ	1	kg/L	Density of reactor feed and product (assumed constant)
19	ΔH_{Rx}	-200	J/mol	Heat of reaction
20				
21	Optimal Conversion Values			
22	T	366.524		K
23	C_A	6.785		mol/L

We can then adjust the temperature of the feed stream and see that it must be raised by 3 K in order to return to the original operating temperature of 368.4 K.

522

	A	B	C	D
1	**Enter Given Information HERE**			Description
2	C_{A0}	9	mol/L	Initial Feed Concentration
3	$C_{A0, 2}$	9	mol/L	Change Initial Feed Concentration (leave the same if this does not occur)
4	t for $C_{A0, 2}$		s	Time when Initial Feed Conc changes
5	Δt	1	s	Time Step
6	ΔC_p	1000	J/kg-K	heat capacity of the reactor feed and product
7	E	166000	J/mol	Activation energy
8	m	4	kg/s	Mass feed rate and product rate
9	k_0	1.97E+20	s^{-1}	Rate constant
10	UA	7000	J/s	Heat Transfer Coefficient and Area
11	T_0	395	K	Feed temperature
12	$T_{0, 2}$	395	K	Change Initial Feed Temperature (leave the same if this does not occur)
13	t for $T_{0, 2}$		s	Time when Initial Feed Temperature Changes
14	T_c	353	K	Coolant temperature
15	$T_{c, 2}$	353	K	Change Initial Coolant Concentration (leave the same if this does not occur)
16	t for $T_{c, 2}$		s	Time when Initial Coolant Temperature Changes
17	V	3000	L	Reactor volume
18	ρ	1	kg/L	Density of reactor feed and product (assumed constant)
19	ΔH_{Rx}	-200	J/mol	Heat of reaction
20				
21	Optimal Conversion Values			
22	T	368.470		K
23	C_A	6.276		mol/L

2: A. Since there is less feed coming into the CSTR, the optimal temperature of the reactor will decrease as well. This can again be seen by using our Excel model from Example 1. We know that originally the temperature of the reactor was at 368.4 K. By decreasing the flow rate to 3 kg/s, we can see that the temperature of the reactor decreases by over 3 K down to 365.1 K.

	A	B	C	D
1	**Enter Given Information HERE**			Description
2	C_{A0}	9	mol/L	Initial Feed Concentration
3	$C_{A0,\,2}$	9	mol/L	Change Initial Feed Concentration (leave the same if this does not occur)
4	t for $C_{A0,\,2}$		s	Time when Initial Feed Conc changes
5	Δt	1	s	Time Step
6	ΔC_p	1000	J/kg-K	heat capacity of the reactor feed and product
7	E	166000	J/mol	Activation energy
8	m	3	kg/s	Mass feed rate and product rate
9	k_0	1.97E+20	s^{-1}	Rate constant
10	UA	7000	J/s	Heat Transfer Coefficient and Area
11	T_0	400	K	Feed temperature
12	$T_{0,\,2}$	400	K	Change Initial Feed Temperature (leave the same if this does not occur)
13	t for $T_{0,\,2}$		s	Time when Initial Feed Temperature Changes
14	T_c	350	K	Coolant temperature
15	$T_{c,\,2}$	350	K	Change Initial Coolant Concentration (leave the same if this does not occur)
16	t for $T_{c,\,2}$		s	Time when Initial Coolant Temperature Changes
17	V	3000	L	Reactor volume
18	ρ	1	kg/L	Density of reactor feed and product (assumed constant)
19	ΔH_{Rx}	-200	J/mol	Heat of reaction
20				
21	Optimal Conversion Values			
22	T	365.139		K
23	C_A	6.610		mol/L

3: B. We see in the equations for the differential equations that the rate constant has a negative effect on the conctration of A. By doubling its value, we will then be decreasing C_A as well. This can also be seen in the Excel model from Example 1. We see that the original concentration of A is 6.3 mol/L. We can then double the value of the rate constant, and we see that the concentration of A drops down to 4.8 mol/L, which is what we had predicted would happen.

	A	B	C	D
1	Enter Given Information HERE			Description
2	C_{A0}	9	mol/L	Initial Feed Concentration
3	$C_{A0,\,2}$	9	mol/L	Change Initial Feed Concentration (leave the same if this does not occur)
4	t for $C_{A0,\,2}$		s	Time when Initial Feed Conc changes
5	Δt	1	s	Time Step
6	ΔC_p	1000	J/kg-K	heat capacity of the reactor feed and product
7	E	166000	J/mol	Activation energy
8	m	4	kg/s	Mass feed rate and product rate
9	k_0	3.94E+20	s^{-1}	Rate constant
10	UA	7000	J/s	Heat Transfer Coefficient and Area
11	T_0	400	K	Feed temperature
12	$T_{0,\,2}$	400	K	Change Initial Feed Temperature (leave the same if this does not occur)
13	t for $T_{0,\,2}$		s	Time when Initial Feed Temperature Changes
14	T_c	350	K	Coolant temperature
15	$T_{c,\,2}$	350	K	Change Initial Coolant Concentration (leave the same if this does not occur)
16	t for $T_{c,\,2}$		s	Time when Initial Coolant Temperature Changes
17	V	3000	L	Reactor volume
18	ρ	1	kg/L	Density of reactor feed and product (assumed constant)
19	ΔH_{Rx}	-200	J/mol	Heat of reaction
20				
21	Optimal Conversion Values			
22	T	368.486	K	
23	C_A	4.818	mol/L	

4.8 References

- Bequette, B. Wayne. *Process Dynamics Modeling, Analysis, and Simulation*, New Jersey: Prentice Hall.
- Fogler, H. Scott. *The Elements of Chemical Reaction Engineering*, Prentice Hall. ISBN 0-13-047394-4
- Kravaris, Costas. *Chemical Process Control: A Time-Domain Approach*, The University of Michigan.
- Riggs, James B.; Karim, M. Nazmul. *Chemical and Bio-Process Control*, Texas Tech University, Ferret Publishing.

Section 5. ODE & Excel model of a simple distillation column

Title: ODE & Excel Modeling of a Simple Distillation Column
Authors: Jennifersue Bowker, George Cater, Kibwe John
Date Presented: 10/3/2006 /Date Revised: 9/24/2006

- First round reviews for this page
- Rebuttal for this page

5.1 Introduction

Distillation is a commonly employed separation technique bases on difference in volatilities. The modern form of distillation as is is known today may be credited to early Arab alchemist, Jabir ibn Hayyan and the development of one of his inventions, the alembic. The distillation apparatus is commonly referred to as a still and consists of a minimum of a reboiler in which mixture to be separated is heated, a condenser in which the vapor components are cooled back to liquid form, and a receiver in which the concentrated liquid component fractions are collected. Ideally, distillation is governed by the principles of Raoult's Law and Dalton's Law. Dalton's Law states that for a mixture, the total vapor pressure is equal to the sum of the individual vapor pressures of the pure components which comprise this mixture. The relationship giving the vapor pressure of a volatile component in a mixture, P_A, is Raoult's Law and is governed by the following equation:

$$P_A = X_A P_A°$$

Where X_A is the mole fraction of component A in the mixture and $P_A°$ is the vapor pressure of pure component A. This ideal model is based on a binary mixture of benzene and toluene but for other mixtures severe deviations from Raoult's Law may be observed due to molecular interactions. For these aforementioned mixtures where the components are not similar the only accurate alternative is obtaining vapor-liquid equilibrium by measurement.

In simple distillation, two liquids with differing boiling points are separated by immediately passing the vapors from the reboiler to a condensing column which condenses the vapor components. As a result the distillate is not usually pure but its composition may be determined by Raoult's Law at the temperature and pressure at which the vapors were boiled off. Consequently, simple distillation is usually used to separate binary mixtures where the boiling temperatures of the individual components are usually significantly different or to separate volatile liquids from non-volatile solids.

A reference of terms used in distillation reference are included at the end of this article.

526

5.2 Distillation Control

Distillation columns comprise an enormous amount of the separation processes of chemical industries. Because of their wide range of uses in these industries and because their proper operation contributes to product quality, production rates and other capital costs, it is clear that their optimization and control is of great importance to the chemical engineer. Distillation control becomes problematic because of the wide variety of thermodynamic factors stemming from the separation process.

For example:

- Separations deviate from linearity of equations as product purity increases.
- Coupling of process variables occurs when compositions are controlled.
- Disturbances occur due to feed and flow agitation
- Efficiency changes in trays lead to non-steady state behavior.

In order to improve upon distillation control you must be able to characterize these potential problems and realize when they occur because they lead to dynamic behavior of the column.

Of key importance to control is the maintenance of material and energy balances and their due effects on the column. Shown below is a schematic of a simple binary distillation column. Using the material balance formulas $D/F = (z-x)/(y-x)$, where z, x, and y are the feed, bottoms and distillate concentrations respectively, you find that as D (Distillate) increases, its purity decreases. This leads to the idea that purity level varies indirectly with the flow rate of that product. Energy input is also key because it determines the vapor flow rate (V) up the column which has direct effects on the L/D ratio (reflux ratio) and hence relates to an increase in the amount of separation occurring. To summarize, energy input determines the amount of separation, while material flow relates the ratio of separation in the products.
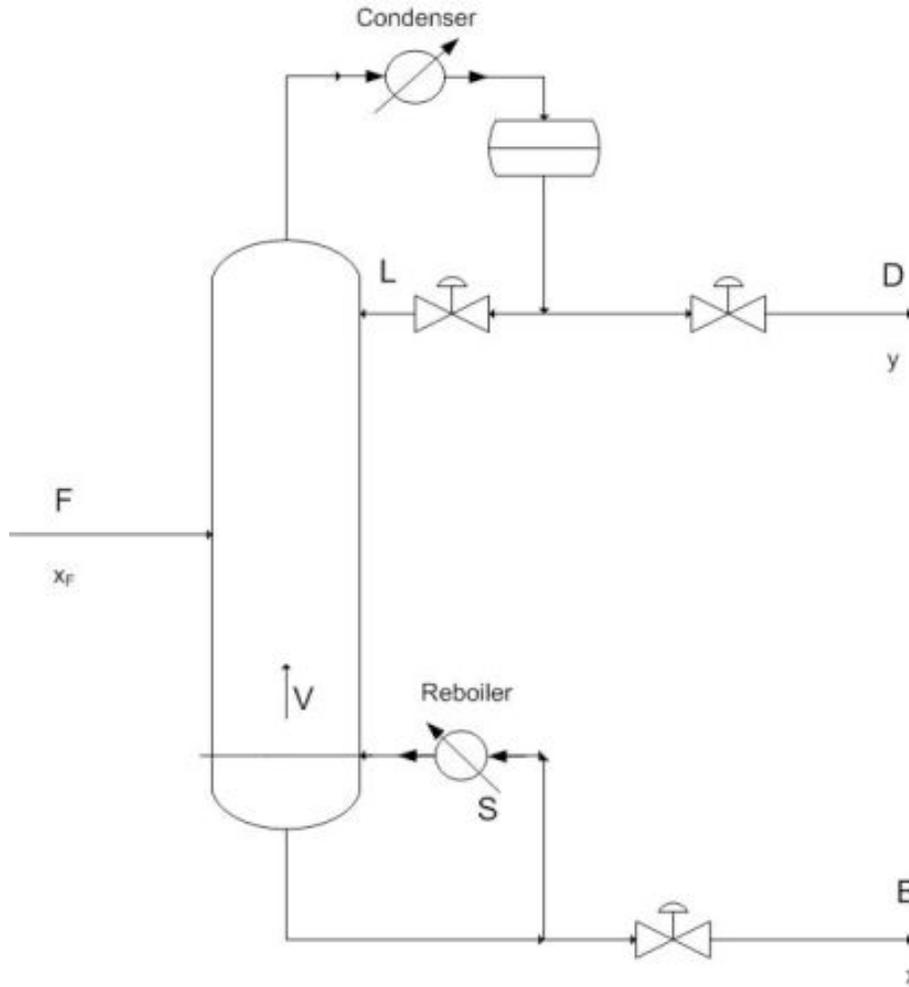
Vapor liquid dynamics within the column also contributes to the theory behind process control because of a few important relations. Changing V (by changing the reboiler energy), causes an extremely rapid response in the overhead composition, while changing the reflux ratio requires a longer response to its effect on the reboiler.

In lower pressure columns, a phenomena known as entrainment or flooding occurs in which liquid is blown up into trays instead of dropping down into trays. This significantly decreases separation efficiency and therefore less product gain occurs. Using a packed column in these low pressure applications provides greater efficiency over tray columns and also allows for the faster accomplishment of a steady state profile. Controlling the occurrence of entrainment in either case is another crucial aspect which should be recognized when designing control systems for columns.

5.2.1 Regulatory Controls

For the distillation process it is imperative that regulatory controls such as level, flow, and pressure controllers are functioning properly to further ensure the effectiveness of the product composition controllers.

In terms of regulatory control, *level controls* are used to maintain specified levels in the reboiler, accumulator, and in the case of a distillation column with two columns due to high tray numbers for a single column, also maintain the level in an intermediate accumulator. Inept use of level controls may lead to problems elsewhere in the distillation process. For example, poor level control on the accumulator and reboiler may lead to problems with composition control for material balance control configurations. Also if the reboiler duty is maintained by one of these level controllers and the controller causes oscillation in the reboiler, consequently cycling may also occur in column pressure.

Flow controllers are used to manipulate and maintain desired flow rates of the reflux, distillate and bottoms products, and the heating medium employed in the reboiler. The setpoints of this specific type of controller is determined by the various composition and level controllers in the process.

Pressure controls are located to the top of the distillation column in the distillate vicinity. Here the column overhead pressure, caused by the accumulation of components in the vapor phase, acts as an integrator causing a change in the level of the accumulator. This pressure may be controlled using a variety of methods.

Pressure Control through Condenser Operation

1) Maximize cooling water flow rate to the condenser (operate at minimum column pressure)

2) Adjust the rate of condensation of the overhead (for example by adjusting the flow rate of refrigerant to the condenser)

3) Adjust level of liquid in the condenser (changes heat transfer area)

Pressure Control through Accumulator Operation

1) Purge vapor from the overhead accumulator

2) Directly changing amount of material phase (for example by pumping in inerts)

5.2.2 Composition and Constraint Control

Because distillation requires a desired product concentration or flowrate, constraint control is used to ensure the desired operation conditions by having setpoints designated for the requirements of the system. The constraint is usually a concentration and the

control of this concentration can vary depending on the application. In almost all industry applications only one product concentration is controlled, while the other is allowed to vary. This is known as Single Composition control and is much easier to achieve and maintain than the Dual composition control which specifies both product concentrations. The advantage to Dual Control however being increased energy efficiency because of increased separation. P&ID of common control placement used for composition control are shown below. Control lines are not included in this diagram because of the various number of control possibilites for a system.
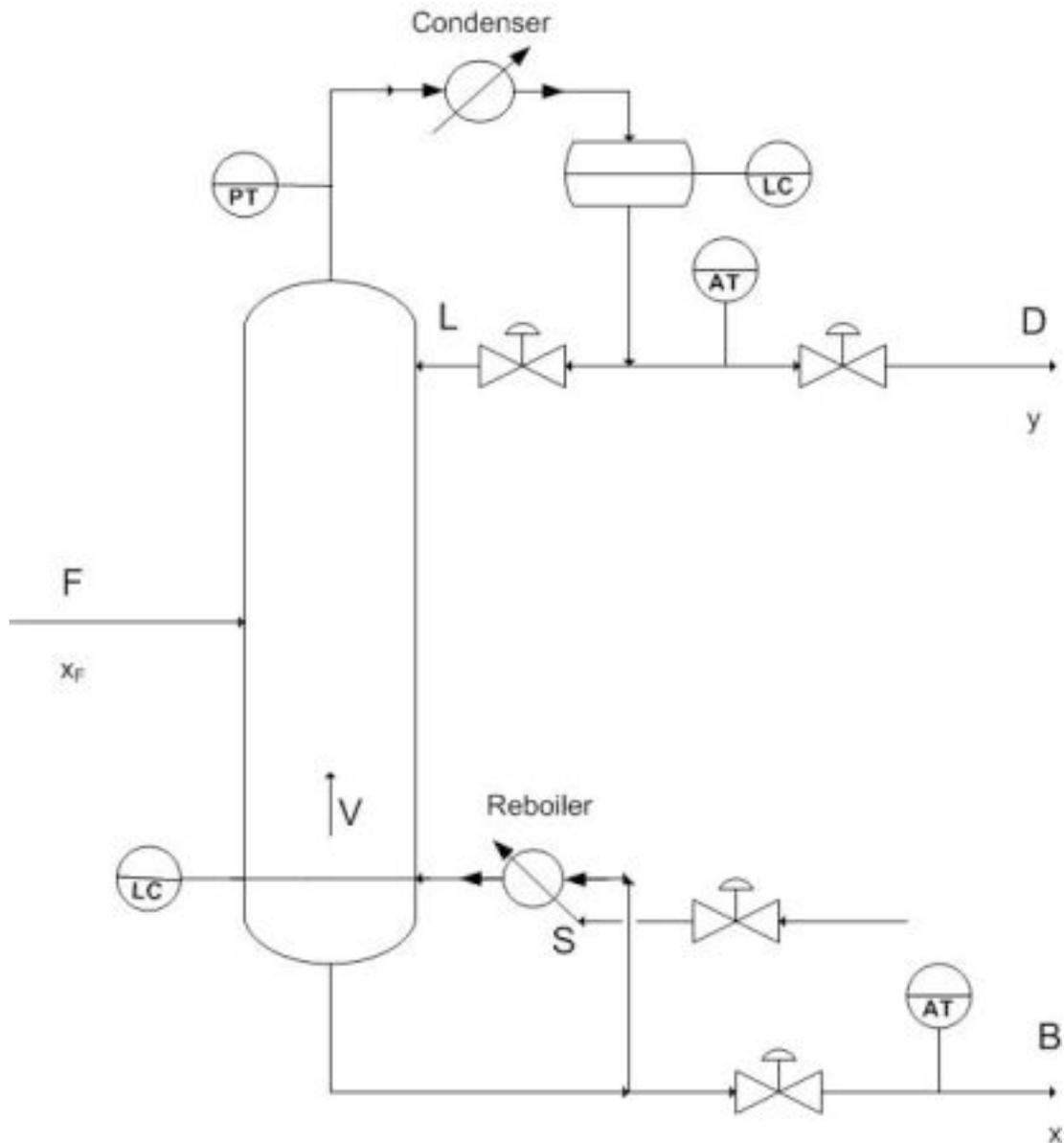
Setting constraints on a column allows for proper control of the product as well as points to issues of safety and maintenance. The most common constraint controls are in the maximum reboiler and condenser duty which results from a number of variables including fouling [[1]], improper valve sizing, and excessive increases in feed. Other

common constraints are flooding and weeping points [[2]] which indicate incorrect L/V ratios and can be compensated for by adjusting the pressure drop across the column.

The most common adjustments for constraint control involve changing the reboiler duty to satisfy the constraints and almost always follow with subsequent fall back mechanisms to ensure product quality.

1) Ensuring single composition control

2) Reducing feed rate

3) Increasing product purity setpoints

5.2.3 Effective Distillation Control

Before doing an in depth analysis of distillation control, it is of paramount importance that the following basics are attended to.

- Firstly, ensure that regulatory controls are indeed functioning congruously.
- For changes in reflux temperature, employ the use of reflux controls.
- Make sure to check and evaluate analyzer deadtime, accuracy, and reliability. This is required to account for the lagtime from the product composition analyzer when using feedback control to control feed flow, reflux ratio and reboiler power. Refer to [[3]] for deadtime. To select proper analyzer refer to [[4]]
- Ensure that any thermistors or RTD's employed to measure tray temperatures for composition inference are fully operational and correctly situated. Care should be taken here to ensure that pressure corrected temperatures are used. Refer to the following for temperature sensors [[5]]
- When streams such as D, B, L, and V are used as manipulated variables for composition control, they should be changed with respect to the measured feed rate when column feed rate changes are a common disturbance.

Summary of Distillation Control Methods

Control Parameter	Example of Control Method
Distillate flow rate (D)	Flow controller (Setpoint controlled by accumulator level)

Bottoms flow rate (W)	Flow controller (Setpoint controlled by level in bottom of column)
Reflux flow rate (L_D)	Flow controller (Setpoint controlled by top tray temperature)
Reboiler Steam flow rate	Flow controller (Setpoint controlled by bottoms composition analyzer)
Feed flow rate	Flow controller (Setpoint manually entered based on unit operations)
Distillate purity (x_D)	Reflux flow controller (Setpoint controlled by top tray temperature)
Bottoms purity (x_W)	Steam flow controller (Setpoint controlled by online analyzer)
Column pressure	Purge flow controller (Setpoint controlled by column pressure)

5.2.4 Control Problems and Disturbances

Feed Composition & Feed Flow Upsets

In order to properly determine the ultimate product purity and flow you must consider the impact of disturbances in the column system. The most significant, however most easily remedial of these disturbances is a feed composition upset, in which there is a change in the feed composition resulting in a major disturbance in the product composition. Hence, configuration of a distillation column must consider the control regulations of such a feed upset. Feed flow upsets are regulated using ratio control of L/F, D/F, V/F and B/F through level and reflux ratio sensors.

Feed and Reflux Enthalpy Disturbances

Feed Enthalpy upsets become an issue for columns operating at low reflux ratios and cause a large deviation from expected product concentrations due to the changes in vapor and liquid flowrates in the column. The usual compensation for this is using a feed heat

exchanger to control the proper enthalpy to the column. In most cases, the feed is preheated before entering the column. By adjusting the duty of the preheater (i.e. decreasing or increasing the heating medium flow rate), a constant vapor/liquid ratio in the feed can be maintained. Rapid changes to the external conditions of a column, (especially large temperature deviations in rainstorms) can cause a subcooled reflux, changing the composition of the products. Reflux control can amend this properly.

Steam and Column Pressure Upsets

The most severe disturbance occurring in a distillation column occurs when there is a loss of steam pressure in the reboiler. A sharp drop in steam pressure results in a drop in reboiler effectiveness and therefore a huge increase in the impurity of the product. This can be avoided using an override control loop for this particular occurrence. Due to the effect of pressure on the relative volatility of the components in the system, a disturbance in column pressure leads to altered product quality. This can be effectively maintained by a composition controller to compensate for these pressure differences.

5.3 ODE Modeling of a Distillation Column

There are two methods used for modeling a distillation column: dynamic modeling and steady state modeling. The key difference between the models is that dynamic modeling is used to monitor changes in the distillation column as a function of time, while the steady state model looks at a given set of conditions at one particular time (i.e. when the column is at steady state).

5.3.1 Dynamic Model

Dynamic modeling of a distillation column may be used for a variety of different reasons: monitoring variations in the column as a result of feed changes, to predict the effects of tray fouling, and to predict when flooding with occur. The dynamic model allows the user to improve distillation control by being able to deal with disturbances that cause upsets to the column's normal operation.

The algorithm for developing a dynamic distillation column model is as follows:
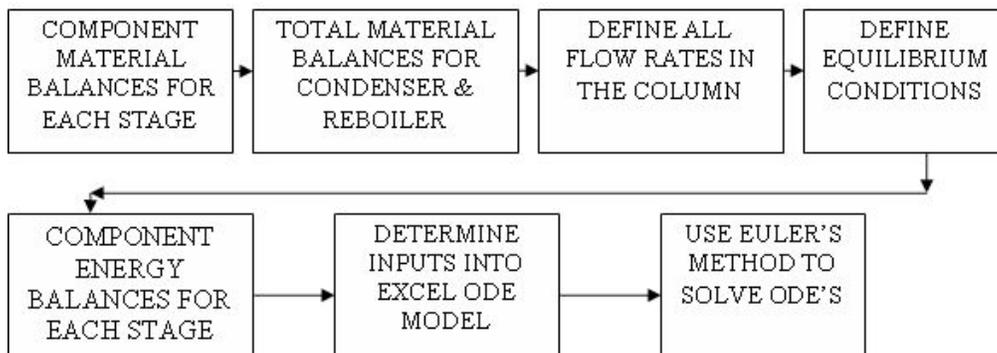
5.3.2 Step 1) Write component material balance for each stage in the column.

$accumulation\ =\ liquid\ entering\ stage\ i + vapor\ entering\ stage\ i - liquid\ leaving\ stage\ i - vapor\ leaving\ stage\ i$

Component material balance for all stages, except the feed tray, overhead condenser, and reboiler:

$$\frac{dM_i x_i}{dt} = L_{i-1}x_{i-1} + V_{i+1}y_{i+1} - L_i x_i - V_i y_i$$

Assumption: For simplicity, accumulation in the each stage is constant; $\quad \dfrac{dM_i}{dt} = 0$

Simplified component material balance for each stage (only composition changes with time):

$$M_i \frac{dx_i}{dt} = L_{i-1}x_{i-1} + V_{i+1}y_{i+1} - L_i x_i - V_i y_i$$

The following are examples of equations used in the Excel Interactive ODE Distillation Column Model, which are provided to help the user understand how the model works.

ODE used to solve for the liquid composition leaving tray 2 (rectifying section):

$$\frac{dx_2}{dt} = \frac{1}{M_2} \left[L_1 x_1 + V_3 y_3 - L_2 x_2 - V_2 y_2 \right]$$

ODE used to solve for the liquid composition leaving tray 5 (stripping section):

$$\frac{dx_5}{dt} = \frac{1}{M_5} \left[L_4 x_4 + V_6 y_6 - L_5 x_5 - V_5 y_5 \right]$$

Overhead condenser component balance:

534

$$\frac{dx_D}{dt} = \frac{1}{M_D} \left[V_1 \left(y_1 - x_D \right) \right]$$

Feed tray component balance:

$$\frac{dx_3}{dt} = \frac{1}{M_3} \left[L_2 x_2 + V_4 y_4 - L_3 x_3 - V_3 y_3 \right]$$

Reboiler component balance:

$$\frac{dx_W}{dt} = \frac{1}{M_W} \left[L_6 x_6 - W x_W - V_7 y_7 \right]$$

5.3.3 Step 2) Write total material balances around condenser and reboiler

Condenser material balance:

Assumption 1: Total condenser (all vapor from the top of the column is condensed into a liquid).

Assumption 2: Overhead accumulator liquid level remains constant.

$$D = \left[V_1 + L_D \right]$$

Reboiler material balance:

$$W = \left[F - D \right]$$

For these equations to work, the user must specify:

-reflux flow rate (mol/min)

-bottoms flow rate (mol/min).

5.3.4 Step 3) Define all flow rates

Vapor Leaving Feed Stage:

$$V_3 = V_4 + F(1 - q_F)$$

Liquid Leaving Feed Stage:

$$L_3 = L_2 + F(q_F)$$

Vapor flow rates in stripping section:

Assumption: Equimolal overflow for vapor in stripping section

$$V_4 = V_5 = V_6 = (V_7)$$

Vapor flow rates in rectifying section:

Assumption: Equimolal overflow for vapor in rectifying section

$$V_1 = V_2 = (V_3)$$

Liquid flow rates in rectifying section:

Assumption: Equimolal overflow for liquid in rectifying section

$$L_2 = L_1 = (L_D)$$

Liquid flow rates in stripping section:

Assumption: Equimolal overflow for liquid in stripping section

$$L_6 = L_5 = L_4 = (L_3)$$

5.3.5 Step 4) Define equilibrium conditions

The binary system used in the Excel ODE model is a benzene-toluene system. The equilibrium data for this system was put in the model and the relative volatilities were calculated for various equilibrium compostions.

Relative Volatility (from equilibrium data):

$$\alpha = \frac{y_{benzene} x_{toluene}}{x_{benzene} y_{toluene}}$$

536

where α is defined as the relative volatility of the two components in the system.

These relative volatilities were plotted against temperature and linear regression was used to fit the data.

Relative volativity as a function of temperature:

$$\alpha = \left[-0.009T + 3.3157\right]$$

This equation models how the separation changes on each tray as a function of tray temperature, which decreases up the column.

Equilibrium Vapor Composition for each stage:

Assumption: Trays in the Column are 100% efficient (vapor and liquid leaving a tray are in equilibrium)

$$y_i = \frac{\alpha x_i}{1 + (\alpha - 1)x_i}$$

Replacing alpha with the temperature dependent equation shows how tray temperature affects the amount of benzene in the vapor leaving each tray.

5.3.6 Step 5) Write component energy balances for each stage.

The ODE energy balances are essential for the dynamic model to run properly. Mass transfer occurs within the column because the temperature varies from the top of the column to the bottom thereby allowing separation of the components in the system.

Since the energy input into the column is added in the reboiler, the reboiler ODE is the first equation entered in the model. In our model, this is given as:

$$\frac{dT_7}{dt} = \frac{1}{M_W}\left[L_6 x_6 - W x_W\right]\left[T_6 - T_7\right] + \frac{q_r}{M_W c_p}$$

The next step is adding energy balances for each subsequent stage in the distillation column. The only stage in the column which has a slightly different energy ODE is the feed stage, given by:

$$\frac{dT_3}{dt} = \frac{1}{M_3}\left[\left[L_2 x_2\right]\left[T_2 - T_3\right] + \left[V_4 y_4\right]\left[T_4 - T_3\right] - \left[L_3 x_3\right]\left[T_2 - T_3\right] + \left[V_3 y_3\right]\left[T_4 - T_3\right] + \left[F x_{feed}\right]\left[T_{feed} - T_3\right]\right]$$

The last energy balance is around the condenser.

Assumption: Reflux return temperature is constant (overhead condenser duty varies to compensate for this).

5.3.7 Step 6) Determine inputs into ODE model

Once all of the equations have been put into the model, all remaining unknown variables must be placed in a section so that the user can specify these input values when running the model.

For the Excel ODE distillation model, the users inputs include:

1) Feed flow rate

2) Mole fraction of light key in the feed

3) Reflux flow rate

4) Condenser, reboiler, and tray levels

5) Phase of the feed (q-value)

6) Feed temperature

7) Integration step size

To model effects of disturbances, the user may also change these input values:

- Feed flow after 200 time steps

- Feed composition after 600 time steps

5.3.8 Step 7) Use Euler's Method to solve the ODE's

This step involves using [Euler's method]. to integrate each ODE over each timestep in the interval to solve for the parameter value at the next time step. Creating a plot of these values versus time allows the user to see how changes in the input values effect parameters such as distillate and bottoms composition or flowrates.

Click the link below to find a summary of the general equations used in dynamic modeling of simple distillation column.

[Media:EquationsforODEModeling.doc]

538

5.3.9 Additional Considerations for Dynamic Distillation Modeling

The Interactive Excel ODE Distillation Column Model does not take into account heat effects within or surrounding the column. It may be beneficial in modeling an actual column to determine the optimal temperatures that feed and reflux should enter at to achieve the greatest possible separation. The energy input into the reboiler is another consideration that may need to be modeled for economic purposes.

Also, to tune the dynamic distillation model with greater precision, additional parameters and equations may be added. One example of this would be to add a K_D or K_U input value to control the levels in the bottom of the column or overhead condenser.

5.3.10 Steady State Model

The steady state model which is described by using a McCabe-Thiele diagram shows the theoretical stages in a binary component distillation column. An example of a McCabe-Thiele diagram is shown below.
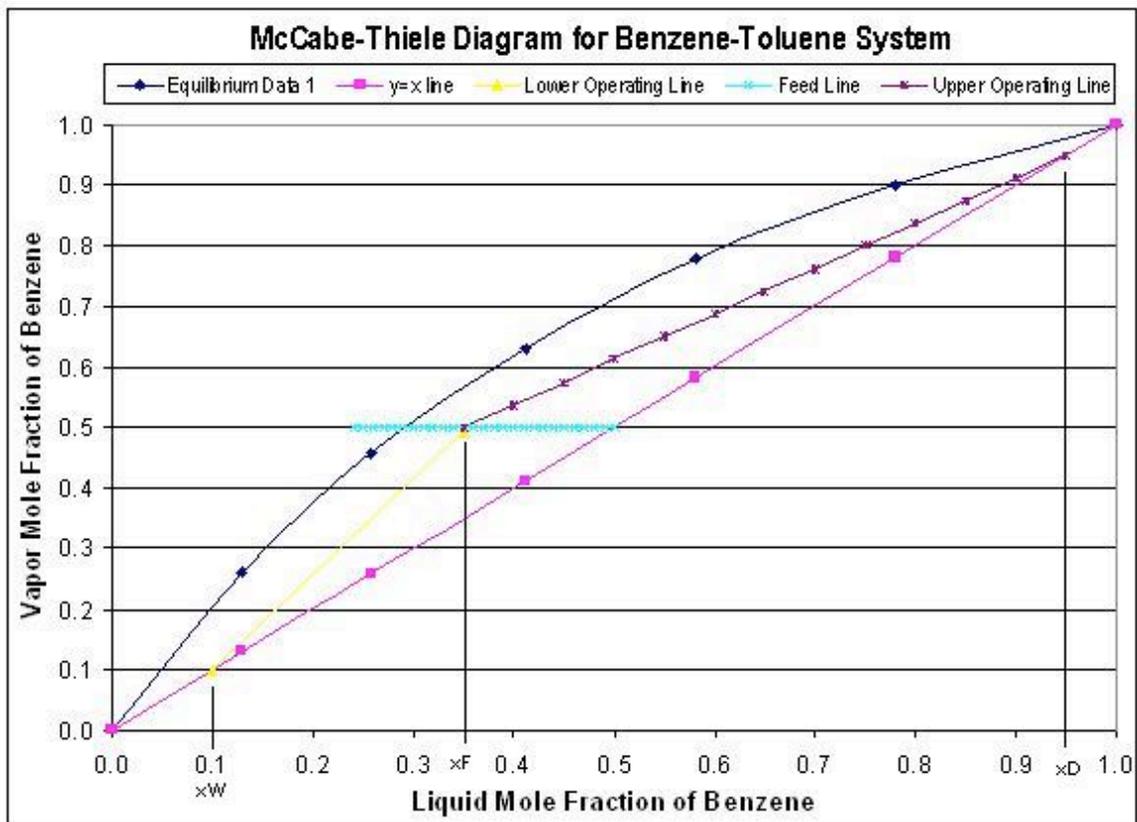
The upper operating line is a graphical representation of the vapor/liquid dynamics in each stage in the rectifying section of the column (above the feed stage), while the lower operating line represents the vapor/liquid dynamics in the stripping section of the column (below the feed stage). The starting point of the upper operating line represents the distillate composition and the bottom point of the lower operating line represents the

539

bottoms compostion. The feed line shows the entering feed composition as well as whether the feed is vapor, liquid, or a combination of the two. To develop this steady state model, one must know the components in the system, so that equilibrium data can be obtained. Also, this model requires that the following parameters must be known:

- reflux ratio
- distillate composition
- bottoms composition
- feed composition
- feed phase

The general equations used in the steady state model are given in the table below:

Media:EquationsforODEModeling.doc

The McCabe-Thiele diagrams are excellent for modeling steady state operation, but they do not describe how disturbances affect the column operation. For each change in a particular parameter, a separate McCabe-Thiele diagram must be made. The dynamic model, although more complex than the steady state model, shows how a column operates during start-ups, when disturbances occurs, and where steady state conditions occur. Therefore, the dynamic ODE model of a distillation column can allow the user to how product purities and flow rates change with time.

5.3.11 Glossary of Terms

M_i = Molar holdup on tray i

L_{i-1} = Liquid molar flowrate into tray i

L_i = Liquid molar flowrate leaving tray i

V_{i+1} = Vapor molar flowrate entering tray i

V_i = Vapor molar flowrate leaving tray i

x_i = mole fraction of light component in the Liquid phase of Tray i

y_i = mole fraction of light component in the Gas phase of Tray i

B = Bottoms flowrate

D = Distillate flowrate

f = Feed flow rate

α = Relative volatility of Benzene-Toluene system.

540

q = Vapor Liquid compostion value

5.4 Our Excel Distillation Model

Below you will find a link for our distillation model in Excel. It is an exceptionally large download so it may take some time if you are not using a high speed internet connection. Because it takes some time for the iterations to run to completion, calculations are not continuous within the spreadsheet. Be patient as it may take a few seconds for the calculations to run when operational parameters change. Also, because of the nature of the iterative process Excel uses to determine the cell values, large deviations may cause the model to crash resulting in #Error# for many cells. If this occurs, close the model and re-open it from the website to try different parameters.

In order to demonstrate dynamic changes occuring during operation, step changes are incorporated into the Excel model. To show the effects of a feed change, input the change into cell C-18. This change will occur at t=2min for a length 2 minutes and then the feed flowrate will revert back to the original value. To demonstrate a change in the feed compostion at t= 6min, input the change into cell C-19 on the Excel model. This compostion change will be constant over the remainder of the time and will not revert to the previous value. Notice the changes occuring in the graphs as these step changes occur. The Excel model can be found here => Media:ODEDistillationModel-Final.zip

This model has a specified control range because typical operation of distillation columns requires operation within a specific range of values where input parameters marginally vary. It is recommended to use the following variables for the initial model.

Feed Flow to Column = 50.00

Feed Composition (xF) = 0.5

Reflux Flow = 15

Bottoms Flow @ T= 0 = 35.00

Mtray = 10

Mcondenser = 50

Mreboiler = 50

qF = 0.4

Feed Temperature = 80

Reflux Temp = 80

Initial Column Temp = 80

Qreboiler = 100

Avoid extreme changes in the initial values, otherwise a significant error may occur causing the model to crash. The principal cause of this is due to the heat effects which are taken into account in the model and the fact that the number of trays on the column is set at 6.

5.5 Terms Commonly Used in Distillation

Active tray area: The region of the tray where the upward moving vapor comes into contact with the downward flowing liquid (where the mass transfer occurs)

Downcomer: Area on side of trays where liquid flows down through the distillation column

Disturbance (wrt distillation): Any minor change in the distillation column caused by an external or internal source that causes product variability

Flooding: Liquid from the active tray area is carried up into the vapor stream (occurs at low L/V ratios)

Ratio Control: Controlled ratio of two manipulated variables

Tray Fouling: Active area of tray is deminished, thus reducing separation efficiency within the column

Upset: Any major change in the distillation caused by an external source that produces erratic column operation that requires manual override to gain control

Weeping Points: Liquid from the active tray area seeps downward through the tray instead of flowing through the downcomer (occurs at high L/V ratios)

5.6 Worked out Example 1

5.6.1 Problem Statement

Using the Excel Model of a Distillation Column posted above, see how a change in the feed concentration upsets the steadystate equilibration of the inital flow and how the column thereby adjusts to a change in feed composition at the time step indicated. Then determine the effect of changing the feed flowrate and reflux flow at the indicated times during the time the column is running. Also, determine the effect of making a larger step size for use in Eulers method.

Remember: an excessive change may cause the distillation column to blow up, so take care in determining reasonable changes. Input the *changes* in the flowrate or concentration into C17-C20 of the spreadsheet. Input the new step size in cell D21.

5.6.2 Solutions

The normal concentration output for the distillation column is shown below. Feed flow rate is 50 mol/min, reflux flowrate is 15 mol/min, feed composition is 0.5 mol fraction of the light component. Step size is 0.01 minutes.
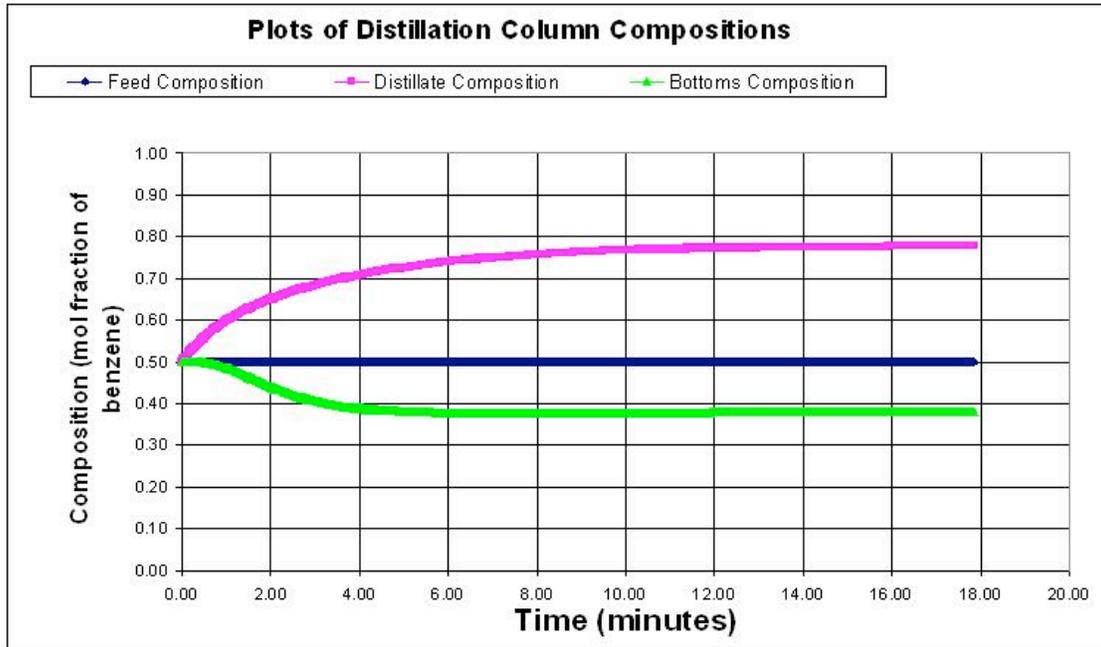
Students should see that an impulse change in flowrates does not really change the steady state profile of the column. However, a change in feed concentration will cause the column to jump and then equilibrate for the new concentration values. Below is a graph with a change in the feed flowrate to 100 mol/min, and a concentration change to 0.6 of the light component.

Plots of Distillation Column Compositions
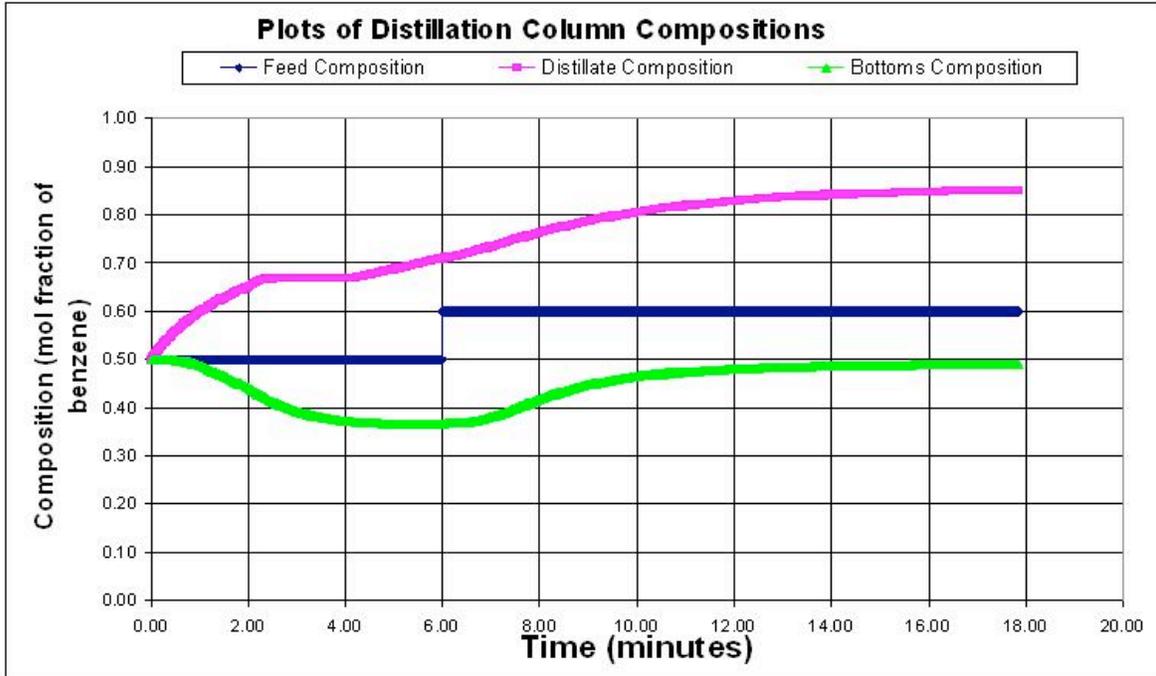
Larger step sizes will cause the model to crash because the error associated with the derivatives becomes larger with each derivative taken. Below the step size is 0.05 minutes. At some point the derivatives become uselsess data and therefore the distillation column model blows up with outrageous values as shown below.

Plots of Distillation Column Compositions

544
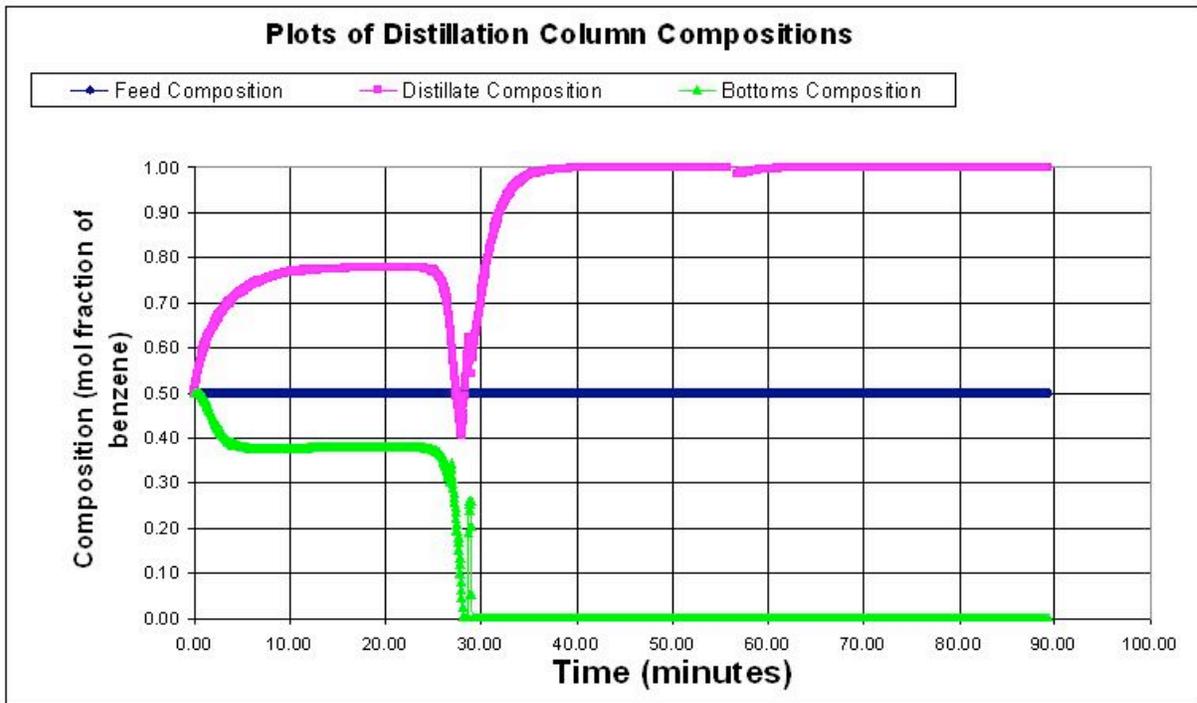
5.7 Worked out Example 2

5.7.1 Qualitative Example

As a new hire at a specialty chemical company you are placed on project where you are asked to design and specify the controls on a state of the art distillation column used to separate a binary system. Your supervisor gives you a few clues to the control mechanism based on the chemistry occurring.

- The salable component is the heavy component and the mixture has a low relative volatility.
- The light component is not usable at all, however it is extremely toxic and must be controlled precisely.
- This is a small operation which requires a low pressure separation.
- The facility has had issues with the onsite steam that is typically used to heat the reboiler.

Draw the P&ID model of how you would model this column, mention all the items you must consider for the process and what you would do to control them.

5.7.2 Answer:

Since the heavy component control is desired you will want a feedback loop based on the bottoms composition controlling reboiler duty. Also note how precise control of the vapor flowrate up the column can be best accomplished through reboiler control. Because it is low pressure you must begin to worry about entrainment and the possibility of flooding the column. Therefore there should be pressure sensors controlling the pressure in the column to maintain it at optimal levels using a packed column instead of a tray column to produce a better separation of the components. Because onsite steam has been shown to be a problem you would probably want to have a redundancy built into the reboiler steam or perhaps have its own autonomic steam delivery because this could potentially jeopardize the quality of your product.
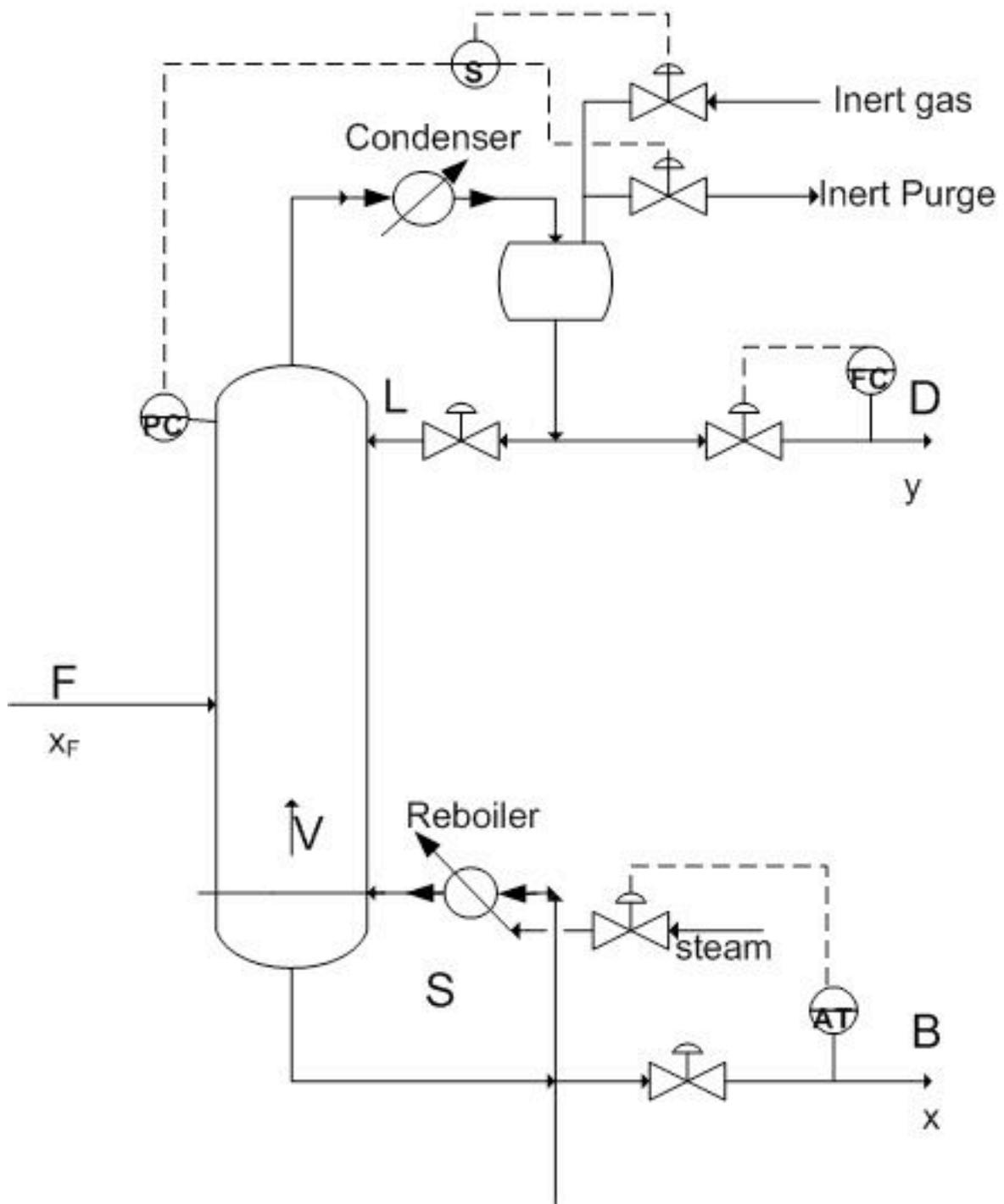
5.8 References

- Bequette, B. Wayne. Process Dynamics Modeling, Analysis, and Simulation, New Jersey: Prentice Hall
- Perry, Robert H. Perry's Chemical Engineers' Handbook, 7th Ed. McGraw Hill. ISBN 0-07-049841-5

- Riggs, James B.; Karim, M. Nazmul. Chemical and Bio-Process Control, 3rd Ed., Texas Tech University, Ferret Publishing

Section 6. ODE & Excel model of a heat exchanger

Title: ODE Modeling in Excel: Heat Exchanger
Authors: Teresa Misiti, Carly Ehrenberger, Danesh Deonarain, Georgina Mang
Date Presented: September 26,2006 Date Revised:

- First round reviews for this page
- Rebuttal for this page

6.1 Introduction

In process industries, heat exchangers are designed to transfer heat from one fluid to another. Heat exchangers have many different applications, especially in chemical processes, air conditioning, and refrigeration. They are classified by their design and type of flow. In some types of heat exchangers, the two fluids are separated by a wall or membrane and the heat transfer takes place by both convection and conduction. In another, less common type of exchanger, the two fluids physically come into contact with each other as the heat transfer occurs.

Since heat exchangers have a wide variety of applications and are commonly used in industry, control of the system is essential. A dynamic model may be created to allow the chemical engineer to optimize and control the heat exchanger. By utilizing this model, predictions can be made about how altering the independent variables of the system will change the outputs. There are many independent variables and considerations to account for in the model. If done so correctly, accurate predictions can be made about the system.

## *6.2 Types of Heat Exchangers*

1. Double - Pipe Heat Exchanger

A double-pipe heat exchanger is the simplest type of heat exchanger and can operate with co-current (Figure 1) or counter-current (Figure 2) flow. The design consists of a single small pipe (tube-side) inside of a larger one (shell-side). A co-current heat exchanger is most commonly used when you want the exiting streams to leave the exchanger at the same temperature. A counter-current heat exchanger is used more often than co-current because they allow for a more efficient transfer of energy.
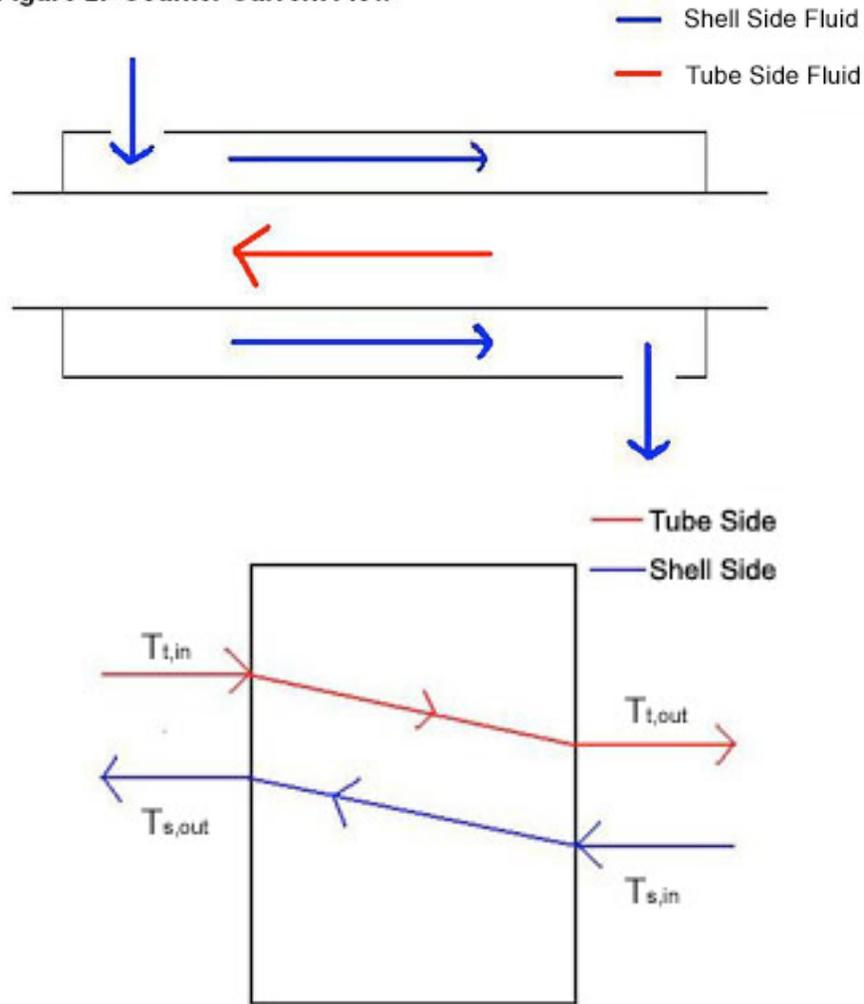
Figure 1: Co-Current Flow
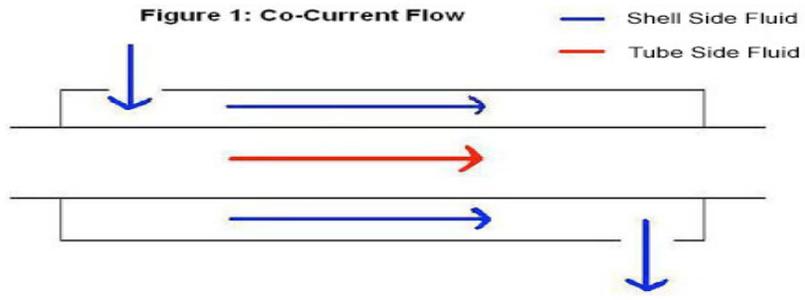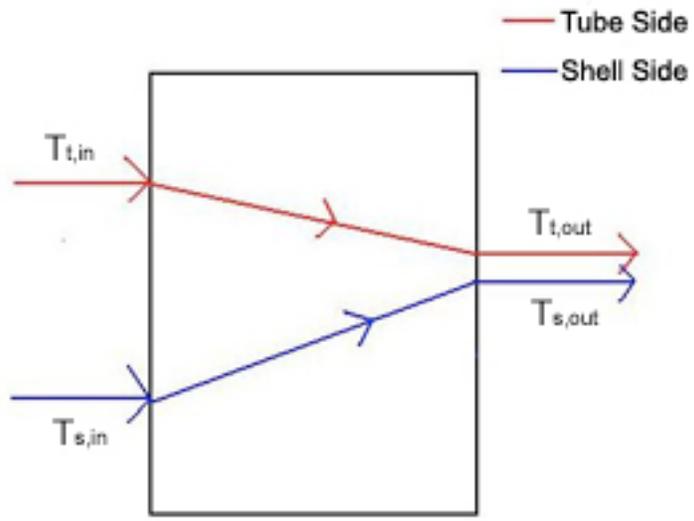
Figure 2: Counter Current Flow

2. Shell-and-Tube Exchanger

A shell-and-tube exchanger is used for larger flows, which are very common in chemical process industries. The design of this exchanger is a shell with a bundle of tubes inside. The tubes are in parallel and a fluid flows around them in the shell. There are many different arrangements such as straight (Figure3) or u-tube (Figure 4). Each arrangement allows for a different type of flow such as co-current, counter-current and cross flow. The tube-side can have one or more passes to increase the energy exchange from the tube-side fluid. The shell-side may contain baffles, or walls, that channel the fluid flow and induce turbulence, and thus, increase energy exchange. Correlations can be developed to predict the increase in energy exchange. These correlations are out of the scope of this article.
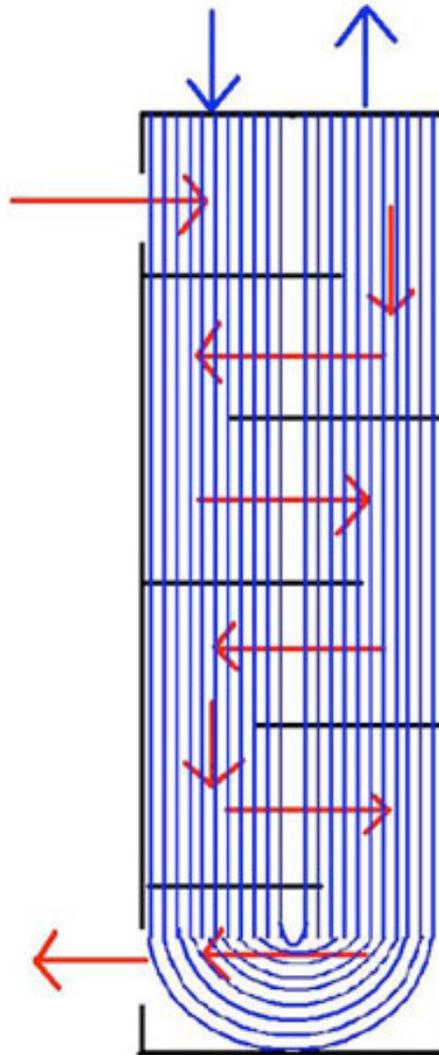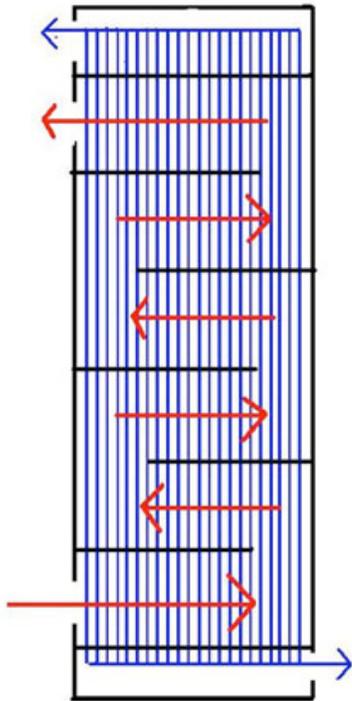
550

Figure 4: U-Tube Heat Exchanger

Figure 3: Straight Tube Heat Exchanger

3. Cross-flow Exchanger

The most common application for a cross-flow heat exchanger is when a gas is being heated or cooled. This device consists of multiple tubes in parallel, usually containing a liquid, and the exterior gas flows across the tubes. In some cases the air is confined in separate flow chambers, such as fins, and in others it is open to flow freely (Figure 5).

Figure 5: Cross-Flow Heat Exchanger (free flow)
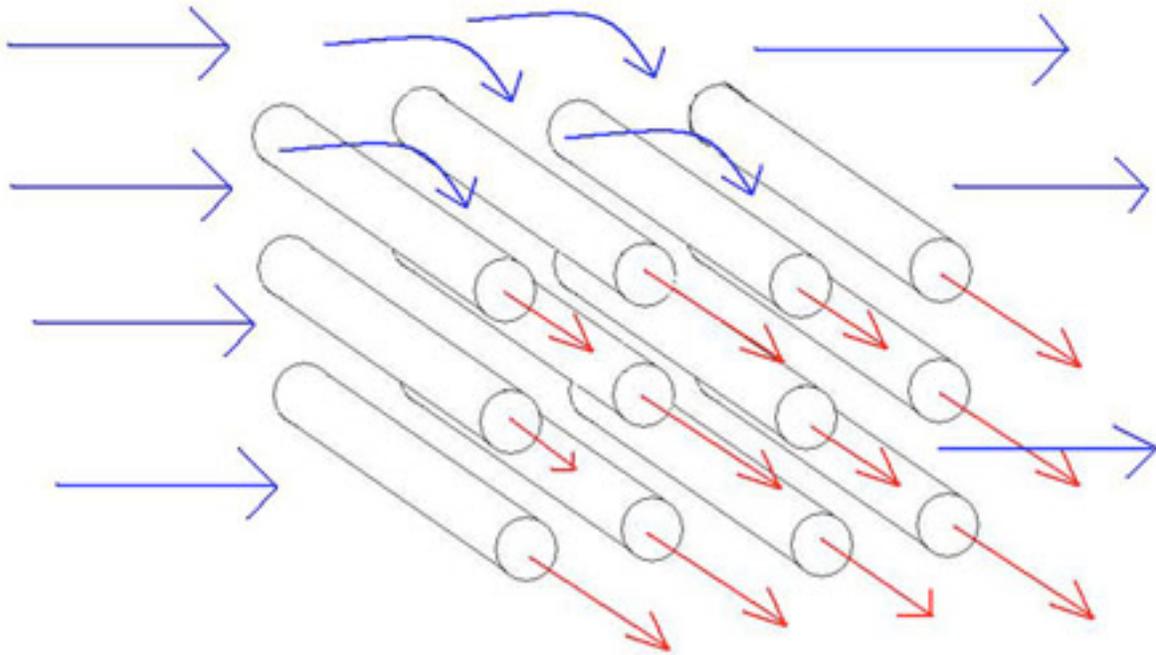
6.3 Dynamic Modeling of Heat Exchangers

Since heat exchangers are so widely used in industry, it necessary for a chemical engineer to be able to optimize and control the system and know how independent variables will affect the outputs from teh system. To do this, a dynamic model is developed and utilized.

A dynamic model of a heat exchanger may be used, for example, to predict how a change in the fluid flowrates or the addition of an insulating jacket will affect the outlet temperature of the product stream. The model uses ordinary differential equations (ODEs) to describe the process and, using a program like Microsoft EXCEL, gives plots of the variables vs. time for the entire process. There are many independent variables in a heat exchanger, which can cause modeling to be very complex since multiple ODEs are required to define all of the process variables.

Some of the independent variables in a heat exchanger system include:

- Shell-side:
    - fluid
    - flowrate
    - temperature

552

- o number of baffles
- Tube-side:
  - o fluid
  - o flowrate
  - o temperature
  - o number of passes
- Flow configuration
  - o Co-current
  - o Counter-current
  - o Cross flow
- Insulating jacket

The primary dependent variable of concern is the outlet temperature of the product stream (usually the tube-side fluid.) The outlet temperature, in a controlled system, is monitored by a sensor. The sensor then transmits a signal to an actuating device of one or more of the independent variables (usually shell-side flow controller) to perform some desired response.

6.4 Modeling with ODEs

The following section outlines the method for developing a dynamic model of a heat exchanger. The model is for a double-pipe heat exchanger that has the ability to flow in co-current or counter-current configurations and the option of an external insulating jacket. It assumes constant fluid properties and perfect heat transfer through the metal of the tubing. The outlet temperature of the tube-side fluid is monitored by a temperature sensor, and the flow rate of the shell-side fluid is controlled by a actuated flow-controlling device.

6.4.1 Energy Balance

An energy balance is first performed on the tube-side fluid.

$$
\begin{array}{c}
\text{Rate of} \\
\text{accumulation} \\
\text{of thermal} \\
\text{energy in} \\
\text{tube-side} \\
\text{fluid}
\end{array}
=
\begin{array}{c}
\text{Rate of} \\
\text{energy} \\
\text{in}
\end{array}
-
\begin{array}{c}
\text{Rate of} \\
\text{energy} \\
\text{out}
\end{array}
-
\begin{array}{c}
\text{Heat} \\
\text{transferred} \\
\text{from} \\
\text{shell-side}
\end{array}
$$

(Eq 1)

The left-most term in the above energy balance is the amount of thermal energy that accumulates in the tube-side fluid and causes a change in its outlet temperature. The terms on the right side of the above energy balance describe the thermal energy of the fluid flowing in and the fluid flowing out and the amount of heat transfer from the shell side fluid. In the term for heat transfer from the shell-side, the temperatures are the temperatures of the outlet streams. The outlet temperatures will change according to whether you are running co-currently or counter-currently. The energy balance is written as:

$$
mc_{p,t}\frac{dT_{t,out}}{dt} = \rho c_{p,t}F_{t,in}T_{t,in} - \rho c_{p,t}F_{t,out}T_{t,out} - \frac{kA_i}{\Delta z}(T_{t,out} - T_{s,out})
$$

(Eq 2)

where,

$m =$ mass of the fluid = $V\rho = \rho A_{crosssectional}\Delta z$

$c_p =$ constant pressure heat capacity of the fluid

$T =$ Temperature

$t =$ time

$k =$ conductive heat transfer coefficent

$A =$ surface area of tube that fluid contacts

$\Delta z =$ length of tube

$\rho =$ density of the fluid

$F =$ volumetric flowrate of the fluid

and subscripts denote

t - tube-side fluid

out - outlet

in - inlet

$\dot{i}$ - inside

An similar energy balance is next performed on the shell-side fluid.

$$\begin{array}{c}\text{Rate of}\\\text{accumulation}\\\text{of thermal}\\\text{energy in}\\\text{shell-side}\\\text{fluid}\end{array} = \begin{array}{c}\text{Rate of}\\\text{energy}\\\text{in}\end{array} - \begin{array}{c}\text{Rate of}\\\text{energy}\\\text{out}\end{array} - \begin{array}{c}\text{Rate of}\\\text{heat}\\\text{transferred}\\\text{to tube-}\\\text{side fluid}\end{array} - \begin{array}{c}\text{Rate of heat}\\\text{loss to the}\\\text{surroundings}\end{array}$$

(Eq 3)

The left-most term in the above energy balance is the amount of thermal energy that accumulates in the shell-side fluid and causes a change in its outlet temperature. The terms on the right side of the above energy balance describe the thermal energy of the fluid flowing in and the fluid flowing out, heat transfer to the tube-side fluid, and also the heat lost by convection to the surroundings. The energy is written as;

$$mc_{p,s}\frac{dT_{s,out}}{dt} = \rho c_{p,s}F_{s,in}T_{s,in} - \rho c_{p,s}F_{s,out}T_{s,out} - \frac{kA_o}{\Delta z}(T_{s,out}-T_{t,out})-hA_{o'}(T_s-T_\infty)$$

(Eq 4)

where,

h = coefficient of convective heat transfer for air

k - conductive heat transfer coefficient.

and subscripts denote

s - shell-side fluid

∞ - air

o - outside of tube

o' - outside of shell

555

6.4.2 Considerations

There are considerations and simplifications you can make in order to solve the differential energy balances. The validity of these assumptions depends on how accurate of a model you require.

1. The heat capacity of the fluid may be temperature dependent. If this is the case, a polynomial equation could be written for the C_p values of each of the fluids. It would take the form-

$$C_p = a + bT + cT^2 + dT^3 \quad \text{(Eq 5)}$$

Values of a, b, c, and d are properties of the fluid and can be found in *Perry's Chemical Engineers' Handbook*.

It should also be noted that if the fluids in the process are gases, their fluid properties will be affected by changes in pressure and the use of a constant C_p value would be inappropriate.

2. The density of the fluid may be temperature dependent. This would be likely if the fluid were a vapor, as in the case of using steam as the shell-side fluid to heat the tube-side process fluid. If this is the case, a differential equation could be written for the ρ value of the fluid and would take the form-

$$\frac{d\rho}{dT} \propto C \quad \text{(Eq 6)}$$

where;

C - is the coefficient of cubic expansion (relates kinetic energy to temperature).

3. Heat loss to the surroundings may be neglected. This would be the case if the heat exchanger is well insulated or if the shell-side fluid is about the same as ambient temperature. In this case-

$$hA(T_s - T_\infty) = 0 \quad \text{(Eq 7)}$$

4. The temperature sensor may have an inherent time delay. This means the

556

temperature output from the sensor lags the actual temperature at the time of the reading. The temperature lag may be accounted for with the differential equation-

$$\frac{dT_{outlet,sensor}}{dt} = \frac{1}{\tau_{Ts}}(T_{outlet,actual} - T_{outlet,sensor})$$

(Eq 8)

where,

τ_{Ts} = time constant for the temperature sensor

τ_{Ts} is a process parameter and usually ranges from 6 to 20 seconds depending on the properties of the sensor. This value would either be given in literature from the manufacturer or would have to be determined through experimentation.

5. The actuator system for the control valve may have a slow dynamic response.
When the actuator system is based on a control valve, the response to the pressure change is slower than the flow through the valve. The flow rate through a control valve can be modeled with the differential equation-

$$\frac{dQ_{actual}}{dt} = \frac{1}{\tau_v}(Q_{setpoint} - Q_{actual})$$

(Eq 9)

where,

τ_v = time constant for the flow control valve

τ_v is a process parameter and usually ranges from 0.5 to 2 seconds depending on the properties of the sensor. This value would have to be determined through experimentation.

6.5 Using EXCEL to Solve ODEs

Since temperature of the heat exchanger varies across the metal as well as along the length of the pipe, a partial derivative for how temperature varies with length must be described. In order to solve this problem using excel, an approximation of length was used. Instead of taking the partial derivative across the length with respect to temperature, the pipe was divided into differential segments, Δz. Ideally Δz is an infinitesimally small cross section of the length of the heat exchanger. We will assume that through this differential segment, the temperature of the liquid that leaves the segment is the same as the temperature of the liquid within the segment. Since we are assuming the same

temperature for the exiting streams as the inside of the segment, the choice of length for these Δz's helps dictate the accuracy of the solution. The combination of these differential units allows us to model a heat exchanger without the use of partial derivatives. The figure below shows an example of the simplification where the heat exchanger is split into three segments (Figure 6).

Figure 6: Heat Exchanger Simplification
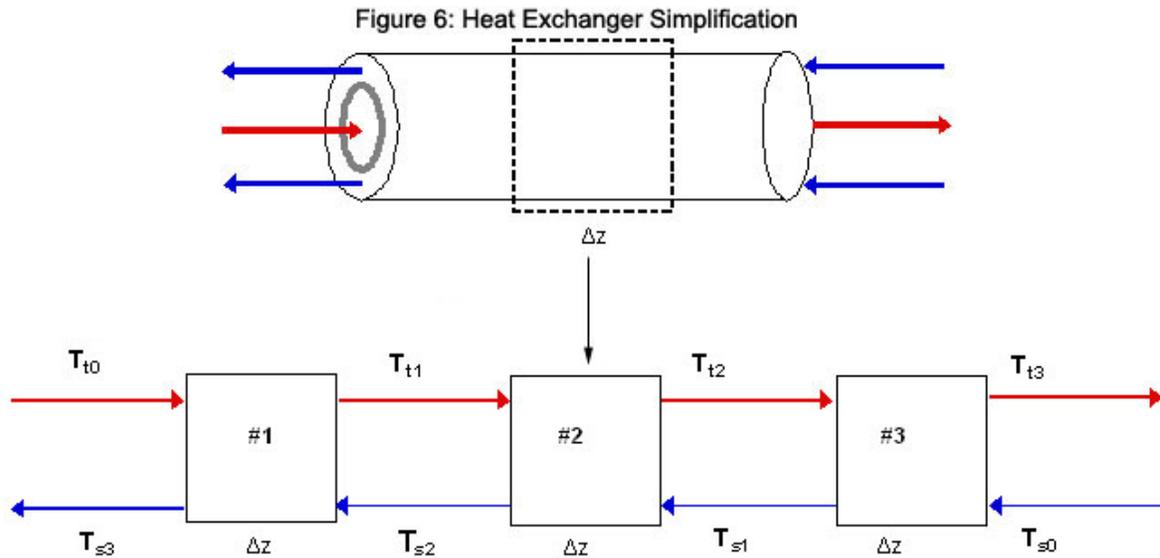
To begin modeling, start by solving the appropriate energy balance (either Eq 2 or Eq 4)

for $\dfrac{dT}{dt}$ and make any simplifications necessary. Then use a method to solve ODEs in Excel, such as Euler's or Heun's. Go here: ExcelModelingODE for further explanation on how to set up an Excel spreadsheet using these methods. For each unit, Δz, and each fluid in that unit of the heat exhanger, you should have an approximation equation. Because ideally Δz represents an infinitesimal section, it follows from the first assumption that we can take the temperature in Δz as being the exit temperature of the hot and cold streams for the Δz's respectively.

The equation for Euler's Method takes the form:

$$T_{h1}\left(t_{i+h}\right) = T_{h1}\left(t_i\right) + \Delta t\left[\frac{dT}{dt}\left(t_i\right)\right]$$

(Eq 10)

558

The equation for Heun's Method takes the form:

$$T_{h_1}(t_{i+h}) = T_{h_1}(t_i) + \Delta t\left[\frac{1}{2}\frac{dT}{dt}(t_i, T_i) + \frac{1}{2}\frac{dT}{dt}(t_{i+h}, T_{i+hk_1})\right]$$

(Eq 11)

The temperatures can then be plotted versus time in order to model how the system inputs affect heat exchange. The following examples walk you through a model of a simple heat exchanger in Excel.

6.6 Worked out Example 1

As a process engineer in a chemical processing company, you were assigned to scale up a process for your company's latest and greatest chemical. Between one of the steps of the reaction, it is necessary to quickly and efficiently cool down the solution from 330K before allowing it to enter the next step. From your previous knowledge, you know that a heat exchanger is necessary and the counter current heat exchanger is the most efficient to achieve a desired temperature. You decide to use cold water entering in at 250K as your coolant since it is relatively inexpensive. In order to allow for the plant engineers in the future to have more control, create a model to determine when this process will go to steady state and graph the temperature profiles of the outlet streams. You can assume the heat exhanger is perfectly insulated and that there is perfect conduction through the metal from shell-side fluid to the tube-side fluid, and this conduction is described by the heat transfer coefficient, k, between the two fluids.

General Relevant Information:
T_{air}=296.15 K (Temp. of the Air)
r_i=0.1 m (Inner Diameter of Inner Tube)
r_o=0.12 m (Outer Diameter of Inner Tube)
r_o'=0.15 m (Outer Diameter of Outer Tube)
Δz=1 m (Increment of length)
k=450000 W/m2*K (heat transfer coefficient between shell and tube-side fluids)

Shell Side Fluid Information (Cold Water)
C_{ps}=4185 J/kg*K (shell side fluid heat capacity)
T_{0s}=250 K (input temp. of shell side fluid)
A_s=0.02543 m (cross sectional area where shell side fluid is present)
ϱ_s=1000 kg/m^3 (density of shell side fluid)
F_s=0.1 m3/s (volumetric flow rate of shell side fluid)

559

Tube Side Fluid Information (Solution)
C_{pt}=1200 J/kg*K (tube side fluid heat capacity)
T_{0t}=330 K (input temp. of tube side fluid)
A_t=0.0314 m (cross sectional area where tube side fluid is present)
ϱ_t=1030 kg/m^3 (density of tube side fluid)
F_t=0.2 m3/s (volumetric flow rate of tube side fluid)

---Solution---

The general Algorithm can be followed to develop the Excel model.

- 1. Make a spreadsheet that has all relevant process information (ie: T inputs, ρ values, etc).
- 2. Perform an overall energy balance for the tube-side and shell-side fluids.

$$\frac{dT}{dt}$$

  Make simplifications and state assumptions. Solve for $\frac{dT}{dt}$
- 3. Divide the heat exchanger into smaller segments of length Δz
- 4. In your spreadsheet, set up a column for time. In the first row, start at t=0 and then increase time by Δh down the column.
- 5. Set up a column for the incoming tube-side temperature, T_{t0}. The value of incoming temperature is inputted by the user. Do the same for the shell-side temperature, T_{s0}.
- 6. Set up a column for the exiting tube-side temperature from each Δz unit. Do the same for the shell-side temperatures. For each of these temperature columns:
  - o a. In the first row, make an initial guess of the temperature. This is usually done by setting the temperature equal to the temperature of the stream coming into the heat exchanger.
  - o b. In subsequent rows, input the Euler's equation (Eq 10) by referencing the appropriate cells for the process information and take all temperatures from time, t_i, which are the temperatures in the row above.
- 7. Make a graph of variables of interest (Temperatures) vs. time.

The graph below illustrates how the exiting hot and cold temperature changes over a short period of time.

560

Outlet Temperatures
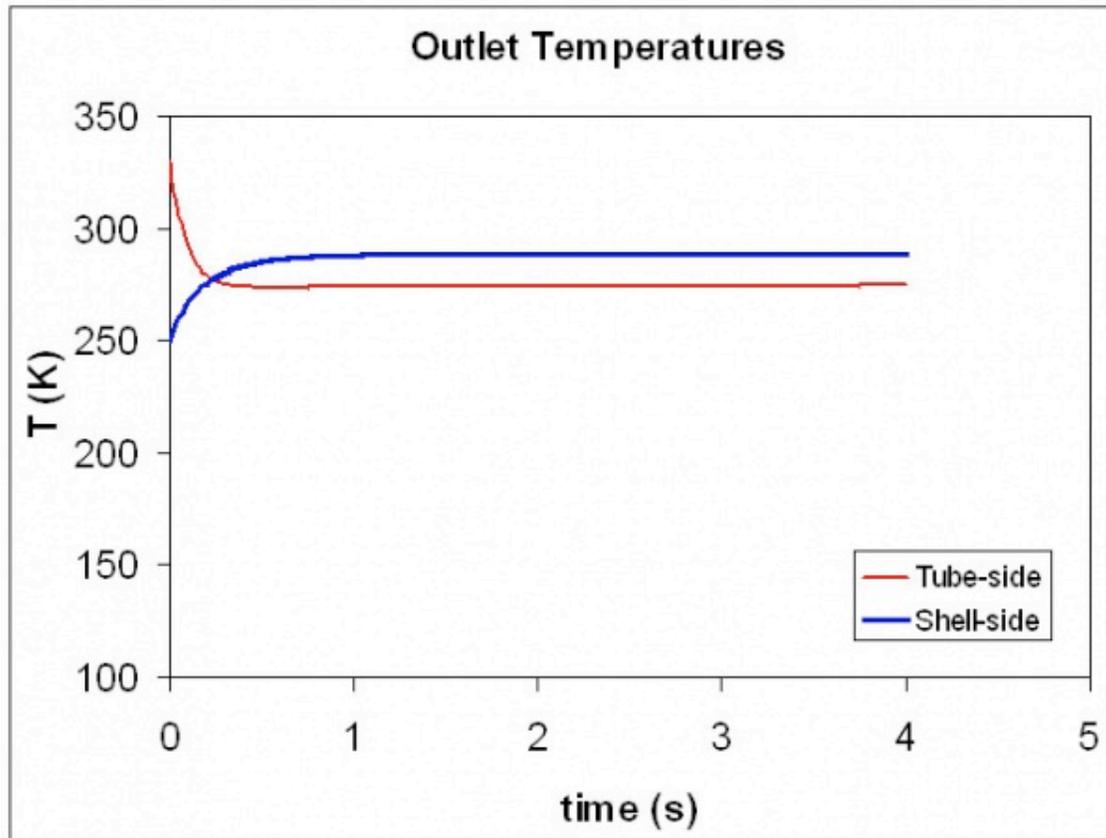
For a more detailed view of this process please see the attached spreadsheet: Euler's Method
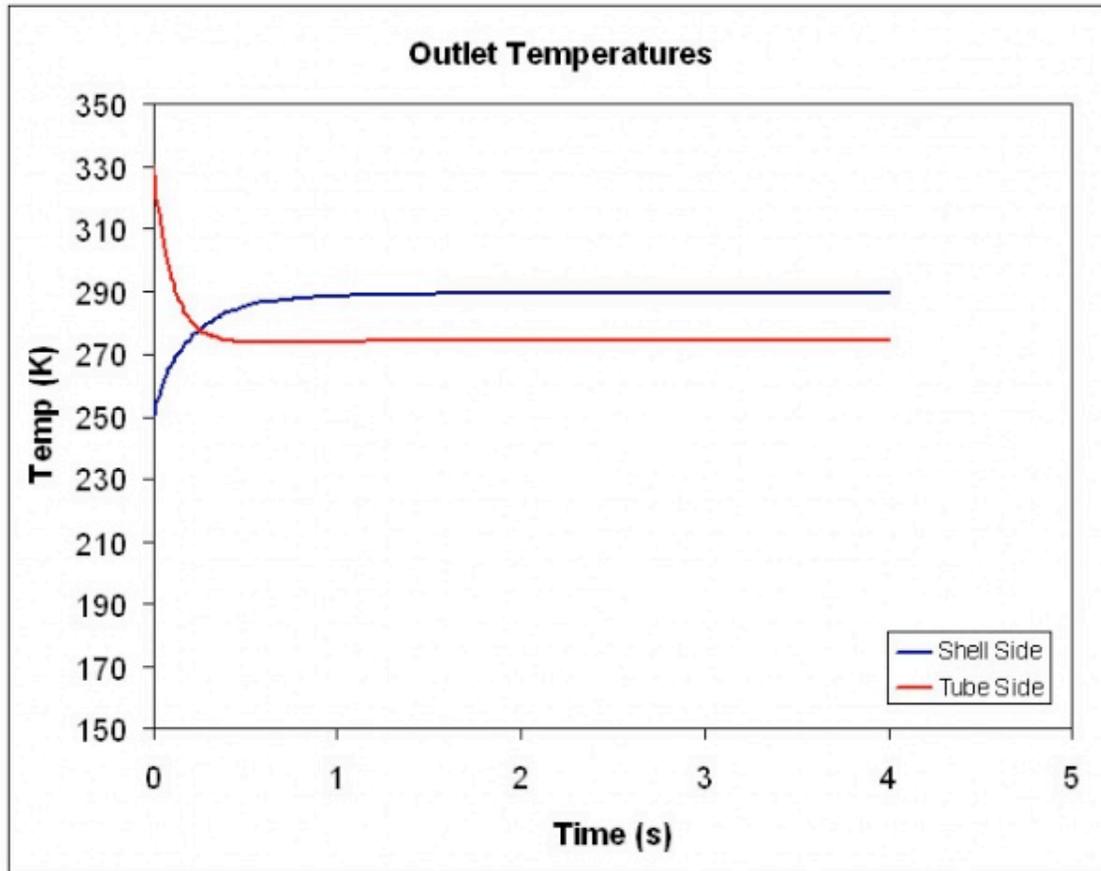
6.7 Worked out Example 2

Again, you are working on the same process, but your supervisor requires you to have a more accurate description of the dynamics of the system. Being as bright as you are, you decide to utilize the 2nd order Runga Kutta method, known as the Heun's method.

---Solution---

Follow the Algorithm in Example 1, except in step 6b, use the equation for Heun's method, Eq 11.

The graph below illustrates how the exiting hot and cold temperature changes over a short period of time.

561

Outlet Temperatures

For a more detailed view of this process please see the attached spreadsheet:

Heun's Method

6.8 Worked out Example 3

A month after you developed a model of the heat exchanger used in the process for your company's latest and greatest chemical, one of the plant engineers calls you and says that at t=2s into the process, the temperature of the incoming fluid changes abruptly from 330K to 300K. How do the temperatures of the outlet streams react to this change. Use the information and model from Example 1.

---Solution---

The Excel spreadsheet can be modified so that at time t=2s, the incoming temperature of the tube-side fluid is 300K. The graph below illustrates how the exiting hot and cold temperature changes over a short period of time.

Outlet Temperatures
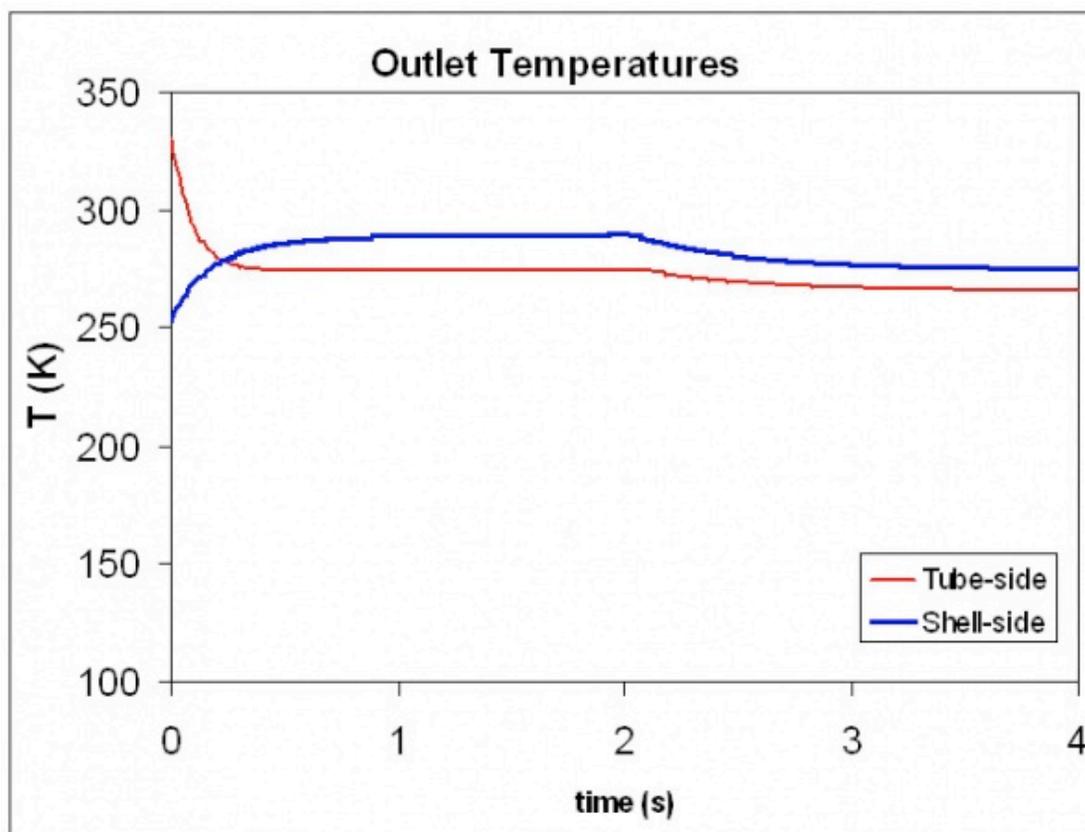
For a more detailed view of this process please see the attached spreadsheet:

Tube-side Temp Change

6.9 References

- Riggs, James B.; Karim, M. Nazmul. Chemical and Bio-Process Control, Texas Tech University, Ferret Publishing.

Section 7. ODE & Excelt model of an adiabatic PFR

Title: ODE and Excel Model of an Adiabatic PFR
Authors: GROUP B => Brian McQuillan, Crystal Miranda, Brandon Quigley, and John Zhang
Date Presented: /Date Revised:

- First round reviews for this page
- Rebuttal for this page

7.1 Introduction

A plug flow reactor (PFR) is a tubular reactor used in chemical reactions. Reactants enter one end of the PFR, while products exit from the other end of the tube. A PFR is useful because of high volumetric conversion and good heat transfer. PFRs carry out power law reactions faster than CSTRs and generally require less volume. Excel modeling for an adiabatic plug flow reactor is useful for estimating conversion as a function of volume. By inputting the values of the constants and iteratively changing the volume, the conversion at a specific volume can be determined. This removes the need for complicated hand calculations and helps the user visualize the reaction and notice trends through the PFR. Alternatively, excel modeling can also determine the temperature inside the reactor as a function of volume. Euler's integration is used to estimate the conversion or temperature through the PFR. By ensuring that the volume intervals are small, the conversion or temperature values generated should be fairly accurate. Refer to the Euler's method section for more information on its implementation.
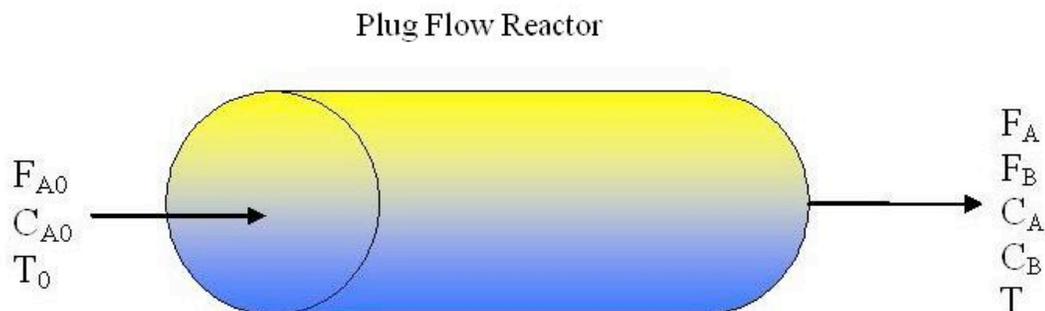
Plug Flow Reactor
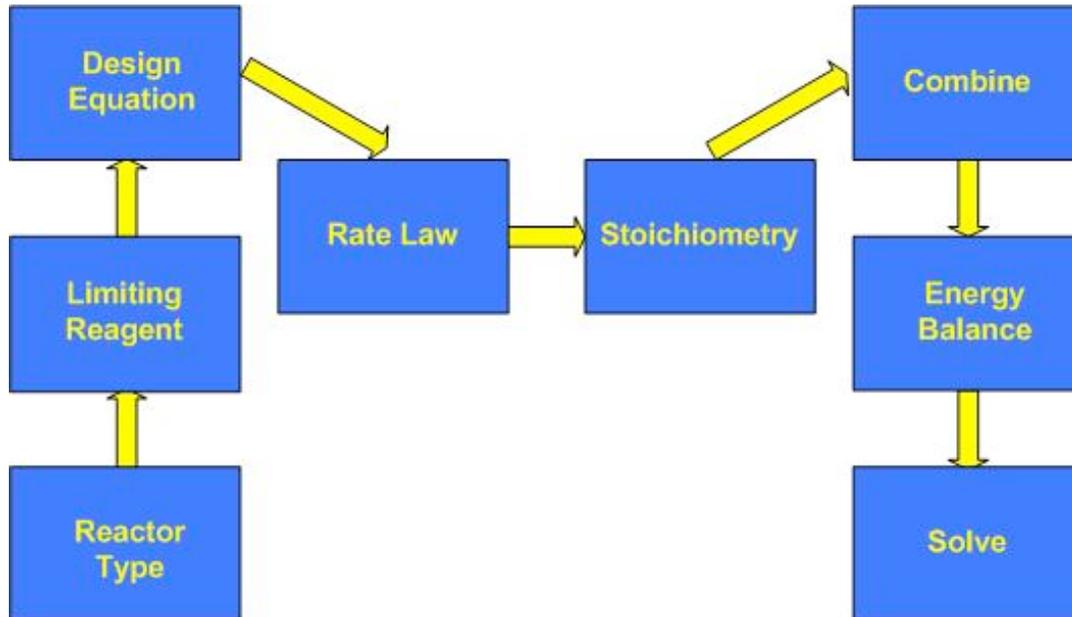
The model presented will include the following assumptions:

-Negligible pressure drop

-Single elementary reaction

-Power law kinetics

-Constant fluid properties (i.e. heat capacity)

To describe the basic algorithm for an adiabatic PFR, a simple case will be considered where pure A enters the reactor:

$aA <--> bB$

1. Reactor Type

PFR

2. Limiting Reagent

The design and rate equations should be expressed in terms of the limiting reagent. Here, A is the limiting reagent.

3. Design Equation

The design equation for a PFR can be expressed in terms of several variables, including conversion, moles, and concentration. It is important to note that the design equation in terms of conversion may only be used if one reaction is occuring. Otherwise, it must be expressed in terms of moles or concentration. The equation to describe conversion as a function of volume, as derived from a simple mole balance, is shown below in Equation 1.

$$\frac{dX}{dV} = \frac{-r_A}{F_{AO}} \quad (1)$$

X= conversion

V= volume

r_A= reaction rate of A

F_{Ao}= initial moles of A

4. Rate Law

The rate law is independent of reactor type. It is expressed in terms of a rate constant and concentration. If the reaction is irreversible, the rate law is modeled according to Equation 2. If the reaction is reversible, the rate law models Equation 3 in which the concentration of B and the equilibrium constant must be accounted for.

$$-r_A = kc_A^a \quad (2)$$

566

$$-r_A = k(c_A^a - \frac{c_B^b}{K_c}) \quad (3)$$

k= rate constant

c_A^a = concentration of A with stoichiometric coefficient a

c_B^b = concentration of B with stoichiometric coefficient b

K_C= equilibrium constant

If the reaction is not isothermal, the rate constant and equilibrium constant should be written in terms of temperature as shown in Equation 4 and 5.

$$k = k_0 exp(\frac{E}{R}(\frac{1}{T_0} - \frac{1}{T})) \quad (4)$$

$$K_C = K_{C1} exp(\frac{\Delta H_{RX}}{R}(\frac{1}{T_1} - \frac{1}{T})) \quad (5)$$

k_0= rate constant at T_0

E= activation energy

R= ideal gas constant

T= temperature

K_{C1}= equilbrium constant at T_1

ΔH_{RX}= standard heat of reaction

5. Stoichiometry

Stoichiometry is used to express concentration of a species in terms of concentration of the limiting reactant. The stoichiometry equation depends on whether the reaction occurs

567

in the liquid or gas phase. If it occurs in the gas phase, Equations 6 and 7 are used. For liquid phase, ε becomes zero and the correlation reduces to Equation 8. There is no pressure term included because it is assumed that pressure drop is negligible, and the initial pressure is equal to the final pressure.

$$C_A = C_{Ao} * \frac{(1-X)\ T_0}{(1+\epsilon X)\ T} \quad (6)$$

$$C_B = C_{Ao} * \frac{(\theta_B + \frac{b}{a}X)\ T_0}{(1+\epsilon X)\ T} \quad (7)$$

$$C_A = C_{Ao} * (1-X) \quad (8)$$

$$C_B = C_{Ao} * (\theta_B + \frac{b}{a}X) \quad (9)$$

$\varepsilon = y_{Ao} *$ (stoichiometric coefficients of products - coefficients of reactants)

y_{Ao} = initial mole fraction of A

θ_i = ratio of initial moles (or concentration) of species i to initial moles (or concentration) of A

6. Combine

The design equation, rate law, and stoichiometric equations are combined. For example, assume the reaction is irreversible, liquid phase, and isothermal, with pure A entering the reactor and has the combined equation as shown below.

$$\frac{dX}{dV} = \frac{k * C_{A0}^{'a}(1-X)^a}{F_{A0}} \quad (10)$$

7. Energy Balance

The energy balance can be used to relate temperature and conversion. It assumes heat flow and shaft work are not present.

$$T = \frac{X(-\Delta H_{RX}) + \Sigma \Theta_i C_{pi} T_o + X \Delta C_p T_r}{\Sigma \Theta_i C_{pi} + X \Delta C_p} \quad (11)$$

C_{pi}= heat capacity of species i

ΔCp= Cp products- Cp reactants

If $\Delta Cp = 0$, the energy balances reduces to the following:

$$T = \frac{X(-\Delta H_{RX}) + \Sigma \Theta_i C_{pi} T_o}{\Sigma \Theta_i C_{pi}} \quad (12)$$

8. Solve

Solve the system of equations in Excel. The ODE is approximated by Euler's method. See the attached spreadsheet for the ODE/Excel modeling of the adiabatic PFR.

7.3 Using The Excel Model

The Excel model can be found here. <u>Adiabatic PFR Model</u>

On the "Reaction Scheme" worksheet enter the various stoichiometric values for your reaction, the specific heats for the various reactants and products, and also the heat of reaction and the reference temperature that the heat of reaction is taken at.

The "Input Conditions" worksheet is fairly straightforward. If the reaction is gas-phase make sure to enter a "1" in the appropriate cell and if liquid-phase enter a "0".

The "Rate Law" sheet contains the definitions for the values you will have to enter on this sheet. Note that T1 is the reference temperature for your rate constant. Make sure to enter a "1" in the cell if your reaction is reversible, and a "0" if it is irreversible.

Make sure not to tamper with the "Numerical Calculations" worksheet! You can, however, scroll downward to extract exact values for volume, conversion, and temperature down the PFR.

Multiple reactions are not included in this model. Having multiple reactions would require a seperate reaction scheme sheet for every reaction, and would require many different adjustments to ensure that all scenarios are accounted for.

7.4 Worked out Example 1

(Note: All data is fictitious)

A plant at the ETF (Equipment Testing Facility) has been acquired in a hostile takeover by Brown Industries, a company for which you are a process engineer. The Engineering Manager, Phil Savage, has assigned you to work on a PFR located within this new plant. The particular PFR in question is used to conduct the following reaction:

A --> B
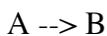
He wants you to provide him with two pieces of information. First, he wishes to know how far down the reactor must the reaction go to achieve 80% conversion. Secondly, he wants to know the temperature at this point.

You decide that first it would be best to ascertain all the thermodynamic and kinetic properties that are inherent to this reaction. Instead of looking them up in Perry's or your old ChemE books, however, you decide to ask the Lead Plant Engineer, Pablo LaValle. Pablo tells you the following:

"Listen wisely to what I say, young grasshopper. This reaction is irreversible, liquid-phase, and elementary. We will feed the reactants into the tube reactor of diameter 50cm at 273 K. We will begin with an initial concentration of A at 1.6 mol/L and an initial molar flow rate of A at 2 mol/min. The heat of reaction at 300 K is -20 kJ/mol A. The rate constant at 300K is 0.2 mol / L * min. The activation energy at 300K is 15.2 kJ/mol. The Cp values for the products and reactants are as follows.

- Cp A = 200 J / mol K
- Cp B = 100 J / mol K

"Go forth and use the Excel sheet that Group B has created and your job will be very easy."

**Solution**

The model for this example can be found [here].

Pablo gave all the data needed to complete the model. After inputing the data, the final page of the spreadsheet was used to determine the reactor conditions Savage desired.

X = 0.8 at a reactor length of 33.4 meters. The temperature at this point of the reactor is 388.4K

7.5 Worked out Example 2

(Note: All data is fictitious)

You have just completed your initiation in the ultra-super secret elite society known as the League of Chemical Engineers which is headquartered high in the Himalayans. They have an adiabatic PFR (diameter = 10cm) in their possession which they use to produce two gases, known as B and C, from a gas-phase reactant known as A. The reversible gas reaction is shown below:

A <--> B + C
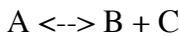
Your sensei wishes to know what the maximum conversion possible is and also what length of reactor would be needed to reach a conversion of 15%. First, however, he orders you to climb K2 where the reaction data is stored by the society. After a perilous 5 day journey you discover the following information:

- Cp_A = 40 kJ/mol K
- Cp_B = 25 kJ/mol K
- Cp_C = 15 kJ/mol K

- C_{AO} = 0.271 mol/L
- F_{AO} = 5 mol/min
- Initial Temp = 450K
- Heat of reaction at 300K = -20 kJ/mol A
- Activation energy = 31.4 kJ/mol
- K_{eq} at 300K = 0.01 mol/L
- k_forward at 300K = 0.133 mol /(L*min)

Solution

After climbing back down K2, and finding a computer, the data obtained can be input into the adiabatic model. The solution can be found here.

Be sure to set the values for both gas phase reaction and reversible reaction to 1.

Using the data table found on the last page of the spreadsheet, the desired information can be extracted.

The maximum conversion that can be achieved is X = 0.187

The reactor requires 64.8 meters to achieve a conversion of 15%.

7.6 References

Fogler, H. Scott. *Elements of Chemical Reaction Engineering*. Pearson Education, Inc., 2006.

Section 8. Cruise control for an electric vehicle

Title: Modeling and PID Controller Example - Cruise Control for an Electric Vehicle
Author: Alex Dowling
Date Prepared: May 3, 2009

8.1 Introduction

Controls principles developed in this course can be applied to non-chemical engineering systems such as automobiles. Some companies, such as NAVTEQ, are developing adaptive cruise control products that use information about the upcoming terrain to shift gears in a more intelligent manner which improves speed regulation and fuel economy. This case study will examine the basics of develop a speed controller for an electric vehicle.

An electric vehicle was chosen for the following reasons:

- Electric vehicles are interesting from an engineering perspective and may become a reality for consumers in the future
- Torque produced by an electric motor is instantaneous (for all practical purposes). Thus actuator lag can be ignored, simplifying the development of said controller.
- Some electric vehicles feature motors directly integrated into the hub of the drive wheel(s). This eliminates the need for a transmission and simplifies vehicle dynamics models.

8.2 Forces

As shown in the free body diagram below, there are six forces acting on the vehicle:

1. Rolling Resistance
2. Aerodynamic Drag
3. Aerodynamic Lift
4. Gravity
5. Normal
6. Motor

573
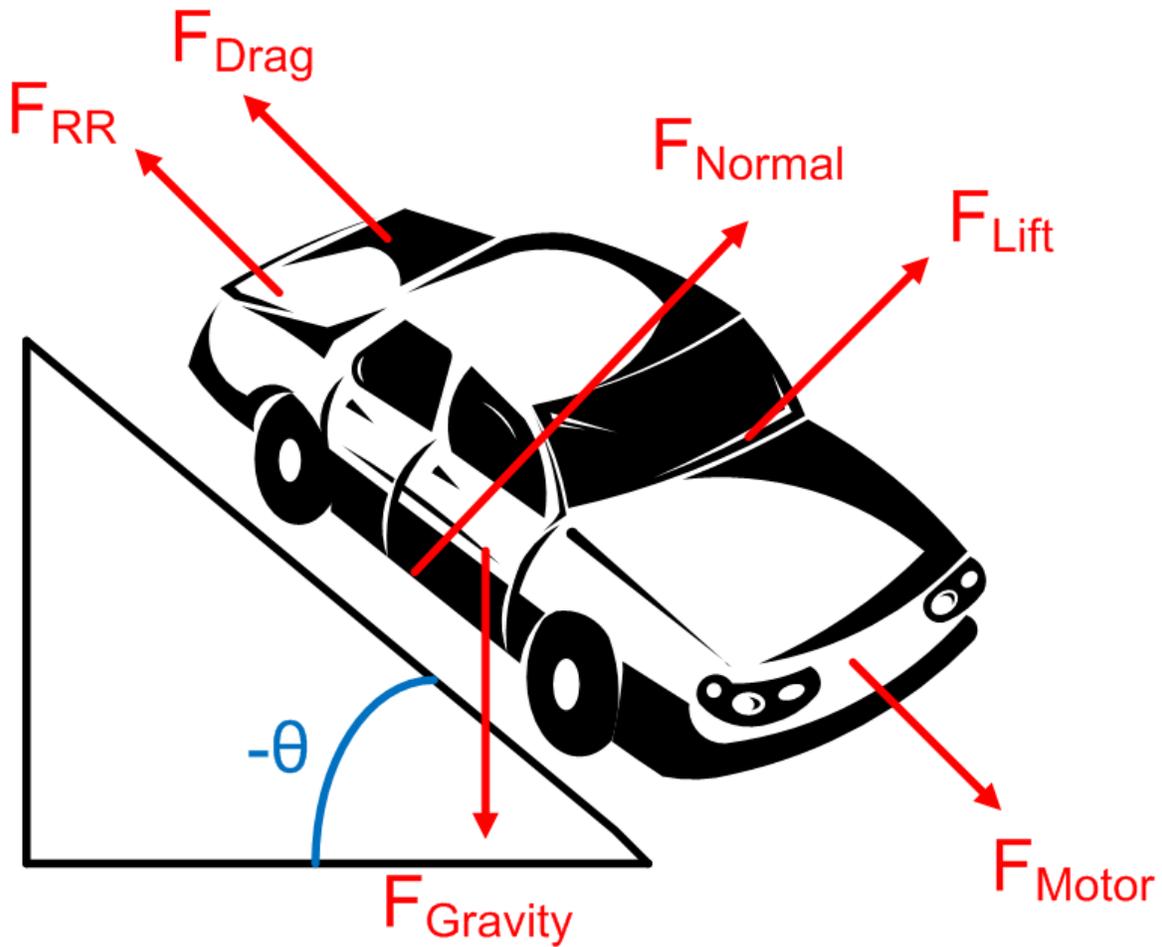
Figure 1: A free body diagram of the forces acting on the vehicle. $-\theta$ is used to denote the grade of the road such that a positive value of θ corresponds to the vehicle traveling uphill.

8.2.1 Rolling Resistance

Rolling resistance is due the tires deforming when contacting the surface of a road and varies depending on the surface being driven on. It can be model using the following equation:

$$F_{RR} = C_{rr1}v + C_{rr2}F_N$$

The two rolling resistance constants can be determined experimentally and may be provided by the tire manufacture. F_N is the normal force.

574

8.2.2 Aerodynamic Drag

Aerodynamic drag is caused by the momentum loss of air particles as they flow over the hood of the vehicle. The aerodynamic drag of a vehicle can be modeled using the following equation:

$$F_{drag} = (1/2)\rho C d A v^2$$

- ρ is the density of air. At 20°C and 101kPa, the density of air is 1.2041 kg/m³.
- CdA is the coefficient of drag for the vehicle times the reference area. Typical values for automobiles are listed here.
- v is the velocity of the vehicle.

8.2.3 Aerodynamic Lift

Aerodynamic lift is caused by pressure difference between the roof and underside of the vehicle. Lift can be modeled using the following equation:

$$F_{lift} = (1/2)\rho C l A v^2$$

- ρ is the density of air. At 20°C and 101kPa, the density of air is 1.2041 kg/m³.
- ClA is the coefficient of lift for the vehicle times the reference area.
- v is the velocity of the vehicle.

8.2.4 Gravity

In the diagram above, there is a component of gravity both in the dimension normal to the road and in the dimension the vehicle is traveling. Using simple trigonometry, the component in the dimension of travel can be calculated as follows:

$$F_{G,travel} = mg\sin(-\theta)$$

- m is the mass of the vehicle.
- g is the acceleration due to gravity.

8.2.5 Normal Force

The normal force is the force excerted by the road on the vehicle's tires. Because the vehicle is not moving up or down (relative to the road), the magnitude of the normal forces equals the magnitude of the force due to gravity in the direction normal to the road.

$$F_N = F_{G,norm.} - F_{lift} = mg\cos(-\theta) - (1/2)\rho ClAv^2$$

- m is the mass of the vehicle.
- g is the acceleration due to gravity.

8.2.6 Motor

The torque produced by an electric motor is roughly proportional to the current flowing through the stater of the motor. In this case study, the current applied to the motor will be controlled to regulate speed. Applying a negative current will cause the vehicle to regeneratively brake.

$$\tau = k_{motor} I$$

$$F_M = \frac{\tau}{r} = \frac{k_{motor} I}{r}$$

- τ is the torque produced by the motor.
- I is the current flowing through the motor.
- r is the radius of the tire.
- k_{motor} is a constant.

8.3 Newton's Second Law

Using Newton's Second Law, a differential for the vehicle's speed can be obtained.

$$ma = \sum F = F_M - F_{drag} - F_{RR} + F_{G,travel}$$

Substituting in the expressions for various forces detailed above yields the following:

$$ma = \frac{\tau}{r} - (1/2)\rho CdAv^2 - C_{rr1}v - C_{rr2}F_N + mg\sin(-\theta)$$

576

Further substituting the expression for normal forces yields the following:

$$ma = \frac{\tau}{r} - (1/2)\rho C dA v^2 - C_{rr1}v - C_{rr2}[mg\cos(-\theta) - (1/2)\rho C l A v^2] + mg\sin(-\theta)$$

Substituting $\dfrac{dv}{dt}$ for a results in the following:

$$m\frac{dv}{dt} = \frac{\tau}{r} - (1/2)\rho C dA v^2 - C_{rr1}v - C_{rr2}[mg\cos(-\theta) - (1/2)\rho C l A v^2] + mg\sin(-\theta)$$

Grouping like turns results in the following:

$$m\frac{dv}{dt} = \frac{\tau}{r} - [(1/2)\rho C dA - C_{rr2}(1/2)\rho C l A]v^2 - C_{rr1}v - C_{rr2}mg\cos(-\theta) + mg\sin(-\theta)$$

In order to simply the remaining analysis, several constants are defined as follows:

$$\alpha = (1/2)\rho r(C dA - C_{rr2}C l A)$$

$$\beta = r C_{rr1}$$

$$\gamma = r C_{rr2}mg\cos(-\theta) - rmg\sin(-\theta)$$

Substituting these into the differential equation results in the following expression:

$$\frac{dv}{dt} = \frac{\tau - \alpha v^2 - \beta v - \gamma(\theta)}{mr}$$

It is important that θ (and thus also γ) is a function of vehicle position.

8.4 PID Controller

One way to regulate vehicle speed is to control the torque generated by the electric motor. If the motor torque is greater than the resistive torque acting on the vehicle (a summation of aerodynamic drag, rolling resistance, etc.) the vehicle will accelerate. If the motor torque is less than the resistive torque, the vehicle will slow down.

8.4.1 Expression for Phase Current

For an electric motor, the phase current flowing through the motor is proportional to the torque produced. Thus one strategy for controlling the vehicles speed is to controller the motor phase current.

Using a PID controller architecture, the expression for motor current is the following:

$$I = K_c(v_{set} - v) + \frac{1}{\tau_I} \int (v - vset)dt + \tau_D \frac{d(v_{set} - v)}{dt} + C_{offset}$$

8.4.2 Differential Equation for Velocity

Substituting this expression into the differential equation for vehicle position results in the following:

$$\frac{dv}{dt} = \frac{k_{motor}\left[K_c(v_{set} - v) + \frac{1}{\tau_I} \int (v_{set} - v)dt + \tau_D \frac{d(v_{set} - v)}{dt} + C_{offset}\right] - \alpha v^2 - \beta v - \gamma(\theta)}{mr}$$

Defining another variable x_1 allows for the removal of the integral from the expression.

$$\frac{dx_1}{dt} = v_{set} - v$$

$$\frac{dv}{dt} = \frac{k_{motor}\left[K_c(v_{set} - v) + \frac{x_1}{\tau_I} + \tau_D \frac{d(v_{set} - v)}{dt} + C_{offset}\right] - \alpha v^2 - \beta v - \gamma(\theta)}{mr}$$

If all changes in v_{set} are gradual then $\dfrac{dv_{set}}{dt} \approx 0$. Applying this simplification results in the following expression:

578

$$\frac{dv}{dt} = \frac{k_{motor}\left[K_c(v_{set} - v) + \frac{x_1}{\tau_I} - \tau_D\frac{dv}{dt} + C_{offset}\right] - \alpha v^2 - \beta v - \gamma(\theta)}{mr}$$

Using a little algebra, an expression for $\frac{dv}{dt}$ can be obtained as follows:

$$\frac{dv}{dt} = \frac{k_{motor}K_c(v_{set} - v) + \frac{k_{motor}}{\tau_I}x_1 + k_{motor}C_{offset} - \alpha v^2 - \beta v - \gamma(\theta)}{k_{motor}t_D + mr}$$

8.4.3 Find Fixed Point

The fixed point can be obtained by setting the derivatives to zero and solving the system of equation.

System of Equations:

$$\frac{dv}{dt} = \frac{k_{motor}K_c(v_{set} - v) + \frac{k_{motor}}{\tau_I}x_1 + k_{motor}C_{offset} - \alpha v^2 - \beta v - \gamma(\theta)}{k_{motor}t_D + mr} = 0$$

$$\frac{x_1}{dt} = v_{set} - v = 0$$

Solution:

$$x_1 = \frac{-\tau_I(k_{motor}C_{offset} - \alpha v_{set}^2 - \beta v_{set} - \gamma(\theta)])}{k_{motor}}$$

$$v = v_{set}$$

As expected, a fixed point exists when the set velocity equals the actual velocity of the vehicle.

8.4.4 Linearize System of ODEs

Before the stability of the system can be (easily) examined, the system must be linearized around a fixed point.

Note: To simply the matrix expressions in section the following notation will be used:

$$y_i' = \frac{dy_i}{dt}$$

Overall a linearized system of ODEs has the following form:

$$\begin{bmatrix} y_1' \\ y_2' \\ \vdots \\ y_n' \end{bmatrix} = \mathbf{J} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} + \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix}$$

The first step of linearizing any system of ODEs to calculate the Jacobian. For this particular system, the Jacobian can be calculated as follows:

$$\mathbf{J} = \begin{bmatrix} \frac{\partial v'}{\partial v} & \frac{\partial v'}{\partial x_1} \\ \frac{\partial x_1'}{\partial v} & \frac{\partial x_1'}{\partial x_1} \end{bmatrix} = \begin{bmatrix} \dfrac{-k_{motor}K_c - mr[2\alpha v + \beta]}{mr + k_{motor}\tau_D} & \dfrac{k_{motor}}{t_I(mr + k_{motor}\tau_D)} \\ -1 & 0 \end{bmatrix}$$

The Jacobian is then evaluated at the fixed point:

$$\mathbf{J} = \begin{bmatrix} \dfrac{-k_{motor}K_c - mr[2\alpha v_{set} + \beta]}{mr + k_{motor}\tau_D} & \dfrac{k_{motor}}{t_I(mr + k_{motor}\tau_D)} \\ -1 & 0 \end{bmatrix}$$

The next step if to calculate the vector of constants. For this particular system, said vector can be calculated as follows:

$$\begin{bmatrix} k_v \\ k_{x_1} \end{bmatrix} = -\mathbf{J}\begin{bmatrix} v \\ x_1 \end{bmatrix}\Bigg|_{fixedpoint} = -\begin{bmatrix} \dfrac{-k_{motor}K_c-2\alpha v_{set}-\beta]}{mr+k_{motor}\tau_D} & \dfrac{k_{motor}}{t_I(mr+k_{motor}\tau_D)} \\ -1 & 0 \end{bmatrix}\begin{bmatrix} v_{set} \\ \dfrac{-\tau_I(k_{motor}C_{offset}-\alpha v_{set}^2-\beta v_{set}-\gamma(\theta)])}{k_{motor}} \end{bmatrix}$$

$$\begin{bmatrix} k_v \\ k_{x_1} \end{bmatrix} = \begin{bmatrix} \dfrac{-(-k_{motor}K_c-2\alpha v_{set}-\beta])v_{set}}{(mr+k_{motor}\tau_D)} + \dfrac{(k_{motor}C_{offset}-\alpha v_{set}^2-\beta v_{set}-\gamma(\theta)])}{(mr+k_{motor}\tau_D)} \\ v_{set} \end{bmatrix}$$

Combining the Jacobian and vector of constants results in the following linearized system:

$$\begin{bmatrix} v' \\ x_1' \end{bmatrix} = \mathbf{J}\begin{bmatrix} v \\ x_1 \end{bmatrix} + \begin{bmatrix} k_v \\ k_{x_1} \end{bmatrix}$$

$$\begin{bmatrix} v' \\ x_1' \end{bmatrix} = \begin{bmatrix} \dfrac{-k_{motor}K_c-2\alpha v_{set}-\beta}{m+k_{motor}]\tau_D} & \dfrac{k_{motor}}{t_I(mr+k_{motor}\tau_D)} \\ -1 & 0 \end{bmatrix}\begin{bmatrix} v \\ x_1 \end{bmatrix} + \begin{bmatrix} \dfrac{-(-k_{motor}K_c-2\alpha v_{set}-\beta])v_{set}}{(mr+k_{motor}\tau_D)} + \dfrac{(k_{motor}C_{offset}-\alpha v_{set}^2-\beta v_{set}-\gamma(\theta)])}{(mr+k_{motor}\tau_D)} \\ v_{set} \end{bmatrix}$$

8.4.5 Stability Analysis

To assess the stability of the controller, the eigenvalues of the Jacobian in the linearized systems of ODEs can be examined. In general, an eigenvalue (*lambda*) is the solution to the following equation:

$$|\mathbf{J} - \lambda\mathbf{I}| = 0$$

Using a computer to solve said equation, the eigenvalues of this particular system can be found to be the following:

$$\lambda = \frac{k_{motor}K_c\tau_I + \beta\tau_I + 2\alpha v_{set}\tau_I \pm \sqrt{\tau_I(k_{motor}^2K_c^2\tau_I + 2k_{motor}K_c\tau_I\beta + 4k_{motor}K_c\tau_I\alpha v_{set} + \beta^2\tau_I + 4\beta\tau_I\alpha v_{set} + 4\alpha^2 v_{set}^2\tau_I - 4k_{motor}mr - 4k_{motor}^2t_D)}}{2\tau_I(mr + k_{motor}\tau_D)}$$

For the system to be stable, the real component of all eigenvalues must be non-positive. The following inequality must be true for a stable controller:

$$k_{motor}K_c\tau_I+\beta\tau_I+2\alpha v_{set}\tau_I < -Real(\tau_I(k_{motor}^2K_c^2\tau_I+2k_{motor}K_c\tau_I\beta+4k_{motor}K_c\tau_I\alpha v_{set}+\beta^2\tau_I+4\beta\tau_I\alpha v_{set}+4\alpha^2 v_{set}^2\tau_I-4k_{motor}mr-4k_{motor}^2t_D))$$

For the system to not oscillate, the imaginary component of all eigenvalues must be zero. The following inequality must be true for a non-oscillating controller:

$$0 > \tau_I(k_{motor}^2K_c^2\tau_I+2k_{motor}K_c\tau_I\beta+4k_{motor}K_c\tau_I\alpha v_{set}+\beta^2\tau_I+4\beta\tau_I\alpha v_{set}+4\alpha^2 v_{set}^2\tau_I-4k_{motor}mr-4k_{motor}^2t_D)$$

581

Interestingly, neither of these criteria depend on the grade of the road (θ). However, during the analysis, it was assumed that θ is constant. For most roads, this is not the case; θ is actually a function of vehicle position. In order to add this additional level of detail, the original system of ODEs needs to be revised:

$$\frac{dv}{dt} = \frac{k_{motor}K_c(v_{set} - v) + k_{motor}x_1 + k_{motor}C_{offset} - \alpha v^2 - \beta v - \gamma(\theta)}{k_{motor}t_D + mr}$$

$$\frac{dx_1}{dt} = v_{set} - v$$

$$\frac{ds}{dt} = v$$

$$\theta = f(s)$$

Unfortunately for any normal road, the grade is not a simple (or even explicate) function of position (s). This prevents an in depth analytical analysis of stability. However for a very smooth road with very gradual changes in grade, the stability of the controller should be unaffected.

8.4.6 Example Electric Vehicle

Simulating the system also for other properties of the controller to be examined. This section presents the results from simulating a specified fictitious electric vehicle.

Parameters

For the fictitious electric vehicle simulated in this analysis, the following parameters were used. These parameters are roughly based on the parameters one would expect to see in a typical electric automobile.

$$m = 1400 \ kg$$

$$CdA = 0.58 \ m^2$$

$$ClA = 0 \ m^2$$

$$C_{rr1} = 2.75 \ \frac{Ns}{m} 0.018$$

$$C_{rr2} = 0.018$$

$$k_{motor} = 1.2 \ \frac{Nm}{A_{rms}}$$

Root Locus Plots

Using the parameters above, Root Locus Plots of the system were constructed to numerically explore the stability. The following controller constants were used when constructing the plots:

$$v_{set} = 25 \ m/s$$

$$K_c = 55$$

$$\tau_I = 0.5$$

$$\tau_D = 1$$

$$C_{offset} = 0$$

Root Locus Plot

Root Locus Plot

584
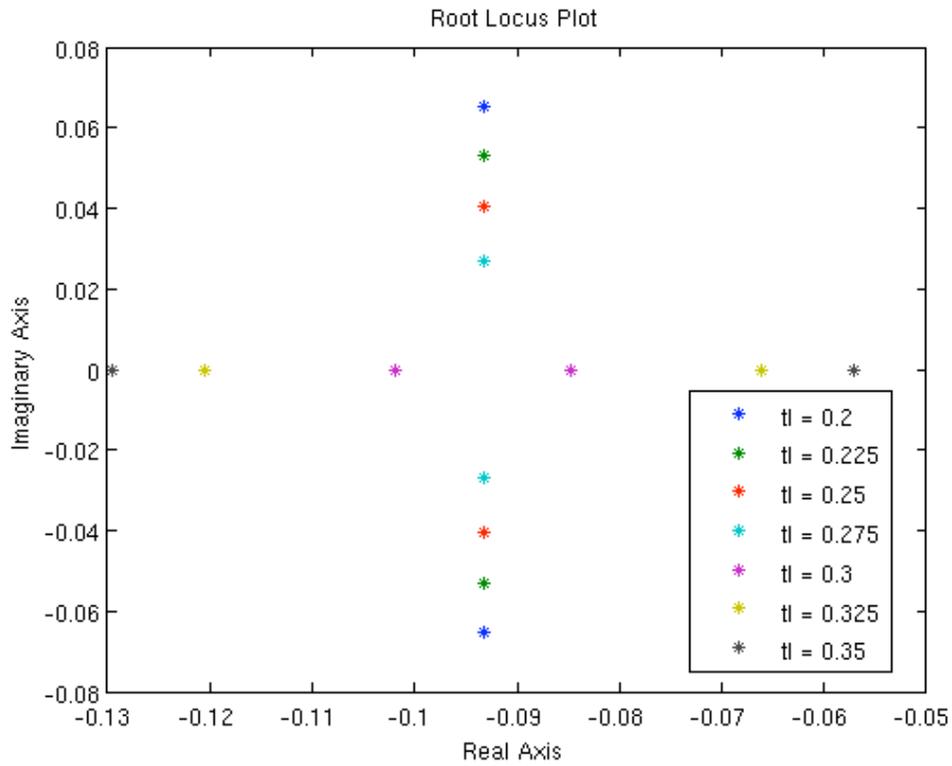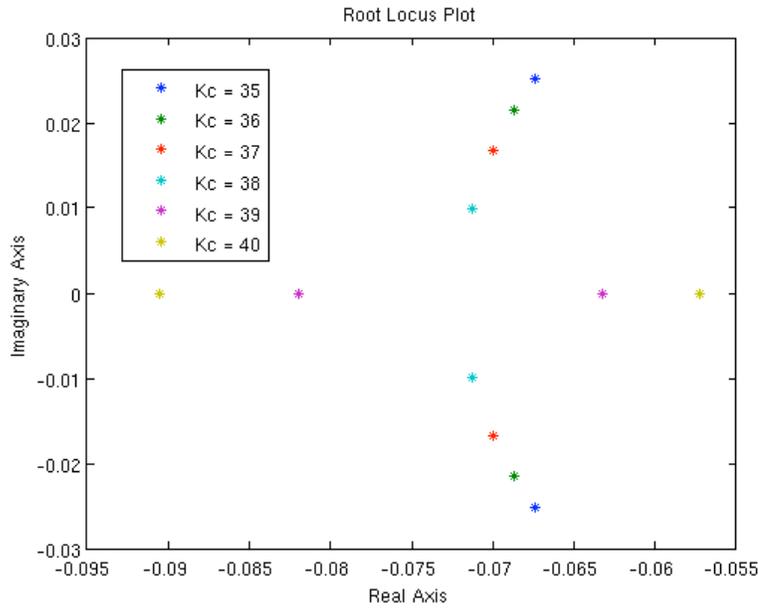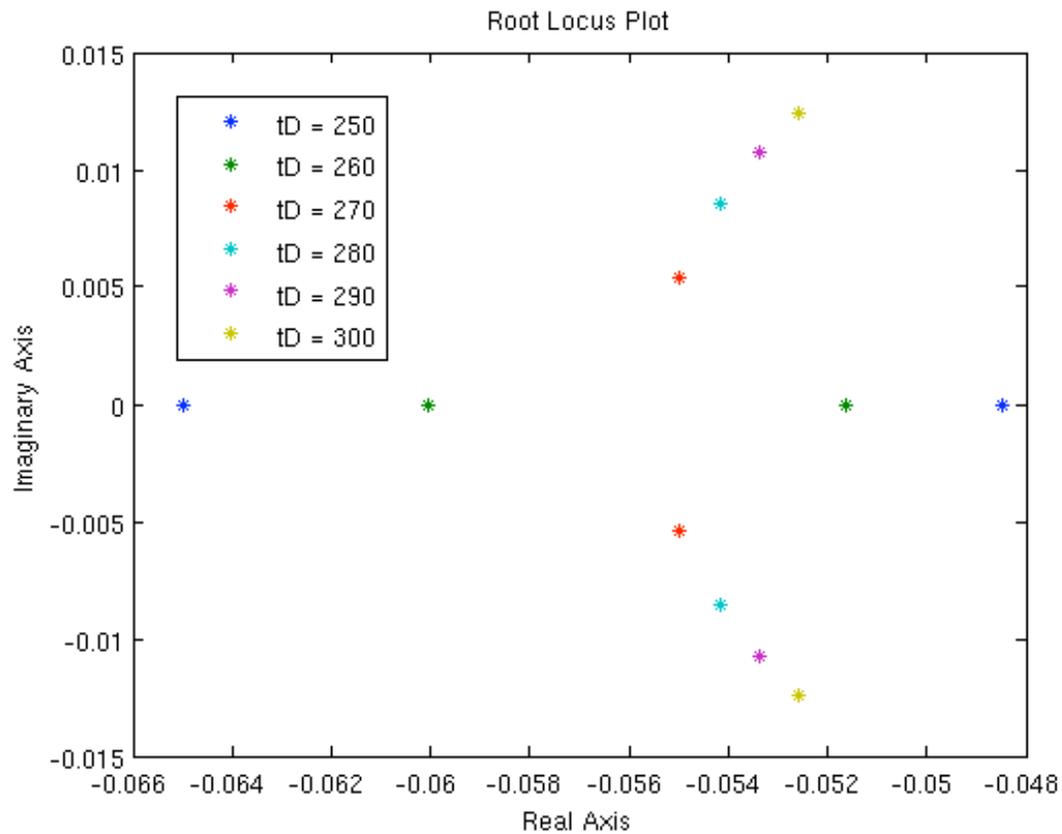
Root Locus Plot

These Root Locus plots show that the controller with said constants is both stable and does not oscillate.

Phase Portrait

A phase portrait of the system with the following parameters was also constructed:

$$v_{set} = 25 \ m/s$$

$$K_c = 55$$

$$\tau_I = 0.5$$

$$\tau_D = 20$$

$$C_{offset} = 0$$

v ' = (6/5 Kc (vset - v) + 6/5/tl x1 + 6/5 coff - 51117/50000(cv² 55203946s/25000))#462 + 6/5 tD)
x1 ' = vset - v   vset = 25   tD = 20

Cursor position: (-1.62, 426)

The backward orbit from (31, 2.9e+02) left the computation window.
Ready.
The forward orbit from (8.9, 1.6e+02) --> a possible eq. pt. near (25, 3e+02).
The backward orbit from (8.9, 1.6e+02) left the computation window.
Ready.
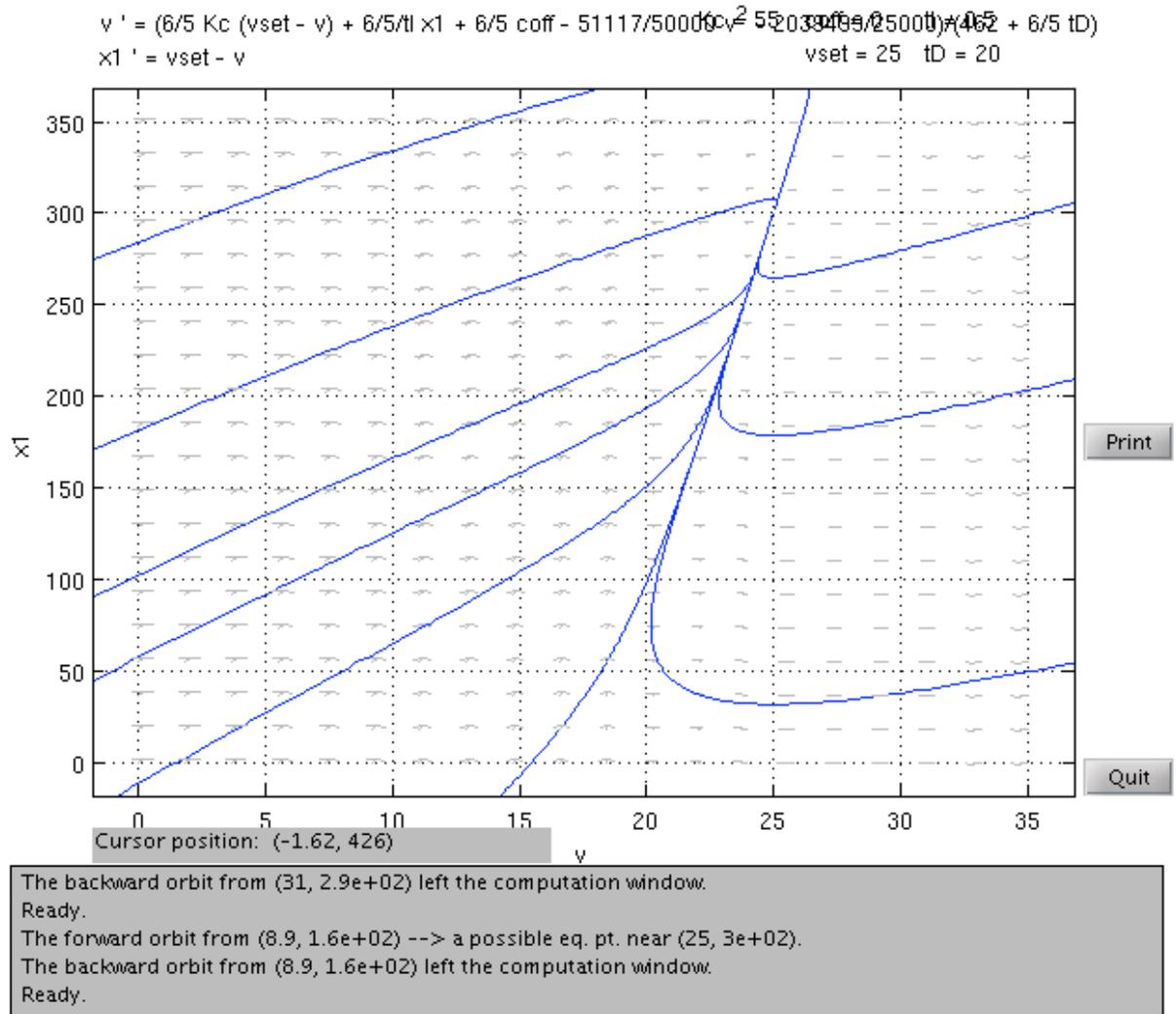
The phase portrait shows that the system is both stable and does not oscillate, as predicted by the Root Locus plots.

Driving Simulation on Level Terrain

The vehicle was simulated starting at 10 m/s and accelerating to 25 m/s via cruise control on level terrain ($\theta = 0$). For this simulation, the following constants were used:

$$v_{set} = 25 \ m/s$$

$$K_c = 55$$

$$\tau_I = 0.5$$

586

$$\tau_D = 20$$

$$C_{offset} = 146 \approx \frac{\alpha v_{set}^2 + \beta v_{set} + \gamma(\theta = 0)}{k_{motor}}$$
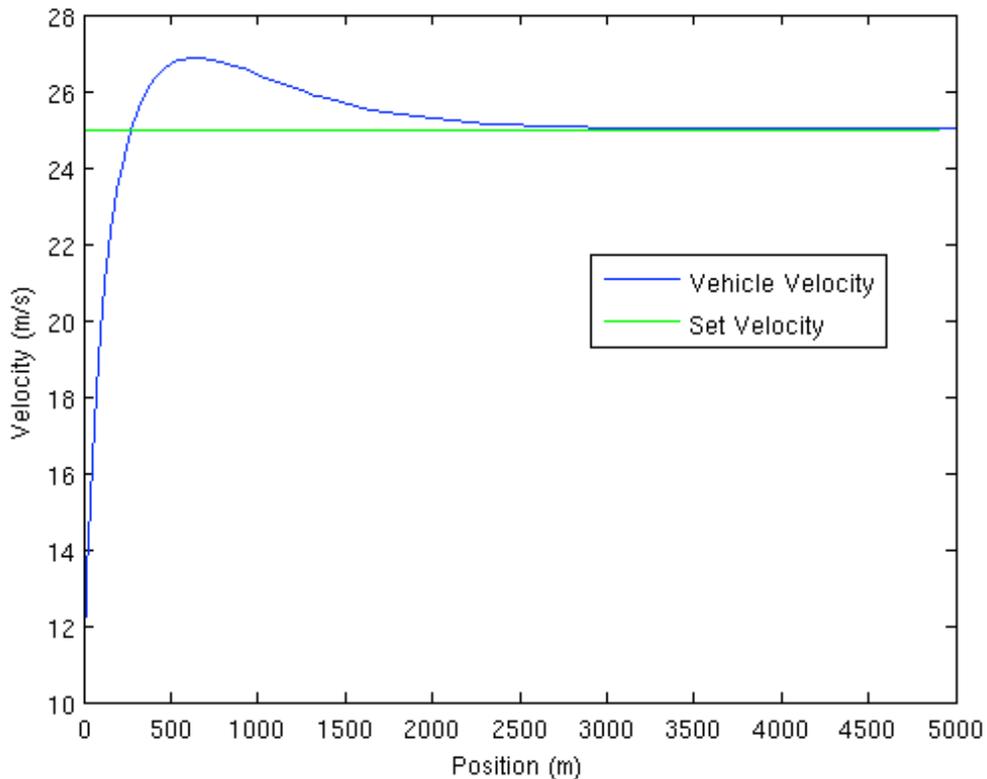
Interestingly this graph shows oscillation, despite the Root Locus plots and phase diagrams. It is important to remember, however, that the Root Locus plot and stability methods involve linearizing the system. It is possible the linearized system is not a good approximation of the system.

It is also important to remember that this particular example involves a large set point change, which can induce oscillations in certain systems.

Driving Simulation on Unlevel Terrain

In order to explore how the controller behaves on a road with a non-zero grade, a route with hills was constructed from the following equation, where *h* is elevation and *s* is position.:

587

$$h = 100 \sin\left(\frac{s}{250\pi}\right) + \frac{s}{2000}$$
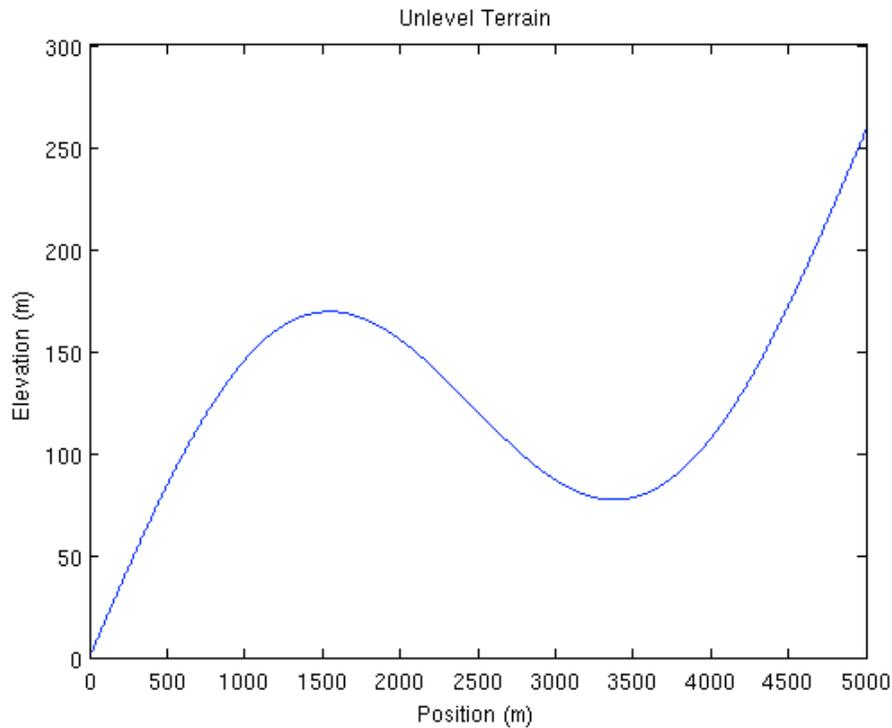
The vehicle was simulated driving said road starting with the following initial conditions and controller constants:

Initial Conditions

$$s(t = 0) = 0$$

$$x_1(t = 0) = 0$$

$$v(t = 0) = v_{set} = 25 \ m/s$$

Controller Constants

$$k_c = 55$$

$$\tau_I = 0.5$$

588

$$\tau_D = 20$$

$$C_{offset} = 146$$
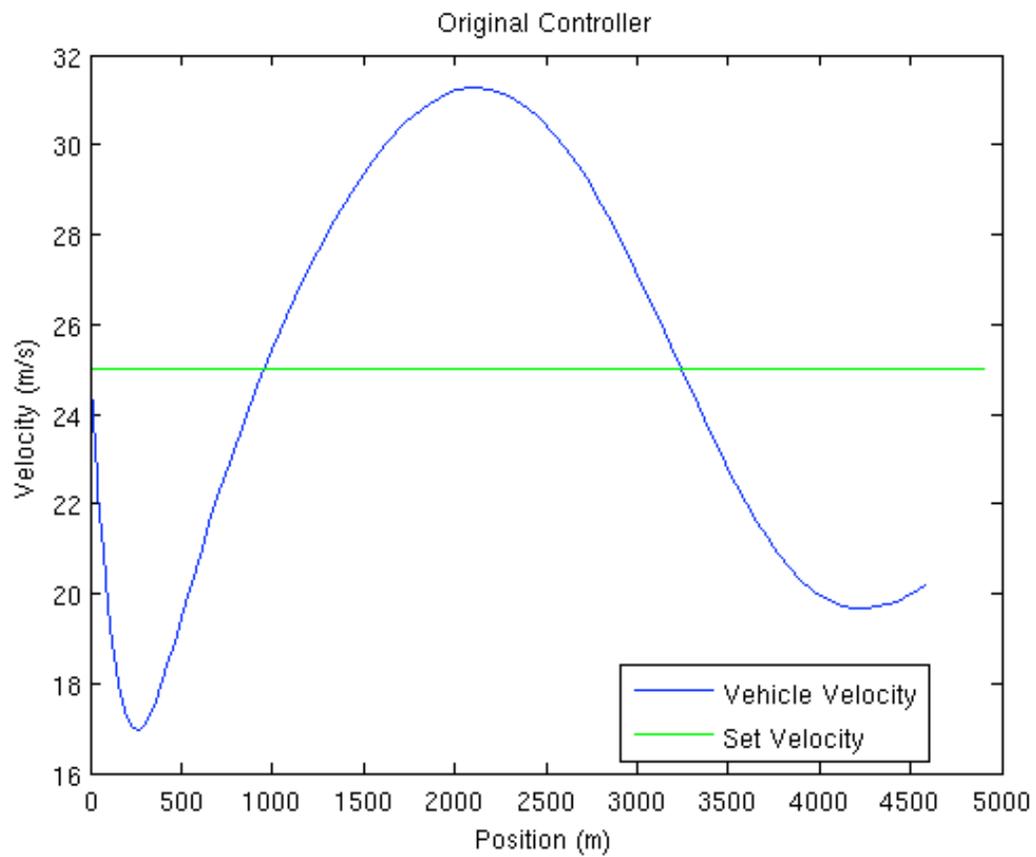
The current controller tuning is inadequate for this road. There are vary large variations in velocity. In order to reduce these variation, the propartional gain K_c was increased by a factor of 5. Below are the results:

Controller with 5x Kc
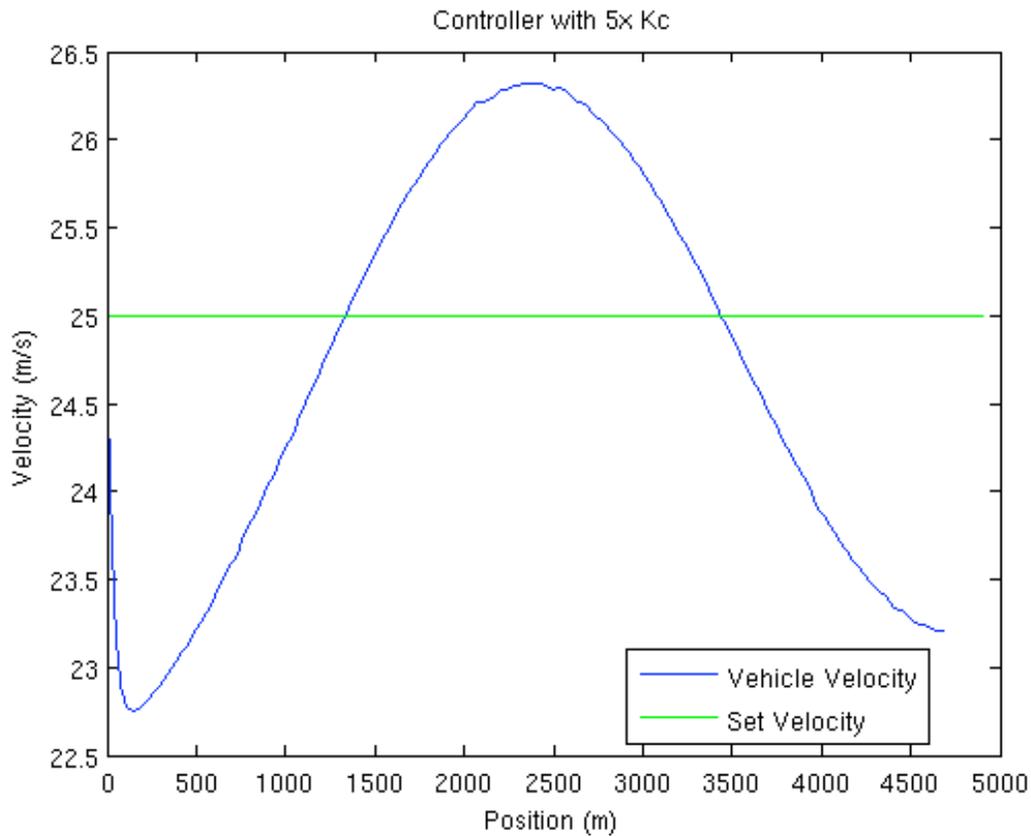
Using the Optimization Toolbox in MATLAB, the controller was optimized by minimizing the sum of the errors (difference between vehicle velocity and set velocity) on this particular segment of road. Below are the optimized controller constants and a plot of the velocity profile:

$$k_c = 232.58$$

$$\tau_I = 0.001$$

$$\tau_D = 0$$

$$C_{offset} = 114.62$$
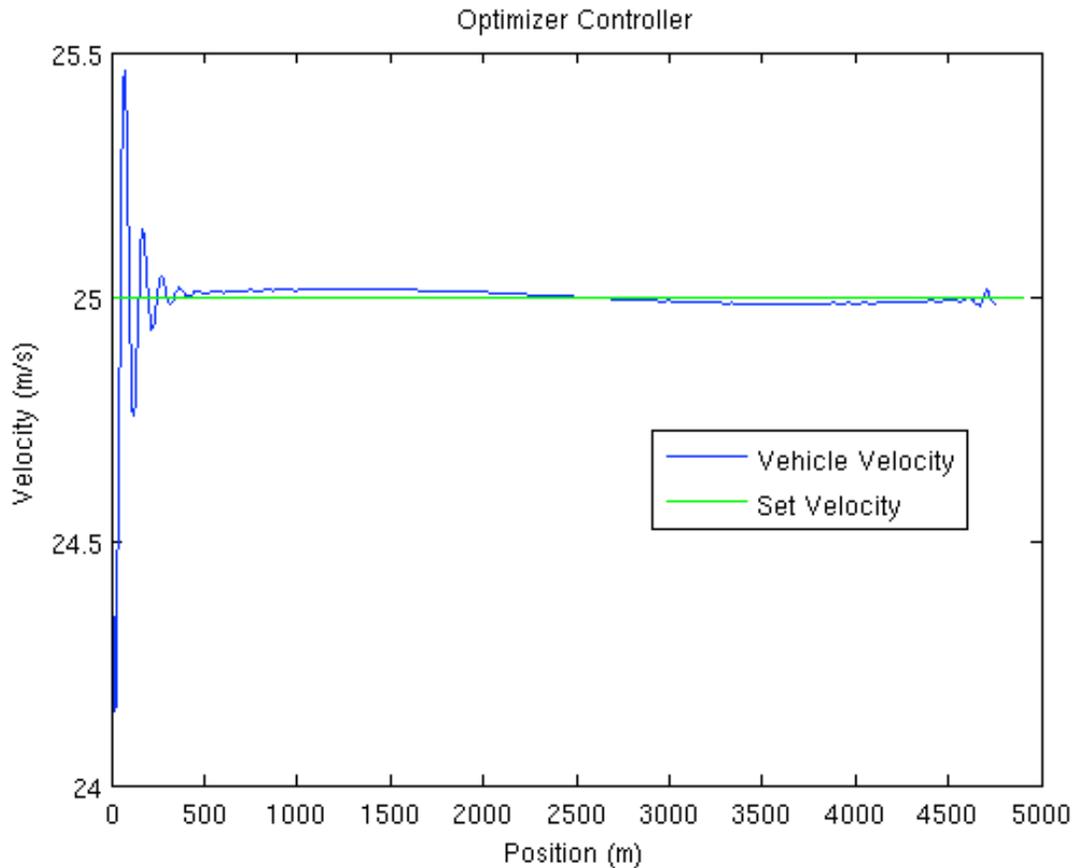
This example shows the power of optimization and model predictive control.

8.5 Summary

This example demonstrates the following:

1. Modeling a simple, non-chemical engineering system.
2. Develop a PID controller for said physical dynamical system.
3. Manipulate said system to develop an system of differential equations.
4. Find fixed point(s) for said system.
5. Linearize said system of differential equations.
6. Find the eigenvalues of said linearized system.
7. Construct Root Locus plots for said linearized system.
8. Construct a phase portrait for said dynamical system.
9. Simulate said dynamical system under various conditions.
10. Demonstrate the idea of model predictive control by optimize said controller under a specific scenario.

Section 9. Blood glucose control in diabetic patients

Written by: Robbert Appel, Jessica Rilly, Jordan Talia

9.1 Background of why insulin control is important for diabetic patients

Diabetes mellitus is a disease of the endocrine system where the body cannot control blood glucose levels. There are two general classifications of diabetes:

Type I (also known as juvenile diabetes)

- Genetic predisposition and/or an autoimmune attack destroys T-cells of pancreas
- Body cannot produce insulin to regulate blood glucose

Type II

- Most common form of diabetes and has reached epidemic status in the United States
- Usually caused by lifestyle
- Obesity reduces body's responsiveness to insulin

Treatment for both types of diabetes may include exercise, dieting, oral medications, or insulin injections. Most insulin dependent diabetics follow a management plan that requires frequent testing of blood glucose levels and then injection of a prescribed dose of insulin based on the blood glucose level. However, the downside of this treatment method is that there is no predictive control. If blood glucose levels are falling and insulin is administered, a hypoglycemic episode may occur. Recent biomedical advancements have resulted in continuous blood glucose monitoring devices as well as insulin pumps. Continuous monitoring allows for finer blood glucose control and can help predict fluctuations in the blood glucose level. Insulin pumps replace the need to administer insulin injections by automatically injecting a prescribed dose, however it requires blood glucose level input from the patient. In the future, insulin pumps and continuous blood glucose monitors may be integrated forming a closed loop control system which will can replace the body's own faulty control system.
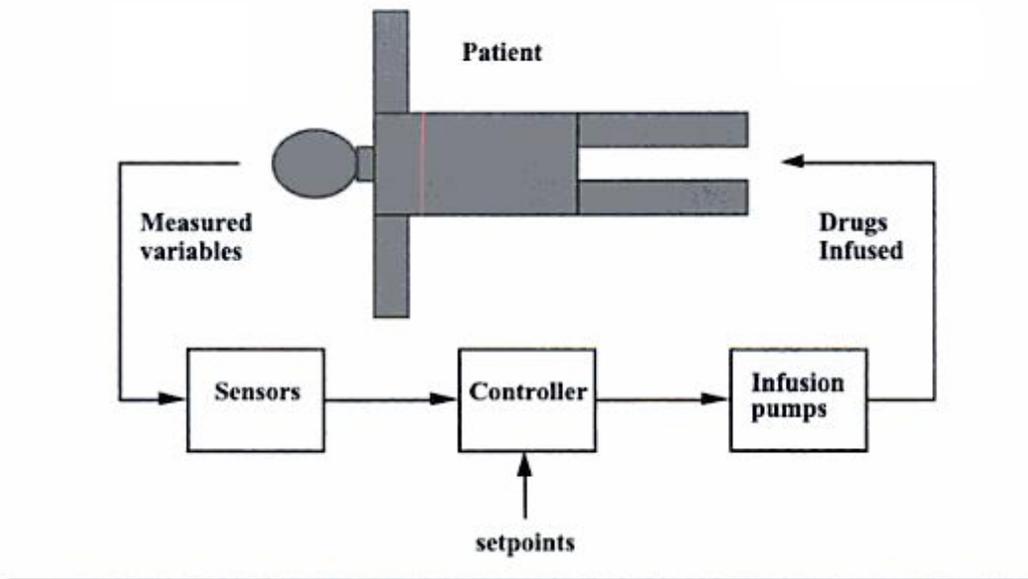
Figure 1: Control Schematic of Insulin Infusion

9.2 Mathematical model for a closed loop insulin delivery system

The following set of differential equations are known as the Bergman "minimal model":

$$\frac{dG}{dt} = -p_1 G - X(G - G_b) + \frac{G_{meal}}{V_1}$$

$$\frac{dX}{dt} = -p_2 X + p_3 I$$

$$\frac{dI}{dt} = -n(I + I_b) + \frac{U}{V_1}$$

Where:

G = deviation variable for blood glucose concentration

X = deviation variable for insulin concentration in a "remote" compartment

I = deviation variable for blood insulin concentration

G_{meal} = a meal disturbance input in glucose

U = the manipulation insulin infusion rate

G_b = steady state value of blood glucose concentration

I_b = steady state value of blood insulin concentration

Blood parameters include p_1, p_2, p_3, n, V_1(blood volume). These are specific to the blood specimen and must be predetermined.

A linear state space model can be used to express the Bergman equations seen above. The general form for a state space model can be seen below:

$$
\begin{bmatrix} \dot{x}_1 \\ \vdots \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nm} \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix}
$$

$$
\begin{bmatrix} y_1 \\ \vdots \\ y_r \end{bmatrix} = \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{r1} & \cdots & c_{rn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} d_{11} & \cdots & d_{1m} \\ \vdots & \ddots & \vdots \\ d_{r1} & \cdots & d_{rm} \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix}
$$

In general:

$$\dot{x} = Ax + Bu$$

y = Cx + Du

Where:

x = states

u = inputs

y = outputs

$$A = \begin{bmatrix} -p_1 & -G_b & 0 \\ 0 & -P_2 & P_3 \\ 0 & 0 & -n \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & \frac{1}{V_1} \\ 0 & 0 \\ \frac{1}{V_1} & 0 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$

$$D = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Using this general formula, we can deconstruct the Bergman equations as a linear state space model. The first input is the insulin infusion and the second input represents the meal glucose disturbance.

First Input:

$$\begin{bmatrix} \dot{G} \\ \dot{X} \\ \dot{I} \end{bmatrix} = \begin{bmatrix} -p_1 & -G_b & 0 \\ 0 & -P_2 & P_3 \\ 0 & 0 & -n \end{bmatrix} \begin{bmatrix} G \\ X \\ I \end{bmatrix} + \begin{bmatrix} 0 & \frac{1}{V_1} \\ 0 & 0 \\ \frac{1}{V_1} & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

Where

$\dot{G}$ = differential blood glucose concentration

$\dot{X}$ = differential insulin concentration in a "remote" compartment

$\dot{I}$ = differential blood insulin concentration

Second Input:

$$y = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} G + \begin{bmatrix} 0 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

Where:

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} U - U_b \\ G_{meal} - 0 \end{bmatrix}$$

9.3 Example Set of Parameters

Gb = 4.5 mmol/liter

Ib = 4.5 mU/liter

V1 = 12 liters

p1 = 0/min

p2 = 0.025/min

p3 = 0.0000013 mU/liter

n = 5/54 min-1

It is important to keep track of units when using these parameters. The concentrations listed above are in mmol/liter, but the glucose disturbance has units of grams. Therefore, it is necessary for us to apply a conversion factor of 5.5556 mmol/grams to the G_{meal} term. Using these variables solve one can solve for the steady states, and calculate the basal

insulin infusion rate (Ub) such that is is equal to 16.1667 mU/min. For these parameters, the resulting state space model is

$$A = \begin{bmatrix} 0 & -4.5 & 0 \\ 0 & -0.025 & 0.000013 \\ 0 & 0 & \frac{-5}{54} \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 0.4630 \\ 0 & 0 \\ \frac{1}{12} & 0 \end{bmatrix}$$

It is common practice in the U.S. to describe glucose concentration in units of mg/deciliter as opposed to mmol/liter. Therefor, the units will be converted from mmol/liter to mg/deciliter. The molecular weight of glucose is 180g/mol, and therefore one it is necessary to multiply the glucose state (mmol/liter) by 18, so that the measured glucose output obtained will be in units of mg/deciliter. The following state-output relationship will handle that:

$$C = \begin{bmatrix} 18 & 0 & 0 \end{bmatrix}$$

$$D = \begin{bmatrix} 0 & 0 \end{bmatrix}$$

Through Laplace transforms it will be found that the process transfer function is :

$$G_p(s) = \frac{-3.79}{(40s + 1)(10.8s + 1)s}$$

The disturbance transfer function due to pole/zero cancellation is simply :

$$G_d(s) = \frac{8.334}{s}$$

However, in reality glucose does not directly enter the blood stream. There is a "lag time" associated with the processing of glucose in the gut. It must first be processed here before entering the blood. However, it can be modeled as a first-order function, with a 20-minute time constant. This modifies the above equation for the disturbance transfer function to include the lag in the gut such that:

$$G_d(s) = \frac{8.334}{s(20s + 1)}$$

9.4 Desired Control Performance

The steady-state glucose concentation of 4.5 mmol/liter corresponds to the glucose concentration of 81 mg/deciliter. A diabetic patient, in order to stay healthy, must keep their blood glucose concentration above 70 mg\deciliter. If the blood glucose concentration falls below 70 mg/deciliter the patient may be likely to faint, as this is a very typical short-term symptom of hypoglycemia. The insulin infusion rate (maniuplated input) cannot fall below 0, and this constraint must be set in simulations. Numerous different control strategies can be used. The equation given for the process function above can be simplified down to the following form and a PD or PID controller can be used.

$$G_p(s) = \frac{k_p}{(tau_p + 1)s}$$

A diabetic will know when they are consuming a meal, and therefore when their blood glucose concentration may rise. Therefore a feed-forward control system may be desired.

9.5 References

- Bequette, B. W. Process Control: Modeling, Design, and Simulation New Jersey: Prentice Hall, 2003. pp 81-83, 694-697.

- Diabetes

Part II
Chemical Process Controls

Chapter 7. Mathematics for Control Systems

Section 1. Dirac delta (impulse) function (10/09)

1.1 Introduction

The Dirac delta function $\delta(t - t_0)$ is a mathematical idealization of an impulse or a very fast burst of substance at $t = t_0$. (Here we are considering time but the delta function can involve any variable.) The delta function is properly defined through a limiting process. One such definition is as a thin, tall rectangle, of width ε:

$$\delta(t - t_0) = \frac{1}{\epsilon} \qquad \text{for} \qquad t_0 - \frac{\epsilon}{2} < t < t_0 + \frac{\epsilon}{2}$$

and zero otherwise, in the limit that $\epsilon \to 0$.

Then, we have

$$\int_a^b \delta(t - t_0)dt = \lim_{\epsilon \to 0}[\epsilon \cdot (\frac{1}{\epsilon})] = 1$$

as long as $a < t_0 < b$. When t_0 is outside the range of (a,b), then the integral is zero.

Likewise, for any function $f(t)$ that is continuous and differentiable (analytic) at t_0,

$$\int_a^b \delta(t - t_0)f(t)dt = \lim_{\epsilon \to 0} \left[\frac{1}{\epsilon} \int_{t_0-\epsilon/2}^{t_0+\epsilon/2} f(t)dt \right] = f(t_0)$$

where the quantity in the square brackets above is just the average value of $f(t)$ in the interval $t_0 - \frac{\epsilon}{2} < t < t_0 + \frac{\epsilon}{2}$. Thus, when $\epsilon \to 0$, it becomes just the value at t_0.

$$\int_0^\infty \delta(t-2)t^2 dt = 2^2 = 4$$

$$\int_3^\infty \delta(t-2)t^2 dt = 0$$

1.3 Delta function at the initial time

Note: if one of the limits of the integral coincides exactly with t_0, then the result is usually cut in half:

$$\int_{t_0}^b \delta(t-t_0)f(t)dt = f(t_0)/2$$

for $b > t_0$. For example:

$$\int_2^\infty \delta(t-2)t^2 dt = (1/2)2^2 = 2$$

However, when we think of an impulse to a system at the initial time t_0, then we really consider that the entire delta function is added to the system - that is, the actual time is an infinitesimal amount beyond t_0; that is, $t = t_0^+$. In that case

$$\int_{t_0}^b \delta(t-t_0^+)f(t)dt = f(t_0)$$

1.4 Definition in terms of a Gaussian

Another equivalent definition of the delta functions are as a Gaussian function:

$$\delta(t-t_0) = \lim_{\sigma \to 0} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-t_0)^2}{2\sigma^2}}$$

All the properties work out to be the same. See the wikipedia page

http://en.wikipedia.org/wiki/Dirac_delta_function

where $a = \sqrt{2}\sigma$ is used.

1.5 Physical applications

In Control, the delta function is used an idealization of a very fast perturbation to the system. For example, if you dump a bucket of water into a tank, then the "flow rate" is essentially a delta function - a very highly peaked function, but with a net integral (the total amount of water in the bucket).

In mechanics, and example of the delta function is the force when hitting an object by a hammer. Say you hit a steel ball with a hammer. It moves with a certain velocity representing the total momentum transferred by the hammer. Rather than talk about the force x time (the net momentum transfer), one talks about an "impulse" which is the net momentum transferred in an infinitesimally short amount of time.

1.6 Relation to the step function

The step function, $\Theta(t - t_0)$, is the integral of the delta function or alternatively, the delta function is the derivative of the theta function, where $\Theta(t - t_0)$ is defined at 1 for $t > t_0$ and 0 for $t < t_0$:

$$\delta(t - t_0) = (d/dt)\Theta(t - t_0)$$

$$\int_a^b \delta(t-t_0)dt = \int_a^b (d/dt)\Theta(t-t_0)dt = \Theta(a-t_0) - \Theta(b-t_0) = 1-0 = 1 \qquad \text{for } a < t_0 < b$$

.

Here, the smooth or gaussian definition of the delta function corresponds to a smooth representation of the Θ function as the integral of a gaussian or equivalently, the error function.

Section 2. First-order differential equations (12/14/09)

2.1 Introduction

We consider the general first-order differential equation:

$$\tau \frac{dy(t)}{dt} + y(t) = x(t)$$

The general solution is given by:

$$y(t) = y_0 e^{-(t-t_0)/\tau} + \frac{e^{-(t-t_0)/\tau}}{\tau} \int_{t_0}^{t} x(t') e^{(t'-t_0)/\tau} dt'$$

where $y_0 = y(t = t_0)$. Note that t' is used to be distinguished from the upper limit t of the integral.

To obtain the general solution, begin with the first order differential equation:

$$\tau \frac{dy(t)}{dt} + y(t) = x(t)$$

Divide both sides by τ :

$$\frac{dy(t)}{dt} + \frac{1}{\tau} y(t) = \frac{1}{\tau} x(t)$$

Rewrite the LHS in condensed form using the integrating factor $e^{-t/\tau}$:

$$e^{-t/\tau} \frac{d}{dt}\left[e^{t/\tau} y(t) \right] = \frac{1}{\tau} x(t)$$

Notice how a chain differentiation will return the LHS to the previous form

Simplify:

$$\frac{d}{dt}\left[e^{t/\tau} y(t) \right] = \frac{1}{\tau} x(t) e^{t/\tau}$$

Integrate both sides:

$$e^{t/\tau} y(t) - e^{t_0/\tau} y(t_0) = \frac{1}{\tau} \int_{t_0}^{t} x(t') e^{t'/\tau} dt'$$

Solve for y(t):

$$y(t) = y_0 e^{-(t-t_0)/\tau} + \frac{e^{-t/\tau}}{\tau} \int_{t_0}^{t} x(t') e^{t'/\tau} dt'$$

2.2 Example Solutions of First Order Differential Equations

Consider:

$$\frac{dy(t)}{dt} = x(t)$$

Multiplying both sides by *dt* gives:

$$\int_{t_0}^{t} \frac{dydt}{dt} = \int_{t_0}^{t} x(t')dt'$$

The general solution is given as:

$$y(t) = y_0 + \int_{t_0}^{t} x(t')dt'$$

Now Consider:

$$\frac{dy(t)}{dt} = -ay(t)$$

Dividing both sides by *y(t)* gives:

$$\frac{1}{y(t)} \frac{dy(t)}{dt} = -a$$

which can be rewritten as:

$$\frac{d}{dt}[ln(y(t))] = -a$$

Multiplying both sides by dt, integrating, and setting both sides of the equation as exponents to the exponential function gives the general solution:

$$y(t) = y_0 e^{-a(t-t_0)}$$

Now Consider:

$$\frac{dy(t)}{dt} = -ay(t) + x(t)$$

The detailed solution steps are as follows:

1. Separate y(t) and x(t) terms

$$\frac{dy(t)}{dt} + ay(t) = x(t)$$

2. Rewrite the LHS in condensed form using the "integrating factor" e^{-at}

$$e^{-at}\frac{d(e^{at}y(t))}{dt} = x(t)$$

Notice how a chain differentiation will return the LHS to the form written in step 1

3. Divide both sides by e^{-at}

$$\frac{d(e^{at}y(t))}{dt} = e^{at}x(t)$$

4. Multiply both sides by dt and integrate

$$e^{at}y(t) - e^{at_0}y_0 = \int_{t_0}^{t} x(t')e^{at'}dt'$$

The general solution is given as:

$$y(t) = y_0e^{-a(t-t_0)} + e^{-at}\int_{t_0}^{t} x(t')e^{a(t'-t_0)}dt'$$

2.3 Step Function Simplification

$$\tau\frac{dy(t)}{dt} + y(t) = x(t)$$

For $x(t) = \Theta(t - t_0)$, which is the step function and $y_0 = 0$

That is, $x(t) = 1$ for $t > t_0$ and $x(t) = 0$ otherwise:

The previously derived general solution:

607

$$y(t) = y_0 e^{-(t-t_0)/\tau} + \frac{e^{-(t-t_0)/\tau}}{\tau} \int_{t_0}^{t} x(t') e^{(t'-t_0)/\tau} dt'$$

reduces to:

$$y(t) = e^{-(t-t_0)/\tau}$$

, since the integral equals 1.

The quantity τ can be seen to be the time constant whereby $y(t)$ drops to $1/e$ of its original value

2.4 Sample Problem

Problem Statement:

Consider the differential equation $\dfrac{dy(t)}{dt} = -0.5y(t) + x(t)$

where $x(t) = 2 + 0.01t$

Assuming $y_0 = 0$, what is the behavior of $y(t)$?

Solution:

General solution was derived previously as:

$$y(t) = y_0 e^{-a(t-t_0)} + e^{-at} \int_{t_0}^{t} x(t') e^{a(t'-t_0)} dt'$$

608

For $a = 0.5$ and $y_0 = 0$, while setting $t_0 = 0$, the solution reduces to:

$$y(t) = e^{-0.5t} \int_0^t (2 + 0.01t')e^{0.5(t')} dt'$$

The following link may be referred to for integral tables: S.O.S. Math

Simplifying the solution gives the following:

$$y(t) = \frac{(e^{-0.5t})((x + 198)(e^{0.5t}) - 198))}{50}$$

Plotting this function displays the following behavior:

Behavior of Response, Y(t)
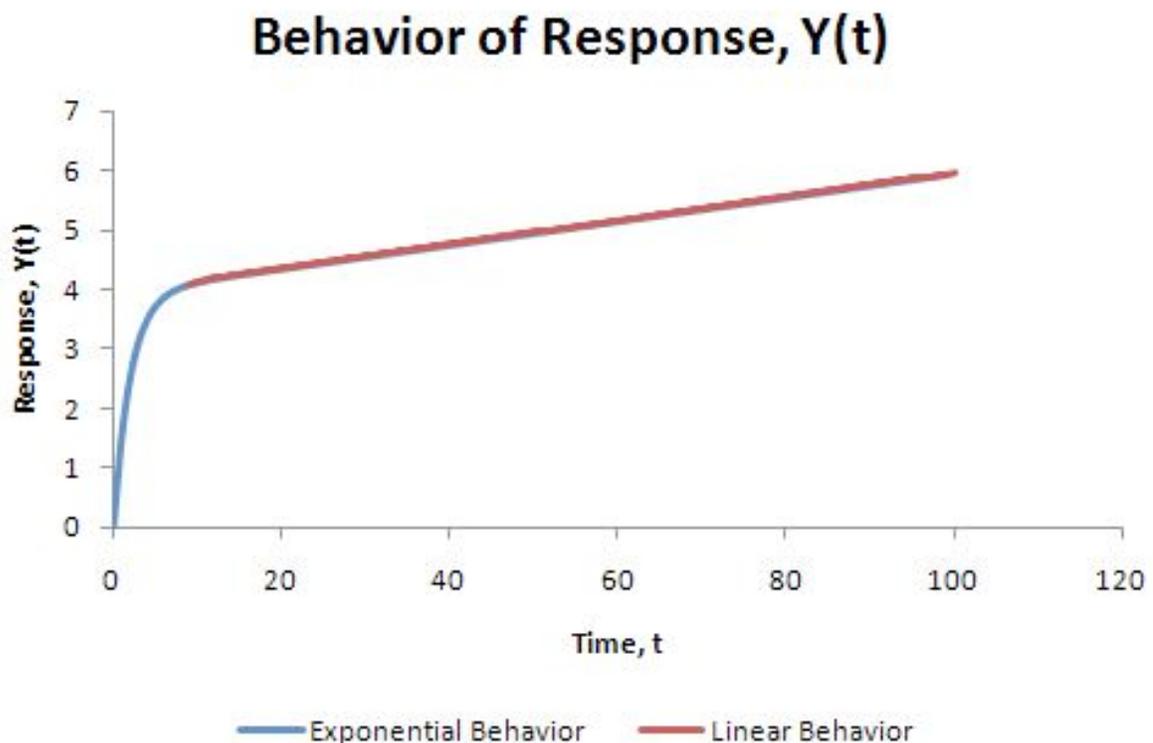
As can be seen clearly from the graph, initially the systemic response shows more exponential characteristics. However, as time progresses, linear behavior dominates.

Section 3. Second-order differential equations (10/13)

3.1 Introduction

We consider the general Second-order differential equation:

$$\tau^2 \frac{d^2Y(t)}{dt^2} + 2\zeta\tau \frac{dY(t)}{dt} + Y(t) = X(t)$$

If you expand the previous Second-order differential equation:

$$\tau_1\tau_2 \frac{d^2Y(t)}{dt^2} + (\tau_1 + \tau_2)\frac{dY(t)}{dt} + Y(t) = X(t)$$

$$(\tau_1 \frac{d}{dt} + 1)(\tau_2 \frac{d}{dt} + 1)Y(t) = X(t)$$

Where:
$$\tau = \sqrt{\tau_1\tau_2}$$
$$\zeta = \frac{\tau_1 + \tau_2}{2\sqrt{\tau_1\tau_2}}$$

Expansion of the differential equation allows you to guess what the shape of the solution (Y(t)) will look like when X(t)=1.

The following rules apply when $\tau_1 = Re(\tau_1) + i*Im(\tau_1)$ and $\tau_2 = Re(\tau_2) + i*Im(\tau_2)$:

- If $Re(\tau_1)$ and $Re(\tau_2)$ are both positive, and there are NO imaginary parts, Y(t) will exponentially decay (overdamped).
- If $Re(\tau_1)$ and $Re(\tau_2)$ are both positive, and there ARE imaginary parts, Y(t) will oscillate until it reaches steady state (underdamped).
- If $Re(\tau_1)$ and $Re(\tau_2)$ are both negative, and there are NO imaginary parts, Y(t) will exponentially grow (unstable).

- If $\text{Re}(\tau_1)$ and $\text{Re}(\tau_2)$ are both negative, and there ARE imaginary parts, $Y(t)$ will oscillate and grow exponentialy (unstable).
- If $\text{Re}(\tau_1)$ and $\text{Re}(\tau_2)$ are both zero, and there ARE imaginary parts, $Y(t)$ will oscillate and neither grow nor decay.
- If τ_1 and τ_2 are both zero, $Y(t)$ is equal to $X(t)$.
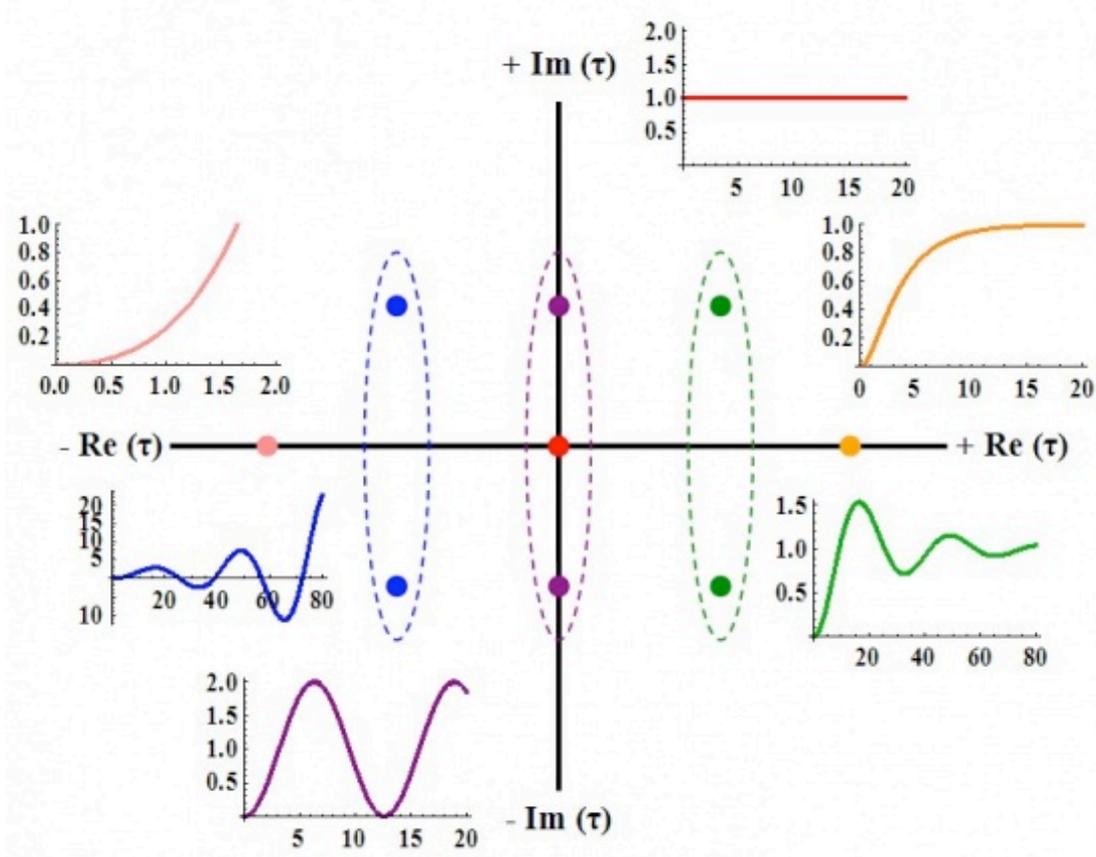
3.2 Solution of the General Second-Order System (When X(t)= ϑ(t))

The solution for the output of the system, $Y(t)$, can be found in the following section, if we assume that the input, $X(t)$, is a step function $\theta(t)$. The solution will depend on the value of ζ. If ζ is less than one, $Y(t)$ will be underdamped. This means that the output will overshoot and oscillate. If ζ is equal to one, $Y(t)$ will be critically damped. This means that the output will reach the steady state value quickly, without overshoot or oscillation. If ζ is greater than one, $Y(t)$ will be overdamped. This means that the output will not reach the steady state value as quickly as a critically damped system, but there will be no overshoot or oscillation.

612

3.2.1 Underdamped (ζ<1)

If $\zeta < 1$, the solution is:

$$Y(t) = 1 - \frac{1}{\sqrt{1-\zeta^2}} e^{-\zeta t/\tau} sin(\sqrt{1-\zeta^2}\frac{t}{\tau} + \phi)$$

Where:

$$\phi = -tan^{-1}(\frac{\sqrt{1-\zeta^2}}{\zeta})$$
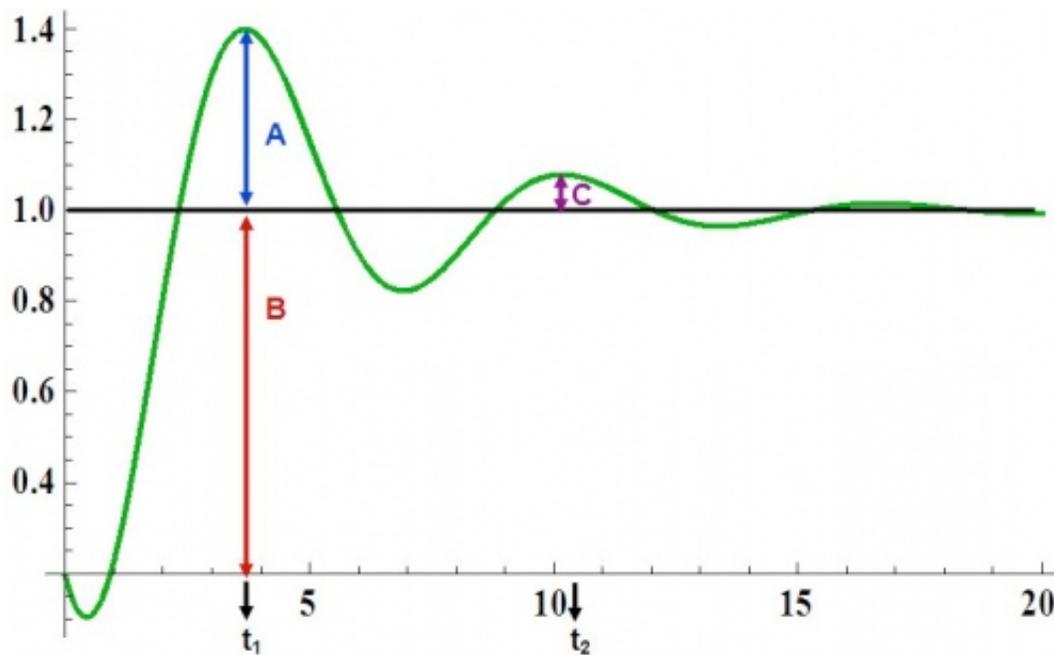
The decay ratio (C/A) can be calculated using the following equation:

$$\frac{C}{A} = e^{-2\pi\zeta/\sqrt{1-\zeta^2}}$$

The overshoot (A/B) can be calculated using the following equation:

$$\frac{A}{B} = e^{-\pi\zeta/\sqrt{1-\zeta^2}}$$

The period (T) and the frequency (ω) are the following:

$$T = t_2 - t_1 = \frac{2\pi\tau}{\sqrt{1-\zeta^2}}$$

$$\omega = \frac{2\pi}{T} = \frac{\sqrt{1-\zeta^2}}{\tau}$$

3.2.2 Critically Damped (ζ=1)

If $\zeta = 1$, the solution is:

$$Y(t) = 1 - (1 + \frac{t}{\tau})e^{-t/\tau}$$
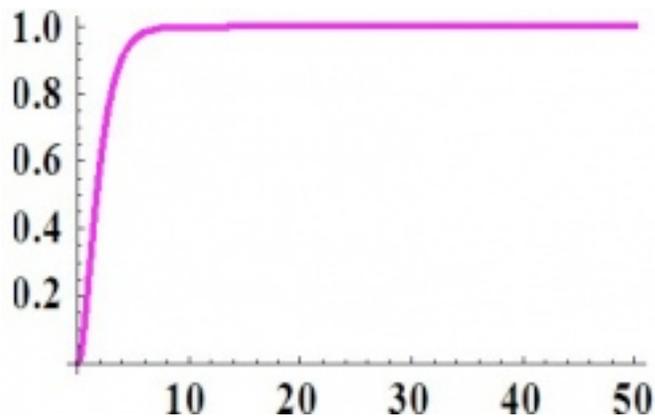
3.2.3 Overdamped (ζ>1)

If $\zeta > 1$, the solution is:

$$Y(t) = 1 - \frac{1}{\sqrt{\zeta^2 - 1}} e^{-\zeta t/\tau} sinh(\sqrt{\zeta^2 - 1}\frac{t}{\tau} + \phi)$$

Where:

614

$$\phi = -tanh^{-1}(\frac{\sqrt{\zeta^2 - 1}}{\zeta})$$
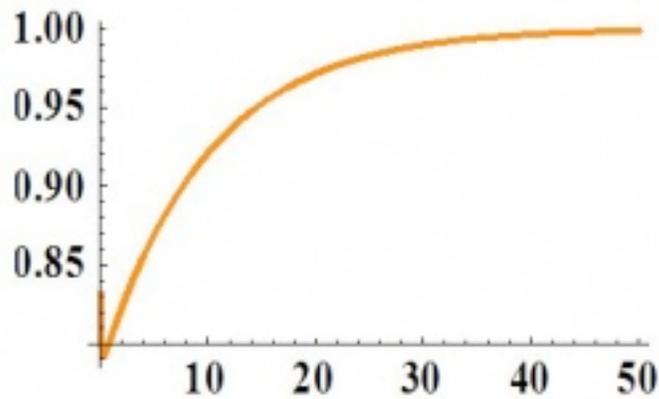
3.3 Example

3.3.1 Problem:

Given:

$A_1 = 1 \ m^2$

$A_2 = 1.5 \ m^2$

$R_1 = 0.25 \ s/m^2$

$R_2 = 0.75 \ s/m^2$

where:

A is the area of the tank

Q is the volumetric flowrate

R is the resistance to the flow of the stream exiting the tank

H is the height of liquid in the tank

develop an expression describing the response of H_2 to Q_{in}. Determine if the system is over, under or critically damped and determine what the graph of the expression would look like using the complex τ plane above. A diagram of the system is shown below:
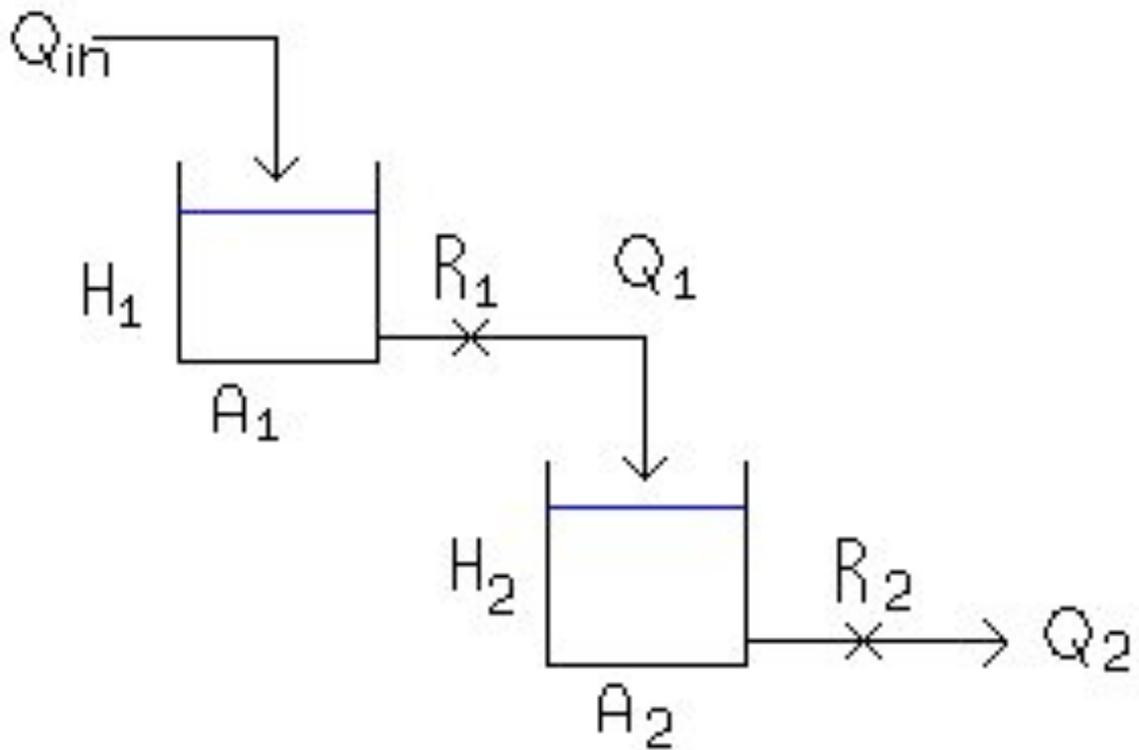
3.3.2 Solution:

Performing a mass balance on each tank:

$$A_1 \frac{dH_1}{dt} = Q_{in} - \frac{H_1}{R_1}$$

(1)

$$A_2 \frac{dH_2}{dt} = \frac{H_1}{R_1} - \frac{H_2}{R_2}$$

(2)

where the left hand terms account for the accumulation in the tank and the right hand terms account for the flow in the entering and exiting streams

Let $\tau_1 = R_1 A_1$ and $\tau_2 = R_2 A_2$

616

Equations (1) and (2) now become

$$\tau_1 \frac{dH_1}{dt} = R_1 Q_{in} - H_1$$

(3)

$$\tau_2 \frac{dH_2}{dt} = \frac{R_2}{R_1} H_1 - H_2$$

(4)

Put like terms on the same side and factor

$$(\tau_1 \frac{d}{dt} + 1)H_1 = R_1 Q_{in}$$

(5)

$$(\tau_2 \frac{d}{dt} + 1)H_2 = \frac{R_2}{R_1} H_1$$

(6)

Apply $(\tau_1 \frac{d}{dt} + 1)$ operator from equation (5) to equation (6)

$$(\tau_1 \frac{d}{dt} + 1)(\tau_2 \frac{d}{dt} + 1)H_2 = (\tau_1 \frac{d}{dt} + 1)\frac{R_2}{R_1} H_1$$

(7)

The $(\tau_1 \frac{d}{dt} + 1)H_1$ term from the left hand portion of equation (5) can be substituted into the right hand side of equation (7)

$$(\tau_1 \frac{d}{dt} + 1)(\tau_2 \frac{d}{dt} + 1)H_2 = R_1 Q_{in} \frac{R_2}{R_1}$$

$$(\tau_1 \frac{d}{dt} + 1)(\tau_2 \frac{d}{dt} + 1)H_2 = Q_{in} R_2$$

This expression shows the response of H_2 to Q_{in} as a second order solution like those pictured above. Here $Y(t)=H_2$ and $X(t)=R_2 Q_{in}$

$$\zeta = \frac{\tau_1 + \tau_2}{2\sqrt{\tau_1 \tau_2}} = \frac{(0.25 * 1) + (0.75 * 1.5)}{2\sqrt{(0.25 * 1)(0.75 * 1.5)}} = 1.296, Overdamped$$

Both values of τ are positive real numbers, and the behavior of the graph of the equation can be found on the complex τ plane above.

3.4 Analogy to Physics: Spring System

It might be helpful to use a spring system as an analogy for our second order systems.
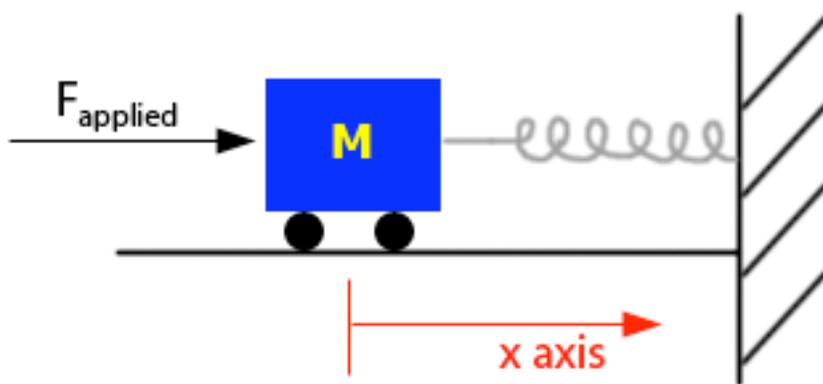
From Newton's second law of motion,

$F = ma$

> Where:
> F is Force
> m is mass
> a is acceleration

For the spring system, this equation can be written as:

$F_{applied} - F_{friction} - F_{restoring} = mx''$

> Where:
> x'' is the acceleration of the car in the x-direction

$F_{applied} - fx' - kx = mx''$

618

Where:

k is the spring constant, which relates displacement of the object to the force applied

f is the frequency of oscillation

$$\frac{m}{k}x'' + \frac{f}{k}x' + x = F_{applied}$$

As you can see, this equation resembles the form of a second order equation. The equation can be then thought of as:

$$\tau^2 x'' + 2\zeta\tau x' + x = F_{applied}$$

$$\tau = \sqrt{\frac{m}{k}}$$

$$\zeta = \frac{f}{2\sqrt{mk}}$$

Because of this, the spring exhibits behavior like second order differential equations:

If $\zeta > 1$ or $f > 2\sqrt{mk}$ it is overdamped

If $\zeta = 1$ or $f = 2\sqrt{mk}$ it is critically damped

If $\zeta < 1$ or $f < 2\sqrt{mk}$ it is underdamped

Section 4. Taylor Series

A **Taylor Series** is a representation of a function in the form of an infinite sum. Each term is calculated from using a derivative of the function as well as a factorial. The following equation is the defining equation of a Taylor series:

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n$$

Where a is the point, around which the function is defined.

The Taylor approximation is most accurate at a, and becomes less precise as the distance from a increases.

The following equation is the same as the equation mentioned above, in summation notation:

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

We will first consider a simple example, the function, e^x, defined around the point, 0:

$$e^x = e^0 + \frac{1}{1!}(x-0) + \frac{1}{2!}(x-0)^2 + \frac{1}{3!}(x-0)^3 + \cdots + \frac{1}{n!}(x-0)^n$$

The numerators of each equation are all 1, becasue any level derivative of e^x is e^x, and e^x, defined at 0 is always 1. The Taylor expansion simplifies to the following equation:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!}$$

Also note that a Taylor series centered around 0 is called a Maclaurin series. In addition to the e^x Maclaurin series, some other common, simplified Maclaurin series are listed here:

$$cos(x) = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \frac{1}{6!}x^6 + \cdots$$

620

$$\sin(x) = \frac{1}{1!}x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7 + \cdots$$

$$\cosh(x) = 1 + \frac{1}{2!}x^2 + \frac{1}{4!}x^4 + \frac{1}{6!}x^6 + \cdots$$

$$\sinh(x) = \frac{1}{1!}x + \frac{1}{3!}x^3 + \frac{1}{5!}x^5 + \frac{1}{7!}x^7 + \cdots$$

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots + x^n$$

$$\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \cdots + \frac{(-1)^{n+1}}{n}x^n$$

These Maclaurin series can be modified to become Taylor series when the equation is not centered around zero.

Expansion of functions using Taylor series may be useful in derivations. Maclaurin series will be useful for when the function is being approximated for small values of x.

Using an infinite number of terms is generally unnecessary to model the function around the center point. The first few terms of either a Taylor or Maclaurin series usually sufficiently approximate the function.

A stimulating discussion of Taylor series may be found in Comtet's "Calcul pratique des coefficients de Taylor d'une fonction algébrique" (Enseign. Math. 10, 267-270, 1964) as well as Whittaker and Watson's landmark treatise, "Forms of the Remainder in Taylor's Series." found in A Course in Modern Analysis, 4th ed.

Source: Wolfram MathWorld: http://mathworld.wolfram.com/TaylorSeries.html.

Section 5. Laplace Transforms

Written By: Paul Rhee, Eric Chuang, and David Mui

5.1 Introduction

Laplace Transform are frequently used in determining solutions of a wide class of partial diffferential equations. Like other transforms, Laplace transforms are used to determine particular solutions. In solving partial differential equations, the general solutions are difficult, if not impossible, to obtain. Thus, the transform techniques sometimes offer a useful tool for attaining particular solutions. The Laplace transform is closely related to the complex Fourier transform, so the Fourier integral formula can be used to define the Laplace transform and its inverse[3]. Integral transforms are one of many tools that are very useful for solving linear differential equations[1]. An integral transform is a relation of the form:

$$F(s) = \int_a^b K(s,t)f(t)\,dt$$

$K(s,t)$ = The kernel of the transformation

The relation given above transforms the function f into another function F, which is called the transform of f. The general idea in using an integral transform to solve a differential equation is as follows:

1) Use the relation to transform a problem for an unknown function f into a simpler problem for F
2) Solve this simpler problem to find F
3) Recover the desired function f from its transform F

Several integral transforms are known to be useful in applied mathematics, but in this chapter we consider only the Laplace transform. This transform is defined in the following way Let $f(t)$ be given for $t \leq 0$. Then the Laplace transform of f, which we will denote by $\mathcal{L}\{f(t)\}$, is defined by the equation:

$$\mathcal{L}\{f(t)\} = \int_0^\infty e^{-st}f(t)\,dt.$$

Whenever this improper integral converges. The Laplace transform makes use of the kernel $K(s,t) = e^-(st)$. Since the solutions of linear differential equations with constant coefficients are bsed on teh exponential function, the Laplace transform is particularly useful for such equations.

622

Signal	Laplace Transform
1. $\delta^{(n)}(t)$	s^n
2. 1 or $u(t)$	$\dfrac{1}{s}$
3. $\dfrac{t^n \exp(-\alpha t)u(t)}{n!}$	$\dfrac{1}{(s + \alpha)^{n+1}}$
4. $\cos \omega_0 t\, u(t)$	$\dfrac{s}{s^2 + \omega_0^2}$
5. $\sin \omega_0 t\, u(t)$	$\dfrac{\omega_0}{s^2 + \omega_0^2}$
6. $\exp(-\alpha t)\cos \omega_0 t\, u(t)$	$\dfrac{s + \alpha}{(s + \alpha)^2 + \omega_0^2}$
7. $\exp(-\alpha t)\sin \omega_0 t\, u(t)$	$\dfrac{\omega_0}{(s + \alpha)^2 + \omega_0^2}$
8. Square wave: $u(t) - 2u\left(t - \dfrac{T_0}{2}\right) + 2u(t - T_0) - \cdots$	$\dfrac{1}{s}\dfrac{1 - e^{-sT_0/2}}{1 + e^{-sT_0/2}}$
9. $(\sin \omega_0 t - \omega_0 t \cos \omega_0 t)u(t)$	$\dfrac{2\omega_0^3}{(s^2 + \omega_0^2)^2}$
10. $(\omega_0 t \sin \omega_0 t)u(t)$	$\dfrac{2\omega_0^2 s}{(s^2 + \omega_0^2)^2}$
11. $\omega_0 t \exp(-\alpha t)\sin \omega_0 t\, u(t)$	$\dfrac{2\omega_0^2(s + \alpha)}{[(s + \alpha)^2 + \omega_0^2]^2}$
12. $\exp(-\alpha t)(\sin \omega_0 t - \omega_0 t \cos \omega_0 t)u(t)$	$\dfrac{2\omega_0^3}{[(s + \alpha)^2 + \omega_0^2]^2}$

5.2 Example Problems

5.2.1 Elementary Functions Using Laplace Transforms

Example 1. Given $f(c) = c$, c is a constant

$$\mathcal{L}[c] = \int_0^\infty e^{-st} c \, dt$$

$$= \left[-\frac{ce^{-st}}{s} \right]_0^\infty$$

$$= \frac{c}{s}$$

Example 2. Given $f(c) = e^{at}$, a is a constant

$$\mathcal{L}[e^{at}] = \int_0^\infty e^{-st} e^{at} \, dt$$

$$= \left[-\frac{e^{-(s-a)t}}{(s-a)} \right]_0^\infty$$

$$= \frac{1}{s-a}, \text{ where } s \geq a$$

Example 3. Given $f(t) = t^2$, Then

$$\mathcal{L}[t^2] = \int_0^\infty e^{-st} t^2 \, dt$$

Integration by parts yields

$$= \left[-\frac{t^2 e^{-st}}{s} \right]_0^\infty + \int_0^\infty \frac{e^{-st}}{s} 2t \, dt$$

Since $t^2 e^{-st} \to 0$ as $t \to \infty$, we have

624

$$\mathcal{L}[t^2] = \frac{2}{s} \left[-\frac{e^{-st}}{s} t \right]_0^\infty + \frac{2}{s} \int_0^\infty \frac{e^{-st}}{s} \, dt$$

$$= \frac{2}{s^3}$$

Example 4. Given $f(t) = \sin \omega t$. Then

$$F(s) = \mathcal{L}[\sin \omega t] = \int_0^\infty e^{-st} \sin \omega t \, dt$$

$$= \left[-\frac{e^{-st}}{s} \sin \omega t \right]_0^\infty + \int_0^\infty \frac{e^{-st}}{s} \omega \cos \omega t \, dt$$

$$= \frac{\omega}{s} \left[-\frac{e^{-st}}{s} \cos \omega t \right]_0^\infty - \frac{\omega}{s} \int_0^\infty \frac{e^{-st}}{s} \omega \sin \omega t \, dt$$

$$F(s) = \frac{\omega}{s} - \frac{\omega^2}{s^2} F(s)$$

Thus, solving for $F(s)$, we obtain

$$\mathcal{L}[\sin \omega t] = \frac{\omega}{(s^2 + \omega^2)}$$

5.2.2 In Class Example Using Laplace Transforms

Solving a differential equation:

$$\frac{dy(t)}{dt} = -ay(t)$$

Rearranging the equation to one side, we have:

$$\frac{dy(t)}{dt} + ay(t) = 0$$

Next, we take the Laplace transform

$$(sy(s) - y_o) + ay(s) = 0$$

where $y(s) = L[y(t)]$ and $y_o = y(0)$ Solving for y(s), we find:

$$y(s) = \frac{y_o}{s + a}$$

$$y(t) = \mathcal{L}^{-1}\{y(s)\} = \mathcal{L}^{-1}\left\{\frac{y_o}{s + a}\right\} = y_o e^{-at}$$

5.3 References

[1] William E. Boyce, *Elementary Differential Equations and Boundary Value Problems* (2005) Chapter 6 pp 307-440.

[2] Dr. Ali Muqaibel, *EE207 Signals & Systems* (2009) <http://tinyurl.com/ycq46qn>

[3] Tyn Myint-U, *Partial Differential Equations for Scientists and Engineers* (2005) pp 337 -341.

Chapter 8. Optimization

Section 1. Optimization

Written by: Brittany Flaherty, Christine Curran, and Lauren Pakan

1.1 Definition

Optimization aims to obtain the best results in a given situation, or to minimize input to maximize benefit [1].

1.2 Engineering Application for Optimization

The following are examples of optimization across all engineering disciplines [1].

- Design of aircraft and aerospace structures for minimum weight
- Finding the optimal trajectories of space vehicles
- Design of civil engineering structures such as frames, foundations, bridges, towers, chimneys, and dams for minimum cost
- Minimum-weight design of structures for earthquake, wind, and other types of random loading
- Design of water resources systems for maximum benefit
- Optimal plastic design of structures
- Optimum design of linkages, cams, gears, machine tools, and other
- mechanical components
- Selection of machining conditions in metal-cutting processes for minimum production cost
- Design of material handling equipment such as conveyors, trucks, and cranes for minimum cost
- Design of pumps, turbines, and heat transfer equipment for maximum efficiency
- Optimum design of electrical machinery such as motors, generators,and transformers
- Optimum design of electrical networks
- Shortest route taken by a salesperson visiting various cities during one tour
- Optimal production planning, controlling, and scheduling
- Analysis of statistical data and building empirical models from experimental
- results to obtain the most accurate representation of the physical phenomenon
- Optimum design of chemical processing equipment and plants
- Design of optimum pipeline networks for process industries
- Selection of a site for an industry

- Planning of maintenance and replacement of equipment to reduce operating costs
- Inventory control
- Allocation of resources or services among several activities to maximize the benefit
- Controlling the waiting and idle times and queueing in production lines to reduce the costs
- Planning the best strategy to obtain maximum profit in the presence of a competitor
- Optimum design of control systems

1.3 Optimization across an Organization

The highest level of optimization includes the optimization of the raw material supply chain and the optimization of the packaging and product distribution chain. The plant wide optimization must consider documentation, maintenance, scheduling, and quality management considerations. Plant wide optimization resolves the conflict of objectives between the unit operations and the envelope strategies required to optimize the entire plant. Within the unit operations level multivariable optimization cannot be achieved when individual processing equipment is defective or when the control loops are not properly tuned. It is important that measurements be sampled fast enough, that controls loops be tuned for fast rates of recovery, and loop cycling be eliminated. When no mathematical model can describe a process, the process can only be optimized experimentally and empirical optimization is required [2].
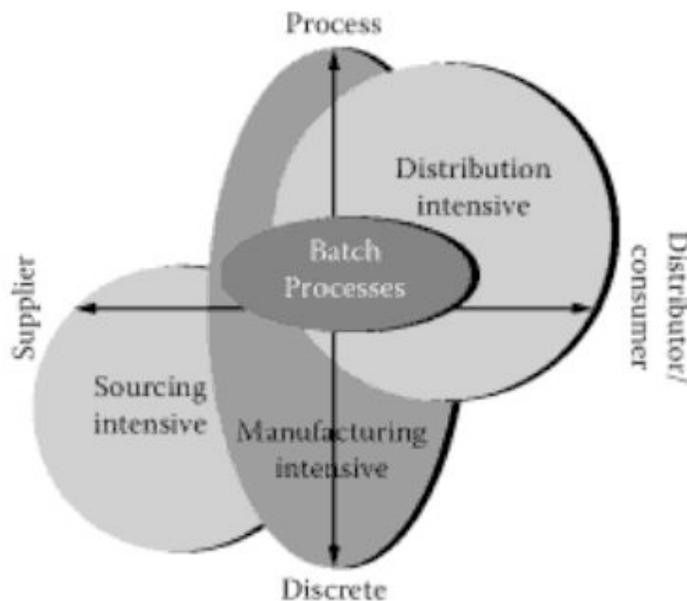
1.4 Optimization Design
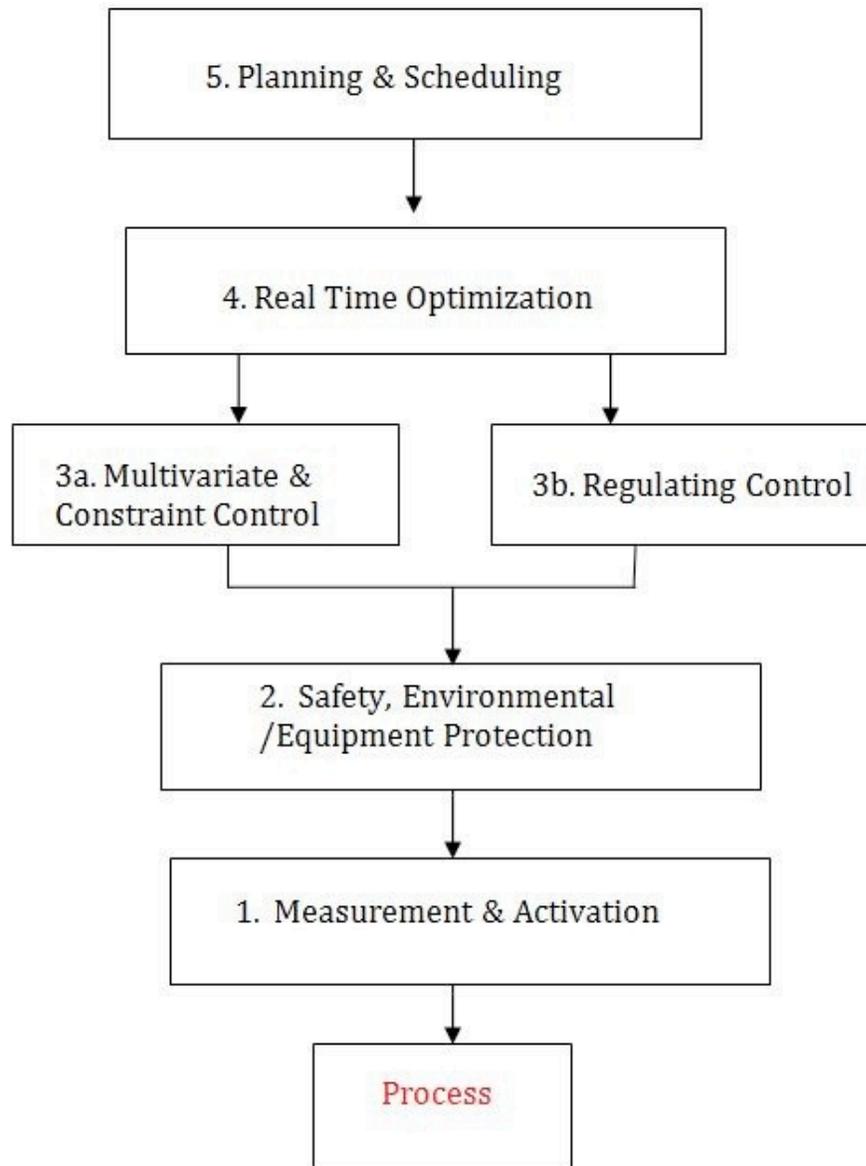
1.5 Constraints

Design constraints are physical limitations or restrictions that must be satisfied to produce an acceptable design [1].

- Operation conditions – safety, environmental
- Equipment constraints – e.g. pump rates
- Storage capacities
- Product quality and impurities

629

1.6 Optimization Situations

The following list describes common reasons for optimization in an industrial plant.

- Sales limited by production (e.g. reduce costs by minimizing downtime)
- Sales limited by market (e.g. be the "low cost producer")
- Plants with high throughput
- High raw material or energy consumption
- Product quality better than specifications
- Loss of valuable or hazardous components through waste stream

1.7 Real-Time Optimization

The following describes the steps in order to optimize a chemical engineering process.

1. Identify process variables
2. Select objective function (e.g. profit $$$)
3. Develop process model and constraints
4. Simplify model to objective function (e.g linearization)
5. Compute the optimum
6. Perform sensitivity study

1.8 Industry Experience

Optimization can be applied to every aspect of a process. For example, at a refinery there are operators that work out in the units. Optimization can be applied to increase the operator/engineer communication by implementing "real-time" computer programs that allow the process engineers to see what is actually happening in the plant. This will help optimize process by allowing engineers to see what the conditions in the plant are in real time.

1.9 Pilot Plant Experience

Optimization of processes in a pilot plant will allow for more efficient scale-up to commercial size. In the ChE 460 course, four unit operations are optimized to produce soybean biodiesel. In order to optimize the reaction conversion, the project engineers vary catalyst concentration, agitation rate, and temperature. A design of experiment (DOE) is used to find the best set of input parameters.

1.10 References

1. Rao, Singiresu S. *Engineering Optimization - Theory and Practice (3rd Edition)*. (pp: 5). John Wiley & Sons. Online version available at: http://knovel.com/web/portal/browse/display?_EXT_KNOVEL_DISPLAY_bookid=1100&VerticalID=0
2. Lipták, Béla G. *Instrument Engineers' Handbook: Process control and optimization.* Boca Raten, FL: CRC Press, 2006.

Section 2. Linear Optimization

written by: Danny Hsiao, Jenny Ou, and Huey Shann Sue

2.1 Introduction

Linear optimization is a method applicable for the solution of problems in which the objective function and the constraints appear as linear functions of the decision variables. The constraint equations may be in the form of equalities or inequalities[1]. In other words, linear optimization determines the way to achieve the best outcome (for example, to maximize profit or to minimize cost) in a given mathematical model and given some lists of requirements represented as linear equations [2].

2.2 Applications

Linear optimization can be applied to numerous fields, in business or economics situations, and also in solving engineering problems. It is useful in modeling diverse types of problems in planning, routing, scheduling, assignment and design [2].

2.2.1 Some examples of applications in different industries

Petroleum refineries

One of the early industrial applications of linear optimization has been made in the petroleum refineries. An oil refinery has a choice of buying crude oil from different sources with different compositions at different prices. It can manufacture different products, such as diesel fuel, gasoline and aviation fuel, in varying quantities. A mix of the purchased crude oil and the manufactured products is sought that gives the maximum profit.

Manufacturing firms

The sales of a firm often fluctuate, therefore a company has various options. It can either build up an inventory of the manufactured products to carry it through the period of peak sales, or to pay overtime rates to achieve higher production during periods of high demand. Linear optimization takes into account the various cost and loss factors and arrive at the most profitable production plan.

Food-processing industry

Linear optimization has been used to determine the optimal shipping plan for the distribution of a particular product from different manufacturing plants to various warehouses.

Telecommunications

The optimal routing of messages in a communication network and the routing of aircraft and ships can also be determined by linear optimization method.

2.3 Characteristics of Linear Optimization

The characteristics of a linear optimization problem are:

1. *The objective function is of the minimization type*

2. *All the constraints are of the equality type*

3. *All the decision variables are non-negative*

Any linear optimization problem can be expressed in the standard form by using the following transformation:

1) The maximization of a function $f(x_1, x_2, ..., x_n)$ is equivalent to the minimization of the negative of the same function.

For example:

Minimize $f = c_1 x_1 + c_2 x_2 + ... + c_n x_n$

is equivalent to

Maximize $f' = -f = -c_1 x_1 - c_2 x_2 - ... - c_n x_n$

Consequently, the objective function can be stated in the minimization form in any linear optimization problem.

2) If a constraint appears in the form of a "less than or equal to" type of inequality as

$$a_{k1} x_1 + a_{k2} x_2 + ... + a_{kn} x_n \leq b_k$$

it can be converted into the equality form by **adding** a non-negative *slack variable* x_{n+1} as follows:

632

$$a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kn}x_n + x_{n+1} = b_k$$

Similarly, if the constraint is in the form of a "greater than or equal to" type of inequality, it can be converted into the equality form by **subtracting** the *surplus variable* x_{n+1}.

3) In most engineering optimization problems, the decision variables represent some physical dimensions, hence the variables x_j will be non-negative.

However, a variable may be unrestricted in sign in some problems. In such cases, an unrestricted variable (which can take a positive, negative or zero value) can be written as the difference of two non-negative variables.

Thus, if x_j is unrestricted in sign, it can be written as $x_j = x_j' - x_j''$, where $0 \le x_j'$ and $0 \le x_j''$

It can be seen that x_j will be negative, zero or positive, depending on whether x_j'' is greater than, equal to, or less than x_j'

2.4 Example 1

This example comes from Seborg, Edgar, and Mellinchamp (with some corrections).

A chemical company is mixing chemical component, A and B to produce product, E and F. The chemical reaction is listed below:

$$A + B = E$$
$$A + 2B = F$$

The conditions of this production is listed as below:

	Cost of Materials	Maximum Available per day	Fixed Cost
A	$0.15/lbs	40,000lbs	350

	Value of Products	Maximum Production Rate	
B	$0.20/lbs	30,000lbs	200
E	$0.40/lbs	30,000lbs	
F	$0.33/lbs	30,000lbs	

The **profit** of this production can be simply described as the function below:

$$\text{Profit} = \sum_s F_s V_s - \sum_r F_r V_r - C.P. - F.C.$$

$$F_s = \text{Flow Rates of Products}$$

$$F_r = \text{Flow Rates of Raw Materials}$$

$$V_s = \text{Values of Products}$$

$$V_r = \text{Values of Raw Materials}$$

$$C.P. = \text{Cost of Production}$$

$$F.C. = \text{Fixed Cost}$$

2.4.1 Constraints

- $0 < F_A < 40,000$
- $0 < F_B < 30,000$
- $0 < F_E < 30,000$
- $0 < F_F < 30,000$
- $\frac{1}{3} F_F + \frac{2}{3} F_E \geq 0$

634

2.4.2 Solution

$$F_A = \frac{1}{2}F_E + \frac{1}{3}F_F$$

$$F_B = \frac{1}{2}F_E + \frac{2}{3}F_F$$

$$Profit = \sum_s F_s V_s - \sum_r F_r V_r - C.P. - F.C.$$

$$= (0.4F_E + 0.33F_F) - (0.15F_A + 0.2F_B) - (0.15F_E + 0.05F_F) - (350 + 200)$$

$$F_A > 0$$

$$\frac{1}{2}F_E + \frac{1}{3}F_F > 0$$

$$F_F > -\frac{3}{2}F_E \dots\dots\dots(1)$$

$$F_A < 40,000$$

$$\frac{1}{2}F_E + \frac{1}{3}F_F < 40,000$$

$$F_F < 120,000 - \frac{3}{2}F_E \dots\dots\dots(2)$$

$$F_B > 0$$

$$\frac{1}{2}F_E + \frac{2}{3}F_F > 0$$

$$F_F > -\frac{3}{4}F_E\ldots\ldots\ldots(3)$$

$$F_A < 30,000$$
$$\frac{1}{2}F_E + \frac{2}{3}F_F < 30,000$$
$$F_F < 45,000 - \frac{3}{4}F_E\ldots\ldots\ldots(4)$$

Solution using Mathematica

INPUT:

profit = 0.25 FE + 0.28 FF - 0.15 FA - 0.2 FB - 350 - 200

sol = Maximize[profit, {FA > 0, FB > 0, FA < 40000, FB < 30000, FE > 0, FE < 30000, FF > 0, FF < 30000, FA == (1/2) FE + (1/3) FF, FB == (1/2) FE + (2/3) FF}, {FE, FF}]

OUTPUT: {3875., {FE -> 30000., FF -> 22500.}}

Solution using "Solver" in Excel.

The above shows an Excel spreadsheet "466optimixation.xls" with the following content:

Row 2: Using Solver in Excel to solve optimization problem. Note yellow is variables to find.

	A	B	C	D	E	F	G
4	stream	A	B	E	F		
6	flow rates	22500	30000	30000	22500		
7	min	0	0	0	0		
8	max	40000	30000	30000	30000		
10	coefficients	-0.15	-0.2	0.25	0.28		
11	fixed costs			350	200		
12	profit	3875					

Formula for profit: SUMPRODUCT(B6:E6,B10:E10)-D11-E11
Formula for FA: =D6/2+E6/3
Formula for FB: =D6/2+2*E6/3

Solver Parameters

Set Target Cell: B12

Equal To: ● Max ○ Min ○ Value of: 0

By Changing Cells: D6,E6

Subject to the Constraints:
B6 <= B8
B6 >= B7
C6 <= C8
C6 >= C7
D6 <= D8
D6 >= D7

Buttons: Solve, Close, Options..., Reset All, Help, Guess, Add, Change, Delete

Result is FE = 30,000, FF = 22500.

If only process ! or process 2 were running a full capacity, the profit would be less.

2.4.3 Linear Optimization

The above is an example of linear optimization. It is often used in oil refinery to figure out maximal profit in response to market competition.

2.5 Example 2

Example of Linear Optimization Problem in Excel

Written by: Jennifer Campbell, Katherine Koterba, MaryAnn Winsemius

2.5.1 Part 1: Organize Given Information

As stated in the Linear Optimization section example above, there are three categories of information needed for solving an optimization problem in Excel: **an objective function, constraints, and decision variables**.

We will use the following example to demonstrate another application of linear optimization. We will be optimizing the profit for Company X's trucking business.

To reach capacity, Company X must move 100 tons of cargo per day by truck. Company X's trucking fee is $250/ton. Besides the weight constraint, the company can only move 50,000 ft^3 of cargo per day due to limited volume trucking capacity. The following amounts of cargo are available for shipping each day:

Cargo Weight (tons)	Weight (tons)	Volume (ft³/ton)
1	30	550
2	40	800
3	50	400

Maximize the profit for Company X. Set up this problem:

Objective Function ($/week):

Profit = 250*(Cargo1 + Cargo2+ Cargo3)

Decision Variables (weight in tons):

Cargo 1

Cargo 2

Cargo 3

Constraints:

Weight: Cargo1 + Cargo2 + Cargo3 ≤ 100

Volume: 550*Cargo1 + 800*Cargo2 +400*Cargo3 ≤ 50000

Amount 1:   Cargo1 ≤ 30

                Cargo1 ≥ 0

Amount 2:   Cargo2 ≤ 40

                Cargo2 ≥ 0

Amount 3:   Cargo3 ≤ 50

                Cargo3 ≥ 0

2.5.2 Part 2: Set Up the Problem Using Excel

Solver is an Add-in for Microsoft Excel. It will be used to optimize Company X's profit. If 'Solver' is not on the 'Tools' menu in Excel, then use the following steps to enable it:

For Windows 2007:

- Click on the Office button at the top left corner of the screen. Click on the "Excel Options" button on the bottom right of the menu.
- Select "Add-ins." Make sure that "Excel Add-ins" is selected in the "Manage" drop down list. Click "Go."
- A new window will appear entitled "Add-ins." Select "Solver Add-in" by checking the box. Click "Go."
- A Configuration window will appear. Allow Office to install the Add-in.
- The solver has been successfully installed. (See Windows Help for more instruction.)

Use the figure below to set up your Excel worksheet.

	A	B	C	D
1				
2				
3	Objective			
4	Profit	0		
5				
6	Variables			
7				
8				
9	Cargo1	0		
10	Cargo2	0		
11	Cargo3	0		
12				
13				
14	Constraints	Formula	Min	Max
15	Weight	0	0	100
16	Volume	0	0	50000
17	Amount1	0	0	30
18	Amount2	0	0	40
19	Amount3	0	0	50

Enter in the following formulas to the cells as shown below:

2.5.3 Part 3: Running Solver

- Click on the "Data" tab and select "Solver". A dialog box will appear.

- Enter the parameters as shown in the figure below.

Detailed steps are as follows:

- In "Set Target Cell," enter the cell corresponding to the company's profit (B4).
- Select "Max" under "Equal To."

- Click on the "Options" tab and check the "Assume Linear Model" box.
- For "By Changing Cells:" select the cells in column B corresponding the cargo amounts (B9, B10, B11).
  - To add constraints, select "Add" under "Subject to the Constraints" A dialog box will open.
  - In the "Cell Reference:" field, enter the cell location of the decision value that is subject to constraint (i.e. B9).
  - Use the pull-down menu in the middle to select the appropriate inequality relation (i.e. <=)
  - In the "Constraint:" field, enter the cell location of the constraint value (i.e. D17).
  - Continue to click the "Add."
  - Repeat above steps until all of the constraints are entered. Then click "OK."
- When all the proper settings have been entered, click "Solve."
- A "Solver Results" box will appear. Select "Keep Solver Solution" and click "OK.".

The solved worksheet is below.

	A	B	C	D
1				
2				
3	Objective			
4	Profit	24218.75		
5				
6	Variables			
7				
8				
9	Cargo1	30		
10	Cargo2	16.875		
11	Cargo3	50		
12				
13				
14	Constraints	Formula	Min	Max
15	Weight	96.875	0	100
16	Volume	50000	0	50000
17	Amount1	0	0	30
18	Amount2	0	0	40
19	Amount3	0	0	50

2.6 Sensitivity Report

Written by Michael Chisholm and Doug Sutherland, Dec. 2009

Excel's solver program allows us to analyze how our profit would change if we had an alteration in our constraint values. These values can change due to a variety of reasons such as more readily available resources, technology advancements, natural disasters limiting resources, etc.

First, it analyzes whether the constraints are binding or non-binding. The binding constraints limit the profit output where the non-binding constraints do not limit the overall process. If the non-binding constraints were changed, the profit would not be effected as long the change in these constraints lies within the allowable increase and decrease that is indicated within the sensitivity report. If the binding constraints are changed, the profit will be directly affected. The affect on the profit is shown with shadow price values, also displayed in the sensitivity report. The shadow price is the resulting increase or decrease in profit per unit increase or decrease in the constraint. This applies as long as the change in constraint remains within the allowable increase or decrease where a linear relationship can be assumed.

The shadow price only analyzes the change in one variable at a time. In order to do two, you must plug the new constraint value for one of the variables and solve using solver. Using the new sensitivity report, analyze the effect that changing the second variable would have with the change being made in the first constraint.

Looking at Example 1 above, we will now walk through the steps on how to create a sensitivity report.

After clicking "solve" in excel, a solver results dialogue box appears as seen below.

There is a list of three options on the right; answer, sensitivity, and limits. Select the sensitivity option before clicking ok. A new tab will be generated in the worksheet titled "sensitivity 1." A view of the sensitivity report within the tab is seen below. As you can see, two tables are generated. For this example, resource A and product F are non-binding as shown with a shadow price of 0 and an infinite allowable increase. The allowable decrease is the amount the capacity changes until the final value is reached. Past this point the constraint would become a binding constraint. For the constraining variables (resource B and product E), their constraints are binding. Regarding resource B, if its constraint was increased by up to 5000 or decreased by up to 15000, this would have a linear effect on profit within this range. For each unit increase or decrease, the profit will change by 12 cents per unit, respectively. The same is true if our capacity for product E changes with it's allowable values and shadow price on the table.

If the constraint on B increased by 5,000 lbs our new profit would be $8,600/day (8,000+.12*5,000). Instead, if our facility could increase production of E by 30,000 lb/day the resulting profit would be $12,950/day (8,000+.165*30,000).

2.7 Solving Linear Optimization Problems Using The Primal Simplex Algorithm

Written by: Tejas Kapadia and Dan Hassing

[Note: needs specific reference, and also solution to the preceding problem by this method would be good -- R.Z.]

Instead of solving linear optimization problems using graphical or computer methods, we

can also solve these problems using a process called the Primal Simplex Algorithm. The Primal Simplex Algorithm starts at a Basic Feasible Solution (BFS), which is a solution that lies on a vertex of the subspace contained by the constraints of the problem. In the Graph in Example 1, this subspace refers to the shaded region of the plot. Essentially, after determining an initial BFS, the Primal Simplex Algorithm moves through the boundaries from vertex to vertex until an optimal point is determined.

The basic procedure is the following:

1. Find a unit basis.
2. Set-up the problem in standard form using a canonical tableau.
3. Check optimality criterion.
    1. If criterion passes, then stop, solution has been found.
4. Select an entering variable among the eligible variables.
5. Perform pivot step.
6. Go back to 1.

For simplicity, we will make the following assumptions:

1. The optimum lies on a vertex and is not unbounded on an extreme half-line.
2. The constraints are equations and not also inequalities.
    1. In the case that the constraints are inequalities, slack variables will need to be introduced. Although the process is not very different in this case, we will ignore this to make the algorithm slightly less confusing.
3. Decision variables are required to be nonnegative.
4. The problem is a minimization problem. To turn a maximization problem into a minimization problem, multiply the objective function by -1 and follow the process to solve a minimization problem.

We will begin with the following example:

Objective Function: Minimize $z = -x_5 - 8x_6$

Subject to the constraints:

$$x_1 - x_5 + x_6 = 2$$

646

$$x_2 + x_5 + x_6 = 1$$

$$x_3 + 2x_5 + x_6 = 5$$

$$x_4 + x_6 = 0$$

$$x_i \geq 0$$

First we begin by finding a unit basis:

A shortcut method to finding this unit basis is putting numbers in for each variable so that every constraint equation is satisfied.

In this case, setting $x_4 = 0$, $x_5 = 0$, and $x_6 = 0$ will satisfy the final equation and also set the values for x_1, x_2, and x_3 to 2, 1, and 5, respectively. Remember, these decision variables must be nonnegative.

Set up the canonical tableau in the following form:

x_1	x_2	x_3	x_4	x_5	x_6	-z	b
1	0	0	0	-1	1	0	2
0	1	0	0	1	1	0	1
0	0	1	0	2	1	0	5
0	0	0	1	0	1	0	0
0	0	0	0	-1	-8	1	0

As you can see, the first four rows correspond to the constraints, while the final row corresponds to the objective function. The "b" column corresponds to the right hand side

(RHS) of the constraints. As you can see, the "-z" column is on the left hand side (LHS) of the equation, rather than the RHS.

First, we should perform pivot steps so that the tableau corresponds to the unit basis we found earlier. By performing pivot steps on x_1, x_2, x_3, and x_4, we will reach the feasible point where (x_1, x_2, x_3, x_4, x_5, and x_6) $= (2, 1, 5, 0, 0, 0)$. Because x_4, x_5, and x_6 all equal zero, the pivot step on x_4 can actually be done on x_5 or x_6, but in this example, we used x_4. These pivot steps can be performed on any row as long as they are all different rows. In this example, we performed pivot steps on $(x_1, 1)$, $(x_2, 2)$, $(x_3, 3)$, $(x_4, 4)$ using the Pivot and Gauss-Jordan Tool at http://people.hofstra.edu/Stefan_Waner/RealWorld/tutorialsf1/scriptpivot2.html. To use this tool, place the cursor on the cell that you wish to pivot on, and press "pivot".

After four pivot steps, the tableau will look like this:

x1	x2	x3	x4	x5	x6	x7	x8
1	0	0	0	-1	1	0	2
0	1	0	0	1	1	0	1
0	0	1	0	2	1	0	5
0	0	0	1	0	1	0	0
0	0	0	0	-1	-8	1	0

As you can see, this is identical to the initial tableau, as x_1, x_2, x_3, and x_4 were set up such that an initial feasible point was already chosen.

The optimality criterion states that if the vector in the bottom left of the tableau is all positive, then an optimal solution exists in the "b" column vector, with the value at the bottom of the "b" column vector as the negative of the value of the objective function at that optimal solution. If this is not true, then a pivot step must be performed. In this example, clearly, a pivot step must be performed.

Next, we need to choose an entering variable. We want to choose an entering variable that has a negative element in the bottom row, meaning that the objective value could be improved if that variable was nonzero in the solution. So, we will choose x_5 in this example. Now, we must calculate ratios of each RHS coefficient divided by the coefficient of the entering variable in that row. In this case, the vector corresponding to

648

this calculation would equal $(2/-1, 1/1, 5/2, 0/0)$. We cannot pivot on a zero element, so we cannot pivot on the fourth row. We want to keep the RHS positive, so we cannot pivot on the first row. We must choose the minimum nonnegative ratio to remain at a feasible solution, so we choose the second row in the x_5 column, which has a ratio of 1/1.

After the pivot step:

x1	x2	x3	x4	x5	x6	x7	x8
1	1	0	0	0	2	0	3
0	1	0	0	1	1	0	1
0	-2	1	0	0	-1	0	3
0	0	0	1	0	1	0	0
0	1	0	0	0	-7	1	1

As we can see, x_6 has a negative coefficient in the bottom row, indicating the same step must be repeated on that column. We calculate ratios for that column, and get: $(3/2, 1/1, 3/-1, 0/1)$. Consequently, we choose to pivot on the fourth row because it corresponds to the minimum nonnegative ratio of 0.

After another pivot step:

x1	x2	x3	x4	x5	x6	x7	x8
1	1	0	-2	0	0	0	3
0	1	0	-1	1	0	0	1
0	-2	1	1	0	0	0	3
0	0	0	1	0	1	0	0
0	1	0	7	0	0	1	1

Because the bottom row is all positive, we are now at an optimal solution. To understand this final tableau, we look at each column for variables that only have one "1" in the column. If the column has only one "1", the RHS value in that row is the value of that

variable. In this case, $x_1 = 3$, $x_3 = 3$, and $x_5 = 1$. Any variable that does not have just a single "1" in the column is equal to zero. So, the optimal solution is (x_1, x_2, x_3, x_4, x_5, and x_6) = $(3, 0, 3, 0, 1, 0)$, and the optimal value is $z = -1$ (z was on the LHS in the tableau).

Now, we have successfully solved a linear optimization problem using the Primal Simplex Algorithm. Verification of the solution can be easily performed in Microsoft Excel.

2.8 References

1. D. E. Seborg, T. F. Edgar, D. A. Mellichamp: Process Dynamics and Control, 2nd Edition, John Wiley & Sons.
2. Rao, Singiresu S. Engineering Optimization - Theory and Practice, 3rd Edition, 129-135, John Wiley & Sons.
3. www.wikipedia.org
4. http://www.math.ualberta.ca/~devries/math372-fall2004/excelSolverTutorial.pdf

Section 3. Non-linear Optimization

Written by: Matthew Goh, Andrew King, and Edwin Yik

3.1 Introduction

Various conditions and situations are not adequately described using linear systems. In this case, nonlinear optimization may be applied. Unlike linear optimization, the optimal operating condition does not exist at the boundaries.

3.2 Quadratic Optimization

$$f(x) = c - x^T b + \frac{1}{2} x^T A x$$

To optimize, it is necessary to find when the gradient of f is equal to zero.

$$\nabla f(x) = 0$$

$$\nabla f(x) = b - Ax$$

$$x_* = A^{-1} b$$

It may be possible to solve the optimal x_* by a linear equation, approximated by a Taylor series [1].

$$f(x_*) = f(x) + (x_* - x)^T \nabla f(x) + \frac{1}{2}(x_* - x)^T \nabla\nabla f(x)(x_* - x) + \dots$$

3.3 Iterative Methods

When direct methods cannot solve the equation (i.e. A is not symmetric positive definite), iterative methods are possible [1].

By starting with an initial guess of x_i , an algorithm may lead to a x_{i+1} that better satisfies the equation. Through iteration, theoretically, $x_\infty = x$.

3.4 Applications

- Finance: Portfolio optimization

- Businesses: Optimize inventory

- Engineering: Rigid body dynamics

- Biochemistry: Kinetic modeling [2]

3.5 Example: Typical Nonlinear 3d Curves

(Image from [1])

As observed, the optimal condition does not necessarily exist at the boundary of the curve.

3.6 Example: Quadratic Optimization

$$f(x) = \vec{c}^{\,T}\vec{x} + \frac{1}{2}\vec{x}^{\,T}Q\vec{x}$$

where

$$\vec{c}^{\,T} = (c_1, c_2, ..., c_n) \quad \vec{x}^{\,T} = (x_1, x_2, ..., x_n)$$

For a quadratic system, n=2, thus, Q (the quadratic term constant) is defined as a symmetric matrix as follows.

$$Q = \begin{bmatrix} Q_1 & Q_3 \\ Q_3 & Q_2 \end{bmatrix}$$

Thus, multiplying out the f ,

$$f(x) = (c_1 x_1 + c_2 x_2) + \frac{1}{2}(Q_1 x_1^2 + 2Q_3 x_1 x_2 + Q_2 x_2^2)$$

3.7 References

1. Lippert, Ross A. "Introduction to non-linear optimization." D.E. Shaw Research, February 25, 2008. http://www.mit.edu/~9.520/spring08/Classes/optlecture.pdf
2. Mendes, Pedro and Kell, Douglas B. "Non-linear optimization of biochemical pathways: application to metabolic engineering and parameter estimation." Journal of Bioinformatics, Volume 14, 869-883. 1998.
3. "Introduction to Non-linear optimization." Georgia Institute of Technology Systems Realization Laboratory. http://www.srl.gatech.edu/education/ME6103/NLP-intro.ppt

Chapter 9. PID Control

Section 1. Constructing Block Diagrams: Visualizing control measurements

It is often convenient to express a control system with a block diagram. These diagrams help to visualize the relationships between each part of the control system. Consider Figure 1 below.

Figure 1. Block diagram representing a control system.

The control system shown is called a closed-loop system or a feedback system because the measured value of the controlled variable is returned or "fed back" to the comparator. In the comparator the controlled variable is compared with the desired value or set point. If there is any difference between the measured variable and the set point, an error is generated. This error enters a controller, which in turns adjusts the final control element to return the controlled variable to the set point [1].

$R(t)$ = "set point", is a synonym for the desired value of the controlled variable
$\varepsilon(t)$ = process error or comparator
$X(t)$ = process variable
$M(t)$ = measurement variable
$Y(t)$ = output variable

The "hat" above the G indicates that it is an operator (not a simple function) and may include integration and differentiation. Thus, you cannot divide by a $\hat{G}$, but have to

multiply both sides by its inverse $\hat{G}^{-1}$ (see below). This is a more general representation that the standard Laplace transform form; here it is not necessary that the operators be linear.

From the block diagram the equations for a system can be constructed. Define the following:

$\hat{G}_c$ =Controller operator

$\hat{G}_p$ =Process operator

$\hat{G}_m$ =Measurement delay operator

The equations for the system in Figure 1 are given below. Note that these equations are always written as "output is equal to operator applied to input."

Process: $Y(t) = \hat{G}_p X(t)$

Controller: $X(t) = \hat{G}_c \epsilon(t)$

Comparator: $\varepsilon(t) = R(t) - M(t)$

Measurement delay: $M(t) = \hat{G}_m Y(t)$

Typically to plot the response of the controller, *Y(t)*, it is necessary to solve these equations to obtain a single equation for *Y(t)*. This can be done by first plugging the measurement delay equation into the compartor equation, or

$$\epsilon(t) = R(t) - \hat{G}_m Y(t).$$

This equation can be put into the controller equation, or

$$X(t) = \hat{G}_c [R(t) - G_m Y(t)].$$

Finally, the equation for *Y(t)* can be obtained by putting the above equation into the equation for the process, or

$$Y(t) = \hat{G}_p \hat{G}_c [R(t) - \hat{G}_m Y(t)].$$

Frequently we have a differential equation for *X(t)* in terms of derivatives of *Y(t)* so we

write

$$X(t) = \hat{G}_p^{-1} Y(t)$$

which implies the form of the general equation

$$\hat{G}_p^{-1} Y(t) = \hat{G}_c[R(t) - G_m Y(t)]$$

If the process, controller, and measurement delay operators are known, Y(t) can be plotted and the response of the system can be observed (i.e. using Mathematica).

That is, we have

$$Y(t) = \hat{G}_p X(t)$$
(1)

$$X(t) = \hat{G}_c \epsilon(t)$$
(2)

$$\epsilon(t) = R(t) - M(t)$$
(3)

$$M(t) = \hat{G}_m Y(t)$$
(4)

$$\hat{G}_p^{-1} Y(t) = \hat{G}_c[R(t) - G_m Y(t)]$$
(5)

Assume no measurement lag: $\hat{G}_m = 1$

$$\hat{G}_p^{-1} = \frac{1}{K_p}\left(\tau_p \frac{d}{dt} + 1\right)$$

Assume first-order system:

$$\hat{G}_c^{-1} = K_c + \frac{K_c}{\tau_I}\int_0^t dt$$

Assume PI controller:

Substituting equations (2) and (3) into (5) gives:

656

$$\frac{1}{K_p}\left(\tau_p\frac{d}{dt}+1\right)Y(t) = (K_c + \frac{K_c}{\tau_I}\int\limits_0^t dt)\epsilon(t) = (K_c + \frac{K_c}{\tau_I}\int\limits_0^t dt)(R(t)-Y(t))$$

Simplifying further gives:

$$\tau_p Y'(t) + Y(t) = K_c(R(t)-Y(t)) + \frac{K_c}{\tau_I}(R(t)-Y(t))$$

Differentiating the above equation gives:

$$\tau_p Y''(t) + Y'(t) = K_c R'(t) - K_c Y'(t) + \frac{K_c}{\tau_I}R(t) - \frac{K_c}{\tau_I}Y(t)$$

Rearranging the above equation:

$$\tau_p Y''(t) + (1+K_c)Y'(t) + \frac{K_c}{\tau_I}Y(t) = K_c R'(t) + \frac{K_c}{\tau_I}R(t)$$

Multiply the above equation by $\dfrac{\dfrac{\tau_I}{K_c}}{}$ to eliminate the co-efficient of Y'(t) gives:

$$\frac{\tau_I \tau_p}{K_c}Y''(t) + \tau_I\frac{1+K_c}{K_c}Y'(t) + Y(t) = \tau_I R'(t) + R(t)$$

[1] Coughanowr, D.R. and S.E. LeBlanc (2009). *Process Systems Analysis And Control*. Third Edition. McGraw Hill.

Section 2. P, I, D, PI, PD, and PID control

see also ECOSSE

Title: P, I, D, PI, PD, PID Control
Note: Video lecture available for this section!
Authors: Ardemis Boghossian, James Brown, Sara Zak
Date Presented: October 19, 2006
Stewards: Ji Sun Sunny Choi, Sang Lee, Jennifer Gehle, Brian Murray, Razili Stanke-Koch, Kelly Martin, Lance Dehne, Sean Gant, Jay Lee, Alex Efta

Date Revised: October 6, 2007

- First round reviews for this page
- Rebuttal for this page

2.1 Introduction

Process controls are necessary for designing safe and productive plants. A variety of process controls are used to manipulate processes, however the most simple and often most effective is the PID controller.

Much more practical than the typical on/off controller, PID controllers allow for much better adjustments to be made in the system. While this is true, there are some advantages to using an on/off controller:

-relatively simple to design and execute

-binary sensors and actuators (such as an on/off controller) are generally more reliable and less expensive

Although there are some advantages, there are large disadvantages to using an on/off controller scheme:

-inefficient (using this control is like driving with full gas and full breaks)

-can generate noise when seeking stability (can dramatically overshoot or undershoot a set-point)

-physically wearing on valves and switches (continuously turning valves/switches fully on and fully off causes them to become worn out much quicker)

To allow for much better control and fine-tuning adjustments, most industrial processes use a PID controller scheme.

The controller attempts to correct the error between a measured process variable and desired setpoint by calculating the difference and then performing a corrective action to adjust the process accordingly. A PID controller controls a process through three parameters: Proportional (P), Integral (I), and Derivative (D). These parameters can be weighted, or tuned, to adjust their effect on the process. The following section will provide a brief introduction on PID controllers as well as methods to model a controlled system in Excel.

The **Process Gain(K)** is the ratio of change of the output variable(responding variable) to the change of the input variable(forcing function). It specifically defines the sensitivity of the output variable to a given change in the input variable.

$$K = \frac{\Delta Output}{\Delta input}$$

Gain can only be described as a steady state parameter and give no knowledge about the dynamics of the process and is independent of the design and operating variables. A gain has three components that include the sign, the value, the units. The sign indicates how the output responds to the process input. A positive sign shows that the output variable increases with an increase in the input variable and a negative sign shows that the output variable decreases with an increase in the input variable. The units depend on the process considered that depend on the variables mentioned.

Example:

The pressure was increased from 21psi to 29psi. This change increased the valve position from 30%vp to 22%vp.

```
K = (29-21)psi / ((22-30)%vp) = -1.0psi/(%vp)
```

Dead Time(t0) is the between the change in an input variable and when the output variable begins. Dead time is important because it effects the controllability of the control system. A change in set point is not immediate because of this parameter. Dead time must be considered in tuning and modeling processes.

2.2 Types of Control

Process controls are instruments used to control a parameter, such as temperature, level, and pressure. PID controllers are a type of continuous controller because they continually adjust the output vs. an on/off controller, when looking at feed forward or feed backward conditions. An example of a temperature controller is shown in Figure 1.

Figure 1. Temperature controller in a CSTR

As shown in Figure 1, the temperature controller controls the temperature of a fluid within a CSTR (Continuous Stirred Tank Reactor). A temperature sensor first measures the temperature of the fluid. This measurement produces a measurement signal. The measurement signal is then compared to the set point, or desired temperature setting, of the controller. The difference between the measured signal and set point is the error. Based on this error, the controller sends an actuating signal to the heating coil, which adjusts the temperature accordingly. This type of process control is known as error-based control because the actuating signal is determined from the error between the actual and desired setting. The different types of error-based controls vary in the mathematical way they translate the error into an actuating signal, the most common of which are the PID controllers. Additionally, it is critical to understand feed-forward and feed-back control before exploring P, I, and D controls. Feed Forward Control Feedback Control

660

2.3 P, I, D, PI, PD, PID Control

As previously mentioned, controllers vary in the way they correlate the controller input (error) to the controller output (actuating signal). The most commonly used controllers are the proportional- integral-derivative (PID) controllers. PID controllers relate the error to the actuating signal either in a proportional (P), integral (I), or derivative (D) manner. PID controllers can also relate the error to the actuating signal using a combination of these controls.

2.3.1 Proportional (P) Control

One type of action used in PID controllers is the proportional control. Proportional control is a form of feedback control. It is the simplest form of continuous control that can be used in a closed-looped system. P-only control minimizes the fluctuation in the process variable, but it does not always bring the system to the desired set point. It provides a faster response than most other controllers, initially allowing the P-only controller to respond a few seconds faster. However, as the system becomes more complex (i.e. more complex algorithm) the response time difference could accumulate, allowing the P-controller to possibly respond even a few minutes faster. Athough the P-only controller does offer the advantage of faster response time, it produces deviation from the set point. This deviation is known as the offset, and it is usually not desired in a process. The existence of an offset implies that the system could not be maintained at the desired set point at steady state. It is analogous to the systematic error in a calibration curve, where there is always a set, constant error that prevents the line from crossing the origin. The offset can be minimized by combining P-only control with another form of control, such as I- or D- control. It is important to note, however, that it is impossible to completely eliminate the offset, which is implicitly included within each equation.

Mathematical Equations

P-control linearly correlates the controller output (actuating signal) to the error (diference between measured signal and set point). This P-control behavior is mathematically illustrated in Equation 1 (Scrcek, *et. al*).

$$c(t) = K_c e(t) + b$$

(1)

$$c(t) = \text{controller output}$$
$$K_c = \text{controller gain}$$
$$e(t) = \text{error}$$
$$b = \text{bias}$$

In this equation, the bias and controller gain are constants specific to each controller. The bias is simply the controller output when the error is zero. The controller gain is the change in the output of the controller per change in the input to the controller. In PID

controllers, where signals are usually electronically transmitted, controller gain relates the change in output voltage to the change in input voltage. These voltage changes are then directly related to the property being changed (i.e. temperature, pressure, level, etc.). Therefore, the gain ultimately relates the change in the input and output properties. If the output changes more than the input, K_c will be greater than 1. If the change in the input is greater than the change in the output, K_c will be less than 1. Ideally, if K_c is equal to infinity, the error can be reduced to zero. However, this infinitesimal nature of K_c increases the instability of the loop because zero error would imply that the the measured signal is *exactly* equal to the set point. As mentioned in lecture, exact equality is never achieved in control logic; instead, in control logic, error is allowed to vary within a certain range. Therefore, there are limits to the size of K_c, and these limits are defined by the system. Graphical representations of the effects of these variables on the system is shown in PID Tuning via Classical Methods.

As can be seen from the above equation, P-only control provides a linear relationship between the error of a system and the controller output of the system. This type of control provides a response, based on the signal that adjusts the system so that any oscillations are removed, and the system returns to steady-state. The inputs to the controller are the set point, the signal, and the bias. The controller calculates the difference between the set point and the signal, which is the error, and sends this value to an algorithm. Combined with the bias, this algorithm determines the action that the controller should take. A graphical representation of the P-controller output for a step increase in input at time t0 is shown below in Figure 2. This graph is exactly similar to the step input graph itself.

Figure 2. P-controller output for step input.

To illustrate this linear P-control relationship, consider the P-only control that controls the level of a fluid in a tank. Initially, the flow into the tank is equal to the flow out of the tank. However, if the flow out of the tank decreases, the level in the tank will increase because more fluid is entering than is leaving. The P-only control system will adjust the flow out of the tank so that it is again equal to the flow into the tank, and the level will once again be constant. However, this level is no longer equal to the initial level in the

662

tank. The system is at steady-state, but there is a difference between the initial set point and the current position in the tank. This difference is the P-control offset.

2.3.2 Integral (I) Control

Another type of action used in PID controllers is the integral control. Integral control is a second form of feedback control. It is often used because it is able to remove any deviations that may exist. Thus, the system returns to both steady state and its original setting. A negative error will cause the signal to the system to decrease, while a positive error will cause the signal to increase. However, I-only controllers are much slower in their response time than P-only controllers because they are dependent on more parameters. If it is essential to have no offset in the system, then an I-only controller should be used, but it will require a slower response time. This slower response time can be reduced by combining I-only control with another form, such as P or PD control. I-only controls are often used when measured variables need to remain within a very narrow range and require fine-tuning control. I controls affect the system by responding to accumulated past error. The philosophy behind the integral control is that deviations will be affected in proportion to the cumulative sum of their magnitude. The key advantage of adding a I-control to your controller is that it will eliminate the offset. The disadvantages are that it can destabilize the controller, and there is an integrator windup, which increases the time it takes for the controller to make changes.

Mathematical Equations

I-control correlates the controller output to the integral of the error. The integral of the error is taken with respect to time. It is the total error associated over a specified amount of time. This I-control behavior is mathematically illustrated in Equation 2 (Scrcek, *et. al*).

$$c(t) = \frac{1}{T_i} \int e(t)dt + c(t_0)$$

(2)

$c(t)$ = controller output
T_i = integral time
$e(t)$ = error
$c(t_0)$ = controller output before integration

In this equation, the integral time is the amount of time that it takes for the controller to change its output by a value equal to the error. The controller output before integration is equal to either the initial output at time t=0, or the controller output at the time one step before the measurement. Graphical representations of the effects of these variables on the system is shown in PID Tuning via Classical Methods.

The rate of change in controller output for I-only control is determined by a number of parameters. While the P-only controller was determined by e, the rate of change for I-

only depends on both e and T_i. Because of the inverse relationship between c(t) and T_i, this decreases the rate of change for an I-only controller.

The I-only controller operates in essentially the same way as a P-only controller. The inputs are again the set point, the signal, and the bias. Once again, the error is calculated, and this value is sent to the algorithm. However, instead of just using a linear relationship to calculate the response, the algorithm now uses an integral to determine the response that should be taken. Once the integral is evaluated, the response is sent and the system adjusts accordingly. Because of the dependence on T_i, it takes longer for the algorithm to determine the proper response. A graphical representation of the I-controller output for a step increase in input at time t0 is shown below in Figure 3. As expected, this graph represents the area under the step input graph.

Figure 3. I-controller output for step input.

2.3.3 Derivative (D) Control

Another type of action used in PID controllers is the derivative control. Unlike P-only and I-only controls, D-control is a form of feed forward control. D-control anticipates the process conditions by analyzing the change in error. It functions to minimize the change of error, thus keeping the system at a consistent setting. The primary benefit of D controllers is to resist change in the system, the most important of these being oscillations. The control output is calculated based on the rate of change of the error with time. The larger the rate of the change in error, the more pronounced the controller response will be.

Unlike proportional and integral controllers, derivative controllers do not guide the system to a steady state. Because of this property, D controllers must be coupled with P, I or PI controllers to properly control the system.

Mathematical Equations

D-control correlates the controller output to the derivative of the error. The derivative of the error is taken with respect to time. It is the change in error associated with change in time. This D-control behavior is mathematically illustrated in Equation 3 (Scrcek, *et. al*).

$$c(t) = T_d \frac{de}{dt}$$

c(t) = controller output
T_d = derivative time constant
de = change in error
dt = change in time

Graphical representations of the effects of these variables on the system is shown in PID Tuning via Classical Methods.

Mathematically, derivative control is the opposite of integral control. Although I-only controls exist, D-only controls do not exist. D-controls measure only the change in error. D-controls do not know where the setpoint is, so it is usually used in conjunction with another method of control, such as P-only or a PI combination control. D-control is usually used for processes with rapidly changing process outputs. However, like the I-control, the D control is mathematically more complex than the P-control. Since it will take a computer algorithm longer to calculate a derivative or an integral than to simply linearly relate the input and output variables, adding a D-control slows down the controller's response time. A graphical representation of the D-controller output for a step increase in input at time t0 is shown below in Figure 4. As expected, this graph represents the derivative of the step input graph.

Figure 4. D-controller output for step input.

Controller Effects on a System

The following images are intended to give a visual representation of how P, I, and D controllers will affect a system.

Description

Figure 5. Stable data sample.

Figure 6. Data disturbance.

25 — We could use a integral controller for
a problem that keeps reoccurring.

20 —

Recall: An I controller adds up
past deviations in an attempt
15 — to reduce future offsets

10 —

- Set point
5 — - Data
- I controller response   But the more dramatic changes
can make it difficult to get back to
the set point

0 —

0   2   4   6   8   10

Figure 8. I-controller effect on data.

25 — We could use a derivative controller
if we are worried about constant
change

20 —
Recall: A D controller only
reacts to changes
15 —

10 —

- Set point
5 — - Data
- D controller response   But it won't necessarily get
you back to your set point

0 —

0   2   4   6   8   10

Figure 9. D-controller effect on data.

Continue reading to see the results of combining controllers.

Proportional-Integral (PI) Control

One combination is the PI-control, which lacks the D-control of the PID system. PI control is a form of feedback control. It provides a faster response time than I-only control due to the addition of the proportional action. PI control stops the system from fluctuating, and it is also able to return the system to its set point. Although the response time for PI-control is faster than I-only control, it is still up to 50% slower than P-only control. Therefore, in order to increase response time, PI control is often combined with D-only control.

Mathematical Equations

PI-control correlates the controller output to the error and the integral of the error. This PI-control behavior is mathematically illustrated in Equation 4 (Scrcek, *et. al*).

$$c(t) = K_c \left(e(t) + \frac{1}{Ti} \int e(t)dt \right) + C$$

(4)

c(t) = controller output
K_c = controller gain
T_i = integral time
e(t) = error
C = initial value of controller

In this equation, the integral time is the time required for the I-only portion of the controller to match the control provided by the P-only part of the controller.

The equation indicates that the PI-controller operates like a simplified PID-controller with a zero derivative term. Alternatively, the PI-controller can also be seen as a combination of the P-only and I-only control equations. The bias term in the P-only control is equal to the integral action of the I-only control. The P-only control is only in action when the system is not at the set point. When the system is at the set point, the error is equal to zero, and the first term drops out of the equation. The system is then being controlled only by the I-only portion of the controller. Should the system deviate from the set point again, P-only control will be enacted. A graphical representation of the PI-controller output for a step increase in input at time t0 is shown below in Figure 5. As expected, this graph resembles the qualitatitive combination of the P-only and I-only graphs.

668

Figure 10. PI-controller output for step input.

Effects of K_c and T_i

With a PI control system, controller activity (aggressiveness) increases as K_c and T_i decreases, however they can act individually on the aggressiveness of a controller's response. Consider Figure 11 below with the center graph being a linear second order system base case.

Figure 11. Effects of K_c and T_i [2]

The plot depicts how T_i and K_c both affect the performance of a system, whether they are both affecting it or each one is independently doing so. Regardless of integral time, increasing controller gain (moving form bottom to top on the plot) will increase controller activity. Similarly, decreasing integral time (moving right to left on the plot) will increase controller activity independent of controller gain. As expected, increasing K_c

and decreasing T_i would compound sensitivity and create the most aggressive controller scenario.

With only two interacting parameters in PI control systems, similar performance plots can still cause confusion. For example, plots A and B from the figure both look very similar despite different parameters being affected in each of them. This could cause further problems and create a wildly aggressive system if the wrong parameter is being corrected. While trial and error may be feasible for a PI system, it becomes cumbersome in PID where a third parameter is introduced and plots become increasingly similar.

Another noteworthy observation is the plot with a normal K_c and double T_i. The plot depicts how the proportional term is practical but the integral is not receiving enough weight initially, causing the slight oscillation before the integral term can finally catch up and help the system towards the set point.

Proportional-Derivative (PD) Control

Another combination of controls is the PD-control, which lacks the I-control of the PID system. PD-control is combination of feedforward and feedback control, because it operates on both the current process conditions and predicted process conditions. In PD-control, the control output is a linear combination of the error signal and its derivative. PD-control contains the proportional control's damping of the fluctuation and the derivative control's prediction of process error.

Mathematical Equations
As mentioned, PD-control correlates the controller output to the error and the derivative of the error. This PD-control behavior is mathematically illustrated in Equation 5 (Scrcek, *et. al*).

$$c(t) = K_c \left(e(t) + T_d \frac{de}{dt} \right) + C$$

(5)

$$c(t) = \text{controller output}$$
$$K_c = \text{proportional gain}$$
$$e = \text{error}$$
$$C = \text{initial value of controller}$$

The equation indicates that the PD-controller operates like a simplified PID-controller with a zero integral term. Alternatively, the PD-controller can also be seen as a combination of the P-only and D-only control equations. In this control, the purpose of the D-only control is to predict the error in order to increase stability of the closed loop system. P-D control is not commonly used because of the lack of the integral term.

670

Without the integral term, the error in steady state operation is not minimized. P-D control is usually used in batch pH control loops, where error in steady state operation does not need to be minimized. In this application, the error is related to the actuating signal both through the proportional and derivative term. A graphical representation of the PD-controller output for a step increase in input at time t0 is shown below in Figure 6. Again, this graph is a combination of the P-only and D-only graphs, as expected.

Figure 12. PD-controller output for step input.

2.3.4 Proportional-Integral-Derivative (PID) Control

Proportional-integral-derivative control is a combination of all three types of control methods. PID-control is most commonly used because it combines the advantages of each type of control. This includes a quicker response time because of the P-only control, along with the decreased/zero offset from the combined derivative and integral controllers. This offset was removed by additionally using the I-control. The addition of D-control greatly increases the controller's response when used in combination because it predicts disturbances to the system by measuring the change in error. On the contrary, as mentioned previously, when used individually, it has a slower response time compared to the quicker P-only control. However, although the PID controller seems to be the most adequate controller, it is also the most expensive controller. Therefore, it is not used unless the process requires the accuracy and stability provided by the PID controller.

Mathematical Equations
PID-control correlates the controller output to the error, integral of the error, and derivative of the error. This PID-control behavior is mathematically illustrated in Equation 6 (Scrcek, *et. al*).

$$c(t) = K_c \left(e(t) + \frac{1}{T_i} \int e(t) dt + T_d \frac{de}{dt} \right) + C$$

(6)

$$c(t) = \text{controller output}$$
$$K_c = \text{controller gain}$$
$$e(t) = \text{error}$$
$$T_i = \text{integral time}$$
$$T_d = \text{derivative time constant}$$
$$C = \text{intitial value of controller}$$

As shown in the above equation, PID control is the combination of all three types of control. In this equation, the gain is multiplied with the integral and derivative terms, along with the proportional term, because in PID combination control, the gain affects the I and D actions as well. Because of the use of derivative control, PID control cannot be used in processes where there is a lot of noise, since the noise would interfere with the predictive, feedforward aspect. However, PID control is used when the process requires no offset and a fast response time. A graphical representation of the PID-controller output for a step increase in input at time t0 is shown below in Figure 7. This graph resembles the qualitative combination of the P-only, I-only, and D-only graphs.

Figure 7. PID-controller output for step input.

In addition to PID-control, the P-, I-, and D- controls can be combined in other ways. These alternative combinations are simplifications of the PID-control.

Note: Order of e(t)
The order of the elements in the e(t) can vary depending on the situation. It could be the fixed element minus the varying element or the other way around. To better illustrate the concept let's go to an example. Let's say you are creating a PID control to control the fluid level in a tank by manipulating the outlet valve. When the fluid level in the tank exceeds your set value, you will want the valve to open up more to allow more flow out of the tank. You are looking for a positive response. Therefore your e(t) should give a positive value when the fluid level is higher than the set. In this case your e(t) will be (V-

672

Vset). The same logic can be used for other systems to determine what the e(t) should be in the PID controls.

2.4 Modeling PID Controllers With Euler in Excel

As with many engineering systems, PID controllers can be modeled in Excel via numerical methods such as Euler's Method. First begin with the initial value for a given parameter. Determine the change in that parameter at a certain time-step by summing the three controllers P, I, and D at that step, which are found using the equations listed in the P, I, D, PI, PD, PID Control section above. Take this change, multiply it by the chosen time-step and add that to the previous value of the parameter of interest. For more detailed information see Numerical ODE Solving in Excel. An example of a chemical engineering problem that uses this method can be seen in Example 4 below.

2.4.1 Troubleshooting PID Modeling in Excel

When setting up an Excel spreadsheet to model a PID controller, you may receive an error message saying that you have created a circular reference. Say you are controlling the flow rate of one reactant (B) to a reactor which is dependent upon the concentration of another reactant (A) already inside the reactor. Your PID equations look as follows:

$$FB = 1 + K_c \left[(A - A_{\text{set}}) + \frac{1}{\tau_i} \int (A - A_{\text{set}}) dt + \tau_d \frac{d(A - A_{\text{set}})}{dt} \right]$$

$$x_i = \frac{d(A - A_{\text{set}})}{dt}$$

After you have set up your columns for A - Aset, d(A - Aset)/dt, xi, and the cells for your parameters like Kc, tau$i$ and tau$d$, you will need to set up your PID column with your PID equation in it. After entering your equation into the first cell of the PID column, you may receive the Circular Reference error message when trying to drag the equation down to the other cells in your column.

There are two things you can do:

1. It is likely that you need to start your PID equation in the second or third cell of your PID column. Enter reasonable values into the first couple of cells before you start the PID equation, and you will find that these values shouldn't affect the end result of your controller.
2. You can also try decreasing the step size (Δt).

2.4.2 Summary Tables

A summary of the advantages and disadvantages of the three controls is shown below is shown in Table 1.

Table 1. Advantages and disadvantages of controls

	Proportional (P)	Integral (I)	Derivative (D)
Advantages	+ Fast response time + Minimizes fluctuation	+ Contains small offset + Returns system to steady state	+ Keeps system at consistent setting + Controls processes with rapidly changing outputs
Disadvantages	- Contains large offset - Does not bring system to desired set point	- Slow response time	- Slow response time - Requires combined use with another controller

A guide for the typical uses of the various controllers is shown below in Table 2.

Table 2. Typical uses of P, I, D, PI, and PID controllers

Controller	Estimates	When to use	Examples
P	Present	Systems with slow response, systems tolerant to offset	Float valves, thermostats, humidistat
I	Back	Not often used alone, as is too slow	Used for very noisy systems
D	Forward	Not used alone because it is too sensitive to noise and does not have set point	None
PI	Present & back	Often used	Thermostats, flow control, pressure control
PID	All time	Often used, most robust, but can be noise sensitive	Cases where the system has intertia that could get out of hand: i.e. temperature and concentration measurements on a reactor to avoid runaway.

A summary of definitions of the terms and symbols are shown below in Table 3.

Table 3. Definitions of terms and symbols.

674

Term or Symbol	Defintion
Bias	Controller output when the error is equal to zero
Change in Error, de	Differential change in error used in the the Derivative Action equation
Change in Time, dt	Differential change in time used in the Deriative Action equatio
Controller	Instrument used to adjust a system property such as temperature, pressure, and level. Controllers vary in three main aspects: (1) The way they correlate error and actuating signal (such as PID controllers) (2) The way they work relative to the data they gather (feedback Vs. feedforward) (3) The way the adjust the system setting (on-off Vs. continuous)
Controller Gain, Kc	Ratio of the change in output to the change in input
Controller Output, c(t)	Voltage output from the controller, usually relating to the actuating signal
Controller Output before Integration, c(t0)	Constant used in the Integral Action equation
Derivative Time Constant, Td	Constant used in the Deriviative Action equation
Error, e(t)	The difference between the measured signal and the set point
Offset	Systematic error that prevents the system from being at the desired set point during steady state
Integral Time, Ti	The time it takes the controller to change the output by a value equal to the error

2.5 Example 1

Hypothetical Industries has just put you in charge of one of their batch reactors. Your task is to figure out a way to maintain a setpoint level inside of the reactor. Your boss wants to use some type regulator controller, but he is not quite sure which to use. Help your boss find the right type of controller. It is extremely important that the level inside the reactor is at the setpoint. Large fluctuation and error cannot be tolerated.

SOLUTION:
You would want to use a PID controller. Because of the action of P control, the system will respond to a change very quickly. Due to the action of I control, the system is able to be returned to the setpoint value. Finally, because it is so critical for the system to remain at a constant setpoint, D control will measure the change in the error, and help to adjust the system accordingly.

2.6 Example 2

You go back to your high school and you notice an oven in your old chemistry class. The oven is used to remove water from solutions. Using your knowledge from ChE 466, you begin to wonder what type of controller the oven uses to maintain its set temperature.

You notice some high school students eager to learn, and you decide to share your knowledge with them in hopes of inspiring them to become Chemical Engineers. Explain to them the type of controller that is most likely located within the oven, and how that controller works.

SOLUTION:
Since the oven is only used to remove water from a solution, fluctuation, error, and lag between the set point and the actual temperature are all acceptable. Therefore, the easiest and simplest controller to use would be the On-Off controller.
The On-Off controller turns on the heating mechanism when the temperature in the oven is below the setpoint temperature. If the temperature of the oven increases above the set temperature, the controller will turn the heating mechanism off.

2.7 Example 3

Having taken your advice, your boss at Hypothetical Industries decides to install a PID controller to control the level in the batch reactor. When you first start up the reactor, the controller initially received a step input. As the reactor achieves steady state, the level in the reactor tends to fluctuate, sending pulse inputs into the controller. For a pulse input, provide a grahical representation of the PID controller output.

Figure 8. Pulse input.

SOLUTION:

The PID-controller output will be a combination of the P-only, I-only and D-only controller outputs. Analogous to the P-controller output for the step input, the P-controller output for the pulse input will exactly resemble the input.

Figure 9. P-controller output for pulse input.

The I-controller output represents the area under the input graph. Unlike the step input, the area under the pulse input graph dropped back down to zero once the pulse has passed. Therefore, rather than continually increase, the I-controller output graph will level off in the end.

Figure 10. I-controller output for pulse input.

The D-controller output represents the derivative of the input graph. The derivative at the first discontinuity of the graph would be positive infinity. The derivative of the second downward discontinuity is negative infinity.

677

Figure 11. D-controller output for pulse input.

Combining the qualitative characteristics of all three graphs we can determine the PID-controller output for a pulse input.

Figure 12. PID-controller output for pulse input.

2.8 Example 4

Different kinds of disturbances are possible when determining the PID controller robustness. These different disturbances are used to simulate changes that might occur within your system. For a CSTR reactor, you decide to heat up your system to account for the cold outside weather. The disturbance in the input temperature is a ramp disturbance, as shown in figure #. If the controller responds to the input temperature, what will be the PID controller output?

678

Figure 13. Ramp input.

SOLUTION: Using a controller with a p-only controller only, we will see a proportional change in the controller output corresponding to the input variable change. See figure 14 below

Figure 14. P-controller output for ramp input.

Using an I-only controller, we will see the controller corresponding to the area under the graph, which in this case, seem to increase exponentially with the ramp geometry.

Figure 15. I-controller output for ramp input.

Using a D-only controller, we will see a step response to the ramp disturbance. This is because the D-component corresponds to the derivative, and a ramp input shows a constant slope (positive in this case) which is different than the starting condition slope (zero usually). See figure 16.

Figure 16. D-controller output for ramp input.

Using a PID controller, the three components all come to play in the controller output. As we would expect, the result will be just a simple addition of the three separate component graphs.

Figure 17. D-controller output for ramp input.

Following is a P&ID of the process A+B-->C.

Figure 18. P&ID for a reaction process.

What is the PID controller expression on V3 controlling the volume in TK001 to a setpoint of 50 liters? Note: The PID controller uses LC1 to measure the volume.

SOLUTION: The general equation for a PID controller is:

681

$$c(t) = K_c \left[e(t) + \frac{1}{T_i} \int e(t)dt + T_d \frac{de}{dt} \right] + C$$

c(t) = controller output
K_c = controller gain
e(t) = error
T_i = integral time
T_d = derivative time constant
C = intitial value of controller

Therefore, for this example, the solution is:

$$F_{\text{out}} = offset + K_c \left[(V_1 - V_{\text{set}}) + \frac{1}{T_i} \int (V_1 - V_{\text{set}})dt + T_d \frac{d(V_1 - V_{\text{set}})}{dt} \right]$$

2.10 Example 6

In this problem, the differential equations describing a particular first-order system with a PID controller will be determined. This is designed to show how well-defined systems can be modeled or explained mathematically. As a supplement to this problem, visit Constructing Block Diagrams. Note that this is an example of solution using integro-differential operators rather than Laplace transforms. Here we have set $K_p = 1$.

Consider a general first-order process:

$$\tau_p Y'(t) + Y(t) = X(t)$$

Where $Y(t)$ is the output of the system and $X(t)$ is the input. Add a PID controller to the system and solve for a single, simple differential equation. The operator or equation for a PID controller is below. Assume there is no dead time in the measurement.

$$G = K_c \left[1 + \frac{1}{\tau_I} \int_0^t d\tau + \tau_D \frac{d}{dt} \right]$$

Solution: Use Constructing Block Diagrams as a reference when solving this problem. Equations defining system,

Process: $\tau_p Y'(t) + Y(t) = X(t)$
Controller: $X(t) = G\varepsilon(t)$
Comparator: $\varepsilon(t) = R(t) - M(t)$
Measurement: $M(t) = Y(t)$

When these equations are combined into one equation, the following equation results. This is achieved by adding the measurement to the comparator to the controller to the process equation.

$$\tau_p Y'(t) + Y(t) = G(R(t) - Y(t))$$

Substituting the controller operator and then evaluating yields:

$$\tau_p Y'(t) + Y(t) = K_c \left[1 + \frac{1}{\tau_I} \int_0^t d\tau + \tau_D \frac{d}{dt} (R(t) - Y(t)) \right]$$

$$\tau_p Y'(t) + Y(t) = K_c \left[R(t) + \frac{1}{\tau_I} \int_0^t R(t) d\tau + \tau_D \frac{d}{dt} R(t) - [Y(t) + \frac{1}{\tau_I} \int_0^t Y(t) d\tau + \tau_D \frac{d}{dt} Y(t)] \right]$$

Because there is an integral in the differential equation, it is necessary to take the derivative with respect to time.

$$\tau_p Y''(t) + Y'(t) = K_c \left[R'(t) + \frac{1}{\tau_I} R(t) + \tau_D R''(t) - [Y'(t) + \frac{1}{\tau_I} Y(t) + \tau_D Y''(t)] \right]$$

$$\tau_p Y''(t) + Y'(t) = K_c(R'(t) - Y'(t)) + \frac{K_c}{\tau_I}(R(t) - Y(t)) + K_c \tau_D(R''(t) - Y''(t))$$

To put this in standard notation for solving a second order differential equation, the Y(t) need to be on one side, and the R(t) terms need to be on the opposite side. Also, the coefficient of the Y(t) term needs to be one.

$$\frac{\tau_I(\tau_p + K_c \tau_D)}{K_c} Y''(t) + \frac{\tau_I(1 + K_c)}{K_c} Y'(t) + Y(t) = \tau_I \tau_D R''(t) + \tau_I R'(t) + R(t)$$

The above equation can then be solved by hand or using a program such as Mathematica. If using a computer program, different values for the control parameters K_c, τ_I, τ_D can be chosen, and the response to a change in the system can be evaluated graphically.

2.11 Multiple Choice Question 1

What type of controller is displayed by the equation below?

$$c(t) = K_c[e(t) + \frac{1}{T_i} \int e(t) dt]$$

a.) Feedforward
b.) PID
c.) Derivative
d.) Proportional Integral

Answer: d

2.12 Multiple Choice Question 2

Which type of controller increases the stability of the system by keeping it at a consistent setting?

a.) Derivative
b.) Proportional
c.) On-Off
d.) Integral

Answer: a

2.13 Multiple Choice Question 3

Which type of controller increases the speed of response to reach the desired set point the fastest while eliminating offset?

a.) On-Off
b.) Proportional
c.) Integral
d.) Proportional-Integral

Answer: d

2.14 Example 4

- Note that the problem and values used in it are fictional!*

A microbiology laboratory discovered a deadly new strain of bacteria, named P. Woolfi, in the city's water supply. In order to sterilize the water, the bacteria has to be heat killed at a temperature of 105 degrees Celsius. However, this is above the boiling point of water at 1 atm and the bacteria is only susceptible to heat when in liquid due to spore formation when in gas or vapor. To accomplish this sterilization it was suggested that an auto-clave be used to keep the water in the liquid state by keeping it at a pressure of 5 atm while heated for 30 seconds. The auto-clave can only handle up to 7 atm of pressure before exploding, so to ensure that the process is running at the desired specifications, a PID Controller Model must be created in Excel. See figure 18 for a visual of the system.

Figure 18. Auto-clave with PID Controls for Temperature and Pressure

Click on this link for the worked out Excel Solution

Explanation:

To simulate the real situation of pressure varying in the system, column B calls an equation to generate a random fluctuation in pressure. Pset is simply the desired specification. Error is the difference between the set pressure and measured signal. du/dt is the sum of the P, I, and D terms. The equations used to calculate each of these can be found in the article, these take into account the error associated with each time-step. dU/dt is the parameter that is varied in order to correct for the difference between the measured pressure and desired pressure.

2.15 Sage's Corner

Super PID Brothers
http://video.google.com/googleplayer.swf?docId=6266024888342893924
Slides without narration

Multiple Choice
http://www.youtube.com/v/p8mw97PAabI

Glucose Level
http://video.google.com/googleplayer.swf?docId=-379606757698801421
Glucose Level Spreadsheet

Glucose Level Slides without narration

2.16 References

- Astrom, Karl J., Hagglund, Tore., "Advanced PID Control", The Instrumentation, Systems and Automation Society.
- Cooper, Douglas J. "Practical Process Control E-Textbook " http://www.controlguru.com
- Scrcek, William Y., Mahoney, Donald P., Young, Brent R. "A Real Time Approach to Process Control", 2nd Edition. John Wiley & Sons, Ltd.
- www.wikipedia.org

Section 3. PID tuning via classical methods

See also ECOSSE

Title: PID Tuning Via Classical Methods

Note: Video lecture available for this section!

Authors: James Bennett, Ajay Bhasin, Jamila Grant, Wen Chung Lim

Stewards: Arthur Edge III, Kathryn Mientel, Renu Rao, Kaveh Saba

Date Presented: 10/19/2006 // Date Revised: 10/16/2007

- First round reviews for this page
- Rebuttal for this page

3.1 Introduction

Currently, more than half of the controllers used in industry are PID controllers. In the past, many of these controllers were analog; however, many of today's controllers use digital signals and computers. When a mathematical model of a system is available, the parameters of the controller can be explicitly determined. However, when a mathematical model is unavailable, the parameters must be determined experimentally. Controller tuning is the process of determining the controller parameters which produce the desired output. Controller tuning allows for optimization of a process and minimizes the error between the variable of the process and its set point.

Types of controller tuning methods include the trial and error method, and process reaction curve methods. The most common classical controller tuning methods are the Ziegler-Nichols and Cohen-Coon methods. These methods are often used when the mathematical model of the system is not available. The Ziegler-Nichols method can be used for both **closed** and **open loop systems**, while Cohen-Coon is typically used for open loop systems. A closed-loop control system is a system which uses feedback control. In an open-loop system, the output is not compared to the input.

The equation below shows the PID algorithm as discussed in the previous PID Control section.

$$u(t) = K_c \left(\epsilon(t) + \frac{1}{\tau_i} \int_0^t \epsilon(t') \, dt' + \tau_d \frac{d\epsilon(t)}{dt} \right) + b$$

u is the control signal

687

ε is the difference between the current value and the set point.

K_c is the gain for a proportional controller.

τ_i is the parameter that scales the integral controller.

τ_d is the parameter that scales the derivative controller.

t is the time taken for error measurement.

b is the set point value of the signal, also known as bias or offset.

The experimentally obtained controller **gain** which gives stable and consistent oscillations for closed loop systems, or the **ultimate gain**, is defined as K_u . K_c is the controller gain which has been corrected by the Ziegler-Nichols or Cohen-Coon methods, and can be input into the above equation. K_u is found experimentally by starting from a small value of K_c and adjusting upwards until consistent oscillations are obtained, as shown below. If the gain is too low, the output signal will be damped and attain equilibrium eventually after the disturbance occurs as shown below.

On the other hand, if the gain is too high, the oscillations become unstable and grow larger and larger with time as shown below.

688

The process reaction curve method section shows the parameters required for open loop system calculations. The Ziegler-Nichols Method section shows how to find K_c, T_i, and T_d for open and closed loop systems, and the Cohen-Coon section shows an alternative way to find K_c, T_i, and T_d.

Open loop systems typically use the quarter decay ratio (QDR) for oscillation dampening. This means that the ratio of the amplitudes of the first overshoot to the second overshoot is 4:1.

3.2 Trial and Error

The trial and error tuning method is based on guess-and-check. In this method, the proportional action is the main control, while the integral and derivative actions refine it. The controller gain, K_c, is adjusted with the integral and derivative actions held at a minimum, until a desired output is obtained.

Below are some common values of K_c, T_i, and T_d used in controlling flow, levels, pressure or temperature for trial and error calculations.

689

3.2.1 Flow

P or PI control can be used with low controller gain. Use PI control for more accuracy with high integration activity. Derivative control is not considered due to the rapid fluctuations in flow dynamics with lots of noise.

K_c = 0.4-0.65

T_i = 6s

3.2.2 Level

P or PI control can be used, although PI control is more common due to inaccuracies incurred due to offsets in P-only control. Derivative control is not considered due to the rapid fluctuations in flow dynamics with lots of noise.

The following P only setting is such that the control valve is fully open when the vessel is 75% full and fully closed when 25% full, being half open when 50% filled.

K_c = 2

Bias b = 50%

Set point = 50%

For PI control:

K_c = 2-20

T_i = 1-5 min

3.2.3 Pressure

Tuning here has a large range of possible values of K_c and T_i for use in PI control, depending on if the pressure measurement is in liquid or gas phase.

Liquid
K_c = 0.5-2

T_i = 6-15 s

Gas
K_c = 2-10

T_i = 2-10 min

3.2.4 Temperature

Due to the relatively slow response of temperature sensors to dynamic temperature changes, PID controllers are used.

K_c = 2-10

T_i = 2-10 min

T_d = 0-5 min

3.3 Process Reaction Curve

In this method, the variables being measured are those of a system that is already in place. A disturbance is introduced into the system and data can then be obtained from this curve. First the system is allowed to reach steady state, and then a disturbance, X_o, is introduced to it. The percentage of disturbance to the system can be introduced by a change in either the set point or process variable. For example, if you have a thermometer in which you can only turn it up or down by 10 degrees, then raising the temperature by 1 degree would be a 10% disturbance to the system. These types of curves are obtained in open loop systems when there is no control of the system, allowing the disturbance to be recorded. The process reaction curve method usually produces a response to a step function change for which several parameters may be measured which include: transportation lag or dead time, τ_{dead}, the time for the response to change, τ, and the ultimate value that the response reaches at steady-state, M_u.

τ_{dead} = transportation lag or dead time: the time taken from the moment the disturbance was introduced to the first sign of change in the output signal

τ = the time for the response to occur

X_o = the size of the step change

M_u = the value that the response goes to as the system returns to steady-state

$$R = \frac{\tau_{dead}}{\tau}$$

$$K_o = \frac{X_o}{M_u} \frac{\tau}{\tau_{dead}}$$

An example for determining these parameters for a typical process response curve to a step change is shown below.

In order to find the values for τ_{dead} and τ, a line is drawn at the point of inflection that is tangent to the response curve and then these values are found from the graph.

To map these parameters to P,I, and D control constants, see Table 2 and 3 below in the Z-N and Cohen Coon sections.

3.4 Ziegler-Nichols Method

In the 1940's, Ziegler and Nichols devised two empirical methods for obtaining controller parameters. Their methods were used for non-first order plus dead time situations, and involved intense manual calculations. With improved optimization software, most manual methods such as these are no longer used. However, even with computer aids, the following two methods are still employed today, and are considered among the most common:

692

3.4.1 Ziegler-Nichols closed-loop tuning method

The Ziegler-Nichols closed-loop tuning method allows you to use the ultimate gain value, K_u, and the ultimate period of oscillation, P_u, to calculate K_c. It is a simple method of tuning PID controllers and can be refined to give better approximations of the controller. You can obtain the controller constants K_c, T_i, and T_d in a system with feedback. The Ziegler-Nichols closed-loop tuning method is limited to tuning processes that cannot run in an open-loop environment.

Determining the ultimate gain value, K_u, is accomplished by finding the value of the proportional-only gain that causes the control loop to oscillate indefinitely at steady state. This means that the gains from the I and D controller are set to zero so that the influence of P can be determined. It tests the robustness of the K_c value so that it is optimized for the controller. Another important value associated with this proportional-only control tuning method is the ultimate period (P_u). The ultimate period is the time required to complete one full oscillation while the system is at steady state. These two parameters, K_u and P_u, are used to find the loop-tuning constants of the controller (P, PI, or PID). To find the values of these parameters, and to calculate the tuning constants, use the following procedure:

Closed Loop (Feedback Loop)

1. Remove integral and derivative action. Set integral time (T_i) to 999 or its largest value and set the derivative controller (T_d) to zero.
2. Create a small disturbance in the loop by changing the set point. Adjust the proportional, increasing and/or decreasing, the gain until the oscillations have constant amplitude.
3. Record the gain value (K_u) and period of oscillation (P_u).

Figure 1. System tuned using the Ziegler-Nichols closed-loop tuning method

4. Plug these values into the Ziegler-Nichols closed loop equations and determine the necessary settings for the controller.

Table 1. Closed-Loop Calculations of K_c, T_i, T_d

	K_c	T_I	T_D
P	$K_u/2$		
PI	$K_u/2.2$	$P_u/1.2$	
PID	$K_u/1.7$	$P_u/2$	$P_u/8$

Advantages

1. Easy experiment; only need to change the P controller
2. Includes dynamics of whole process, which gives a more accurate picture of how the system is behaving

Disadvantages

1. Experiment can be time consuming
2. Can venture into unstable regions while testing the P controller, which could cause the system to become out of control

3.4.2 Ziegler-Nichols Open-Loop Tuning Method or Process Reaction Method:

This method remains a popular technique for tuning controllers that use proportional, integral, and derivative actions. The Ziegler-Nichols open-loop method is also referred to as a process reaction method, because it tests the open-loop reaction of the process to a change in the control variable output. This basic test requires that the response of the system be recorded, preferably by a plotter or computer. Once certain process response values are found, they can be plugged into the Ziegler-Nichols equation with specific multiplier constants for the gains of a controller with either P, PI, or PID actions.

Open Loop (Feed Forward Loop)

To use the Ziegler-Nichols open-loop tuning method, you must perform the following steps:

1. Make an open loop step test

694

2. From the <u>process reaction curve</u> determine the transportation lag or dead time, τ_{dead}, the time constant or time for the response to change, τ, and the ultimate value that the response reaches at steady-state, M_u, for a step change of Xo.

$$K_o = \frac{X_o}{M_u} \frac{\tau}{\tau_{dead}}$$

3. Determine the loop tuning constants. Plug in the reaction rate and lag time values to the Ziegler-Nichols open-loop tuning equations for the appropriate controller—P, PI, or PID—to calculate the controller constants. Use the table below.

Table 2. Open-Loop Calculations of K_c, T_i, T_d

	K_c	T_i	T_d
P	K_0		
PI	$0.9K_0$	$3.3\tau_{dead}$	
PID	$1.2K_0$	$2\tau_{dead}$	$0.5\tau_{dead}$

Advantages

1. Quick and easier to use than other methods
2. It is a robust and popular method
3. Of these two techniques, the Process Reaction Method is the easiest and least disruptive to implement

Disadvantages

1. It depends upon purely proportional measurement to estimate I and D controllers.
2. Approximations for the K_c, T_i, and T_d values might not be entirely accurate for different systems.
3. It does not hold for I, D and PD controllers

3.5 Cohen-Coon Method

The Cohen-Coon method of controller tuning corrects the slow, steady-state response given by the Ziegler-Nichols method when there is a large dead time (process delay) relative to the open loop time constant; a large process delay is necessary to make this method practical because otherwise unreasonably large controller gains will be predicted.

This method is only used for first-order models with time delay, due to the fact that the controller does not instantaneously respond to the disturbance (the step disturbance is progressive instead of instantaneous).

The Cohen-Coon method is classified as an 'offline' method for tuning, meaning that a step change can be introduced to the input once it is at steady-state. Then the output can be measured based on the time constant and the time delay and this response can be used to evaluate the initial control parameters.

For the Cohen-Coon method, there are a set of pre-determined settings to get minimum offset and standard decay ratio of 1/4(QDR). A 1/4(QDR) decay ratio refers to a response that has decreasing oscillations in such a manner that the second oscillation will have 1/4 the amplitude of the first oscillation . These settings are shown in Table 3..

Table 3. Standard recommended equations to optimize Cohen Coon predictions

	Kc	Ti	Td
P	(P/NL)*(1+(R/3))		
PI	(P/NL)*(0.9+(R/12))	L*(30+3R) / (9+20R)	
PID	(P/NL)*(1.33+(R/4))	L*(30+3R) / (9+20R)	4L / (11+2R)

where the variables P, N, and L are defined below.

P	Percent Change of Input
N	Percent Change of Output/τ
L	τ_{dead}
R	τ_{dead}/τ

Alternatively, K_0 can be used instead of (P/NL). K_0, τ, and τ_{dead} are defined in process reaction curve section. An example using these parameters is shown here [1].

The process in Cohen-Coon turning method is the following:

1. Wait until the process reaches steady state.

2. Introduce a step change in the input.

3. Based on the output, obtain an approximate first order process with a time constant τ delayed by τ_{dead} units from when the input step was introduced.

696

The values of τ and τ_{dead} can be obtained by first recording the following time instances:

t0 = time at input step start point t2 = time when reaches half point t3 = time when reaches 63.2% point

4. Using the measurements at t0, t2, t3, A and B, evaluate the process parameters τ, τ_{dead}, and K_o.

5. Find the controller parameters based on τ, τ_{dead}, and K_o.

3.5.1 Advantages

1. Used for systems with time delay.
2. Quicker closed loop response time.

3.5.2 Disadvantages and Limitations

1. Unstable closed loop systems.
2. Can only be used for first order models including large process delays.
3. Offline method.
4. Approximations for the K_c, τ_i, and τ_d values might not be entirely accurate for different systems.

These are other common methods that are used, but they can be complicated and aren't considered classical methods, so they are only briefly discussed.

3.6.1 Internal Model Control

The Internal Model Control (IMC) method was developed with robustness in mind. The Ziegler-Nichols open loop and Cohen-Coon methods give large controller gain and short integral time, which isn't conducive to chemical engineering applications. The IMC method relates to closed-loop control and doesn't have overshooting or oscillatory behavior. The IMC methods however are very complicated for systems with first order dead time.

3.6.2 Auto Tune Variation

The auto-tune variation (ATV) technique is also a closed loop method and it is used to determine two important system constants (P_u and K_u for example). These values can be determined without disturbing the system and tuning values for PID are obtained from these. The ATV method will only work on systems that have significant dead time or the ultimate period, P_u, will be equal to the sampling period.

3.7 Examples

3.7.1 Worked out Example 1
Problem

You're a controls engineer working for Flawless Design company when your optimal controller breaks down. As a backup, you figure that by using coarse knowledge of a classical method, you may be able to sustain development of the product. After adjusting the gain to one set of data taken from a controller, you find that your ultimate gain is 4.3289.

From the adjusted plot below, determine the type of loop this graph represents; then, please calculate K_c, T_i, and T_d for all three types of controllers.

Sinusoidal Curve

Solution

From the fact that this graph oscillates and is not a step function, we see that this is a closed loop. Thus, the values will be calculated accordingly.

We're given the Ultimate gain, $K_u = 4.3289$. From the graph below, we see that the ultimate period at this gain is $P_u = 6.28$

Sinusoidal Curve

From this, we can calculate the K_c, T_i, and T_d for all three types of controllers. The results are tabulated below. (Results were calculated from the Ziegler-Nichols closed-loop equations.)

				K_c	T_I	T_D
K_U	4.3289		P	2.1645	-	-
P_U	6.28		PI	1.9677	5.2333	-
			PID	2.5464	3.1400	0.7850

3.7.2 Worked out Example 2
Problem

Your partner finds another set of data after the controller breaks down and decides to use the Cohen-Coon method because of the slow response time for the system. They also noticed that the control dial, which goes from 0-8, was set at 3 instead of 1. Luckily the response curve was obtained earlier and is illustrated below. From this data he wanted to calculate K_c, T_i and T_d. Help him to determine these values. Note that the y-axis is percent change in the process variable.

Solution

In order to solve for K_c, T_i and T_d, you must first determine L, ΔC_p, and T. All of these values may be calculated by the reaction curve given.

From the process reaction curve we can find that:
L = 3
T = 11

$\Delta C_p = 0.55$ (55%)

Now that these three values have been found N and R may be calculated using the equations below.

$$N = \frac{\frac{\Delta C_p}{T}}{\frac{L}{T}} = \frac{NL}{\Delta C_p}$$

R =

Using these equations you find that
N = .05
R = 0.27
We also know that since the controller was moved from 1 to 3, so a 200% change.
P = 2.00

We use these values to calculate K_c, T_i, and T_d, for the three types of controllers based on the equations found in Table 3.

	Kc	Ti	Td
P	14.53		
PI	12.30	6.42	
PID	18.68	6.65	1.04

3.7.3 Tuning the parameters of a PID controller on an actual system

There are several ways to tune the parameters of a PID controller. They involve the following procedures. For each, name the procedure and explain how the given measured information is used to pick the parameters of the PID controller.

Questions
a. The controller is set to P only, and the system is operated in "closed-loop", meaning that the controller is connected and working. The gain is tuned up until a resonance is obtained. The amplitude and frequency of that resonance is measured.

b. The system is kept in "open-loop" mode, and a step-function change is manually made to the system (through a disturbance or through the controller itself). The resulting response of the system is recorded as a function of time.

Answers
a. We will use the Ziegler-Nichols method.
Ki=0.5Ku

702

Ku is the ultimate gain when the system started oscillating.

b. We will use the Cohen-Coon method.
We can locate the inflection point of the step function and draw a tangent. T_{dead} is located at the crossing of that tangent with t, and T is located at the cross of the tangent with M(t)

3.7.4 Multiple Choice Question 1

Which of the following do you RECORD during the Ziegler-Nichols Method?

    a. K_c
    b. τ_i
    c. K_o
    d. τ_d

Answer:C

3.7.5 Multiple Choice Question 2

For the Ziegler-Nichols Method, it is important to:

    a. Find a gain that produces damped oscillation
    b. Set P and I controllers to zero
    c. Record the period of oscillation
    d. Calculate T_c

Answer:A,C

3.8 Sage's Corner
http://video.google.com/googleplayer.swf?docId=-5436955221139210544
Powerpoint Slides without narration

http://www.youtube.com/v/8ZJGErQwGAQ
Powerpoint Slides without narration

3.9 References

- Svrcek, William Y., Mahoney, Donald P., Young, Brent R. A Real Time Approach to Process Control, 2nd Edition. John Wiley & Sons, Ltd.
- Astrom, Karl J., Hagglund, Tore., Advanced PID Control, ISA, The Instrumentation, Systems and Automation Society.
- "ACT Ziegler-Nichols Tuning," http://ourworld.compuserve.com/homepages/ACTGMBH/zn.htm
- Ogata, Katsuhiko. System Dynamics, 4th Edition. Pearson Education, Inc.

Section 4. PID tuning via Frequency Responses w/ Bode Plots

written by: Tony Martus, Kegan Lovelace, Daniel Patrick, Merrick Miranda, Jennifer DeHeck, Chris Bauman, Evan Leonard
Edited by: Alfred Chung (Derivations), Ran Li (Determining Stable Controller Gain), Nirala Singh (Determining Slope of Bode Plot), Katy Kneiser and Ian Sebastian (synthesis with 2006 wiki "Bode Plots"), Robert Appel, Jessica Rilly and Jordan Talia (Formatting)

4.1 A Description of Amplitude Ratio, Phase Shift and Frequency Response

A good way to understand the physical meaning of amplitude ratio and phase shift and how they change as a function of the frequency of the input stream is through the analysis of a simple system. A well-insulated continuously-stirred tank reactor (CSTR) with an input flow equal to its output flow is a good example (Figure 1). For this description we consider inlet and outlet temperature, but inlet concentration or feed rate could also be used.

Figure 1: CSTR example

If the CSTR is initially at 55°C, and the inlet flow has a temperature that fluctuates sinusoidally between 50°C and 60°C, the outlet temperature will also vary sinusoidally. If the fluid has a large residence time in the tank, and the frequency of the temperature variation of the input fluid is high relative to the residence time. Then the outlet temperature will also vary quickly in a sinusoidal fashion, but with a significantly smaller amplitude than that of the inlet. The large hold up in the well mixed tank dampens the fluctuations in the inlet temperature. In contrast, if the the frequency of the inlet

704

temperature sine wave is low relative to the fluid residence time, the output will also be a sine wave with temperatures ranging from 50°C to 60°C.

In both cases described above there will be a lag or phase shift between the input and the output of the system (see the figure immediately below). In our example, the phase shift is controlled by the average residence time of fluid in the tank. The time constant of this system is the average residence time of the tank. A system with a small residence time will respond very quickly to changes in input flow rate, and large residence times will result in sizable lag time.

As frequency of the temperature variation increases, the period of the input decreases and the phase shift becomes larger. This can be seen in the following graphs: </br>

Solid line = input, dashed line = output (graphs adapted from Bequette)

Imagine that we have a sensitive reaction that must be maintained with 3°C of 55°C, it now becomes of utmost importance that we understand the frequency response of this

705

CSTR system. Herin lies the utility of Bode plots; they depict a range of different frequency responses in two plots. This allows a relatively rapid determination of system robustness. In addition, frequency response and Bode plot analysis can be used to tune PID control systems.

4.2 Frequency Response

For a given process described by $Y(t) = \hat{G}X(t)$, one considers a sinusoidal input

$$X(t) = sin(\omega t)$$

After a long time, the solution will also be sinusoidal, with

$$Y(t) \sim ARsin(\omega t + \phi)$$

where AR is the amplitude ratio (must be a positive number) and ϕ is the phase lag ($\phi < 0$) or phase lead ($\phi > 0$).

Sometimes, to solve for $Y(t)$ using an ODE solver, it is useful to have the process equation in the form $X(t) = \hat{G}^{-1}Y(t)$ and define $\hat{G}^{-1}$ instead of $\hat{G}$.

4.2.1 Expressions for AR and phi
First-Order System

$$\hat{G}^{-1} = \tau_p(\partial/\partial t) + 1$$

$$AR = \frac{1}{\sqrt{1 + \omega^2\tau_p^2}}$$

$$\phi = \tan^{-1}(-\omega\tau_p)$$

Second-Order System

$$\hat{G}^{-1} = \tau_p^2(\partial/\partial t)^2 + 2\zeta_p\tau_p(\partial/\partial t) + 1$$

706

$$AR = \frac{1}{\sqrt{(1 - \omega^2 \tau_p^2)^2 + (2\zeta\omega\tau_p)^2}}$$

$$\phi = \tan^{-1}\left(\frac{-2\zeta_p\omega\tau_p}{1 - \omega^2\tau_p^2}\right)$$

Dead Time

$$AR = 1$$
$$\phi = -\omega\tau_{dead}$$

Systems in series

$$AR = AR_1 \cdot AR_2 \cdot ...$$
$$\log AR = \log AR_1 + \log AR_2 + ...$$
$$\phi = \phi_1 + \phi_2 + ...$$

P Controller

$$\hat{G}_c = K_c$$
$$AR = K_c$$
$$\phi = 0$$

PI Controller

$$\hat{G}_c = K_c\left(1 + \frac{1}{\tau_I}\int_0^t dt'\right)$$

$$AR = K_c\sqrt{1 + 1/(\omega\tau_I)^2}$$

$$\phi = \tan^{-1}\left(-\frac{1}{\omega\tau_I}\right)$$

PD Controller

$$\hat{G}_c = K_c(1 + \tau_D(\partial/\partial t))$$
$$AR = K_c\sqrt{1 + \omega^2\tau_D^2}$$
$$\phi = \tan^{-1}(\omega\tau_D)$$

PID Controller

$$\hat{G}_c = K_c(1 + \frac{1}{\tau_I}\int_0^t dt' + \tau_D(\partial/\partial t))$$
$$AR = K_c\sqrt{1 + \omega\tau_D - 1/(\omega\tau_I)^2}$$
$$\phi = \tan^{-1}(\omega\tau_D - \frac{1}{\omega\tau_I})$$

4.3 Introduction to Bode Plots

Because it's impossible to perfectly model a real chemical process, control engineers are interested in characterizing the robustness of the system – that is, they want to tune controllers to withstand a reasonable range of change in process parameters while maintaining a stable feedback system. Bode (pronounced boh-dee) plots provide an effective means to quantify the system's stability. Bode plots describe an open or closed-loop system as a function of input frequency and give a picture of the system's stability. If all inputs into a system were constant, it would be a relatively simple task to control the system and its output. However, if an input value, such as temperature, varies sinusoidally, then the output should also be describable as a sinusoidal function. In this case Bode plots are a useful tool for predicting the response of the system. There are two Bode plots used to describe a system. The first shows amplitude ratio as a function of frequency. The amplitude ratio is the amplitude of the output sine wave divided by the input sine wave. The second plot graphs phase shift as a function of frequency, where phase shift is the time lag between the output and the input sine curves divided by the period of either of the sine curves and then multiplied by -360° (or -2π radians) to express this shift as an angle.

4.4 Constructing Bode Plots

Bode plots concisely display all relevant frequency input and output information on two plots: amplitude ratio as a functions of frequency and phase shift as a function of frequency. The amplitude ratio plot is a log-log plot while the phase angle plot is a semilog (or log-linear) plot.

708

To construct a Bode plot, an engineer would have empirical data showing input and output values that vary as sinusoidal functions of time. For instance, there might be inlet temperature data that varies sinusoidally and the outlet temperature data that also varies sinusoidally. For example, given a heat exchanger that is used to cool a hot stream, we can vary the flow rate of the hot stream and model it sinusodially. (This can be done using ODEs & an Excel model of a heat exchanger.)

Figure 2 below includes inlet flowrate (process fluid stream) varying as a sinusoidal function and the hot temperature exiting the heat exchanger likewise varying as a sinusoidal function. Portions of this graph are not yet at steady-state.

Figure 2: Inlet Flowrate modeled by $F_s = 0.1 + 0.1\sin(\dfrac{\pi}{4t})$.

Bode plots are constructed from steady-state data. Figure 3 shows part of the steady-state region of the same data used for Figure 2.

Figure 3: Steady-state inlet flow rate and outlet temperatures varying as sunusoidal functions.

To collect a single data point for a Bode plot we will use the information from a single period of the inlet flow rate and the corresponding temperature from the hot exiting stream.

The amplitude ratio, AR, is the ratio of the amplitude of the output sinusoidal curve divided by the amplitude of the input sinusoidal curve.

$$AR = \frac{output\,amplitude}{input\,amplitude}$$

The value of the amplitude ratio should be unitless so if the units of the input frequency and the units of the output frequency are not the same, the frequency data should first be normalized. For example, if the input frequency is in °C/min and the output frequency is also in °C/min, then AR is unitless and does not need to be normalized. If, instead, the input frequency is in L/min and the output frequency is in °C/min, then AR = °C/liters. In this case the inlet and outlet frequencies need to be normalized because the ratio AR =

710

°C/L doesn't say anything about the physical meaning of the system. The value of AR would be completely different if the units of AR were Kelvin/gal.

To find the phase shift, the periods of the input and output sine curves need to be found. Recall that the period, P, is the length of time from one peak to the next.

$$P = \frac{1}{f} = \frac{2\pi}{\omega}$$

$$f = frequency$$

$$\omega = frequency(rad/sec)$$

The phase shift is then found by

$$\phi = -2\pi(\Delta P/P)$$

Using these values found from multiple perturbations in feed flow rate it is possible to construct the following Bode plots:

711

Figure 4: Bode Plot (Amplitude Ratio) [Log-Log Plot]

Figure 5: Bode Plot (Phase Shift) [Semi-Log]

4.4.1 Rules of Thumb when analyzing Bode plots

Generally speaking, a gain change shifts the amplitude ratio up or down, but does not affect the phase angle. A change in the time delay affects the phase angle, but not the amplitude ratio. For example, an increase in the time delay makes the phase shift more negative for any given frequency. A change in the time constant changes both the amplitude ratio and the phase angle. For example, an increase in the time constant will decrease the amplitude ratio and make the phase lag more negative at any given frequency.

Table 1. Rules of Thumb

Change	Effect
Gain Change	Amplitude ratio is shifted up or down; no change in phase angle
Time Delay	Phase angle changes, but no change in amplitude ratio

Prior to the advent of powerful computer modeling tools, controls engineers would model systems using transfer functions. Readers interested in learning more about how these were used to construct Bode plots should refer to Bequette's Process Control: Modeling, Design, and Simulation (see References). This wiki assumes that the engineer already has data in Excel, etc, that shows the sinusoidal behavior of input and outputs.

4.5 Bode Stability Criterion

A control system is unstable if the open-loop frequency response exhibits an AR exceeding unity at the frequency for which the phase lag is 180 degrees. *This frequency is called the* crossover frequency. *Adjusting the gain so that the AR < 1 for - phi < (180 minus phase margin) results in acceptable stability if the phase margin is greater than 30 degrees.*

As previously mentioned, the controls engineer uses Bode plots to characterize the stability of the system. If you were to find that the amplitude ratio is greater than 1 the system would be unstable at that frequency. An important stability criterion is that AR < 1 when phase shift = -180 degrees (or -pi radians). This is called the Bode stability criterion and if it holds for a closed-loop system, then that system will be stable. If this criterion is not satisfied and you were to put a feedback controller on this process, as done in PID tuning via optimization [PID_HeatExchange.xls – Note there are still errors in this file and a link will be provided once the file is operational.], the process would explode due to the response lag. The reason for this is that the system might think temperature is too low when it's actually to high, so it would decrease cooling water rather than increase it. Also, with an AR < 1, an error in input is amplified in the output, so the system will lose control due to amplification of error. This could obviously be a dangerous situation, for example in an exothermic CSTR, which could explode. This can also be seen by following the procedure above for the PID_HeatExchange.xls file.

4.6 Frequency Response in Mathematica

Bode plot for PD controller, $T_D = 1$:

AR = Sqrt[1 + omega^2]

phi = 180/Pi*ArcTan[omega]

LogLogPlot[AR,{omega,0.01,100}]

LogLinearPlot[phi,{omega,0.01,100}]

```
In[5]:= AR = Sqrt[1 + omega^2]
        phi = 180 / Pi * ArcTan[omega]
        LogLogPlot[AR, {omega, 0.01, 100}]
        LogLinearPlot[phi, {omega, 0.01, 100}]
```

$$Out[5]= \sqrt{1 + omega^2}$$

$$Out[6]= \frac{180 \, ArcTan[omega]}{\pi}$$

Out[7]=

Out[8]=

Bode plot for PID controller, using the \Manipulate" function to watch the effect of varying the values of both Tau_I and Tau_D:

Manipulate[LogLogPlot[Sqrt[1 + (omega*tauD - 1/(tauI*omega))^2],{omega,0.01,100}],{tauI,0.1,10},{tauD,0,10}]

Manipulate[LogLinearPlot[180/Pi*ArcTan[omega*tauD - 1/(tauI*omega)],{omega,0.01,100}],{tauI,0.1,10},{tauD,0,10}]

Bode plot for open-loop operation of third-order system with a P controller:

716

AR = Sqrt[1/(1+omega^2)]*Sqrt[1/(1+omega^2)]*Sqrt[1/(1+25*omega ^ 2)] phi = 180/Pi*(ArcTan[-omega] + ArcTan[-omega] + ArcTan[-5*omega]) LogLogPlot[AR,{omega,0.01,100}] LogLinearPlot[phi,{omega,0.01,100}]

In[11]:= AR = Sqrt[1 / (1 + omega^2)] * Sqrt[1 / (1 + omega^2)] * Sqrt[1 / (1 + 25 * omega^2)]

phi = 180 / Pi * (ArcTan[-omega] + ArcTan[-omega] + ArcTan[-5 * omega])

LogLogPlot[AR, {omega, 0.01, 100}]

LogLinearPlot[phi, {omega, 0.01, 100}]

Out[11]= $\dfrac{\sqrt{\dfrac{1}{1+25\,omega^2}}}{1 + omega^2}$

Out[12]= $\dfrac{180\,(-2\,ArcTan[omega] - ArcTan[5\,omega])}{\pi}$

Out[13]=

Out[14]=

717

Find crossover frequency:

FindRoot[phi==-180,{omega,1}]

This yields omega = 1.18322 which is the crossover frequency, and can be used to determine the amplitude ratio (ask mathematica for AR) the gain (Kc=1/AR)

To determine what the gain should be set to at a phase margin of 30 degrees (safe operating conditions) use the following in mathematica:

FindRoot[phi==-150,{omega,1}]

The phase margin of Cohen-Coon and other frequency analysis (such as Zielger-Nicholas) can be calculated by using the gain (Kc) value gotten from the method a typing the following into mathematica:

NSolve[Kc*AR==1,omega]

[Click here to learn more about Bode Plots](#)

4.7 Derivations

Assume a first-order process with input $X(t)$, output $Y(t)$, and operator

$$\hat{G}^{-1} = (\tau \frac{\delta}{\delta t} + 1)$$

where

$$X(t) = \hat{G}^{-1}Y(t) = (\tau \frac{\delta}{\delta t} + 1)Y(t) = \tau Y'(t) + Y(t)$$

(1)

For Bode Plot Analysis, we assume a periodic input to the system: $X(t) = sin(\omega t)$

Based on this input, we can hypothesize that: $Y(t) = AR * sin(\omega t + \phi)$

From this hypothesis, we can replace equation (1) with our functions for $X(t)$ and $Y(t)$. This is shown below.
$X(t) = \tau Y'(t) + Y(t)$ (1)

718

$$sin(\omega t) = \tau \frac{\delta}{\delta t}(ARsin(\omega t + \phi)) + ARsin(\omega t + \phi)$$
(2)

$sin(\omega t) = AR(\tau\omega cos(\omega t + \phi)) + sin(\omega t + \phi)$ (3)

Using trigonometric identities, we can replace $cos(\omega t + \phi)$ with $cos(\omega t)cos(\phi) - sin(\omega t)sin(\phi)$ and $sin(\omega t + \phi)$ with $sin(\omega t)cos(\phi) + cos(\omega t)sin(\phi)$ to get:

$sin(\omega t) = AR(\omega\tau(cos(\omega t)cos(\phi) - sin(\omega t)sin(\phi)) + sin(\omega t)cos(\phi) + cos(\omega t)sin(\phi))$ (4)

To solve for AR and ϕ, we can match the coefficients for the $sin(\omega t)$ term and $cos(\omega t)$ term on the left and right side of equation (4).

Coefficient of $cos(\omega t)$: $0 = AR(\omega\tau cos(\phi) + sin(\phi))$ (5)

Coefficient of $sin(\omega t)$: $1 = AR(-\omega\tau sin(\phi) + cos(\phi))$ (6)

From equation (5), we can divide both sides by $ARcos(\phi)$ and rearranging to get:

$$tan(\phi) = -\omega\tau \quad \text{or} \quad \phi = tan^{-1}(-\omega\tau)$$
(7)

Based on this relationship, we can draw a triangle with the angle ϕ and the relationship shown in equation (7). A picture of the triangle is shown below.

From equation (6) and the triangle above, we can solve for AR:

$$AR = \frac{1}{-\omega\tau(sin(\phi)) + cos(\phi)}$$

$$AR = \frac{1}{-\omega\tau(\frac{-\omega\tau}{\sqrt{1+\omega^2\tau^2}}) + \frac{1}{\sqrt{1+\omega^2\tau^2}}}$$

$$AR = \frac{1}{\sqrt{1 + \omega^2\tau^2}} \quad (8)$$

4.8 Determining Stable Controller Gain and Phase Margin

Tuning via phase margin is a more precise method than tuning through Cohen-Coon Rule and Ziegler-Nichols Tuning Rule. We can find the gain and phase margin by using bode plots or by an analytical method.

Graphical Method

In order to find a safe controller gain, use the following steps:
1) Construct Bode Plot of LogAR versus Log $\omega\tau$ and ϕ versus Log $\omega\tau$
2) Use the above plot to find $\omega\tau$ such that $\phi = $ -180° + ϕ_{PM}.
3) Use $\omega\tau$ found from above step to find AR in LogAR versus Log $\omega\tau$ plot.

4) $$K_{Csafe} = \frac{1}{AR}$$

In order to find the phase margin, use the following steps:
1) Construct Bode Plot of LogAR versus Log $\omega\tau$ and ϕ versus Log $\omega\tau$
2) Use the above plot to find $\omega\tau$ such that AR=1.
3) Use $\omega\tau$ found from above step to find ϕ in ϕ versus Log $\omega\tau$ plot.
4) $\phi_{PM} = 180 - | \phi_{AR=1} |$

Analytical Method
Finding the gain with a phase margin:
1) Use your expression for Φ, set Φ to the desired value and find the root for ω.
2) Input the value obtained for ω in your expression for AR.

3) $$K_C = \frac{1}{AR}$$

Finding the phase margin with a gain:

720

$$K_C = \frac{1}{AR}$$

1) Use your expression for AR, set $K_C = \frac{1}{AR}$ and solve for ω.

2)Input the value obtained for ω in your expression for Φ.

4.9 Determining the Slope of a Bode Plot

Sometimes it is necessary to determine the relative shape of a Bode Plot without the use of a computer or ODE equation solver. In sketching the relative shape, there are three main parts of a basic Bode Plot that must be considered.

1. Crossing point

2. Starting value

3. Slope

4.9.1 Crossing Point

The crossing point of the Bode Plot is when the frequency, ω, is equal to $1/\tau$. If the x-axis of the Bode Plot is $\omega\tau$, the crossing point will be at x-axis $= 1$. If there are multiple values of τ, there may be multiple crossing points, each at the frequency where ω is equal to $1/\tau$. If there are multiple τ values, and the x-axis is $\omega\tau$, it must be specified which τ the x-axis is referring to.

4.9.2 Starting Value

The starting value depends on the K_C and K_P values contributing to the process.

4.9.3 Slope

The slope for an uncontrolled process is equal to the negative of the order of the process. For example, if there is a first-order process with no controller the slope of the Bode Plot would be -1 (-1 for first order), after the crossing point has been reached. A picture of this Bode Plot is shown below.

For a process controlled by a PD controller, the AR is essentially the inverse of the first-order system AR, meaning that the slope addition from a PD controller is a +1, instead of a -1 as in a first-order process. For a second-order system with a PD controller the final slope will be -1 (-2 from the second order process, +1 from PD). However, there may be multiple crossing points, whose location depends on the value of τ_P and τ_D. Remember that the crossing point is where $\omega\tau$ equals one. Since the slope contribution only comes after the crossing point, a Bode Plot for τ_D less than τ_P would look like the picture below.

Crossing point 1 is where $\omega\tau_P$ is 1 and crossing point 2 is where $\omega\tau_D$ is 1.

Similarly, if τ_D is greater than τ, the slope would first go to +1, then to -1.

For a PI controller, the slope contribution from the controller comes before the crossing point, and then goes to zero after the crossing point. A PID controller would therefore look like the picture shown below, assuming $\tau_D = \tau_I$.

4.10 Example 1 – Determine the amplitude ratio

You are a controls engineer and wish to characterize a heat exchanger used in a chemical process. One of the many things you are interested in knowing about the system is how the hot outlet temperature responds to fluctuations in the inlet flow rate. Using data for a particular inlet flow rate, you graphed normalized (why?) flow rate and normalized hot outlet temperature vs. frequency (rad/min). Use this graph to determine the amplitude ratio.

Solution:

Because flow rate units and temperature do not have the same units, these values needed to be normalized before calculating the amplitude ratio. To normalize flow, use the following equation:

$$Fnorm = \frac{F - Fmin}{Fmax - Fmin}$$

where F is the flow rate of a particular data point.

In this problem temperature, T, is in °C, so

$$Tnorm = \frac{T - 273.15}{100}$$

To find the amplitude of both wave functions, first recall that the amplitude of a wave is the maximum height of a wave crest (or depth of a trough). For one steady-state wave produced from a column of values in Excel, you could quickly calculate the amplitude by using the max( ) and min( ) Excel functions. [This can be found using Excel help.] If you opted to find the amplitude this way, then the amplitude for a single wave function would be

$$Amplitude = \frac{[max() - min()]}{2}$$

Note that this is just one way to find the amplitude of a wave. It could also be found simply by reading the values off the graph of a wave.

Once the amplitudes of the inlet and outlet waves are found, use the values to find the amplitude ratio, AR, as follows:

$$AR = \frac{outlet\,streams\,amplitude}{inlet\,streams\,amplitude} \underset{=}{\frac{0.0486}{0.499}} = 0.0974$$

724

The following graph shows how inlet flow and both the hot and cold outlet temperatures vary as sinusoidal functions of time. This graph was generated using the same data for the heat exchanger of Example 1. Use this graph to find the phase shift between the inlet flow and the hot outlet stream.

Solution:

Determine the period (P) – This can be done by finding the time from one peak to the next of a given wave. In this case, we want to know the period of the inlet flow rate, so P = 1.14s.

Determine the lag (delta P) – This can be done by finding the time difference between the peak of the inlet flow rate and the peak of the hot outlet stream. (Remember that the hot outlet wave lags the wave of the inlet flowrate).

$deltaP = 0.87s$

$$Phaseshift = \frac{0.87s}{1.14s} * -2pi = -4.80$$

Since we're only concerned with time values for finding the phase shift, the data doesn't need to be normalized.

4.12 References

- Bequette, B. Wayne. Process Control: Modeling, Design, and Simulation. pp 172-178, 214-235. Upper Saddle River, N.J. Prentice Hall, PTR 2003.
- Liptak, Bela G. Process Control and Optimization. Volume 2. Instrument Engineers Handbook. Fourth Edition. CRC Press, Boca Raton Fl. 2003. *Insertformulahere*

Section 5. PID tuning via optimization

Note: ⊞Video lecture available for this section!

Authors: Andrew MacMillan, David Preston, Jessica Wolfe, Sandy Yu

Stewards: YooNa Choi, Yuan Ma, Larry Mo, Julie Wesely

Date Presented: October 19, 2006; Revised: October 18, 2007

- First round reviews for this page
- Rebuttal for this page

5.1 Introduction

Tuning a controller is a method used to modify the effect a process change will have on the piece of equipment being controlled. The goal of tuning a system is to construct the most robust process possible. The method chosen for tuning a system varies depending on the parameter being measured, the sensitivity of the materials, the scale of the process, and many other variables unique to each process. This chapter discusses the basics of tuning a controller using predictive methods. To learn more about using an effect based method see the Classical Tuning section.

5.2 Optimization

When tuning a PID controller, constants K_c, T_i, and T_d need to be optimized. This can be accomplished using the following equation, Equation 1, from the Classical Tuning section:

$$MV = K_c \left(e(t) + \frac{1}{T_i} \int_0^t e(\tau)\, d\tau + T_d \frac{de(t)}{dt} \right)$$ Equation 1

MV - manipulated variable (i.e. valve position) that the controller will change

K_c - accounts for the gain of the system

T_i - accounts for integrated error of the system

T_d - accounts for the derivative error of the system

Tuning by optimization uses computer modeling programs, such as Microsoft Excel, to

find the optimal values for the coefficients K_c, T_i, and T_d to yield the minimum error. The Solver function in Excel can be used in this situation.

For more information about the use of these parameters and their overall effects on the control system see the [P, I, D, PI, PD, and PID control](#) section.

5.3 Excel Modeling

For instructions on how to install Solver in Excel, see the [Adding in the Solver Application in Excel 2007](#) section.

For instruction on how to use Solver in Excel, see the [Excel's Solver Tool](#) section.

Equation 1 can be used in Excel to optimize the values of K_c, T_i, and T_d. The following are the steps to optimize these constants:

- Pick a set point. Examples of set points include the temperature at which a reaction is expected to stay and the flow rate of a cooling stream.
- Make a column for error. Create a function in each cell so that the error is equal to the set point minus the actual value.
- Make a column that will calculate Equation 1.
- At the end of the error column, use the sum function to add up all of the errors in the column.
- Make three cells for K_c, T_i, and T_d, and insert the initial values.
- Open the Solver function in Excel.
- In Solver, set the target cell to the summed errors cell. Set the target cell equal to the minimum value possible. Set the changing cells to K_c, T_i, and T_d.
- Click 'Solve' to minimize the sum of the error and therefore optimize the K_c, T_i, and T_d constants.

Tips for using Excel Solver to optimize PID controller parameters:

- When choosing initial values for your parameters, try starting with $K_c=1$, $T_i=1000$, and $T_d=1$. Since Solver only does a limited number of iterations (usually set at 100), it may settle on an incorrect answer.
- Change the values for K_c manually and optimize with Solver until you get the lowest value in the target cell. This is the most efficient as K_c usually has the largest effect on a PID controller.
- Make sure to set constraints on your parameters in the 'Subject to the Constraints:' box in Solver (e.g. $K_c > 0$).
- Make sure you incorporate any physical limitations (e.g. maximum heater temperature) into your Excel model. Even though your system may not be close to the limit initially, Solver may try and go past that limit if it is not programmed into your model.

5.4 Industrial Application

The parameters (K_c, T_i, T_d) found through optimization will not necessarily give the best control. To optimally tune a controller, a technician must adjust equipment to tune the process. So why is it necessary to tune by optimization? The parameters found by optimization give an accurate starting point for the technician.

5.5 Example of PID Tuning by Optimization

This example is based on the example in the CSTR Heat Exchange Model section.

The goal is to optimize the PID controller for the coolant temperature.

729

Follow the fully worked out example at: [Media:Optimal PID.xls Optimal PID]

As shown in the Excel file, the coolant temperature changes in response to variations from the set point by a PID type controller. Changing the K_c, T_i, and T_d values using Solver, to minimize the total error, provided the values in the green box. On the second sheet, labeled 'disturbances', it can be seen that once optimized, the parameters should be fit to the system, regardless of changes in operating conditions.

It is important to note that the starting values you choose for the PID parameters will greatly affect your final results from using an excel solver. So you should always use your intuition to judge whether your starting values are reasonable or not. According to ExperTune , starting PID settings for common PID control loops are:

Loop Type: Flow ; P = 50 to 100 ; I = 0.005 to 0.05 ; D = none ;

Loop Type: Liquid Pressure ; P = 50 to 100 ; I = 0.005 to 0.05 ; D = none

734

Loop Type: Gas Pressure ; P = 1 to 50 ; I = 0.005 to 0.05 ; D = 0.02 to 0.1 ;

Loop Type: Liquid Level ; P = 1 to 50 ; I = 1 to 100 ; D = 0.01 to 0.05 ;

Loop Type: Temperature ; P = 2 to 100 ; I = 0.2 to 50 ; D = 0.1 to 20 ;

The above values are rough, assume proper control loop design, ideal or series algorithm and do not apply to all controllers. These information should only be used as a possible consideration and should not be taken as an absolute starting value for all PID controls. For more information about the above values please go to http://www.expertune.com/tutor.html

As mentioned before, to optimize the parameters in a PID-controlled system the solver function can be used in excel. However the solver function can sometimes run into its own problems because of the process it uses to solve for these "optimal" values. To ensure that solver gives you the correct optimal values, some manual "optimization" is needed. Using your excel sheet, the controller on the system can be separated into its different components P, I, and D. The behavior of each component can be monitored as the parameters are changed manually. Theoretically you would choose values that have every component reach steady-state over time. The excel solver can then be used as a precision tool. If your values are already closer to the optimal values, then solver should have no problem making them more precise. It is recommended that the graphs of the components be checked after using solver with your initial guesses. You may not have to go through this long and most of the time painful process. This is an example of how the P, I, and D plots change with different parameters. Note that the combined PID plot will also reach steady-state if each of the components reach steady-state.

(Parameters are not optimized)

(Optimized parameters)

5.6 Example: Optimization of a Heat Exchanger

This example is based on the ODE Modeling in Excel of the Heat Exchange Model.

Find the best values of K_c, T_i, and T_d in the given spreadsheet Media:PID_HeatExchange.xls.

Note the three different tabs:

- the first tab shows the problem to be solved
- the second is a step by step method for creating this sheet on your own
- the third is the answer to the problem

After using Solver:

- Click through the cells in one row to determine which parameters are affected by MV changes
- Change the values of K_c, T_i, and T_d to see how they will effect the overall temperature profile. Which parameter has the greatest effect? Based on the equation, does this make sense?

736

5.7 Multiple Choice Question

What do you want to minimize when optimizing the constants K_c, T_i, and T_d?

   a. time
   b. response period
   c. error
   d. nothing

5.8 Answers

5.8.1 Example

Look at the third tab on the Excel file to find the optimization answer. K_c (gain) will have the greatest effect on the output of the controller.

5.8.2 Multiple Choice Question

c) error

Tuning by optimization uses computer modeling programs to find the optimal values for the coefficients Kc, Ti, and Td to yield the *minimum error*.

5.9 Sage's Corner

A walkthrough of the Heat Exchanger Example above

```
http://video.google.com/googleplayer.swf?docId=8598561635696386040
```

slides for this talk

5.10 References

- Svrcek, William Y., Mahoney, Donald P., Young, Brent R. "A Real Time Approach to Process Control", 2nd Edition. John Wiley & Sons, Ltd.

Section 6. PID downsides and solutions

See also [ECOSSE](#)

Title:PID Downslides & Solutions

Note: Video lecture available for this section!

Authors: Ashwini Miryala, Kyle Scarlett, Zachary Zell, Brandon Kountz

Stewards: Brian Hickner, Lennard Gan, Addison Heather, Monique Hutcherson

Date Presented: 10/19/2006; Revised: 10/8/2007

- First round reviews for this page
- Rebuttal for this page

6.1 Introduction

A proportional-integral-derivative (PID) controller is one of the most common algorithms used for control systems. It is widely used because the algorithm does not involve higher order mathematics, but still contains many variables. The amount of variables that are used allows the user to easily adjust the system to the desired settings. The algorithm for the PID uses a feedback loop to correct the difference between some measured value and the setpoint. It does this by calculating and outputting some action that will correct this error in the system. A PID controller has a proportional, integral and a derivative control which handles the current, past and predicted future of the signal error. For more information about PID, please refer to PID Intro. The PID controller can operate systems that run in a linear or nonlinear fashion. Tuning processes are done to the controller to tackle the possible nonlinear system. Limitations arise within the system because tuning is limited to only three different parameters (proportional, integral, and derivative controls). Additional information on tuning of PID can be found at [1] or [2]. The most common limitations that occur within the PID control specifically involve the integral control. The following article addresses some of the common limitations faced by each control type, with an emphasis on the integral control, and some solutions to overcome each of these limitations.
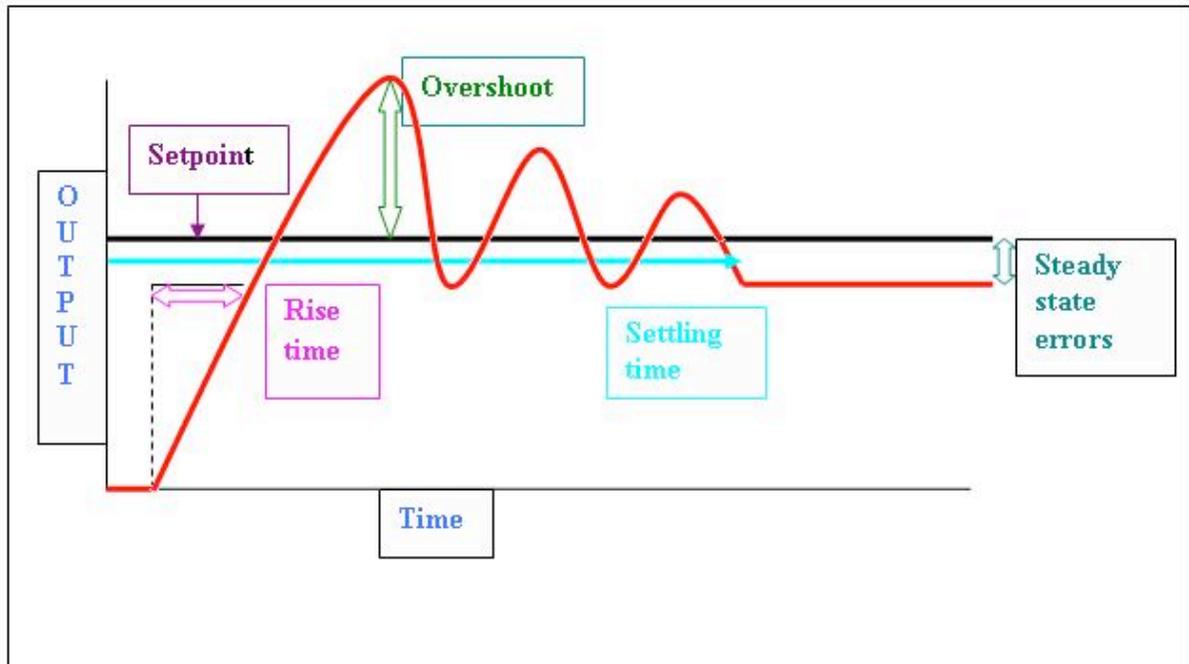
6.2 Proportional Control

The main purpose of the proportional control is minimize the fluctuations that occur within the system.
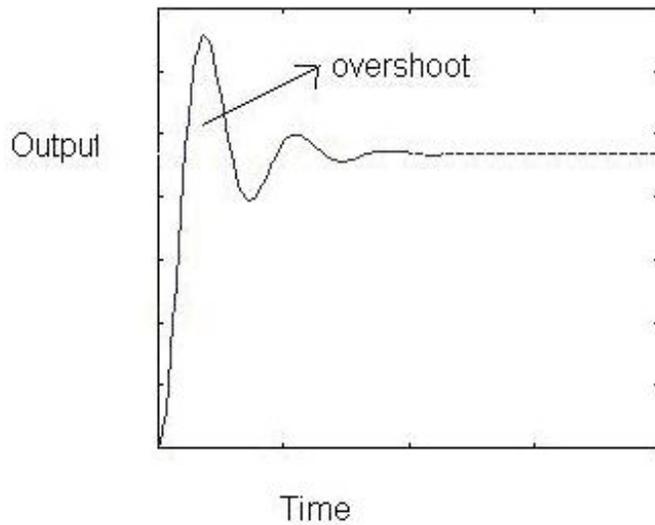
6.2.1 Limitations

The P-controller usually has steady-state errors (the difference in set point and actual outcome) unless the control gain is large. As the control gain becomes larger, issues arise with the stability of the feedback loop. For instance, reducing the rise time implies a high proportional gain, and reducing overshoot and oscillations implies a small proportional gain. This is not possible to achieve in all systems.

Below is a general process outcome diagram showing the terminology used above.

Below is a sample diagram of process output of proportional control.

The way to eliminate these steady-state errors is by adding an integral action. The integral term in the equation drives the error to zero. Higher Integral constant ($1 / T_t$) drives the error to zero sooner but also invites oscillations and instability. Read on the integral control section below to know more about limitations associated with this integral term.

Below is a sample process output diagram when integral control is added.

The above picture shows the reduction of overshoots and oscillations compared to the picture before adding the integral action.

6.3 Integral Control

The contribution from the integral term is proportional to both the magnitude of the error and the duration of the error. Summing the instantaneous error over time (integrating the error) gives the accumulated offset that should have been corrected previously. The accumulated error is then multiplied by the integral gain and added to the controller output. The magnitude of the contribution of the integral term to the overall control action is determined by the integral gain, Ki.

The integral term is given by:

$$D_{out} = K_d \frac{de}{dt}$$

Where

- D_{out}: **Derivative output**

- K_d: **Derivative Gain**, a tuning parameter
- e: **Error** = $SP - PV$
- t: **Time** or instantaneous time (the present)

6.3.1 Limitations
Windup

A basic knowledge of the concept of windup is useful before describing a specific type. Windup is defined as the situation when the feedback controller surpasses the saturation (i.e. maximum) limits of the system actuator and is not capable of instantly responding to the changes in the control error. The concept of the control variable reaching the actuator's operation limits is reasonable considering the wide variety of operating conditions that are possible. When windup occurs the actuator constantly runs at its saturation limit despite any output the system might have. This means that the system now runs with an open loop instead of a constant feedback loop.

Integrator Windup

The most common type of windup that occurs is integrator windup. This occurs when the input into the system receives a sudden positive step command and causes a positive error when the system first responds to the actuator. If the rate of integration is larger than the actual speed of the system the integrator's output may exceed the saturation limit of the actuator. The actuator will then operate at its limit no matter what the process outputs. The error will also continue to be integrated and the integrator will grow in size or "wind up". When the system output finally reaches the desired value, the sign of the error reverses (e.g. $y > y_{sp}$) and causes the integrator to "wind down" until things return back to normal. Through the wind down the control signal is still maximum for a long period of time and the response becomes delayed. The integrator takes a long period of time to fully recover to the operating range of the actuator. Integrator windup may occur from large set point changes, significant disturbances, or equipment malfunctions.

The illustration is a flowchart showing the specific steps that take place through the integrator controller. It shows the input (u) and output of the system (y) , along with the integrator, the actuator, the system, and sensor involved in the process. The sigma used in each flowchart is used to represent the summation of all variables inputed to it.

Measurement signal (y)
and set point (ysp)

ysp

y

Control Signal

u desired

u_{max}

u actual

time

The graphs above are illustrations of integrator windup. The error (e) is shown in the top graph as $e = y_{sp} - y$, where y_{sp} is the set point and y is the measured signal. The bottom graph displays the control signal (u). In this specific case of integrator windup, a set point change occurs and the control signal becomes saturated to its maximum amount u_{max} . The system's error is still not eliminated though because the control signal is too small to make the error go to zero. In turn, this causes the integral to increase, which causes the desired control signal to increase. Therefore, there continues to be a difference between the desired control signal and the true control signal. After a period of time, the set point is lowered to a level where the controller is able to eliminate the control error. At this point $y > y_{sp}$, which causes a change in the sign of the error. This change in sign causes the control signal to begin decreasing. The true signal (u_{out}) is stuck at u_{max} for a while due to desired control signal u being above the limit u_{max} . The changes in setpoints throughout this specific example occur because the input is changed in order to get a minimal error in the system.

The images above are another display of integrator windup. The top most shows the error, the middle shows the control signal, and the bottom shows the integral portion. The integral term begins to decrease, but remains positive, as soon as the error becomes negative at t_1. The output stays saturated due to the large integral termal that developes. The output signal remains at this level until the error becomes negative for a sufficiently long time (t_2). The control signal then fluctuates several times from its maxmimum to its minimum value and eventually settles down on a specific value. At every place where the control signal reaches its maximum the integral value has a large overshoot, and it follows that every place the control signal reaches a minimum the integral has a dampened oscillation. The integral term accounts for the removal of error in the oscillating system, which in turn causes the system's output to eventually reach a point that is so close to the setpoint that the actuator is no longer saturated. At this point, each graph begins to behave linearly and settles down. This example will be used later to show a solution to integral windup. The following section lists several ideas of how to prevent windup.

6.3.2 Solutions

The solutions involved in integral windup are termed as anti-windup schemes. Anti-windup is a method of counteracting the windup that happens in the integration that occurs in the integral controller of the PID.

1.Back-Calculation and Tracking

When the output is saturated, the integral term in the controller is recomputed so that it too is at the saturation limit. The integrator changes over time with the constant T_t. The difference between the output controller (v) and the actuator output (u) is found, and this is defined as an error signal. This error signal is then fed back to the input of the integrator with a gain of $\dfrac{1}{T_t}$. The controller output is proportional to the amount of time that the error is present. The error signal is only present when the actuator is saturated, therefore it will not have any effect on normal operation. With this method, the normal feedback path around process is broken and a new feedback path around the integrator is formed. The integrator input becomes:
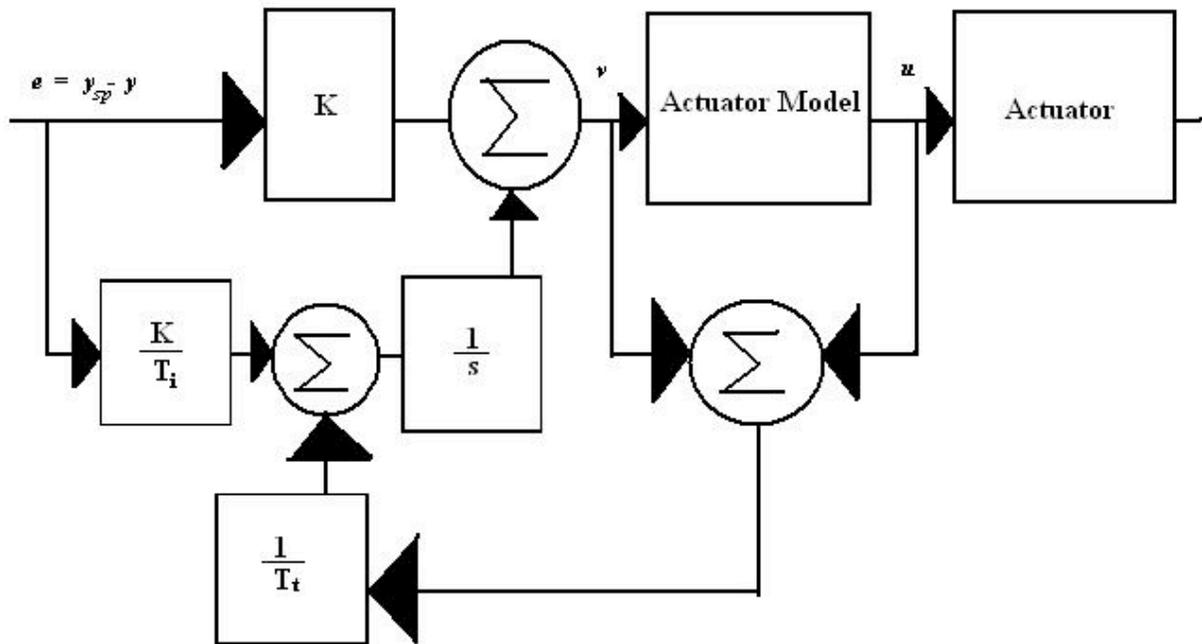
$$\frac{1}{T_t}e_s + \frac{K}{T_i}e$$

744

$$e_s = \frac{-KT_t}{T_i}e$$

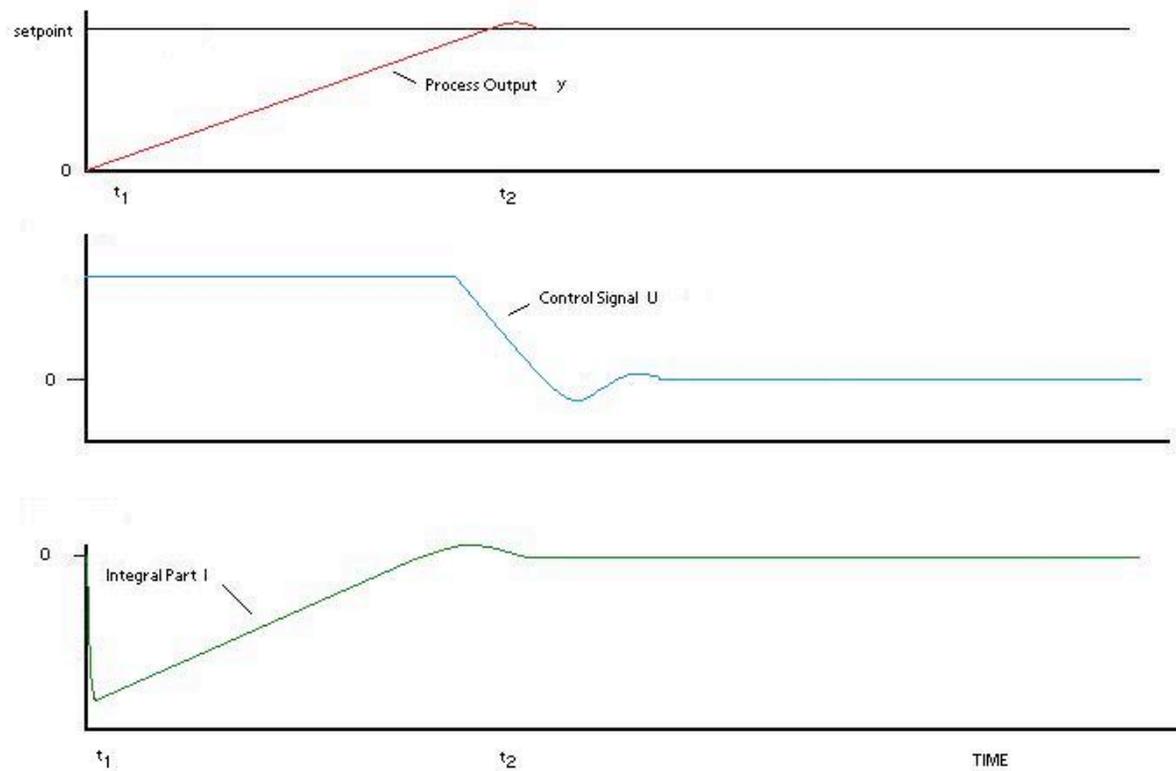Knowing that $e_s = u - v$, the following equation can be made:

$$v = u_{lim} + \frac{KT_t}{T_i}e$$

u_{lim} is the control variable when saturated. The signs for u_{lim} and e are always the same; therefore v is always bigger than u_{lim}. This in turn prevents the windup in the integrator. The feedback gain is $\frac{1}{T_t}$ and is the rate at which the controller output is reset. T_t is called the tracking time constant and is used to tune the amount of anti-windup. The flowchart shown below shows a PID controller with anti-windup implemented. The error signal in this process is only non-zero when the actuator is saturated. When this occurs the normal feedback path around the process is broken because the process input remains constant. A new feedback path is then formed around the integrator and causes the integrator output to be driven towards a value such that the integrator input becomes zero. Back-calculating never allows the input to the actuator to reach its actual saturation level because it forecasts what will actually go into the actuator model beforehand.

The following plots describe what happens to the same system (second example) described in the introduction to Integrator Windup section, only the controller now has anti-windup. The integrator output is at a value that causes the controller to be at the saturation limit, however, the integral is intially negative when the actuator is saturated (t_1), which contrasts with the original behavior of the integrator. This has a positive effect on the process output, as it converges to the set point much quicker than the normal PI controller. There is only a slight overshoot in the process output (t_2).

2.Conditional Integration

Conditional integration operates similarly to back-calculating. In this method the integration is turned off when the control is far from steady state. Integral action is then only used when certain conditions are met, otherwise the integral term is held constant. There are two possible conditions that may cause the integral to be set as a constant value. First, if the controller output is saturated (integration is shut off if error is positive but not negative), then the integrator input is set to a constant value. A second approach involves making the integrator constant when there is a large error. There is a disadvantage to the second approach because the controller may get stuck at a nonzero control error if the integral term has a large value at the time of switch off. For this reason, the first approach looks at the saturated controller output, not the saturated actuator output because referring to the actuator output will generate the same disadvantage[1] .

There is very little difference performance wise between integration and tracking, but they move the proportional bands differently. Read on to learn about proportional bands.

To demonstate what it means to turn the integral term off, here is the equation (with logic) representing a control with conditional integration.

$$u(t) = K\left(e(t) + \frac{1}{T_i} \int e(t)dt\right) + u(t_0)$$

If $e(t) < \dfrac{e_{max}}{K}$, then $T_i = T_i$

Else $T_i = 1,000,000$

3.Series Implementation

To avoid windup in an interacting controller a model of the saturation can be incorporated into the system. The model of saturation will limit the control signal indirectly. Too hard a limitation will cause an unnecessary limitation of the control action. Too weak a limitation will cause windup. More flexibility is provided if the saturation is positioned as follows:

With this setup it is possible to force the integral part to assume other preload values during saturation. The saturation function is replaced by the nonlinearity shown. This anti windup procedure is sometimes called a "batch unit" and is considered a type of conditional integration. It is mainly used for adjusting the overshoot during startup when there is a large set point change.

6.3.3 The Proportional Band

The proportional band is useful to understand when dealing with windup and anti windup schemes. The proportional band is a range such that if the instantaneous value of the process output or predicted value on the interval falls within the range, the actuator does not saturate. The proportional band (defined as the area between y_{low}, y_{high}) is given by the following equations:

Control signal for PID control without derivative gain limitation:

$$u = K * (b * y_{sp}) + I - K * T_d * \frac{dy}{dt}$$

The predicted process output is given by

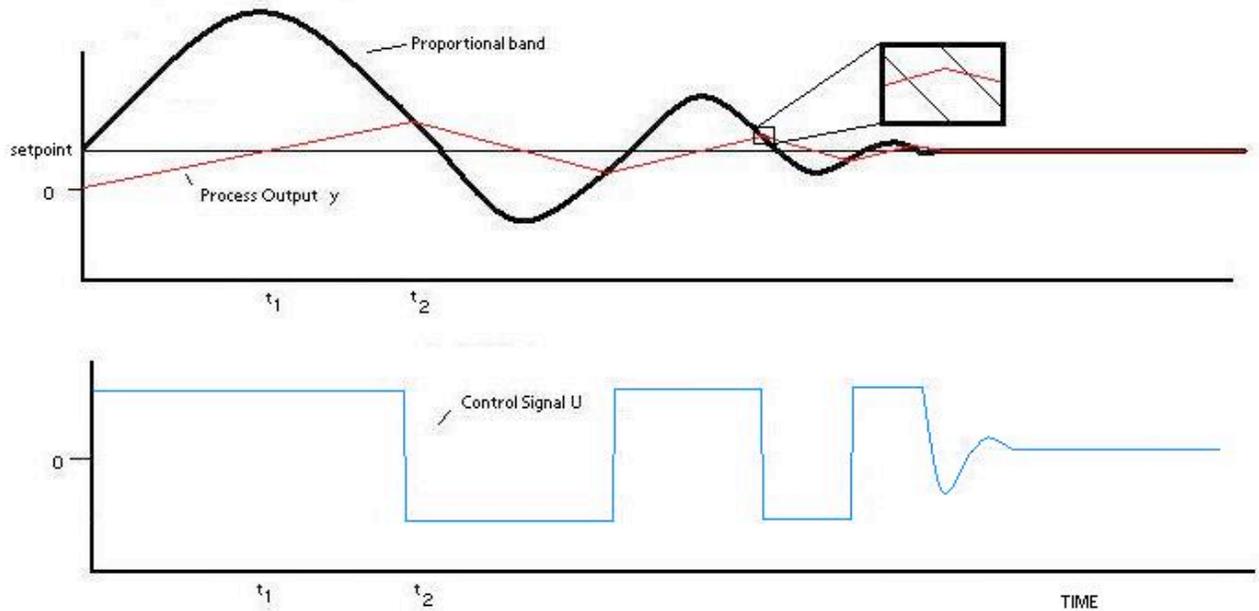$$y_p = y + T_d * \frac{dy}{dt}$$

748

Combining the above equations for a maximum and minumum control signal (u_{max} and u_{min})which corresponds to the points for which the actuator saturates,

$$y_{low} = b * y_{sp} + \frac{I - u_{max}}{K}$$

$$y_{high} = b * y_{sp} + \frac{I - u_{min}}{K}$$

If the predicted output is in the proportional band, the controller operates linearly. The control signal saturates when the predicted output is outside of the proportional band. The following plot illustrates how the proportional band is useful in understanding windup for the same example. At time equals zero, the proportional band is increased, indicating that the integral part is increasing. Immediately after t_1 , the output is greater that the setpoint, which causes the integral part to start to decrease. As can be seen, the output value, y does not reach the proportional band until it is greater than the setpoint, y_{sp} .

When the output finally reaches the proportional band at t_2 , it goes through the band very quickly, due to the band changing at a high rate. The control signal u decreases immediately and saturates in the opposite direction, causing the output to decrease and approach the setpoint. This process repeats as the output value converges with the setpoint.

The tracking time constant has a major influence on the proportional band. The proportional band is moved closer to the output because of tracking. The tracking time constant determines the rate that the proportional band moves at.

Real Life Scenario

Imagine a situation where a water storage tank that is needed to store water and supply for use in a plant and it is necessary to keep the water level constant. A level sensor is installed to keep track of the water level in the tank, and it sends this data to the controller. The controller has a set point for instance 75% full. The rate of water entering the tank is increased by opening valve and decreased by closing the valve. In the case of integral windup, the valve is all the way closed or open and the actuator sends the signal to close it more or open it more which is not possible.

6.4 Derivative Control

Derivative action can be thought of as making smaller and smaller changes as one gets close to the right value, and then stopping in the correct region, rather than making further changes. Derivative control quantifies the need to apply more change by linking the amount of change applied to the rate of change needed. For example, an accelerator would be applied more as the speed of the car continues to drop. However, the actual speed drop is independent of this process. On its own, derivative control is not sufficient to restore the speed to a specific value. Pairing the match in change with a proportionality constant is enough to properly control the speed.

The derivative control is usually used in conjunction with P and/or I controls because it generally is not effective by itself. The derivative control alone does not know where the setpoint is located and is only used to increase precision within the system. This is only control type that is open loop control (also known as feedforward loop). The derivative control operates in order to determine what will happen to the process in the future by examining the rate of change of error within the system. When the derivative control is implemented the following general equation is used:

$$u(t) = \frac{1}{T_d}\frac{d}{dt}e(t) + u(t_0)$$

Where:

$u(t)$ = control signal
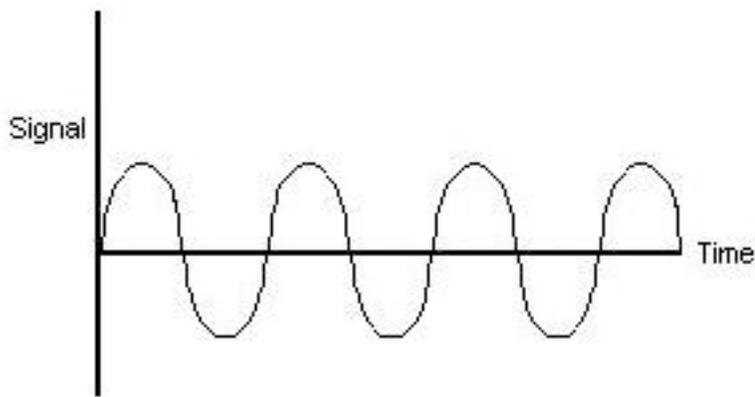
T_d = derivative time

$e(t)$ = error

750

$u(t_0)$ = initial control signal

This equation shows the derivative control is proportional to the change in error within the system.

6.4.1 Limitations

The major problem associated with this control is the noise problem. When the frequency within the system is high (change in error of the system is large), taking the derivative of this signal may amplify the signal greatly. Therefore, small amounts of noise within the system will then cause the output of system to change by a great amount. In these circumstances, it is often sensible use a PI controller or set the derivative action of a PID controller to zero.

The following is a graphical represenation of some sinusoidal output signal function of a system.

6.4.2 Solutions

To eliminate/minimize this problem, an electronic signal filter can be included in the loop. Electronic signal filters are electronic circuits which perform signal processing functions, specifically intended to remove unwanted signal components and/or enhance wanted ones. Electronic filters can be: passive or active, analog or digital, discrete-time (sampled) or continuous-time, linear or non-linear, etc. The most common types of electronic filters are linear filters, regardless of other aspects of their design.

For more information on electronic signal filters, reference the following website:

http://en.wikipedia.org/wiki/Electronic_filter

The derivative term in the equation is usually changed by putting a first order filter on the term so that the derivative does not amplify the high frequency noise that is attenuated. Below is a sample outcome figure of a possible derivative of the output signal shown above along with the filtered signal.

As shown, it is possible for the amplitude to be magnified when the derivative is taken for a sinusoidal function. A filter is usually a set of complicated equations that are added to the derivative that effect the function as shown.

6.5 Robustness of PID controllers

When designing a PID controller it is important to not only create a controller that will correct the error in an efficient manor, but also a robust controller. No matter how good the controller is if it cannot handle the magnitude of certain disturbances then it will not be able to perform well and complete the job. The robustness of a controller is the ability of the controller to withstand disturbances and still operate effectively. In order to calculate controller robustness, a weighted sum of the integral of the squared controller effort (ISC) and the integral of the squared controller error (ISE) is calculated. The smaller the weighted sum the more robust the controller.

$ISE = 0 \int \infty\ e(t)2\ dt$

$ISC = 0 \int \infty\ [u(\infty)\text{-}u(t)]2\ dt$

$IT = w\ ISC + ISE$

In these equations u is the input variable, w is the weight factor which is based on downstream processes and IT is the total error and controller effort. The weight factor is used to assign a higher significance to the ISC or the ISE. The significance depends on the design and whether it is more important to have minimal error and tight control or
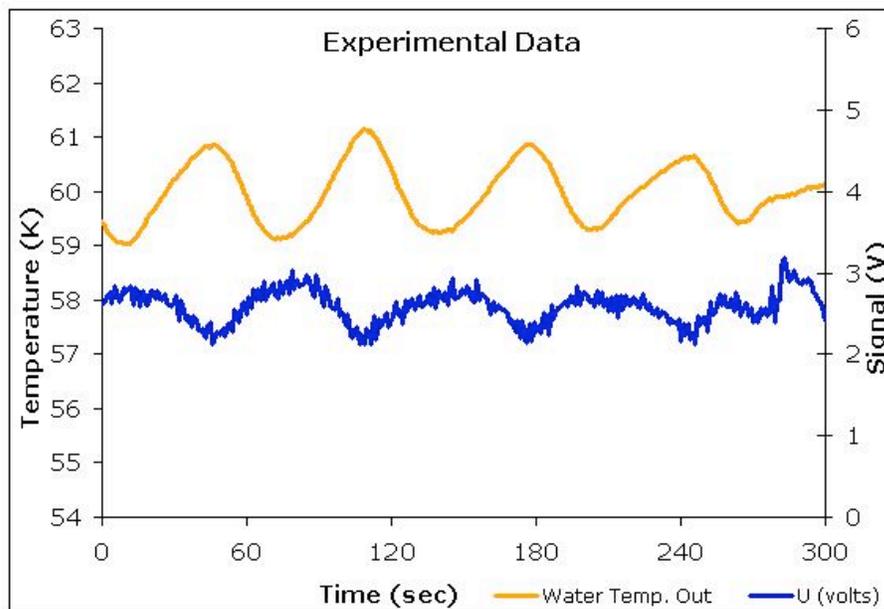
less controller effort with more error. Low values of w are used when tight control is desired. Robustness calculations are determined experimentally. The robustness depends a lot on the type of control used, whether it is feedback, feed- forward, cascade, or a combination.
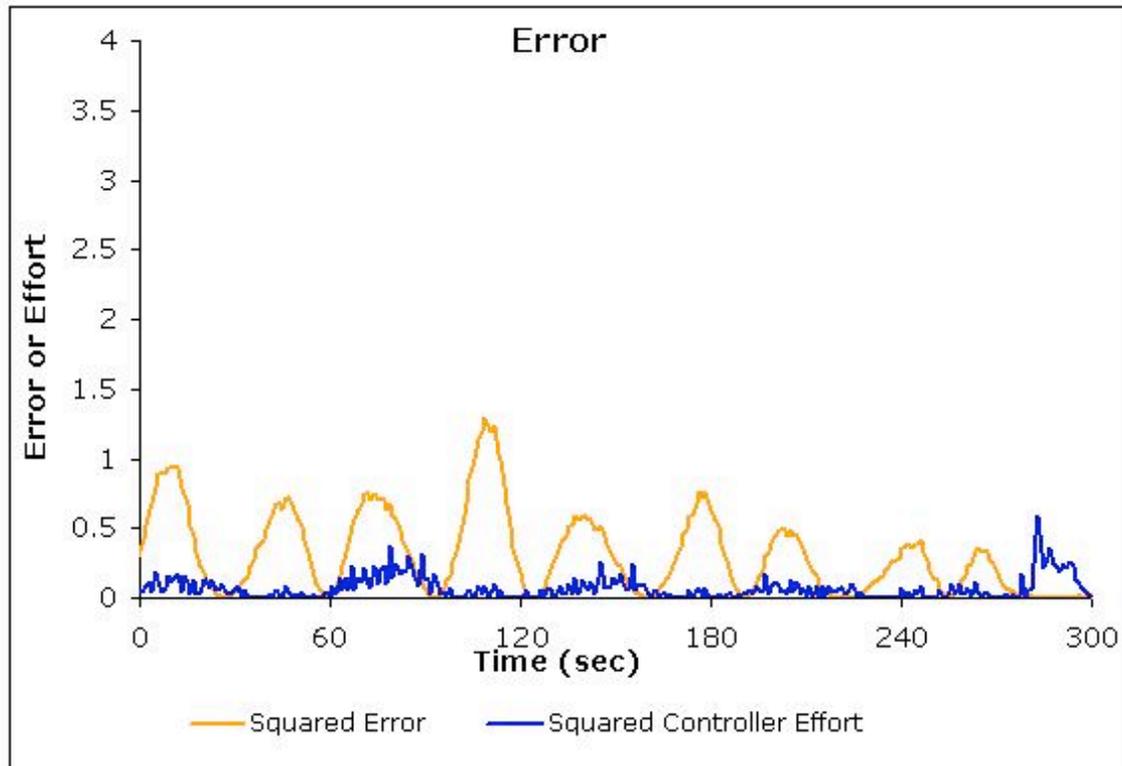
6.5.1 Notes on robustness

It is important to realize that when calculating the robustness of a feed forward system, the system will reach a steady state, but it is unlikely that the system will reach the original set point. As a result, the feed forward system will most likely be less robust than a feedback system. The reason the feed-forward system doesn't return to the set point but finds a new steady state is because it uses open loop control and there are no sensors to inform it that its adjustment to the disturbance did not work correctly. For this reason, feed-forward controllers are sometimes coupled with a feedback system. When combined the feed-forward controller can now see if its corrections achieved the desired result or not, resulting in a more robust controller.

6.6 Example Robustness Calculation

In this system, the controller is regulating the temperature of the feed to a certain downstream process. After collecting experimental data from your controller, the following graph was made.

Using this graph the squared controller effort and the squared error can also be graphed and the ISE and ISC can be calculated. The graph of the corresponding effort and error can be seen below.

Using the squared controller effort and the squared error graph, the robustness can be calculated. To find the integral of the squared error and the squared controller effort the trapezoid rule can be used.

IT = w ISC + ISE

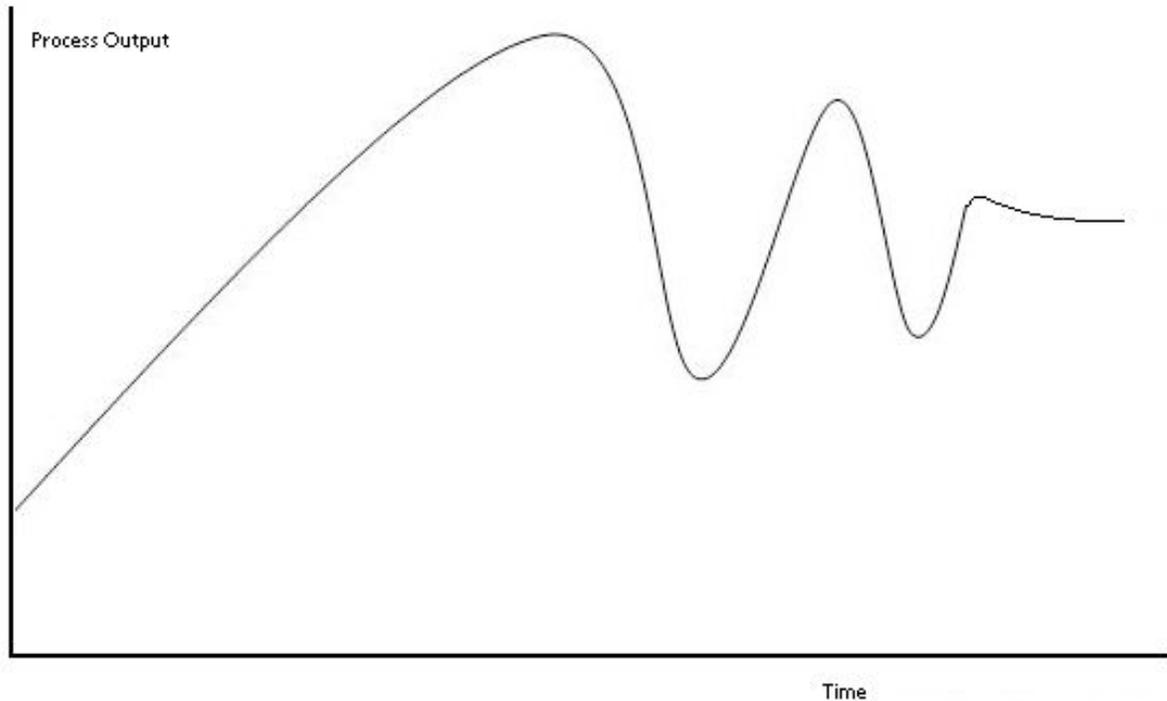ISC= 21 ISE=94

Assuming a weight factor of 1, the robustness of the controller is 115.
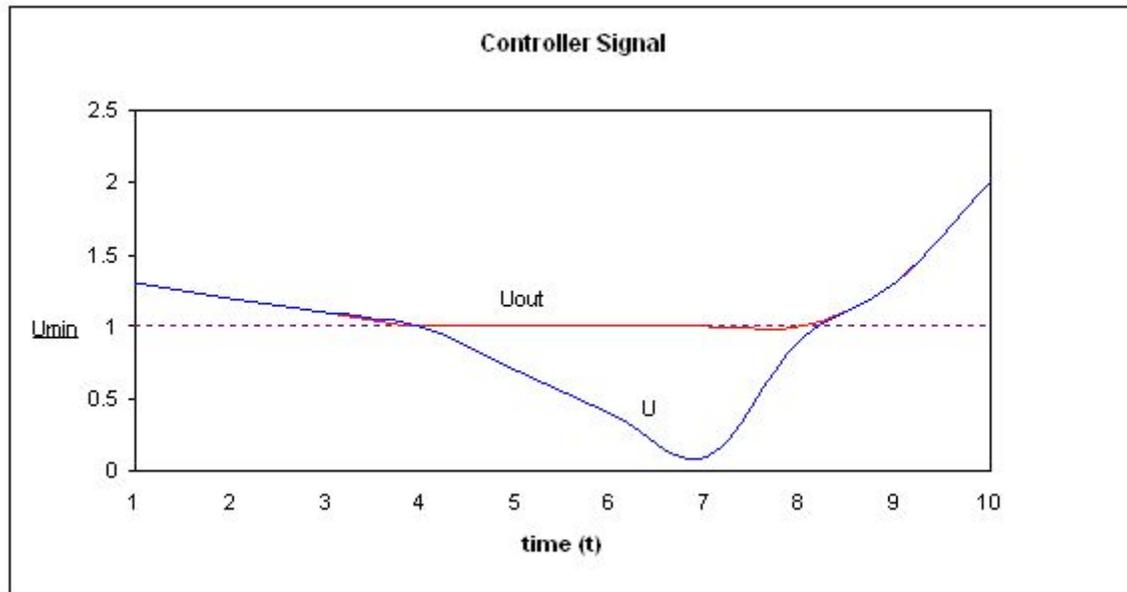
6.7 Example problems

6.7.1 Problem 1

Consider the below output from using one of the controllers. The general equation of the output signal, U = K(e). What problems do you see in the below picture and how can you correct them?

754

Process Output

Time

Solution: The response time is very high, and it has a high overshoot and oscillation problems as well as steady-state errors. The equation confirms that it is a P-controller and we can correct the above situation by adding a integral term to eliminate steady-state errors and by adding a Derivative term to eliminate the overshoot and oscillations.

6.7.2 Problem 2

A heat exchanger uses a PID control system to control the temperature of the outlet of tube side of the heat exchanger. It's required to keep water at 20 C for the cooling of an effluent from an exothermic reaction. The specific controller has a min output of 1 volt and a max output of 6 volts. The experimental data shows a large lag time in the controller's ability to cool the incoming fluid (the lower portion of the sinusoidal curve). The following data shows the correlation between the controller output and the calculated signal. Describe the type of problem and a possible solution to the problem.
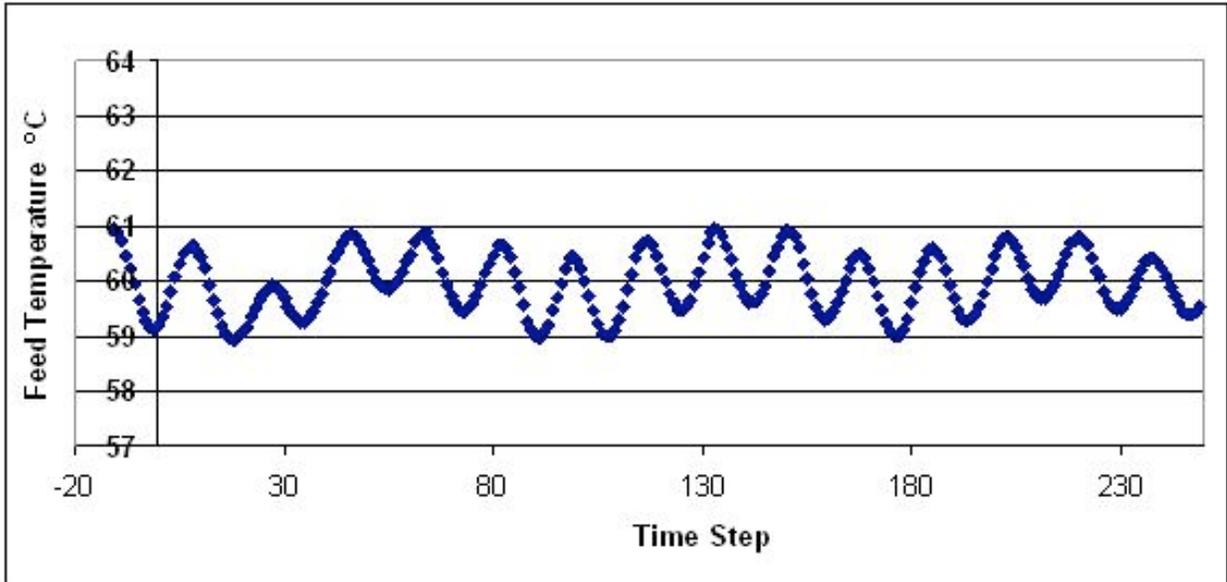
Controller Signal

Solution: The Problem shown is an example of integrator-windup. The controllers calculated signal is lower than the min output that it can handle. This explains the lag time seen for the lower portion of the sinusoidal curve. To fix this problem the integral control can be recomputed using back track calculation.
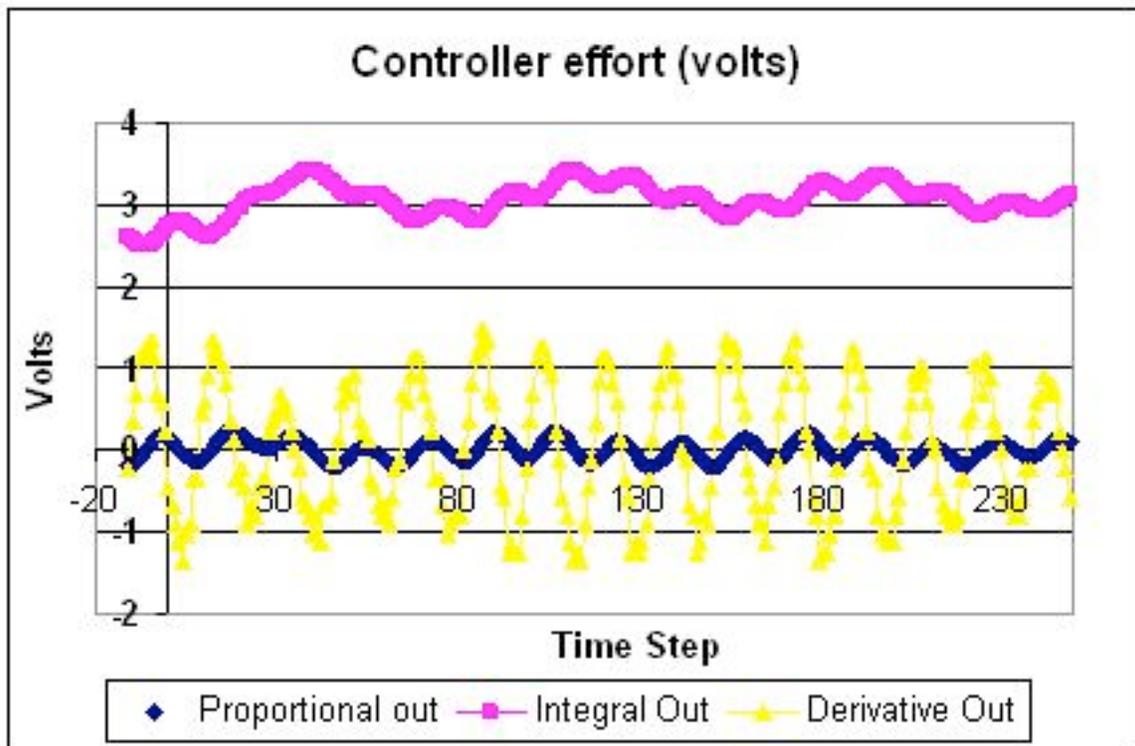
6.7.3 Problem 3

(a) You are using a PID controller to maintain the feed to a reactor at a constant temperature. It is imperative that the feed remain within 1 °C of the set-point. You've tuned your PID control system according to the Ziegler-Nichols open loop system and obtained the following values for K_c, τ_I, and τ_D: $K_c = 0.218$, $1/\tau_I = 0.0536$, $\tau_D = 4.67$

You want to know if these are good control values or if they can be improved. Your control variable output looks like this:
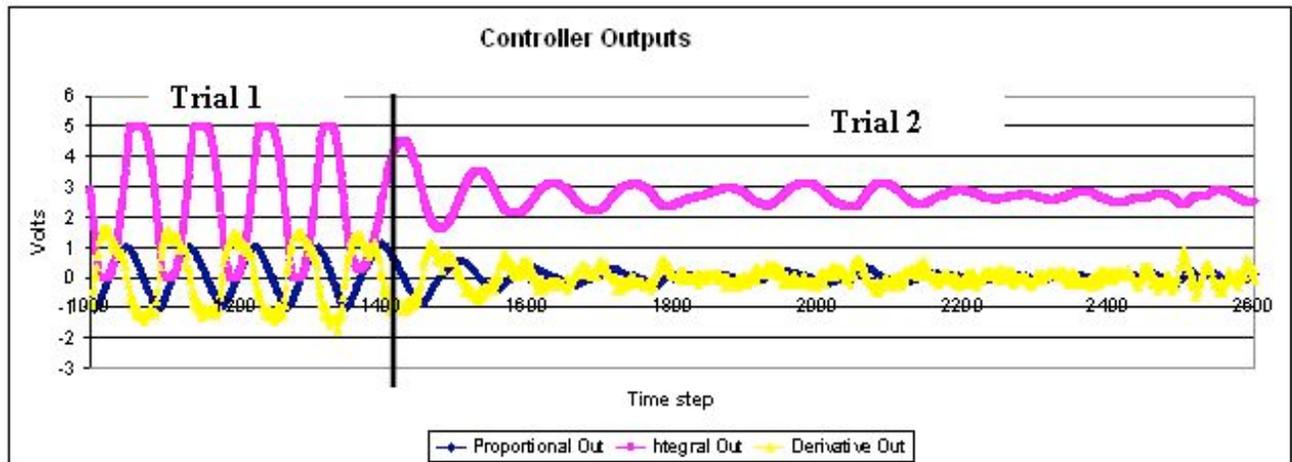
Additionally, you have available the following plot of the voltage output from your control parameters:

Is this system in control? If so, explain why. If not, explain why and what (qualitative) improvements to the PID gains will bring the system into better control.

Solution: The system is in control in that it appears to roughly stay within the defined acceptable boundaries of ±1°C. However, it does display oscillation, indicating that the system is working a little harder than maybe it has to. Looking at the controller voltage output, we see that the derivative gain is giving high oscillations. Thus maybe decreasing this gain would keep the system in control with a little less wear and tear on the control valves.

(b) At another flow rate, you observe the following behavior in your controller:

Is the system under control in Trial 1? In Trial 2? Why or why not? What physically happened between trials to give the observed change in output?

Solution: In trial 1, without even looking at the direct system output, we can guess that it is probably not in control. The controller output voltage indicates that the integral gain is way too high, and its wild oscillations will doubtlessly be causing oscillations in the system. Physically, this means that the system is opening and closing the valve constantly, creating a lot of wear and tear on the machinery, and probably overcompensating for small changes and further destabilizing the system. In trial 2, we can see that the integral voltage oscillations have damped, indicating that the gain has been reduced. As these oscillations level out, the oscillations in the other control gains also damp down, as the system comes into better control and none of the terms are working as hard as they had been before.

6.7.4 Problem 4

Note this is an open-ended problem meant to give readers an intuition about PID controllers

Imagine the classic differential equation problem of two tanks in a series governed by the differential equations:

758

$$\frac{dh_1}{dt} = PID - K_1\sqrt{h_1}$$

$$\frac{dh_2}{dt} = K_1\sqrt{h_1} - K_2\sqrt{h_2}$$

$$PID = 1 + K_p e(t) + \frac{1}{t_i}\int_0^t e(\tau)\,d\tau + t_d\frac{de(t)}{dt}$$

h_1 = height of tank 1
h_2 = height of tank 2
K_i = valve constant associated to tank 1 or 2

The valve constants $K_1 = 5$ and $K_2 = 5$ and the set point is 3. What values of K_p, t_i, and t_d will cause the two tanks to reach steady state the fastest? Also what values will give the most oscillations?

Solution:
Click on this link for the worked out Mathematica solution

Other questions that can be explored with slight modifications of the code are:
1. Will the values of K_p, t_i, and t_d yield similar results if the PID controller was placed on tank 2? What new values of K_p, t_i, and t_d will cause the two tanks to reach steady state the fastest?

2. For the values of K_p, t_i, and t_d that causes the two tanks o reach steady state the fastest, what happens if you increase K_1? What will happen if you increase K_2?
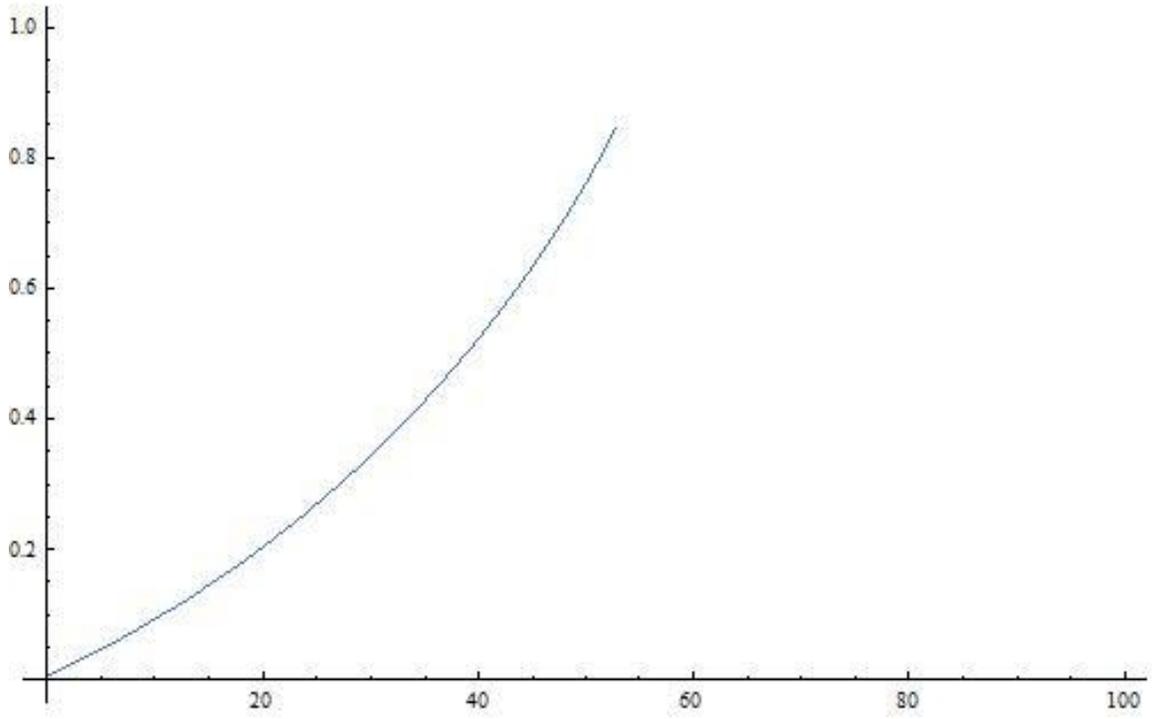
6.7.5 Problem 5

Label the following general response charts as overdamped, critically damped, underdamped, undamped, and unstable.
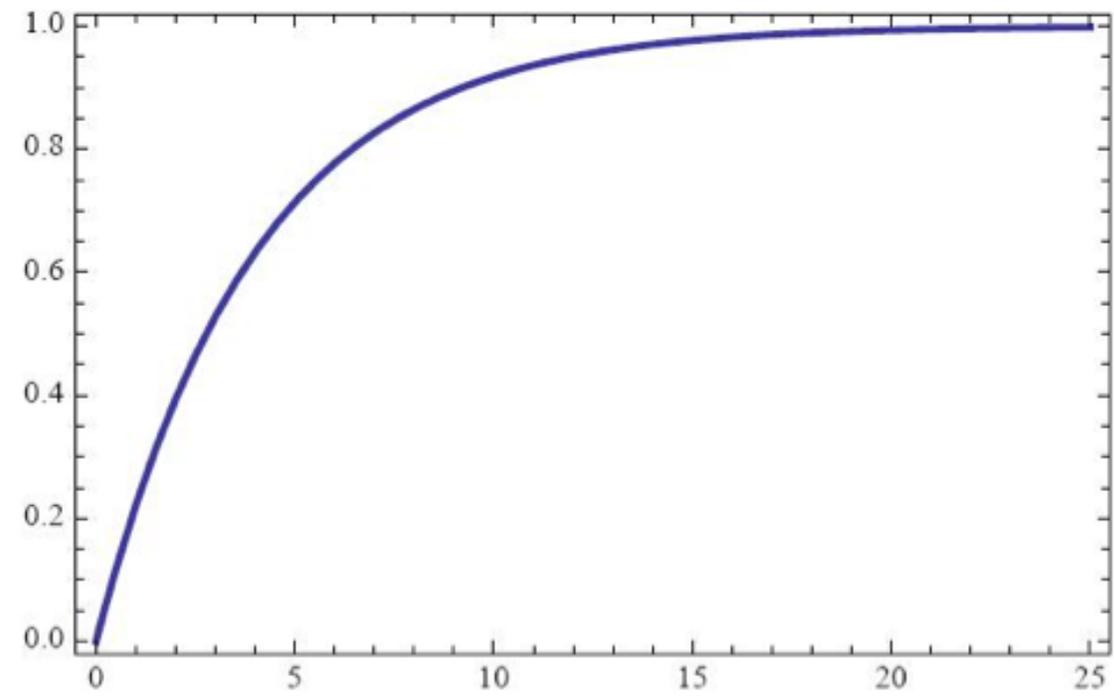
(a)

Step Response

760

(b)

(c)

(d)

(e)

762
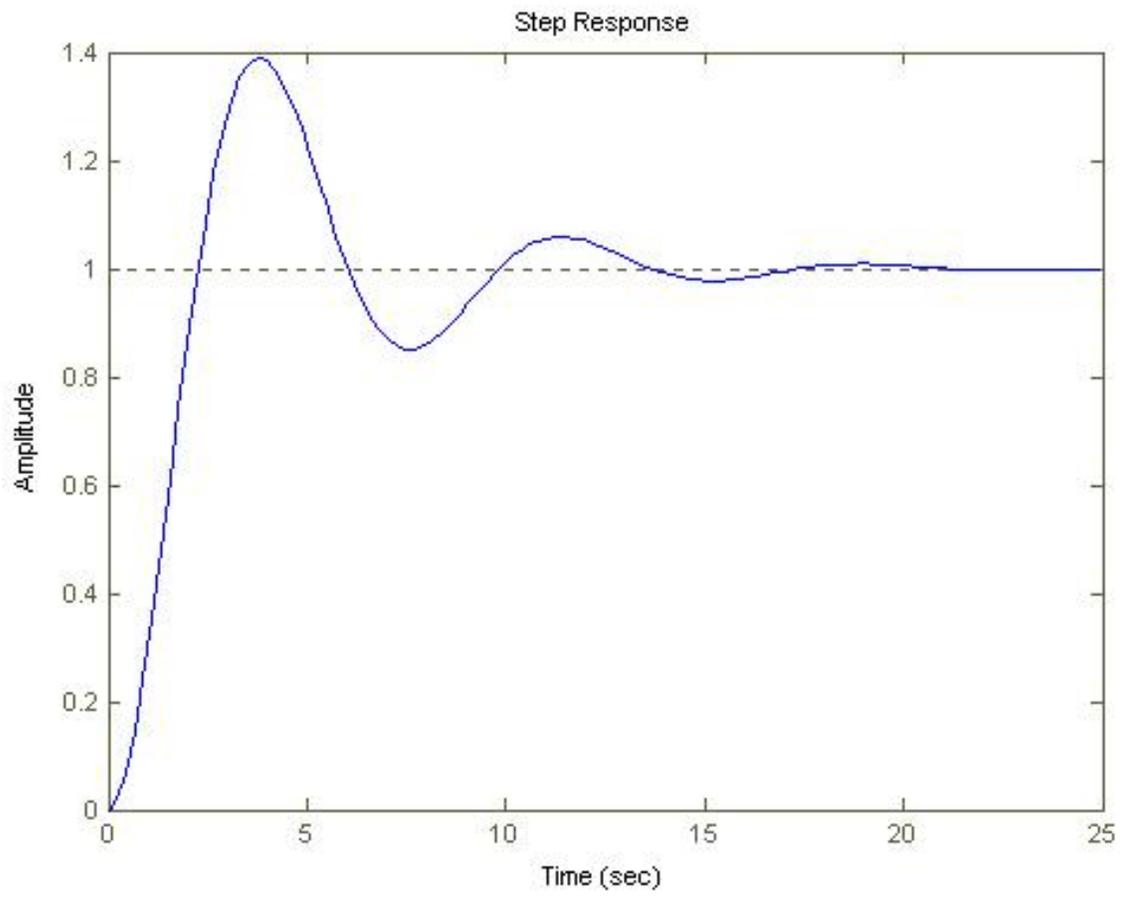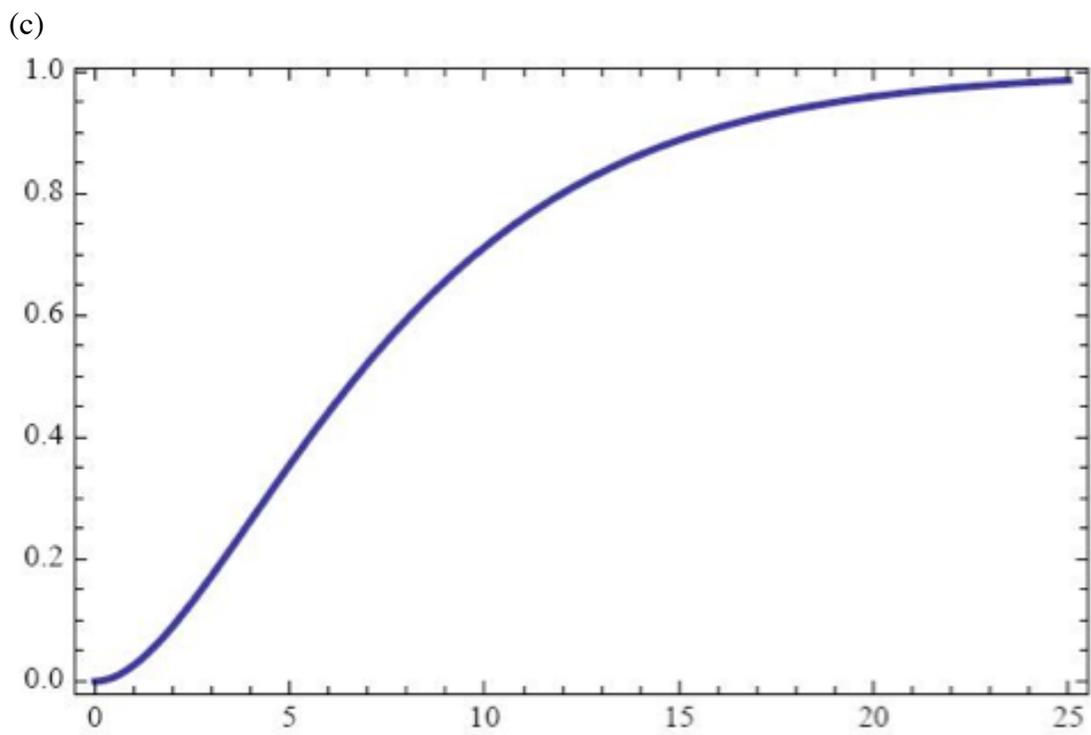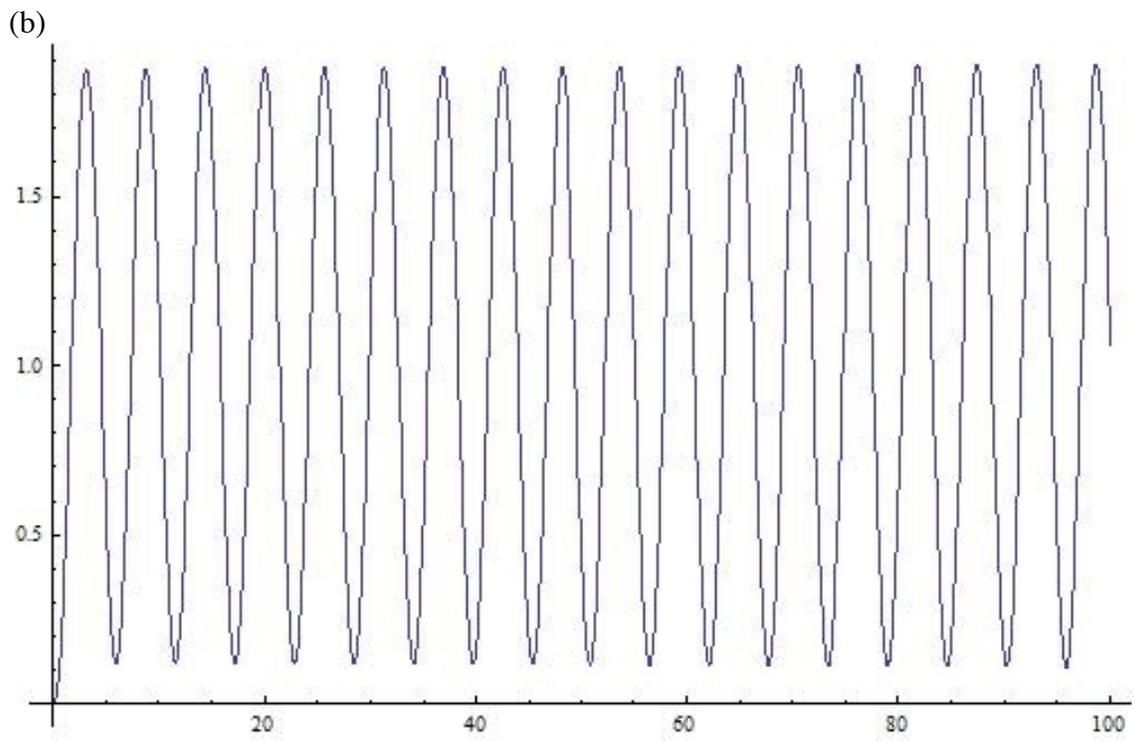
Answers: a)Underdamped b)Undamped c)Critically Damped d)Unstable e)Overdamped

6.8 Multiple Choice Question 1

1. What is a back-calculation used for?

(a) To reduce noise problems

(b) To reduce steady-state errors

(c) To reduce errors in integral term

(d) To reduce errors in proportional term

6.9 Multiple Choice Question 2

2.What is a proportional band?

(a) A way to eliminate errors in integral term

(b) A way to eliminate errors in proportional term

(c) A way to eliminate errors in derivative term

(d) A way to understand anti-windup problem

6.10 Sage's Corner
http://video.google.com/googleplayer.swf?docId=-1994051404153478750

6.11 References

- Bhattacharyya, Shankar P., Datta & Silva, PID Controllers for Time-Delay Systems, 2005 [1]
- Astrom, Karl J., Hagglund, Tore., Advanced PID Control, ISA, The Instrumentation, Systems and Automation Society.
- U of Texas chE
- Wikipedia Article, PID Controller, http://en.wikipedia.org/wiki/PID_control