3 Continuous Control

By Harold Wade

Topic Highlights

Process Characteristics
Feedback Control
Controller Tuning
Advanced Regulatory Control

3.1 Introduction

Continuous control refers to a form of automatic process control in which the information—both from sensing elements and to actuating devices—can have any value between minimum and maximum limits. This is in contrast to discrete control, where the information normally is in one of two states, such as on-off, open-closed, run-stop, etc.

Continuous control is organized into feedback control loops, as shown in Figure 3-1. In addition to a controlled process, each control loop consists of a sensing device that measures the value of a controlled variable, a controller that contains the control logic plus provisions for human interface, and an actuating device that manipulates the rate of addition or removal of mass or energy or some other property that can affect the controlled variable. (In emerging technology, the control logic may be located at either the sensing or the actuating device.)

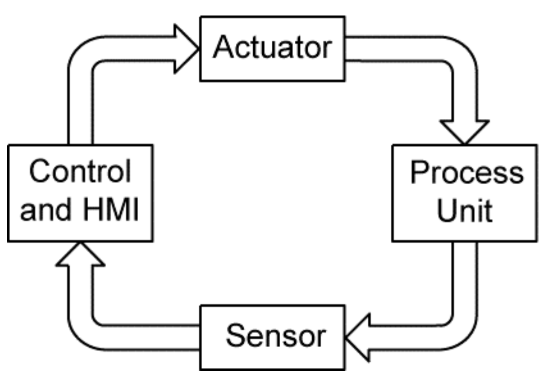


Figure 3-1: Components and Information Flow in a Feedback Control Loop

Continuous process control is used quite extensively in industries where the product is in a continuous, usually fluid, stream. Representative industries are petroleum refining, chemical and petrochemical, power generation, and municipal utilities. Continuous control can also be found in processes in which the final product is produced in batches, strips, slabs, or as a web in, for example, the pharmaceutical, pulp and paper, steel and textile industries. There are also applications for continuous control in the discrete industries—for instance, a temperature controller on an annealing furnace, or motion control in robotics.

Since the components in a feedback control loop may be physically separated by 50 meters (m) to 500 m or more, some form of signal communication must be employed. An early technology for signal transmission used in the process industries employed pneumatic signals, where 3 pounds per square inch (psi) [20.684 kiloPascals (kPa)] represented the minimum value and 15 psi (103.42 kPa) represented the maximum value. This technology has largely been replaced by electric signal transmission, utilizing a signal range of 4-to-20 milliamps (mA) direct current (DC). Electric signal transmission is the dominant technology used today, although this is being replaced by shared digital signals and by wireless communication.

The central device in a control loop, the controller, may be built as a stand-alone device or may exist as shared components in a digital system, such as a distributed control system (DCS) or programmable logic controller (PLC).

Process control systems, plus the related functions of measurements and alarms, are represented using special symbols on "piping and instrument diagrams" (P&IDs). A P&ID shows the outline of the process units and connecting piping as well as a standard symbolic representation of the instrumentation and control (I&C) systems. Figure 3-2 is an example of a small P&ID for a heat exchanger. In actual practice, however, a typical P&ID will encompass many process units, measurements and controls, and will be densely drawn on one or more large sheets of paper.

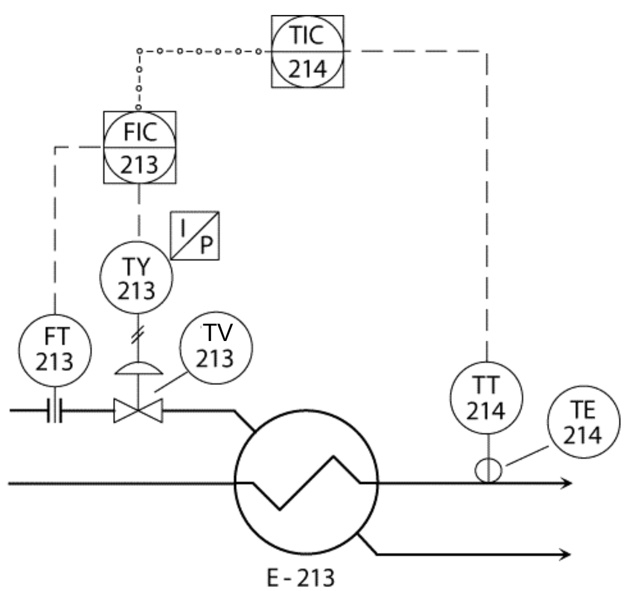


Figure 3-2: Piping and Instrumentation Diagram (P&ID)

The symbols used to represent various types of instrumentation devices, the type of communication between devices, and the nomenclature for device identification are defined by the standard ISA-5.1-1984 (R1992) -*Instrumentation Symbols and Identification*. Figure 3-3 shows the symbols used for continuous controllers and other general instrumentation, as defined in this standard.

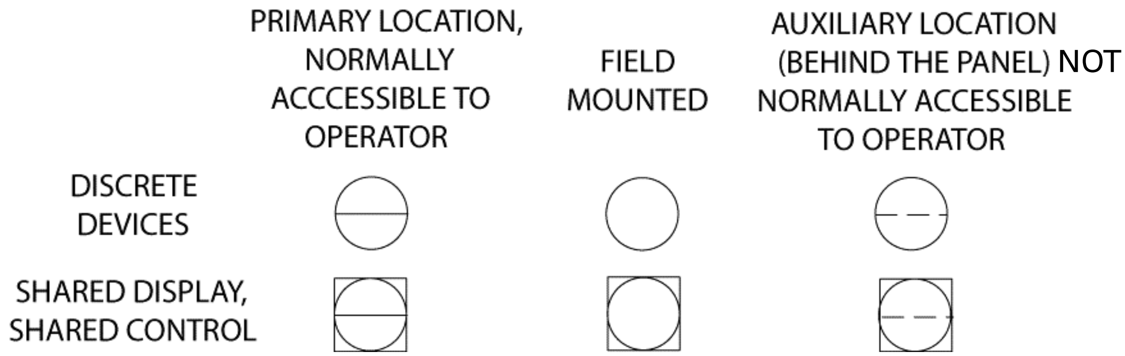


Figure 3-3: Symbols for Continuous Controllers and Other General Instruments, from ISA-5.1-1984 (R1992) - Instrumentation Symbols and Identification

3.2 Process Characteristics

In order to understand feedback control loops, one must understand the characteristics of the controlled process. Listed below are characteristics of almost all processes, regardless of the application or industry.

- Industrial processes are non-linear; that is, they will exhibit different responses at different operating points.
- Industrial processes are subject to random disturbances, due to fluctuations in feedstock, environmental effects, and changes or malfunctions of equipment.
- Most processes contain some amount of dead time; a control action will not produce an immediate feedback of its effect.
- Many processes are interacting; a change in one controller's output may affect other process variables besides the intended one.
- Most process measurements contain some amount of noise.
- Most processes are unique; processes using apparently identical equipment may have individual idiosyncrasies.

A typical response to a step change in signal to the actuating device is shown in Figure 3-4.

In addition, there are physical and environmental characteristics that must be considered when selecting equipment and installing control systems.

- The process may be toxic, requiring exceptional provisions to prevent release to the environment.
- The process may be highly corrosive, limiting the selection of materials for components that come in contact with the process.
- The process may be highly explosive, requiring special equipment housing or installation technology for electrical apparatus.

3.3 Feedback Control

The principle of feedback control is that, if a controlled variable deviates from its desired value (setpoint), corrective action moves a manipulated variable (the controller output) in a direction that

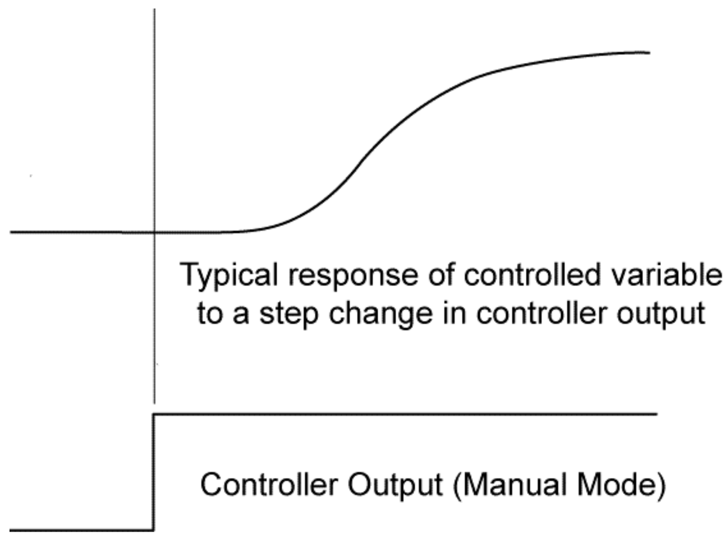


Figure 3-4: Typical Response of Controlled Variable to Step-Change in Controller Output (Manual Mode)

causes the controlled variable to return toward setpoint. Most feedback control loops in industrial processes utilize a proportional-integral-derivative (PID) control algorithm. There are several forms of the PID. There is no standardization for the names. The names "ideal," "interactive," and "parallel" are used here, although some vendors may use other names.

3.3.1 Ideal PID Algorithm

The most common form of PID algorithm is the *ideal* form (also called the "ISA" form). This is represented in mathematical terms by Equation 3-1, and in block diagram form by Figure 3-5:

$$m = K_C\left(e + \frac{1}{T_I}\int edt + T_D\frac{de}{dt}\right) \tag{3-1}$$

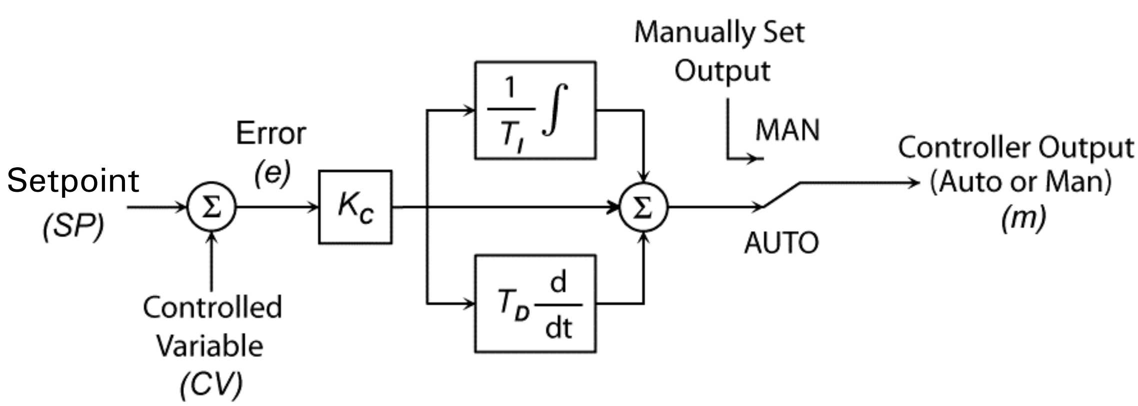


Figure 3-5: Block Diagram of "Ideal" PID Algorithm,
also Showing Functional Form of Automatic – Manual Switch

Here, m represents the controller output; e represents the error (difference between setpoint and controlled variable). Both m and e are in percent of span. The symbols K_C (controller gain), T_I (integral time) and T_D (derivative time) represent tuning parameters that must be adjusted for each application.

The terms in the algorithm represent the proportional, integral, and derivative contributions to the output. The proportional mode is responsible for most of the correction. The integral mode assures that, in the long-term, there will be no deviation between setpoint and controlled variable. The derivative mode may be used for improved response of the control loop. In practice, the proportional and integral modes are almost always used; the derivative mode is often omitted, simply by setting $T_D = 0$.

There are other forms for the tuning parameters. For instance, controller gain may be expressed as proportional band (PB), defined as the amount of measurement change (in percent of measurement span) required to cause 100% change in the controller output. The conversion between controller gain and proportional band is shown by Equation 3-2:

$$K_C = \frac{100}{PB} \qquad PB = \frac{100}{K_C} \tag{3-2}$$

The integral mode tuning parameter may be expressed in reciprocal form, called *reset rate*. Whereas T_I is normally expressed in "minutes per repeat," reset rate is expressed in "repeats per minute." The derivative mode tuning parameter, T_D, is always in time units, usually in minutes. (Traditionally, the time units for tuning parameters has been "minutes." Today, however, many vendors are expressing the time units in "seconds.")

3.3.2 Interactive PID Algorithm

The *interactive* form, depicted by Figure 3-6, was the predominant form for analog controllers and is used by some vendors today. Other vendors provide a choice of the ideal or interactive form. There is essentially no technological advantage to either form; however, the required tuning parameters differ if the derivative mode is used.

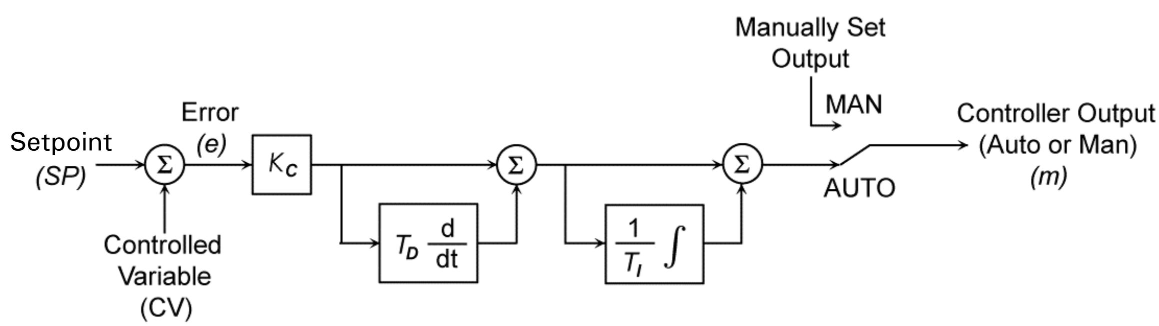


Figure 3-6: Block Diagram of Interactive PID Algorithm

3.3.3 Parallel PID Algorithm

The *parallel* form, shown in Figure 3-7, uses independent gains on each mode. This form has traditionally been used in the power generation industry and in such applications as robotics, flight control, motion control, etc. Other than power generation, it is rarely found in the continuous process industries. With compatible tuning, the ideal, interactive, and parallel forms of PID produce identical performance; hence no technological advantage can be claimed for any form. The tuning procedure for the parallel form differs decidedly from that of the other two forms.

3.3.4 Time Proportioning Control

Time proportioning refers to a form of control in which the PID controller output consists of a series of periodic pulses whose duration is varied to relate to the normal continuous output. For example, if the fixed cycle base is 10 seconds, a controller output of 30% will produce an "on" pulse of 3 seconds and an "off" pulse of 7 seconds. An output of 75% will produce an "on" pulse of 7.5 seconds and an "off"

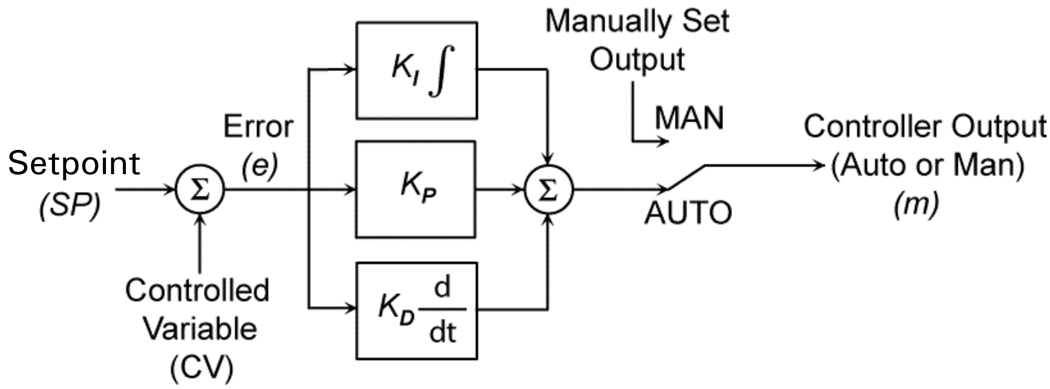


Figure 3-7: Block Diagram of Parallel Algorithm (Showing Independent Gains on Each Mode)

pulse of 2.5 seconds. This type of control is usually applied where the cost of an on-off final actuating device is considerably less than the cost of a modulating device. In a typical application, the "on" pulses apply heating or cooling by turning on a resistance heating element, an SCR (silicon controlled rectifier) or a solenoid valve. The mass of the process unit, say, a plastics extruder barrel, acts as a filter to remove the low-frequency harmonics and apply an even amount of heating or cooling to the process.

3.3.5 Manual-Automatic Switching

It is desirable to provide a means for process operator intervention in a control loop in event of abnormal circumstances, such as a sensor failure or a major process upset. Figures 3-5, 3-6, and 3-7 show a manual-automatic switch that permits switching between manual and automatic modes. In the manual mode, the operator can set the signal to the controller output. However, when the switch is returned to the automatic position, the automatic controller output must match the operator's manual setting or else there will be a "bump" in the controller output. (The term *bumpless transfer* is frequently used.) With older technology, it was the operator's responsibility to prevent bumping the process. With current technology, bumpless transfer is built into most control systems; some vendors refer to this as initializing the control algorithm.

3.3.6 Direct and Reverse Acting

For safety and environmental reasons, most final actuators such as valves will close in the event of loss of signal or power to the actuator. There are instances, however, when the valve should open in the event of signal or power failure. Once the failure mode of the valve is determined, the action of the controller must be selected. Controllers may be either *direct acting* (DA) or *reverse acting* (RA). If a controller is direct acting, an increase in the controlled variable will cause the controller output to increase. If the controller is reverse acting, an increase in the controlled variable will cause the output to decrease. Since most control valves are fail-closed, then the majority of the controllers are set to be reverse acting. The setting—DA or RA—is normally made at the time the control loop is commissioned. With some DCSs, the DA/RA selection can be made without considering the failure mode of the valve; then a separate selection is made as to whether to reverse the analog output signal. This permits the human-machine interface (HMI) to display all valve positions in a consistent manner, 0% for closed and 100% for open.

3.3.7 Activation for Proportional and Derivative Modes

Regardless of algorithm form, there are certain configuration options that every vendor offers. One configuration option is the DA/RA setting. Other configuration options pertain to the actuating signal for the proportional and derivative modes. Note that, in any of the forms of algorithms, if the derivative mode is being used ($T_D \neq 0$), a setpoint change will induce an undesirable spike on the controller output. A configuration option permits the user to make the derivative mode sensitive only to changes in the controlled variable, not to the setpoint. This choice is called "derivative-on-error" or "derivative-on-measurement."

Even with derivative-on-measurement, on a setpoint change, the proportional mode will cause a step change in controller output. This, too, may be undesirable. Therefore, a similar configuration option permits the user to select "proportional-on-measurement" or "proportional-on-error." Figure 3-8 shows both proportional and derivative modes sensitive to measurement changes alone. This leaves only the integral mode on error, where it must remain, since it is responsible for assuring the long-term equality of setpoint and controlled variable. In event of a disturbance, there is no difference in the responses of derivative-on-measurement, proportional-on-measurement, and the configuration with all modes on error.

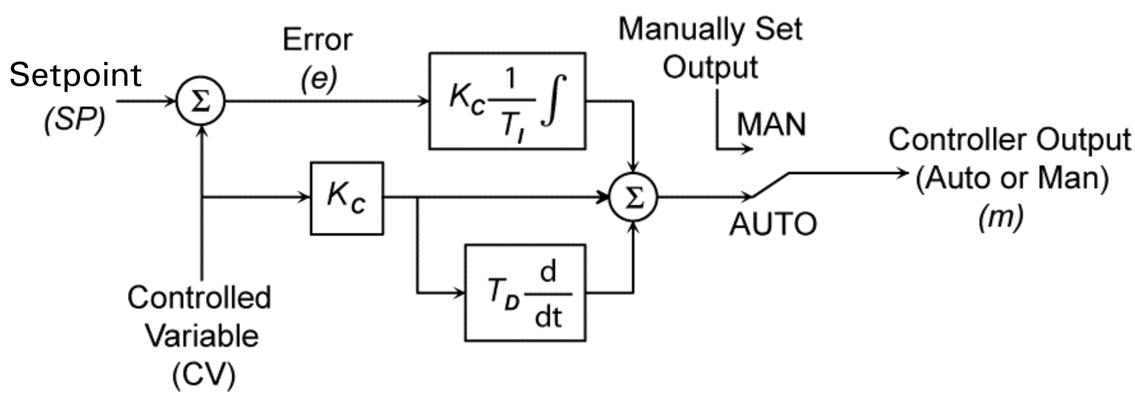


Figure 3-8: Block Diagram of Ideal PID Algorithm with Both Proportional and Derivative Modes on Measurement

3.3.8 Discrete Forms of PID

The algorithm forms presented above, using calculus symbols, are applicable to analog controllers that operate continuously. However, control algorithms implemented in a digital system are processed at discrete sample instants (for instance, one-second intervals), rather than continuously. Therefore, a modification must be made to show how a digital system approximates the continuous forms of the algorithm presented above. Digital processing of the PID algorithm also presents an alternative that was not present in analog systems. At each sample instant the PID algorithm can calculate either a new position for the controller output or the increment by which the output should change. These forms are called the *position* and the *velocity* forms, respectively. Assuming that the controller is in the automatic mode, the following equations are processed at each sample instant for the position algorithm. The subscript "*n*" refers to the nth processing instant, "*n-1*" to the previous processing instant, and so on.

Compute the error: $$e_n = SP_n - CV_n \tag{3-3}$$

Increment sum of errors: $$S_n = S_{n-1} + e_n$$

Compute controller output: $$m_n = K_C\left[e_n + \frac{\Delta T}{T_I}S_n + \frac{T_D}{\Delta T}\left(e_n - e_{n-1}\right)\right]$$

Save S_n and e_n values for the subsequent processing time.

The velocity mode or incremental algorithm is similar. It computes the amount by which the controller output should be changed at the n^{th} sample instant.

Compute change in controller output:

$$\Delta m_n = K_C\left[\left(e_n - e_{n-1}\right) + \frac{\Delta T}{T_I}e_n + \frac{T_D}{\Delta T}\left(e_n - 2e_{n-1} + e_{n-2}\right)\right] \tag{3-4}$$

Add the incremental output to the previous value of controller output, to create new value of output:

$$m_n = m_{n-1} + \Delta m_n \tag{3-5}$$

Save m_n, e_{n-1}, and e_{n-2} values for the subsequent processing time.

From a user point of view, there is no advantage of one form over the other. Vendors, however, may prefer a particular form due to the ease of incorporation of features of their system, such as tuning and bumpless transfer.

The configuration options—DA/RA, derivative- and proportional-on-measurement, or error—are also applicable to the discrete forms of PID. In fact, there are more user configuration options offered with digital systems than were available with analog controllers.

3.4 Controller Tuning

In the previous section, it was mentioned the parameters K_C, T_I (or their equivalents, proportional band and reset rate), and T_D must be adjusted so the response of the controller matches the requirements of a particular process. This is called "tuning the controller." There are no hard and fast rules as to the performance requirements for tuning. These are largely established by the particular process application and by the desires of the operator or controller tuner.

3.4.1 Acceptable Criteria for Loop Performance

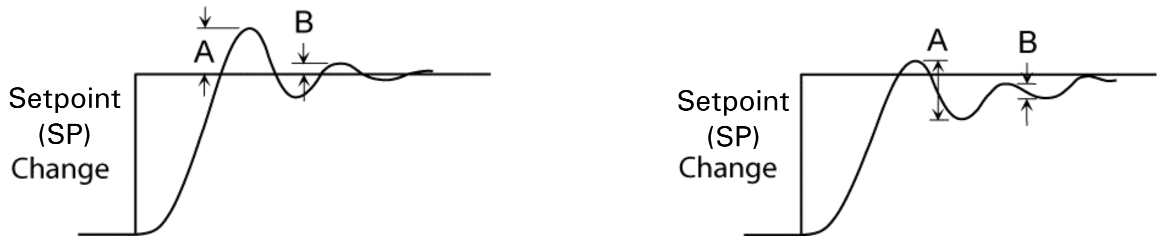
One widely used response criterion is the loop should exhibit a quarter-amplitude decay following a setpoint change. See Figure 3-9. For many applications, however, this is too oscillatory. A smooth response to a setpoint change with minimum overshoot is more desirable. A response to setpoint change that provides minimum overshoot is considered less aggressive tuning than quarter-amplitude decay. The penalty for less aggressive tuning is that a disturbance will cause a greater deviation from setpoint or a longer time to return to setpoint. The controller tuner must decide the acceptable criterion for loop performance before actual tuning.

Controller tuning techniques may be divided into two broad categories: those that require testing of the process, either with the controller in automatic or manual, and those that are less formal, often called *trial-and-error* tuning.

3.4.2 Tuning from Open Loop Tests

The open-loop process testing method uses only the manually-set output of the controller. A typical response to a step change in output was shown in Figure 3-4. It is often possible to approximate the response with a simplified process model containing only three parameters—the process gain (K_p), the dead time in the process (T_d), and the process time constant (τ_p). Figure 3-10 shows the response of a first-order-plus-dead-time (FOPDT) model that approximates the true process response.

Figure 3-10 also shows the parameter values, K_p, T_d and τ_p. There are a number of published correlations for obtaining controller tuning parameters from these process parameters. The best known is based upon the Ziegler-Nichols reaction curve method. Correlations for P-only, PI, and PID controllers are given here in Table 3-1.



$$Decay\ Ratio = \frac{Second\ deviation\ from\ SP}{First\ deviation\ from\ SP} = \frac{B}{A}$$

a. Decay Ratio Definition in Normal Situations

$$Decay\ Ratio = \frac{Second\ (\ peak - valley\ )}{First\ (\ peak - valley\ )} = \frac{B}{A}$$

b. Decay Ratio Definition in Abnormal Situations
(Also can be used to determine decay ratio following a disturbance)

Figure 3-9: Decay Ratio Definitions

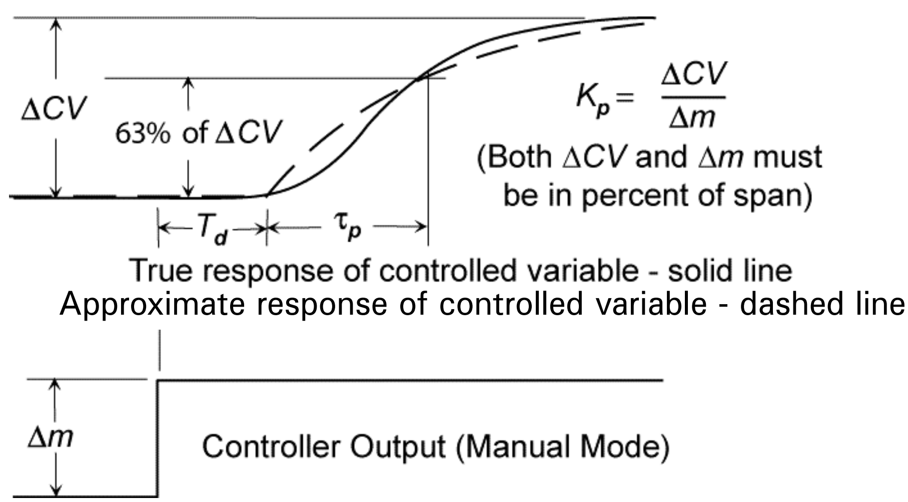


Figure 3-10: Approximating the Open-Loop (Controller in Manual) Response to Step Change in Output with a Simplified Process Model

Table 3-1: Controller Tuning Parameters Based Upon Open Loop Test Data

	P-Only	**PI**	**PID**
K_C	$\frac{\tau_p}{K_p T_d}$	$\frac{0.9\tau_p}{K_p T_d}$	$\frac{1.2\tau_p}{K_p T_d}$
T_I	---	$3.33T_d$	$2.0T_d$
T_D	---	---	$0.5T_d$

Another tuning technique that uses the same open-loop process test data is called "lambda tuning." The objective of this technique is for the setpoint response to be an exponential rise with a specified time constant, λ. This technique is applicable whenever it is desired to have a very smooth setpoint response, at the expense of degraded response to disturbances.

There are other elaborations of the open-loop test method, including multiple valve movements in both directions, numerical regression methods for obtaining the process parameters, etc. Despite its simplicity, the open-loop method suffers from the following problems:

- It may not be possible to interrupt normal process operations to make the test.
- If there is noise on the measurement, it may not be possible to get good data, unless the controlled variable change is at least five times the amplitude of the noise. For many processes, that may be too much disturbance.
- The technique is very sensitive to parameter estimation error, particularly if the ratio of T_d/τ_p is small.
- The method does not take into consideration the effects of valve stiction.
- The actual process response may be difficult to approximate with an FOPDT model.
- A disturbance to the process during the test will severely deteriorate the quality of the data.
- For very slow processes, the complete results of the test may require one or more working shifts.
- The data is valid only at one operating point. If the process is nonlinear, additional tests at other operating points may be required.

Despite these problems, under relatively ideal conditions—minimal process noise, minimal disturbances during the test, minimal valve stiction, etc.—the method provides acceptable results.

3.4.3 Tuning from Closed Loop Tests

Another technique is based on testing the process in the closed-loop. (Ziegler-Nichols referred to this as the "ultimate sensitivity" method.) To perform this test, the controller is placed in the automatic mode, integral and derivative actions are removed (or a proportional-only controller is used), a low controller gain is set, then the process is disturbed—either by a setpoint change or a forced disturbance—and the oscillating characteristics are observed. The objective is to repeat this procedure with increased gain until sustained oscillation (neither increasing nor decreasing) is achieved. At that point, two pieces of data may be obtained: the value of controller gain (called the "ultimate gain," K_{CU}) that produced sustained oscillation, and the period of the oscillation, P_U. With this data, one can enter Table 3-2 and calculate tuning parameters for a P-only, PI, or PID controller.

Table 3-2: Controller Tuning Parameters Based Upon Closed Loop Test Data

	P-Only	**PI**	**PID**
K_C	$0.5K_{CU}$	$0.45K_{CU}$	$0.6K_{CU}$
T_I	---	$0.83P_U$	$0.5P_U$
T_D	---	---	$0.125P_U$

There are also problems with the closed-loop method.

- It may not be possible to subject the process to a sustained oscillation.

- Even if that were possible, it is difficult to predict or to control the magnitude of the oscillation.
- Multiple tests may be required, resulting in long periods of interruption to normal operation.

Despite these problems, there are certain advantages to the closed-loop method.

- Minimal uncertainty in the data. (Frequency, or its inverse, period, can be measured quite accurately.)
- The method inherently includes the effect of a sticking valve.
- Moderate disturbances during the testing can be tolerated.
- No *a priori* assumption as to the form of the process model is required.

A modification of the closed-loop method, called the *relay method*, attempts to exploit the advantages while circumventing most of the problems. The relay method utilizes the establishment of maximum and minimum limits for the controller output. For instance, if controller output normally is 55%, the maximum output can be set at 60% and the minimum at 50%. While this does not establish hard limits for excursion of the controlled variable, persons familiar with process will feel comfortable with these settings or will reduce the difference between the limits.

The process is then tested by a setpoint change or a forced disturbance, using an on-off controller. If an on-off controller is not available, then a P-only controller with a maximum value of controller gain can be substituted. The controlled variable will oscillate above and below setpoint, with the controller output at either the maximum or minimum value, as shown in Figure 3-11.

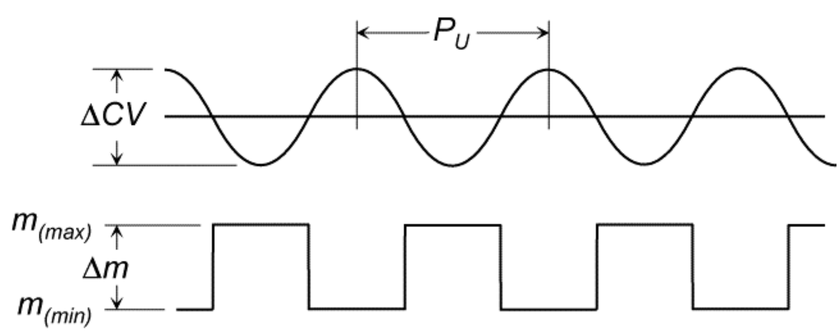


Figure 3-11: On-Off Controller Output and Oscillating Controlled Variable in a Relay Test

If the period of time when the controller output is at the maximum setting exceeds the time at the minimum, then both the maximum and limits should be shifted upward by a small but identical amount. After one or more adjustments, the output square wave should be approximately symmetrical. At that condition, the period of oscillation, P_U, is the same as would have been obtained by the previously described closed-loop test. Furthermore, the ultimate gain can be determined from a ratio of the controller output and CV amplitudes:

$$K_{CU} = \frac{4}{\pi}\frac{\Delta m}{\Delta CV} \tag{3-6}$$

Thus the data required to enter Table 3-2 and calculate tuning parameters has been obtained in a much more controlled manner than the unbounded closed loop test.

While the relay method is a viable technique for manual testing, it can also be easily automated. For this reason, it is the basis for some vendors' self-tuning techniques.

3.4.4 Trial-and-Error Tuning

Despite these tools for formal process testing for determination of tuning parameters, many loops are tuned by trial-and-error. That is, an unsatisfactory loop closed-loop behavior is observed, and an estimate (often merely a guess) is made as to which parameter(s) should be changed and by how much. Good results often depend upon the person's experience. Various methods of visual pattern recognition have been described but, in general, such tuning techniques remain more of an art than a science.

A recently published technique, Wade's *Basic and Advanced Regulatory Control: System Design and Application* (see 3.6.1) called *improving as-found tuning*, or "intelligent trial-and-error tuning," attempts to place controller tuning on a more methodological basis. The premise of this technique, which is applicable only to PI controllers, is that a well-tuned controller exhibiting a slight oscillation (oscillations that are decaying rapidly) will have a predictable relation between the integral time and period of oscillation. The following relation has been found to provide acceptable results:

$$1.5 \le \frac{P}{T_I} \le 2.0 \tag{3-7}$$

Further insight into this technique can be gained by noting that the phase shift through a PI controller, from error to controller output, depends very strongly on the ratio P/T_I and only slightly on the decay ratio. For a control loop with a quarter-amplitude decay, the limits above are equivalent to specifying a phase shift of approximately 15°.

If a control system engineer or instrumentation technician is called upon to correct the errant behavior of a control loop, then (assuming that it is a tuning problem and not some external problem) the "as-found" behavior is caused by the "as-found" tuning parameter settings. The behavior can be characterized by the decay ratio *(DR)* and the period *(P)* of oscillation. The as-found data set—K_C, T_I, *DR*, *P*—represents a quanta of knowledge about the process. If either an open-loop or closed-loop test were made in an attempt to determine tuning parameters, then the existing knowledge about the process would be sacrificed.

From Equation 3-7, upper and lower limits for an acceptable period can be established.

$$1.5\,T_I \le P \le 2.0\ T_I \tag{3-8}$$

If the as-found period *P* meets this criteria, the implication is the integral time is acceptable. Hence, adjustments should be made to the controller gain K_C until the desired decay ratio is obtained. If the period is outside this limit, then the present period can be used in the inverted relation to determine a range of acceptable new values for T_I:

$$0.5\ P \le T_I \le 0.67\,P \tag{3-9}$$

Wade's *Basic and Advanced Regulatory Control: System Design and Application* (see 3.6.1) and "Trial and error: an organized procedure" (see 3.6.2) contain more information, including a flow chart, describing this technique.

3.4.5 Self-Tuning

Although self-tuning, auto-tuning and adaptive-tuning have slightly different connotations, they will be discussed collectively here.

There are two different circumstances where some form of self-tuning would be desirable:

1. If a process is highly nonlinear and also experiences a wide range of operating points, then a technique that automatically adjusts the tuning parameters for different operating conditions would be highly beneficial.

2. If a new process unit with many control loops is to be commissioned, it would be beneficial if the controllers could determine their own best tuning parameters.

There are different technologies that address these situations.

For initial tuning, there are commercial systems that in essence automate the open-loop test procedure. On command, the controller will revert to the manual mode, test the process, characterize the response by a simple process model, then determine appropriate tuning parameters. Most commercial systems that follow this procedure display the computed parameters and await confirmation before entering the parameters into the controller. An automation of the relay tuning method described previously falls into this category.

The simplest technique addressing the nonlinearity problem is called *scheduled tuning.* If the nonlinearity of a process can be related to a key parameter such as process throughput, then a measure of that parameter can be used as an index to a lookup table (schedule) for appropriate tuning parameters. The key parameter may be divided into regions, with suitable tuning parameters listed for each region. Note that this technique depends upon the correct tabulation of tuning parameters for each region. There is nothing in the technique that evaluates the loop performance and automatically adjusts the parameters based upon the evaluation.

There are also systems that attempt to recognize features of the response to normal disturbances to the loop. From these features, heuristic rules are used to calculate new tuning parameters. These may be displayed for confirmation, or they may be entered into the algorithm "on the fly." Used in this manner, the system tries to adapt the controller to the random environment of disturbances and setpoint changes as they occur.

There are also "third party" packages, typically running in a notebook computer, that access data from the process, such as by transferring data from the DCS data highway. The data is then analyzed and advisory messages are presented that suggest tuning parameters and provide an indication of the "health" of control loop components, especially the valve.

3.5 Advanced Regulatory Control

If the process disturbances are few and not severe, feedback controllers will maintain the average value of the controlled variable at setpoint. But in the presence of frequent or severe disturbances, feedback controllers permit significant variability in the control loop. This is because a feedback controller *must* experience a deviation from setpoint in order to change its output. This variability may result in an economic loss. For instance a process may operate at a safe margin away from a target value to prevent encroaching on the limit and producing off-spec product. Reducing the margin of safety will produce some economic benefit, such as reduced energy consumption, reduced raw material usage or increased production. Reducing the variability cannot be done by feedback controller tuning alone. It may be accomplished by the use of more advanced control loops such as ratio, cascade, feedforward, decoupling, and selector control.

3.5.1 Ratio Control

Often, when two or more ingredients are blended or mixed, the flow rate of one of the ingredients paces the production rate. The flow rates for the other ingredients are controlled to maintain a specified ratio to the pacing ingredient. Figure 3-12 shows a ratio control loop. Ratio control systems are found in batch processing, fuel oil blending, combustion processes where the air flow may be ratioed to the fuel flow, and many other applications. The pacing stream is often called the "wild" flow, since it may or may not be provided with an independent flow rate controller—only a measurement of the wild flow stream is utilized in ratio control.

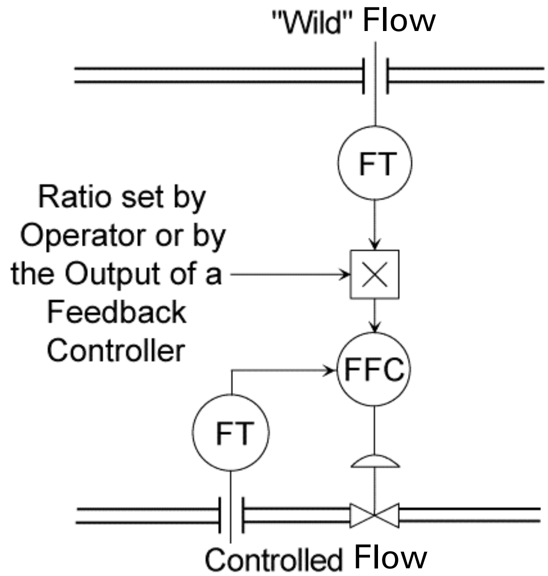


Figure 3-12: Ratio Control Strategy

The specified ratio may be manually set, automatically set from a batch recipe, or adjusted by the output of a feedback controller. An example of the latter is a process heater that uses a stack oxygen controller to adjust the air-to-fuel ratio. When the required ratio is automatically set by a higher-level feedback controller, the ratio control strategy is merely one form of feedforward control.

3.5.2 Cascade Control

Cascade control refers to control schemes that have an inner control loop nested within an outer loop. The feedback controller in the outer loop is called the "primary" controller. Its output sets the setpoint for the inner loop controller, called the "secondary." The secondary control loop must be significantly faster than the primary loop. Figure 3-13 depicts an example of cascade control applied to a heat exchanger. In this example a process fluid is heated with a hot oil stream. A temperature controller on the heat exchanger output sets the setpoint of the hot oil flow controller.

If the temperature controller directly manipulated the valve, there would still be a valid feedback control loop. Any disturbance to the loop, such as a change in the process stream flow rate or a change in hot oil supply pressure, would require a new position of the control valve. Therefore, a deviation of temperature from setpoint would be required to move the valve.

With the secondary loop installed as shown in Figure 3-13, a change in hot oil supply pressure will result in a change in hot oil flow. This will be rapidly detected by the flow controller which will then make a compensating adjustment to the valve. The momentary variation in hot oil flow will cause minimal, if any, disturbance to the temperature control loop.

In the general situation, all disturbances within the secondary loop—a sticking valve, adverse valve characteristics, or (in the example) variations in supply pressure—are confined to the secondary loop and have minimal effect on the primary controlled variable. A disturbance that directly affects the primary loop, such as a change in process flow rate in the example, will require a deviation at the primary controller for its correction regardless of the presence or absence of a secondary controller.

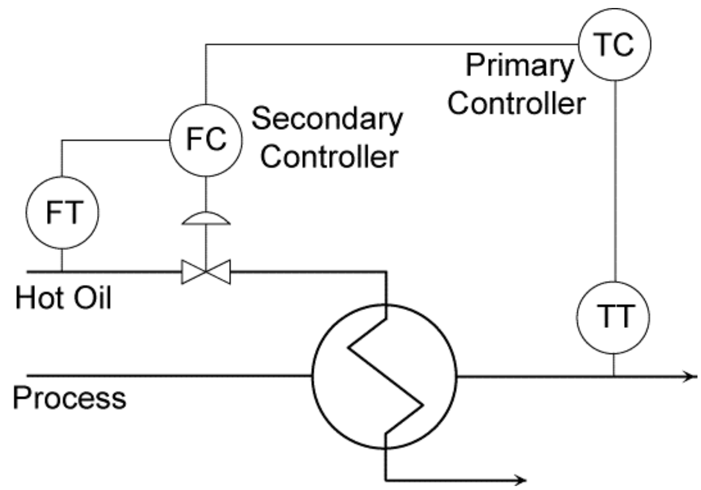


Figure 3-13: Example of Cascade Control Strategy

When examining a process control system for possible improvements, consider whether intermediate control loops can be closed to encompass certain of the disturbances. If so, the effect of these disturbances will be removed from the primary controller.

3.5.3 Feedforward Control

Feedforward control is defined as the manipulation of the final control element—valve position or setpoint of a lower-level flow controller—using a measure of a disturbance rather than the output of a feedback controller. In essence, feedforward control is open loop control. Feedforward control requires a process model in order to know how much and when correction should be made for a given disturbance. If the process model were perfect, feedforward control alone could be used. In actuality, the process model is never perfect; therefore, feedforward and feedback control are usually combined.

The example in the previous section employed cascade control to overcome the effect of disturbances caused by variations in hot oil supply pressure. It was noted, however, that variations in process flow rate would still cause a disturbance to the primary controller. If the process and hot oil flow rates varied in a proportionate amount, there would be only minimal effect on the process outlet temperature. Thus a ratio between the hot oil and process flow rates should be maintained. While this would eliminate most of the variability at the temperature controller, there may be other circumstances, such as heat exchanger tube scaling, that would necessitate a long-term shift in the required ratio. This can be implemented by letting the feedback temperature controller set the required ratio as shown in Figure 3-14.

Ratio control, noted earlier as an example of feedforward-feedback control, corrects for the steady-state effects on the controlled variable. Suppose that there is also a difference in dynamic effects of the hot oil and process streams on the outlet temperature. In order to synchronize the effects at the outlet temperature, *dynamic compensation* may be required in the feedforward controller.

To take a more general view of feedforward, consider the generic process shown within the dotted lines in Figure 3-15. This process is subject to two influences (inputs)—a disturbance and a control effort. The control effort may be the signal to a valve or to a lower level flow controller. In this latter case, the flow controller can be considered as a part of the process. Transfer functions *A(s)* and *B(s)* are mathematical abstractions of the dynamic effect of each of the inputs on the controlled variable. A feedforward controller *C(s),* a feedback controller, and the junction combining feedback and feedforward are also shown in Figure 3-15.

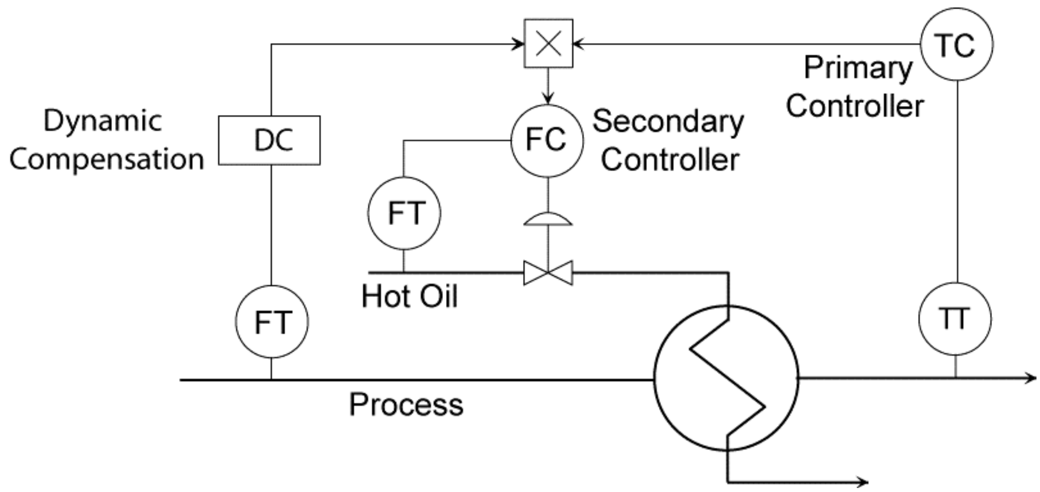


Figure 3-14: Combined Feedback-Feedforward Control with Dynamic Compensation

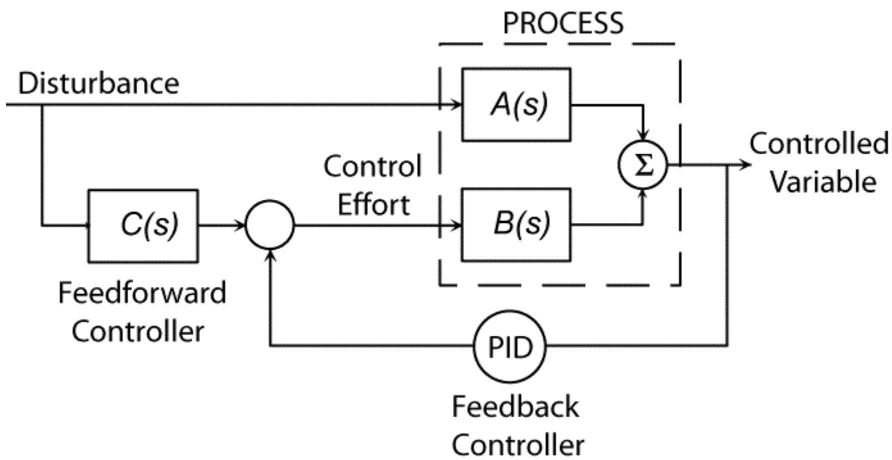


Figure 3-15: Generic Feedback-Feedforward Control Structure

There are two paths of influence from the disturbance to the controlled variable. If the disturbance is to have no effect on the controlled variable (that is the objective of feedforward control), these two paths must be mirror images that cancel out each other. Thus the feedforward controller must be the ratio of the two process dynamic effects, with an appropriate sign adjustment. The correct sign will be obvious in any practical situation. That is:

$$C(s) = -\frac{A(s)}{B(s)} \tag{3-10}$$

If both A(s) and B(s) have been approximated as FOPDT models (see Section 3.4.2), then *C(s)* is comprised of, at most, a steady-state gain, a lead-lag and a dead-time function. These functions are contained in every vendor's function block library. The dynamic compensation can often be simpler than this. For instance, if the dead times through *A(s)* and *B(s)* are identical, then no dead-time term is required in the dynamic compensation.

Now consider combining feedback and feedforward control. Figure 3-15 shows a junction for combining these two forms of control but does not indicate how they are combined. In general, feedback and feedforward can be combined by adding or by multiplying the signals. A multiplicative combination is essentially the same as ratio control. In situations where a ratio must be maintained between disturbance and control effort, multiplicative combination of feedback and feedforward will provide a relatively constant process gain for the feedback controller. If the feedback and feedforward were combined additively, variations in process gain seen by the feedback controller would require frequent retuning. In other situations, it is better to combine feedback and feedforward additively, a control application often called "feedback trim."

Regardless of the method of combining feedback and feedforward, the dynamic compensation terms should be only in the feedforward path, not the feedback path. It would be erroneous for the dynamic compensation terms to follow the combining junction in Figure 3-15.

Feedforward control is one of the most powerful control techniques for minimizing variability in a control loop. It is often overlooked due to lack familiarity with the technique.

3.5.4 Decoupling Control

Frequently in industrial processes, a manipulated variable—a signal to a valve or to a lower-level flow controller—will affect more than one controlled variable. If each controlled variable is paired with a particular manipulated variable through a feedback controller, interaction between the control loops will lead to undesirable variability.

One way of coping with the problem is to pair the controlled and manipulated variables so as to reduce the interaction between the control loops. A technique for pairing the variables, called *relative gain analysis*, is described in most texts on process control, as well as in both books referenced in 3.6.1. If, after applying this technique, the residual interaction is too great, the control loops should be modified for the purpose of decoupling. With decoupled control loops, each feedback controller output affects only one controlled variable.

Figure 3-16 shows a generic process with two controlled inputs—a signal to valves or setpoints to lower-level flow controllers—and two controlled variables. The functions P_{11}, P_{12}, P_{21} and P_{22} represent dynamic influences of inputs on the controlled variables. With no decoupling, there will be interaction between the control loops. However, decoupling elements can be installed so that the output of PID#1 has no effect on CV#2, and PID#2 output has no effect on CV#1.

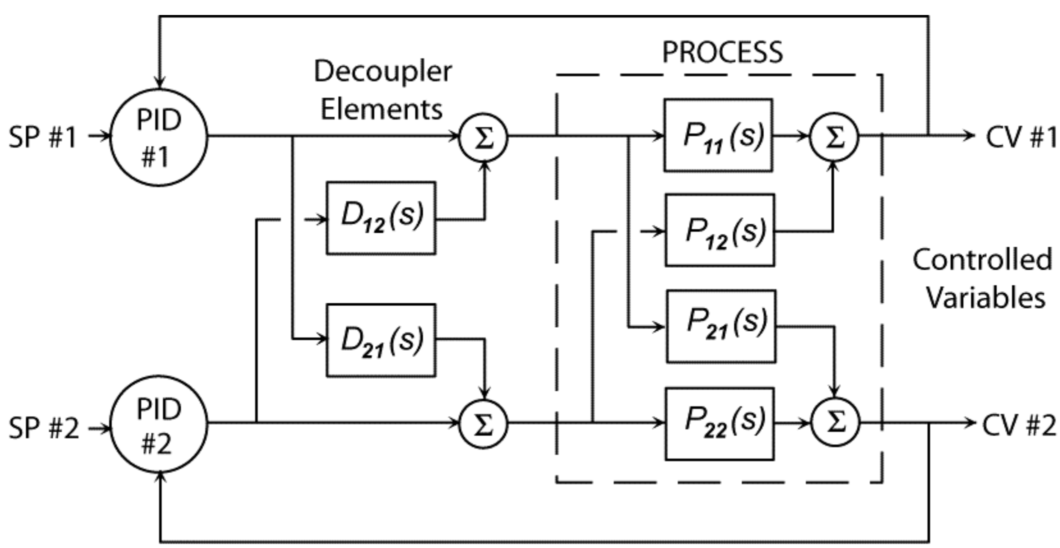


Figure 3-16: Multiple-Input, Multiple-Output (2x2) Process with Decoupled Feedback Control Loops

Using an approach similar to feedforward control, note that there are two paths of influence from the output of PID#1 to CV#2. One path is through the process element $P_{21}(s)$. The other is through the decoupling element $D_{21}(s)$ and the process element $P_{22}(s)$. For the output of PID#1 to have no effect on CV#2 these paths must be mirror images that cancel out each other. Therefore, the decoupling element must be

$$D_{21}(s) = -\frac{P_{21}(s)}{P_{22}(s)} \tag{3-11}$$

In a practical application, the appropriate sign will be obvious. In a similar fashion, the other decoupling element is given by

$$D_{12}(s) = -\frac{P_{12}(s)}{P_{11}(s)} \tag{3-12}$$

If the process elements are approximated with FOPDT models as in Section 3.4.2, the decoupling elements are, at most, comprised of gain, lead-lag and dead-time functions, all of which are available from most vendors' function block library.

The decoupling technique described here can be called "forward decoupling." Inverted decoupling, an alternative described in Wade's "Inverted Decoupling, A Neglected Technique" (see 3.6.2), has certain advantages as well as possible disadvantages.

If one variable is of greater priority than the other, partial decoupling should be considered. Suppose that CV#1 in Figure 3-16 is a high-valued product and CV#2 is a low-valued product. Variability in CV#1 should be minimized, whereas variability in CV#2 can be tolerated. Therefore the decoupling element $D_{12}(s)$ can be implemented and $D_{21}(s)$ omitted.

3.5.5 Selector (Override) Control

Selector control, also called "override" control, differs from the other techniques because it does not have as its objective the reduction of variability in a control loop. It does have an economic consequence, however, because the most economical operating point for many processes is near the point of encroachment on a process, equipment, or safety limit. Unless a control system is present that prevents such encroachment, the tendency will be to operate well away from the limit, at a less-than-optimum operating point. Selector control permits operating closer to the limit.

As an example, Figure 3-17 illustrates a process heater. In normal operation, an outlet temperature controller controls the firing rate of the heater. During this time, a critical tube temperature is below its limit. Should, however, the tube temperature encroach on the limit, the tube temperature controller will override the normal outlet temperature controller and reduce the firing rate of the heater. The low-signal selector in the controller outputs provides for the selection of the controller that is demanding the lower firing rate.

If ordinary PI or PID controllers are used for this application, one or the other of the controlled variables will be at its setpoint, with the other variable less that its setpoint. The integral action of the non-selected controller will cause it to wind up—that is, its output will climb to 100%. In normal operation, this will be the tube temperature controller. Should the tube temperature rise above its setpoint, its output must unwind from 100% to a value that is less than the other controller's output before there is any effect on heater firing. Depending upon the controller tuning, there may be a considerable amount of time when the tube temperature is above its limit.

When the tube temperature controller overrides the normal outlet temperature controller and reduces heater firing, there will be a drop in heater outlet temperature. This will cause the outlet temperature

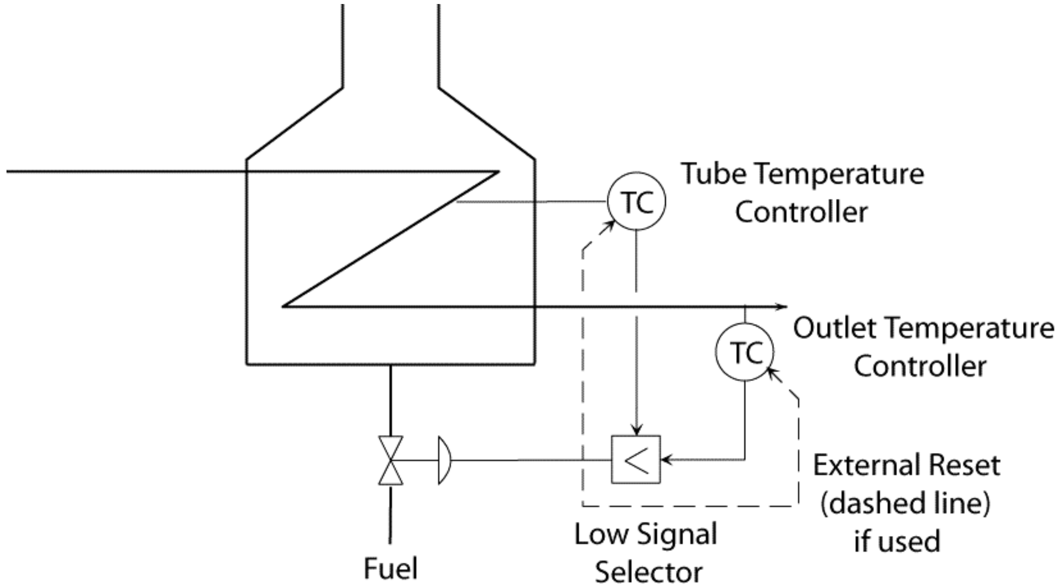


Figure 3-17: Application Example of the Use of Selector (Override) Control

controller to wind up. Once the tube temperature is reduced, returning to normal outlet temperature control is as awkward as was the switch to tube temperature control.

These problems arise because ordinary PID controllers were used in the application. Most vendors have PID algorithms with alternative functions to circumvent these problems. Two techniques will be briefly described.

Some vendors formulate their PID algorithm with "external reset." The integral action is achieved by feeding the output of the controller back to a positive feedback loop that utilizes a unity-gain first-order lag. With the controller output connected to the external feedback port, the response of a controller with this formulation is identical to that of an ordinary PID controller. Different behavior occurs when the external reset feedback is taken from the output of the selector, as shown in Figure 3-17. The non-selected controller will not wind up. Instead, its output will be equal to the selected controller's output plus a value representing its own gain times error. As the non-selected controlled variable (for instance, tube temperature) approaches its limit, the controller outputs become more nearly equal, but with the non-selected controller's output being higher. When the non-selected controller's process variable reaches the limit, the controller outputs will be equal. Should the non-selected controller's process variable continue to rise, its output will become the lower of the two—hence it will be selected for control. Since there is no requirement for the controller to unwind, the switch-over will be immediate.

Other systems do not use the external feedback. The non-selected controller is identified from the selector switch. As long as it remains the non-selected controller, it is continually initialized so that its output equals the other controller output plus the value of its own gain times error. This behavior is essentially the same as external feedback.

There are many other examples of selector control in industrial processes. On a pipeline, for instance, a variable speed compressor may be operated at the lower speed demanded by suction and discharge pressure controllers. For distillation control, reboiler heat may be set by the lower of the demands of a composition controller and a controller of differential pressure across one section of a tower, indicative of tower flooding.

3.6 References

3.6.1 Books

The Automation, Systems, and Instrumentation Dictionary, Fourth Edition. ISA, 2003.

Shinskey, F.G. *Process Control Systems: Application Design and Tuning.* 4th Ed. McGraw-Hill, 1996.

Wade, H.L. *Basic and Advanced Regulatory Control: System Design and Application*. ISA, 2004.

3.6.2 Articles

Wade, H.L. "Inverted Decoupling, A Neglected Technique." *ISA Transactions,* Vol. 36, No. 1. 1997.

Wade, H.L. "Trial and error: An organized procedure." *InTech,* May 2005.

3.6.3 Standards

ISA-5.1-1984 (R1992), *Instrumentation Symbols and Identification.*

About the Author

Harold Wade, Ph.D., is president of Wade Associates, Inc., a Houston-based consulting firm specializing in control systems engineering, instrumentation, and process control training. He has more than 40 years of instrumentation and control industry experience with Honeywell, Foxboro, and Biles and Associates. A Senior Member of ISA and a licensed professional engineer, he is the author of *Basic and Advanced Regulatory Control: System Design and Application*, 2nd Ed., published by ISA in 2004. He started teaching for ISA in 1987. He was a 2002 inductee into *Control Magazine's* "Automation Hall of Fame."