CHAPTER 14

Process Control Fundamentals

14.1 INTRODUCTION

Processing natural gas to profitably produce end products requires precision and is potentially hazardous. Small changes in a process can have a large impact on the end result. Conditions such as pressure, temperature, flow, composition, and many other factors must be controlled carefully and consistently within specified limits to ensure quality and safety. Process control enables the safe and reliable production of natural gas and its derivatives while reducing manufacturing costs such as raw materials and energy.

This chapter introduces readers to a better understanding of the fundamental concepts in dynamics and process control, which include the ability to select a process control configuration for specific applications, decide on suitable instruments for process monitoring, develop input-output relationships, understand the dynamic open-loop and closed-loop characteristics, and design and tune feedback controllers.

14.2 DYNAMIC PROCESS CHARACTERISTICS

The dynamic process characteristics of most processes are described by three elements: resistance, capacitance, and dead time contributions. These elements will determine how the process responds to changes.

14.2.1 Resistance-Type Processes

A characteristic of a resistance element is the ability to transfer material or energy. Flow through a pipe is the most common example of a resistance-type process. A strictly resistance-type process has a proportional-only response. Any change in the resistance—for example, a control valve opening—will result in an immediate proportional change in the flow. The amount of change is a function of the process gain. Any change in the load—for example, the upstream and downstream pressure in this case of flow through a pipe—will also result in an immediate proportional change in the flow. The amount of change is again a function of the process gain.

14.2.2 Capacitance-Type Processes

A characteristic of a capacitance element is the ability to store energy and mass. Thermal capacitance is simply the product of the mass of an object and its specific heat. Energy can be stored in a heat exchanger. Mass capacitance is the accumulation of mass. In other words, input does not always equal output. Separators and surge dampers are common examples of strictly mass capacitance processes.

The gas capacitance of a tank is constant. The liquid or solid capacitance equals the cross-sectional area of the tank at the surface. If the cross-sectional area is constant, then the capacitance is constant at any head.

An example of a strictly capacitive process is a tank containing liquid with liquid removed at a constant rate—for instance, a positive displacement pump. The change in level of the tank is the product of the difference between the inlet and outlet flow rate and the time that the tank has been filling divided by the capacitance, or cross-sectional area, of the tank. Since the capacitance of the tank can change with level and therefore time—for instance, a sphere or horizontal tank with elliptical heads—the capacitance is a differential.

The larger the vessel is in relation to the flows, the larger the capacitance and hence the more slowly the controlled variable changes for a given change in the manipulated variable. The capacitance of the process tends to attenuate disturbances and hence makes control less difficult. The capacitance of the process is its time constant. The time constant is calculated from the differential equations used to model the process, but as an approximation is roughly equal to the process residence time.

14.2.3 Process Dead Time

Another contributing factor to the dynamics of many processes is a transportation lag or dead time. Dead time is the delay in time for an output to respond to an input. For example, if the flow into a pipeline is increased, a period of time elapses before a change in the outlet flow is detected. This elapsed time is a function of the line length and fluid velocity. Since dead time is the time required to move material from one point to another, it is frequently referred to as "transportation lag." Some examples of dead time include static mixers and conveyor belts. Dead time is also encountered when energy or mass sources must flow to the destination where temperature, composition, flow, or pressure is sensed. Delay also arises when a control signal requires time to travel between two points.

Dead time interferes with good control inasmuch as it represents an interval during which the controller has no information about the effect of a load change or control action already taken. In the design of a control system, every attempt should be made to minimize this delay by properly locating instruments and sampling points, ensuring sufficient mixing, and minimizing transmission lags.

Dead time is particularly difficult for feedback controllers for a variety of reasons. Feedback controllers are designed for immediate reaction, and dead time varies with throughput.

14.2.4 Inertia-Type Processes

The motion of matter as described by Newton's second law typifies inertia-type processes. These effects are important when fluids are accelerated or decelerated as well as in mechanical systems with moving components.

14.2.5 Combinations of Dynamic Characteristics

Few processes are strictly resistance or capacitance types. Systems that contain a resistance element and a capacitance element result in a single time-constant process. As noninteracting resistance and capacitance elements are added to the system, then multiple time constants result. When inertial effects or interactions between first-order resistance and capacitance elements are introduced, then processes exhibit a second-order response.

Single time-constant processes are described with a first-order differential equation containing a constant gain and a time constant. The gain is independent of the time characteristics and relates to the process amplification or ultimate output divided by input. The time constant describes the time required for the system to adjust to an input. This time constant is simply the product of the resistance and the capacitance.

The following equation describes a first-order process response without dead time:

$$\text{Change in Output} = \text{Gain} \times \text{Change in Input}\left(1 - e^{-\mathrm{t}/\text{time constant}}\right) \tag{14-1}$$

Two distinct principles apply to all first-order processes without dead time. First, the initial rate of change is the maximum, and, if unchanged, the system would reach the ultimate output in a period equal to the time constant. Second, the actual response is exponential with the output reaching 63.2% of the ultimate value at a time lapse equal to the time constant of the system. Dead time just shifts this response to begin after the delay has elapsed.

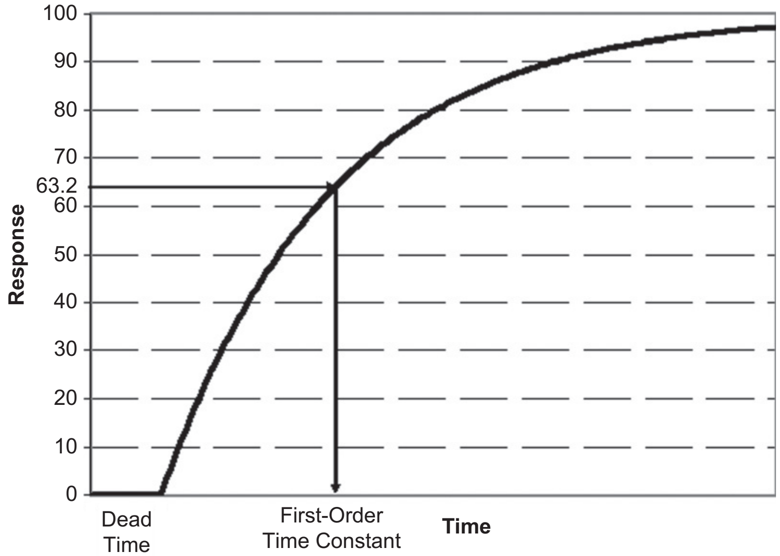

Figure 14-1 First-order open-loop process response.

Combinations of first-order lag with dead time, as shown in Figure 14-1, adequately estimate the response of most processes for purposes of simulation and control.

14.2.6 Examples: Simple Systems

14.2.6.1 Pipelines

Pipelines are controlled to provide a target pressure or flow at the end of the line. Pressure at the beginning of the pipeline is independent, as determined by a gas source or a compressor. The source pressure and the controlled pressure determine the flow rate.

The static gain (G) of flow to upstream pressure (P_1) is

$$G = F * P_1 / \left(P_1{}^2 - P_2{}^2\right) \tag{14-2}$$

where P_2 is the downstream pressure.

The sum of the dead time and time constant can be determined empirically as (Liptak, 1995)

$$K * V * MW * (P_1 - P_2) / T * Z * F \tag{14-3}$$

where K is 0.015 for English units; V is the volume of the pipeline in cubic feet; MW is the molecular weight of the gas; P is the pressure in psia; T is the gas temperature in °K; Z is the compressibility factor; and F is the flow in lb/sec.

For the conditions of $F = 1{,}500$ lb/sec, $P_1 = 2{,}000$ psia, $P_2 = 1{,}000$ psia, $V = 3{,}226{,}200$ cubic feet (approximately 48 inches diameter by 50 miles pipeline), $MW = 20$, $T = 50°F$ (10°C or 283°K), and $Z = 0.95$, the process gain will be 1 lb/sec per psi, and the dead time plus first-order time constant will be 2,400 seconds, or 40 minutes.

Pumps, compressors, and turbines have static relationships between the process variables, so constant-speed rotating equipment introduces no dynamics to the process. Variable-speed equipment usually introduces a very small, usually negligible, lag.

14.2.6.2 Vessels and Piping

Systems of vessels with connecting piping are simple examples of combined capacitance and resistance processes. The piping contributes a resistance element, and the vessels contribute a capacitance element to the system. A first-order system accurately describes these combinations.

For the example of a vessel filled with gas and incoming piping, the process gain between vessel pressure and incoming flow is a function of the pressure drop in the line to the vessel and the flow rate. The time constant is calculated rigorously from the following (Liptak, 1995):

$$2 * V * MW * (P_1 - P_2)/(R * T * F) \tag{14-4}$$

where V is volume of the vessel; MW is molecular weight of the gas; $P_1 - P_2$ is the pressure drop; R is ideal gas constant; T is temperature; and F is flow rate.

For a vessel with volume of 100 cubic feet, gas with molecular weight of 29, a vessel pressure of 100 psia (P_2), a source pressure of 105 psia (P_1), gas constant (10.732 psia-ft^3/lbmole-°R), gas temperature of 62.33°F (522°R), and gas inlet flow rate of 10 lb mass per minute, the time constant will be about 0.5 minute, or 30 seconds. For fluid temperature and composition variations, the length of pipe introduces a dead time.

14.2.6.3 Concurrent Heat Exchanger

Gas processing strategies commonly employ concurrent heat exchangers. There are six process variables involved: cold stream flow, hot stream flow, cold stream inlet temperature, cold stream outlet temperature, hot stream inlet temperature, and hot stream outlet temperature. Four of the variables are independent, whereas two are dependent. The flow rate of one stream and the inlet temperatures are based on other system considerations. This leaves the outlet temperatures and one flow rate as potential independent variables. Most often, an outlet temperature is the controlled variable, and an inlet flow rate or bypass flow is the manipulated variable.

Assuming that the temperature change of the tube wall dominates the transfer function, the time constant is approximated by the mass of the tubes and the specific heat of the tubes divided by the product of the flow rates and specific heats of the two process streams. The following empirical formula for the time constant can be derived (Liptak, 1995):

$$K * M_t * Cp_t/(F_w * Cp_w * F_c * Cp_c) \tag{14-5}$$

where $K = 100$ for English units.

For the case in which M_t (mass of tubes) is 4,400 lbm, Cp_t (specific heat of tubes) is 0.112 Btu/lb °F, F_w (flow of warm stream) is 55 lb/sec, Cp_w (specific heat of warm stream) is 0.70 Btu/lb °F, F_c (flow of cold stream) is 80 lb/sec, and Cp_c (specific heat of cold stream) is 0.80 Btu/lb °F, the time constant will be 20 seconds.

Note should be made that the dynamic considerations for furnaces and boilers are similar to heat exchangers.

14.2.7 Effects of Variable Conditions

In the preceding examples, the time constant depends on flow rate, which varies. Therefore, the time constant will vary with flow. The process gain and dead time will change with flow as well. Flow control valves can also introduce nonlinear responses due to changes in inlet or outlet pressure as well as the characteristics of the valve itself.

Instrumentation can also introduce gain variations. Temperature measurements typically introduce a time lag while composition analyzers normally introduce a dead time. Pneumatic control valves without positioners can introduce delays. Although digital control systems significantly reduce the delays experienced with analog systems, delays can occur due to scan frequency.

14.3 CLOSED-LOOP CONTROL

The most important aspect of control is the dynamic relationship between the measured, dependent, or controlled variable and the correcting, independent, or manipulated variable. In order to control the desired outcome of the measured variable, the response of the corrective action must be expressed as a function of time. In other words, the control dynamics must be known. The dynamic behavior of a closed-loop system depends on the gain or transfer function of the process and the gain or transfer function of the control system, including the influences of valve movements and control signal delays. The size and configuration of the equipment such as piping, vessels, heat exchangers, distillation columns, compressors, and pumps as well as the laws of chemistry and physics establish the process gain.

Knowledge of the process gain is necessary in order to provide stable control of the loop. Systems of differential equations can accurately calculate the process gain. The solution of the differential equation system with properly defined initial conditions provides the values of the process variables over time.

In the absence of a dynamic simulation, we can use the first-order lag with a dead-time model to approximate the process gain.

14.3.1 Degrees of Freedom

In the development of a control system for a process, it is desirable to know precisely the number of process variables, which the system designer is entitled to attempt to regulate, commonly known as the degrees of freedom of the process. In fact, an analysis of the dynamics of a system must begin with an analysis of the degrees of freedom. The degrees of freedom are simply the number of variables that describe the system less the independent relationships that exist among the variables. The degrees of freedom dictate the number of variables to specify when properly simulating a system. Once the plant is specified, the design of a control structure requires that the control degrees of freedom be known. This is the number of variables that are controllable. It is very easy to calculate this number, even for quite complex processes, because it is equal to the number of manipulated variables (the number of control valves in the process). For control system design, the degrees of freedom limit the number of automatic controllers. However, these variables are different from the design optimization variables. When optimizing, the degrees of freedom determine the number of variables that must be included in the objective function to yield a single solution.

Once a variable is determined to be a manipulated or specified variable, then it becomes an independent variable and uses a degree of freedom. Other variables that have a relationship to this variable are dependent variables.

14.3.2 Controllers

In order to choose controls properly, one must know the requirements of the process and the corresponding characteristics of the controller. The controller characteristics vary with the control mode. The controller mode ultimately chosen depends on the process dynamics, control objectives, and controller cost.

Feedback or closed-loop controllers compare a measurement (controlled variable) to its desired value (setpoint) and generate a correction signal (change in the controller output) to a control element (manipulated variable) based on the difference between the measurement and the desired value. The difference between the setpoint and measurement is the error. Negative feedback controllers act to eliminate the error. A controller can operate on the error in a variety of ways.

Control action is determined by the way the controller output responds to the error. In "direct" action, the controller output signal increases when the controller input exceeds the setpoint. Conversely, with "reverse" action, the controller output signal decreases when the controller input rises. Many controllers have a switch that selects direct or reverse action.

Controller, valve, and process must match with the correct choice of action. The control loop and the valve failure position must be determined first because these two decisions dictate the correct controller action.

There are two basic modes of control:

- Discrete (On/Off): The valve is either fully open or fully closed, with no intermediate state. Alternatively, a manipulated variable such as a motor is "on" or "off."
- Continuous: The valve can move between fully open or fully closed, or held at any intermediate position.

Some common modes of continuous control responses include proportional (throttling), integral (reset), derivative (rate), and combinations of these modes.

14.3.3 On/Off Control

The most simple control strategy, and consequently the least expensive in terms of initial cost and maintenance, is the on/off control. A room thermostat is an excellent example of on/off control. In the field, an on/off control is known as "snap-acting." Level controls on low-pressure separators and temperature controls on indirect fired heaters are often snap-acting. The controller will turn the manipulated variable on when the measurement deviates from setpoint and will turn the manipulated variable off when the measurement is at setpoint. In practical applications, a small error, or dead band, is allowed to eliminate constant cycling of the manipulated variable.

In many cases, the irregular value of the controlled or manipulated variable resulting from a snap-acting control creates serious operating and control problems downstream of the processes. In these cases, a proportional control is used.

14.3.4 Proportional Control

Proportional (P) control is the basic action employed in all controllers not using snap-action. Proportional control is a condition when the change in the controller output (*m*) is proportional to the change in the error (*e*). The constant magnitude of the change in the controller output to the change in error is the controller proportional gain (K_c):

$$m = K_c \cdot e + b \tag{14-6}$$

A bias (*b*) adjusts the output—for instance, 4 milliamps for electronic loops. If the controller output is 0%–100%, then there is no bias.

The proportionality constant is sometimes expressed in terms of percent proportional band (P_b) where

$$P_b = 100/K_c \tag{14-7}$$

In some control systems, proportional action is adequate to meet the control objectives. Proportional control will result in deviations from setpoint except when the setpoint is at 50% of span. When the gain is increased, the error is reduced but is not eliminated. Most processes become unstable as the error is eliminated unless the time constant is extremely high (very slow processes).

14.3.5 Integral Control

Integral (I) action is added to the controller to minimize or eliminate error. Integral control is the condition when the change in the controller output is proportional to the integral of the error. The prime purpose of the integral control is to prevent error and keep the controlled variable at the control point even as the process load changes. The integral algorithm constantly calculates the accumulated error and corrects for past over- or undercompensation with an integral time setting (T_i):

$$m = 1/T_i \cdot \int e\,dt + b \tag{14-8}$$

The integral time expressed in repeats per minute or minutes per repeat is the reset time, representing the time (in minutes) for the integral action to repeat.

Integral-only control is sometimes used in flow control systems and is suitable for pure time-delay processes. There are conditions when an integral controller will continue to accumulate error even when it is not desirable and the output will go to the maximum or minimum value. Some examples of undesirable action are (1) the process is not active, (2) a surge is not encountered for a surge controller, and (3) a cascade master is switched off cascade. An external reset or antireset windup feature is necessary to protect against saturation in the idle state.

Integral mode introduces a lag into the system and is most often used in conjunction with proportional mode. A single algorithm combines these modes where the proportional gain is applied to the integral error as well as the current error. The resulting algorithm is

$$m = K_c * (e + 1/T_i * \int e\,dt) + b \tag{14-9}$$

The integral rate requires careful adjustment. When integral rate is set correctly, valve movement occurs at a rate that the process can respond stably and robustly. If set too fast, cycling will result because the valve moves faster than the process and the measurement. If set too slow, the process will not recover quickly enough, resulting in sluggish control.

Integral action lags proportional action. The proportional action provides the quick response to correct for the upset, whereas the integral action provides the gradual correction to bring the controlled variable back to the setpoint.

Proportional-integral control has the advantage of reducing the initial error that is encountered with integral-only control while eliminating error.

14.3.6 Derivative Control

Proportional plus integral control does not provide correction that is rapid enough for certain processes. A derivative (D) response anticipates a change in process load and transmits a corrective signal to minimize the lag. This action corrects based on the rate of change of the deviation from the setpoint.

Derivative action leads proportional in that it causes the valve to move faster and further than it ordinarily would with proportional action alone. Temperature control systems normally require derivative action. These systems incur large process and measurement lags. Special applications such as antisurge systems also use derivative action because a rapid valve response is imperative.

Derivative control has the property that the change in the controller output (m) is proportional to the rate of change in the error (e). The derivative mode predicts errors and takes corrective action prior to occurrence of the error proportional to the derivative time (T_d):

$$m = T_d\, de/dt \tag{14-10}$$

The derivative time is the duration of time that the algorithm will look into the future. Larger derivative times will contribute larger corrective action. This corrective action will also protect against overshoot of the setpoint.

As the size of processing equipment increases, the momentum makes it difficult to control without derivative action. However, there are several limitations of derivative control. When the change in the error is constant, derivative mode takes no action. In addition, derivative control will react to setpoint changes and step changes in discontinuous measurements, such as chromatographs. For these reasons, derivative control is never used alone. In a few instances, proportional–derivative control may be used, such as special instances of slave controllers in a temperature cascade system and batch pH control.

14.3.7 Proportional-Integral-Derivative Control

While proportional and integral (PI) control can eliminate error and provide stability, considerable time may occur before the error returns to zero. Proportional, integral, and derivative (PID) controllers are used in processes with large capacitance or long time constants or slowly changing process outputs. Temperature and concentration loops are the most common examples.

One algorithm describes PID modes:

$$m = K_c * (e + 1/T_i \cdot \int e\, dt + T_d\, de/dt) + b \tag{14–11}$$

This algorithm can simultaneously respond to current error, eliminate error, and anticipate error. Much research has been conducted to determine optimum settings for proportional gain, integral time, and derivative time.

Sample and hold algorithms are used when the dead time of a loop is greater than the time constant. Conventional PID algorithms are not adequate in these instances. This algorithm utilizes the standard PID algorithm part of the time by switching the controller between automatic and manual. The period for which the controller is switched to manual is set by a timer to exceed the dead time and then alternates to automatic for an output update. In order to affect the required magnitude of change, the integral setting must be increased.

14.4 CONTROL LOOP TUNING

Controller tuning, which is the adjustment of the controller parameters to match the dynamic characteristics (or personality) of the entire control loop, has been referred to as the most important, least understood, and most poorly practiced aspect of process control. The tuning of feedback controllers is part of the overall commissioning of plants and machines. Often "trial and error" is used to achieve an acceptable combination of the tuning parameters for a particular process. "Good" control is a matter of definition and depends on such factors as individual preference, process disturbances and interactions, and product specifications.

The following sections discuss control quality and the performance criteria to consider when tuning a controller. They also examine methods of tuning PID controllers and the considerations for selection of the PID settings.

14.4.1 Quality of Control

If the PID controller parameters (the proportional, integral, and derivative terms) are chosen incorrectly, the controlled process input can be unstable; i.e., its output diverges, with or without oscillation, and is limited only by saturation or mechanical breakage. Tuning a control loop is the adjustment of its control parameters (gain/proportional band, integral gain/reset, derivative gain/rate) to the best values for the desired control response.

Good process control begins in the field, not in the control room. Sensors and measurements must be in appropriate locations, and valves must be sized correctly with appropriate trim. The final control elements, such as control valves, execute the changes required to manipulate the preferred process parameters such as flow, temperature, pressure, level, and ratio. If the instruments in the field do not function as required, then one cannot expect the overall process control to perform optimally. Tuning should be modified as the process and equipment change or degrade.

The controllability of a process depends on the gain that can be used. Higher gain yields greater rejection of disturbance and a faster response to setpoint changes. The predominant lag is based on the largest lag in the system. The subordinate lag is based on the dead time and all other lags. The maximum gain that can be used depends on the ratio of the predominant lag to the subordinate lag. From this ratio, we can draw two conclusions: (1) decreasing the dead time increases the maximum gain and the controllability, and (2) increasing the ratio of the longest to the second longest lag increases the controllability. In general, for the tightest loop control, the dynamic controller gain should be as high as possible without causing the loop to be unstable.

Controller tuning requires setting the three constants in the PID controller algorithm to provide control action designed for specific process requirements. The performance of the controller is evaluated in terms of the responsiveness of the controller to an error, the degree to which the controller overshoots the setpoint, and the degree of system oscillation. Note that the use of the PID algorithm for control does not guarantee optimal control of the system.

The best response to a process change or setpoint change varies depending on the application. Some processes must not allow an overshoot of the process variable beyond the setpoint if, for example, this would be unsafe. Other processes must minimize the energy expended in reaching a new setpoint. Generally, stability of response (the reverse of instability) is required, and the process must not oscillate for any combination of process conditions and setpoints. Some processes have a degree of nonlinearity, so parameters that work well at full-load conditions do not work when the process is starting up from no load.

14.4.2 Controller Response

Depending on the process to be controlled, the first consideration is to decide what type of response is best, or at least acceptable. The three possible general extremes of response that exist are overdamped, critically damped, and underdamped. When one is examining the response, there are several common performance criteria used for controller tuning that are based on characteristics of the system's closed-loop response. Some of the more common criteria include overshoot, error, rise time, and decay ratio. Of these simple performance criteria, control practitioners most often use decay ratio.

The decay ratio is the ratio of the amplitude of an oscillation to the amplitude of the preceding oscillation, as shown in Figure 14-2.

Most control engineers believe that for many control loops the optimum tuning is a 1/4 wave decay. This response provides robust disturbance rejection. In fact, the quarter decay ratio typically provides a good trade-off between minimum deviation from the setpoint after an upset and the fastest return to the setpoint.

14.4.3 Controller Performance from Plot Data

Through examination of the response plots, the relative performance of a specific controller can be compared and relative performance can be judged. Some specific performance metrics include rise time, peak overshoot ratio, settling time, and decay ratio. These and similar terms permit comparisons among the range of performance available.

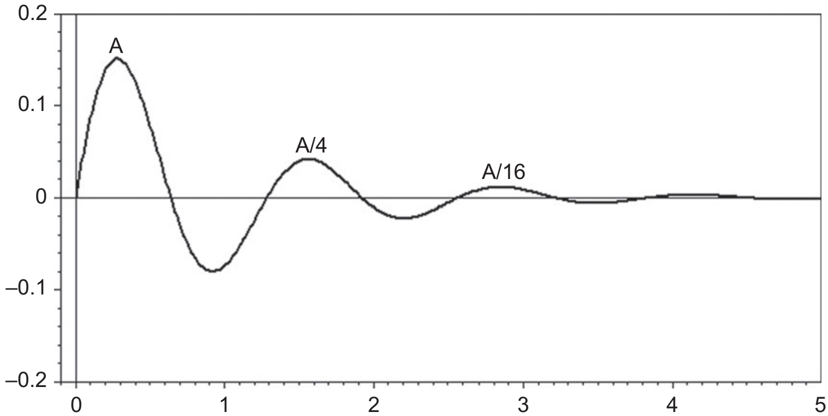

Figure 14-2 Second- or higher-order response to a setpoint change.

14.4.3.1 Peak-Related Criteria

Analysis of the magnitude of difference between setpoint and process variable without regard to time is one method of determining controller performance. These criteria for performance are the peak-related criteria. The popular peak related criteria include:

Peak overshoot ratio (POR) = Height of the first peak (B)/Size of the setpoint step (A)

Decay ratio = Height of the second peak (C)/Height of the first peak (B)

Figure 14-3 shows a setpoint step response plot with labels indicating peak features. In Figure 14-3, the process variable (PV) was initially at 20%, and a setpoint step moves it to 30%. Applying the peak-related criteria by reading off the PV axis: A = (30 − 20) = 10%, B = (34.5 − 30) = 4.5%, and C = (31 − 30) = 1%. For this response, POR = 4.5/10 = 0.45 or 45%, and decay ratio = 1/4.5 = 0.22 or 22%.

An old rule of thumb is that a 10% POR and 25% decay ratio (sometimes called a quarter decay) are popular values. Yet in today's industrial practice, many plants require a "fast response but no overshoot" control performance.

No overshoot means no peaks, and thus, B = C = 0. This increasingly common definition of "good" performance means the peak-related criteria discussed previously are not useful or sufficient as performance-comparison measures.

14.4.3.2 Time-Related Criteria

An additional set of measures focuses on time-related criteria. Figure 14-4 shows the same setpoint response plot, but with the time of certain events labeled.

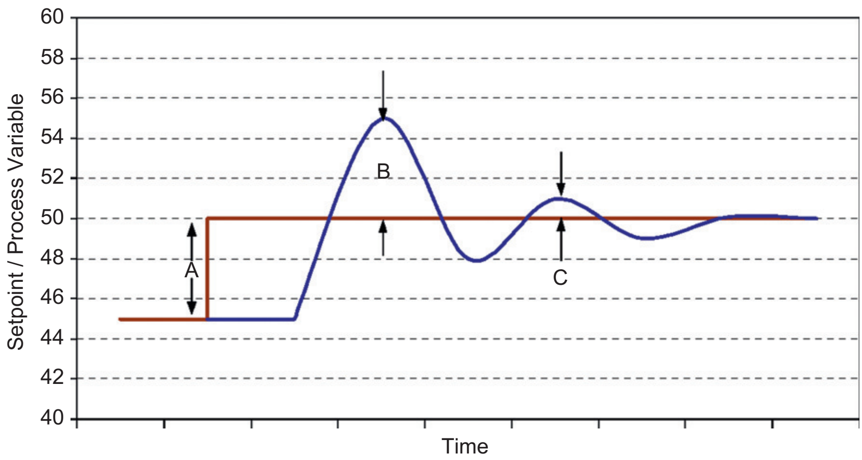

Figure 14-3 Peak-related performance criteria (A = size of the setpoint step, B = height of the first peak, C = height of the second peak).

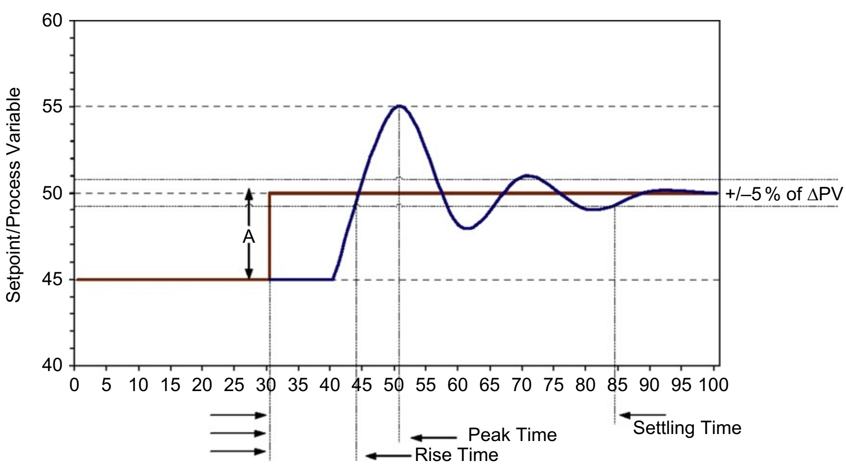

Figure 14-4 Time-related performance criteria.

The clock for time-related events begins when the setpoint is stepped, and as shown in the plot, includes

- Rise time = time until the PV first crosses the setpoint
- Peak time = time to the first peak
- Settling time = time to when a PV first enters and then remains within a band whose width is computed as a percentage of the total change in PV (or ΔPV)

The 5% band used to determine settling time in the preceding plot was chosen arbitrarily. Other percentages are equally valid depending on the situation.

The setpoint was stepped at time $t = 30$ minutes. The time-related criteria are then computed by reading off the time axis as rise time = (44 − 30) = 14 minutes, peak time = (51 − 30) = 21 minutes, and settling time = (85 − 30) = 55 minutes for a ±5% of ΔPV band.

14.4.3.3 When There Is No Overshoot

We should recognize that the peak and time criteria are not independent for the following:

- A process with a large decay ratio will likely have a long settling time.
- A process with a long rise time will likely have a long peak time.

In situations in which we seek moderate tuning with no overshoot in our response plots, there is no peak overshoot ratio, decay ratio, or peak time to compute. Even rise time, with its asymptotic approach to the new steady state, is a measure of questionable value.

In such cases, settling time, or the time to enter and remain within a band of width we choose, remains a useful measure.

The plot illustrated in Figure 14-5 shows the identical process as that in the previous plots. The only difference is that in this case, the controller is tuned for a moderate response.

We compute for this plot:

$$\text{Settling Time} = (68 - 30) = 38 \text{ min for a} \pm 5\% \text{ of } \Delta\text{PV band}$$

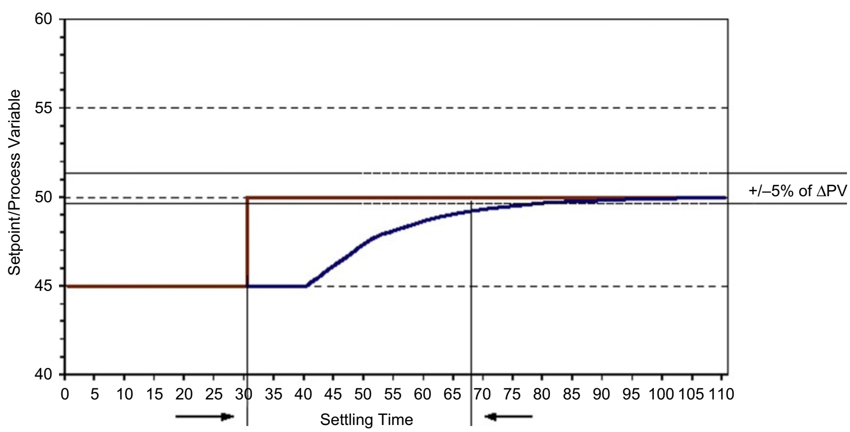

Figure 14-5 Error performance criteria.

14.4.4 Error Performance Criteria

Simple performance criteria, such as decay ratio, use only a few points in the response and are simple to use. However, more complicated error performance criteria describe the entire response of the process. Several criteria or objectives have been proposed. Among the most popular are minimum integral of square error (ISE), minimum integral of absolute error (IAE), and minimum integral of time and absolute error (ITAE).

The integrated square error (ISE) criterion uses the square of the error, thereby penalizing larger errors more than smaller errors. This gives a more conservative response, i.e., faster return to the setpoint. In mathematical terms, with e representing the error as a function of time, we can write

$$\mathrm{ISE} = \int_0^\infty e\,(t)^2 dt \tag{14–12}$$

Integrated absolute error (IAE) essentially takes the absolute value of the error:

$$\mathrm{IAE} = \int_0^\infty |e\,(t)| dt \tag{14–13}$$

The integrated time absolute error (ITAE) criterion is the integral of the absolute value of the error multiplied by time. ITAE results in errors existing over time penalized even though they may be small, which results in a more heavily damped response. The mathematical expression for this criterion follows:

$$\mathrm{ITAE} = \int_0^\infty |e\,(t)| \cdot t\; dt \tag{14–14}$$

14.4.5 Tuning Methods

There are a number of methods for tuning single-loop controllers. A few of the methods described are based on simple experiments or simple models and do not require any frequency domain analysis (although such analysis may enhance understanding of the resulting closed-loop behavior).

Many techniques are used for tuning control loops, including experience and a sense for the adequacy of control, heuristics, complex mathematics, and self-tuning systems. Control loop tuning may be accomplished in closed loop where the controller is set on automatic or in open loop where the controller is set on manual.

14.4.5.1 Process Reaction Curve Methods

In process control, the term "reaction curve" is sometimes used to characterize a step response curve. In the process reaction curve methods, a trend is generated in response to a change such as the plot shown in Figure 14-6. This trend reveals the gain, integral time, and derivative time of the process. These methods are performed in an open loop, so no control action occurs and the process response can be isolated.

To generate a process reaction curve, the process reaches steady state or as close to steady state as possible. Then, in an open loop, so there is no control action, one introduces a small disturbance and records the reaction of the process variable.

Methods of process analysis with forcing functions other than a step input are possible and include pulses, ramps, and sinusoids. However, step function analysis is the most common, as it is the easiest to implement.

In this method, the necessary data are generated by introducing a disturbance into the system and analyzing the resulting process reaction curve. When the system reaches steady state, one introduces a disturbance, X_o. The disturbance introduced to the system is a change in either the setpoint or process variable.

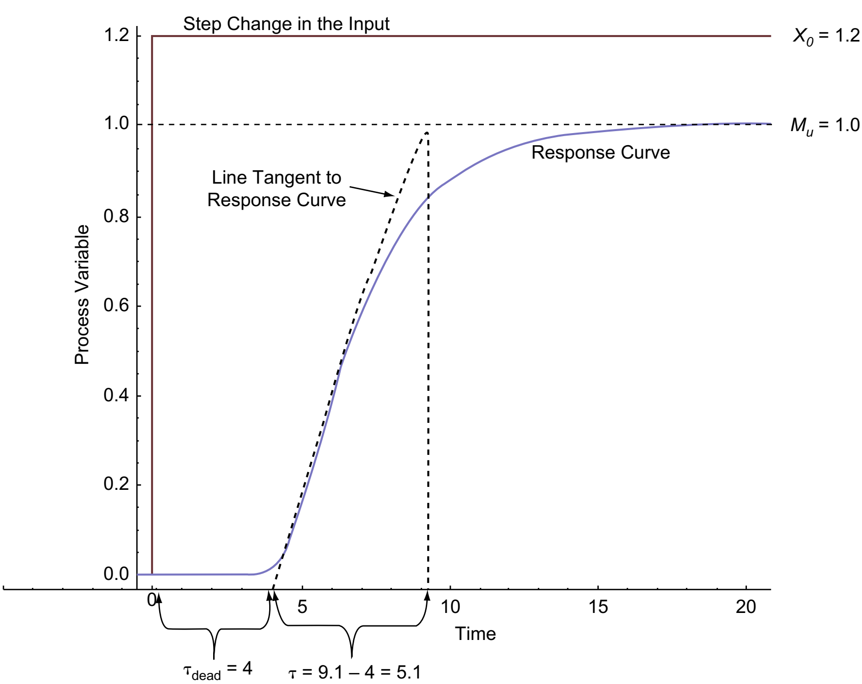

Figure 14-6 Open-loop response to a step change.

The process reaction curve method usually produces a response to a step function change for which several parameters may be measured, revealing transportation lag or dead time, τ_d; the time for the response to change, τ; and the ultimate value that the response reaches at steady state, M_u.

14.4.5.1.1 Ziegler–Nichols Open-Loop Procedure

Ziegler and Nichols (1942) developed controller tuning equations based on field measurements of the ultimate gain and ultimate period. The Ziegler–Nichols tuning method was one of the first formal methods that found wide adoption. This method is also known as the "reaction curve" method. If one wants to use the Ziegler–Nichols method, the process must be stable. With the controller in manual, one changes the output by a small amount outside the noise band. Then one can observe the effect and verify the validity of the dynamic process data collected.

This procedure consists of the two following steps:

1. Determination of the dynamic characteristics, or personality, of the control loop.
2. Estimation of the controller tuning parameters that produce a desired response for the dynamic characteristic determined in the first step; in other words, matching the personality of the controller to that of the other elements in the loop.

In this method, the dynamic characteristics of the process are represented by the ultimate gain of a proportional controller and the ultimate period of oscillation of the loop.

For a manual tuning test, the derivative time is set to zero, and the integral time is set at least 10 times larger than normal so that most of the controller response is from the proportional mode. One increases the controller gain to create equal, sustained oscillations. The controller gain at this point is the ultimate gain (K_u), and the oscillation period is the ultimate period (T_u). In practice, the gain is increased only until decaying oscillations first appear to reduce the disruption to the process. Auto tuners and adaptive controllers can eliminate the need for manual controller tuning. The relay or on/off method is used by auto tuners to automatically compute the ultimate period and gain by switching the controller output when it crosses and departs from a noise band centered on the setpoint (Blevins et al., 2003; McMillan, 2005).

For the Ziegler–Nichols ultimate oscillation method, the controller gain is a fraction of the ultimate gain, and the integral time is a fraction of the ultimate period, as follows for a PI controller.

Defining the open-loop gain as K_o produces

Table 14-1 Tuning Parameters for the Open-loop Ziegler–Nichols Method

	K_c	T_i	T_d
P	K_o		
PI	$0.9K_o$	$3.3\tau_d$	
PID	$1.2K_o$	$2\tau_d$	$0.5\tau_d$

$$K_o = (X_o/M_u) * \tau/\tau_d \tag{14-15}$$

The tuning parameters derived by Ziegler and Nichols (1942) are given in Table 14-1.

The advantages of the Ziegler–Nichols open-loop tuning method are

1. It is quick and easy.
2. The process reaction curve method is the least disruptive to implement.

The disadvantages of the Ziegler–Nichols open-loop tuning method are

1. It depends on a purely proportional measurement to estimate I and D parameters.
2. Approximations for the K_c, T_i, and T_d values might not be entirely accurate for different systems.
3. It is not applicable for I, D, and PD controllers.

14.4.5.1.2 Cohen–Coon Tuning Method

Cohen and Coon (1952) modified the Ziegler–Nichols (1942) open-loop tuning rules in 1952. The modifications are insignificant when the dead time is small relative to the time constant but can be important for large dead time. The Cohen and Coon (1952) tuning parameters are as follows:

For P-only control,

$$K_c = (1 + \mu/3)/(\tau_d * RR) \tag{14-16}$$

For PI control,

$$K_c = (1 + \mu/11)/(\tau_d * RR) \tag{14-17}$$

$$T_i = 3.33 * \tau_d * ((1 + \mu/11)/(1 + 11\mu/5)) \tag{14-18}$$

For PID control,

$$K_c = 1.35 * (1 + \mu/3)/(\tau_d * RR) \tag{14-19}$$

$$T_i = 2.5 * \tau_d * ((1 + \mu/5)/(1 + 3\mu/5)) \tag{14-20}$$

$$T_d = .37 * \tau_d/(1 + \mu/5) \tag{14-21}$$

In the preceding equations,

$$\mu = \tau_d/\tau \tag{14-22}$$

$$RR = X_o/M_u * \tau \tag{14-23}$$

As with the Ziegler–Nichols open-loop method recommendations, one can adjust the Cohen–Coon values in closed-loop mode to achieve the quarter decay ratio.

The advantages of the Cohen–Coon method over the Ziegler–Nichols open-loop method are

1. It is useful for systems with a large dead time.
2. It provides faster closed-loop response time.

The disadvantages and limitations of the Cohen–Coon method are

1. It can produce an unstable closed-loop system, since this is an open-loop method.
2. It is useful only for first-order models, including large process delays.
3. Approximations for the K_c, τ_i, and τ_d values might not be entirely accurate for different systems.

14.4.5.1.3 Internal Model Control Tuning Rules

Internal model control (IMC) tuning rules have proven to be robust and yield acceptable performance when used in the control of common processes. In general, analytical IMC tuning rules are derived for PI and PID compensators by matching an approximate process model to a low-dimensional reference model. The IMC controller structure depends on two factors: the complexity of the model and the performance requirements stated by the designer.

The IMC philosophy relies on the international model principle, which states that a feedback regulator under external disturbances may regain regulation and stability provided a suitably reduplicated model of the disturbance signal is adapted in the feedback path. In other words, if one develops the control scheme based on an exact model of the process, perfect control is theoretically possible. Internal model control uses open-loop step-response Laplace transfer functions with process gain and time constant to predict a measurement change due to a change in setpoint. Use of the IMC philosophy can also generate settings for conventional PI or PID controllers. The algorithm includes a model to bias the setpoint to remove the steady-state error, which becomes a feedback adjustment method. One can filter the biased setpoint to obtain a reference trajectory. A single tuning factor filters the time constant. High tuning factors lead to gentle control action, whereas low tuning factors lead to aggressive control action.

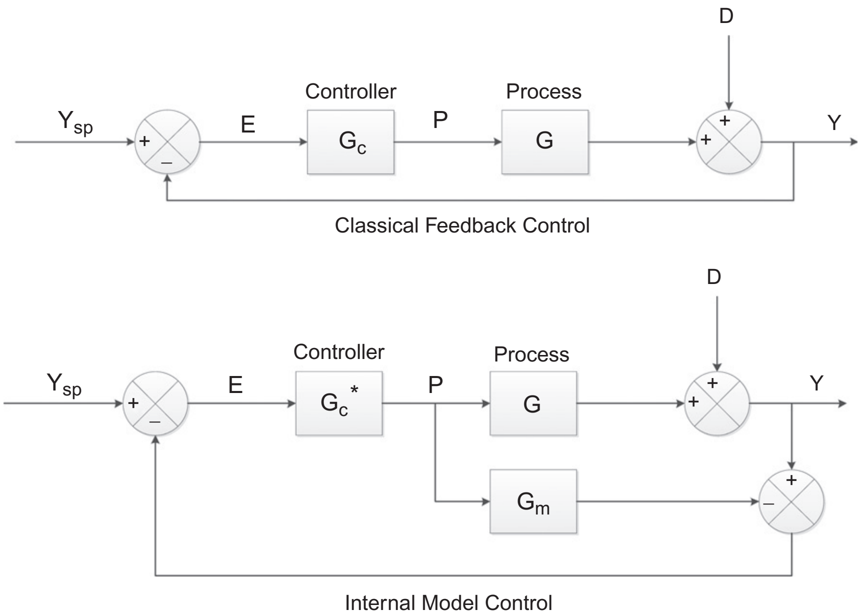

Figure 14-7 Comparison of classical feedback control to internal model control.

Figure 14-7 shows a comparison of the classical feedback control algorithm to the IMC algorithm where Y_{sp} is setpoint, Y is process output, E is setpoint error (the difference between process variable and setpoint), G_c is controller transfer function, P is manipulated input, G is process transfer function, D is system disturbance, G_m is plant model, and G_c^* is internal model controller.

Since the general IMC method is unnecessarily complicated for processes that are well approximated by first-order dead time or integrator dead-time models, simplified IMC rules were developed by Fruehauf et al. for PID controller tuning (Chien and Fruehauf, 1990; Fruehauf et al., 1994).

The IMC-PID tuning rules often apply in industry. However, the widely published IMC tuning rules, while providing adequate suppression of output disturbances, do a poor job suppressing load disturbances when the process dynamics are significantly slower than the desired closed-loop dynamics. Morari and Zafiriou (1989) proposed to address this problem by including an additional integrator in the output disturbance while performing the IMC design procedure. This method was found to provide adequate load disturbance suppression for many processes and has been applied to model predictive control (MPC). However, the resulting controllers do not have PID structure.

IMC models are inherently stationary and linear. Like PID controllers, IMC must be tuned for changes in process gain or time constant. The model should be reparameterized when the process dynamics change substantially.

14.4.5.2 Constant Cycling Methods

14.4.5.2.1 Ziegler–Nichols Closed-Loop Method

The Ziegler–Nichols closed-loop tuning method is probably the most well-known tuning method. It requires a simple closed-loop experiment, using proportional control only. One can increase the proportional gain until a sustained oscillation of the output occurs (neither grows nor decays significantly with time). Then one records the proportional gain that gives the sustained oscillation and the oscillation period (time). The proposed tuning parameters are given in Table 14-2. The ultimate gain, K_u, is the gain at which the oscillations continue with a constant amplitude. The period of these oscillations is the ultimate period, P_u, as shown in Figure 14-8.

In most cases, increasing the proportional gain will provide a sufficient disturbance to initiate the oscillation. Measurement noise may also initiate oscillation. Only if the output is very close to the setpoint will it be necessary

Table 14-2 Tuning Parameters for the Ziegler–Nichols Closed-Loop Method

	Gain	Reset	Derivative
P	$0.5\ K_u$	—	—
PI	$0.45\ K_u$	$1.2/P_u$	—
PID	$0.6\ K_u$	$2/P_u$	$P_u/8$

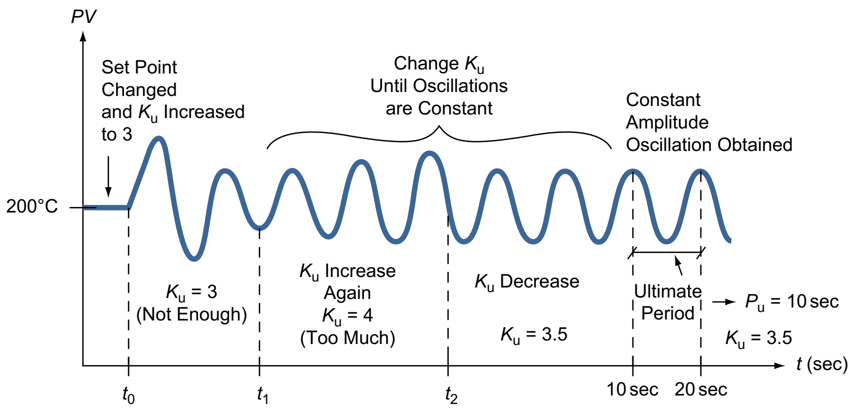

Figure 14-8 Example of determination of the ultimate period.

to introduce a setpoint change after increasing the gain, in order to initiate an oscillation. Note that for controllers giving positive output signals, i.e., controllers giving output signals scaled in the range 0–1 or 0%–100%, a constant bias must be included in the controller in addition to the proportional term, thus allowing a negative proportional term to have an effect. Otherwise, the negative part of the oscillation in the plant input will be cut off and affect the oscillation of the output. In this case, both the input and output of the plant may still oscillate but would show a more complex behavior than the single-frequency sinusoids that the experiment should produce.

The steps required for the Ziegler-Nichols closed-loop methods are

1. Place the controller into automatic with low gain and no reset or derivative.
2. Gradually increase the gain by making small changes in the setpoint until oscillations start.
3. Adjust the gain to force the oscillations to continue with constant amplitude.
4. Note the gain and period.
5. The ultimate gain, G_u, is the gain at which the oscillations continue with a constant amplitude. The period of these oscillations is the ultimate period, P_u.

Advantages of the Ziegler–Nichols closed-loop tuning method are as follows:

1. It is an easy experiment that requires a change only in the gain of a controller.
2. The tuning considers the dynamics of the whole system, which gives a more accurate picture of how the system behaves.

Disadvantages of this method are as follows:

1. The experiment can be time consuming.
2. Unstable regions may be encountered while testing the controller, which could cause the system to become out of control.

14.4.5.3 Autotune Variation Technique

The objective of autotuning methods is to obtain a PID controller capable of satisfying typical requirements such as rapid following, zero steady-state error, and overshoot suppression by means of a practical and robust method. The autotune method has several advantages over open-loop pulse testing methods such that (1) no prior knowledge of the system time constant is needed; and (2) the method is closed-loop tested, so the process will not drift away from the setpoint.

Relay feedback autotuning and frequency domain autotuning by magnitude and phase calculation are some of the most common methods to adjust the parameters of a PID controller. Relay feedback autotuning uses the ultimate frequency. This frequency is a reference, and the design operating frequency will be a fraction of the reference.

Åström and Hägglund (1995) proposed a method for determining the ultimate frequency and ultimate gain commonly called autotune variation (ATV). This method consists of an approximate method called the harmonic balance method based on relay feedback autotuning. Figure 14-9 shows a process transfer function with a feedback ideal relay. In parallel with a feedback relay, there is a PID controller, the achieved virtual controller.

An approximate condition for oscillation can be determined by assuming that there is a limit cycle with period T_u and frequency $w_u = 2\pi/T_u$ such that the relay output is a periodic symmetric square wave. If the relay amplitude is d, a simple Fourier series expansion of the relay output shows that the first harmonic component has the amplitude $4d/\pi$.

The ATV method determines the ultimate gain and period in a manner similar to that of the ultimate method, but ATV tests can be implemented without unduly upsetting the process. Controller settings are calculated, and the controller is then tuned online to meet the selected tuning criterion. The ultimate controller gain, K_u, is calculated by

$$K_u = 4h/\pi a \tag{14-24}$$

where h is the relay height; and a is the amplitude of the process variable. The ultimate period can be directly calculated from the process variable, also.

With K_u and P_u calculated, settings for a PI controller can be calculated using the Ziegler–Nichols method.

An adaptive controller has been achieved through use of the ATV autotuning technique or relay feedback autotuning, which supplies the frequency of the limit cycle and its associated gain, known as ultimate frequency and ultimate gain (Åström and Wittenmark, 1995).

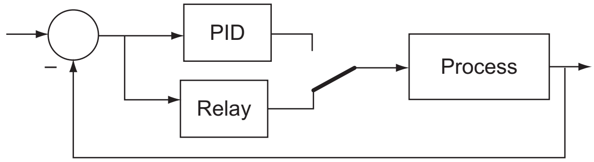

Figure 14-9 Block diagram of a relay autotuner.

14.4.5.4 Lambda Tuning

Another class of tuning method that became popular with the increased use of computing power is lambda tuning. Lambda tuning refers to all tuning methods where the control loop speed of response is a selectable tuning parameter. The closed-loop time constant is "lambda." Lambda tuning, which originated with Dahlin (1968), is based on the same internal model control (IMC) theory as the model predictive control, is model-based, and uses a model inverse and pole-zero cancellation to achieve the desired closed-loop performance.

The lambda tuning equations were developed for simplicity and practicality. For self-regulating processes, the equations are as follows:

$$\lambda = \lambda_f * \tau_1 \tag{14-25}$$

$$T_i = \tau_1 \tag{14-26}$$

$$K_c = T_i / (K_o * (\lambda + \tau_d)) \tag{14-27}$$

$$T_d = \tau_2 \tag{14-28}$$

where K_c is controller gain; K_o is open-loop gain; λ is lambda (closed-loop time constant); λ_f is lambda factor; τ_d is total loop dead time; τ_1 is largest open-loop time constant; τ_2 is second largest open-loop time constant; T_i is integral time setting; and T_d is derivative time setting.

The lambda factor is the ratio of closed-loop time constant to the open-loop time where the open-loop time constant is the largest time constant (τ_1). For maximum load-rejection capability, a lambda equal to the total loop dead time ($\lambda = \tau_d$) can be used, and the loop will still be stable.

The goal of lambda tuning is to match the setpoint response to the first-order time constant or lambda. The response is initially delayed by the process dead time. Since lambda tuning is a model-based method, the tuning parameters are derived from a model of the process. Given a model, the tuning method for an ideal-type PID controller is simply a proper conversion of the units. Parallel- and series-type controllers require different tuning. For PI controllers, series tuning and ideal tuning are the same.

The design concept behind lambda is to cancel the process with the controller and then use a first-order filter to obtain the desired response.

Once the field devices have been checked and corrected as required, an open-loop response test with the controller in manual operation is performed to understand the dynamics of the process. Testing is performed over a range of normal operating conditions. The collected data should be fitted to a simple dynamic model (common models include first order plus dead time and integrator plus dead time).

A unification of lambda, internal model control, and Ziegler-Nichols reaction curve and ultimate oscillation tuning methods has been achieved. The controller tuning equations from diverse methods reduce to a common form, where the maximum controller gain is proportional to the time constant to dead-time ratio (τ_1/τ_d) and is inversely proportional to the open-loop gain (K_o). This common form is easy to remember and provides insight as to the relative effects of process dynamics on tuning and hence on loop performance (Boudreau and McMillan, 2006).

14.4.6 PID Tuning Software

Most modern industrial facilities no longer tune loops using the manual calculation methods discussed previously. Instead, PID tuning and loop optimization software is used to ensure consistent results. These software packages will gather the data, develop process models, and suggest optimal tuning. Some software packages can even develop tuning by gathering data from reference changes.

Mathematical PID loop tuning induces an impulse in the system and then uses the controlled system's frequency response to design the PID loop values. In loops with response times of several minutes, mathematical loop tuning is recommended because trial and error can take too much time to find a stable set of loop values. Optimal values are harder to find. Some digital loop controllers offer a self-tuning feature in which very small setpoint changes are sent to the process, allowing the controller itself to calculate optimal tuning values.

When dynamic process data are generated, it is important that the change in the controller output signal causes a response in the measured process variable that clearly dominates the measurement noise. One way to quantify the amount of noise in the measured process variable is with a noise band. A noise band is based on the standard deviation of the random error in the measurement signal when the controller output is constant and the process is at steady state. Plus or minus three standard deviations of the measurement noise around the steady state of the measured process variable (99.7% of the measurements are contained within the noise band) is conservative when used for controller tuning.

When dynamic process data are generated, the change in controller output should cause the measured process variable to move at least 10 times the size of the noise band. In other words, the signal-to-noise ratio should be greater than 10. For instance, if the noise band is 1°F, then the controller output should be moved enough during a test to cause the measured exit temperature to move at least 10°F.

With today's data historian capabilities, much high-frequency information can be captured for calculating tuning factors. Trending capabilities are also quite helpful for these analyses.

14.4.7 Power Spectrum Analysis

Power spectrum analysis is a technique commonly used by PID tuning software and applies a fast Fourier transform (FFT) to the variation of a particular signal to compute its frequency spectrum. The result is presented as a plot of signal power against frequency and is referred to as its power spectrum. The power spectrum of a signal indicates the relative magnitudes of the frequency components that combine to make up the signal.

The data used to determine the power spectrum must reflect sufficient excitation in the signal. Ideally, the signal should be subject to some form of random excitation in order to generate appropriate data (for example, by the application of a pseudo-random binary sequence to the plant at some appropriate point). The power spectrum is useful for determining the degree of noise that is associated with the signal and also deciding on appropriate sampling rates (following the guidelines of Shannon's sampling theorem).

The power spectrum analysis graphically indicates the frequency content of the PV signal. It does this by performing a fast Fourier transform on the PV signal and plotting the magnitude of each frequency analyzed. Figure 14-10 shows an example of a Power Spectrum plot with frequency. Plots of power over period are more applicable to the process control environment where we would rather think in terms of seconds and minutes instead of hertz, as shown in Figure 14-11.

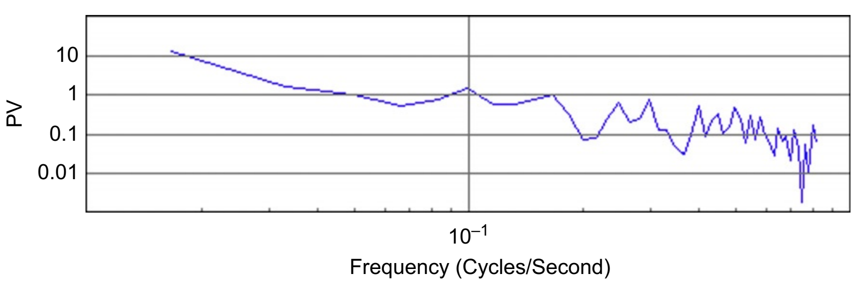

Figure 14-10 Power spectrum (PV) versus frequency.

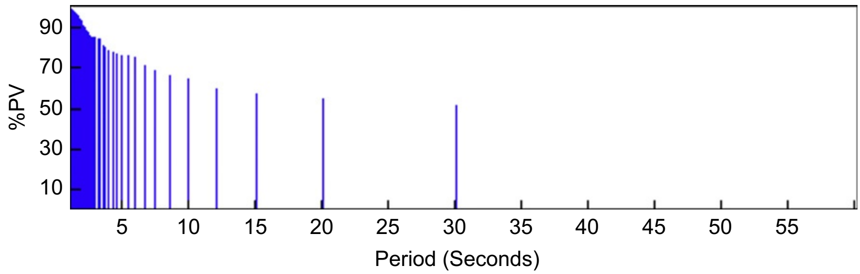

Figure 14-11 Power spectrum (PV) versus period.

The vertical axis of a power spectrum plot is the power detected at each frequency. The span is from 0 Hz to a threshold frequency (f_{thres}) which corresponds to half the logging frequency that is associated with the signal:

$$f_{\text{thres}} = 1/2T \tag{14-29}$$

where T is the log interval in seconds.

According to Shannon's theorem, any information in the signal that relates to a frequency greater than (f_{thres}) cannot be extracted at the particular log interval of sampling. This criterion is, in fact, idealistic, and in practical situations, the threshold frequency is halved in order to guarantee the representative recovery of information:

$$f < f_{\text{limit}} = 1/4T \tag{14-30}$$

A cycle might be completely hidden by the random noise content of the PV signal.

14.4.8 Choosing a Tuning Method

The most popular tuning alternatives employed in today's operations are shown in Table 14-3.

The choice of method will depend largely on whether or not the loop can be tuned offline and the response time of the system. If tuning the system offline, the best method often involves subjecting the system to a step change in input, measuring the output as a function of time, and using this response to determine the control parameters.

If the system must remain online, one tuning method is to first set the I and D values to zero. The P should be increased until the output of the loop oscillates; then the P should be left set to approximately half of that value for a "quarter amplitude decay" type response. Then I should be increased until any error is corrected in sufficient time for the process. However, too much I

Table 14-3 Most Popular Tuning Alternatives

Method	Advantages	Disadvantages
Ziegler–Nichols	Proven online method.	Process upset, some trial and error, very aggressive tuning
Tune by feel	No math required. Online method.	Erratic, not repeatable
Software tools	Consistent tuning. Online or offline method. May include valve and sensor analysis. Allows simulation before downloading.	Some cost and training involved
Cohen–Coon	Good process models.	Some math. Offline method. Only good for first-order processes.

will cause instability. Finally, D should be increased, if required, until the loop is acceptably quick to reach its reference after a load disturbance. However, too much D will cause excessive response and overshoot. A fast PID loop tuning usually overshoots slightly to reach the setpoint more quickly; however, some systems cannot accept overshoot, in which case an "overdamped" tune is required, which will require a P setting significantly less than half that of the P setting causing oscillation.

14.4.8.1 Tuning Flow Loops

Flow loops are too fast to use the standard methods of analysis and tuning. Proper tuning of flow loops depends on whether the control is analog or digital. Some flow loops using analog controllers are tuned with high gain, but this will not work with digital control. With an analog controller, the flow loop has a predominant lag (L) of a few seconds and no subordinate lag. The scan rate of a digital controller can be considered dead time. Although this dead time is small, it is large enough when compared to L to force a low gain. Table 14-4 shows the typical PID settings depending on the type of loop.

14.4.9 Advanced Control Methods

The following sections discuss some advanced control methods, including feedforward control, cascade control, override and selectors, interaction and decoupling, nonlinear control, adaptive control, model predictive control, model-based control, optimizing control, and self-tuning controllers.

Table 14-4 Typical PID Settings for Different Loops

Loop Type	P_b %	Integral min/rep	Integral rep/min	Derivative min	Valve Type
Flow	50–500	0.005–0.05	20–200	none	Linear or modified percentage
Liquid pressure	50–500	0.005–0.05	20–200	none	Linear or modified percentage
Gas pressure	1–50	0.1–50	0.02–10	0.02–0.1	Linear
Liquid level	1–50	1–100	0.1–1	0.01–0.05	Linear or modified percentage
Temperature	2–100	0.2–50	0.02–5	0.1–20	Equal percentage
Chromatograph	100–2,000	10–120	0.008–0.1	0.1–20	Linear

14.4.9.1 Feedforward Control

Feedforward control differs from feedback control in that the load or primary disturbance is measured and the manipulated variable is adjusted so that deviations in the controlled variable from the setpoint are minimized. The controller can then reject disturbances before they affect the controlled variable. For accurate feedforward control, steady-state or dynamic analysis should be the basis for models that relate the effect of the manipulated and disturbance variable on the controlled variable. Since the model is an approximation and not all disturbances are measured, feedforward control should always be used in conjunction with feedback control. This combination will allow compensation for measured and unmeasured disturbances as well as model mismatch.

14.4.9.2 Cascade Control

Processes that respond to disturbances with long time delays or lags are difficult to control with a single feedback controller. One relatively simple way to improve the dynamic response is to use a secondary measurement and controller. The secondary measurement and controller should recognize the effect of a disturbance before the primary controlled variable is affected. Cascade control, where the primary controller (also called master controller) output becomes the setpoint for the secondary controller (also called slave controller), is a readily configured strategy in most computer control systems.

Cascade control is commonly used on distillation towers for composition control. The primary measurement of a key component typically has a time delay, and feedback control is difficult. Temperature of a tray that is sensitive to compositional changes can be used as a secondary measurement. The tray temperature usually responds faster than the analyzer.

In the case of distillation composition control, the response can be further improved by configuring the heat medium flow to a reboiler or reflux rate as a secondary control for the appropriate tray temperature control. As a rule of thumb, the primary loop should be five times faster than the secondary loop.

14.4.9.3 Override and Selectors

One approach to addressing more controlled variables than manipulated variables is to use a selector. The most common types of selective controls are high and low selectors. This allows control action to be based on a selection criterion for multiple measurements.

Override is another type of selective control. In this case, control is dictated by a process variable that reaches a predetermined high or low limit.

14.4.9.4 Interaction and Decoupling

An interactive system exists when a manipulated variable for one controlled variable affects the controlled variable of another loop. Additional feedback loops occur between unpaired manipulated and controlled variables in these cases and destabilize the system.

Relative gain analysis reveals interactions and the degree of interaction. Pressure and level are functions of flow differences, and temperature and composition are functions of flow ratios. These differences and ratios can often be used to decouple and minimize interactions. This technique is rational decoupling. Decoupling is similar to feedforward control except the load variable replaces a manipulated variable.

When ratios and differences are not effective in eliminating interactions, then the controllers can be detuned to restore stability. This detuning will lead to loss in control performance.

Another technique includes linear decouplers calculated from process gains. The issues with this approach are constraints and initialization. When a manipulated variable is constrained, then two controllers compete for an unconstrained variable. One controller will wind down, whereas the other controller winds up. Decoupling using measured values of the manipulated variables overcomes the initialization and constraint issues. Even with these enhancements, decouplers are difficult to match and remain matched with the process, leading again to stability issues.

Partial decoupling can be used to stabilize interactive loops. Best practice is to protect the least likely controlled variable to change setpoint, the slowest, or the most important controlled variable.

Other techniques for decoupling are using an adaptive multivariable controller on feedforward loops and dynamic matrix methods. These methods are described in subsequent sections.

Blending and distillation systems often present interactive systems that require decoupling for effective control.

14.4.9.5 Nonlinear Control

Conventional process control systems utilize linear dynamic models. For highly nonlinear systems, control techniques directly based on nonlinear models provide significantly improved performance.

Most real processes display some nonlinear behavior. The process gain and dead time can change with load, time with equipment degradation, and dead time with transportation lag. In many cases, linear controllers provide adequate control performance. As the degree of nonlinearity increases, then improved control performance may be necessary and desired.

There are two classes of nonlinear control: discontinuous and continuous. The discontinuous methods include on-off and three state devices. These discontinuous methods are adequate only when accurate regulation is not essential. Continuous nonlinear control methods include fuzzy logic, output filtering, characterization, and dead-band or gap action.

14.4.9.6 Adaptive Control

Programmed and self-adaptive controls comprise the two classes of adaptive controls. Programmed adaptive applies when a measured variable yields a predictable response from the control loop. Programmed adjustments are continuous or discontinuous, including gap action and switching controller gains. Adaptive methods include self-tuning, model reference, and pattern recognition.

A classic example of variable, predictable gain on a control loop occurs in plug flow situations commonly found in heat exchangers. At low flow, the dead time, steady-state gain, and dominant time constant are proportionately higher than high flow. For instance, at 50% flow, these parameters are all twice as high as at 100% flow.

The equation for a flow-adapted PID controller is

$$m = K_c * (w * e + w^2/T_i * \int e\,dt + T_d * de/dt) + b \tag{14-31}$$

where K_c, T_i, and T_d are the proportional, integral, and derivative settings at full-scale flow; and w is the fraction of full-scale flow.

Gap action, mentioned in the nonlinear control section, can also be adaptive. A dead band is placed around the error used in the calculation of the proportional, integral, and derivative action. No action is taken within the deadband. Level control of surge tanks, pH control, and other systems where stability may be sacrificed with too much control action are typical examples.

Gain scheduling is another form of nonlinear control or programmed adaptation. Variable breakpoint control is another description of this type of control. This method is often similar to gap action. Table 14-5 shows a comparison between gap action and gain scheduling.

Another common method of controller gain adjustments is an error-squared formulation. This produces a nonlinear controller with similar characteristics to the typical implementation of gain scheduling. Little action is taken at small deviations from setpoint compared to the action taken at larger deviations. This method should be used with caution, and limits placed as instability may occur at large deviations from setpoint.

When control loop response is impacted by unknown or immeasurable disturbances, then programmed adaptation is not possible. The industry practices several forms of self-adaptive methods, including self-tuning regulator, model reference adaptive, and pattern recognition.

All self-tuning systems have common elements of an identifier, controller synthesizer, and controller implementation. Types of self-tuning regulators include dead beat, stability theory, fuzzy logic, pole-placement, generalized predictive, and minimum variance algorithms.

The system identifier estimates the parameters of the process. It determines the response of the controlled variable due to a change in the manipulated

Table 14-5 Typical PID Settings for Different Type Loops

Situation	Gap Action	Gain Scheduling
Close to setpoint	No action within a predetermined dead band.	Typically very little action with low controller gain.
Further from setpoint	One set of predetermined PID tuning parameters used outside the dead band.	One or more additional sets of predetermined PID tuning parameters used with higher gain the further the departure from setpoint.

variable. These changes may be deliberately introduced or may be transients that normally occur.

A synthesizer calculates the controller parameters based on a control objective. Recursive estimators determine the optimal proportional, integral, and derivative gains. A function block updates the controller parameters calculated at a predetermined frequency or subject to heuristics.

Model reference adaptive methods are classified as optimal or response specification type algorithms. Minimum variance types employ a least squares method comparing the process variable to its setpoint. The minimum variance adaptive controllers often become unstable when the time delay varies or is unknown. Generalized predictive control uses predictive horizon techniques to overcome the time delay issues experienced with minimum variance methods.

Pole placement and pole assignment routines are common response specification type adaptive controllers. These algorithms use a desired closed-loop frequency to determine preferred controller parameters. This technique requires an excitation of the process that is normally performed in open loop.

Pattern recognition does not use a model to self-adapt a controller. When a change in setpoint occurs, the dead time, sensitivity, and steady-state gain are determined from a positive and negative load response. After identification of the measurement noise, the oscillation, damping ratio, and overshoot are evaluated for desired parameters based on integrated absolute error. Neural network algorithms can also be used for pattern recognition. The radial basis function form of neural networks is ideal due to its ability to analyze time series data.

Optimization can be attempted with single loop controllers using several techniques. However, multivariable algorithms are much more effective for optimization solutions.

Evolutionary optimization and gradient search methods are two approaches with single loop controllers. With evolutionary optimization, online experimentation is conducted to determine the optimum. An independent variable is changed and the response of the optimization objective is measured to determine whether or not there is an improvement. If the response has a negative impact on the objective, then the independent variable will be moved in the opposite direction. If the response has a positive impact on the objective, then the independent variable will continue movement in the same direction. If there is no response, then it is assumed that the optimum has been found.

An example is minimization of a cold separator for a cryogenic natural gas liquids recovery operation. An inlet gas split ratio can be increased or decreased, and the response of the cold separator temperature monitored. If an increase in the split ratio causes an increase in the cold separator temperature, then the split ratio will be decreased. This decrease will continue until there is no change in the cold separator temperature or until the temperature experiences an increase.

Gradient search methods including steepest ascent algorithms are sometimes implemented for optimizing control. Complications arise when a constraint is encountered. Typically, penalty terms for active constraints are imposed on the objective function to force control back within a feasible region.

14.4.9.7 Model Predictive Control

MPC has become the most popular advanced control technique for difficult control problems. The main idea of MPC is to choose the control action by repeatedly solving an optimal control problem online. This technique aims at minimizing a performance criterion over a future horizon, possibly subject to constraints on the manipulated inputs and outputs, where the future behavior is computed according to a model of the plant. Model predictive controllers rely on dynamic models of the process, most often linear empirical models obtained by system identification. The models are used to predict the behavior of dependent variables (i.e., outputs) of a dynamic system with respect to changes in the process-independent variables (i.e., inputs).

If a reasonably accurate dynamic model of the process is available or can be derived, then the model updated with current measurements can be used to predict the future process behavior. The values of the manipulated inputs are calculated so that they minimize the difference between the predicted response of the controlled outputs and the desired response.

Despite the fact that most real processes are approximately linear within only a limited operating window, linear MPC approaches are used in the majority of applications with the feedback mechanism of the MPC compensating for prediction errors due to structural mismatch between the model and the plant. In model predictive controllers that consist only of linear models, the superposition principle of linear algebra enables the effect of changes in multiple independent variables to be added together to predict the response of the dependent variables. This simplifies the control problem to a series of direct matrix algebra calculations that are fast and robust.

Model predictive control offers several important advantages:

1. The manipulated inputs are adjusted based on how they affect the outputs.
2. Inequality constraints on process variables are easily incorporated into the control calculations.
3. The control calculations can be coordinated with the calculation of optimum setpoints.
4. Accurate model predictions provide an indication of equipment problems.

Model predictive control is widely considered the control method of choice for challenging multivariable control problems and normally provides very quick payout.

Identification of the dynamic process model is key to the successful implementation of MPC techniques. In most MPC applications, special plant tests are required to develop empirical dynamic models from input-output data. Although these plant tests can be disruptive, the benefits from implementing MPC generally justify this short-term disruption to normal plant production.

14.5 REFERENCES

Åström, K.J., Hägglund, T., 1995. PID Controllers: Theory, Design, and Tuning, second ed. ISA Press, Instrumentation, Automations, and Systems (ISA) Society, Research Triangle Park, NC.

Åström, K.J., Wittenmark, B., 1995. Adaptive Control, second ed. Addison-Wesley, Dover, NY.

Blevins, T.L., McMillan, G.K., Wojsznis, W.K., Brown, M.W., 2003. Advanced Control Unleashed—Plant Performance Management for Optimum Benefits. ISA Press, Instrumentation, Automations, and Systems (ISA) Society, Research Triangle Park, NC.

Boudreau, M.A., McMillan, G.K., 2006. New Directions in Bioprocess Modeling and Control. ISA Press, Instrumentation, Automations, and Systems (ISA) Society, Research Triangle Park, NC.

Chien, I.-L., Fruehauf, P.S., 1990. Consider IMC Tuning to Improve Controller Performance. Chem. Eng. Prog. 86 (10), 33–41.

Cohen, G.H., Coon, G.A., 1952. Theoretical Consideration of Retarded Control. Trans. ASME 75, 827–834.

Dahlin, E.B., 1968. Designing and Tuning Digital Controllers. Instrum. Control Syst. 41 (6), 77.

Fruehauf, P.S., Chien, I.L., Lauritsen, M.D., 1994. Simplified IMC-PID Tuning Rules. ISA Trans. 33, 43–59.

Liptak, B.G., 1995. Instrument Engineers Handbook, Process Control, third ed. Chilton Book Company, Radnor, PA.

McMillan, G.K., 2005. Good Tuning—A Pocket Guide, second ed. ISA Press, Instrumentation, Automations, and Systems (ISA) Society, Research Triangle Park, NC.

Morari, M., Zafiriou, E., 1989. Robust Process Control. Prentice-Hall, Englewood Cliffs, NJ.

Ziegler, J.G., Nichols, N.B., 1942. Optimum Settings for Automatic Controllers. Trans. ASME 64, 759–768.