

7.0 context and direction

Chemical processing plants are characterized by large time constants and time delays. For control engineering, we can often approximate these high-order systems by the FODT (first-order-dead-time) model. Dead time in a process increases the difficulty of controlling it.

DYNAMIC SYSTEM BEHAVIOR**7.1 big and slow - high-order overdamped systems**

We began our study of process control by considering a mixed tank. Applying a material or energy balance to a well-mixed tank produces a first-order lag system. We subsequently combined two balances to produce a second-order system. In one case, two material balances described storage of material in two tanks. In another case, a single tank stored both material and energy. Energy and material balances show that the tank causes a dynamic lag between input and output, because it takes time to adjust the amount of mass or energy distributed throughout the tank. We might thus expect that more storage elements would lead to higher-order behavior, and require higher-order equations to describe them.

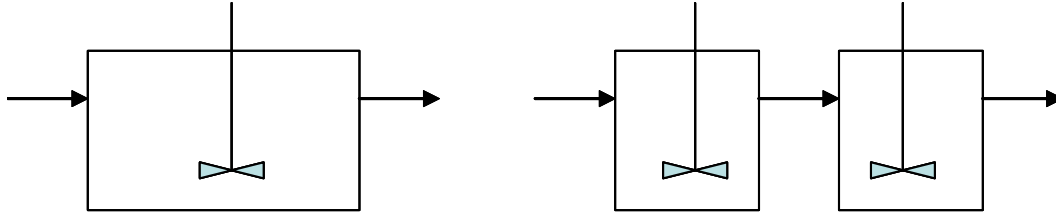
The classic illustration of a high-order system is a set of n tanks in series: each tank feeds the next, and a change in the inlet stream composition C_{A0} must propagate through multiple tanks to be felt at the output C_{An} . The individual tank models are

$$\frac{d}{dt} V_i C'_{Ai} = F C'_{Ai-1} - F C'_{Ai} \quad (7.1-1)$$

They are combined by eliminating the interior stream variables to produce a single transfer function between input and output.

$$\frac{C'_{An}(s)}{C'_{A0}(s)} = \frac{1}{(\tau_1 s + 1)(\tau_2 s + 1) \cdots (\tau_n s + 1)} \quad (7.1-2)$$

Let us illustrate high-order behavior and (7.1-2) by first imagining a single well-mixed overflow tank of time constant τ . If we introduce a step increase in the inlet concentration, we will (by the well-mixed assumption) immediately detect a rise in the outlet stream – the familiar first-order lag response. If we have instead two tanks in series, each half the volume of the original, we will detect a second-order, sigmoid response at the outlet.



Each tank has a smaller individual time constant, and their sum is the time constant τ of the original tank. If we continue to increase the number of tanks in the series, always maintaining the total volume, we observe a slower initial response with a faster rise around the time constant. This behavior is shown in Figure 7.1-1.

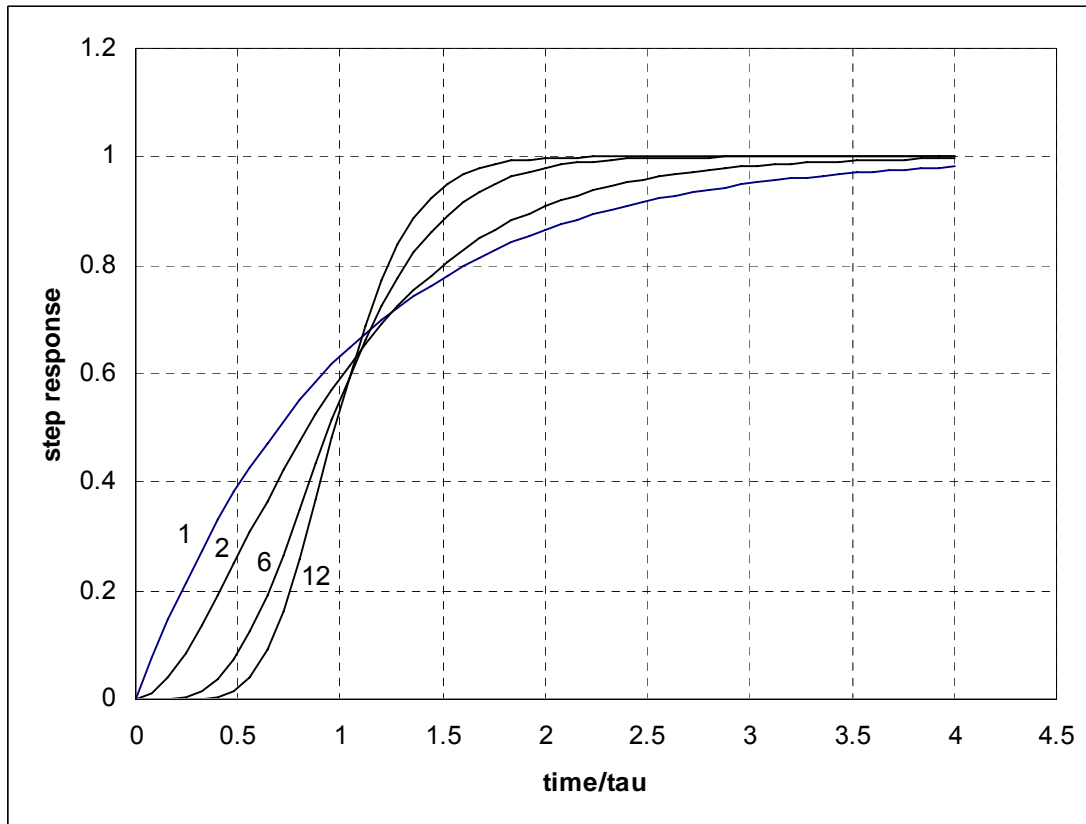
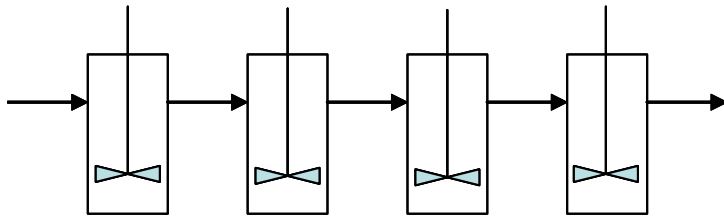


Figure 7.1-1. Step response for tanks in series; equal time constants

The step response shows that high-order systems have a longer start-up period before rising toward the final value.

7.2 the FODT approximation to high-order step response

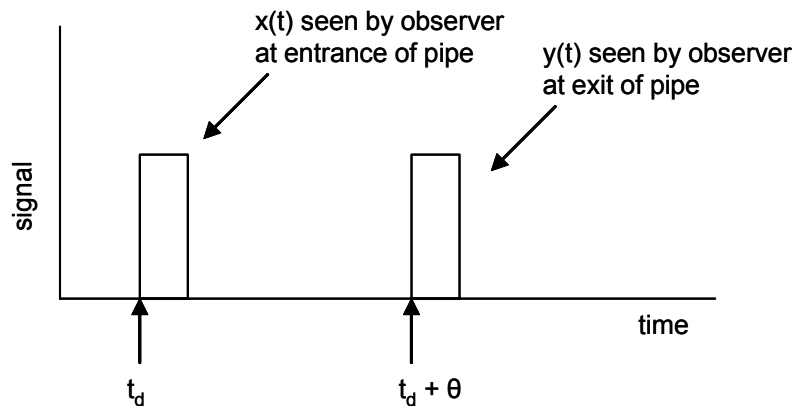
There may be occasions when detailed dynamic analysis of a process is warranted. Often, however, it is sufficient to obtain a simplified dynamic model that gives a reasonable approximation to the process behavior.

Figure 7.1-1 indicates that high-order step responses might be represented as a first-order rise following a period of delay. The dynamic model that would behave this way is called the FODT (first-order-dead-time) model; it turns out to be suitable for describing the dynamic response of many chemical processes.

7.3 dead time is delay

Before we examine the FODT model, we will look at dead time by itself. Chemical processes require that material be moved from one location to another: in conduits, on conveyer belts, through vessels. The transportation time is finite; it implies a delay between the onset of a disturbance at one location and its observation at another. This delay is often called dead time; it is familiar to anyone who has waited at the faucet for the hot water to arrive.

Consider a pipe carrying a liquid. A pulse of solute added at the entrance will be observed at the exit only after the solute is transported through the pipe.



The transit time depends on the liquid velocity and the length of the pipe. The figure indicates faithful transmission of the input signal $x(t)$ from inlet to outlet, as if every particle in the pipe moved at the same velocity. However, in real chemical processes the solute pulse $y(t)$ would become distorted through diffusive and dispersive effects. Nonetheless, a simple description of transport using the average fluid velocity is often sufficient to represent dead time in a process:

$$\theta = \frac{L}{\langle v \rangle} = \frac{V}{F} \quad (7.3-1)$$

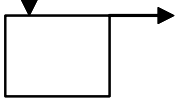
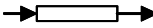
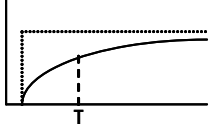
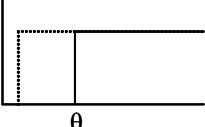
Thus the dead time is the residence time in the pipe.

Consider again Figure 7.1-1 and the series of tanks. If taken to the limit of an infinite number of tanks, each infinitesimally small, we finally obtain a pure delay, in which at time τ the full step disturbance appears at the outlet.

7.4 dead time and lag are different

In casual conversation, one might not distinguish between lag and delay; however, in process control these two terms have distinct meanings. A lag process is illustrated by a mixed tank, and a delay (or dead time) process by a pipe.

Table 7.4-1. Comparison of lag and dead time processes

	first-order lag	pure dead time
representative process		
defining equation	$\tau \frac{dy'}{dt} + y' = x'(t)$	$y' = x'(t - \theta)$
transfer function	$G(s) = \frac{y'(s)}{x'(s)} = \frac{1}{\tau s + 1}$	$G(s) = \frac{y'(s)}{x'(s)} = e^{-\theta s}$
step response		

In the tank step response, the output lags behind the input, but there is no delay between input and a response at the output. In the pipe, by contrast, the output is delayed.

7.5 frequency response of a dead time process

The sinusoidal input

$$x(t) = A \sin \omega t \quad (7.5-1)$$

is faithfully reproduced at the output of the dead time process, but will be delayed. Thus the amplitude ratio is unity, and the phase lag depends on the input frequency and the dead time. Inserting $j\omega$ into the Laplace transform in Table 7.4-1,

$$R_A = |e^{-\theta j\omega}| = 1$$

$$\phi = \angle e^{-\theta j\omega} = -\theta\omega \quad (7.5-2)$$

Thus dead time delays a signal in an unbounded manner, but does not diminish its amplitude. From what we recall of the Bode criterion for evaluating closed loop stability, we might speculate that dead time could be particularly significant.

7.6 the FODT model

To represent the dynamic behavior of a complex process as first order plus dead time is to say that the many storage elements and pipelines in the process can be represented by a single tank and pipe in series (the order does not matter). The FODT model is a first order ODE with a delayed input:

$$\tau \frac{dy'}{dt} + y' = Kx'(t - \theta) \quad (7.6-1)$$

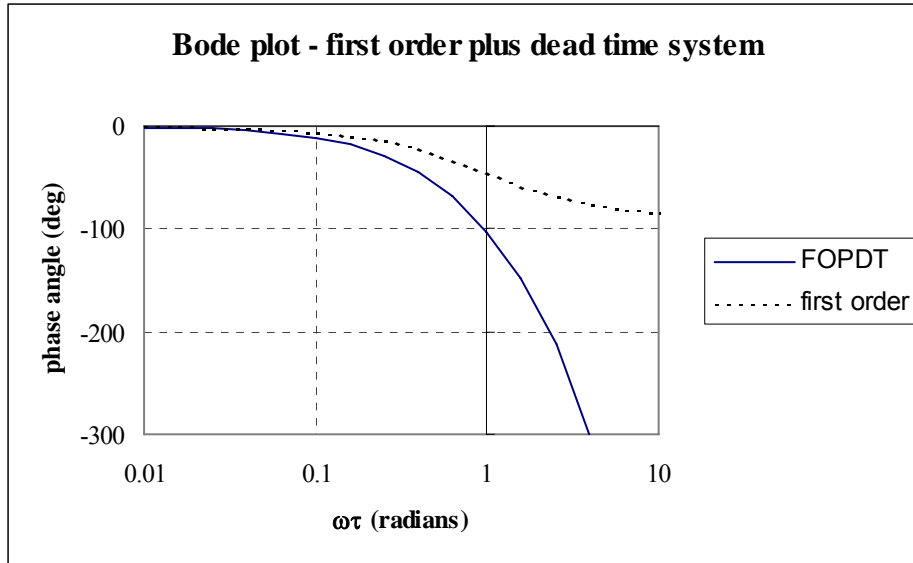
The FODT time constant τ comes from the tank, and the dead time θ from the pipe. Taking the Laplace transform of (7.6-1)

$$\frac{y'(s)}{x'(s)} = G(s) = \frac{K}{\tau s + 1} e^{-\theta s} \quad (7.6-2)$$

7.7 Bode plot of FODT

With dead time, we finally see how the phase angle can become significant: although the first-order lag phase angle is limited to -90° , the dead time contribution to phase lag is unbounded. In the plot, the dead time has been set equal to the time constant.

If a FODT process is placed in a feedback loop with a controller, a crossover frequency will inevitably be found, and the controller settings will be limited by the onset of instability..



7.8 identification - obtaining an FODT to represent a process

Arriving at a suitable approximate model (both form and parameter values) is known as identifying a model. Most often, the FODT approximation would be derived from an experimental test of the dynamic system. For example, a step input $x_{\text{data}}(t)$ would be contrived, and the response $y_{\text{data}}(t)$ measured. Then the experimental data would be compared with the FODT step response $y_{\text{model}}(t)$, and the parameters K , τ , and θ adjusted to achieve a satisfactory match. The match might be done by eye; alternatively, a least-squares criterion could provide an objective comparison between different sets of parameter values. Marlin (2000) describes further methods for obtaining parameter values from the data.

If the step input can be maintained sufficiently long to see the response become virtually constant, the gain K is relatively easy to determine. However, the response variable remains subject to other disturbances, which may distort the experimental data through noise and confounding inputs. Such confounding inputs may force a given step test to be short. Furthermore, in some systems it may be impossible to approximate a step input. In others, it may be undesirable to force a system away from the desired operating condition for a lengthy period. Thus other forms of input disturbance suitable for identifying a system model are discussed by Seborg et al (2004).

Figure 7.8-1 shows experimental data for a system with negative gain. The noise in the response variable trace makes the choice of dead time somewhat uncertain. Furthermore, the short duration of the run means that different combinations of gain and time constant can be used to fit the observed trace; a longer run would have distinguished these two parameters more clearly.

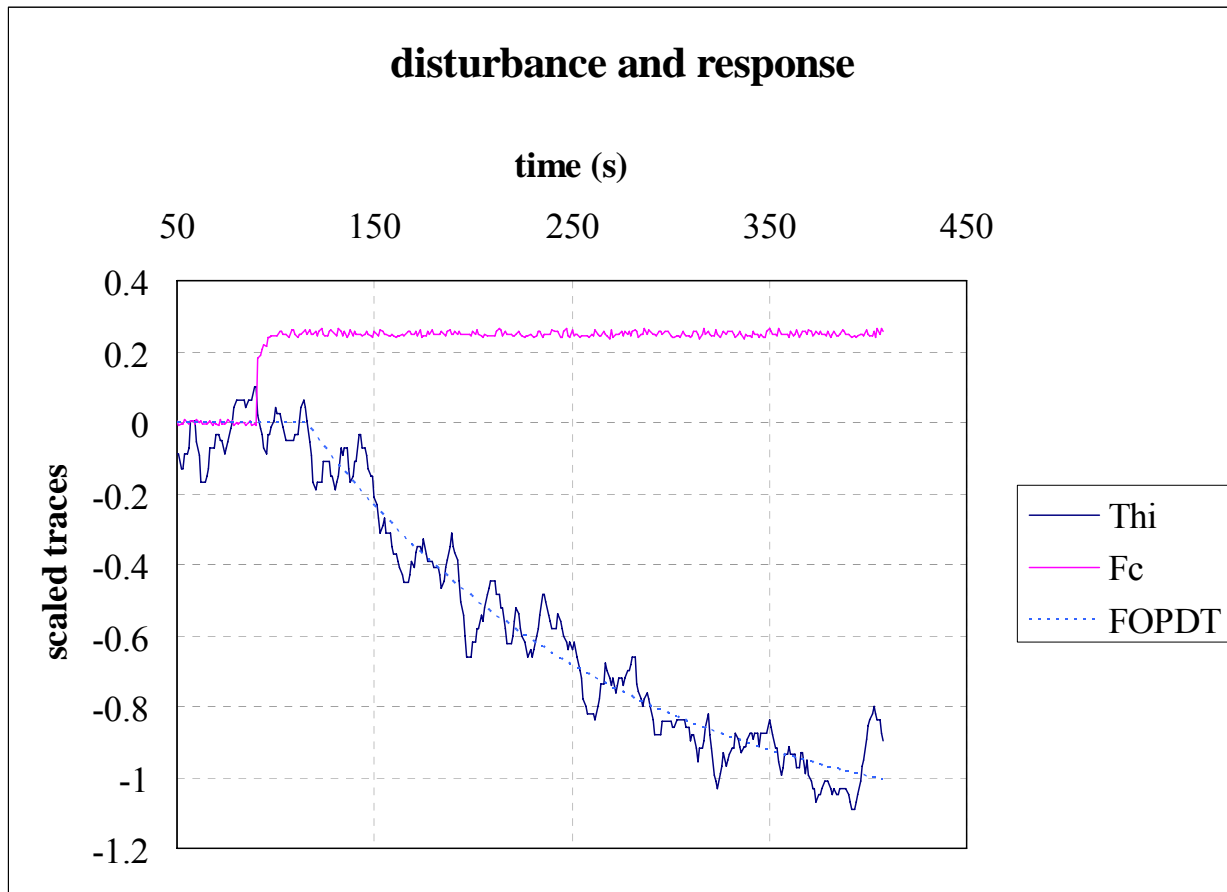


Figure 7.8-1 Experimental step response and FODT fit

We make use of a process model in closed loop calculations. For example, we have used the manipulated variable transfer function $G_m(s)$ to compute the loop transfer function. In an experimental test, however, we are likely to obtain G_m in combination with the sensor and valve transfer functions. This is because manipulating the valve with the controller output in manual is the most common way of creating a step input in the manipulated variable. Furthermore, we obtain the value of the response variable only by measuring it, usually with the installed sensor. Hence the FODT model actually includes the behavior of valve and sensor, as well. Notice in Figure 7.8-1 that the input is not quite a step, but rather a rapid first-order response. These dynamics become part of the ostensible process model.

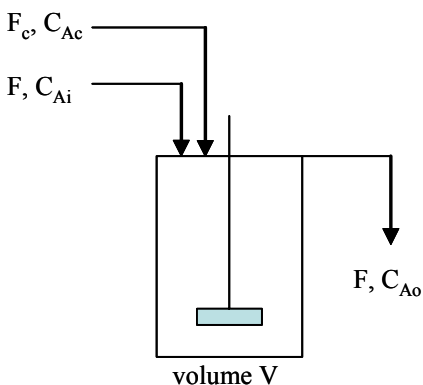
In Lesson 6 we encountered a process with underdamped behavior. In such a case, it may be possible to identify a suitable second-order-dead-time model, thus augmenting the more varied responses of second order with dead time. We will not pursue this idea further in this lesson.

7.9 digital calculation of dead time

Dead time is computed by defining arrays for the input and output variables. The arrays represent the variation of the inputs and outputs over time. The dead time is represented by a difference in the array index between input and output variables. For example, suppose that the variables are computed every five seconds, so that $x_{\text{input}}(i+1)$ represents the value of the input 5 s after that stored in $x_{\text{input}}(i)$. Then a dead time of 20 s would require that a change in the input variable occurring at index value i not be introduced to the output variable array until the index reaches $i+4$.

7.10 an example process with dead time

Consider the blending tank we studied in Lesson 3; however, recognize that the composition sensor may be placed some distance from the tank, so that there is a delay between achieving a composition in the tank and presenting its value for measurement.



The system model, adapted from Lesson 3, is

$$\tau \frac{dC'_{Ao}(t)}{dt} + C'_{Ao}(t) = C'_{Ai}(t - \theta) + \frac{C_{Ac}}{F} F'_c(t - \theta) \quad (7.10-1)$$

where the time constant τ is given by the ratio of tank volume to volumetric flow, and the dead time θ by the ratio of the length of pipe between tank and sensor to the average velocity of the outlet stream in the pipe. Notice that in this particularly simple case, we have derived the model from our understanding of the process, not experimental data.

The responses to changes in either of the inputs C'_{Ai} or F'_c will resemble those computed in Lesson 3, except that they will be observed only after a time interval of θ has followed the occurrence of the input.

CONTROL SCHEME

7.11 the control schemestep 1 - specify a control objective for the process

Our control objective is to maintain C_A at set point C_{As} .

step 2 - assign variables in the dynamic system

The controlled variable is C_A . From (7.10-1) we see that the manipulated variable F_c affects the controlled variable through a transfer function that includes dead time. The disturbance transfer function for input C_{Ai} includes dead time, as well. For feedback control, that does not matter - we observe that a disturbance has occurred only when the controlled variable begins to deviate from set point. (In other control schemes, however, we will be able to make use of measurements of the disturbance variable. In this case, it will be useful to know the dead time for disturbances.)

step 3 - PID (proportional-integral-derivative) controller algorithm

Derivative mode can help to stabilize the loop. However, we expect to find that dead time will force us to detune (that is, make less aggressive) the controller.

step 4 - choose set points and limits

The parameter values, also from Lesson 3, are

$$\begin{aligned} V &= 6 \text{ m}^3 \\ F &= 0.02 \text{ m}^3 \text{ s}^{-1} \\ \tau &= 300 \text{ s} \\ \theta &= 60 \text{ s} \\ F_{cs} &= 10^{-4} \text{ m}^3 \text{ s}^{-1} \\ C_{Ais} &= 8 \text{ kg m}^{-3} \\ C_{Aos} &= 10 \text{ kg m}^{-3} \\ C_{Ac} &= 400 \text{ kg m}^{-3} \end{aligned}$$

We will suppose that C_{Ai} may vary between 4 and 10 around its reference value of 8 kg m^{-3} . Therefore, from the steady state material balance, F_c must be capable of varying between 0 and $2 \times 10^{-4} \text{ m}^3 \text{ s}^{-1}$ to maintain C_A at set point.

EQUIPMENT**7.12 sensor transmitter range**

The sensor must respond to the controlled variable (as mercury rises in the glass), and its associated transducer or transmitter must convert this response to an input signal for the controller. For old-generation pneumatic controllers, this would be an air pressure; for electronic controllers and computers, it is often transmitted as an electric current that

varies between 4 and 20 mA. The controller will convert the current to a percentage of the maximum input. Thus we assign to the sensor a gain with dimensions of

$$K_s (=) \frac{\text{mA}}{\text{unit of controlled variable}} \frac{\%in}{\text{mA}} \quad (7.12-1)$$

The magnitude of K_s depends on the sensitivity, or the range, of the sensor. Increasing the sensitivity of the sensor will cause a larger signal to the controller for a given deviation in controlled variable. An increase in sensitivity, and thus sensor gain, increases loop gain in the same way as increasing the controller gain.

Some processes must employ multiple ranges. For example, a process with a set point of 500°C would have a low sensitivity (wide range) for monitoring the controlled variable during start-up from ambient conditions and a higher sensitivity (narrow range) for normal operation near the set point. Such an application might be done with multiple sensors, with variable sensitivity, or with multiple control loops, according to the particular case.

7.13 valve saturation

There is not unlimited power to manipulate a process: even though a control algorithm might calculate an output greater than 100%, this can direct the valve to be no more than fully open, or fully closed. The control engineer must size equipment so that the manipulated variable is sufficient to exert a countervailing influence to the anticipated disturbances.

For example, if the controlled variable is a reactor temperature and the manipulated variable is the flow of cooling water to the reactor jacket, the maximum flow rate must be sufficient to cool the reactor under the most unfavorable anticipated disturbance conditions. Then the valve must be selected, along with other piping components, to supply this flow. No controller tuning can compensate for a valve that is too small.

7.14 proportional gain and proportional band

We have worked with the gain of the controller, K_c . An alternative convention is also used: the proportional band. Proportional band is the inverse of the non-dimensional controller gain, multiplied by 100.

$$PB = \frac{100}{K_c} \quad (7.14-1)$$

Thus a gain of 1 %out %in⁻¹ would be equivalent to a proportional band of 100. Large proportional band implies low gain.

7.15 integral time and reset rate

Just as proportional mode has an inverse convention, the integral time is sometimes replaced with the reset rate. The reset rate is simply the reciprocal of the integral time, and has dimensions of repeats time⁻¹. Using a large reset rate is equivalent to a small integral time, implying aggressive integral mode control.

7.16 reset windup in integral mode

The integral mode integrates the error over time; a persistent error leads to a growing integral-mode contribution to the controller output. In our definition of the standard PID algorithm, there is no limit to the growth of this contribution. Hence under some circumstances the computed controller output could become significantly greater than 100%. The valve would be no more than fully open (or closed) as described in Section 7.13, but the controller output would be increasing. If the error were finally reversed, requiring the valve direction to reverse, the controller would be unable to direct it to do so until the integral mode contribution had been reduced by persistent error of the opposite sign.

Such a condition is called reset windup, and could occur under prolonged severe input step disturbances, or a fault in the loop (such as a mistakenly closed manual valve) that prevented the manipulated variable from affecting the controlled variable. Controller devices and algorithms generally include windup protection to prevent the unbounded growth of the integral mode.

7.17 protecting set point changes from derivative spikes

The standard PID algorithm defined the derivative mode in terms of the time derivative of the error. However, derivative mode need not be employed in response to set point changes, so the definition is changed to apply derivative only to changes in the controlled variable. Thus the algorithm becomes

$$x_{co} - x_{co,bias} = K_c \left(\varepsilon + \frac{1}{T_I} \int_0^t \varepsilon dt - T_D \frac{dy}{dt} \right) \quad (7.17-1)$$

Here y represents the controlled variable, and the sign in front of T_D has become negative, because error is always defined as the difference between set point and controlled variable.

Algorithm (7.17-1) is the basis of practical controller algorithms. However, for derivations and some illustrations we will still use the Laplace transform of original algorithm.

7.18 filtering the derivative mode

The derivative mode, by reacting to the rate of change of the controlled variable, is subject to noise. Applying derivative mode to a noisy signal

can introduce disturbances by needless variation of the manipulated variable. The effect of noise can be reduced by filtering the derivative mode; this is usually shown by altering the derivative-mode term in the PID algorithm:

$$\frac{T_D s}{\alpha T_D s + 1} \quad (7.18-1)$$

Equation (7.17-1) may be thought of as the transfer function that describes sending input $\varepsilon(s)$ first through a differentiator $T_D s$ followed by a first-order filter $(\alpha T_D s + 1)^{-1}$. In the time domain, this would be represented as

$$\alpha T_D \frac{d\varepsilon_{df}}{dt} + \varepsilon_{df} = T_D \frac{d\varepsilon}{dt} \quad (7.18-2)$$

where the variable ε_{df} is the differentiated, filtered error signal that is combined with the original error and the integrated error in the PID algorithm. The filter parameter α is often set between 0.1 and 0.2. If α were zero, ε_{df} would be the unfiltered derivative of the error ε in the ideal PID algorithm.

CLOSED LOOP BEHAVIOR

7.19 closed loop transfer function

The familiar closed loop diagram can be drawn, and the closed loop transfer function derived. For the disturbance

$$\frac{C'_A(s)}{C'_{Ai}(s)} = \frac{\frac{K_d}{\tau s + 1}}{1 + K_s K_v K_c \left(1 + \frac{1}{T_I s} + T_D s \right) \frac{K_m}{\tau s + 1} e^{-\theta s}} \quad (7.19-1)$$

In (7.19-1) we have omitted the dead time from the disturbance transfer function in the numerator because we (presume that we) have no independent measure of C_{Ai} and thus only know that a disturbance has occurred when we see the response in C_A . We proceed as before to resolve the fractions, and obtain

$$\frac{C'_A(s)}{C'_{Ai}(s)} = \frac{K_d s}{\tau s^2 + s + K_s K_v K_c K_m \left(T_D s^2 + s + \frac{1}{T_I} \right) e^{-\theta s}} \quad (7.19-2)$$

Here, we grind to a halt, because none of our Laplace transform experience has prepared us to deal with the exponential term in the

denominator. It is possible to substitute a polynomial approximation (called a Padé approximation) for the exponential term, and thus obtain an approximate transfer function that can be inverted, but we will not do this.

7.20 Bode criterion for closed loop stability

Instead, let us examine the stability of the closed loop by the Bode criterion. Extracting the loop transfer function from the denominator in (7.19-1) we obtain the frequency response.

$$R_A = |G_L(j\omega)| = K_s K_v K_c K_m \sqrt{1 + \left(T_D \omega - \frac{1}{T_I \omega}\right)^2} \frac{1}{\sqrt{1 + (\tau \omega)^2}} \quad (7.20-1)$$

$$\phi = \angle G_L(j\omega) = \tan^{-1}(-\omega\tau) + \tan^{-1}\left(T_D \omega - \frac{1}{T_I \omega}\right) - \theta\omega \quad (7.20-2)$$

From (7.20-2) we see that the crossover frequency depends on the system model parameters τ and θ , and it may be further influenced by controller parameters T_I and T_D . These controller settings also influence the amplitude ratio, along with the controller gain K_c .

Choosing sample process and controller parameters

$$\begin{aligned} K_s K_m K_v &= 0.5 \\ \tau &= 3 \text{ min} \\ \theta &= 1 \text{ min} \\ T_I &= 10 \text{ min} \\ T_D &= 2 \text{ min} \end{aligned}$$

the crossover frequency is computed to be 3.088 radians min^{-1} and the controller gain at instability to be 2.994. The Bode plot for the loop transfer function is shown in Figure 7.20-1. Amplitude ratios are plotted for gains below, at, and above the stability limit (which is indicated by a marker).

Notice that the amplitude ratio becomes level at high frequencies, instead of dropping off, as we have previously seen. This is because the decreasing amplitude of the first order process is balanced by the increasing amplitude of the derivative controller mode. The derivative mode filter described in Section 7.18 would prevent this high derivative gain at high frequencies.

The phase angle actually rises over some range of frequency, due to the contributions of the controller, but ultimately the process dead time dominates, and the phase angle crosses the -180° stability criterion.

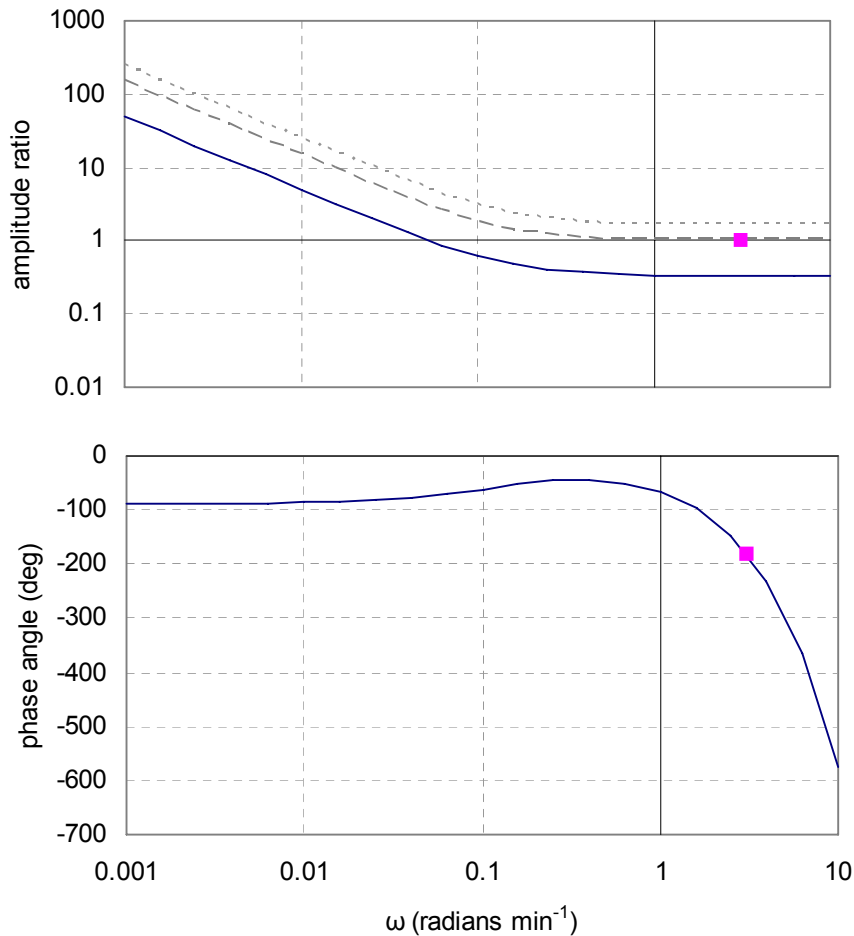


Figure 7.20-1. Bode plot for loop transfer function at gains 1, 3, and 5

A numerical calculation of closed loop response to a tiny step disturbance is given in Figure 7.20-2. The system is initially stable, but the 1% disturbance initiates a cycle of oscillation that increases in amplitude. In a perfectly linear system, the amplitude would increase without bound. In a practical closed loop, the valve output would oscillate between fully open and fully closed, and the process would move into regions of operation not well described by the model and parameters in use. While not strictly unbounded, the practical system response is nonetheless undesirable, so that the linear stability analysis has indicated a controller setting to be avoided.

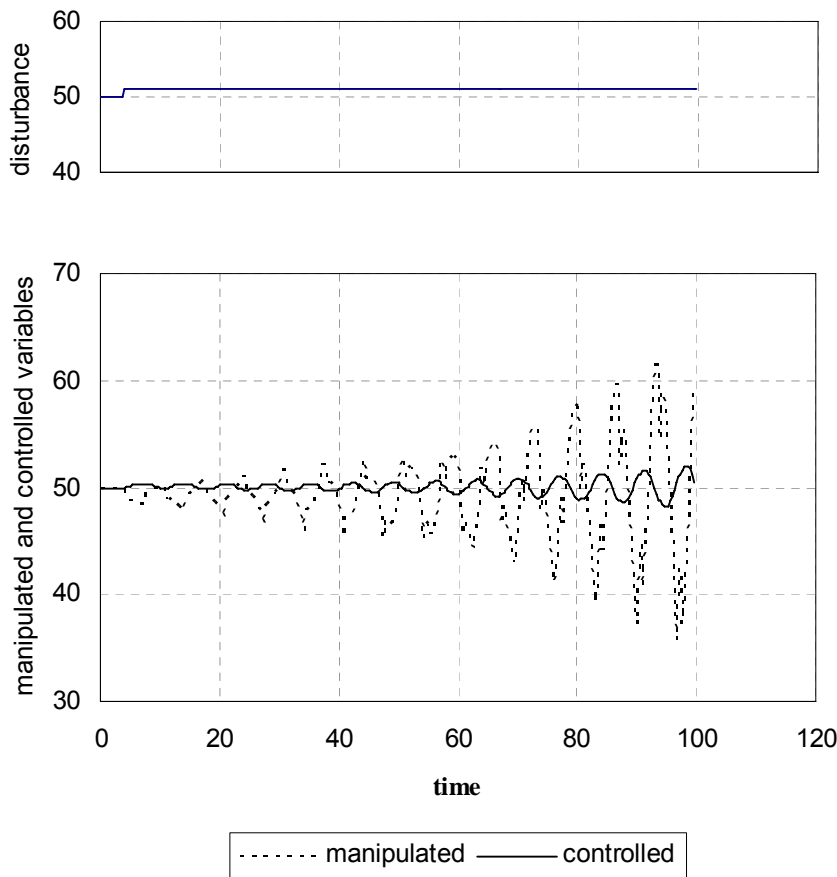


Figure 7.20-1. Unstable response to small disturbance at gain of 3

7.21 tuning the controller via correlations

The Bode criterion has given us an estimate of the stability boundary; we must still decide where to set the controller. We have previously tuned controllers by several methods:

- stability preservation - adjusting gain to realize particular gain and phase margins
- direct simulation with a process model to minimize integral measures of error, such as IE and IAE, for various inputs
- direct simulation with a process model using less comprehensive criteria, such as minimizing time to return to less than 2% deviation.

Now we introduce two correlations for setting controller parameters. In each case, the method represents the process as an FODT and specifies controller settings as a function of the model parameters. The correlations give different results because their authors had different ideas of what constituted “good control”.

Zeigler-Nichols

Probably the best known, it actually comes in two flavors. The first (open loop) defines controller parameters in terms of the FODT process parameters (p.347 in Marlin). The second (closed loop) depends on knowing the controller gain that will push the closed loop to the point of instability (constant amplitude oscillation), if only proportional mode is used. Controller parameters are then defined in terms of the gain and the frequency of the sustained oscillation (p.330 in Marlin). The independent variables in these methods can be determined from an existing FODT model, or from empirical plant testing (careful with that instability test, however!)

Ciancone

This is Marlin's recommendation. Read Chapter 9 for the story of its development. Read particularly Example 9.5 for a good illustration of its use. In short, the correlation provides robust tuning for an FODT process (and thus for anything reasonably approximated by FODT). The closed loop equations (FODT and PID controller) were solved numerically to predict controlled variable response to disturbance and set point steps. Controlled variable IAE was minimized by varying controller parameters K_c , T_i , and T_d . However, this optimization was broadened beyond a single operating case. It accounted for model error and changes in operating conditions by computing IAE for multiple cases, in which model parameters were varied. In addition, constraints were placed on variation in the manipulated variable. The correlation is presented as graphical relations among nondimensional parameters.

$$K_c K_p = f_1\left(\frac{\theta}{\tau + \theta}\right) \quad \frac{T_i}{\tau + \theta} = f_2\left(\frac{\theta}{\tau + \theta}\right) \quad \frac{T_d}{\tau + \theta} = f_3\left(\frac{\theta}{\tau + \theta}\right) \quad (7.21-1)$$

The controller gain is normalized by the product of the gains in the remainder of the loop (recall that the product of the gains around the feedback loop is dimensionless). The integral and derivative times are normalized by the sum of the FODT time constant and dead time. The independent variable is the fraction dead time in the process.

Other correlations could be used (many are available; Seborg et al (2004) give a good overview), but the general idea should be clear: first get knowledge of the process itself, such as a dynamic model and the values of its parameters, and use this knowledge to choose controller settings. The variety of results that can be obtained in practical systems is due partly to the richness of the mathematical form of the PID algorithm in a closed feedback loop, and partly to the approximate nature of our process models - the real process simply has more potential for complexity than our mathematical models can predict. Marlin (2000) gives further discussion of the effects of error in process models.

7.22 conclusion

We have asserted that an enormous variety of chemical processes may be represented by the FODT approximation. With the FODT process we have introduced the crucial issue of dead time. Dead time is one of the distinguishing features of chemical process control, and exerts strong influence on how well the process can be controlled.

Having such a standard system model available allows the development of correlations for tuning controllers. The correlations point out that good knowledge of the process is the basis for successful control. We will find that as more detailed understanding of the process is obtained, more sophisticated control strategies can be employed.

7.23 references

Marlin, Thomas E. *Process Control*. 2nd ed. Boston, MA: McGraw-Hill, 2000. ISBN: 0070393621.

Seborg, Dale, Thomas Edgar, and Duncan Mellichamp. *Process Dynamics and Control*. 2nd ed. Hoboken, NJ: Wiley, 2004. ISBN: 0471000779.