# Limit Cycles with Imperfect Valves: Implications for Controllability of Processes with Large Gains

**Antonio C. B. de Araújo and Sigurd Skogestad***

*Department of Chemical Engineering, Norwegian University of Science and Technology (NTNU), Trondheim, Norway*

There is some disagreement in the literature in regard to whether or not large plant gains are a problem when input−output controllability issues are involved. In this paper, controllability requirements are derived for two types of input errors, namely, restricted (low) input resolution (e.g., caused by a sticky valve) and input disturbances. In both cases, the controllability is limited if the plant gain is large at high frequencies. Limited input resolution causes limit cycle behavior (oscillations) similar to that observed with relay feedback. The magnitude of the output variations is dependent on the plant gain at high frequency but is independent of the controller tuning. Provided frequent input (valve) movements are acceptable, one may reduce the output magnitude by forcing the system to oscillate at a higher frequency, for example by introducing a faster local feedback (e.g., a valve positioner) or by pulse modulating the input signal.

## 1. Introduction

The main goal of feedback control is to the keep the plant output ($y$) within specifications, despite disturbances, errors, and uncertainty. A fundamental question arises: Is the process input−output controllable? There are many factors that must be considered, and one of them is the magnitude of the process gain. The gain is dependent on the frequency and, for multivariable plants, it also is dependent on the input direction. To quantify this, the singular values $\sigma_i(G(j\omega))$ of the process transfer function $G(s)$ are considered. Of particular interest are the maximum and minimum singular values, denoted as $\bar{\sigma}(G)$ and $\underline{\sigma}(G)$, respectively. In this paper, for simplicity, mainly single input−single output (SISO) systems are considered, where $\bar{\sigma}(G(j\omega)) = \underline{\sigma}(G(j\omega)) = |G(j\omega)|$.

It is well-accepted that small process gains may cause problems. For example, the requirement for avoiding input saturation is

$$\underline{\sigma}(G) \geq 1$$

that is, a minimum gain of 1 is required.[1] This assumes that the desired output changes (setpoints) are of magnitude 1 and the allowed inputs are also of magnitude 1, both expressed in terms of the 2-norm.

It is less clear whether large process gains pose a problem. Skogestad and Postlethwaite[2] considered the condition number, which is defined as

$$\gamma(G) = \frac{\bar{\sigma}(G)}{\underline{\sigma}(G)}$$

and make the following statement: *A large condition number may be caused by a small value of $\underline{\sigma}(G)$, which is generally undesirable. On the other hand, a large value of $\bar{\sigma}(G)$ is not necessarily a problem.*

On the other hand, Moore[3] claims that high sensitivity (high gains) can be a problem, because of the low input resolution in valves and actuators. He states: *Valves and other actuators all have a minimum resolution, with respect to positioning. These limitations restrict the fine adjustments that are often necessary for high gain processes to attain steady operation. If the fine adjustment that is necessary for steady state is less than the resolution of the valve, sustained oscillations are likely to occur. Consider, for example, a steam valve with resolution of ±1.0%. If a valve position of 53.45% is necessary to meet the target temperature, then the valve will, at best, settle to a limit cycle that hunts over a range from* ∼55% *to 53%. If the process gain is 10, the hunting of the valve will cause a limit cycle in the control temperature of 20%.* In this paper, we confirm that limit cycles are unavoidable under such conditions, but we also find that it is the process gain at the frequency of the limit cycles, and not at steady-state, that matters for controllability.

McAvoy and Braatz[4] argued, along the same lines as Moore,[3] and state that, for control purposes, the magnitude of steady-state process gain ($\bar{\sigma}(G)$) should not exceed ∼50.

In this paper, two main types of input errors are discussed. We first consider the input oscillations caused by restrictions of the input (valve) resolution. Later, in section 7, we consider input (load) disturbance, which is not related to the valve resolution problems. Most of the results are derived for first-order plus delay processes. When possible, more general derivations are presented.

## 2. Restricted Input Resolution and Limit Cycles

As mentioned by Moore[3] and proved below, feedback control with restricted (low) input resolution results in limit cycles (hunting). A simple representation of restricted (low) input resolution is to use a quantized input, as depicted in Figure 1. The output $u_q$ from the quantizer is

$$u_q = q \cdot \mathbf{round}\left(\frac{u}{q}\right) \qquad (1)$$

where $q$ is the quantization step and the **round** function takes its argument to the nearest integer. This may, for example, represent restricted valve resolution and, to some extent, valve stiction and valve dead band.[5] An extreme case with only one quantization step is an on−off valve.

Figure 2 shows a feedback system with a quantizer. Here, $G(s)$ is the plant transfer function model, $K(s)$ the controller, $y$

* To whom correspondence should be addressed: Tel.: +47-7359-4154. Fax: +47-7359-4080. E-mail: skoge@chemeng.ntnu.no.

10.1021/ie060323p CCC: $33.50 
Published on Web 11/18/2006

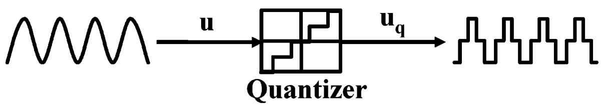


**Figure 1.** Quantization of a smooth signal.

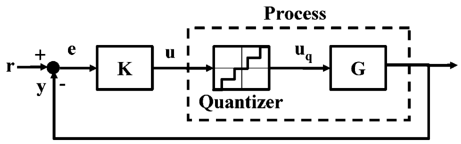


**Figure 2.** Feedback control of a process with restricted input resolution (quantizer).

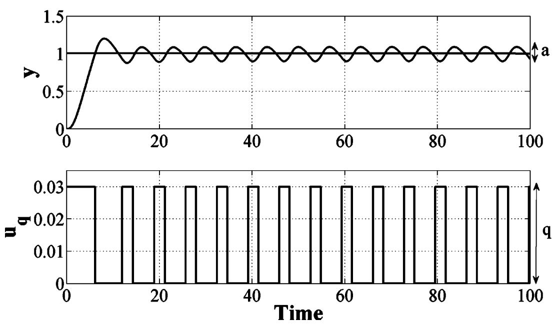


**Figure 3.** Simulation results for the system in example 1.

the plant output with reference $r$, and $u$ the manipulated variable (for simplicity, the Laplace variable $s$ is often omitted). The low input resolution results in a stepwise input "disturbance" of magnitude $q$, and this again results in oscillations in the plant output $y(t)$ of magnitude $a$. Note that $a$ here is defined as the "total" amplitude from the bottom to the top of the oscillations.

*Theorem 1. For the feedback system with a quantizer in Figure 2, limit cycles are inevitable if there is integral action in the controller, such that the output, on average, has no steady-state offset.*

*Proof. At steady state, the average value of the output $y$ is equal to the reference $r$; that is, $y_{ss} = r$, where $y_{ss}$ denotes the average ("steady-state") value of $y(t)$ as $t \to \infty$. To achieve this, the input $u$ must, on average, equal the following value:*

$$u_{ss} = \frac{y_{ss}}{G(0)} = \frac{r}{G(0)} \tag{2}$$

*where $G(0)$ denotes the steady-state plant gain. Except for the special case that $u_{ss}$ happens to correspond exactly to one of the quantizer levels $q_i$ (which, in practice, with measurement noise, will not occur), the quantized input $u_q$ must then cycle between at least two of the quantizer levels.*

Let us consider the most common case where the output cycles between the two neighboring quantizer stabilizes to $u_{ss}$, which, here, are denoted $q_1$ and $q_2$. Let $f$ and $(1 - f)$ denote the fraction of time spent at each of the two levels. Then, at steady state (as $t \to \infty$), $u_{ss} = fq_1 + (1 - f)q_2$ and we have the following expression for the fraction of time $u$ spends at level $q_1$:

$$f = \frac{q_2 - u_{ss}}{q_2 - q_1} \tag{3}$$

Note that the closer $u_{ss}$ is to one of the quantizer levels, the longer the time $u_q$ will remain on it.

**Example 1.** As an example, consider the system simulated in Figure 3, where $q_1 = 0$ and $q_2 = 0.03$ (this may represent an on/off valve). The third-order plant model is

$$G(s) = \frac{100}{(10s + 1)(s + 1)^2} \tag{4}$$

and we use a proportional−integral (PI) controller,

$$K(s) = K_c\left(\frac{\tau_I s + 1}{\tau_I s}\right); \quad K_c = 0.04, \quad \tau_I = 10 \tag{5}$$

Note that the integral time is chosen so that we cancel the dominant pole in $G(s)$ (the IMC tuning rule). The steady-state plant gain is $G(0) = 100$. Initially, the system is at steady state, with $u_q = q_1 = 0$ and $y = r = 0$. We then make a step change $r = 1$. The steady-state plant gain is $G(0) = 100$; therefore, to achieve $y_{ss} = 1$, the required average input is $u_{ss} = 1/100 = 0.01$ which is closer to $q_1 = 0$ than $q_2 = 0.03$. The fraction of time that $u_q$ remains at $q_1 = 0$ is $f = (0.03 - 0.01)/0.03 = 0.67$. As expected, this agrees with the simulations.

**Example 2.** A similar simulation example with $q_1 = 0$ and $q_2 = 0.03$ is shown in Figure 3, but for a first-order with delay plant:

$$G(s) = \frac{ke^{-\theta s}}{\tau s + 1} \tag{6}$$

with $k = 100$, $\theta = 1$, and $\tau = 10$. We use the same PI controller as in eq 5 with $\tau_I = \tau = 10$ and $K_c = 0.04$. The main difference, in comparison to example 1, is that the step reference change is much smaller, $r = 0.2$, such that the input stays at the upper quantizer level of $q_2 = 0.03$ a much shorter time. The steady-state plant gain is $k = G(0) = 100$; therefore, to achieve $y_{ss} = 0.2$, the required average input is $u_{ss} = 0.2/100 = 0.002$. From eq 3, the fraction of time $u_q$ remains at $q_1 = 0$ is $f = (0.03 - 0.002)/0.03 = 0.93$. Again, this is consistent with the simulations.

For the simulated system in Figure 3 (see example 1), the magnitude of limit cycles (oscillations) in $y$ is $a = 0.189$ and the period is $T = 6.72$ s. The oscillations in $y(t)$ are observed to be quite similar to sinusoidal. For the simulated system in Figure 4 (see example 2), we have $a = 0.3$ and $T = 16.07$ s. However, in this case, the oscillations in $y(t)$ are far from sinusoidal.

We next want to derive analytic expressions for $a$ and $T$. We first make the simplifying assumption that the resulting limit cycles are sinusoidal, and then we study the more general case.

## 3. Describing Function Analysis of Oscillations (Assuming Sinusoids)

The quantizer (nonlinearity) that causes the limit cycles can be regarded as a relay without hysteresis and, in the following, is treated as such. As an approximation, the amplitude of the oscillations can then be determined analytically from an harmonic linearization or describing function analysis of the nonlinearity. This analysis is exact if the resulting limit cycle is sinusoidal. For the feedback system in Figure 2, the condition for oscillation is given by[7]

$$N(a_u)L(j\omega) = -1 \tag{7}$$

where $N(a_u)$ is the describing function of the nonlinearity (quantizer), which is a function of the amplitude $a_u$ of the oscillations in $u(t)$ at the quantizer input, and $L = GK$ is the

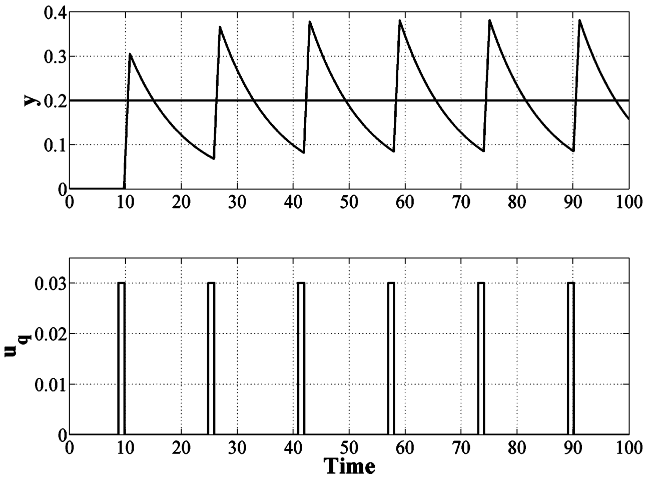


**Figure 4.** Simulation results for the system in example 2.

loop-transfer function (excluding the quantizer). For a relay without hysteresis, the describing function is[6]

$$N(a_u) = \frac{4q}{\pi a_u} \tag{8}$$

where $q$ is the relay amplitude (quantization step). Because, according to eq 8, $N(a_u)$ is a real number, it follows from eq 7 that $\omega$ is actually the ultimate frequency ($\omega_{L,180}$) and

$$N(a_u) = \frac{1}{|L(j\omega_{L,180})|} = \frac{4q}{\pi a_u} \tag{9}$$

The amplitude of the corresponding oscillations at the plant output are $a = a_u/|K(j\omega_{L,180})|$, which leads to

$$a = \frac{4q|G(j\omega_{L,180})|}{\pi} \tag{10}$$

$$T = \frac{2\pi}{\omega_{L,180}} \tag{11}$$

where $T$ is the period of oscillation. This is exact if the limit cycles are sinusoidal.

**Example 1 (continued).** For the system given by eqs 4 and 5, $\angle L(j\omega_{L,180}) = -\pi/2 - 2\arctan(1 \cdot \omega_{L,180}) = -\pi[\text{rad}] = -180°$, which yields $\omega_{L,180} = 1$ rad/s and $|G(j\omega_{L,180})| = 4.999$. From a describing function analysis, the period of oscillation is then $T = 2\pi/\omega_{L,180} = 6.28$ s, and, from eq 10, $a = (4/\pi)q|G(j\omega_{L,180})| = 0.191$. This is in good agreement with the simulation results ($T = 6.72$ s, $a = 0.189$).

**First-Order with Delay Process.** Consider a first-order with delay plant $G()$ controlled by a PI controller with $\tau_I = \tau$,

$$G(s) = \frac{ke^{-\theta s}}{\tau s + 1} \tag{12}$$

$$K(s) = K_c\left(\frac{\tau_I s + 1}{\tau_I s}\right), \quad \tau_I = \tau \tag{13}$$

For this system, we have $\angle L(j\omega_{L,180}) = -(\pi/2) - \omega_{L,180}\theta = -\pi$, which gives $\omega_{L,180} = (\pi/2)(1/\theta)$ and $|G(\omega_{L,180})| = k/\sqrt{[(\pi/2)(\theta/\tau)]^2+1}$. From the describing analysis in eqs 10 and 11, we then have

$$a = \frac{4}{\pi}\frac{qk}{\sqrt{[(\pi/2)(\tau/\theta)]^2 + 1}} \text{ and } T = 4\theta \tag{14}$$

For small delays ($\theta/\tau \ll 1$), this gives $a \approx (8/\pi^2)q(k/\tau)\theta$, and we see that amplitude of the oscillations increases proportionally with $k'k/\tau$ (which is the initial slope of the step response) and $\theta$. For large delays ($\theta/\tau \gg 1$), $a \approx (4/\pi)qk$, and we see that amplitude of the oscillations increases proportionally with $k$ (steady-state gain) and is independent of $\theta$. In all cases $a$ increases proportionally with $q$.

**Example 2 (continued).** With $k = 100$, $\theta = 1$, $\tau = 10$, and $q = 0.03$, eq 14 gives $T = 4$ s and $a = 0.243$. This should be compared with the actual value from the simulations which are $T = 16.1$ s and $a = 0.296$. Taking into account that the oscillations in $y(t)$ are far from sinusoidal, the value of $a$ in eq 14 obtained from the describing function analysis is quite good (~20% too small). However, the period $T$ is a factor of 4 too small.

The two examples suggest that the amplitude of $a$ in eq 17 from the describing function analysis is quite accurate, but they also indicate that the actual period may be much larger. This conclusion is confirmed by an exact analysis for a first-order with delay plant, which is presented next.

## 4. Exact Analysis of Oscillations for a First-Order Plus Delay Process

In this section, exact results for non-sinusoidal quantized responses are derived for a first-order with delay plant that is controlled by a PI controller with $\tau_I = \tau$. The following theorem is based on the work by Wang et al.[10]

*Theorem 2. For a system given by eqs 12 and 13, set up according to the configuration of Figure 2 with a quantizer level q, the amplitude and period of the limit cycle oscillations are*

$$a = kq\frac{1 - e^{-t_1/\tau} + e^{-T/\tau} - e^{-(T-t_1)/\tau}}{1 - e^{-T/\tau}} \tag{15}$$

and

$$T = \theta\left(\frac{1}{1-f} + \frac{1}{f}\right) \tag{16}$$

*where* $t_1 = \theta/(1 - f)$ *and f is calculated from* $u_{ss} = fq_1 + (1 - f)q_2$.

*Proof. See the Appendix.*

**Example 2 (continued).** With $f = 0.933$, the amplitude and period of oscillation calculated using eqs 15 and 16 are $a = 0.2962$ and $T = 16.07$ s, respectively, which matches exactly the observed results in Figure 4.

Note that the assumption $\tau_I = \tau$ is the reason why $a$ and $T$ are independent of the controller settings $K_c$ and $\tau_I$.

In Figure 5, the amplitude $a/(kq)$ from eq 15 is plotted as a function of $\theta/\tau$ for various values of $f$. For small delays ($\theta \ll \tau$), $a$ increases almost proportionally $\theta/\tau$; however, for large values of $\theta$, it levels off at a constant value of $a = kq$. Note that $a$ is dependent only weakly on $f$.

To compare, the dashed line in Figure 5 represents eq 14 from the describing function analysis. The agreement is generally very good, with a maximum difference of 27% for large values of $\theta/\tau$.

On the other hand, note that the period of oscillation can be very different from that which is observed with the describing function analysis. From eq 16, the period $T$ increases proportionally with the delay $\theta$, which agress with the value $T = 4\theta$ in eq 16 from the describing function analysis. However, in the exact analysis, $T$ also is dependent on $f$ and goes to infinity as $f$ approaches 0 or 1. From eq 16, the minimum value $T = 4\theta$

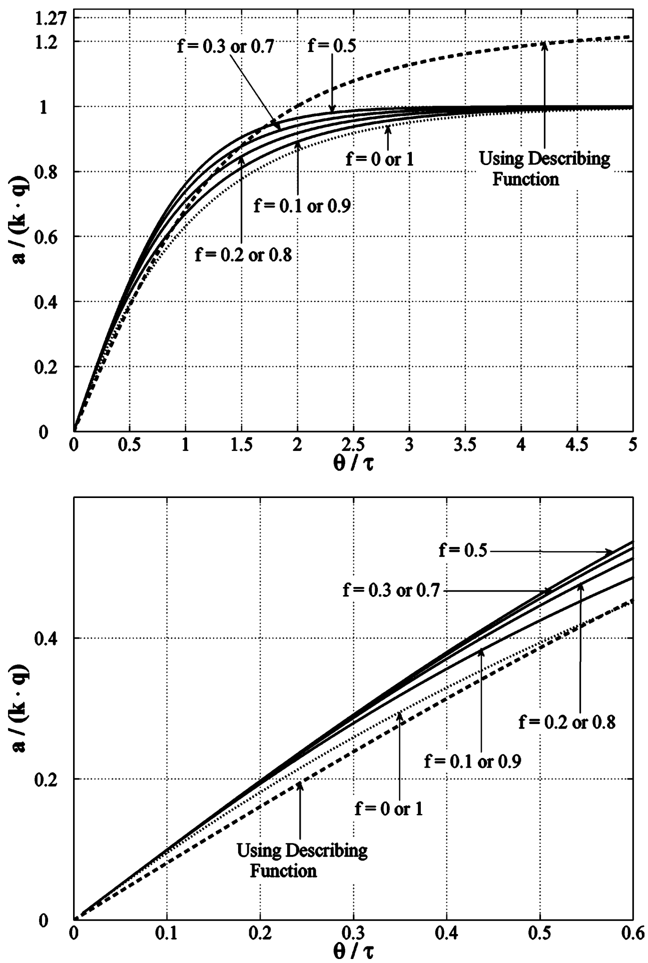


**Figure 5.** Amplitude $a$ in eq 15, plotted against $\theta/\tau$, for first-order plus delay processes. The lower figure is a magnified view of the upper figure for small values of $\theta/\tau$.

is obtained when $f = 0.5$, and only this limiting value agrees with the describing function analysis. This is not too surprising, because the input is most similar to "sinusoidal" when $f = 0.5$.

## 5. Controllability Requirements for Systems with Restricted Input Resolution

Consider a feedback system with restricted input resolution (quantized input), as shown in Figure 2. Assume there is integral action in the controller, such that there are limit cycles (Theorem 1). Let $a_{\max}$ denote the maximum allowed amplitude of the limit cycles (oscillations) in $y$. Then, from eq 10, the following approximate controllability requirement applies:

$$|G(j\omega_{L,180})| < \frac{\pi}{4}\left(\frac{a_{\max}}{q}\right) \qquad (17)$$

Note that this condition is dependent on the plant only and, more specifically, the plant gain at frequency $\omega_{L,180}$.

**Remark 1.** The controllability condition (eq 17) is approximate because it is based on a describing function analysis, which is exact only for sinusoidal oscillations. Nevertheless, the results in the previous section indicates that the gain from the describing function analysis is surprisingly accurate. For a first-order plus delay process, the maximum deviation was only 27% (for large values of $\theta/\tau$). Thus, eq 17 is expected to provide a tight controllability condition.

**Remark 2.** The controller has some effect on the condition, because $\omega_{L,180}$ is the frequency where the sum of the phase lag in the controller $K$ and plant $G$ is $180^o$. However, for a well-tuned controller, we typically have $\omega_{L,180} \approx 1.57/\theta$; that is, $\omega_L$ is dependent only on on the effective delay $\theta$ in the plant. Specifically, this value applies for a first-order (or second-order) plant that has been tuned with an SIMC PI(D) controller[8] (the value is exact when $\tau_1$ is smaller than $\sim 8\theta$, where the SIMC rule is $\tau_I = \tau_1$, and also applies well for the case when $\tau_1$ is large and the SIMC rule is $\tau_I = 8\theta$).

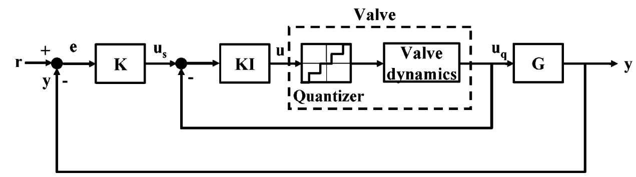


**Figure 6.** Frequency modulation generated using a valve position controller (KI).

**Remark 3.** Persistent oscillations are generally undesirable. Therefore, the allowed $a_{\max}$ value for oscillations is typically considerably much smaller ($\sim$10%) than the maximum allowed output deviation, $y_{\max}$; i.e., $a_{\max} = 0.1y_{\max}$.

## 6. How To Mitigate Oscillations Caused by Restricted Input Resolution

From the describing function analysis, the magnitude $a$ of the output oscillations for the system in Figure 2 is given by eq 14. The magnitude can be reduced, for example, by the following means:

(a) Change the valve so that the resolution is better (use a smaller quantization level $q$).

(b) Redesign the process or the measurement devices to get a smaller effective delay $\theta$.

(c) Introduce fast, forced cycles at the input with a higher frequency than those generated "naturally". For example, one may use high-frequency pulse modulation or add a high-frequency "dither" signal (e.g., amplified measurement noise caused by derivative action in controller).

(d) "Valve positioner": Use a measurement of $u_q$ and add a local feedback at the input to generate faster cycling (see Figure 6). This may be viewed as a combination of cases (b) and (c).

The problem with approaches (b), (c), and (d) is that frequent input moves may be undesirable, for example, because the valve cannot be moved so quickly or because of excessive wear.

**Frequency (Pulse) Width Modulation.** Let us consider, in more detail, approaches (c) and (d). A system with restricted (low) input resolution and no (average) steady-state offset is bound to cycle (Theorem 1) and the amplitude $a$ of the oscillations is given by the process gain at the frequency of oscillations (e.g., see eq 10). So far, we have let the system cycle at its "natural" frequency, $\omega_{L,180}$, as given by eqs 11 and 16. However, because the gain $|G(j\omega)|$ for most processes is lower at high frequencies, an attractive alternative is forcing the system to cycle at a higher frequency.

One approach, (d), uses a valve position controller that is based on measuring $u_q$, as shown in Figure 6. Here, the controller $K$ sets the setpoint $u_s$ for the valve position (input), and the "internal" valve position controller (KI) adjusts the input $u$ signal such that the actual input $u_q$ matches the desired input $u_s$ (at least on average). The valve position controller (KI) should have integral action, or a sufficiently high proportional gain, such that the internal loop cycles. The frequency of the cycling is determined by the effective delay in the "internal" valve

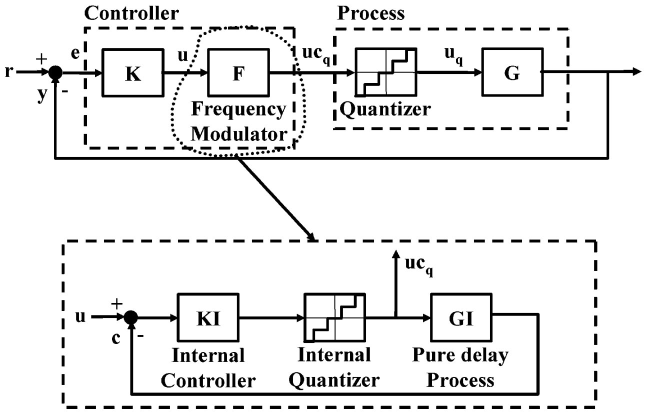


**Figure 7.** System with frequency modulation. The box shows one way of generating high-frequency oscillations. Alternatively, for an on/off valve, a clock may be used to set the frequency while the controller sets the pulse width.

position loop, which generally is much smaller than the delay in the overall outer loop. The results is that the frequency of the oscillations is much higher and the resulting amplitude *a* of the output is much smaller. This is consistent with the recommendations in the *Instrument Engineers' Handbook*,[11] where it is noted that a positioner can reduce the dead band of a valve/actuator combination from as much as 5% to <0.5%.

However, one may not have a measurement of the actual input $u_q$, and a valve position controller is, in fact, not necessary to reduce the effect of low input resolution. A more general approach, (c), is to introduce forced pulsing by adding a frequency modulator *F* at the output of the controller. One realization for *F* is an internal feedback loop, as depicted in Figure 7. This is similar to the valve positioner controller, except that we need an internal quantizer, because there is no measurement of $u_q$. The modulator forces the system to cycle at a higher frequency than that which follows "naturally". For example, forced pulsing is commonly used for on/off valves in small-scale plants where the valve may open or close every second and the controller sets the average position.

**Example 3.** Through the use of a valve position controller, as shown in Figure 6, the response of the system in eqs 12 and 13 is depicted in Figure 8. The valve dynamics is assumed to be a delay of 0.1, and the remaining process (*G*) has a delay of 0.9. As it can be seen, the output amplitude is drastically reduced at the expense of high-frequency input oscillations.

**P-Control.** Another potential approach to eliminate oscillations is to use a P controller (with a sufficiently low controller gain). However, in practice, this approach is not acceptable, because it results in an unacceptable steady-state offset. Consider a set-point change *r*, for which the desired input to achieve no offset is $u_{ss} = r/G(0)$ (see eq 2). Assume that *r* is such that $u_{ss}$ is between two quantization levels for the input. For any non-oscillating control system, including feedforward, we then have $\Delta u = |u_q - u_{ss}| \geq q/2$ and the resulting offset in the output is

$$|y - r| = |G(0)|\cdot|u_q - u_{ss}| \geq |G(0)|\frac{q}{2} \qquad (18)$$

From this observation, we conclude that the offset $|y - r|$ will be large for a plant with a large steady-state gain, $|G(0)|$, so P control is, in practice, not recommended as a method to mitigate oscillations.

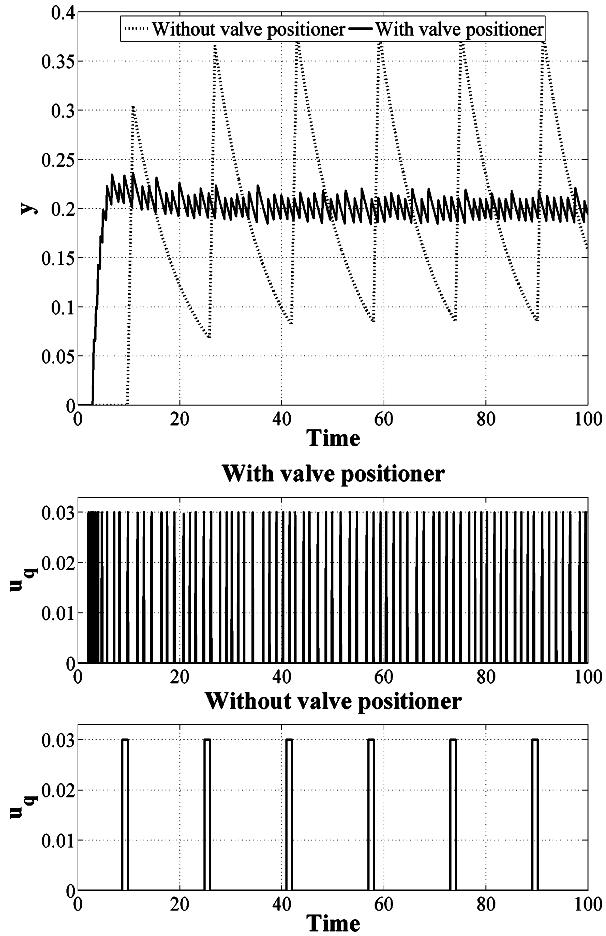


**Figure 8.** Effect of using valve position control for the system in example 2.

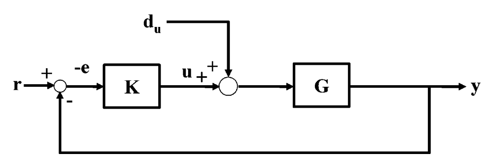


**Figure 9.** Block diagram of a feedback control system with disturbance at the input of the plant.

## 7. Input (Load) Disturbance

Consider a plant model in deviation variables

$$y(s) = G(s)u(s) + G_d(s)d(s) \qquad (19)$$

where *G* is the plant model, $G_d$ the disturbance model, *y* the plant output, *u* the manipulated variable, and *d* the disturbance (for simplicity, the Laplace variable *s* is often omitted). Without control, the effect of disturbances on the output is $y = G_d(s)d$, and the term "large" disturbances means that the product $|G_d d|$ is large, such that the output deviation $|y|$ will be large, unless we apply control. In this section, input disturbances are mainly considered, i.e., $G_d = G$. This case is illustrated in Figure 9, where $d = d_u$ is the disturbance at the plant input.

**Feedforward Control.** As mentioned in the Introduction, a large plant gain, especially at steady state, is a problem with feedforward control. As an example, consider a plant $y = G(u + d)$, where $d = d_u$ is the input (load) disturbance. Clearly, if $|G|$ is large, then $|u + d|$ must be small, to avoid a large $|y|$. With feedforward control, *u* is adjusted based on measuring *d*. First, an accurate measurement of *d* is required and it must be possible to adjust *u* such that $|u - d|$ is small. The latter is not possible with restricted input resolution. For example, returning to the example of Moore[3] from the Introduction: $|u - d| = 2\%$ and $|G| = 10$ gives $|y| = 20\%$, all at steady state.

**Feedback Control.** On the other hand, with feedback control, "large" disturbances are not necessarily a problem, at least not at steady state. Consider a single disturbance $d$. Without control, the steady-state sinusoidal response from $d$ to the output is $y(\omega) = G_d(j\omega)d(\omega)$, where phasor notation is used and $|d(\omega)|$ denotes the magnitude of the disturbance at frequency $\omega$. We assume that the magnitude is independent of the frequency, i.e., $|d(\omega)| = d_0$, and we assume that the control objective is that the output is less than $y_{max}$ at any given frequency, i.e., $|y(\omega)| < y_{max}$. From this, one can immediately draw the conclusion that no control is needed if $|G_d(j\omega)d_0| < y_{max}$ *at all frequencies* (in which case the plant is said to be "self-regulating"). If $|G_d(j\omega)d_0| > y_{max}$ at some frequency, then control is needed. With feedback control ($u = -Ky$), we get $y(s) = S(s)G_d(s)d(s)$, where $S$ is the sensitivity function ($S = (I + GK)^{-1}$). The requirement $|y(\omega)| < y_{max}$ then becomes

$$|S(j\omega)|\cdot|G_d(j\omega)||d(\omega)| < y_{max} \quad \forall\omega \tag{20}$$

With integral action in the controller, $|S|$ is zero at steady state; therefore, generally, it does not matter if $|G_d|$ is large at steady state (provided there is no problem with input saturation, but this is mainly a design issue rather than a control issue). However, $|S|$ increases with frequency and crosses a value of 1 at the bandwidth frequency $\omega_S$: $|S(j\omega_S)| = 1$. At this frequency, the requirement described by eq 20 gives the controllability requirement

$$|G_d(j\omega_S)| < \frac{y_{max}}{|d(\omega_S)|} \tag{21}$$

**Input Disturbance.** However, the purpose of this paper is not to consider plants for which $|G_d|$ is large, but rather plants for which $|G|$ is large (in practice, these are often related, because all plants have disturbances at the input to the plant). To this effect, we consider input (load) disturbances $d_u$ for which $G_d(s) = G(s)$ (see Figure 9). Hence, eq 21 gives the following controllability bound on the allowed plant gain at frequency $\omega_S$:

$$|G(j\omega_S)| < \frac{y_{max}}{|d_u(\omega_S)|} \tag{22}$$

This bound is independent of the controller and, thus, provides a fundamental controllability requirement. In most cases, $|G|$ is smaller at high frequency, so the bound is easier to satisfy if $\omega_S$ is increased. However, for stability reasons, the value of $\omega_S$ is limited, and a typical upper bound is $\omega_S \approx 0.5/\theta$, where $\theta$ denotes the "effective delay" around the feedback loop.[2]

Input disturbances are very common, but what is the expected value of $|d_u|$? This is difficult to answer, because input disturbances have many sources. For example, in many cases, the input is a valve that receives its power from a hydraulic system (e.g., the brakes of a car) or from pressured air (many process control applications). A change (disturbance) in the power system will then cause an input disturbance. The value of $|d_u|$ will vary, depending on the application. If it is assumed that the system has been scaled such that the largest expected input $u$ is of magnitude 1, then it seems reasonable that $|d_u|$ is at least 0.01, and that a typical value is 0.1 or larger.

**Steady-State Implications.** The condition described by eq 22 provides a bound on the plant gain at frequency $\omega_S$. The implications, in terms of the steady state, are discussed next by considering a first-order with delay plant,

$$G(s) = G_d(s) = \frac{k e^{-\theta s}}{\tau s + 1} \tag{23}$$

where $k = |G(0)|$ is the steady-state gain of the plant. The high-frequency asymptote is $|G(j\omega)| \approx k/\tau\omega = k'/\omega$, where $k' = k/\tau$ is the initial slope of the step response. With $\omega_S \approx 0.5/\theta$, eq 22 gives the controllability requirement:

$$\frac{k}{\tau} = k' < 0.5\,\frac{y_{max}}{\theta|d_u|} \tag{24}$$

Equation 24 may seem to indicate that a plant with a large steady-state gain $k$ is fundamentally difficult to control (see case 1 below). However, as discussed in case 2, this is not always true, because, according to eq 22, it is the gain at frequency $\omega_S$ that should be small and a process can have a large steady-state gain while having a small gain at high frequency.

**Case 1.** In some cases, a large steady-state gain $k$ implies a large gain at high frequencies, resulting in not being able to satisfy the controllability requirement in eq 21. A physical example is a pH neutralization process, as studied in chapter 5 in the work of Skogestad and Postlethwaite.[2] The component balance for the excess of acid $y$ gives the model $\tau_h s y(s) = 1/\epsilon\, u(s) - y(s)$. where $\tau_h$ is the residence time and $u$ the neutralization flow. This is of the form of eq 23, with $k = 1/\epsilon$ and $\tau = \tau_h$. The reason for the small value of $\epsilon$ is that the desired concentration in the tank ($y$) can be on the order of $10^6$ smaller than in the neutralization inflow. Because of the large high-frequency gain, this plant is not controllable, according to eq 22, and a design change is required, for example, where the neutralization is done in several steps (tanks) rather than in a single step.

**Case 2.** As an example of a case where a large steady-state gain does not imply control problems, consider a near-integrating process:

$$G(s) = \left(\frac{k'}{s + \epsilon}\right) e^{-\theta s} \tag{25}$$

This is of the form of eq 23, with $k = k'/\epsilon$ and $\tau = 1/\epsilon$. Thus, as $\epsilon \to 0$, the steady-state gain $G(0) = k'/\epsilon$ approaches infinity, but the high-frequency slope of the gain $k'$ remains finite, because it is independent of $\epsilon$, so eq 24 may not impose any controllability limitation. A physical example is a liquid level where $\epsilon$ represents the self-regulating effect. The mass balance may be written as $s\Delta V(s) = \Delta q_{in} - \Delta q_{out}$, where the linearized outflow is $\Delta q_{out} = k'\Delta Z(s) + \epsilon\Delta V(s)$ and $Z$ is the valve position. $\epsilon \to 0$ for the case when the outflow is only dependent weakly on $V$. With $y = \Delta V$, $u = \Delta Z$, and $d = \Delta q_{in}$, this results in a model of the form in eqs 25 and 23.

## 8. Discussion

We have derived expressions for the amplitude and period of oscillations that result with feedback control of a system with restricted input resolution (quantizer). Importantly, the amplitude and period were determined (under certain assumptions about the integral time) to be independent of the controller gain. However, note that the time before cycling actually starts may be considerably longer than the period $T$ of the oscillations, and that this start-up time is dependent on the controller gain. By detuning the controller (reducing the controller gain), generally a longer time is required for the oscillations to start.

This is confirmed by the simulations in Figure 3 in McAvoy and Braatz,[4] where a detuned controller gives no oscillations with a simulation time of 80 s. However, it is easily confirmed that oscillations do indeed develop if the simulation time is extended to 95 s or more.

In this paper, we have considered the effect of input (valve) inaccuracy and input load disturbances, with the following corresponding controllability requirements:

$$|G(j\omega_{L,180})| < \frac{\pi}{4}\left(\frac{a_{\max}}{q}\right) \qquad (17)$$

$$|G(j\omega_S)| < \frac{y_{\max}}{|d_u(\omega_S)|} \qquad (22)$$

Which condition is the more restrictive? There is no general answer, but let us first consider two reasons for why the latter (input disturbance) may be more restrictive. First, the input disturbance $|d_u|$ is normally larger than the quantization step $q$. Second, the bound for input load disturbance occurs at a lower frequency $\omega_S$, where the gain $|G(j\omega)|$ is generally larger than that at frequency $\omega_{L,180}$. Especially, assume that the magnitude of the first-order plus delay plant in the high-frequency range can be approximated by $|G(j\omega)| = k/\tau\omega$. Then, taking the typical values $\omega_S = 0.5/\theta$ and $\omega_{L,180} = 1.5/\theta|$, we get

$$\frac{|G(j\omega_S)|}{|G(j\omega_{L,180})|} \approx \frac{\omega_{L,180}}{\omega_S} \approx 3 \qquad (26)$$

This leads to the conclusion that the output deviation caused by an input disturbance is likely to be larger than the sustained output variations caused by restricted input resolution. On the other hand, we are less likely to accept sustained oscillations ($a_{\max}$) than short-time deviations ($y_{\max}$), so one could argue that $a_{\max}$ is usually smaller than $y_{\max}$ (a typical value may be $a_{\max} = 0.2y_{\max}$). In summary, it is not clear which is the more restrictive.

McAvoy and Braatz[4] have stated that, for control purposes, the magnitude of the steady-state process gain ($k = \bar{\sigma}(G(0))$) should not exceed 50. In this paper, we have derived controllability conditions, eqs 17 and 22, that limit the plant gain at frequencies $\omega_{L,180}$ and $\omega_S$, respectively. These conditions have some implications for the steady-state gain which, in special cases, may provide some justification for the rule-of-thumb of McAvoy and Braatz.[4] Specifically, the expression described by eq 18 for steady-state offset with P control gives $k \le (2|y - r|)/q$. For example, with $q = 0.02$ and $|y - r|_{\max} = a_{\max} = 0.2$, this requires $k < 20$. Thus, P control should only be used for plants with a small steady-state gain. Furthermore, eq 22 may be rewritten, as in eq 24, to get $k < 0.5y_{\max}(\tau/\theta|d_u|)$. If we select $|y_{\max}| = 1$, $|d_u| = 0.1$, and $\tau/\theta = 10$ (similar to that used in the simulation in McAvoy and Braatz[4]), we then derive a bound of $k < 50$. However, note that the bounds described by eqs 18 and 24 do not imply that large steady-state gains are always a problem for control. First, eq 24 is derived for a first-order with delay model where $k$ and $\tau$ are assumed independent, whereas they often are coupled (e.g., see eq 25). Second, eq 18 applies to P control and the implication is that integral action must be added for control of such processes.

In the Introduction, we referred to a case by Moore[3] that seems to prove that a large steady-state gain (i.e., large gain at zero frequency) gives large output variations (poor control) when we have restricted valve resolution. However, in practice, the system will not cycle at a low frequency, but rather at a higher frequency ($\omega_{L,180}$), where the process gain is smaller and the resulting output variables are therefore smaller. We may also introduce forced cycling or use valve position control to reduce the output variation further.

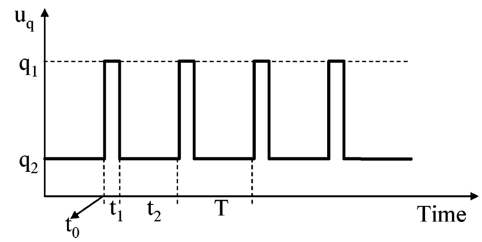


**Figure 10.** Input to be applied to the system in eq 15.

## 9. Conclusion

In this paper, controllability requirements are derived for two types of input errors, namely (i) restricted input resolution (e.g., that caused by valve inaccuracy) and (ii) input disturbances.

In regard to type i, limited input resolution with integral feedback control (no steady-state offset) causes limit cycle behavior (oscillations) (see Theorem 1). The magnitude of the oscillations can be reduced by pulse-modulating the input signal or using valve position control, but this assumes that frequent input movements are acceptable. The controllability requirement derived from an approximate describing function analysis, assuming no forced oscillations, is

$$|G(j\omega_{L,180})| < \frac{\pi}{4}\left(\frac{a_{\max}}{q}\right) \qquad (17)$$

where $L = GK$ and, typically, $\omega_{L,180} \approx 1.5/\theta$ (where $\theta$ is the effective delay in the loop). The variable $a_{\max}$ is the allowed magnitude for the resulting sustained output oscillations (limit cycles). This expression agrees well (within 27%) with an exact nonlinear analysis for a first-order plus delay process. With forced oscillations (pulse modulating the input signal), we can select the frequency $\omega$ to be much higher than the "natural" cycling frequency $\omega_{L,180}$, and the controllability limitations are generally less restrictive.

In regard to type ii, for input (load) disturbances of magnitude $|d_u|$, the controllability requirement is

$$|G(j\omega_S)| < \frac{y_{\max}}{|d_u(\omega_S)|} \qquad (22)$$

where $y_{\max}$ is the allowed magnitude of the resulting short-term output deviation and, typically, $\omega_S \approx 0.5/\theta$.

In summary, large gains at frequencies around the closed-loop bandwidth ($\omega_S$, $\omega_{L,180}$) may cause problems with feedback control. There is no controllability condition that involves the steady-state gain $k = |G(0)|$ only, so a large steady-state gain is not, by itself, a problem for feedback control.

## 10. Appendix. Proof of Theorem 1

Consider the first-order plus delay process in eq 12. Now, assume this process is excited by a periodic and persistent input (it is applied because $t > 0$) of the form given by Figure 10. It represents the signal generated from a relay without hysteresis in which $q_1$ and $q_2$ are the limit values, $t_1$ is the time interval where $u_q$ remains in $q_1$, and $T$ is the period of oscillation ($T = t_1 + t_2$). This signal can be represented in the Laplace domain as a series of steps delayed in time. Assume now, without a loss of generality, that $q_2 = 0$ and $q_1 = q$. The resulting transformed signal is given by

$$u_q(s) = \frac{q}{s}(1 - e^{-t_1 s} + e^{-Ts} - e^{-(t_1 + T)s} + e^{-2Ts} - e^{-(t_1 + 2T)s} + ...) \quad (27)$$

When this signal is applied to the process in eq 12, oscillations result in the output.

The set of maximum (or minimum) values of these oscillations are such that $t = \{t|t = t_1 + mT + \theta, \forall m \in \mathbb{N}\}$ and the minimum (or maximum) values are found in the set $t = \{t|t = mT + \theta, \forall m \in \mathbb{N}\}$.

The maximum (or minimum) at $\theta + T < t \leq \theta + t_1 + T$ is

$$y(s) = \left(\frac{k}{\tau s + 1}\right)e^{-\theta s}\left(\frac{q}{s}\right)(1 - e^{-t_1 s} + e^{-Ts}) \quad (28)$$

which, inverted to the time domain, gives

$$y(t) = kq(1 - e^{-(t-\theta-T)/\tau} + e^{-(t-\theta-t_1)/\tau} + e^{-(t-\theta)/\tau}) \quad (29)$$

Thus, the maximum (or minimum) is

$$y(t_1 + T + \theta) = kq(1 - e^{-t_1/t} + e^{-T_1/\tau} + e^{-(t_1+T)/\tau}) \quad (30)$$

Hence, the maximum (or minimum) amplitude, $y_{ext1}$, can be extended to

$$y_{ext1} = kq(1 - e^{-t_1/\tau} + e^{-T/\tau} - e^{-t_1+T/\tau} + e^{-2T/\tau} - ...) \quad (31)$$

which can be written as

$$y_{ext1} = kq[(1 - e^{-t_1/\tau})(1 + e^{-T/\tau} + e^{-2T/\tau} + e^{-3T/\tau} + ...)] \quad (32)$$

The infinite sum in eq 32 is given by

$$\lim_{n\to\infty} \sum_{j=0}^{n} (e^{-T/\tau})^j = \frac{1}{1 - e^{-T/\tau}} \quad (33)$$

where the fact that $(e^{-T/\tau})^n$ converges to zero as $n$ approaches infinity is used.

Accordingly,

$$y_{ext1} = kq\left(\frac{1 - e^{-t_1/\tau}}{1 - e^{-T/\tau}}\right) \quad (34)$$

The minimum (or maximum) at $\theta + t_1 + T < t \leq \theta + 2T$, $y_{ext2}$, is determined by following the same development process used to derive $y_{ext1}$, i.e.,

$$y_{ext2} = kq\left[\frac{e^{-T/\tau}(1 - e^{-t_1/\tau})}{1 - e^{-T/\tau}}\right] \quad (35)$$

The amplitude is calculated by $a = y_{ext1} - y_{ext2}$, or

$$a = kq\left(\frac{1 - e^{-t_1/\tau} + e^{-T/\tau} - e^{-(T-t_1)/\tau}}{1 - e^{-T/\tau}}\right) \quad (36)$$

The formula in eq 36 is dependent on $t_1$ and $T$, which must be determined.

From Figure 2,

$$u(s) = K(s)[r(s) - y(s)] \quad (37)$$

where $K(s)$ is given by eq 13, $r(s)$ is a step change in reference ($r(s) = r_0/s$), and $y(s) = K(s)G(s)u_q(s)$, where $G(s)$ is given by eq 12.

In the limit when $t \to \infty$, the quantizer behaves exactly as the relay depicted in Figure 10 and, assuming that $q_1$ and $q_2$ are arbitrary values, the first three terms of $u_q$ are

$$u_q(s) = \frac{q_2}{s} + \frac{q_1 - q_2}{s}(e^{-t_1 s} - e^{-(t_1+t_2)s}), \quad (38)$$

where the fact that $T = t_1 + t_2$ is used.

Consider a PI controller. Substituting eq 38 into eq 37 and inverting it to the time domain, the following equation for $u(t)$ in the interval $\theta \leq t < t_0 + \theta$ is observed:

$$u(t) = \frac{K_c}{\tau_I}\{r_0(t + \tau_I) - kq_2[(\tau_I - \tau)(1 - e^{-(t-\theta)/\tau}) + t - \theta]\} \quad (39)$$

For the interval $\theta + t_0 \leq t < t_0 + t_1 + \theta$, $u(t)$ is given by

$$u(t) = \frac{K_c}{\tau_I}\{r_0(t + \tau_I) - kq_2[(\tau_I - \tau)(1 - e^{-(t-\theta)/\tau}) + t - \theta] - k(q_1 - q_2)[(\tau_I - \tau)(1 - e^{-(t-t_1-\theta)/\tau}) + t - t_1 - \theta]\} \quad (40)$$

Similarly, for the interval $\theta + t_0 + t_1 \leq t < t_0 + t_1 + t_2 + \theta$,

$$u(t) = \frac{K_c}{\tau_I}\{r_0(t + \tau_I) - kq_2[(\tau_I - \tau)(1 - e^{-(t-\theta)/\tau}) + t - \theta] - k(q_1 - q_2)[(\tau_I - \tau)(1 - e^{-(t-t_1-\theta)/\tau}) + t - t_1 - \theta] + k(q_1 - q_2)[(\tau_I - \tau)(1 - e^{-(t-t_1-t_2-\theta)/\tau}) + t - t_1 - t_2 - \theta]\} \quad (41)$$

So far, no assumptions on the controller settings ($K_c$ and $\tau_I$) have been made. The expressions described by eqs 39−41 become drastically simplified if the integral time is selected as $\tau_I = \tau$, which is an appropriate setting for many plants.[9]

Furthermore, for a relay without hysteresis, its output ($u_q(t)$) changes because its input ($u(t)$) is equal to zero, and, because the quantizer behaves as a relay when $t \to \infty$, the following equations give relations for $t_1$ and $t_2$.

For $t = t_0$:

$$r_0(t_0 + \tau_I) = kq_2(t_0 - \theta) \quad (42)$$

For $t = t_0 + t_1$:

$$r_0(t_0 + t_1 + \tau_I) = kq_2(t_0 + t_1 - \theta) - k(q_1 - q_2)(t_0 - \theta) \quad (43)$$

For $t = t_0 + t_1 + t_2$:

$$r_0(t_0 + t_1 + t_2 + \tau_I) = kq_2(t_0 + t_1 + t_2 - \theta) - k(q_1 - q_2)(t_0 + t_2 - \theta) + k(q_1 - q_2)(t_0 - \theta) \quad (44)$$

Combining eqs 42−44, the following expressions give the period $T$ of the oscillations:

$$t_1 = \frac{k(q_1 - q_2)\theta}{kq_1 - r_0} \tag{45}$$

$$t_2 = \frac{k(q_1 - q_2)\theta}{r_0 - kq_2} \tag{46}$$

$$T = t_1 + t_2 \tag{47}$$

On average, the input must equal the steady-state value, $u_{ss} = y_{ss}/G(0) = r_0/k$ (where $k = G(0)$), and if this does not happen to exactly correspond to one of the quantizer level, the quantized input $u_q$ will cycle between the two neighboring quantizer levels ($q_1$ and $q_2$). Let $f$ and $(1 - f)$ denote the fraction of time spent at each level. Then, at steady state, $u_{ss} = r_0/k = fq_1 + (1 - f)q_2$ and, from this expression, $f$ is determined to be

$$f = \frac{r_0 - kq_2}{k(q_1 - q_2)} \tag{48}$$

From eq 48,

$$t_1 = \frac{\theta}{1 - f} \tag{49a}$$

$$T = \theta\left(\frac{1}{1 - f} + \frac{1}{f}\right) \tag{49b}$$

which completes the proof.

## Literature Cited


(1) Morari, M. Design of Resilient Processing Plants III—A General Framework for the Assessment of Dynamic Resilience. *Chem. Eng. Sci.* **1983**, *38*, 1881.

(2) Skogestad, S.; Postlethwaite, I. *Multivariable Feedback Control: Analysis and Design*, 2nd Edition; Wiley: Chichester, U.K., 2005.

(3) Moore, C. F. Selection of Controlled and Manipulated Variables. In *Practical Distillation Control*; Van Nostrand Reinhold: New York, 1992.

(4) McAvoy, T. J.; Braatz, R. D. Controllability of Process with Large Singular Values. *Ind. Eng. Chem. Res.* **2003**, *42*, 6155.

(5) Shoukat Choudhory, M. A. A.; Thornhill, N. F.; Sha, H. N. L. Modelling valve stiction. *Control Eng. Pract.* **2005**, *13*, 641.

(6) Slotine, J. E.; Li, W. *Applied Nonlinear Control*; Prentice Hall International Editions: Englewood Cliffs, NJ, 1991.

(7) Aström, K. J.; Hägglund, T. *Automatic Tuning of PID Controllers*; Instrument Society of America: Research Triangle Park, NC, 1988.

(8) Skogestad, S. Simple Analytic Rules for Model Reduction and PID Controller Tuning. *J. Process Control* **2003**, *13*, 291.

(9) Smith, C. L.; Corripio, A. B.; Martin, J., Jr. Controller Tuning from Simple Process Models. *Instrum. Technol.* **1975**, 39.

(10) Wang, Q.; Hang, C.; Zou, B. Low-Order Modelling from Relay Feedback. *Process Des. Control* **1997**, *36*, 375.

(11) Liptak, B. G. *Instrument Engineers' Handbook. Volume II: Process Control and Optimization*, 4th edition; CRC Press: Boca Raton, FL, 2006.



*Received for review* March 17, 2006
*Revised manuscript received* October 5, 2006
*Accepted* October 9, 2006

IE060323P