Comparison of PI Controller Tuning Methods

Michael W. Foley,*,† Navin R. Ramharack,† and Brian R. Copeland‡

Department of Chemical Engineering, University of the West Indies, St. Augustine, Trinidad, and Department of Electrical and Computer Engineering, University of the West Indies, St. Augustine, Trinidad

The development of model-based methods for tuning proportional−integral (PI) and proportional−integral−derivative (PID) controllers is a topic of renewed research interest. A number of techniques have appeared in the last five years aimed at improving upon the standard "λ-tuning" or direct synthesis (DS) approaches. This paper compares the Skogestad internal model control (SIMC), direct synthesis for disturbance rejection (DS-d), and Wang−Shao (WS) tuning algorithms with the IMC improved PI technique as implemented on first-order-plus-deadtime systems. The main objective was to assess the relative control effort and robustness of these proportional−integral controllers when tuned for the same level of performance. Recommendations were provided for selecting the most appropriate tuning method for a given application based upon the primary function of the feedback loop (servo versus regulatory, etc.) and the plant deadtime-to-time constant ratio.

1. Introduction

It is a remarkable fact that the majority of processes in the chemical industries can be satisfactorily controlled using proportional or proportional−integral (PI) feedback. Luyben[1,2] has listed default settings for flow, level, pressure, and temperature control loops which represent good initial values of the controller settings. In most instances, an experienced control engineer or instrument mechanic can fine-tune these values by trial-and-error and quickly obtain an acceptable, if not optimal, closed-loop response.

However, there exist many control loops for which this simple approach may not be sufficient, namely, those characterized by significant time delays, process nonlinearities, time-varying dynamics, and/or interaction with other regulators in a multivariate system. For these difficult problems, it is often worthwhile to generate the initial settings with a model-based tuning strategy. A comprehensive summary of model-based PI and PID (proportional−integral−derivative) tuning techniques was recently published by O'Dwyer.[3] Undoubtedly, the most well-known is the method of Ziegler and Nichols.[4] Though a standard component of undergraduate control courses in the electrical, mechanical and chemical engineering disciplines, the Ziegler−Nichols rules are rarely employed in the process industries. This is because they were designed to produce an underdamped (quarter-decay) response, which implies very aggressive controller settings.

By contrast, the "λ-tuning" or direct synthesis (DS) methods[5] are becoming increasingly popular with plant engineers. This may be attributed to the fact that they possess a single tuning parameter (λ) which has a predictable effect on closed-loop behavior. The user is able to tailor the response to attain a suitable tradeoff between high performance on one hand versus smooth valve adjustments and minimal sensitivity to modeling errors on the other. Furthermore, the λ-tuning method is computationally straightforward (requires no numerical iteration) and is based upon simple, low-order approximations of the process dynamics such as the first-order-plus-deadtime (FOPDT) model. It has long been recognized that FOPDT approximations can adequately explain the behavior of a wide range of processes, and many industrially proven techniques are now available for fitting these models to plant data.

The purpose of this paper is to present a practical assessment of the following PI tuning methods when applied to the control of FOPDT processes: (i) SIMC (Skogestad internal model control, Skogestad[6]), (ii) DS-d (direct synthesis for disturbance rejection, Chen and Seborg[7]), (iii) WS (Wang and Shao[8]), and (iv) IMC (internal model control) improved PI (Rivera et al.[9]).

Section 2 covers the mathematical preliminaries. The tuning formulas associated with each strategy are stated in Section 3 and analyzed by comparing their integral-time expressions. Section 4 documents a series of simulation experiments carried out to examine the control effort and robustness of each algorithm assuming that they have been tuned to provide equal performance when implemented on a digital controller. These results are summarized in Section 5, and guidelines are developed to assist the user in selecting the most appropriate tuning strategy for a particular application.

2. Definitions

Figure 1 is a schematic diagram of a feedback control loop. The symbols r, Y, and e represent the setpoint, controlled variable, and control error, respectively. The load disturbance d enters at the process input; the manipulated variable is denoted by U. It is assumed that the combined dynamics of the process, final control element, and sensor are well-approximated by the first-order-plus-deadtime transfer function:

$$G(s) = \frac{K\,e^{-Ds}}{\tau s + 1} \qquad (1)$$

* Corresponding author. Tel.: (868) 662-2002 ext. 3412. Fax: (868) 662-4414. E-mail: mfoley@uwi.tt.

† Department of Chemical Engineering.

‡ Department of Electrical and Computer Engineering.

10.1021/ie040258o CCC: $30.25
Published on Web 00/00/0000 **PAGE EST: 10**

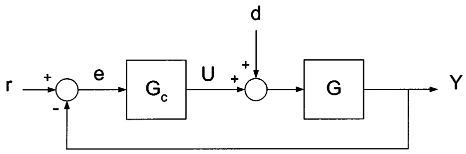

Figure 1. Feedback control loop.

The main task of the proportional−integral feedback controller

$$G_C(s) = K_C\left(1 + \frac{1}{\tau_I s}\right) \tag{2}$$

is to reject the effect of the load disturbance and maintain the controlled variable Y at setpoint, thereby keeping the magnitude of the control error as small as possible.

There are many accepted mathematical definitions of performance, robustness, and control effort. In this report, performance was quantified as the ratio of the variance of the control error σ_e^2 to $\sigma_{e,MV}^2$, the minimum error variance attainable with a PI structure:[10]

$$\text{performance (\%)} = \frac{\sigma_{e,MV}^{\ 2}}{\sigma_e^{\ 2}} \times 100 \tag{3}$$

Accordingly, the performance of a PI controller is bounded by 0 (unstable) and 100% (best achievable). Closed-loop variances were computed using a contour integration approach, analogous to that described by Horton et al.[11] for PI level controllers. When evaluating regulatory performance, the setpoint was held constant while the load disturbance varied as a random walk process. For the servo case, the load variable was kept at its initial steady-state value and the command input was modeled as a random walk. As explained by MacGregor et al.,[12] σ_e^2 may be interpreted as the mean square control error instead of the error variance if one is interested in the response to deterministic step changes in setpoint or load. The mean square control error is closely related to the integral of the squared error (ISE), a well-known performance metric for analog control systems.

Control effort was measured by reference to the variance of the valve adjustments exhibited by the minimum variance PI controller, $\sigma_{\Delta U,MV}^2$:

$$\text{control effort (\%)} = \frac{\sigma_{\Delta U}^{\ 2}}{\sigma_{\Delta U,MV}^{\ 2}} \times 100 \tag{4}$$

Notice that it is entirely possible for the control effort of a given loop to exceed 100%. This would simply indicate that the variance of its control moves is even greater than that required by the PI controller, which minimizes the variance of the error. A small control effort is obviously desirable since it implies smooth increments in the final control element (e.g., valve). Larger values suggest increased wear-and-tear on the valve as well as a greater likelihood of encountering hard manipulated variable constraints.

The robust stability of each controller was assessed by simultaneously increasing the steady-state gain (K), time constant (τ), and deadtime (D) of the *actual* plant until one or more poles of the discrete-time characteristic equation were found to lie outside the unit circle. Let K_M, τ_M, and D_M represent the nominal or estimated values of K, τ, and D. Robustness was defined as the largest deviation, δ_{max}, for which closed-loop stability was maintained, where

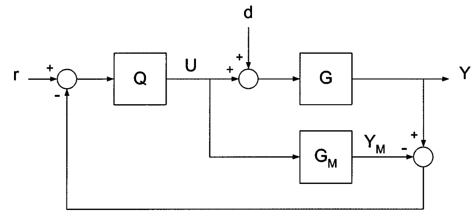

Figure 2. IMC block diagram.

$$\delta\ (\%) \equiv \frac{K - K_M}{K_M} \times 100 = \frac{\tau - \tau_M}{\tau_M} \times 100 = \frac{D - D_M}{D_M} \times 100 \tag{5}$$

The idea of simulating modeling errors by perturbing the FOPDT parameters in this manner was taken from Marlin.[13] The effect is analogous to decreasing throughput in a continuous stirred-tank heater. For instance, it is apparent from a dynamic energy balance that, if the feed flowrate were to decrease by half, the gain, time constant, and time delay of the process transfer function would double, resulting in a correlated error $\delta = 100\%$.

2.1. IMC Analysis of PI Control. The internal model control (IMC) structure is illustrated in Figure 2. This block diagram was employed by Garcia and Morari[14] as a means of comparing several well-known model-based control strategies. The symbol G_M represents the model of the true process G, which is typically obtained from the process reaction curve (see Section 7.2 of ref 5). A given model-based algorithm may be analyzed according to how it constructs the IMC controller Q. The feedback signal $Y - Y_M$ is the sum of the disturbance (Gd) and model mismatch ($[G - G_M]U$) effects. A satisfactory tradeoff must therefore be established between performance and robustness in the selection of Q.

When the plant model is perfect, we find by comparison of Figures 1 and 2 that

$$Q(s) = \frac{G_C(s)}{1 + G_C(s)G(s)} = \frac{(\tau_I s + 1)/K}{(\tau_F + D)s + \dfrac{(\tau_I s + 1)}{(\tau s + 1)}\, e^{-Ds}} \tag{6}$$

where

$$\tau_F \equiv \frac{\tau_I}{K_C K} - D \tag{7}$$

Tuning a PI controller then amounts to the specification of τ_I and τ_F. An important decision in the design of any feedback compensator is whether to cancel the dominant pole(s) of the open-loop process with the controller zero(s). The implications of doing so were discussed by Åström and Hagglund[15] and Middleton and Graebe.[16] If the integral time τ_I is equated with τ as in conventional direct synthesis, the IMC form of the PI controller reduces to

$$Q(s) = \frac{(\tau_I s + 1)/K}{(\tau_F + D)s + e^{-Ds}} \approx \frac{(\tau_I s + 1)/K}{\tau_F s + 1} \tag{8}$$

when the delay term e^{-Ds} is replaced with the first-order Taylor series $1 - Ds$. The numerator of $Q(s)$ is an "approximate inverse" of the plant transfer function $G(s)$. The form of the denominator suggests that τ_F ought to be regarded in much the same way as the "IMC filter constant", at least for designs which cancel or nearly cancel the plant pole. The relationship between the IMC filter constant and the sensitivity of a model-based controller to changes in the process dynamics is well-established.[9,17]

3. Model-Based PI Tuning Methods

This section presents, in order of increasing computational complexity, the tuning formulas obtained when the four methods are applied to first-order-plus-dead-time plants. It was assumed that the process steady-state gain $K > 0$, which would imply that the correct controller action is increase−decrease or reverse. The tuning rules are easily adapted for negative-gain plants by replacing K with its absolute value and configuring a direct-acting controller.

3.1. IMC Improved PI (Rivera et al.[9]). Rather than approximating the time delay as a first-order Taylor series, Rivera and co-workers omitted the factor e^{-Ds} from eq 1 and attempted to account for its effect on the process dynamics by inflating the time constant to $\tau + D/2$. The controller settings

$$K_C = \frac{\tau + (D/2)}{K\epsilon} \tag{9}$$

$$\tau_I = \tau + (D/2) \tag{10}$$

were then shown to produce a FOPDT closed-loop response to step changes in setpoint. The user-specified parameter ϵ can be interpreted as the desired closed-loop time constant. Rivera et al. suggested that ϵ be chosen greater than $\max(0.1\tau, 1.7D)$. Notice that, for this algorithm, the approximate IMC filter constant $\tau_F = \tau_I/(K_C K) - D = \epsilon - D$. Decreasing ϵ generally improves performance at the cost of greater control effort and reduced robustness.

3.2. SIMC (Skogestad[6]). In the SIMC approach, the proportional gain is computed as

$$K_C = \frac{\tau}{K(\lambda + D)} \tag{11}$$

and the integral time as the lesser of τ_I^{SIMC1} and τ_I^{SIMC2}, where

$$\tau_I^{SIMC1} = \tau \tag{12}$$

$$\tau_I^{SIMC2} = 4(\lambda + D) \tag{13}$$

λ is a positive constant selected by the user. When $\tau_I = \tau_I^{SIMC1}$, SIMC coincides with the direct synthesis technique in which $\lambda = \tau_F$ plays the role of the approximate closed-loop time constant. A PI controller configured in this way can be tuned for excellent servo behavior, but its regulatory response may be unacceptably sluggish for lag-dominant processes, i.e., when the deadtime-to-time constant ratio $D/\tau \ll 1$. When high-performance regulatory control of such a process is required, SIMC offers the flexibility of choosing $\tau_I = \tau_I^{SIMC2} < \tau$ in order to avoid the pole-zero cancellation. The particular form of eq 13 was designed to produce a critically damped response for a pure integrator, an extreme example of a lag-dominant process.

3.3. DS-d (Chen and Seborg[7]). The primary task of most chemical process control loops is regulation, not setpoint tracking. This motivated Chen and Seborg to reformulate the direct synthesis equations to generate a critically damped regulatory closed-loop transfer function with time constant τ_c. The DS-d (direct synthesis for disturbance rejection) settings are

$$K_C = \frac{\tau^2 + \tau D - (\tau_c - \tau)^2}{K(\tau_c + D)^2} \tag{14}$$

$$\tau_I = \frac{\tau^2 + \tau D - (\tau_c - \tau)^2}{\tau + D} \tag{15}$$

As is the case with ϵ and λ in the IMC and SIMC methods, an increase in τ_c yields a slower closed-loop response (lower performance). However, there is a limit to the extent to which the DS-d controller can be detuned. One must restrict

$$\tau_c < \tau + \sqrt{\tau^2 + \tau D} \tag{16}$$

in order that K_C and τ_I remain positive.

3.4. WS (Wang and Shao[8]). The Wang and Shao (WS) approach was based upon a frequency-domain model of the process dynamics. The mathematical objective was to fix the maximum distance of the Nyquist curve $G(j\omega)G_C(j\omega)$ from the imaginary axis to a user-defined value, $1/\alpha$. This provides the designer with a direct handle on robustness because it guarantees that the gain margin GM > α and the phase margin PM > $\cos^{-1}(1/\alpha)$; the authors recommended that α be chosen in the range (1.5, 2.5). Huang[18] employed this controller as a performance benchmark for the assessment of PI(D) control schemes.

The WS tuning rules can be written for a FOPDT process as

$$K_C = \frac{(1 + 2\tau^2{\omega_{90}}^2)\sqrt{1 + \tau^2{\omega_{90}}^2}}{\alpha K\omega_{90}[\tau + D(1 + \tau^2{\omega_{90}}^2)]} \tag{17}$$

$$\tau_I = \frac{1 + 2\tau^2{\omega_{90}}^2}{{\omega_{90}}^2[\tau + D(1 + \tau^2{\omega_{90}}^2)]} \tag{18}$$

The symbol ω_{90} denotes the frequency ω at which the open-loop phase angle $\phi(\omega) = -\pi/2$, where

$$\phi(\omega) = -\tan^{-1}(\tau\omega) - D\omega \tag{19}$$

ω_{90} must be determined using an iterative numerical procedure (e.g., Newton's method) or by means of a commercially available root-finding routine such as the "Goal Seek" and "Solver" functions in Microsoft Excel. The initial guess is not of critical importance, since the phase angle (eq 19) decreases monotonically with frequency; $\pi/(2(\tau + D))$ is a suitable starting value.

3.5. Discussion. It is interesting to note from eqs 10, 12, and 18 that the integral times associated with the IMC, SIMC1, and WS techniques are unaffected by their respective tuning parameters. (The acronym "SIMC1" will be used to refer to the SIMC method when τ_I is obtained from eq 12, and "SIMC2" will be used when τ_I is set via eq 13.) It is demonstrated in the Appendix that the WS integral time (eq 18) lies in the interval (τ, $\tau + (4/\pi^2)D$). The DS-d integral time depends on the choice of τ_c, but it is clearly $\leq\tau$. Hence,

$$\tau_I^{DS-d} \leq \tau_I^{SIMC1} < \tau_I^{WS} < \tau_I^{IMC} \tag{20}$$

The SIMC2 algorithm is intended for high-performance control of lag-dominant processes, i.e., when $4(\lambda + D) < \tau$. Thus τ_I^{SIMC2} is also bounded above by the open-loop time constant τ.

When the controllers are tuned for the same level of performance, it is expected that the controller with the largest reset time τ_I would compensate by employing the strongest proportional action, i.e.,

$$K_C^{DS-d} \leq K_C^{SIMC1} < K_C^{WS} < K_C^{IMC} \tag{21}$$

One would further anticipate that an IMC improved PI controller would prove more noise-sensitive than the others because its high-frequency gain (K_C) is largest. In terms of the time-domain dynamics, a PI controller tuned using this method will exhibit the strongest initial response or "proportional kick" to step changes in setpoint or load. The remaining error will be integrated to zero most slowly, because this strategy generally assigns least penalty to low-frequency deviations.

As pointed out by Wang and Cluett,[19] the transient response of a first-order-plus-deadtime system operating under model-based PI control depends only on the ratio D/τ and the user-specified tuning parameter. This can be verified by replacing the Laplace transform variable s with s/τ in eqs 1, 2, and 9−19, which has the effect of normalizing the time scale with respect to τ. Zhuang and Atherton[20] developed tuning correlations for PI controllers by minimizing various integral performance criteria. For step setpoint changes, the ISE-optimal integral time is

$$\tau_I = \frac{\tau}{0.690 - 0.155(D/\tau)} \tag{22}$$

when $0.1 < D/\tau < 1$. For step changes in load,

$$\tau_I = 1.87\tau\left(\frac{D}{\tau}\right)^{0.586} \tag{23}$$

Equation 22 implies that $\tau_I > \tau$ for all $D/\tau \in (0.1, 1)$. On the other hand, eq 23 yields $\tau_I > \tau$ for $D/\tau > 0.344$. These results lead one to predict that the IMC improved PI and WS schemes will be capable of delivering higher performance than that of SIMC and DS-d in servo applications and better regulatory control of delay-dominant processes. (The correlations given by Zhuang and Atherton for the range $D/\tau \in (1.1, 2)$ further support these conclusions.) Chen and Seborg[7] observed that the servo performance of the SIMC and DS-d methods can be significantly improved through the use of a two-degrees-of-freedom control structure. However, this functionality is rarely implemented in industrial control systems (Section 3.8 of ref 15) and was not included in the simulation experiments described below.

4. Simulation Results

The main purpose of this section is to compare the DS-d, SIMC, WS, and IMC improved PI tuning methods on the bases of control effort, robustness, and achievable performance. The vast majority of chemical process control loops are now implemented on digital hardware; hence, the simulations were carried out in discrete-time using the velocity form of the PI controller and a modified z-transform representation of the plant dynamics.[21] This approach also simplified the variance calculations, because the Padé approximation of the process deadtime was not required. To compensate for the sampling delay when computing the controller parameters, the deadtime was raised to $D + T/2$ in eqs 9−19 (Franklin et al.[22]). According to Marlin,[13] the closed-loop response of a digital PI controller will closely resemble that of an identically tuned analog system when the sampling/control interval T is chosen $<0.05(D + \tau)$. In the simulation studies which follow, T was specified as $0.03(D + \tau)$. Trends obtained with faster sampling rates were virtually indistinguishable from those presented here.

4.1. Servo Control. Figure 3 illustrates the behavior of the four algorithms when applied to servo control of the FOPDT process.

$$G(s) = \frac{e^{-s}}{10s + 1} \tag{24}$$

The control effort (variance of changes in the manipulated variable) generally increases (Figure 3a) and robustness decreases (Figure 3b) as performance improves. Note that, to facilitate the comparison, the tuning guidelines reported in Section 3 for the IMC and WS algorithms were not utilized. Instead, the adjustable parameter associated with each strategy was reduced until the best achievable performance was attained, i.e., such that further decreases would cause performance to degrade. The diamond marker on the SIMC curve divides the SIMC1 and SIMC2 algorithms; points to the right were generated using the SIMC2 integral time (eq 13).

It is apparent from eqs 10, 12, and 20 that, for $D/\tau \ll 1$,

$$\tau = \tau_I^{SIMC1} \approx \tau_I^{WS} \approx \tau_I^{IMC}$$

which explains why these curves almost coincide in Figure 3. The WS and IMC improved PI controllers can be configured for the highest levels of servo performance: 99.0% for WS and 99.2% for IMC. Hence, if required, the IMC scheme could be tuned to track setpoint changes with an error variance of only 0.8% greater than that which would be exhibited by an optimal PI controller minimizing σ_e^2.

Differences between the four tuning methods are more easily discerned as the deadtime-to-time constant ratio increases. Figure 4 displays the results obtained for the delay-dominant process.

$$G(s) = \frac{e^{-20s}}{10s + 1} \tag{25}$$

An unexpected tradeoff evidently arises in the servo control of FOPDT processes. The tuning method with the greatest robustness, namely, IMC improved PI, also exerts the greatest control effort. Conversely, the DS-d controller would implement the smoothest valve adjustments but prove most sensitive to changes in the plant dynamics. This result is counterintuitive, because one

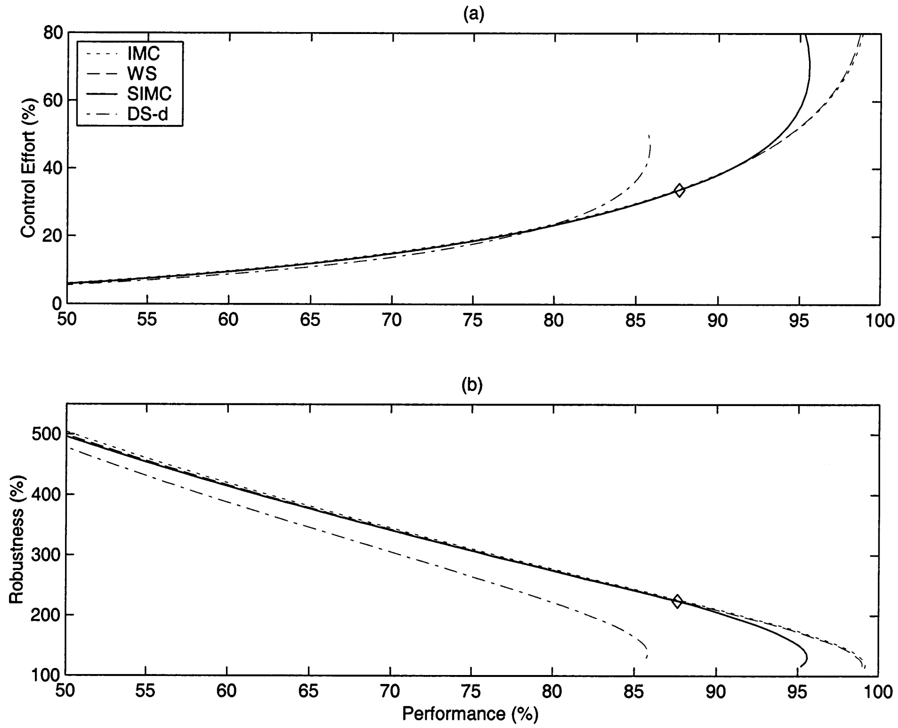

Figure 3. Servo control of lag-dominant process ($D/\tau = 0.1$).

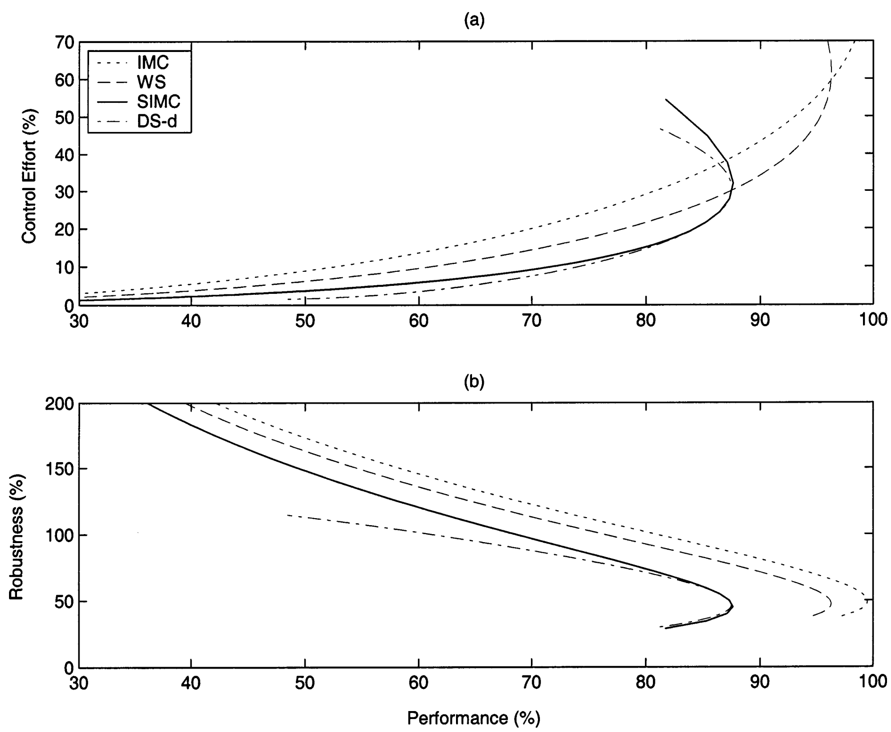

Figure 4. Servo control of delay-dominant process ($D/\tau = 2$).

would naturally assume that, if a controller is moving the manipulated variable in a more conservative manner than it would in an alternate scheme tuned for the same performance, it would place less stringent requirements on model accuracy. The Wang and Shao[8] tuning rules arguably yield the best compromise between these competing objectives.

As previously noted by Chen and Seborg,[7] the SIMC and DS-d methods give similar responses when designed for high-performance control of delay-dominant systems. This can be explained by recalling from Zhuang and Atherton[20] that the ISE-optimal integral time is $>\tau$ when $D/\tau > 0.344$. The best performance attainable with the DS-d algorithm occurs when $\tau_c = \tau$, since then τ_I^{DS-d} assumes its maximum value of $\tau = \tau_I^{SIMC}$. The DS-d curve truncates when the performance falls to 48.5% because further increases in τ_c would cause K_C and τ_I to become negative.

4.2. Regulatory Control. Figures 5 and 6 show that minimizing control effort and maximizing robustness are not conflicting goals in the regulatory case. Rather, for low-to-medium performance control of lag-dominant

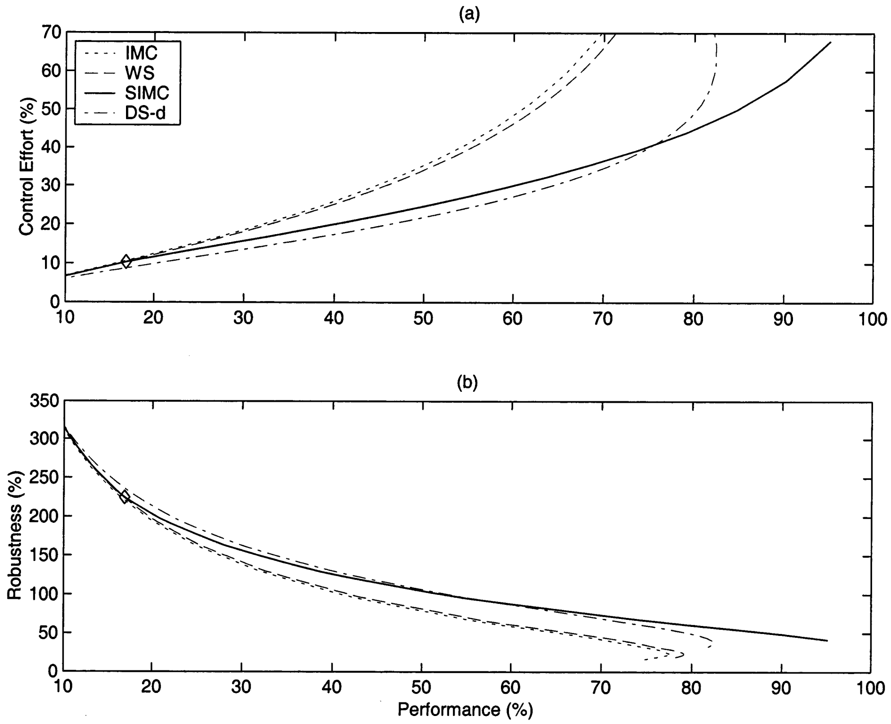

Figure 5. Regulatory control of lag-dominant process ($D/\tau = 0.1$).

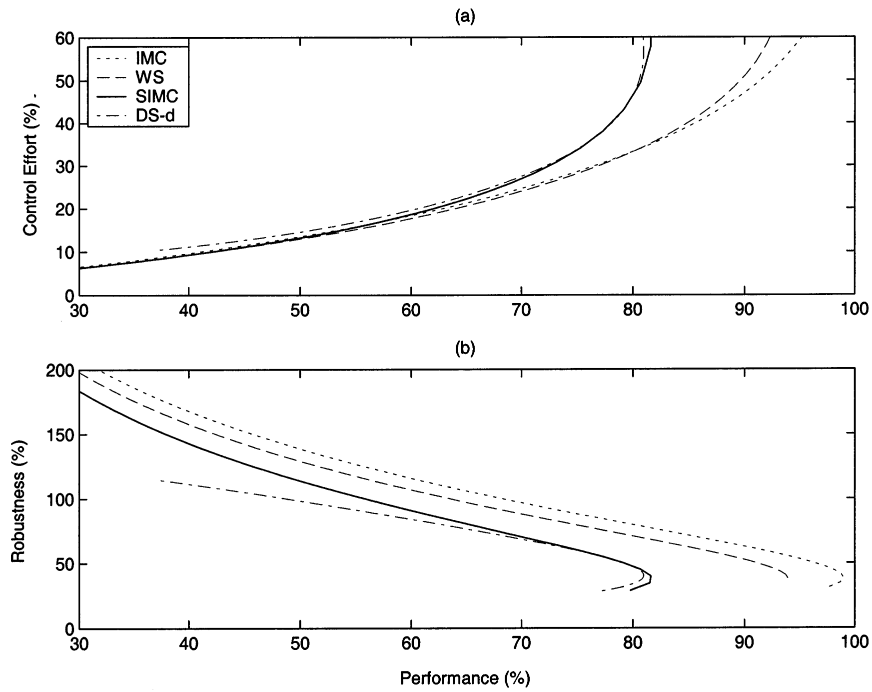

Figure 6. Regulatory control of delay-dominant process ($D/\tau = 2$).

processes (Figure 5), DS-d is an improvement on the other algorithms in terms of both of these criteria. SIMC is the appropriate choice when near-minimum-variance regulation is necessary.

Middleton and Graebe[16] concluded that the decision to cancel—rather than shift—slow, stable open-loop poles involves a tradeoff between input disturbance rejection and robustness. They proved that, for any shifting design, there exists an "extreme frequency equivalent" canceling design with a globally superior stability robustness margin. No such tradeoff is evident in Figure 5, where the shifting designs (SIMC2 and DS-d) exhibit better robustness than the canceling schemes (IMC and WS). This may be attributed to the fact that the loops have been tuned for equal control error variance instead of extreme frequency equivalence. In the context of PI control of a FOPDT plant, the latter concept applies to controllers possessing equal proportional gain, which was not the basis of this comparison.

From Figure 6, one arrives at a different recommendation for regulation of delay-dominant systems. The IMC improved PI approach clearly yields the best robustness and achievable performance. Its control effort is the least among the four methods when tuned for high performance and is very nearly the least at lower performance specifications.

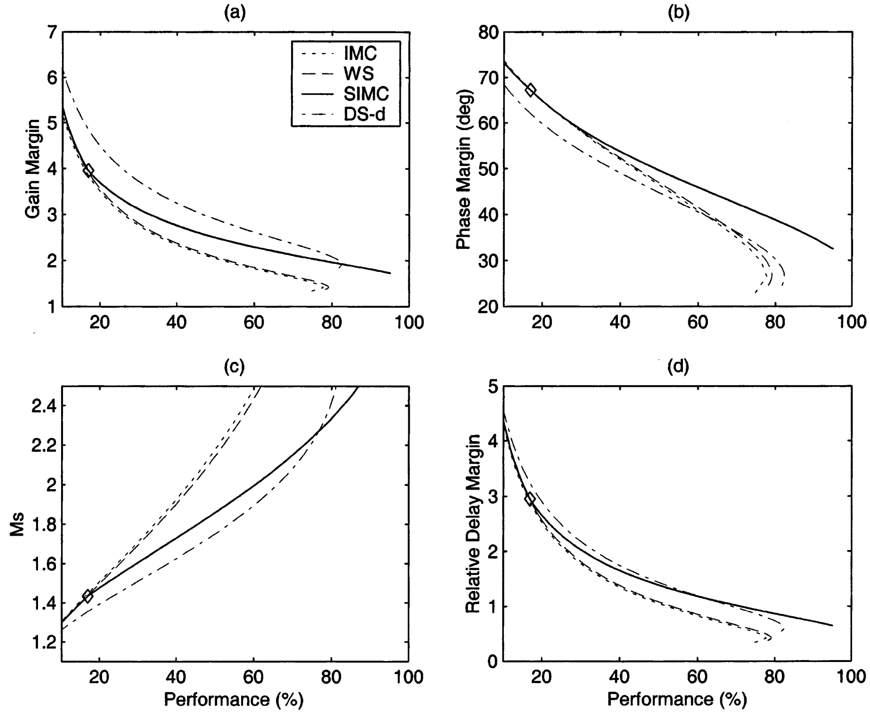

Figure 7. Robustness measures for PI regulation of lag-dominant process.

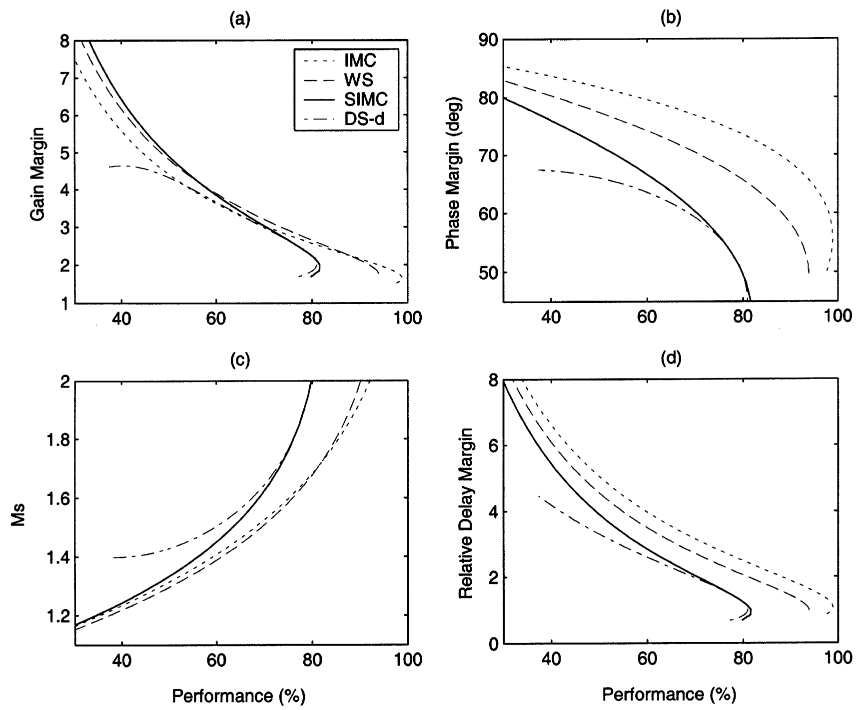

Figure 8. Robustness measures for PI regulation of delay-dominant process.

Parts a–d of Figure 7 trend four widely accepted measures of controller robustness: gain margin, phase margin, M_s, and relative delay margin (RDM). The variable M_s refers to the maximum magnitude (or infinity-norm) of the sensitivity function, i.e.,

$$M_s = ||S(j\omega)||_\infty = \sup_{\omega} \frac{1}{|1 + G_C(j\omega)G(j\omega)|} \tag{26}$$

The relative delay margin (Wang and Cluett[19]) is given by

$$\mathrm{RDM} \equiv \frac{D_{\max} - D_M}{D_M} = \frac{\pi \mathrm{PM}}{180 D_M \omega_{cg}} \tag{27}$$

where ω_{cg} is the gain crossover frequency and the phase margin, PM, is expressed in degrees. $D_{\max}$ represents the maximum deadtime for which the feedback loop remains stable, assuming that the other system parameters (e.g., K and τ) stay at their nominal values. For the SIMC1 method, it can be shown that RDM =

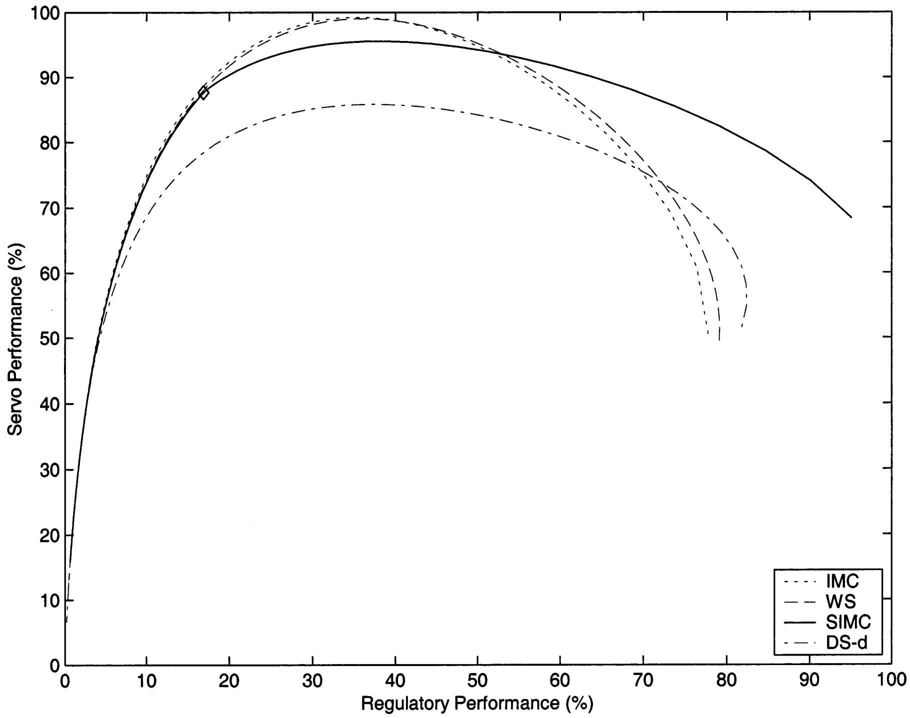

Figure 9. Servo-regulatory performance for lag-dominant process ($D/\tau = 0.1$).

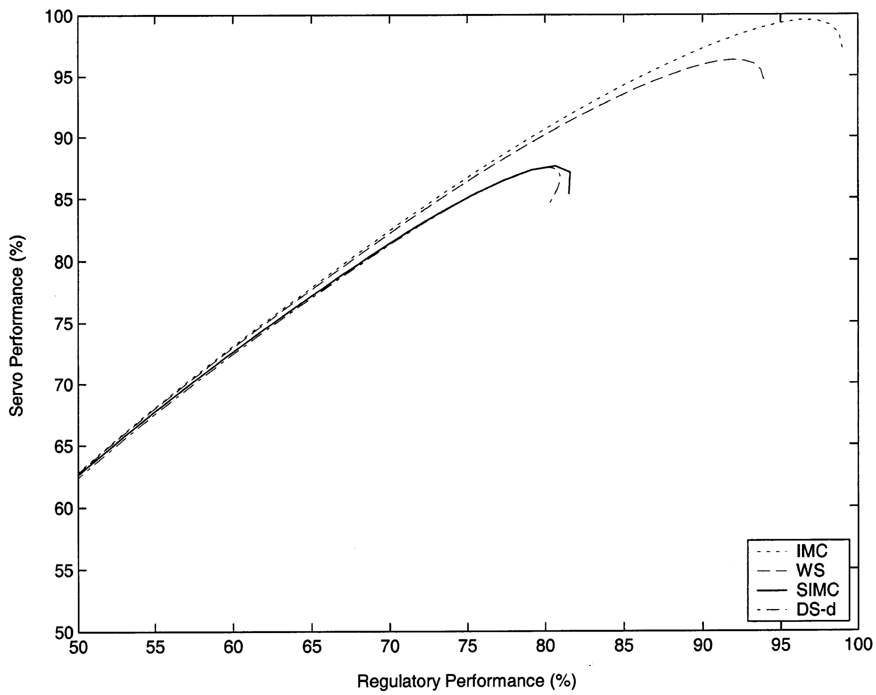

Figure 10. Servo-regulatory performance for delay-dominant process ($D/\tau = 2$).

$\pi\tau_F/(2D_M) + \pi/2 - 1$, which again highlights the connection between τ_F and controller robustness.

Figures 5b and 7d demonstrate that the relative delay margin correlates remarkably well with stability robustness. On the other hand, M_s and control effort vary with performance in virtually identical fashion (see Figures 5a and 7c). This is not entirely surprising, since for unit step disturbances we may derive the following from Figure 1:

$$sU(s) = -\frac{G_C(s)G(s)}{1 + G_C(s)G(s)} = S(s) - 1 \qquad (28)$$

That is, the rate of change of the manipulated variable is completely determined by the sensitivity function. This observation is further justified by the fact that the RDM contains gain and phase information, which are both pertinent to the issue of robustness, whereas M_s (and the variance of the control moves) depends only on the magnitude of $S(s)$. The exclusive use of M_s as a robustness measure has been previously questioned by Kristiansson and Lennartson.[23]

The gain margin is a reliable indication of the robustness of a feedback system to low-frequency modeling errors. However, this is not normally of primary concern when fitting FOPDT models to step response data collected from high-order chemical processes. The shape of the gain margin curves in Figure 7a is similar to the stability limits of Figure 5b, but gain margin does not correlate particularly well with robustness for delay-dominant systems (cf. Figures 8a and 6b). The phase margin plot of Figure 7b also provides less accurate predictions of comparative robustness than does the relative delay margin; it would lead one to rank the throughput robustness of the tuning methods in the wrong order. Similar trends were obtained for the servo case (results not shown), though the correspondence between M_s and control effort became less obvious as the process deadtime was increased.

4.3. Servo-Regulatory Performance. The following recommendations are made by way of summarizing the foregoing analysis: (i) servo control = WS; (ii) regulatory control ($D/\tau < 0.35$) = DS-d; and (iii) regulatory control ($D/\tau > 0.35$) = IMC improved PI. It remains to consider situations where the controller is required to simultaneously provide close setpoint tracking and good disturbance rejection. Clearly, none of the four tuning procedures provides superior results for every application. Figures 9 and 10 were constructed in an effort to identify the strategy which best facilitates the compromise between servo and regulatory performance.

Figure 9 implies that, for lag-dominant plants, this design objective is most closely realized using the SIMC tuning rules. The upper-right portion of the plot reveals that the SIMC controller can be tuned to provide good servo and regulatory response with the same settings. Since the time delay in secondary control loops is normally small, it is suggested that slave loops be tuned using the SIMC technique when the trial-and-error approach fails. From Figure 10, it was concluded that IMC improved PI is the better method for servo-regulatory control of delay-dominant processes. Notice that the performance curves have shifted toward the line $y = x$, indicating greater consistency between servo and regulatory responses for plants with a large dead-time-to-time constant ratio.

5. Conclusions

This paper has outlined a new methodology for the analysis of model-based PI tuning strategies. The IMC improved PI (Rivera et al.[9]), WS (Wang and Shao[8]), SIMC (Skogestad[6]), and DS-d (Chen and Seborg[7]) tuning rules were presented for first-order-plus-deadtime processes and compared with respect to robustness, control effort, and achievable performance. Each of these algorithms is reliable and easy to implement. The computational requirements of the WS technique are a little greater, since iteration is involved in the calculation of ω_{90}.

The correct choice of tuning method depends on the process control objective as well as the plant deadtime-to-time constant ratio, D/τ. The WS scheme was recommended when the response to setpoint changes is of greatest concern. For purely regulatory applications, DS-d gave the best overall results for lag-dominant processes ($D/\tau < 0.35$). IMC improved PI is better-suited for the regulation of delay-dominant systems ($D/\tau > 0.35$). When setpoint-following and disturbance rejection are of roughly equal importance (as for slave loops in a cascade structure), SIMC should be used if D/τ is small and IMC improved PI should be used if D/τ is large.

A secondary objective of this research was to evaluate the utility of the following robustness metrics: gain margin, phase margin, M_s (peak magnitude of the sensitivity function), and relative delay margin. Robustness was defined by the maximum joint perturbation in gain, time constant, and deadtime of the true process for which the digital control loop remained stable. The relative delay margin was found to provide the most accurate prediction of the comparative robustness of PI controllers tuned for the same level of performance. M_s is a better reflection of control effort than robustness, at least in regulatory applications.

Future work will focus upon the analysis of PID tuning rules. The implications of these results upon the selection of appropriate benchmarks for the assessment of PI(D) control loops should also be considered.

Acknowledgment

The authors would like to thank the staff of the U.W.I. Engineering Library for their assistance with this research. The useful suggestions of Ms. Rhonda Julien are also gratefully acknowledged.

Appendix: Derivation of Bounds for the Wang and Shao[8] Integral Time

Since $\tan(x) > x$ in the open interval $x \in (0, \pi/2)$, we have

$$\tan(D\omega_{90}) > D\omega_{90} \tag{A.1}$$

for $\omega_{90} \in (0, \pi/(2D))$. The symbol ω_{90} denotes the frequency at which the phase lag of the open-loop process equals $-90°$, i.e.,

$$\begin{aligned} &\phi(\omega_{90}) = -\tan^{-1}(\tau\omega_{90}) - D\omega_{90} = -\frac{\pi}{2} \\ &\Rightarrow \tan\left(\frac{\pi}{2} - D\omega_{90}\right) = \tau\omega_{90} \Rightarrow \tan(D\omega_{90}) = \frac{1}{\tau\omega_{90}} \end{aligned} \tag{A.2}$$

Substitution of eq A.2 into eq A.1 gives

$$\begin{aligned} &{\omega_{90}}^2\tau D < 1 \Rightarrow 1 + \tau^2{\omega_{90}}^2 > {\omega_{90}}^2\tau D(1 + \tau^2{\omega_{90}}^2) \\ &\Rightarrow 1 + 2\tau^2{\omega_{90}}^2 > {\omega_{90}}^2\tau^2 + {\omega_{90}}^2\tau D(1 + \tau^2{\omega_{90}}^2) \\ &\Rightarrow \tau_I = \frac{1 + 2\tau^2{\omega_{90}}^2}{{\omega_{90}}^2[\tau + D(1 + \tau^2{\omega_{90}}^2)]} > \tau \end{aligned} \tag{A.3}$$

Hence, the integral time is bounded below by the process time constant for all $D, \tau > 0$.

To derive an upper bound for τ_I, we first observe that the function

$$f(x) \equiv \frac{x}{\tan(x)} + \frac{4x^2}{\pi^2} - 1 \tag{A.4}$$

is positive in $(0, \pi/2)$. Thus, for $\omega_{90} \in (0, \pi/(2D))$,

$$\frac{D\omega_{90}}{\tan(D\omega_{90})} + \frac{4D^2{\omega_{90}}^2}{\pi^2} > 1$$

$$\Rightarrow D\tau{\omega_{90}}^2 + \frac{4D^2{\omega_{90}}^2}{\pi^2} > 1, \quad \text{using eq A.2}$$

$$\Rightarrow 1 + \tau^2{\omega_{90}}^2 < \left(D\tau + \frac{4D^2}{\pi^2}\right){\omega_{90}}^2(1 + \tau^2{\omega_{90}}^2)$$

$$\Rightarrow 1 + \tau^2{\omega_{90}}^2 < D\tau{\omega_{90}}^2(1 + \tau^2{\omega_{90}}^2) + \frac{4D^2}{\pi^2}{\omega_{90}}^2(1 + \tau^2{\omega_{90}}^2) + \frac{4D{\omega_{90}}^2}{\pi^2}\tau$$

$$\Rightarrow 1 + 2\tau^2{\omega_{90}}^2 < \tau{\omega_{90}}^2[\tau + D(1 + \tau^2{\omega_{90}}^2)] + \frac{4D}{\pi^2}{\omega_{90}}^2[\tau + D(1 + \tau^2{\omega_{90}}^2)]$$

$$\Rightarrow \tau_I = \frac{1 + 2\tau^2{\omega_{90}}^2}{{\omega_{90}}^2[\tau + D(1 + \tau^2{\omega_{90}}^2)]} < \tau + \frac{4D}{\pi^2} \qquad \text{(A.5)}$$

Nomenclature

d = load disturbance
D = process deadtime
DS-d = direct synthesis for disturbance rejection
FOPDT = first-order-plus-deadtime
GM = gain margin
$G(s)$ = Process transfer function
$G_C(s)$ = controller transfer function
IMC = internal model control
ISE = integral of the squared error
K = process steady-state gain
K_C = proportional gain
M_s = peak magnitude of sensitivity function
PI = proportional−integral
PID = proportional−integral−derivative
PM = phase margin
Q = IMC form of feedback controller
r = setpoint
RDM = relative delay margin
s = Laplace transform variable
sup = supremum
$S(s)$ = sensitivity function
SIMC = Skogestad IMC
SIMC1 = SIMC using integral expression (eq 12)
SIMC2 = SIMC using integral expression (eq 13)
U = manipulated variable
WS = Wang−Shao
Y = controlled variable

Greek Symbols

α = tuning parameter in WS strategy
δ = correlated parameter estimation error
ϵ = tuning parameter in IMC improved PI strategy
λ = tuning parameter in SIMC strategy
${\sigma_e}^2$ = control error variance
${\sigma_{e,MV}}^2$ = minimum error variance achievable with PI control
${\sigma_{\Delta U}}^2$ = variance of adjustments in manipulated variable
${\sigma_{\Delta U,MV}}^2$ = variance of control moves exhibited by minimum variance PI controller
τ = process time constant
τ_c = tuning parameter in DS-d strategy
τ_F = approximate IMC filter time constant
τ_I = integral time
ω = angular frequency
ω_{cg} = gain crossover frequency
ω_{90} = frequency at which plant phase angle equals −90°

Subscripts

max = maximum value for which closed loop remains stable
M = nominal value

Literature Cited

(1) Luyben, W. L. *Process Modeling, Simulation and Control for Chemical Engineers*, 2nd ed.; McGraw-Hill: New York, 1990.
(2) Luyben, W. L. *Plantwide Dynamic Simulators in Chemical Processing and Control*; Marcel-Dekker: New York, 2002.
(3) O'Dwyer, A. *Handbook of PI and PID Tuning Rules*; Imperial College Press: London, 2003.
(4) Ziegler, J. G.; Nichols, N. B. Optimum Settings for Automatic Controllers. *Trans. ASME* **1942**, *65*, 433.
(5) Seborg, D. E.; Edgar, T. F.; Mellichamp, D. A. *Process Dynamics and Control*, 2nd ed.; Wiley: New York, 2004.
(6) Skogestad, S. Simple Analytic Rules for Model Reduction and PID Controller Tuning. *J. Process Control* **2003**, *13*, 291.
(7) Chen, D.; Seborg, D. E. PI/PID Controller Design Based on Direct Synthesis and Disturbance Rejection. *Ind. Eng. Chem. Res.* **2002**, *41*, 4807.
(8) Wang, Y.-G.; Shao, H.-H. Optimal Tuning for PI Controller. *Automatica* **2000**, *36*, 147.
(9) Rivera, D. E.; Morari, M.; Skogestad, S. Internal Model Control. 4. PID Controller Design. *Ind. Eng. Chem. Process Des. Dev.* **1986**, *25*, 252.
(10) Ko, B.-S.; Edgar, T. F. Assessment of Achievable PI Control Performance for Processes with Dead Time. *Proc. Am. Control Conf.* **1998**, 1548.
(11) Horton, E. C.; Foley, M. W.; Kwok, K. E. Performance Assessment of Level Controllers. *International Journal of Adaptive Control and Signal Processing* **2003**, *17*, 663.
(12) MacGregor, J. F.; Harris, T. J.; Wright, J. D. Duality between the Control of Processes Subject to Randomly Occurring Deterministic Disturbances and ARIMA Stochastic Disturbances. *Technometrics* **1984**, *26*, 389.
(13) Marlin, T. E. *Process Control: Designing Processes and Control Systems for Dynamic Performance*; McGraw-Hill: New York, 1995.
(14) Garcia, C. E.; Morari, M. Internal Model Control. 1. A Unifying Review and Some New Results. *Ind. Eng. Chem. Process Des. Dev.* **1982**, *21*, 308.
(15) Åström, K. J.; Hagglund, T. *PID Controllers: Theory, Design and Tuning*, 2nd ed.; Instrument Society of America: Research Triangle Park, NC, 1995.
(16) Middleton, R. H.; Graebe, S. F. Slow Stable Open-Loop Poles: To Cancel or Not To Cancel. *Automatica* **1999**, *35*, 877.
(17) Morari, M.; Zafiriou, E. *Robust Process Control*; Prentice Hall: Englewood Cliffs, 1989.
(18) Huang, B. A Pragmatic Approach towards Assessment of Control Loop Performance. *International Journal of Adaptive Control and Signal Processing* **2003**, *17*, 589.
(19) Wang, L.; Cluett, W. R. *From Plant Data to Process Control: Ideas for Process Identification and PID Design*; Taylor and Francis: London, 2000.
(20) Zhuang, M.; Atherton, D. P. Automatic Tuning of Optimum PID Controllers. *IEE Proc. D* **1993**, *140*, 216.
(21) Palmor, Z. J.; Shinnar, R. Design of Sampled-Data Controllers. *Ind. Eng. Chem. Process Des. Dev.* **1979**, *18*, 8.
(22) Franklin, G.; Powell, J.; Workman, M. *Digital Control of Dynamic Systems*, 2nd ed.; Addison-Wesley: Reading, MA, 1990.
(23) Kristiansson, K.; Lennartson, B. Robust and Optimal Tuning of PI and PID Controllers. *IEE Proc. Control Theory Appl.* **2002**, *149*, 17.

Received for review October 11, 2004
Revised manuscript received April 4, 2005
Accepted June 20, 2005

IE040258O