Lee, W. and W. V. Weekman, Jr., "Advanced Control Practice in the Chemical Process Industry: A View from Industry," *AIChE J.*, **22**, (1976).
Maarleveld, A. and J. E. Rijnsdorp, "Constraint Control on Distillation Columns," *Automatica*, **6**, p. 51 (1970).
Morari, M., "Studies in the Synthesis of Control Structures of Chemical Processes," Ph.D. Thesis, University of Minnesota (Dec., 1977).
Morari, M., Y. Arkun and G. Stephanopoulos, "An Integrated Approach to the Synthesis of Process Control Structures," JACC Proceedings, Philadelphia, PA (1978).
Morari, M., Y. Arkun and G. Stephanopoulos, "Studies in the Synthesis of Control Structures for Chemical Processes. Part I. Formulation of the Problem. Process Decomposition and the Classification of the Control Tasks. Analysis of the Optimizing Control Structures," *AIChE J.*, **26**, p. 220 (1980a).
Morari, M. and G. Stephanopoulos, "Studies in the Synthesis of Control Structures for Chemical Processes. Part II. Structural Aspects and the Synthesis of Alternative Feasible Control Schemes," *AIChE J.*, **26**, p. 232 (1980b).
Morari, M. and G. Stephanopoulos, "Studies in the Synthesis of Control Structures for Chemical Processes. Part III. Optimal Selection of Secondary Measurements Within the Framework of State Estimation in the Presence of Persistent Unknown Disturbances," *AIChE J.*, **26**, p. 247 (1980c).
Motard, R. L., H. M. Lee, R. W. Barkeley, D. M. Ingels, *CHESS, Chem. Eng. Simulation System: System's Guide*, Technical Publ. Co., Houston, TX (1968).
Prett, D. M. and R. D. Gillette, "Optimization and Constrained Multivariable Control of a Catalytic Cracking Unit," 86th National AIChE Meeting (April, 1979).
Rijckaert, M. J., Martens, X. M. and J. Desarnauts, "Ethylene Plant Optimization by Geometric Programming," *Computers in Chem. Eng.*, **2**, p. 93 (1978).
Rijnsdorp, J. E., "Chemical Process Systems and Automatic Control," *Chem. Eng. Prog.*, **63** (7), p. 97 (July, 1967).
Rivas, J. R., D. F. Rudd and L. R. Kelly, "Computer-Aided Safety Interlock Systems," *AIChE J.*, **20** (2), p. 311 (1974).
Roffel, B., H. J. Fontein, "Constraint Control of Distillation Processes," *Chem. Eng. Sci.*, **34**, p. 1007 (1979).
Shah, M. J. and R. E. Stillman, "Computer Control and Optimization of a Large Methanol Plant," *Ind. and Eng. Chemistry*, **62** (12), p. 59 (Dec., 1970).
Umeda, T. and T. Kuriyama, "A Logical Structure for Process Control System Synthesis," Proc. of IFAC Congress, Helsinki, Finland (1978).
Webb, P. U., B. E. Lutter and R. L. Hair, "Dynamic Optimization of FCC Units," AIChE Annual Meeting, New York (1977).
Weber, R. and C. Brosilow, "The Use of Secondary Measurements to Improve Control," *AIChE J.*, **18**, p. 614 (1972).
Westerberg, A. W., "Decomposition Methods for Solving Steady-State Process Design Problem," in *Decomposition of Large-Scale Problems*, ed., D. M. Himmelblau, North Holland Pub. Co. (1973).
Westerberg, A. W. and C. J. DeBrosse, "An Optimization Algorithm for Structured Design System," *AIChE J.*, **19**, p. 335 (1973).
Wismer, D. A., ed., *Optimization Methods for Large-Scale Systems . . . with Applications*, McGraw-Hill (1971).

Manuscript received October 5, 1979; revision received August 14, and accepted August 20, 1980.

Design of Advanced Process Controllers

Z. J. PALMOR

Mechanical Engineering, Technion
Israel Institute of Technology
Haifa, Israel

and

R. SHINNAR

Chemical Engineering Department
City College of New York
New York, NY

This paper deals with defining the relation between modeling of process units and controller design. In the process industries, it is often difficult to obtain accurate process models. It is shown that advanced algorithms such as dead-time compensators require more detailed process information for design than do with *PI* and *PID* controllers. A method is described to systematically evaluate the relation between process identification and controller design, stability, and performance. It is shown that conventional gain and phase margins do not provide proper safety margins for dead-time compensators and optimal controllers. Methods for safe design of dead-time compensators are derived, and the approach could be useful in a wide class of problems.

To illustrate the approach presented, the design of *PI* controllers by conventional methods is analyzed. The conditions, as well as the exceptions, are specified in which methods such as described by Ziegler and Nichols or Cohen and Coon will give good results. It is shown that for any stable nonlinear system with an input/output function $y = G_p^*(u)$, a linearized design function G'_{pd} can be constructed and identified that guarantees stability and reasonable performance of the controller (Eqs. 21 and 22 and Figures 4-7) over a wide range of operating conditions. A rigorous framework for identifying the exceptions and understanding the reason why such methods work is presented.

SCOPE

The paper deals with a fundamental problem of chemical engineering design, the relation between the quality of the information required for a specific design and the performance of the system. Chemical process units are normally nonlinear systems, the exact mathematical description of which is complex and difficult to obtain. We therefore have to ask ourselves what is the type and accuracy of information really required for a specific design problem and how much will better information pay off in increased performance. Here, we try to approach this question for continuous controllers with a single measured and manipulated variable, starting with simple *PI* controllers and then looking at control loops incorporating a dead time compensator.

The emphasis is not on tuning recipes, but on trying to understand how strongly simplified linearized models permit design of controllers for nonlinear complex systems and to define the quality of the information required for design in a rigorous way. The approach taken uses the concept of model space defined in a previous paper. It is explained in detail with a few examples and related to present methods of controller design. It should be a useful tool in a number of other applications in chemical reactor design as well as in the design of control systems in general.

0001-1541-81-4867-0793-$2.00

CONCLUSIONS AND SIGNIFICANCE

To contribute appreciably to system performance, advanced controllers, like those derived from optimal control algorithms, require a much better identification of G_p^* than do simple controllers. As an example, the dead-time compensator that appears in any optimal control algorithms dealing with a process having inherent delays is discussed. Process models of this type are noninvertible, non-minimum-phase functions. One of the features of optimal algorithms is that the controller contains an inversion of G_{pd}, which makes sense, since for invertible transfer functions K/G_{pd} is an excellent controller. However, even for noninvertible process models, the algorithms contain the equivalent of inversion (see Figure 3b). The algorithm uses strong control action whenever the phase lag of G_pG_c equals multiples of 2π. As the phase lag is not known accurately for high frequencies, such strong control is not used, but rather what is known about G_p^* is relied on.

It is shown that, for optimal controllers, phase and gain margins give no indication of the stability margin (or robustness) of the controller. It is also shown that the type and quality of information about G_p^* that is required to achieve a better performance of the controlled system can be specified exactly. This is accomplished by defining a set of linear transfer function G_{pd}^j that bound the real linearized response within a specified range of frequencies. A similar set can be used to characterize the nonlinear behavior of the system. Using this approach, the dead-time compensator can be modified to give satisfactory performance.

Finally, the approach presented here is related to stability analysis using sensitivity to parameter variation, and it is shown that the results are similar but that the method under discussion allows a quantitative definition of the parameter space over which the system should perform satisfactorily. It also defines the minimum complexity of the process model required. In summary, it is shown that there is a systematic way of comparing simplified and complex models, and it is possible by simulation to judge the merits of more detailed process information for the purposes of controller design. It is shown that for successful advanced controller design, one must take into account the knowledge one possesses about the system which is different from case to case. It is also shown that the main reason straightforward application of optimal design methods sometimes fail is that the identification and modeling procedures are incompatible with the design method.

INTRODUCTION

Many industrial processes contain inherent delays (or dead times) between the effect of process inputs on outputs. If such delays are large, effective control of the process is difficult to achieve. One good way of improving control in such cases is to use a dead-time compensator (abbreviated *DTC*), also known as a Smith predictor (O. Smith, 1959). Such a predictor is in some form the inherent result whenever one applies a minimum-variance optimization algorithm to a process having a delay (or dead-time in our terminology, abbreviated *DT*) (Palmor and Shinnar, 1978, 1979; Palmor, 1980a). It can be very effective (Alevisokis et al., 1974; Palmor and Shinnar, 1978, 1979; Lupfer and Oglesby, 1961; Nielsen, 1969; Meyer et al., 1977), but it is far less commonly used than its performance merits. One possible reason for its infrequent use might be the difficulty of properly adjusting the controller (Palmor and Shinnar, 1979; Ross, 1977). The *DTC* eliminates the *DT* from stability consideration, and consequently users may tend to adjust the controllers coupled with the *DTC* by conventional techniques (O. Smith, 1959; Alevisakis et al., 1974; Lupfer and Oglesby, 1961; Donoghue, 1977; Ross, 1977; and others) as if the *DT* did not exist.

This leads in many cases, and especially when the *DT* is large, to unsatisfactory performance of the controlled system. The main reason for this inadequate performance is that dead-time compensators are sensitive to differences between the behavior of the process model assumed and that of the real process (Buckley, 1960; Palmor and Shinnar, 1978; Palmor, 1980a, 1980b). This is a common problem in many optimal design algorithms (Edgar, 1976). In the following discussion a coherent design strategy for applying such controllers will be derived. The emphasis is not on giving a straightforward recipe for tuning, which will be presented in a following paper, but rather on developing an understanding of the design method itself. The *DTC* is used as an example that illustrates a powerful method for the design of complex process controllers in general. Efficient design of advanced controllers depends strongly on understanding the quality of information required for a specific design.

The quality of the information known or that can be obtained for a given process changes from case to case. For a successful design the more complex and efficient the design method, the better the understanding needed of the nature and quality of the information about the process. It is the relation between the quality of information required for an advanced design procedure and the incremental improvement in performance that is considered here.

Structure of Dead-Time Compensators

Let us start by reviewing the derivation of a *DTC* and discussing the cause of the difficulties in designing it. Consider a linear process having a transfer function $G_p(s)$ of the form:

$$G_p(s) = \frac{y(s)}{u(s)} = G_{po}(s)e^{-\theta s} \tag{1}$$

where $G_{po}(s)$ is a lumped parameter function, $y(s)$ is the output, $u(s)$ is the manipulated variable, and θ is the *DT*. If θ is of the same order of magnitude as the dominant time scale of $G_{po}(s)$, the permitted gain of any controller is strongly reduced. For example, if we make a change in the set point, the output sensor will not respond until a time θ has elapsed.

If we chose a large gain for the controller, we will continue to change u and overcorrect. One way of overcoming this limitation would be to build a predictor that computes the effect of the control action that has been taken during the interval θ and deduct it from the measured error being transmitted to the controller. If the basic controller has the form $G_{co}(s)$, without a *DTC* the control action $u(s)$ is $-G_{co}(s)y(s)$. If we deduct from $y(s)$ the effect of $u(s)$ during the delay, we get

$$u(s) = -G_{co}(s)\{y(s) - u(s)\, G_{po}(s)(1 - e^{-\theta s})\} \tag{2}$$

which can be written

$$G_c(s) = \frac{u(s)}{e(s)} = \frac{G_{co}(s)}{1 + G_{co}(s)G_{po}(s)(1 - e^{-s\theta})} \tag{2a}$$

Equation 2a is presented schematically in the diagram in Figure 1. The term "predictor" might be clearer from an equivalent representation of the feedback scheme described in Figure 1, which is given in Figure 2. We can look at any control problem, especially of a system with a DT, in the following way. If we introduce a control action now, it will be felt in the future. Let us now, at time t, look at a future point in time $t + \Delta t$. If the DT is θ, then the only time we can affect the output is for values of Δt larger than θ.

Based on the previous history of the process and knowledge of the disturbance, we can try to predict what the values of the output variable y would be if no control action is taken. We often have a reasonable idea about the time scale and nature of the disturbance affecting a process, and we can therefore try to build a predictor for $y(t)$ for values of Δt greater than θ. We can now formulate a control policy $u(t)$ that will hold $y(t)$ at time $t + \Delta t$ at some desired value. Two predictions are matched here: One for the output in the absence of any control, $y_1(t + \Delta t)$; and one for the result of the policy $u(t)$ on the output, $y_2(t + \Delta t)$. Since any such estimation uses the past history of $y(t)$, it is completely equivalent to a feedback controller, which also acts on the system based on the past values of $y(t)$.

Instead of formulating a control strategy G_c (or controller transfer function), we formulate the structure of a predictor for future $y(t)$ based on the history of $y(t)$ and $n(t)$. This is very useful as it relates G_c to both the nature of the inputs and the constraints of a real process. This relationship was discussed in detail by Palmor and Shinnar (1979) for sampled data processes, and the conclusions apply equally well here. The reader is referred to this work for a more detailed discussion.

There is one conceptual difference between the approach in Figure 2 and the standard concept of feedback control. The control policy presented here is based not only on the history of $y(t)$, but also on the history of $u(t)$. In trying to match the two predictions, the procedure must include all previous actions that will affect the process at time $t + \Delta t$. As any action taken during the time interval $(t - \theta, t)$ has had no effect on $y(t)$, the concept of taking into account past control actions that have as yet not affected $y(t)$ was added to the standard feedback strategy. This leads automatically to the scheme given in Figure 1. In that sense, a dead-time compensator is also a predictor. It will therefore automatically appear in any optimal control policy based on schemes that are represented in Figure 2, in which G_p contains a delay.

Now let us look at the closed-loop transfer function that is obtained from such a controller. The overall transfer function from y to r is:

$$\frac{y(s)}{r(s)} = \frac{G_{co}(s)G_{po}(s)e^{-\theta s}}{1 + G_{co}(s)G_{po}(s)} \tag{3}$$

Note that while the response of the system still contains the delay, the characteristic function does not. This might lead to the conclusion that one can chose and tune a controller using only $G_{po}(s)$ (O. Smith, 1959; Alevisokis et al., 1974; Donoghue, 1977). This inference is misleading. If we follow this argument to its logical conclusion, if $G_p(s)$ is invertible we could choose a controller $G_c(s) = K_c/G_p(s)$, and we would end up with a characteristic equation, $1 + K_c$ that is stable for any K_c. It is well known that such arguments are not allowed (for example, Rosenbrock, 1974; Ragazzini and Franklin, 1959).

The mathematical description of a practical industrial process will normally be highly complex and in most cases nonlinear. G_p^* is designated as the real model of the process, with G_p' its linear approximation. In most cases a simplified approximation of G_p' is used, which will be designated G_{pd}. However, always remember that although G_{pd} is clearly defined, G_p' and G_p^* are often not completely known, since theoretical modeling as well as empirical identification procedures have their limitations.

Even with these limitations, the form of G_p' that results from theoretical modeling is often too complex for control design purposes, and therefore approximated forms are used for G_{pd} (Gould, 1969; Koppel, 1968).

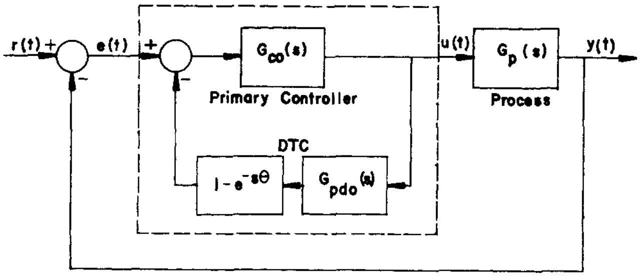

Figure 1. Closed-Loop System with a Dead-Time Compensator

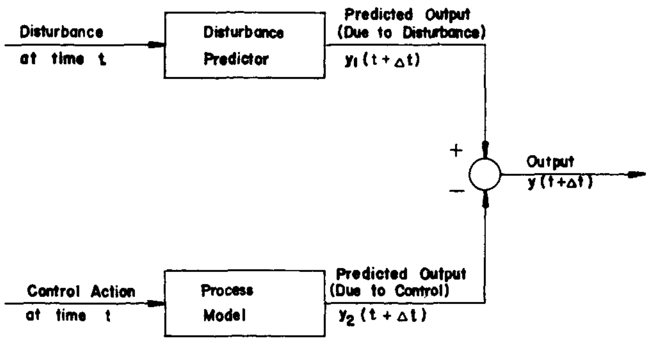

Figure 2. Interpreting the Control Problem as Matching Two Predictions

Assume that G_p' can be written in the simple form given in Eq. 1 ($G_p' = G_{po}'(s)e^{-\theta s}$) but $G_{po}' \neq G_{pdo}$. If we recompute Eq. 3, we get

$$\frac{y(s)}{r(s)} = \frac{G_{co}(s)G_p'(s)}{1 + G_{co}(s)G_{pdo}(s)(1 - e^{-\theta s}) + G_{co}(s)G_p'(s)} \tag{4}$$

Even though the delay is assumed to be known exactly, the characteristic equation still contains the delay term, since G_{po}' and G_{pdo} are not the same. Though we cannot demand a controller to be totally insensitive to such differences, we must require that it not be oversensitive to the exact form of G_{pd} and allow for some differences between G_{pd} and G_p'.

In reality the relation between G_{pd} and G_p' is even more complex, and this relationship will be discussed later. Even if the real delay is slightly different from our estimate ($\theta' \neq \theta$) (while the structure and the other parameters match exactly, Eq. 4 becomes:

$$\frac{y(s)}{r(s)} = \frac{G_{co}(s)G_p'(s)}{1 + G_{co}(s)G_{po}(s)(1 - e^{\theta s} + e^{-\theta' s})} \tag{5}$$

Note that though Eq. 3 did not contain the delay in the characteristic equation, both Eqs. 4 and 5 do.

Before going to the stability analysis, it will be useful to elaborate the concept of inversion of a transfer function.

Inversion can appear in several forms. A transfer function G_p is called invertible if it is minimum phase. A transfer function containing a delay does not have a physically realizable inverse, at least theoretically. However, one can design physically realizable control algorithms that eliminate the delay from the characteristic equation. A similar case of eliminating terms from the characteristic equation appears in multivariable controller design. For example, the concept of noninteracting control is based on the fact that control circuits can be designed that minimize (or theoretically eliminate) the interaction between control loops. Thus, for example, if plant outputs Y are related to the manipulated variables U by:

$$Y(s) = G_p(s)\,U(s)$$

where Y and U are vectors and $G_p(s)$ is a transfer matrix, a noninteracting controller G_c can be designated by choosing

$$G_c(s) = G_p(s)^{-1}K(s)$$

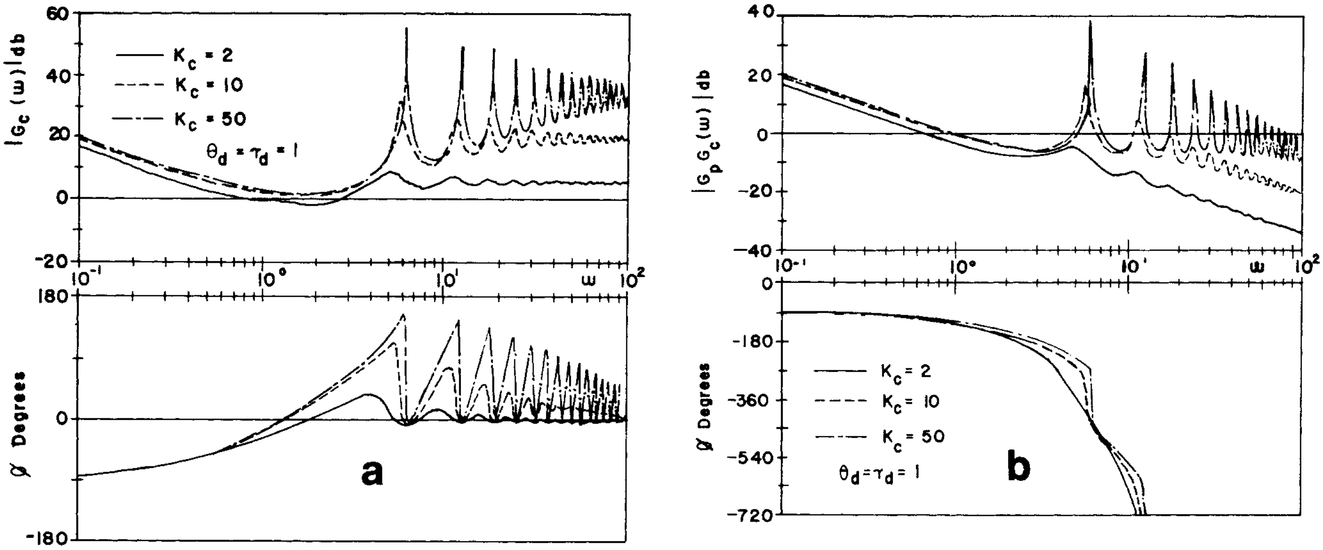

Figure 3. a. Bode Plots for Controller (7) with Different Gains b. Open-Loop Frequency Responses for Process (6) and Controller (7)

where $K(s)$ is a diagonal matrix of transfer functions. Although specific design methods of multivariable controllers are outside the scope of this paper, in many cases they involve an inverse matrix of the process transfer-function matrix $G_p(s)$. To get diagonal matrices, compensation loops must be introduced that cancel terms in the elements of $G_p(s)$. The problems here are similar. G_p must be checked to see if it is invertible. If not, most modern design methods are able to handle noninvertibility in some way.

But, in all cases, terms of the characteristic equations are eliminated, and thus a wrong picture on the stability of the system can result. Inversion of the transfer function is a valuable tool to get a good controller design, but a method to judge stability is needed. In the following section such a method for a simple example of a single-input, single-output system is devised. The general approach should, however, also be valid for more complex and multivariable systems.

STABILITY ANALYSIS OF CONTROLLER WITH DTC: CONCEPT OF INVERSION

The lack of the time delay in the characteristic equation (Eq. 3) is equivalent to the inversion of an invertible transfer function. In some sense optimal control algorithms are able to invert a noninvertible transfer function in an intelligent way. Let us consider the case where:

$$G_{pd} = \frac{1}{1 + \tau s}\, e^{-\theta s} \tag{6}$$

which is a stirred tank (or first-order filter) with a *DT*. The process gain here has been normalized to unity. In this case G_{pdo} is invertible and the only noninvertible part of G_{pdo} is the delay (*DT*).

For this example the controller is a *PI* controller $[G_{co}(s) = K_c(1 + 1/\tau s)]$ coupled with a dead-time compensator. (Note that the integral time chosen is equal to τ.) The overall controller transfer function is:

$$G_c(s) = \frac{K_c(1 + 1/(\tau s))}{1 + K_c\{1 + 1/(\tau s)\}(1 - e^{-\theta s})/(\tau s + 1)} = \frac{K_c(1 + \tau s)}{\tau s + K_c(1 - e^{-\theta s})} \tag{7}$$

Equation 7 can be derived by solving the unconstrained minimum-output variance problem for a controller acting on Eq. 6 with a certain type of nonstationary noise (Palmor, 1980a; Palmor and Shinnar, 1979). The general form of this type of controller for a system with $G_p(s) = G_{po}(s)e^{-\theta s}$ is:

$$G_c(s) = \frac{K_c/(sG_{po}(s))}{1 + \{K_c/(sG_{po}(s))\}\, G_{po}(s)(1 - e^{-\theta s})} \tag{8}$$

The closed-loop transfer function from y to r is therefore:

$$\frac{y(s)}{r(s)} = \frac{K_c}{s + K_c}\, e^{-\theta} s \tag{9}$$

Note that K_c can be chosen to be as large as desired. Even if $K_c \to \infty$ the controllers given by both Eqs. 7 and 8 appear to have a finite gain, since for $K_c \to \infty$ $G_c(s)$ in Eq. 7 reduces to:

$$G_c(s) \to \frac{1 + \tau s}{1 - e^{-\theta s}} \qquad K_c \to \infty \tag{10}$$

and $G_c(s)$ in Eq. 8 to:

$$G_c(s) \to \frac{(G_{po})^{-1}}{1 - e^{-s\theta}} \qquad K_c \to \infty \tag{11}$$

For both Eqs. 10 and 11, Eq. 9 becomes:

$$\frac{y(s)}{r(s)} \to e^{-\theta s} \qquad K_c \to \infty \tag{12}$$

which for $\theta \to 0$ simply gives $y(s) = r(s)$, or perfect control, as would be expected for an invertible transfer function with a properly chosen controller. As the controller is a physically realizable transfer function, it of course cannot eliminate the delay from the overall response. In fact, it eliminated from the overall transfer function only the invertible part of G_{pd}. But, the problems faced in the design of this controller are exactly the same as if G_{pd} itself would be invertible. Still, a very large K_c should not be chosen as we do not rely on the fact that G_c cancels all the poles and zeros of G_p. In other words, it is recognized that $G_{pd} \neq G_p'$.

The problem here is the same. In terms of stability analysis the *DTC* eliminated the delay from the characteristic equation. If the function is invertible, the algorithm allows infinite gains at all frequencies. For a noninvertible function it uses a very high gain for those frequencies at which it is safe to do so and a low gain at the frequencies for which a large gain would lead to instability. This can be illustrated by considering the Bode diagram of $G_pG_c(j\omega)$ (see Figures 3a and 3b). For the example given here $G_pG_c(s)$ is simply:

$$G_pG_c(s) = \frac{1}{1+\tau s}\frac{K_c(1+\tau s)e^{-\theta s}}{\tau s + K_c(1-e^{-\theta s})} = \frac{K_c e^{-\theta s}}{\tau s + K_c(1-e^{-\theta s})} \tag{13}$$

and for large $K_c(K_c \to \infty)$ reduces to:

$$G_pG_c(s) \to e^{-\theta s}/(1 - e^{-\theta}s) \tag{13a}$$
$$K_c \to \infty$$

Rewritten in terms of magnitude and phase angle, Eqs. 10 and 13a become:

$$|G_c(j\omega)| = \sqrt{(1+\tau^2\omega^2)}\Big/\left(2\left|\sin\left(\frac{\theta\omega}{2}\right)\right|\right) \tag{14a}$$

$$\phi_{G_c} = \tan^{-1}(\tau\omega) - \pi/2 + \frac{1}{2}R\{\theta\omega\} \tag{14b}$$

and

$$|G_cG_p(j\omega)| = \frac{1}{2}\left\{\left|\sin\left(\frac{\theta\omega}{2}\right)\right|\right\}^{-1} \tag{15a}$$

$$\phi_{G_cG_p} = -\pi/2 - \frac{1}{2}R\{\theta\omega\} \tag{15b}$$

respectively, where $R\{\alpha\}$ denotes a residue of α modolu 2π.

Equation 14b indicates that $G_c(s)$ in Eq. 10 gives considerable phase advance. Using Eq. 15b the phase crossover frequencies, ω_{co}, at which $\phi_{G_cG_p} = -(2n+1)\pi$, may be determined. The result:

$$R\{\theta\omega_{co}\} = \pi$$

The amplitude ratio at these frequencies is thus:

$$|G_cG_p(j\omega_{co})| = 1/2 \tag{16}$$

Similarly, the gain crossover frequencies, ω_g, for which $|G_cG_p(j\omega)| = 1$, may be easily found from Eq. 15a to be:

$$R\{\theta\omega_g\} = \pi/3$$

The corresponding phase lags are:

$$\phi_{G_cG_p}(j\omega_g) = -2\pi/3 \tag{17}$$

Equations 16 and 17 indicate that the closed-loop system with the controller (Eq. 10) has a gain margin (designated *g. m.*) of 2 and a phase margin (designated *p. m.*) of 60°. On the other hand, it is evident from Eq. 15a that $|G_cG_p(j\omega)|$ becomes infinite at the frequencies;

$$R\{\theta\omega_\infty\} = 0$$

which are located at the maximum distance possible from the various phase crossover frequencies, ω_{co}.

The *g. m.* and *p. m.* given above have been derived for the *PI* plus *DTC* controller in Eq. 7 with infinite gain. In the appendix it is proven that these properties hold for any controller given by Eq. 8 and that for finite gains the *g. m.* > 2 and the *p. m.* > 60°. The fact that the *DTC* acts like a lead network and might have infinite amplifications has also been noted by Astrom (1977).

If G_p' were really equal to G_{pd}, it would be an ideal controller. Let us go back to the invertible case. If G_{pd} is invertible then by choosing $G_c = K_c/G_{pd}$, $\phi_{G_pG_c}$ has been made equal to zero for all ω. There is therefore no need to worry about crossover frequencies and any K_c may be chosen. In the case of the type of noninvertible function considered here, this cannot be done because the phase lag decreases continuously with ω and will attain a value of $-(\pi + 2n\pi)$ an infinite number of times. The controller specified does the best that can be done. It puts all the strong control action at frequencies that have phase lags far from $\phi_{\omega co}$.

It does so while still keeping a proper phase and gain margin. But it can be immediately seen that this action is meaningless. If there is only a slight mistake in θ, then for high frequencies the peak in $|G_cG_p(j\omega)|$ will occur at a different phase lag. Not only is perfect knowledge of the *DT*, but also of $G_{po}(s)$, assumed here. Exactly the same illusion of stability is present here as when the term $1/G_{pd}$ in G_c is included for an invertible transfer function. This is a problem common to many optimal control algorithms with or without delays, single or multi-variable, using a stochastic design approach (Palmor and Shinnar, 1979) or a deterministic formulation. Such algorithms will often include indirect inversion even if the transfer functions are noninvertible. Proper optimal methods avoid this problem and lead to realizable controllers.

However, the algorithm makes use of the information contained in G_{pd} in a way that is somewhat similar to inversion of an invertible function. Phase margins are here meaningless as can be seen from the example and the Appendix. For an infinite K_c the slightest deviation of G_p' from G_{pd} would make the system unstable, but the design still shows the proper phase and gain margins. There are several ways of overcoming this instability. For example, K_c can be reduced. Figures 3a and 3b, give the Bode diagram of both $G_c(s)$ and $G_pG_c(s)$ for different values of K_c. An optimal algorithm can also be formulated that leads to exactly this type of controller. (See Eq. 8 and Palmor, 1980b.)

For finite values of K_c the peaks of $|G_c(j\omega)|$ and $|G_pG_c(j\omega)|$ decrease as ϕ increases. $|G_pG_c(j\omega)|$ will always have an infinite peak at $\omega \to 0$ regardless of K_c. This peak is due to the inclusion of an integrator in G_c(Eqs. 7 and 8). A high value of $|G_c(j\omega)|$ and $|G_pG_c(j\omega)|$ for $\omega \to 0$ is a requirement for any controller. If K_c is reduced enough, $|G_pG_c(j\omega)|$ will not show any additional peaks that exceed unity. There are a number of other ways to deal with this problem. For example, we can demand that the controller is reasonably insensitive to parameter variation (Kestenbaum et al., 1976; Palmor and Shinnar, 1979; Shinnar, 1976; Rosenbrock, 1974). A constraint can be put on the control effort (Palmor and Shinnar, 1979; Palmor, 1980b; Shinnar, 1976), which will have similar results, or a filter can be put on the dead-time compensator or on G_c that reduces the peaks in $|G_pG_c(j\omega)|$ for high frequencies. But these are mechanistic approaches. The control effort per se is not a major concern.

In process control the main cost related to the level of the control effort is due to the steady-state control. And if it were only the uncertainties of the parameters in G_{pd} that were of concern, one could try to measure them more accurately and update the measurements by some adaptive procedure (Astrom and Wittenmark, 1973). The next section therefore tries to deal with the heart of the problem, namely the inaccuracy of the model itself.

MODEL IDENTIFICATION IN CONTROLLER DESIGN: CONCEPT OF "MODEL SPACE"

The fact that the design model. G_{pd}, used for design is not equal to the real model is a problem that also exists in conventional controller design, in fact in most controller design problems in the process industry (Shinnar, 1978). It is often encountered without the engineer even being aware of it and dealt with in an intuitive way. Shinnar (1978) introduced the concept of "model space" as an attempt to formalize what a skilled designer does intuitively.

Let us define the problem. In a chemical process the real model G_p^* is nonlinear, often complex, and inexactly known. The degree to which it is accessible to identification and mathematical formulation varies greatly from process to process. A much more accurate model is possible for a bubble-tray distillation column or some heat exchangers than for a hydrocracker or a fluid catalytic cracker. The complexity of the latter is not simply a question of the mathematical model but rather a problem of identifying G_p^*.

Shinnar (1978) discussed the fact that process models involve two types of complexity. One is purely mathematical in the sense that if the proper model becomes too complex, it must be simplified to get any analytical solution or even to allow simulation in a reasonable amount of computer time. Industrial pro-

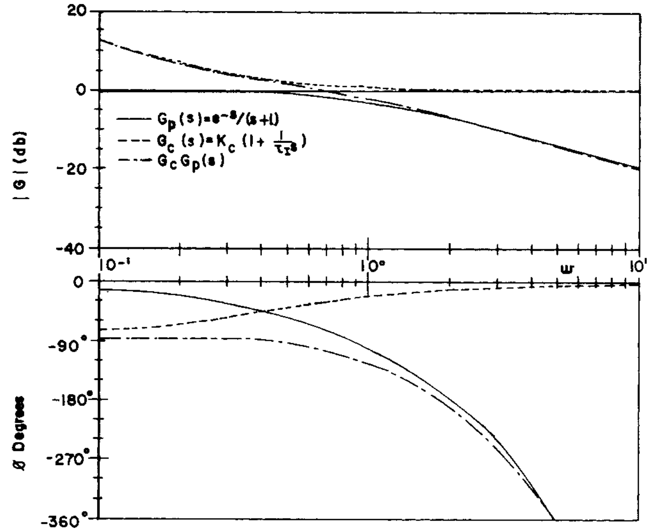

Figure 4. Bode Diagrams for $G_p(s)$, $G_c(s)$, and $G_pG_c(s)$

cesses, especially in the chemical industry, possess a different type of complexity. The proper model is unknown and proper identification involves a large experimental effort. The real description of the system may also contain a large number of constants and it is hard to simultaneously identify more than two or three constants. An approximated simplified model, with a small number of constants that can be identified, must therefore suffice.

This does not mean that good experimental process identification is not wanted. Good measurements of step responses and frequency response (for example, Hougen, 1979) are extremely valuable in constructing a process model. Often far too little effort is spent on good identification. But though good identification will lead to improved controller design, it does not mean that $G_{pd} = G_p^*$, especially in nonlinear cases. In many design problems and especially in control, the work required to get a close fit between G_p^* and G_{pd} often cannot be afforded. Furthermore, in a nonlinear system G_p^* itself is a function of many other frequently unknown inputs and variables. If a design method based on a linearization of G_p^* is used, the linearized version of G_p^*, G_p', changes with operating conditions. The range of these changes varies from problem to problem.

What is normally meant by saying that a function G_{pd} is a good representation of G_p^*? Assume that for a given system G_p^* is a matrix function that relates the vector y to a vector u such that:

$$y = G_p^*(u) \tag{18}$$

G_{pd} is a useful fit of G_p^* if $\|G_{pd}(u) - G_p^*(u)\|$ is smaller than a given value over a certain range of u. If this difference is very small over all possible values of u, then for design purposes G_{pd} is identical to G_p^*. This is seldom true for complex systems, and we therefore have to be more specific. The range of u for which we want a prediction and the precision of the prediction desired both must be specified.

The concept of model space is based on the assumption that for any given unknown G_p^*, a set of models $\{G_{pd}\}$ can be found such that:

$$\|\{G_{pd}(u)\} - y\| < E$$

where the values of y are the measurements of $G_p^*(u)$ for a wide range of experimental values of u, and E contains both the experimental error and the deviation between G_{pd} and G_p^*. This model space is a vaguely defined set, as the form and order of $\{G_{pd}\}$ are not specified exactly. For a given G_p^* there exists a number of reasonable formulations of the set G_{pd} that give acceptable errors over a space u.

Proper formulation of G_{pd} depends on knowledge of the system. In a system built of many known and clearly defined elements, such as servomechanisms, electronic filters, or even a bubble-tower distillation column, the complexity is due to the accumulation of errors in the functions describing the individual elements. In chemical reactors and many other process systems, the complexity is in the nature of the nonlinear system itself. The forms of the permissible function G_{pd} are limited mainly by what is known about the system. Often a simpler G_{pd} is intentionally chosen even though a more complex version could be derived from process modeling.

Better identification of a complex system is equivalent to narrowing the properties of the set $\{G_{pd}\}$ that is consistent with the experiments and increasing the space of u over which the error of the prediction remains small. But very often the designer cannot afford to identify G_p^* that closely, as it is too time-consuming and expensive. He must then ask himself what properties of G_p^* are really affecting the design.

Let us diverge for a moment to try to understand the relation between G_{pd} and G_p' when conventional *PI* and *PID* controllers are used. Consider for example a system in which G_p^* is complex but linear and constant with time. If a *PI* controller is designed, a complete description of the system is not needed. Let us go back to the open-loop Bode diagram. If, for example, we know that $|G_p^*(j\omega)|$ is monotonously decreasing for high values of ω, or more exactly for all values of ω greater than the first ω_{co}, then for the purpose of stability we need only know $G_p^*(\omega)$ in the frequency range in the vicinity of ω_{co}. The integral controller can be designed such that the effect of the integral mode close to the crossover frequency is minor.

Figure 4 gives as an example the open-loop Bode diagrams for the process

$$G_p(s) = \frac{e^{-1}}{s+1} \tag{19}$$

for the controller

$$G_c(s) = K_c\left(1 + \frac{1}{\tau s}\right) \tag{20}$$

and for $G_cG_p(s)$.

Equation 19 has been suggested as a useful approximation for all overdamped transfer functions in process control.

If G_p^* itself is linear (or equal to G'_p in our definition) but different from G_{pd} all that is required for the system to be as stable as that designed for G_{pd} is that:

$$|G_p'(j\omega)| < |G_{pd}(j\omega)| \tag{21}$$

for all ω and

$$\phi_{G_p'} > \phi_{G_{pd}} \tag{22}$$

for all ω for which $|G_pG_c(j\omega)| > 0.5$.

Figure 5 gives, as an example, the Bode diagrams for three different transfer functions that satisfy conditions (Eqs. 21 and

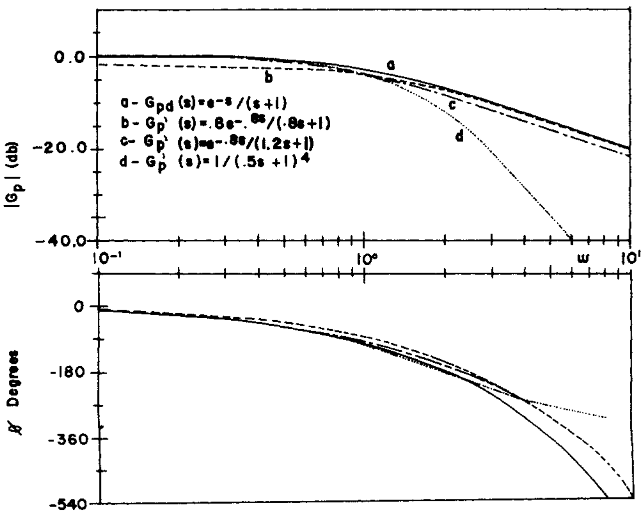

Figure 5. Bode Plots for Processes Satisfying Eqs. 21 and 22.

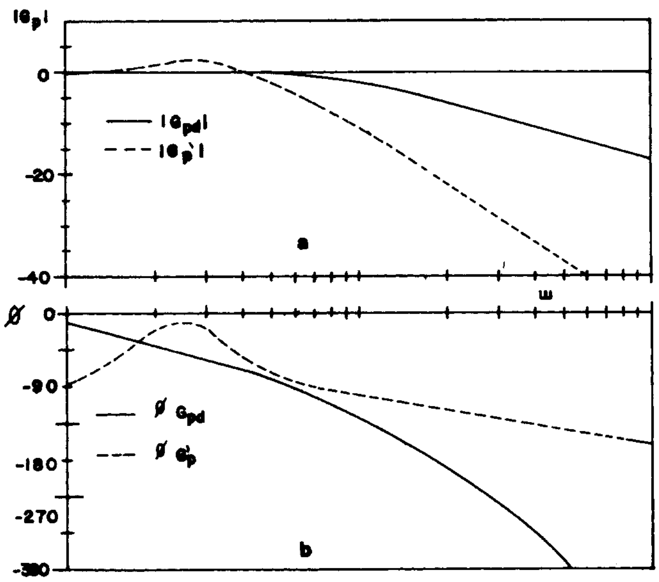

Figure 6. a. An Example of $G'_p(s)$ Not Fulfilling Eq. 21 at Low Frequencies b. An Example of $G'_p(s)$ Not Fulfilling Eq. 22 at Low Frequencies

22) if $G_{pd}(s)$ is given by Eq. 19. At high frequencies Eq. 21 may be relaxed such that $|G'_p(\omega)|$ does not have to be less than $|G_{pd}(\omega)|$ but only

$$|G'_p(\omega)| \leq 0.5\ |G_c(\omega)|^{-1} \tag{23}$$

If the process possesses an overdamped transfer function a function $e^{-\theta s}/(1 + \tau s)$ can always be found that fulfills the criterion. That can also be done from the step response by fitting it as closely as possible by a linear transfer function, drawing the Bode diagram, and then fitting it by a simpler design function that fulfills criteria (Eqs. 21 and 22). Equivalent criteria could be found directly for the response to a step function, but that is not the concern in this paper. Actual design will be discussed in follow-up papers. It should be pointed out, however, that the reaction curve method proposed by Ziegler and Nichols (1949) to fit Eq. 19 to experimental step responses does not guarantee a fit for criteria (Eq. 21) and 22 and therefore can lead to unstable designs (Kestenbaum et al., 1976).

Under certain conditions Eq. 21 may also be relaxed for low frequencies. If Eq. 22 is fulfilled, for low frequencies $|G'_p(j\omega)|$ can be larger than $|G_{pd}(j\omega)|$ and could be underdamped, Figure 6a. On the other hand, cases where condition (Eq. 22) is not satisfied at low frequencies must be treated with caution. An example in which the phase lag has a local minimum is shown in Figure 6b. In such cases the integrator may cause instability unless both Eqs. 21 and 22 are fulfilled. Thus, for certain types of transfer function all that must be known is the behavior near the crossover frequency.

The original Ziegler and Nichols empirical tuning method takes advantage of this fact. For such cases, a *PI* controller can be designed by just knowing the crossover frequency and the amplitude ratio $|G_p(\omega)|$ at the crossover frequency. Both parameters can be directly determined experimentally by using a proportional controller and increasing K_c until the system starts to become unstable, measuring both the critical value of K_c and the frequency of the oscillations. What is being done here is the following. The proportional controller is adjusted such that it fulfills criterion (Eq. 23), and τ_I is chosen such that the integral controller only affects the behavior of the Bode diagram at frequencies safely below ω_{co}. This procedure is always safe as long as the phase lag in the Bode diagram does not look like Figure 6b.

This procedure gives good setting for certain types of simple systems, but it can be seen readily from Eqs. 21 and 22 why it works and what the limitations are. Also, note that for a *PI* controller this method does not require any information about $G'_p(s)$ for frequencies higher than the first cutoff frequency, provided $G'_p(j\omega)$ decreases monotonously. The value of τ_I in the Ziegler-Nichols method is:

$$\tau_I = \frac{\pi}{\omega_{co} \times 0.6}$$

The amplitude ratio due to the integral mode is thus

$$\sqrt{(\tau_I\omega_{co})^2 + 1}/(\tau_I\omega_{co}) \sim 1.02$$

The phase-lag contribution of the *I* mode is approximately $-10°$. That is to say, the crossover frequency, ω_{co}, is practically unaffected by Ziegler-Nichols tuning for τ_I. It was noted earlier that there are cases for which tuning based solely on K_c and ω_c will not work. In general a design model, G_{pd}, based on a reasonably good experimentally verified knowledge of G'_p (or G^*_p) will lead to superior results.

For *PID* controllers the case is more complex. The amplification of an ideal derivative controller increases with frequency and as a good knowledge of $G'_p(j\omega)$ at high values of ω is lacking, one could run into difficulty. A way out of this difficulty is to use a high-frequency filter for the derivative controller action. A lead compensator of the form:

$$\frac{1 + \tau_D s}{1 + (\tau_D/\gamma)s} \tag{24}$$

can be used. This type of compensator is common to electronic controllers. By proper choice of γ and τ_D it can be assumed again that the design does not require a knowledge of G'_p in the high-frequency range, Figure 7. But note that a knowledge of G'_p over a wider range of frequencies is needed than is the case for the *PI* controller. The fact that a lead compensator reduces the phase lag and thereby increases ω_{co} is helpful. If the Ziegler-Nichols method for choosing τ_D is used, some assumptions are again made about the slope of $G'_p(\omega)$ at high frequencies. By proper choice of γ the range of frequencies can be clearly defined, which requires a detailed knowledge of the upper bound of $|G'_p(\omega)|$. For values of ω much larger than $1/\gamma$, all that is needed is that $|G'_p(\omega)|$ be less than $0.5\gamma/K_c\tau_D$.

The fact that no information is needed about either the exact phase lag or the amplitude ratio of G'_p is very important, as such information is extremely hard to obtain in any real process. The advantage of a *PI* controller is that it requires such limited information. When one says that very simple models are sufficient to tune *PI* controllers, one really means that a *PI* controller is sensitive only to certain properties of the transfer function.

Now let us elaborate on Eqs. 21 and 22, both in their exact and in their relaxed form for a nonlinear system. Assume that the

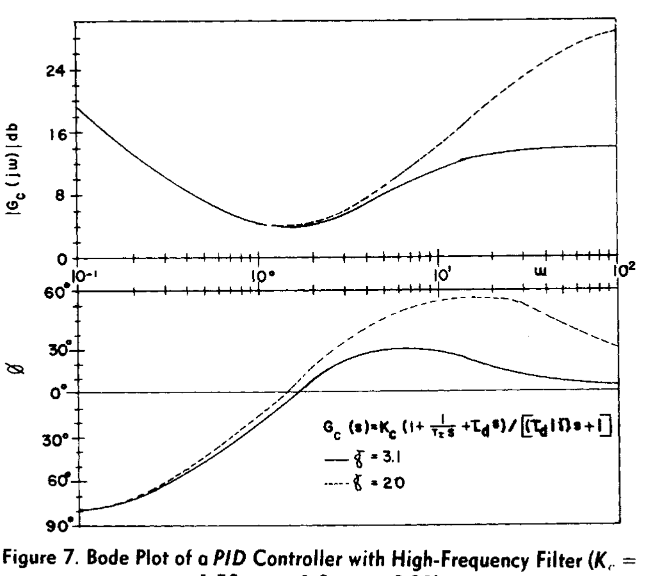

Figure 7. Bode Plot of a *PID* Controller with High-Frequency Filter (K_c = 1.58, τ_I = 1.8, τ_D = 0.31)

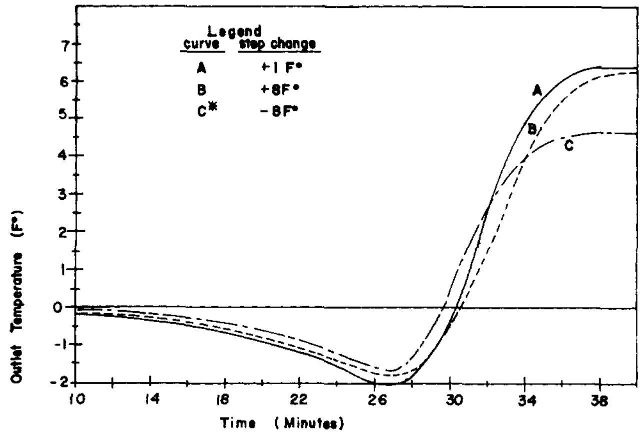

Figure 8. Response to a Step Change in Inlet Temperature of a Packed Reactor with an Exothermic Reaction (from Silverstein and Shinnar, 1977)

relation $y = G_p^*(u)$ is bounded and asymptotically stable in the range of values of interest, such that y reaches a steady-state value for any steady-state value of u.

The nonlinearity will express itself in several ways. First, the form and magnitude of the response of $y(t)$ to a step change in $u(t)$ will depend on the magnitude of such a step. Second, the function $y = G_p'(u)$ will also depend on other inputs. The second dependence can be expressed as a dependence of $G_p'(u)$ on time, provided the rate of changes in G_p' are slow compared with the time scale of $G_p'(u)$ itself.

Let us first examine the case where the steady-state changes but the perturbations in y and u are so small that linearization around each steady state can be justified. For each steady state a different G_p' is obtained, and together these form a set $G_p'^j$. For each member in $G_p'^j$, a design function, G_{pd}^j can be found that fulfills conditions of Eqs. 21 and 22. In many cases the members of the set will be very similar, with only the constants changing. Then an overall design function G_{pd}^* may be formulated so as to fulfill conditions of Eqs. 21 and 22 for the set G_{pd}^j (and clearly for $G_p'^j$). Satisfying these conditions will sometimes harm performance, but the designer often has a reasonable idea how much G_p' changes.

Here, we are still in the domain of accepted classical control theory. However, when the deviations in y and u from steady state are large enough such that nonlinearities become important, the case is more difficult. (An example is given in Figure 8 from Silverstein and Shinnar, 1977.) The fact that in such cases controllers designed by linear process identification and design methods perform very satisfactorily despite the strong nonlinearities of the system can be explained as follows. Let us look at a series of large control actions $u(t)$ and the resulting trajectories $y(t)$. For an open-loop globally stable system a set of linear functions $\overline{G}_p^j$ that approximates each trajectory $y(u,t)$ as:

$$y(t) = \overline{G}_p^j(u(t))$$

can be found.

Again, in many cases the set $\overline{G}_p^j$ will have very similar properties and can be approximated by a set $\overline{G}_{pd}^j$ in which only the constants change. This method is not a general one. The nonlinear system must have certain properties to allow a set G_{pd} to be found that is a good approximation of the phase-space trajectories of the nonlinear system.

Here, the cases are limited to those in which the nonlinear system itself is stable. But even more important, the system must have a natural dampening capacity for high-frequency inputs. This prevents the nonlinear system from generating higher harmonics with a large amplitude. Otherwise it would be impossible to find a reasonable linear approximation for the real response to a step input. Fortunately, this is a property of many chemical reactor and process units in which high-frequency disturbances are damped by heat and mass-transfer processes. It is not a general property of nonlinear systems. Generally, the approach discussed here is not used for reactors or units that lack that property. For more complex cases linear design methods will not surface.

Again, if we choose a $\overline{G}_{pd}$ such that Eqs. 21 and 22 are fulfilled for the set $\overline{G}_{pd}^j$, then the controller will perform satisfactorily over the whole space of u and y. For complex systems, satisfactory performance is achievable only over a limited range of u and y. There are obviously nonlinear systems for which even that is not possible. But, in the chemical process industries, high frequencies are normally damped, limiting the effect of higher harmonics and often allowing the approximation of nonlinear responses with pseudolinear trajectories. The latter is extremely similar to, though slightly more general than the approach taken by Rosenbrock (1974, p. 103), and we thus have a heuristic criterion for the limitations of linear controller design for nonlinear systems. Consider, for example, a process similar to that described in Figure 4, in which the throughput changes. This variation will change the time parameters of the system, and $|G_p'(j\omega)|$ will move either to the right or left. The only concern is decreases in ω_{co}.

In very extreme cases such as reactors with multiple steady states, even the sign of the steady-state gain of $G_p'(u)$ will change. Good process modeling is helpful in pointing out the existence of such possibilities; fortunately, they are not that common. For a limited range of changes in operating conditions, G_p' will vary but in a reasonably predictable way. For example, changes in throughput will change the time constants and the crossover frequency. Again we can demand that G_{pd} is chosen such that all forms and values of the set $\{G_p'\}$, which now includes both changes in operating conditions and level of response, fulfill Eqs. 21 and 22. The wider the possible changes in $\{G_p'\}$ (or the larger the set in model space), the worse will be the performance of the controller compared with one designed for a single form of G_p'.

A worst G_{pd} can therefore be defined by measuring the response at the highest throughput expected. A relation between the critical values and throughput would also be attained and the controller automatically adjusted accordingly.

The advantage of the *PI* controller is that Eqs. 21 and 22 are one-sided and require minimum information with respect to the phase angle.

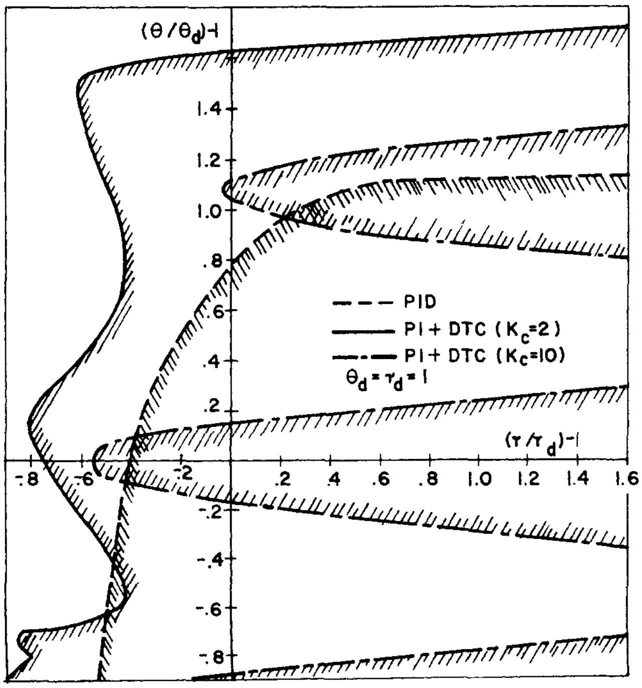

Figure 9. Stability Limits for *PI* plus *DTC* and *PID* Controllers for Process Parameter Variation

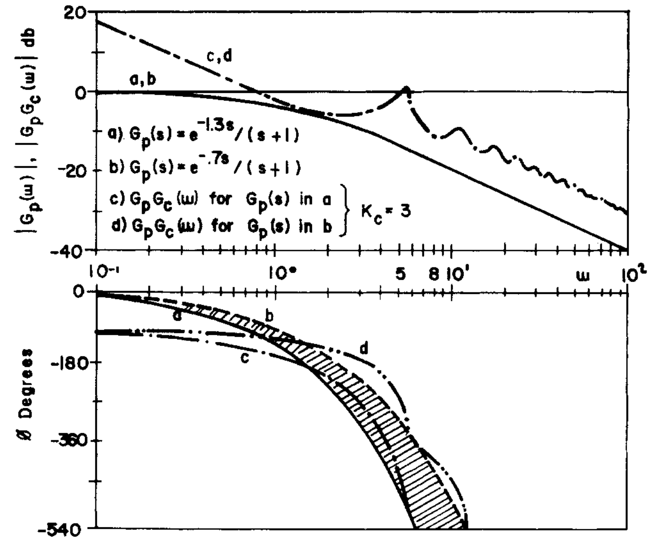

Figure 10. Bounds for G_p', Determined by Margins in θ (Controller (7))

SENSITIVITY TO PARAMETER CHANGES IN G_{pd}

One standard method used to check the sensitivity of the controller to changes in the transfer function is to choose a specific G_{pd} and check the sensitivity of performance and stability to relatively large changes in the parameters of G_{pd}. For example, if we choose the simple two-parameter function given in Eq. 19, we can vary both τ and θ and get stability limits in τ and θ. For a *PID* controller these limits are one-sided, Figure 9. For a given specific controller, there is a relation between τ and θ such that for any given θ, τ must be larger the $\tau_{\min}$; and vice versa, for a given τ of the system, θ must be less than $\theta_{\max}$.

Now let us remember that the real G_p' is not given by Eq. 19. A reasonable sensitivity to parameters means that there is a whole space of possible G_p' different in order and form from G_{pd} that is also stable with the controller designed for the nominal process model. Conditions 21 and 22 specify this exactly. It is easily seen that in this case there is a certain equivalence between parameter sensitivity and model sensitivity. However, if the model is not accurate, it does not help that the parameters of G_{pd} can be measured accurately, as this will not guarantee good performance. A reasonably wide range of parameters of G_{pd} for which the system is performing well will also ensure that this type of controller will perform well over a wider range of transfer functions if they are similar to G_{pd} in the range of frequencies for which G_p' is not strongly damped.

For more complex controllers (or controllers that amplify beyond the first ω_{co}), more detailed information of G_p' is needed than is in the first example. The dead-time compensator operating on an overdamped system is a good example.

It was noted earlier that with an unconstrained dead-time compensator both $|G_p(j\omega)|$ and $|G_cG_p(j\omega)|$ have a large number of peaks at high frequencies for high gains, Figures 3a and 3b. If K_c in Eq. 7 or 8 is reduced, the number of peaks decreases. If K_c is reduced enough, all peaks (aside from peak at $\omega \to 0$, which is due to the integral mode) will be eliminated. If the second peak is reduced sufficiently below unity, the result is equivalent to a *PI* controller. A function G_{pd} can then be defined such that the criteria given in Eqs. 21 and 22 guarantee stability for all transfer functions that fulfill it. The dead-time compensator with $K_c = 2$ and the transfer function (Eq. 19) permit a set of transfer functions for which criteria of Eqs. 21 and 22 guarantee stability.

However, the efficiency of the dead-time compensator is then also reduced. It is clear that a large number of peaks cannot be allowed to occur at high frequencies, as there is no way the required accuracy in G_{pd} could be obtained. Even if it could be, the fact that G_p' changes with operating conditions would make such a close fit meaningless. The most that could be practically afforded is one such peak. Figure 10 gives the frequency response $G_pG_c(j\omega)$ of a *PI* plus *DTC* with $K_c = 3$ designed for the process specified by Equation (19). The peak reaches unity close to $\phi_{G_pG_c} = -2\pi$.

To achieve a reasonable stability margin, transfer functions can be allowed only for which $|G_cG_p(\omega)|$ is $0.5 - 0.6$ at any ω where $\phi = -(\pi + 2n\pi)$. Under these circumstances Eq. 22 is not sufficient to guarantee stability, and Eq. 21 is still necessary as an upper bound on $|G_cG_p(j\omega)|$ in the region of frequencies of interest is needed. Condition 22 must be replaced by a more stringent one. Thus $\phi_{G_p} > \phi_{G_{pd}}$ no longer ensures stability. For example, if ϕ at $\omega = 5$ turns out to be $-\pi$, the system has no stability margin, despite the fact that condition 22 is fulfilled. Condition 22 must therefore be changed to a two-sided condition. A simple way to get such a condition in this case is to vary the *DT* in the design model, G_{pd}, in both directions.

A bounded region arrived at in this way is shown in Figure 10. Thus, any $G_p'(s)$, for which condition 21 holds and the phase angle for $\omega < 8$ falls between curve a and curve b in Figure 10 is stable for the dead-time compensator. Curves c and d in Figure 10 are the open-loop frequency responses, $G_cG_p(j\omega)$, for the controller ($K_c = 3$), with each of the two models being used to generate the region bounded by the phase-lag curves a and b. Thus, the phase lags, $\phi G_p'G_c$, for any $G_p'(s)$ fulfilling the two-sided criterion falls between the phase lags given by curves c and d. If, in addition, condition 21 is satisfied, the overall design for any $G_p'(s)$ will be stable with appropriate stability margins.

Note that more precise knowledge of $G_p'(\omega)$ for a wider range of frequencies is required here compared with the *PI* controller alone. A *PID* controller requires a more detailed knowledge of the process than does a *PI* controller and the demands for more information are even more extensive for the *DTC*. Note, too, that there are other ways to constrain a *DTC*. If detailed information is known about G_p' for a frequency range beyond the first peak, a higher peak could have been allowed and all subsequent peaks cut off by means of a filter. The height of the peak that can be allowed depends greatly on the quality of the information on G_p^*, and this quality varies highly from case to case. It also depends on how much G_p' itself changes with time because of changes in operating conditions. Remember that a far better knowledge of G_p^* and G_p' is needed that when a *PI* or *PID* controller is used.

The stability analysis here is presented in given terms of the Bode diagrams. It can also be given in terms of the Nyquist or the inverse Nyquist diagram. (See, for example, Figure 11

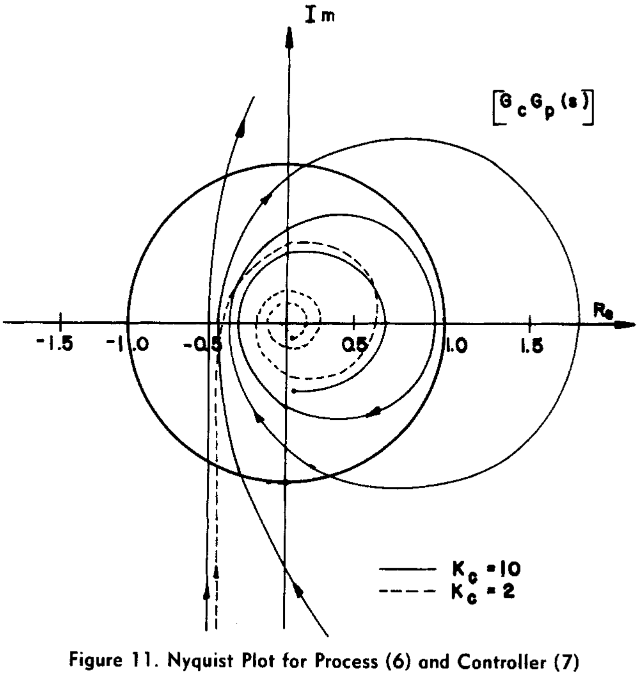

Figure 11. Nyquist Plot for Process (6) and Controller (7)

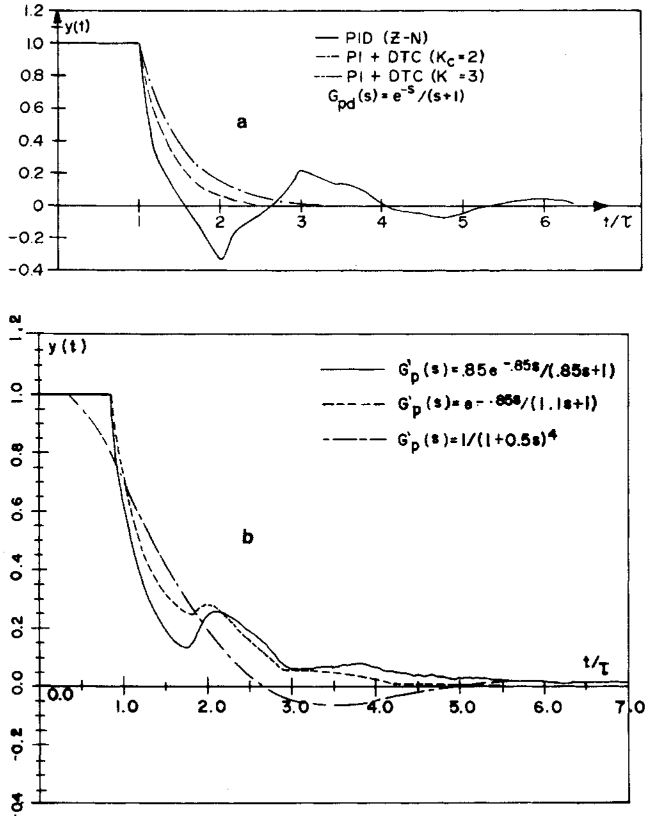

Figure 12. a. Responses to Unit Step Change in Set Point of the *PID* and *PI* + *DTC* Controllers ($G'_p + G_{pd}$) b. Responses to Unit Step Change in Set Point of the *PI* + *DTC* Controller ($G'_p \neq G_{pd}$)

which is a replot of Figure 2b). We can now include in the conventional Nyquist criterion (for open-loop stable system) a practical additional requirement. The Nyquist plot $G_pG_c(j\omega)$ must stay inside the unit circle for all frequencies above the range for which G'_p is not accurately known. Furthermore, this means that for a practical control system, $G_pG_c(j\omega)$ must converge to a region inside a circle with a radius ρ where $\rho < 1$ and is determined by safety criteria.

There are several questions that must be asked:

1. Is G'_p measurable with sufficient accuracy?
2. How much does G'_p vary with nonlinearities and changes in process throughput and other operating conditions?
3. How much improvement in performance can be obtained?

Figure 12a gives the response of the controller with a *DTC* to a setpoint change for the exact model and compares it with the response of a system using only a *PID* controller. The *PI* plus *DTC* controller using $K_c = 2$ gives a far better response than the *PID* controller. This is important, as the controller with $K_c = 3$ is an improvement and is not that far from the performance of the unconstrained controller (a delayed unit step), which is obviously not permissible.

One must also ask how much would the performance deteriorate if the transfer function changes or is different from the one used in the design. Figure 12b gives additional several step responses of the *DTC* for which $G'_p \neq G_{pd}$. All of the G'_p chosen fulfill the criteria given for stability, and the response to a set-point change is still significantly better than for a *PID* controller. The range of permitted transfer functions is, however, rather narrow for $K_c = 3$ and much wider for $K_c = 2$. The main advantage of the dead-time compensator in this instance is a smoother response with low overshoot. If one wants to tune the *PID* controller with no overshoot, it would be significantly slower.

On the other hand, if conventional criteria, such as rise time or band width are used, the difference will be much smaller. The advantage will depend on the specific case and is not the main subject of this paper. In Figure 12b the examples for $G'_p \neq G_{pd}$ are for nonlinearities that express themselves in a change of the parameters or the order of the linearized transfer function. Figure 13a gives the response to a set-point change of a simple nonlinear system: stirred tank reactor with a second-order reaction and delayed input, which has the equation

$$\tau \frac{dy}{dt} = u(t - \theta) - y - k\tau y^2 \tag{25}$$

When linearized and transformed, this becomes

$$\tau s y(s) = u(s)e^{-\theta s} - y(s) - 2k\tau y_s s(s) \tag{26}$$

The linearized transfer function is

$$\frac{y(s)}{u(s)} = \frac{\dfrac{1}{1 + 2k\tau y_s}}{1 + \dfrac{\tau}{1 + 2k\tau y_s}} e^{-\theta s}$$

If we set $\theta = 1$; $\tau = 5$; $2k = 4$; $u_s = 0.6$ and $y = 0.2$, $\tau/1 + 2k\tau y_s = 1$ and we get the same linearized transfer function as in Figure 12a, but with $K_p = 1/5$). Here, limits on the possible changes in y_s and y for set-point changes and operating conditions must be specified.

If they are specified as $0.12 < y_s < 0.28$ and $0.10 < y < 0.30$ they can be used to obtain a set of G'_p. It will be found that for the G_{pd} under discussion the controller chosen is stable for the set G'_p. The response is given in Figure 13a for a set-point change and is compared to that of the linear case. Again, we would have to include a stability margin for changes in throughput. The response in Figure 13a for $K_cK_p = 3$ is satisfactory and the controller with this gain and with the nonlinear system is not more sensitive to inaccuracies in the estimated *DT* than with the linear one as can be seen from Figure 13b.

Testing final designs by simulation using a nonlinear transfer function is desirable for complex systems and can give valuable insights about the sensitivity of the design. However, one should remember that nonlinear transfer functions are also approximations. A very simple example was used here to illustrate the basic principle that advanced design methods such as *DTC* require better identification of the process model than conventional controllers. Also, the quality of information required was discussed. The same problem can be looked at in terms of variations in parameter space.

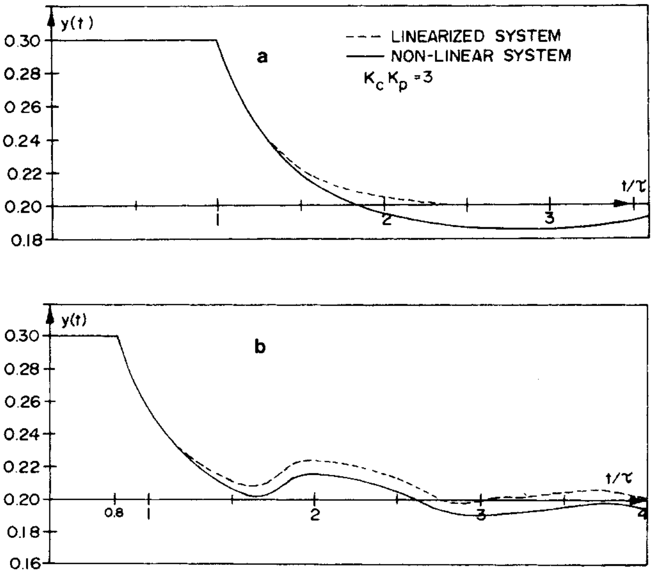

Figure 13. Step Responses of the Nonlinear and Linear Systems (Eqs. 25 and 26) Controlled by the *PI* + *DTC* a. Exact Parameters b. −20% Error in the Estimated *DT*.

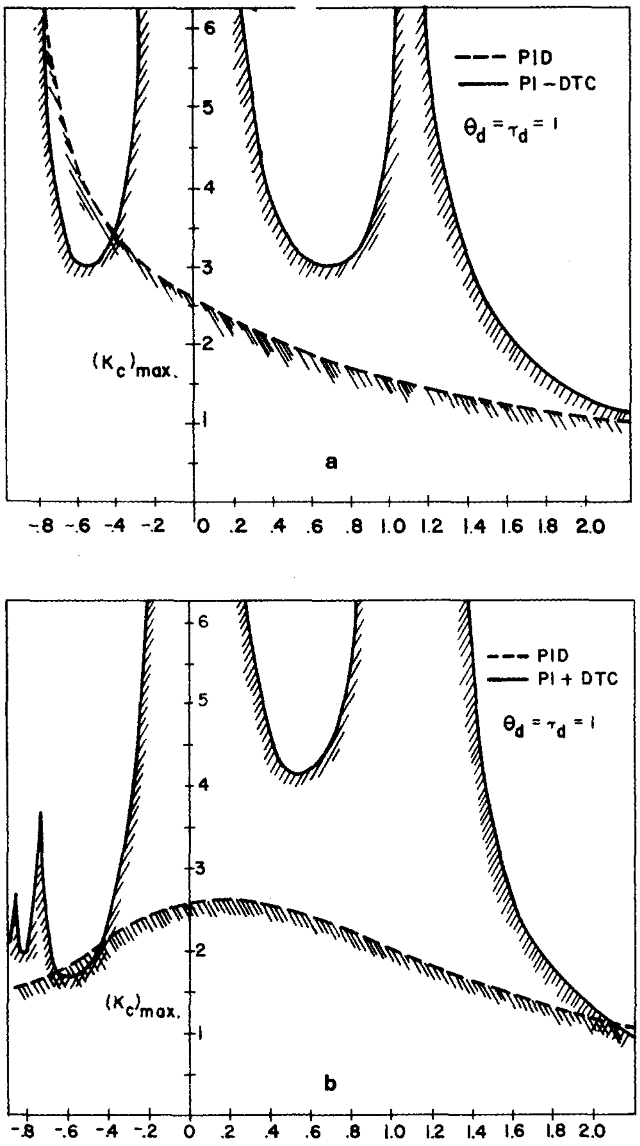

Figure 14. a. Maximum Gain as Function of Variations in Dead Time b. Maximum Gain as Function of Throughput Variations

If the model is simple and, as here, has only two parameters, the stability limits for a given controller can be drawn in the plane $\theta - \tau$. These limits are shown for various values of K_c in Figure 9. The stable region is more complex and narrower than in the case of the *PID* controller. When $K_c = 2$ the stability range is wider than with a *PID* controller, whereas when $K_c = 10$ the range is complex and very narrow. Figures 14a and 14b present those limits along two lines in the plane $\theta - \tau$. Figure 14a goes along the coordinate $\beta = \theta/\theta_d - 1$, which is the relative change in the dead time. It gives the permissible errors in θ, the *DT* (in terms of β) for a given K_c, or the maximum permissible K_c if a tolerance to a certain variability of θ only is demanded. In practice only half these ranges could be used to ensure reasonable safe margins. As noted before, stability margins in terms of gain or phase margins are meaningless here, but a margin in θ has a definite meaning in terms of the sensitivity of the controller to variations in the transfer function. The curve in Figure 14a has a more general applicability. It applies to any G_{pd} if the controller is designed using Eq. 8 and therefore if Eq. 7 is used instead) but K_c has to be replaced by $K_c\theta$. Note that the value $K_c = 3$ is the value for which the system becomes insensitive to changes in θ.

But as a safety margin in the gain is also needed either $K_c\theta$ must be reduced further or the method outlined earlier to make sure that G_p' is sufficiently well defined in the region of ω of interest. Figure 14b gives the result along the diagonal in the plane $\theta/\theta_d - 1$, and $\tau/\tau_d - 1$. This diagonal refers to simultaneous changes in θ and τ, designated by a parameter α given by $\alpha = \theta/\theta_d - 1 = \tau/\tau_d - 1$. Changes along this line approximately described changes in G_p' due to variations in throughput. Although the sensitivity of the system is symmetric for both positive and negative changes in θ, it is much more sensitive to negative changes in α, which is understandable if one looks at Figure 3b. Changes in θ move only the phase lag, and in this case the symmetric sensitivity is due to the fact that the phase lag can be moved either $-\pi$ or -3π to the frequency region of the first peak. In cases of multiple peaks ($K_c > 3$) the changes in θ will usually move the larger phase lags, $-(2n + 1)\pi$, to the frequencies of the further peaks. Changing α moves both the phase angle and the frequency at which the peak occurs. A negative α increases the range of frequencies for which $|G_cG_p(\omega)|$ can exceed unity.

Palmor and Shinnar (1979) have shown that curves like that in Figures 14a and 14b can be used to ensure stable performance of the controller. Those cases however, dealt with sampled data control with infrequent sampling and showed that most transfer functions can be approximated by a second-order model plus a *DT* with a very small error, provided the sampling interval is large enough. Here the case is different, as the model fit cannot be improved in a similar way.

In some sense, checking parameter sensitivity is justified here, too, since the parameters of G_p' change with operating conditions even if the structure of the model is valid. In a broader sense, it has been shown here how reasonable insensitivity to parameter variation also ensures a tolerance to a deviation between G_{pd} and G_p^*. This is important because the problem that $G_{pd} \neq G_p' \neq G_p^*$ exists simultaneously with the problem of parameter variation. With a *PI* or *PID* controller these problems can often be avoided because the model space is wide and it is easy to formulate methods for choosing G_{pd} such that the design is robust. In the *DTC* it has been shown that the choice was either to choose a robust design or to improve our knowledge about G_p^* such that an improved controller design could be used.

The approach taken here also clearly delineates what precision in the model in the frequency domain is required for a specific performance. The Bode diagram was used here since in this kind of description it is easier to translate the information about G_p' into physical models. It is a straightforward task to translate these conditions into the polar plane.

SUMMARY AND CONCLUSIONS

The focus of this paper was to show a systematic approach to advance controller design suitable for process control systems in which the transfer function is normally more complex and harder to define than in servomechanism or electronic system design. The approach was demonstrated by a very simple example for didactic purposes. The basic principle examined is that in advanced controller design for the chemical process industries one cannot rely on conventional stability analysis to give a robust controller design. Most modern algorithms guarantee stability if the plant is exactly equal to the desired transfer function, but they give little quantitative information about the stability margin when the real nonlinear plant deviates from the linearized transfer function used. Gain and phase margins are valid and useful criteria for *PID* controllers but insufficient or misleading (if not properly interpreted) for more complex compensators.

The main reason for this problem is that to obtain good performance most advanced design methods try to achieve the equivalent of a inversion of the transfer function. This is a sensible policy for design of the controller itself. What is needed for successful design is a different approach to stability analysis. Although the design presented here is based on the most likely model of the plant, stability analysis has to include the confidence limits of our knowledge about the real behavior of the plant. Modern controller design therefore requires not only good identification but also a knowledge about the variability of plant conditions and an understanding about the uncertainties of the process identification and the effect of nonlinearities. Both the uncertainties in the model and the variability with time change from case to case but are generally larger in process industries than in servomechanisms.

An example was given to demonstrate some potential approaches to the general problem that allow a quantitative definition of both model uncertainty and plant variability. It was also shown that for many cases one can find a suitable model in which these two uncertainties can be defined by a suitable parameter space for the basic model itself. The design will be made for the most likely values of these parameters, but it should function well for the whole parameter space. For other cases other definitions of plant variability might be more useful.

Each design method has to be matched to an identification method, and if one wants to use more advanced controller designs successfully and safely, we have to understand the way improved information leads to better design. Until recently, controller design for process control was limited by the types of controller available. Today, with microprocessors about any complex control function can be designed at almost no penalty or cost, and coupled with better process identification superior performance will likely result.

Interactive graphics terminals allow complex functions to be dealt with more easily than before. The problem is to derive proper design methods. One possible approach to systematically evaluate the information required for design of dead-time compensators was demonstrated. Fixed tuning prescriptions were specifically avoided. There are quite a number of ways in which the same goal can be achieved. What was wanted here was to show a way in which the quality of the information required about the process model can be evaluated systematically. Advanced controller design should be an iterative and interactive process in terms of controller design and in terms of system identification. In any good design procedure the cost of obtaining additional information should be evaluated and compared with the value of the improved dynamic performance obtainable.

Most experienced designers do this intuitively based on their experience. It is a challenge for academic engineering research to attack this basic problem of design and to try to evaluate systematically the value of increased information in design.

For example, the results about the *PID* controller presented here show that in many cases the information required to design a well-functioning *PID* controller is more limited than when a dead-time compensator is used. In some cases the dead-time compensator designed with proper process identification that takes into account the variability of the process will result in superior performance.

The concept of model space presented in this paper can be used in a dual way, as a guide in thinking about the behavior of the system and as a quantitative tool in system identification by providing bounds on the accuracy of process model for specific controllers.

In several previous papers the authors have commented on the shortcomings of optimal control algorithms if inappropriately used. Once their pitfalls are understood, they can be powerful tools in suggesting improved algorithms. However, as shown here, they cannot be used as straight algorithms. They must be tested and evaluated by methods that are outside the algorithms themselves. This paper presented some of the reasons and suggested a general approach to overcome these difficulties. The approach is not limited to the *DTC* or other optimal algorithms, but includes any complex control strategy applied to imperfectly known processes.

APPENDIX

Proof is given here that the overall closed-loop system consisting of the process (Eq. 1) and of the controller (Eq. 8) with any gain K_c has a gain margin (*g. m.*) larger than 2 and a phase margin (*p. m.*) of at least 60°.

The open-loop transfer function of the system is

$$G_cG_p(s) = \frac{K_c e^{-\theta s}}{s + K_c - K_c e^{-\theta s}} = \frac{e^{-\theta s}}{1 + \frac{s}{K_c} - e^{-\theta s}} \tag{A.1}$$

First, it is shown that Eq. A.1 is stable for any finite K_c. The R.H.P. of the complex s plane is enclosed by a simple closed curve, c (which, for example, may consist of the imaginary axis $j\omega = jIm(s)$, $-\infty < \omega < \infty$, $\omega \neq 0$, and a large semicircle $s = Re^{-j\phi}$ where $R \to \infty$ and $\pi/2 < \phi < -\pi/2$) since Eq. A.1 has a pole at $s = 0$, this point is excluded from curve c. Both $1 + s/K_c$ and $e^{-\theta s}$ are analytic in and on curve c. However, $|1 + s/K_c| > 1$ and $|e^{-\theta s}| \leq 1$ for all s on c. That is, $|1 + s/K_c| > |e^{-\theta s}|$ on curve c. Applying Rouche' theorem, we immediately conclude that Eq. A.1 is stable. (Clearly, when $K_c \to \infty$ all the poles of Eq. A.1 lie on the $j\omega$ axis.)

To determine the stability properties of the closed loop, it therefore suffices to examine the modified Nyquist plot, i.e., the mapping $G_cG_p(j\omega)$ in the $G_cG_p(s)$ complex plane. Thus, replacing s in Eq. A.1 by $j\omega$, we get

$$G_cG_p(j\omega) = \frac{e^{-\theta\omega j}}{1 + j\frac{\omega}{K_c} - e^{-\theta\omega j}} = \frac{e^{-\theta\omega j}}{1 - \text{Cos}\theta\omega + j\left(\frac{\omega}{K} + \text{Sin}\theta\omega\right)} \tag{A.2}$$

Denoting the denominator of Eq. A.2 by $D(j\omega)$, this Equation may be written

$$G_cG_p(j\omega) = \frac{e^{-\theta\omega j}D^*(j\omega)}{|D(j\omega)|^2} \tag{A.3}$$

where $D^*(j\omega)$ is the complex conjugate of $D(j\omega)$. It can be easily verified that

$$|D(j\omega)|^2 = 2\left(1 - \text{Cos}\theta\omega - \frac{\omega}{K_c}\text{Sin}\theta\omega\right) + \frac{\omega^2}{K_c^2} \tag{A.4}$$

and that

$$Re\{e^{-\theta\omega j}D^*(j\omega)\} = -\left(1 - \text{Cos}\theta\omega - \frac{\omega}{K_c}\text{Sin}\theta\omega\right) - \frac{1}{2}\frac{\omega^2}{K_c^2} + \frac{1}{2}\frac{\omega^2}{K_c^2} \tag{A.5}$$

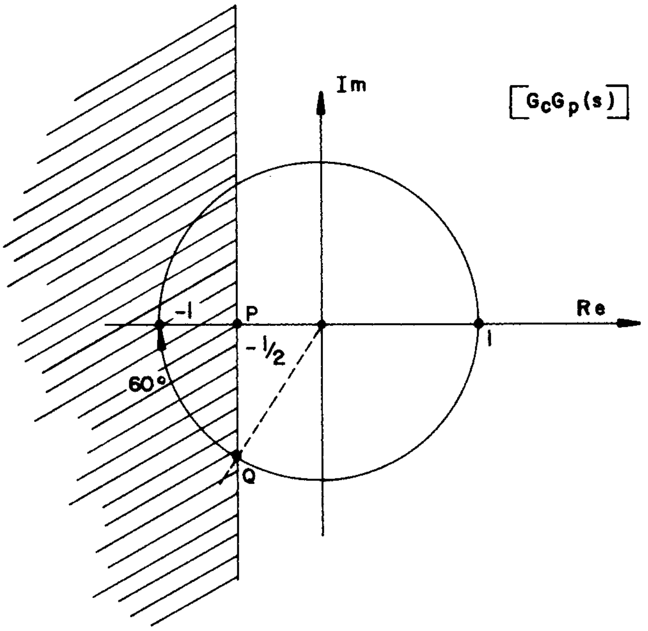

Figure 15. Points and Regions in the $\{G_cG_p(s)\}$ Plane Illustrating the Proof in the Appendix

Substitution of Eqs. A.4 and A.5 into Eq. A.3 results in

$$Re\{G_cG_p(j\omega)\} = \frac{1}{2} + \frac{\frac{1}{2}(\omega^2/K_c^2)}{|D(j\omega)|^2} \geq -\frac{1}{2} \quad \text{(A.6)}$$

Equation A.6 indicates that the Nyquist plot, $G_cG_c(j\omega)$, avoids the region $Re\{G_cG_p(s)\} \leq -1/2$, Figure 15. This means that $G_cG_p(j\omega)$ may cross the negative real axis in the $G_cG_p(s)$ plane to the right of $-1/2 + jo$ only. It is therefore evident that for any K_c the closed-loop system is stable with the $g.m. \geq 2$.

Recall that the phase margin is the minimum amount of negative phase shift that must be introduced to make the points on the Nyquist plot having unit magnitude pass through $-1 + jo$. Since all Nyquist plots are located to the right of the straight line $Re\{G_cG_p(s)\} = -1/2$, Figure 15, the worst $p.m.$ arises when the Nyquist plot passes through the point Q in Figure 15. Hence, the $p.m. \geq 60°$.

For the infinite gain cases, Eq. A.6 reduces to

$$Re\{G_pG_c(j\omega)\} = -\frac{1}{2} \quad \text{(A.7)}$$

The Nyquist plot is simply infinite times the straight line in Eq. A.7. In the infinite gain cases, the closed-loop systems have a $g.m. = 2$ and a $p.m. = 60°$.

NOTATIONS

- e = error signal
- DT = dead time
- DTC = dead-time compensator
- $g.m.$ = gain margin
- G_c = overall controller (G_{co} coupled with DTC) transfer function
- G_{co} = controller transfer function
- G_p = process transfer function
- G_{po} = invertible part of G_p
- G_p^* = true process transfer function
- G_p' = linearization of G_p^*
- G_{pd} = process model used for design
- $\{G_p\}$; $G_p'^j$; $\bar{G}_p^j$ = sets of linearized process models
- K_c = controller gain
- n = disturbance
- $p.m.$ = phase margin
- r = set point
- s = laplace operator
- t = time
- u = control action
- y = output

Greek Letters

- α = relative change in throughput
- β = relative change DT
- θ = dead time
- θ_d = design value of θ
- τ = time constant
- τ_d = design value of τ
- τ_I = integral time
- τ_D = derivative time
- ϕ = phase lag
- ω = frequency
- ω_{co} = phase crossover frequency
- ω_g = gain crossover frequency
- ω_∞ = frequency at infinite peaks

ACKNOWLEDGMENT

The work of Z. J. Palmor was done while staying with Taylor Instrument Co., Rochester, New York during Summer, 1979. The support and involvement of the Engineering Division and the Systems Engineering Department are greatly appreciated.

LITERATURE CITED

Alevisakis, G., and D. C. Seborg, "Control of Multivariables Systems Containing Time Delays using a Multivariables Smith Predictor," *Chem. Eng. Sci.*, **29**, p. 373 (1974).

Astrom, K. J., "Frequency Domain Properties of Otto Smith Regulators," *Int. J. of Control*, **36**, No. 2, p. 307 (August, 1977).

Astrom, K. J., and B. Wittenmark, "On Self Turning Regulators," *Automatica* 9 (1973).

Bucklly, P. S., "Automatic Control of Processes with Dead Time," Proc. 1st IFAC Congress, Moscow (1960).

Cohen, J. H. and G. A. Coon, "Theoretical Consideration of Retarded Control," *Transactions of the ASME*, p. 827 (July, 1953).

Donoghue, J. F., "A Comparison of the Smith Predictor and Optimal Design Approaches for Systems with Delay in the Control," *IEEE Transactions on IECI*, **24**, No. 1, p. 109 (Feb. 1977).

Edgar, T. F., "Status of Design Methods for Multivariable Control," *AIChE Symp. Ser.* **72**, No. 159, p. 99 (1976).

Gould, I. A., *Chemical Process Control: Theory and Application*, Addison-Wesley Publishing Co. (1969).

Hougen, J. O., "Measurement and Control Applications," ISA (1979).

Kestenbaum, A., R. Shinnar, and F. F. Thau, "Design Concepts for Process Control," I&EC Process Design and Development, **15**, p. 2 (Jan., 1976).

Koppel, L. B., *Introduction to Control Theory*, Prentice-Hall Inc. (1968).

Lupfer, D. E., and H. W. Oglesby, "Applying Dead Time Compensation for Linear Predictor Process Control," *ISA Transactions*, **8**, No. 11, p. 53 (1961).

Meyer, C., D. E. Seborg, and R. K. Wood, "A Comparison of the Smith Predictor and Conventional Feedback Control," *Chem. Eng. Sci.*, **31**, pp. 775-778 (1976).

* Nielsen Reference Omitted.

Palmor, Z. J., and R. Shinnar, "Design of Sampled Data Controllers," *Ind. Eng. Chem. Process Des. Dev.*, **18**, No. 1, pp. 8-30 (1979).

Palmor, Z. J., "Structural, Stability and Sensitivity Properties of Optimal Stochastic Control Systems for Dead Time Processes," 1980a, to appear in *Automatica*.

Palmor, Z. J., "Stability Properties of Smith Dead Time Compensator Controllers," 1980b, to appear in *Int. J. of Control*.

Ragazzini, J. R., and G. F. Franklin, "Sampled-Data Control System," McGraw-Hill, Inc., p. 157, (1959).

Rosenbrock, H. H., "Computer-Aided Control System Design," Academic Press, 1974. G. W. Ross, *ISA Transactions*, **16**, No. 3, p. 25 (1977).

Shinnar, R., "The Interface between Industry and the Academic World," *AIChE Symposium Series, No. 159*, **72**, pp. 155, 166 (1976).

Shinnar, R., "Chemical Reactor Modeling—the Desirable and the Achievable," *ACS Symp. Ser.*, **72** (1978).

Silverstein, J. L., and R. Shinnar, "The Effect of Design on the Stability and Control of a Hydrocracker," AIChE Meeting, New York (1977).

Smith, O. J. M., "A Controller to Overcome Dead time," *ISA J.*, **6**, No. 2, p. 28 (1959).

Ziegler, J. G. and N. B. Nichols, "A Controller to Overcome Dead Time," *Transactions of the ASME* p. 759 (Nov., 1942).

Manuscript received October 24, 1979; revision received October 30, 1980 and accepted November 21, 1980.