AUTOMATIC PROCESS CONTROL

Ernest F. Johnson

Department of Chemical Engineering
Princeton University, Princeton, New Jersey

I. Introduction

A. Scope

The field of automatic process control has only recently come to the attention of the chemical engineer as an area wherein a quantitative treatment offers interesting possibilities for technological advance. This brief survey undertakes to orient automatic process control in the chemical engineering picture and to identify and evaluate the possibilities of intelligent development of the field. The important broad concepts are presented in a general description of the nature of control. These concepts are applied to the solution of typical process control problems in an elucidation of available techniques.

A brief history of the development of modern control theory is included to provide some indication of what kind of help may be obtained from developments in other control areas such as the field of servomechanisms. The final section of the chapter deals with the future trends in the field of automatic process control and how chemical engineers can contribute materially to the acceleration of these trends.

B. Control as a Basic Concept in Chemical Engineering

Chemical engineering deals with the design, development, and operation of processes and plants in which bulk raw materials are converted by chemical and physical means into bulk products of varying degrees of finish. The purely technical problems of design and analysis of processes and equipment are solved readily on the basis of a few fundamental concepts. If we regard smooth and stable operation of process plants as an important quality of a successful design, we must regard the general concept of feedback control as one of our basic concepts, coequal in importance with such concepts as those of equilibrium, conservation, and rate processes.

The technical side of chemical engineering may be contemplated most broadly in terms of the most immediately creative aspect of the field, namely, the design of a plant. We may regard such design as involving a succession of integrating steps in which first of all the basic properties of matter are used in conjunction with basic rate expressions to describe the component rate processes; next the rate equations are combined with the constraints imposed by thermodynamics, e.g., conservation of mass, momentum, and energy, and chemical and physical equilibrium to permit sizing the process units; and finally the process units are fitted into the overall plant sequence with such control instrumentation as is necessary for satisfactory operating performance. In this picture of a plant design there are four distinct basic concepts: *conservation*, *equilibrium*, *rate process*, and *control*. Each must be considered in maximizing the economic gain to be realized from the plant. A sound understanding of these concepts and how they are applied is essential to successful design.

C. Automatic Control in the Process Industries

A typical characteristic of much of the chemical process industry is that the operating plants appear to require little human supervision. The process units are often extensively instrumented and the desired operating conditions are maintained rigidly by arrays of control instruments. Automatic control is clearly an important feature of modern processing.

The process industries were among the first to make widespread use of automatic controls with the attendant advantages of continuous operation, improved yields and product quality; and the general adaptability of the many processes and operations to automatic regulation has contributed greatly to the rapid growth of the industry.

Because of this ready adaptability of the processes and process equipment to automatic control there has been little incentive to develop a sound quantitative theory of automatic process control. Chemical engineering ingenuity has been applied to more immediately profitable process problems. Processes involving difficult control problems have been rejected in favor of easily controllable processes without any concerted effort to solve the control problem. As a result, chemical engineers have made little contribution to control theory, and even today most chemical engineers regard control problems as lying a little outside their field.

However it is obvious in many operations that improved overall performance can result from a clearer understanding of the control problems. Furthermore, the trends toward greater integration of operations and the exploitation of faster processes require that the control problems be handled in a quantitative manner by process engineers. The nature of control problems and what is involved in handling them, are discussed in Parts II and III.

II. Nature of Control and Control Problems

A. Definition

In the broadest sense *control* is *regulation for some purpose.* Automatic process control is the mechanical (or other non-manual) regulation of processes for the purpose of producing high quality products at high rates and low cost. Generally this regulation involves holding process conditions relatively constant in the face of external disturbances. By contrast, the control action in the field of servomechanisms is designed to maintain a close agreement between a varying desired behavior and the actual behavior, as for example, in steering devices.

Control, of course, is not limited to machines and process units but is a significant feature of living organisms, corporate enterprises, and sociological organizations. Regardless of the nature of the control system the basic characteristics are the same, and the broad principles of handling control problems are equally applicable. Only the terminology and degree of sophistication of treatment are different.

B. General Characteristics of Control Systems

1. *Information Processing*

All control systems whether they be process plants, steering systems, living organisms, or industrial corporations may be regarded as information processing systems. They obtain information, analyze it, and on the basis of their analysis generate new information to take some kind of action.

Consider the simple heat exchange process shown in Fig. 1 wherein a reaction kettle is heated by steam condensing in the jacket of the kettle. The temperature in the kettle is measured by a thermometer which is connected to a controller. The controller in turn actuates a control valve in the steam line thereby ultimately causing changes in the reactor temperature. In this example information flows from the kettle to the thermometer; from the thermometer to the controller; from the controller to the control valve; from the control valve to the kettle; and so on around the control loop. The information flows as various kinds of signals, all of which may be regarded as measures of kettle temperature either actual or potential. Thus by virtue of heat exchange between the kettle

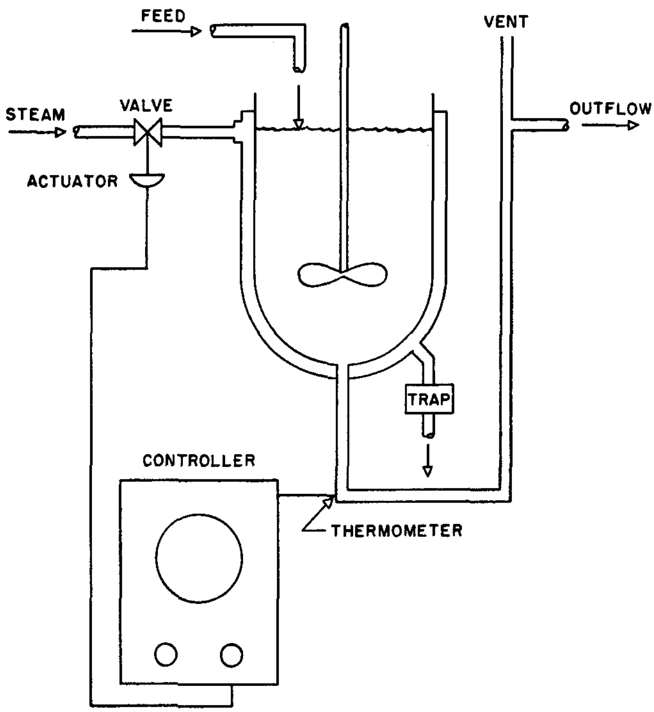

FIG. 1. Jacketed kettle control system.

contents and thermometer, the thermometer attains a temperature indicative of but not necessarily equal to the kettle temperature. Depending on the nature of the thermometer, it sends to the controller a pressure variation, a volume change, or an electrical impulse, which in the controller may be converted to mechanical displacements, air pressures or electrical signals. The controller sends out air pressures or electrical signals which become valve positions. These positions in turn moderate flows which affect steam pressures in the jacket and ultimately affect temperatures in the kettle. The signals flowing from the kettle to the controller are all measures of the actual kettle temperature; the signals flowing from the controller to the kettle are all *potential* kettle temperatures.

Since the information processing involved in control systems is the propagation of signals, the appropriate pictorial representation of such systems is the *signal flow diagram.* Such a diagram indicates not only where signals go but how they are related to each other. A simple kind of signal flow diagram for the heat exchange system just described is shown

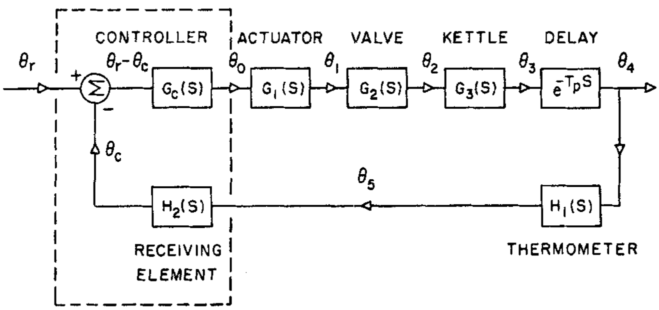

FIG. 2. Block diagram of jacketed kettle control system.

in Fig. 2. Campbell (C1) shows how primitive signal flow diagrams can be derived for process control systems and how they reveal weak process designs.

2. *Control System Components*

All process control systems comprise (a) measuring elements, (b) controlling elements, and (c) the controlled system or *plant,* as it is frequently called. The measuring elements produce signals related to the plant performance, and the controlling elements on the basis of these signals regulate the plant performance. In the example of Fig. 1, the measuring elements are shown as two blocks in the feedback path, one block for the primary measuring element which in this case is the temperature-sensing device in the kettle, and the other for the receiving element or secondary measuring element, which usually is located in the controller housing and which converts the signal from the primary element into something intelligible to the controller.

The controlling elements include the controller and its accessories and the final control element, which in this case and in most process cases is a valve. A single block represents the controlled system, which is the jacketed, agitated kettle.

In more complicated systems each of the blocks of Fig. 2 might be expanded to various multi-block diagrams.

3. *Systems Concept*

Figure 2 emphasizes an important characteristic of control systems, namely, that every component in the control circuit is important to the overall behavior of the system. It is not the size of a component, nor the amount of the flows of matter and energy to or from a component, that determines the contribution of the component to the system behavior. Rather it is the dynamic or time-dependent behavior of the component and how the various components interact with each other. An overall appraisal of the control system as an integrated unit is necessary in control technology.

Systems engineering is the name given to engineering activity which considers the overall behavior of a system, or more generally which considers all factors bearing on a problem, and the systems approach to control engineering problems is correspondingly that approach which examines the total dynamic behavior of an integrated system. It is concerned more with quality of performance than with sizes, capacities, or efficiencies, although in the most general sense systems engineering is concerned with overall, comprehensive appraisal.

4. *Feedback*

Control systems may be classified from their signal flow diagrams as either *open-loop* systems or *closed-loop* systems depending on whether the output of the primary control circuit is fed back to the controlling component. As Fig. 2 suggests, the typical control circuit consists of sequential arrays of components deployed about the process under control. If the controller is not apprised of the behavior of the controlled variable, the control system is an open-loop one. Conversely, if the measuring means on the controlled variable sends its signals back to the controller so that the behavior of the controlled variable is always under the scrutiny of the controller, the system is a closed-loop or feedback control system.

A simple example of an open-loop control system would be a steam-jacketed resin kettle very much like that in Fig. 1 except that the steam *pressure* is regulated automatically by the behavior of the measured jacket pressure but not by the actual temperature of the resin batch in the kettle. In the corresponding closed-loop system the steam pressure is regulated by the temperature of the resin batch as in Figs. 1 and 2. The only way open-loop control can be precise is through a close calibration between steam pressure and batch temperature. Since this close calibration can be maintained inexpensively only in the absence of load changes of any kind, it is obvious that the field of application of open-loop control is limited. In the example of Fig. 1, load changes would result from changes in flow rate or temperature of the feed to the kettle. Changes in either of these quantities would cause changes in the kettle temperature which would have no effect on the controller.

With feedback control, on the other hand, the controller action is dictated by the behavior of the controlled variable, and precise control is generally possible even though load changes occur.

An elaboration of open-loop control which sometimes can be competitive with closed-loop control is *feed-forward control,* in which the controller is apprised of factors which affect the key process variable to be controlled, but is not directly apprised of the behavior of this variable. Thus for the kettle of Fig. 1, it would be possible to measure the flow rate and temperature of the feed and have a simple computer analyze the information and on the basis of this information set the steam pressure to be maintained by the controller. So long as the computer considered all pertinent factors, the control would be satisfactory, but an unaccounted factor, such as a change in agitator speed, might make the control ineffectual.

5. *Stability*

Most processes or operations of importance in chemical engineering are by themselves stable, that is, in response to a small perturbation they do not run away or oscillate, but, rather, tend to level out at some new condition. When fitted into a feedback control system, however, any real process can be made to oscillate merely by sufficient amplification of the signals going around the loop. Since instability cannot be tolerated in a control system, the condition at which the system just becomes unstable is a critical limit in operating performance, and the degree of amplification required to produce the condition is a prime control characteristic of the system. This characteristic, called the *ultimate proportional gain,* provides a convenient basis for approximating a good control system design. In operating plants it is used frequently to make the final adjustments to control instruments.

C. Control Problems

1. *Analysis*

The analytic control problem deals with the appraisal of existing control systems. What is the overall dynamic behavior of the system, and what individual contributions are made to the overall behavior by the various components in the system? How may processes be characterized as to their dynamic behavior in terms useful for control system analysis? What are the control characteristics of typical instruments such as the measuring devices, controllers, and regulating units?

2. *Synthesis*

On the other hand the synthetic problem is the design of the whole control system, including in its broadest implication the design of process, as well as the specification of, control instruments. Before the synthetic problem can be tackled intelligently, the criteria of satisfactory control must be identified. These criteria are different for different systems, but most usually they are described in terms of the response of the system to certain stimuli. Having established the criteria of control, the problem of synthesis is one of optimizing the selection of control system components and their disposition in the control loop, so that the criteria are met.

III. Treatment of Control Problems

A. History

Although automatic control has existed as long as living matter has existed, control problems were not treated quantitatively and successfully much before the early 1920's.

The obvious formal approach to a quantitative treatment of automatic control is to write rigorous, general differential equations for the dynamic behavior of all the components in the system; combine these equations into a single description of the system; solve this single equation for the system response to a typical disturbance or stimulus; compare this response with the desired response; and make such adjustments in manipulable parameters as will produce the desired response. This kind of approach was applied successfully to the design of steering systems for ships and to the design of positioning systems for naval guns as early as 1922. It was also applied in the 1930's and 1940's to a variety of simple process systems. Unfortunately in the general case, this classical approach breaks down because rigorous general equations cannot be written for any but the simplest processes, or if they can be written, the overall equation becomes unwieldy and usually unsolvable. Furthermore the matching of actual response and desired response must be a trial-and-error procedure, since there are no simple means of gauging the total effect of changing the characteristics of individual components.

By the start of World War II, a new approach to control system synthesis was being developed from Nyquist's theoretical treatment (N1) of feedback amplifiers in 1932. This approach utilized the response of components and systems to steady-state sinusoidal excitation or *frequency response* as it is more usually called. The frequency response approach provides an important basis for present-day methods of handling control problems by affording a simply manipulable characterization which avoids the need for obtaining the complete solutions of system equations.

Since World War II, frequency response techniques have been applied to an increasing variety of control problems, and although the bulk of progress prior to the war came from treating servomechanisms, there has been heightened effort in process control problems since the war. The works of Rutherford (R3), Aikman (A1), and Young (Y1) in Great Britain have been notable in this respect.

A comprehensive summary with extensive bibliographies of what has been done with frequency response techniques is contained in the book edited by Oldenburger (O3), which includes in addition to a few solicited articles, all the papers presented at the international symposium on the subject sponsored by the Instruments and Regulators Division of the American Society of Mechanical Engineers in 1953. The use of frequency response in process control problems is described in an introductory fashion by Johnson in a series of two articles (J2, J3). Ceaglske's book (C4) is an elementary exposition of frequency response principles for chemical engineers. More practical but with somewhat alien terminology, is the book of Young (Y1) which describes the methods developed and used by Imperial Chemical Industries in Great Britain. Also alien and more theoretical is the short but surprisingly comprehensive book by Farrington (F1).

All of the foregoing deal with automatic process control. The great bulk of books in the field, however, are concerned more with servomechanisms of one kind or another. Among such books which are helpful for automatic process control are Brown and Campbell (B4), Chestnut and Mayer (C5), and Draper *et al.* (D1), to mention but a few. More advanced texts are those of Truxal (T3) and Tsien (T4).

Despite the growing body of knowledge of process control theory most process plants are instrumented empirically today. Some useful empirical procedures are described in Section III, C, 3.

B. Analysis

The analysis problem in control can be divided into an analysis of individual components and an analysis of integrated loop behavior, the latter depending on the former. Typical process control systems may be regarded as combinations of process elements, measuring elements, controllers and computing elements, final control elements, and transmission lines. They may also be regarded as combinations of time lags, time delays, amplifiers, summers, differentiators, integrators, and other simple functional units.

In analyzing individual components, what is required is the relationship between signal inputs (forcings) and outputs (responses). In the analysis of a whole process control loop what is required is the behavior of the controlled process variable in response to the typical disturbances imposed on the system. Trimmer (T2) gives an interesting and general treatment of the response of physical systems.

1. *Process Elements*

Typical process elements are flow systems, heat exchangers, contacting systems for diffusional operations, and chemical reactors. Each of these elements has to some extent the properties of storing energy (including matter) and of amplifying or attenuating signals.

a. First Order Lag. Consider the simple jacketed kettle of Fig. 1. At constant feed rate and feed temperature, the relationship between outlet liquid temperature and jacket steam temperature is obtained readily by combining the overall rate equation and a heat balance on the liquid. For a well stirred kettle, the bulk liquid temperature in the kettle and the outlet temperature of the liquid are the same and

$$W_f C_{p_f} \frac{d\theta_o}{dt} = UA(\theta_i - \theta_o) \tag{1}$$

where W = mass
C_p = heat capacity at constant pressure
θ = temperature
U = overall heat transfer coefficient
A = heat transfer surface area related to U
d = differential operator
t = time

Subscripts:
f = liquid feed in the kettle
i = steam in kettle jacket
o = liquid leaving kettle

This equation is a description of the dynamical behavior of the kettle in that it relates temporally the input or forcing temperature of the steam jacket to the response or output temperature of the kettle contents. It is an oversimplification in that the heat capacity of the kettle wall has been neglected.

Since many equations are involved in a control system analysis, it is desirable that each equation be written as simply as possible. Operational calculus provides a useful notation, and in particular the Laplace transformation permits a very simple treatment if the differential equations are linear. A further simplification results if the same types of initial conditions are taken for all problems, or if only steady-state sinusoidal behavior is considered. Churchill (C6) and Carslaw and Jaeger (C2) give particularly useful presentations and applications of the Laplace transformation.

Fortunately, process control problems are most usually concerned with maintaining operating variables constant at particular values. Most disturbances to the process involve only small excursions of the process variables about their normal operating points with the result that the system behaves linearly regardless of how nonlinear the descriptive equations may be. Thus Eq. (1) is a nonlinear differential equation since both C_{p_f} and U are functions of θ_o; but for small changes in θ_o average values of C_{p_f} and U may be regarded as constants, and the equation becomes the simplest kind of first order linear differential equation.

When the Laplace transform of a differential equation is taken, a term must be included for the initial conditions, i.e., for the conditions at $t = 0$. This term is zero for the special case when the dependent variable and all its derivatives with respect to time are zero at $t = 0$. Now the typical process control situation is that the process variables are normally constant with time except for occasional disturbances which temporarily derange the system. Thus the period of time which is of especial interest in control analysis is the time from the start of a disturbance until the system returns to its normal controlled condition. At $t = 0$ for this period all the time derivatives of the dependent variable are zero since all conditions are steady, and for convenience the steady initial value of the variable can be taken as zero. Hence for control purposes the initial conditions terms in the Laplace transformation may be eliminated.

The Laplace transform of Eq. (1) then, is

$$W_f C_{p_f} s\theta_o(s) = UA(\theta_i(s) - \theta_o(s)) \tag{2}$$

where s = Laplacian complex variable and the $\theta(s)$ terms are transformed variables. This equation can be rearranged to give the ratio of output to input transforms,

$$\frac{\theta_o(s)}{\theta_i(s)} = \frac{1}{1 + Ts} \tag{3}$$

where $T = (W_f C_{p_f})/UA$, the *time constant* of this stage (element). The operational expression for the ratio of output to input, in this case $1/(1 + Ts)$, is called the *transfer function* of the stage.

Process elements which are describable by Eq. (3) are often called *first order time lags* or *first order RC stages*, since they cause the output signal to lag behind the input signal and since the time constant in each case is the product of a resistance term and a capacitance term. For the jacketed kettle, the resistance $R = 1/(UA)$ and the capacitance $C = W_t C_{p_f}$. Among the process elements which may behave like first order time lags are mixers, single stage contactors, liquid storage tanks, gas pressure reservoirs, and stirred reactors.

Although the transfer function gives a complete dynamic description of process elements, some interpretation is necessary. The response of the element to any kind of forcing function can be determined from the transfer function, but only a very few kinds of forcing are of any importance in control problems. These important forcing functions are: (a) step; (b) pulse; (c) ramp; (d) steady state sine wave; and (e) random.

Step Forcing. In step forcing, the input which has been constant for all $t < 0$, suddenly jumps to a new value at $t = 0$ and for all time thereafter remains at the new value. The response to a step function is called the *indicial response* or the *response to a constant* or the *transient response* or merely the *step response.* Since step forcing is as severe a kind of forcing as can be imposed on a process, there is considerable advantage in handling control problems in terms of the indicial response insofar as is practicable. Furthermore this response is useful in specifying criteria of control performance. It is disadvantageous, however, in that its computation requires a total solution of the differential equations describing the element. Such a solution becomes difficult and cumbersome for all real elements except the very simplest.

The step reponse of a first order lag may be obtained readily by substituting in Eq. (3) the transform of $\theta_i = A$, which is $\theta_i(s) = \frac{A}{s}$, where A is the magnitude of the step [see tables of transforms in references (B4), (C2), and (C6)].

Thus

$$\theta_o(s) = \frac{A}{s(1 + Ts)} \tag{4}$$

and the solution in the time domain may be written directly from the same tables of transforms referred to previously. For the conditions that the process element is initially at rest, the response to a rise of A in the input signal expressed as a departure from the initial condition is

$$\theta_o = A(1 - e^{-\frac{t}{T}}) \tag{5}$$

The important characteristics of this response are that the slope at $t = 0$ is A/T, and at $t = T$ the response is 63% of its final value. Large values of the time constant T correspond to sluggish response. Obviously, large time constants are undesirable in measuring instruments, whereas in process elements the sluggishness arising from large time constants may in special cases be advantageous in providing a stabilizing inertia in the control system.

Pulse Forcing. In pulse forcing, a shot of energy of some kind is put into the element. The simplest kind of pulse to visualize (but not the simplest to generate) is the *rectangular* pulse which consists of a succession of two step inputs of equal magnitude but opposite sign so that the input returns to its initial condition at the end of the pulse. Other pulses are the *cosine pulse*, which is half a cosine wave, and the *unit impulse* which has infinite magnitude but a time lapse such that the input is a unit of energy. Pulse responses are harder to interpret than step responses, but they can be obtained experimentally with less total upset to the system

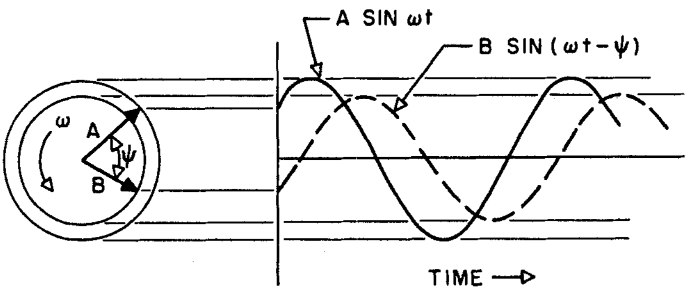

FIG. 3. Vector representation of sine waves.

than step responses. Lees and Hougen (L1) describe the successful use of a displaced cosine pulse to determine the dynamic characteristics of a heat exchange process.

Ramp Forcing. In ramp forcing the input changes at constant rate. This kind of forcing is of greater pertinence to servomechanisms (steering controls, for example) than to process systems.

Sinusoidal Forcing. From the standpoint of control analysis the most useful forcing function is the steady state since wave. If a steady-state, low amplitude sinusoidal variation is imposed on some property of an inlet process stream, the same property of the corresponding outlet stream will also vary sinusoidally and at the same frequency. For most process components the output wave will lag behind the input wave, and the output amplitude will be less.

The *magnitude ratio* (amplitude ratio, modulus) of the output wave to the input wave and the *phase angle* (argument) between the output and the input waves, both given as functions of frequency, constitute the frequency response characteristics of the component. Figure 3 gives a vector representation of the input and output sinusoids. The two sine waves may be regarded as being generated by the ordinate projections of the vectors as they rotate counterclockwise at an angular velocity of ω radians per unit time and displaced ψ radians with respect to each other. For these two waves the magnitude ratio is B/A and the phase angle is $-\psi$, the negative sign indicating that the output lags the input. If amplitudes are small, the amplitude ratios of a series of connected components are multiplicative and the phase angles are additive, providing there is no interaction between components. Thus the overall frequency response characteristics of a system of components can be determined readily from the frequency response characteristics of the individual components. The limitation regarding small amplitudes arises from the fact that the simple relationship between overall behavior and individual behavior is only valid for linear systems, i.e., for systems describable by linear differential equations. Fortunately, as has been intimated, although most real systems are nonlinear, their behavior for small perturbations is linear.

The frequency response characteristics of a process element or a group of elements can be computed readily from the corresponding transfer function merely by substituting $j\omega$ for s, where j is the imaginary number, $\sqrt{-1}$, and ω is the angular velocity. Thus the frequency response characteristics of a simple first order lag are given by

$$\frac{\theta_o(j\omega)}{\theta_i(j\omega)} = \frac{1}{1 + j\omega T} \tag{6}$$

Equation (6) may be split into real and imaginary terms by multiplying the numerator and denominator by $1 - j\omega T$ to get

$$\frac{\theta_o(j\omega)}{\theta_i(j\omega)} = \frac{1}{1 + \omega^2T^2} - j\frac{T}{1 + \omega^2T^2} \tag{7}$$

As shown in Fig. 4, this equation can be plotted for all frequencies on a complex plane having an imaginary vertical axis and a real horizontal axis. The result, which is a semicircle in the fourth quadrant, is the vector locus of the ratio of output to input. At any frequency ω the corresponding sine waves may be produced by rotating the input vector (magnitude 1.0) and the output vector counterclockwise at ω radians per unit time at constant ψ and taking the imaginary parts in a manner similar to that suggested in Fig. 3. Note that the magnitude ratio is merely the square root of the sum of the squares of the coordinates,

$$\left|\frac{\theta_o(j\omega)}{\theta_i(j\omega)}\right| = \frac{1}{(1 + \omega^2T^2)^{1/2}} \tag{8}$$

and the phase angle is

$$\psi = -\tan^{-1}\omega T \tag{9}$$

Note further that the magnitude ratio is 1.0 at zero frequency and zero at infinite frequency and that the corresponding phase angles are zero and -90 degrees respectively.

Figures like Fig. 4, called Nyquist diagrams, are widely used in frequency response characterization.

Farrington (F1) hangs most of his penetrating treatment of control fundamentals on the *inverse Nyquist diagram,* which is a complex plane plot of the ratio of input to output. On this diagram the first order time lag is given by a straight vertical line through Imaginary $= 0$, Real $= 1.0$.

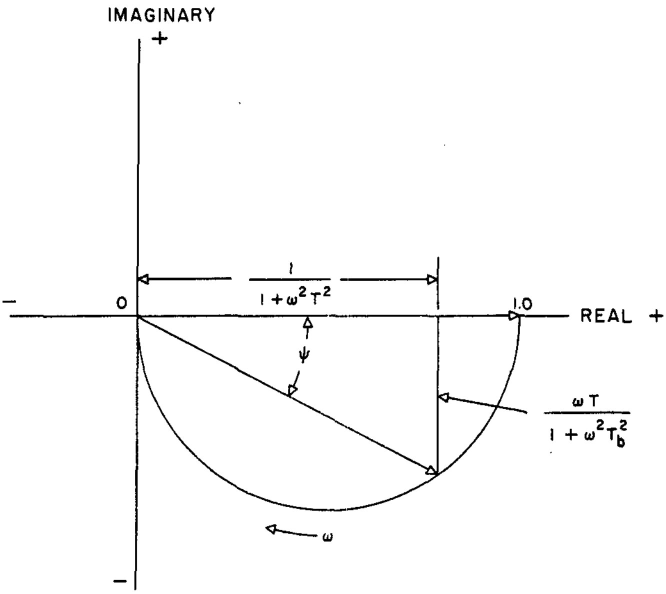

FIG. 4. Vector locus, output/input, first order lag.

A more convenient representation than either of the foregoing is the *Bode diagram,* in which the logarithm of the magnitude ratio and the numerical value of the phase angle, usually in degrees, are plotted against the logarithm of the frequency. The convenience of this plot lies in the fact that the frequency response characteristics of a series of coupled components can be determined readily by a simple graphical addition of the characteristics of the individual components. Furthermore for many kinds of loop components the characteristics can be approximated by a few straight lines. For example, Fig. 5 shows Bode diagrams for a single first order time lag and also for two lags in series. The straight line approximations, which are shown as dashed lines, are found readily from the following facts, which in turn are readily derivable from Eqs. (8) and (9):

(a) At low frequencies the magnitude ratio is unity, and at high frequencies it decreases with increasing frequency at a slope of $-n$ on a log-log plot where n is the order of the lag (or the number of first order lags in series).

(b) For a first-order time lag the high frequency approximation, with a slope of -1, intersects 1.0 magnitude ratio at a frequency of $1/T$ radians per unit time, where T is the time constant of the lag. This frequency is called the *corner frequency* or *break frequency*. The actual magnitude ratio at this frequency is $1/\sqrt{2} = 0.707$.

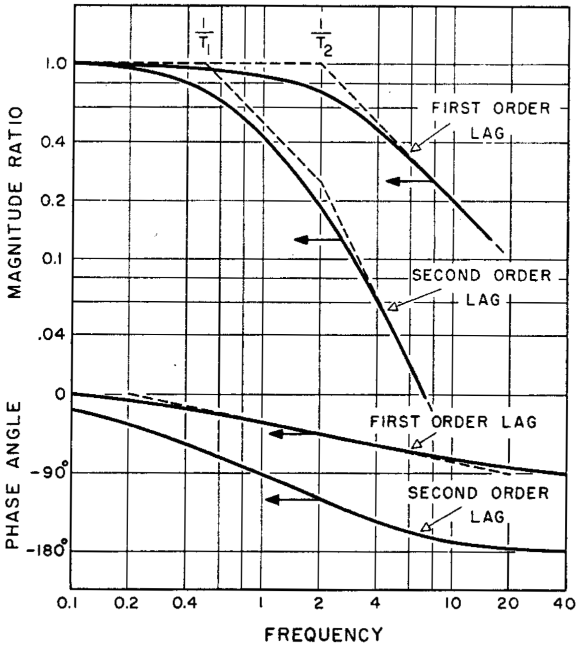

FIG. 5. Frequency response characteristics of time lags.

(c) For a series of first order lags in which the individual components are non-interacting, that is, in which the behavior downstream of a given lag has no effect on that lag, the break frequencies occur at each $1/T$ beginning at the one corresponding to the largest time lag (lowest frequency). At each break the straight line approximation increases in negative slope by one.

(d) The straight line approximation for the phase angle of a single, first order time lag is drawn through $-45°$ at $\omega = 1/T$ from 0 at $\omega = 1/(10T)$ and ending at $-90°$ at $\omega = 10/T$. At frequencies below

$1/(10T)$ the phase angle is zero, and above $10/T$ the phase angle is $-90°$. For a series of lags the individual straight-line approximations may be added graphically.

b. Series of First Order Lags. There are no real control systems so simple that they behave like a single first-order lag. Most systems behave like three or more lags coupled in series.

For three first order lags connected so that the output of the first lag becomes the input of the second lag and the output of the second becomes the input to the third, the overall transfer function relating the output of the third stage to the input to the first is

$$\frac{\theta_4(s)}{\theta_1(s)} = \frac{K_1}{1 + T_1 s} \cdot \frac{K_2}{1 + T_2 s} \cdot \frac{K_3}{1 + T_3 s} \qquad (10)$$

where the individual $K/(1 + Ts)$ terms are the respective transfer functions of the lags. This equation may be derived simply by writing Eq. (3) for each stage and combining the resulting expressions algebraically. If the stages interact, that is, if what happens in stage 2 affects stage 1, etc., the individual time constants in Eq. (10) must be replaced by effective time constants which are related to the interacting resistances and capacitances (F1).

Interaction occurs in most systems of coupled stages. Thus heat transfer through a number of resistances in series may be treated as involving a series of interacting RC (resistance-capacitance) stages. An equation similar to Eq. (10) can be written for this system, and by equating the coefficients on like powers of s after expanding the equations, effective time constants can be determined which permit using the simpler equation of the non-interacting system.

Non-interacting, coupled stage systems are called *cascaded* systems. A simple example would be a series of liquid level tanks arranged so that the discharge of one is the feed to another and so that the flow from a given tank depends only on the liquid level in that tank. Systems of serial RC stages may be effectively non-interacting if the downstream capacitances are very small relative to upstream capacitances. Thus, for all practical purposes, a system composed of a small thermocouple immersed in a large bath would be non-interacting since the bath temperature would be insensitive to changes in the thermocouple temperature.

By substituting $j\omega$ for s in Eq. (10), the overall frequency response characteristics of a third order lag are found to be

$$\left|\frac{\theta_4(j\omega)}{\theta_1(j\omega)}\right| = \left|\frac{\theta_2(j\omega)}{\theta_1(j\omega)}\right| \cdot \left|\frac{\theta_3(j\omega)}{\theta_2(j\omega)}\right| \cdot \left|\frac{\theta_4(j\omega)}{\theta_3(j\omega)}\right| \qquad (11)$$

for magnitude ratio and

$$\psi = \psi_1 + \psi_2 + \psi_3 \qquad (12)$$

for phase angle. The convenience of the Bode diagram derives from these two equations and Eqs. (8) and (9). Note that on a Bode diagram at high frequencies the magnitude ratio of an n-th order lag decreases at a slope of $-n$ and the phase angle approaches $-90n$ degrees as a limit.

Random Forcing. All of the foregoing kinds of forcing whether step, pulse, or periodic are arbitrary functions. Yet in a typical plant or process under automatic control, the usual forcing function is a random one. Theoretically the output from a process component or an array of components must be related to the input via the transfer function of the component even though the input is purely random and the output seemingly random. Because of the high level of noise in the random signals it is not practical to attempt to relate input and output by performing Fourier analyses on input and output records. However the random input signals from a fairly long record, e.g., 2 minutes for a flow process, can be correlated statistically (*auto-correlation*), and a *cross correlation* for the same period of time can be obtained between input and output. The transfer operator relating the cross-correlation function and the auto-correlation function, is the same as the operator relating output and input in real time. Goodman and Reswick (G1) describe the application of their delay line synthesizer in determining the relation between input and output from auto-correlation and cross-correlation functions.

Up to the present time it has not been possible to demonstrate the ultimate reliability of characterizations based on random disturbances. However, the use of random disturbances offers great potential advantage in studying existing process control systems where upsets like step disturbances cannot be tolerated. Because of the extensive calculation required to reduce the random operating records to statistical-correlation functions, high speed digital computation is essential in this treatment.

c. Second-Order Elements. The transfer function of a second-order element may be written as

$$\frac{\theta_o(s)}{\theta_i(s)} = \frac{1}{As^2 + Bs + 1} \tag{13}$$

where the coefficients A and B depend on the nature of the element. Because some generalizations may be made regarding the behavior of second order elements, it is customary to write Eq. (13) in the form

$$\frac{\theta_o(s)}{\theta_i(s)} = \frac{1}{\dfrac{s^2}{\omega_n^2} + \dfrac{2\zeta}{\omega_n} s + 1} \tag{14}$$

where ω_n is the natural angular frequency of the element in radians per unit time and ζ is the damping ratio, dimensionless.

The two roots of the auxiliary equation of the differential equation corresponding to the above expression are

$$s = \omega_n[-\zeta \pm \sqrt{\zeta^2 - 1}] \tag{15}$$

If the damping ratio, ζ, is zero, the roots are purely imaginary, and the response to a step disturbance would be a steady state cycling (*zero damped oscillation*) at a frequency of ω_n. If the damping ratio is unity, the element would be critically damped, and the transient response would be the swiftest possible recovery without overshoot. At damping ratios intermediate between 0 and 1.0, the transient response is a damped sinusoid, and at damping ratios greater than 1.0, the response is the aperiodic response of coupled *RC* stages.

d. Distributed Parameter Components. Systems which may be regarded as a series of time lags are *lumped parameter systems* in that the properties, such as resistance and capacitance which determine their dynamic characteristics, are presumed to operate at specific points in the system. The jacketed kettle, or a thermometer in a protecting well, or a compressed air system involving tanks separated by flow restrictions, may be treated adequately as lumped parameter systems. On the other hand, systems like a double-pipe heat exchanger, or a transmission line, or a tubular flow reactor, cannot be treated reliably as lumped systems, and consequently they are *distributed parameter systems.* Strictly speaking all systems are distributed and only under certain conditions may they be approximated as lumped systems.

There are two methods of tackling the distributed parameter case: (1) by arbitrarily dividing the system into n lumped parameter stages, where n is taken large enough (based on experience) to ensure a good approximation; and (2) by using operational calculus directly. Oldenbourg and Sartorius (O2) describe both the lumping approach and the direct approach. A straightforward general treatment of the direct approach is presented by Cohen and Johnson (C9) in an analysis of double-pipe heat exchangers. Takahashi (T1) has derived and tabulated transfer functions for a wide variety of heat exchange systems.

As an example of what is involved in the direct approach, consider a double-pipe heat exchanger in which condensing steam in the outer jacket heats up cold water flowing at constant mass rate in the inner pipe. It is desired to find the transfer function relating the temperature of the water leaving the inner pipe (as output) and the temperature of steam in the jacket (as input). Assuming for the purposes of this exposition that the steam pressure is constant throughout the length of the exchanger (but not necessarily constant with time), and that the metal walls have negligible thermal resistance, the combined enthalpy balances and heat transfer equations for a differential length of the exchanger are

$$\frac{\partial \theta_{fl}}{\partial t} = -u_{fl}\frac{\partial \theta_{fl}}{\partial x} + \frac{1}{T_1}(\theta_w - \theta_{fl}) \tag{16}$$

and

$$\frac{\partial \theta_w}{\partial t} = \frac{1}{T_{22}}(\theta_{st} - \theta_w) - \frac{1}{T_{12}}(\theta_w - \theta_{fl}) \tag{17}$$

where u_{fl} = fluid velocity
x = distance in the direction of fluid flow
$T_1 = \dfrac{A_{fl}C_{p_{fl}}\rho_{fl}}{h_{fl}\pi D_i}$
$T_{12} = \dfrac{A_w C_{p_w}\rho_w}{h_{fl}\pi D_i}$
$T_{22} = \dfrac{A_w C_{p_w}\rho_w}{h_s \pi D_o}$
A = cross-sectional area normal to flow
C_p = heat capacity at constant pressure
D = diameter
h = film coefficient for heat transfer at wall surface
ρ = density

Subscripts:

fl = fluid
i = inside
o = outside
st = steam (condensing)
w = wall

These equations are partial differential equations rather than ordinary differential equations, but through use of the Laplace transformation they may be converted to ordinary differential equations. Thus, Eq. (16) transformed is

$$s\theta_{fl}(s) = -u_{fl}\frac{d\theta_{fl}(s)}{dx} + \frac{1}{T_1}(\theta_w(s) - \theta_{fl}(s)) \tag{18}$$

If Eqs. (16) and (17) are each written as the sum of a steady-state part and a transient part, and only the varying part is transformed, the resulting equations may be solved simultaneously to give

$$\theta_{fl,L}(s) = \left[\theta_{fl,0}(s) - \frac{b}{a}\theta_{st}(s)\right] e^{-\frac{L}{u_{fl}}a} + \frac{b}{a}\theta_{st}(s) \tag{19}$$

where
$$a = s + \frac{1}{T_1} - \frac{T_{22}}{T_1(T_{12}T_{22}s + T_{12} + T_{22})}$$
$$b = \frac{T_{12}}{T_1(T_{12}T_{22}s + T_{12} + T_{22})}$$
$$L = \text{length of exchanger}$$

Subscripts:

0 = point in exchanger where $x = 0$, i.e. inlet
L = point in exchanger where $x = L$

For constant inlet water temperature ($\theta_{\mathrm{fl},0}$ = constant), Eq. (19) may be written

$$\frac{\theta_{\mathrm{fl,L}}(s)}{\theta_{\mathrm{st}}(s)} = \frac{b}{a}\left[1 - e^{-\frac{L}{u_{\mathrm{fl}}}a}\right] \tag{20}$$

This ratio is the transfer function between the outlet water temperature and the steam temperature in the jacket. For constant steam temperature, the transfer function between outlet water temperature and inlet water temperature is

$$\frac{\theta_{\mathrm{fl,L}}(s)}{\theta_{\mathrm{fl},0}(s)} = e^{-\frac{L}{u_{\mathrm{fl}}}a} \tag{21}$$

As shown by Cohen and Johnson, Eq. (20) leads to frequency response characteristics which on a Bode diagram exhibit resonances both in the magnitude ratio and phase angle. The first resonance occurs at a period approximating the residence time of a slug of water in the inner pipe.

In general, distributed process systems are characterized by magnitude ratios and phase angles which decrease without limit as frequency increases. If the sinusoidal forcing is applied in a distributed manner, the magnitude ratios and phase angles decrease in a periodic or resonating manner.

e. Time Delays. Equation (21) describes a kind of *time delay* or *dead time* in that the effect of a change is not felt or observed until a finite time, L/u_{fl}, has elapsed. These time delays appear frequently in process systems as sampling lags in cases where the analytical or measuring means cannot be located near the process stream, and as distance-velocity lags in cases where the measuring point in the process flow stream must be located an appreciable distance from the point of real interest. The term *distance-velocity lag* is applied to delays for which the time constant or time lapse is given by the ratio of distance to velocity as in Eq. (21).

Consider a measuring point situated in a reactor effluent line a distance L from the reactor. If the fluid velocity in the line is u_{fl}, the time constant for the distance-velocity lag is $L/u_{\mathrm{fl}} = T_{\mathrm{D}}$, and the transfer function is

$$\mathcal{L}\left[\frac{\theta(t-T_D)}{\theta(t)}\right] = \frac{\theta_o(s)}{\theta_i(s)} \tag{22}$$

where $\mathcal{L}$ = Laplace transform, θ = measured variable, e.g., concentration or temperature; and subscripts o and i refer to outlet and inlet respectively, i.e. to $L = L$ and $L = 0$ respectively.

The transformation in Eq. (22) gives

$$\frac{\theta_o(s)}{\theta_i(s)} = e^{-T_D s} \tag{23}$$

Substituting $s = j\omega$ gives the frequency response characteristics, which are that the magnitude ratio is unity for all frequencies, and the phase lag (negative phase angle) increases with increasing frequency without limit. Since the characteristic of unlimited phase angle promotes system instability, time delays are undesirable and should be minimized whenever possible.

f. Nonlinear Process Components. All the components discussed above have been treated as linear systems. Some components which are definitely nonlinear may also be treated as linear systems on the basis that controlled behavior necessarily results in minor fluctuations about mean operating points. No techniques are available for dealing with nonlinear elements generally, but there are two quite different methods which have had some success in dealing with nonlinear elements within certain limitations. These methods are described in Section III, B, 6.

2. *Measuring Elements*

Although the process elements are the heart of the process, they are only a part of the control system, neither more important nor less important than the other elements of the loop. Measuring elements are certainly as important as process elements, for no control is possible if there is no measure of what is going on in the process. In what follows, the emphasis will be on primary sensing elements, although strictly speaking the measuring elements in typical process control systems include both the sensing elements at the point of measurement and the receiving element in the controller.

The great majority of automatic process control systems involve one or more of only five process variables, namely, *pressure*, *temperature*, *flow rate*, *composition*, and *liquid level.* Many of these variables are measured by the same kind of instrument, and indeed, all of them under certain circumstances can be evaluated in terms of pressures. Thus temperature can be measured by the pressure exerted by a confined gas in the gas thermometer; the differential pressure across a restriction in a flow line is a measure of flow rate; the pressure exerted by a boiling liquid mixture at constant temperature is a measure of liquid composition; and the hydrostatic pressure at the bottom of a tank is a measure of the height of liquid in the tank.

A comprehensive summary of measuring elements appears in Perry (P1); and Eckman (E1) gives general descriptions of measuring instruments and their characteristics. *Industrial and Engineering Chemistry* devotes a monthly column to instrumentation, which deals primarily with new developments in the measurement of process variables.

The desired characteristics in measuring elements—including both the primary sensing element and secondary receiving elements—are accuracy and response speed consistent with low overall cost. Absolute accuracy is seldom a requirement, but it is important that the measuring means provide a reliable measure of the process variable. The speed of response should be high relative to the response speed of the process being monitored. A low overall cost requires that the first cost and also all associated operating and maintenance costs be low. What constitutes a low cost for any given problem will depend of course on a total economic appraisal.

Some of the more important measuring methods are catalogued briefly below:

a. Pressure. Pressures are most usually measured by balancing the pressure against a column of liquid as in the *manometer* or against a *pressure spring* of some sort such as a *bourdon tube, diaphragm, bellows,* or *helix.* Where it is desirable to have an electrical signal as the measure of pressure, the *strain gage* and differential transformer is useful.

Manometers and pressure springs may be described dynamically to a first approximation by second-order differential equations for which the roots of the characteristic equation are conjugate complex. As shown in Section III, 8, 1c, since the roots are complex, these systems have an oscillatory mode, and the response of the system to step forcing, for example, is a damped sinusoid.

b. Temperature. Temperature is perhaps the most widely measured process variable, yet it is in principle the most difficult to measure since it cannot be measured directly.

The three most important types of thermometers are expansion-type thermometers (pressure thermometers), electrical thermometers, and radiation thermometers. In expansion-type thermometers the primary sensing element is a bulb containing an expansible fluid. The bulb is connected to a pressure spring through capillary tubing. Expansion of the thermometric fluid with rising temperature causes expansion of the pressure spring, which in turn is converted to a mechanical displacement as the final measure of temperature. The response of these thermometers is determined by the dynamics of the sensing bulb and usually can be regarded as involving a pair of coupled *RC* stages: two resistances given by the external and internal fluid films, and two capacitances given by the bulb wall and the bulb fluid. Transmission lags in the capillary tubing and the behavior of the pressure spring are of secondary importance.

It is frequently desirable to protect the sensing bulb by sheathing it in a thermowell with the result that at least another *RC* stage is added to the measuring system and the response becomes more sluggish.

Because of the ease with which electric signals can be transmitted and manipulated it is not surprising to find that electrical thermometers are the most widely used thermometers in control systems. Both the thermoelectric type thermometer (thermocouple) and the resistance thermometer can be made small, and hence quite high response speeds are realizable. Even sheathed couples can be obtained which have time constants as low as three seconds. In these instruments the hot junction of the couple is welded directly to the sheath.

c. Flow. Most process flow streams are metered for continuous automatic control by orifice meters. The pressure drop across the orifice is sensed either by an enlarged leg manometer or by a pneumatic differential pressure cell. In both cases the response is rapid, usually far more rapid than is required for typical flow control problems.

d. Liquid Level. Liquid level can be measured directly by floats or indirectly by measuring hydrostatic head or by measuring the buoyancy of a submerged mass (displacement float). Thus the problem is essentially one of measuring either a mechanical displacement or a pressure.

e. Composition. There are many ways of monitoring composition including measuring properties such as density, viscosity, refractive index, thermal conductivity, absorption and emission spectra, dielectric constant, and the like. An increasingly popular method is mass spectrometry, despite greater expense than conventional methods.

Most of these methods are rapid in their response to the particular property which is the measure of composition, but frequently their overall behavior is sluggish because of the time delays in leading the stream samples to the test cells.

In some cases, for example where mass spectrometers are used in analyzing complex mixtures, high speed computing equipment is necessary for interpretation, since a large number of simultaneous equations must be solved. The total time lapse for such cases may run into minutes.

3. *Controllers*

The signal from the measuring elements in control systems goes to the controller, where it is compared with some measure of the command signal to the controller. In process control terminology the command signal is called the *set point.* Related but differing quantities are the *desired value* and the *control point.* The former is the value at which it is desired to hold the process variable being controlled; the latter is the point at which the variable is actually controlled.

Most usually the controller takes the difference between the set point and the controlled variable and on the basis of this difference or *deviation,* generates a control action. The relationship between control action and deviation provides a convenient means of classifying controllers.

In process control only a few types of control action (control modes) are important, namely: (1) on-off or two-position control; (2) proportional control; (3) integral control or automatic reset; (4) derivative or rate action.

With the exception of derivative action any of these control modes may be used alone in certain applications. Integral and derivative actions are most usually combined with proportional control to give *proportional plus integral control* (*proportional control with automatic reset*); *proportional plus derivative control;* or *three-mode control,* which is *proportional plus integral plus derivative.*

These various control actions may be generated electrically, mechanically, pneumatically, or hydraulically. Pneumatic and electrical controllers are most widely used. The former are often preferred in petroleum refining and similar industries where fire hazards make electrical devices dangerous unless specially protected. Present trends are toward increased use of electrical controls, particularly where signals must be transmitted more than 200 ft.

a. Two-Position Control. The least expensive and most widely used mode is two-position or on-off control. As the name implies the controller output is one value for all positive deviations from the set point and another value for all negative deviations. Usually these output values are the extremes of full output and zero output. Since this type of control is discontinuous and produces cycling in the controlled variable, it is most applicable to stable systems characterized by large energy capacities and small load changes. Household heating systems and constant temperature baths are familiar examples.

A mathematical treatment of discontinuous automatic control systems with special attention to high-speed position control problems is given by Flügge-Lotz (F2).

To avoid excessive wear due to too frequent action of the on-off controller, it is customary to provide a fixed dead zone or differential around the set point of perhaps 3% of full variable range. The effect of the dead zone is to decrease the frequency of cycling and to increase the phase lag between controller output and input to slightly more than 180°.

b. Proportional Control. In proportional control the controller output is proportional to the deviation, that is

$$\theta_o - \theta_a = K_p(\theta_r - \theta_c) \tag{24}$$

where θ_o = controller output
θ_a = controller output for zero deviation
K_p = controller proportional gain
θ_c = measured variable as seen by the controller, per cent of full scale
θ_r = set point, per cent full scale

Since the controller output must counteract the measured variable, these two quantities, θ_o and θ_c, are of opposite sign and hence are inherently 180° out of phase. In commercial proportional controllers this 180° phase shift and also any set gain K_p, are constant for all practical ranges of frequency. Thus the frequency response characteristics of a proportional controller are a magnitude ratio of K_p and a phase lag of 180°.

Instrument manufacturers sometimes present the scales on proportional action as *per cent proportional band* which is $100/K_p$.

Proportional control is the basic control mode. Indeed on-off control may be regarded as a limiting case of proportional control at very high K_p. Proportional control produces an immediate and opposing particular output for every particular deviation. It has the disadvantage that it will tolerate an *offset*, i.e., a steady-state error or sustained deviation, since in general, the particular deviation at which the corresponding controller output balances the process, will not be zero. The offset produced by a given load change can be shown to be equal to $1/(1 + K_p)$ times the deviation that would result if no control action were taken to correct for the load change. This computation is valid only if K_p is expressed in terms of the potential corrections resulting from the actual controller outputs.

It is clear from the above expression that offset can be minimized with proportional controllers by setting the proportional gain at high values. At high gains, however, control systems become increasingly unstable, and for some systems instabilities occur at relatively low gains, so that proportional control alone is unsatisfactory. In order to eliminate offset without promoting system instability, an additional mode of control such as integral control must be used.

c. Integral Control. Integral control has also been called *proportional speed floating control.* The former name derives from the fact that the controller output is proportional to the time integral of the deviation;

the latter from the fact that the control action is always changing, hence floating.

In differential form the expression for integral control is

$$\frac{d\theta_o}{dt} = \frac{1}{T_{\text{int}}}(\theta_r - \theta_o) \tag{25}$$

where T_{int} is the integral time (reciprocal reset rate). Because of the floating action the integral controller takes increasing action so long as there is any deviation, hence offset is impossible.

From the Laplace transformation of Eq. (25), the transfer function for an integral controller is $1/(sT_{\text{int}})$ and by substituting $j\omega$ for s the corresponding frequency response characteristics are found to be a phase angle of $-90°$ and a magnitude ratio of $1/\omega T_{\text{int}}$.

d. Derivative Action. Offset may also be reduced by adding a stabilizing control mode which permits the use of higher gains in the proportional mode. Such a stabilizing mode is derivative action, or *rate* action as it is sometimes called.

In derivative control, the controller output is proportional to the rate of change of the input with time, or in differential form

$$\frac{d\theta_o}{dt} = T_d \frac{d^2}{dt^2}(\theta_r - \theta_o) \tag{26}$$

where T_d = derivative time.

Unlike integral control this action by itself provides no control, since it gives a sustained controller output for a constantly changing input. The output does not depend on the amount of deviation but only on how fast the deviation is changing.

Derivative action is used occasionally with proportional control alone, but more frequently it is used with proportional plus integral control.

The transfer function for derivative control is sT_d, and the frequency response characteristics are a phase angle of $+90°$ and an amplitude ratio of ωT_d. Control stability results from the leading phase angle.

e. Proportional Plus Integral Control. The most important control mode combinations are proportional plus integral and three-mode control. Proportional plus derivative control finds occasional use, but the combination of integral plus derivative control has limited practical application because at certain frequencies the two actions cancel each other and no control results.

Proportional plus integral control (also called proportional plus reset) has the transfer function

$$\frac{\theta_o(s)}{(\theta_r - \theta_o)(s)} = K_p\left(1 + \frac{1}{sT_{\text{int}}}\right) \tag{27}$$

The frequency response characteristics are sketched in Fig. 6 at low frequencies. Note that the integral action provides infinite gain at zero frequency.

f. Three-Mode Control. The transfer function for ideal three-mode control is

$$\frac{\theta_o(s)}{(\theta_r - \theta_c)(s)} = K_p\left(1 + \frac{1}{sT_{\text{int}}} + sT_d\right) \tag{28}$$

Although proportional plus integral controllers follow quite closely the idealized behavior of Eq. (27), most commercially available three-mode

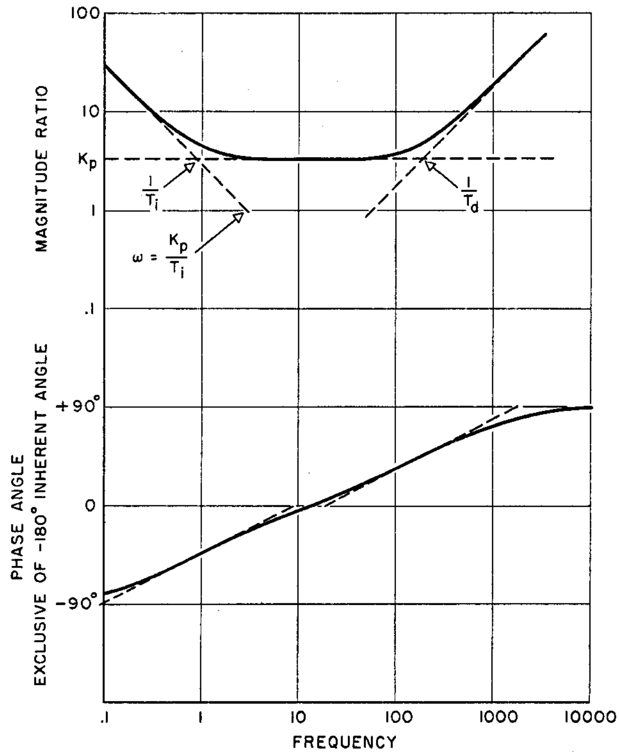

FIG. 6. Frequency response characteristics of three-mode controller.

controllers only very approximately follow Eq. (28). The reason for this departure from ideal behavior is that the derivative action cannot readily be generated independently of the other actions, and particularly in pneumatic controllers the resulting interaction alters the overall performance of the controller. Typical equations for various commercial three-mode controllers are given by Farrington (F1) and Young (Y1).

Figure 6 gives the frequency response characteristics of ideal three-mode control. The chief contributions of the derivative action are to decrease the phase lag of the controller and provide high gain at high frequencies.

It can be seen from the ideal equation that the reset rate, $1/T_{\text{int}}$, is the number of times per minute that the integral action repeats the proportional action, and the rate time, T_d, is the time that the derivative mode advances the control action over that of the proportional mode alone.

4. *Final Control Element*

The controller output becomes the input to the final control element or *regulating unit* as it is often called.

a. Valve Types. For most process control problems the final control element is a valve. Depending on whether the controller output is an air pressure or a voltage the drive for the valve will be a diaphragm motor or a reversing electric motor (or for on-off control, a solenoid).

The most widely used control valves for liquid and high pressure gas streams are the various sliding stem valves such as the *bevel plug*, *V-port*, *parabolic plug*, and for small flows the *needle* valve. For corrosive materials and slurries, pinch-type valves like the *Saunders patent* valves are used. Rotary stem valves include plug types and for gas flows *butterfly valves*, *dampers*, and *louvers*.

The bevel plug valve, which is essentially a globe valve, is used with on-off and narrow band proportional control. For more difficult control systems, *characterized valves* like the V-port and parabolic plug are preferred because the flow-lift (valve position) behavior is more uniform over the full range of the valve.

Where there are large pressure drops across the valve, *double-seated* valves are used to balance the pressures on the plug and ensure easy operation. To avoid erratic control action due to valve sticking the controller output may be applied to a *valve positioner*, which is a servomechanism (position controller) designed to fix the valve position as required by the controller.

b. Valve Characteristics. In characterizing valves for control purposes it is necessary to examine both the valve itself and the valve motor or actuator. Valves are usually characterized by their manufacturers in terms of the relationship between valve stem position (lift) and flow through the valve at constant pressure drop. Thus a *linear valve* is one for which the flow at constant pressure drop is linearly dependent on the lift. A *semilogarithmic* characteristic produces a straight line when the logarithm of the flow is plotted against lift. The latter characteristic is also called *equal-percentage*, since equal increments of valve lift produce equal percentage flow changes (percentage of previous flow). Equal-percentage valves are desirable in control circuits which are subject to large load changes because precise adjustment of flow rate is possible whether the valve is almost fully open or almost fully closed. With linear valves on the other hand, a given shift in valve position will cause a very high percentage change in flow rate at low flows and negligible percentage change at high flows.

For many process control applications it is possible that the pressure drop across the valve will have more effect on flow rate than will the valve position. The relationship between pressure drop and flow rate at constant valve position is the square-root expression for flow through an orifice. This relation introduces a nonlinearity in the control system which complicates the analysis of the system except insofar as it can be linearized. For equal-percentage valves the effect of changing pressure drop in the lift-flow characteristic of the valve is small, but for linear valves the effect is large.

The selection of valve size is something of an art in that a compromise is usually required. In general, the valve should be small enough to take up most of the pressure drop in the supply line but large enough to pass the maximum possible control agent flow that might be needed by the process. An incorrectly sized linear valve can seriously affect the stability of the control loop, whereas the size of an equal-percentage valve has little effect on stability.

Flow ranges of control valves frequently are characterized as *rangeability*, which is the ratio of maximum to minimum flow rates, and *turndown*, which is the ratio of normal maximum operating flow rate to minimum flow rates. The minimum flow rate for control valves is usually not zero to preclude hysteresis effects that might result from total closing.

c. Valve Actuators. Valve motors in pneumatic systems are most usually spring-loaded diaphragms. Their behavior can be characterized by second-order equations, but in most process control systems their response is sufficiently rapid to permit characterizing them as simple first-order lags.

Electric motors may be variable-speed, reversible motors or simple solenoids (for on-off control). Because the former are expensive, it is a common practice to transduce the electric signals to pneumatic signals and use inexpensive diaphragm valves.

Valve actuators may be classified broadly as *directional*, such as solenoids and electric motors; *force-balanced*, such as spring-loaded diaphragms, and *position-balanced*, such as valve positioners. On the basis of this classification, Close (C7) describes the important types of actuators in commercial use, their dynamic characteristics, and how they are integrated into process systems.

5. *Transmission Lines*

Transmission lines in process control systems rarely make any significant contribution to the overall loop characteristics. Where signals are transmitted electrically, there is no detectable signal attenuation for any frequencies characteristic of the process components, and even for pneumatic transmission lines as long as 200 feet there is little loss of signal. Transmission lines have distributed properties, but according to Bradner (B3) who has studied pneumatic transmission lines extensively, they can be approximated as second-order systems.

6. *Loop Behavior*

a. Linear Systems. If the dynamic characteristics of all the components in the control system are known, it is possible in principle to predict how the whole system will behave. Although the closed-loop can be analyzed by formal solution of the system equations or by means of the Nyquist diagram or the inverse Nyquist diagram (F1) or by the root-locus method of Evans (E2), for process control systems it is usually more convenient to use the Bode diagram, particularly for systems involving distance-velocity lags.

Consider the simple control loop of Fig. 2. The open-loop characteristics for a typical system exclusive of the controller, are a magnitude ratio and phase angle which fall off with increasing frequency. When the loop is closed through a controller, the magnitude ratio of the closed loop has a resonance peak if the controller gain is sufficiently great.

If the transfer function for all the components in the process control loop lumped together exclusive of the controller is $G(s)$, and the transfer function of the controller (exclusive of the summer which takes the difference between set point and measured variable) is $G_c(s)$, the transfer function of the closed loop is

$$\frac{\theta_c(s)}{\theta_r(s)} = \frac{G_c(s)G(s)}{1 + G_c(s)G(s)} \tag{29}$$

The closed-loop frequency response characteristics may be obtained by replotting the open-loop characteristics on a graph having contours of constant closed-loop magnitude ratios and angles. Points for the closed-loop characteristics are obtained from the intersections of the open-loop locus and the contours. Either polar diagrams such as the transfer function plot (Nyquist diagram) and the inverse Nyquist diagram, or a plot of log magnitude versus phase (Nichols or Black chart) (B4) may be used.

It can be shown by a simple algebraic manipulation that the loci on a Nyquist diagram of constant closed-loop magnitude ratio, M, are eccentric circles of radius $M/(1 - M^2)$ having centers along the real axis at $-M^2/(M^2 - 1)$ from the origin. These M circles are for the closed-loop magnitude ratio of controlled variable to set point, corresponding to the transfer function given by Eq. (29).

In an analysis of the closed-loop behavior of an automatic process control system three items are of particular interest: (a) How stable is the system in recovering from upsets? (b) How quickly does it recover? (c) What is its steady-state deviation after recovery?

The question of stability can be considered most readily from the Bode diagram for the open-loop characteristics of the whole system excluding the controller. Because of the summing mechanism in the controller which gages the deviation, there is inherent in any controller a basic phase lag of 180°. Taking the deviation involves reversing the sign of the signal going around the loop, and this reversal corresponds to a phase shift of exactly 180°.

If a sinusoidal input signal of sufficient frequency is impressed on the process components in the loop so that the corresponding output signal from the process (input to the controller summer) lags by 180°, the addition of the controller to the loop will bring the total phase shift to 360° and the signals at this particular frequency will be in phase around the loop. If the gain in the controller is just sufficient to offset the attenuation in the process, the signals will cycle around the loop continuously without abatement. Higher gains in the controller lead to instability and runaway, while lower gains result in damped response. Thus the frequency at which the process phase lag is 180°, is the critical frequency or *ultimate frequency* of the system. Similarly the proportional control gain which just produces steady-state sinusoidal oscillation is the *ultimate proportional gain.* At this controller gain, the total loop gain, which is the product of controller and overall process gains, is unity. A step disturbance anywhere in the loop will cause sustained oscillations of the controlled variable at the ultimate frequency since the step disturbance excites all frequencies.

The ultimate frequency response properties of a control loop are given special names. *Gain crossover* is the point at which the overall open-loop magnitude ratio is unity. *Phase crossover* is the point at which the overall open-loop phase angle is $-360°$. Since the controller has an inherent phase angle of $-180°$, it is a common practice to disregard this angle and handle the analysis on the basis of the overall open-loop characteristics exclusive of the additional 180° lag. Thus, *phase crossover* occurs at the frequency for which the total phase angle of the open-loop system—exclusive of the $-180°$ inherent shift in the controller—is $-180°$, i.e., at the ultimate frequency. *Phase margin* is the difference between the phase angle at gain crossover (exclusive of the $-180°$ in the controller) and $-180°$. For example, if the phase angle at gain cross-over is $-120°$, the phase margin is 60°. Phase margin may be regarded simply as the margin of stability in terms of phase angle that the system has at the critical condition of unity magnitude ratio. Similarly, *gain margin* is the measure of stability in terms of magnitude ratio that the system has at the critical condition of ultimate frequency. Gain margin is expressed as a factor. Thus for a control loop with an overall magnitude ratio of 0.4 at phase cross-over, the gain margin is 1/0.4 or 2.5.

Control loops with gain margins exceeding 2.0 and phase margins not less than 30° are reasonably stable, although for some cases somewhat higher minima may be desirable, At these margins the loop response to disturbances usually will be oscillatory with fairly rapid damping.

How much oscillation and how much damping there will be can be determined reliably only by solving the descriptive equations of the system for the case of a typical disturbance. Since such solutions are tedious at best and impossible at worst, it is desirable to make use of approximations derived from the frequency response characteristics insofar as practicable. It is generally true for simple systems, that the initial overshoot of the controlled variable in response to a step change will be approximately equal to the resonance peak of the corresponding closed-loop magnitude ratio, provided the peak is about 1.2–1.3. That is, a 20–30% overshoot (based on the amount of the step change) will occur if the peak in the closed-loop magnitude ratio is 1.2–1.3. For higher magnitude ratios the peak overshoot will be higher but not proportional to the closed loop peak. Furthermore the frequency of the damped oscillation will approximate the resonance peak frequency, and for typical stable gain margin and phase margin, e.g., 2.5 and 45° respectively, the subsidence ratio (ratio of a succeeding peak to the peak of the preceding cycle) will approximate one-third.

From the foregoing it is possible to sketch in the transient response except for the final steady value of the controlled variable. The speed of recovery is most conveniently expressed in terms of the *settling time*, which is the time required to come within some limit of the final steady value, usually within 5%. A system with a 1/3 subsidence ratio would have a settling time for 5% limit of roughly two times the period of oscillation of the transient.

The steady-state error of the system is the deviation at infinite time after the disturbance. For a unit change in set-point the error is

$$\left| \frac{\theta_r - \theta_c}{\theta_r} (j\omega) \right|_{\omega=0} = 1 - \left| \frac{\theta_c(j\omega)}{\theta_r(j\omega)} \right|_{\omega=0} \tag{30}$$

But from Eq. (29)

$$\left| \frac{\theta_c(j\omega)}{\theta_r(j\omega)} \right|_{\omega=0} = \left| \frac{G_c(j\omega)G(j\omega)}{1 + G_c(j\omega)G(j\omega)} \right|_{\omega=0} \tag{31}$$

and since all θ are actual or potential values of the controlled variable, $|G(j\omega)|_{\omega=0} = 1.0$ and therefore

$$\left| \frac{\theta_c(j\omega)}{\theta_r(j\omega)} \right|_{\omega=0} = \left| \frac{G_c(j\omega)}{1 + G_c(j\omega)} \right|_{\omega=0} \tag{32}$$

With proportional control $G_c(j\omega) = K_p$ and $|G_c(j\omega)| = K_p$ so that Eq. (32) becomes

$$\left| \frac{\theta_c(j\omega)}{\theta_r(j\omega)} \right|_{\omega=0} = \frac{K_p}{1 + K_p}$$

and Eq. (30) becomes

$$\left| \frac{\theta_r - \theta_c}{\theta_r} (j\omega) \right|_{\omega=0} = \frac{1}{1 + K_p} \tag{33}$$

as stated in Section III, B, 3b.

With the addition of integral control to give proportional plus reset control, $G_c(j\omega) = K_p \left(1 + \frac{1}{j\omega T_{int}}\right)$ and the magnitude ratio at zero frequency becomes infinite with the result that $\left| \frac{\theta_r - \theta_c}{\theta_r} (j\omega) \right|_{\omega=0} = 0$ and no sustained deviation is possible.

The same result obtains with three-mode control. On the other hand the combination of proportional and derivative control gives the same steady state error as proportional alone, since the derivative contribution disappears at low frequency.

b. Nonlinear Systems. The foregoing discussion of closed loop behavior has been limited for the most part to linear systems. For such systems the problem of analysis is relatively simple because frequency response techniques are applicable, and the individual component characteristics can be combined readily to give the overall system characteristics.

With nonlinear systems, however, all simplicity disappears. No general methods of solving even the simplest, nonlinear differential equations are known. Frequency response characterization is useless since sinusoidal forcing will not produce sinusoidal response. The only recourse other than arbitrary linearization of the equations is to utilize approximation methods like those of describing function analysis and phase-plane analysis.

The describing function, first suggested in this country by Kochenburger (K1), is merely the ratio of the fundamental component of the output of the nonlinear element to the amplitude of the sinusoidal input. If there is only one nonlinear element in the control system, all harmonics higher than the fundamental will be attenuated in the linear parts of the system, and the describing function may be used as the frequency response characterization of the nonlinear element. The only other restriction is that any analyses based on frequency response must be at particular signal levels in the nonlinear element.

A serious limitation on describing function analysis is that there is no means of gaging its reliability. For example, it might be expected that describing function analysis would be particularly useful in identifying the existence of limit cycles in nonlinear systems. (Limit cycles are the conditions of sustained oscillation that obtain in nonlinear systems. The familiar condition of ultimate frequency and ultimate proportional gain in a process control system, represents a limit cycle since it is not possible to adjust the system exactly to the ultimate gain and as a result a truly linear system would run away.) Yet there are frequent examples of cases where describing function analysis has failed to indicate limit cycles which actually existed (T3).

In general, the transient response of stable nonlinear systems cannot be inferred from the frequency response characteristics derived from describing function analysis.

Phase-plane analysis of nonlinear systems, suggested by MacColl (M1), is based on the characteristics of the equation

$$\frac{d^2\theta}{dt^2} + A\left(\theta, \frac{d\theta}{dt}\right)\frac{d\theta}{dt} + B\left(\theta, \frac{d\theta}{dt}\right)\theta = 0 \qquad (34)$$

The phase-plane representation is a plot of $d\theta/dt$ vs. θ as families of curves for a given system with initial conditions $d\theta/dt(0)$ and $\theta(0)$ as parameters. This plot or phase portrait, provides a useful indication of the transient response of a nonlinear system. It cannot be applied to sinusoidal or other continuing forcing functions. Furthermore, the method is limited to second-order systems or systems that can be handled as second-order systems.

A succinct appraisal of phase-plane analysis, its possibilities and limitations, is given by Truxal (T3).

c. Process Systems. In recent years a number of process control systems have been studied theoretically, experimentally, and by analogy. No attempt will be made here to list all these studies, but a few of the more quantitative studies will be cited by way of illustration.

Williams *et al.* (W1), describe the results of studies of the automatic control of continuous fractional distillation. These studies were made on an analog computer which could simulate a five-plate tower. The effects of column design, varying feed rate, imperfect sampling, and quality of feed and reflux on controllability were evaluated. An earlier article by Rose and Williams (R2) on the same system compares various schemes for controlling fractionation columns. One interesting conclusion is that derivative control action cannot improve the control for any of the various combinations of measurement and regulation that were studied.

Bilous and Amundson (B1, B2), also using an analog computer, examined the stability and sensitivity of continuous, stirred reactors and tubular reactors. For the former type of reactor the methods of small perturbations used in nonlinear mechanics are directly applicable, but for the latter the partial differential equations necessary for the basic description make a rigorous solution impossible. Both the methods of small perturbations and of frequency response were used in the study. For large perturbations, the solutions for the stirred reactor were obtained on the analog computer. It was concluded that the natural behavior of stirred reactors is not to approach unstable states. The parametric sensitivity, i.e., the sensitivity to changes in operating conditions, of tubular reactors was predicted semi-quantitatively from frequency responses and transient responses approximated by linearizations about fixed operating points.

In one of four articles on the design of gas flow systems Schwent *et al.* (S2), present a theoretical analysis of a wind tunnel installation and a comparison of theoretical and observed performance. All possible simplifying assumptions and approximations are used to advantage.

d. Sampled-Data Systems. The analysis up to this point has considered systems in which the flow of signals is continuous. Virtually all process systems involve continuous measurement and recording of process variables and continuous regulation and control of processes. However, there are a few process systems in which the controlled variable is measured intermittently. In some fractional distillation operations, for example, an expensive process stream analyzer may be shared between two separate units, analyzing first one unit stream and then the other. Such systems are called sampled-data systems, because the flow of signals is carried in samples at regular time intervals. Because of the time lapse between signals, the sampled-data systems are potentially less stable than the corresponding continuous systems.

Fortunately, frequency response techniques are applicable with slight modification to sampled-data systems. Ragazzini and Zadeh (R1) describe the analysis of these systems. Applications in process control with large distance-velocity lags are considered by Oldenbourg (O1) and Sartorius (S1).

7. *Complex Systems*

Complex, multiple-loop, process control systems do not pose any new problems of analysis. The closed-loop characteristics of the smaller loops become the individual component characteristics of successively larger loops.

8. *Effect of Disturbance Location*

Although it is common practice to analyze control loops in terms of the response of the controlled variable to changes in set point, the usual disturbances in process control systems occur at various points in the process rather than at points in the controlling instruments. No special techniques of analysis are required to determine the response of the controlled variable to disturbances applied anywhere in the loop. It is only necessary to manipulate the component transfer functions algebraically, until the ratio of controlled variable to disturbance is found.

Little is known about the nature of process disturbances, so some arbitrary upset like a step change must be used in the analysis. In general, the resulting upset in the controlled variable will be more severe, the closer the point of disturbance is to the point of measurement of the controlled variable. For the usual types of process components, the effect of interposing components between the disturbance and the controlled variable is to attenuate the disturbance. A convenient control system property, which gives a clear picture of the effect of disturbance location, is the *deviation reduction factor*. This factor is the ratio of the change in the controlled variable when the system is disturbed to the change that would have resulted had there been no control action. European engineers (J1, Y1) have made greater use of this quantity than have American engineers.

C. Synthesis

1. *Criteria of Control*

Before an engineer can undertake a design he must have a clear picture of the purpose of the design. In the case of process control systems, the ultimate objective is an operating process plant which will maximize the return on the allowed investment. A more narrow specification might be to design an operating process plant of given capacity and maximized return on investment. An even narrower specification but one more typical of the kind given to the process control system designer, would be to design a plant of a given capacity with an operating performance specified as to tolerable limits of critical process variables.

Since process plants most usually operate continuously under steady-state conditions, the criteria for good control may be specified realistically in terms of the allowable behavior of the process variables following typical upsets in operations. Hence the transient response of process variables to step changes in loads or step changes in set point provides a convenient basis for fixing control performance specifications.

For some processes no cycling of the controlled variable is desirable, for example, because the oscillations might be amplified in process units downstream. In other processes, continuous cycling is not objectionable so long as the controlled variable always lies within a specified range. In still other processes a temporary oscillation which damps out quickly may be required.

For some processes, it is not necessary for the controlled variable to line out right on the set point. In most cases, however, little if any offset is permissible.

A generally used set of criteria for good control is that the controlled variable in response to a unit step change in set point (a) overshoot by not more than 20 per cent of the step and (b) damp out with a subsidence ratio of about one-third. This behavior is approximated by many systems if the closed-loop frequency response and the corresponding open-loop frequency response have certain simple characteristics. Since the closed-loop frequency response characteristics can be determined readily from the open-loop frequency response, the latter characteristics of simple control systems can be used as a convenient basis for design.

2. *Control System Synthesis from Frequency Response Characteristics*

Generally recommended frequency response characteristics for automatic control systems are given by Oldenburger (O3) as follows:

(a) Phase margin should be at least 30 degrees and the gain margin at least 2.5.

(b) Maximum closed-loop magnitude ratio should be less than 2.0, and preferably about 1.3. The resonant frequency at the peak magnitude ratio should be as high as possible.

In the absence of other specifications, the problem of control system synthesis is one of adjusting the characteristics of the various loop components until the recommended frequency response characteristics are obtained. All the control components including measuring elements should have negligible lagging and attenuation characteristics over the range of frequencies which are of importance for the control system. The process components should have as little phase lag as possible. A large attenuation in the process magnitude ratios, however, is not objectionable and in many cases makes for simple and precise control.

Minimizing the overall phase angle of a typical process system, maximizes the frequency at phase crossover and hence maximizes the response speed of the system. Large phase lags in process components, for example resulting from distance-velocity lags, can be counteracted by the phase lead characteristic of derivative control action.

To determine the controller gain required for a particular closed-loop resonance peak in magnitude ratio a simple graphical proportionality can be applied on the Nyquist diagram. The procedure is to draw the open-loop vector locus of the ratio of the controlled variable to the command for some arbitrary controller gain, e.g., one equivalent to a gain margin of 2.5. If the desired resonance peak in the corresponding closed-loop magnitude ratio is 1.3, draw in the $M = 1.3$ circle. Estimate where the open-loop locus would be just tangent to the M circle. From the point of expected tangency, draw a vector to the origin through the arbitrary open-loop locus. The necessary controller gain for a 1.3 resonance peak, will be given by multiplying the arbitrarily chosen controller gain by the ratio of the length of the vector between the tangent and the origin to the length between the origin and the intersection of the vector with the open-loop locus.

If the proportional gain of the controller for reasonably stable control is too low so that there is excessive steady-state deviation, integral control action must be added. Ordinarily the reset rate is kept low enough to avoid changing the phase margin of the system. A commonly used rule of thumb (Y1), is that the integral time should be set equal to the ultimate period (reciprocal ultimate frequency). In cases where the initial phase margin is reduced by the integral action, it is necessary to reduce also the proportional gain to ensure the stability of the control system.

Flow systems have very small capacitances and as a result are most effectively controlled by proportional plus integral controllers having low proportional gains and very high reset rates. Such controllers are essentially proportional speed floating controllers (pure integral) and usually the proportional gain is non-adjustable at a value near unity. Catheron and Hainsworth (C3) show that for liquid flow systems the process response is so rapid that the controlling components are limiting and hence the controller characteristics can be preset and never adjusted further.

Improvement of the speed of response of control systems, requires that the frequency at gain crossover be increased by increasing the controller proportional gain. Since such an increase in gain crossover frequency decreases both the phase margin and the gain margin, the stability of the system will decrease. What is needed is the phase-leading characteristic of derivative action to increase the ultimate frequency, and consequently increase both the gain margin and the phase margin of the system at increased proportional gain. The maximum derivative time is fixed by the allowable gain margin, since the derivative action contributes increasing gain with increasing frequency.

3. *Empirical Methods of Controller Specification*

The more usual problem in the synthesis of automatic process control systems is simply the selection of appropriate controlling components

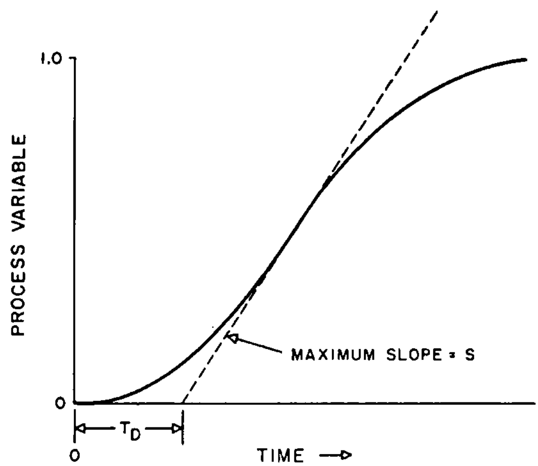

FIG. 7. Process reaction curve.

including measuring and regulating instruments which, when hung on the process equipment, will provide an adequate control performance in response to the normal disturbances imposed on the process.

There are a couple of simple empirical approaches for estimating the optimum controller settings for a particular process. Both approaches require data on the response of the existing process to simple stimuli: one the open-loop response to a step; the other the behavior of the closed-loop at the condition of ultimate gain.

a. Reaction Curve (Signature Curve). Theoretically, the transient response of a system to a step-forcing function contains all possible information about the dynamic characteristics of the system, since the step contains the whole spectrum of frequencies. Unfortunately this information cannot be elicited reliably because much of it is contained in the response near zero time.

Ziegler and Nichols (Z1) on the basis of studies of a variety of process systems, proposed that systems generally can be characterized by the *apparent dead times* and the *maximum reaction rates* of their transient responses (reaction curves or signature curves as they are often called in this context). The reaction curve can be obtained readily if the process may be subjected to a step input with the loop open, and the two quantities can be taken from the curve as indicated in Fig. 7.

The maximum reaction rate is the slope at the point of inflection, S, and the apparent dead time is the abscissa intercept of the maximum reaction rate line, T_D.

Optimum controller settings based on these quantities are as follows:

For proportional control only

$$K_p = \frac{1}{ST_D} \tag{35}$$

For proportional plus integral

$$K_p = \frac{1}{1.1ST_D}$$

and (36)

$$T_{int} = \frac{T_D}{0.3}$$

For three-mode control with action similar to the Taylor Fulscope Controller,

$$\begin{aligned} K_p &= \frac{1}{0.83ST_D} \\ T_{int} &= 2T_D \\ T_d &= \frac{T_D}{2} \end{aligned} \tag{37}$$

where all times are in minutes, T_D is the apparent dead time, and S is the maximum slope of the reaction curve.

Cohen and Coon (C8) present a similar set of rules based on a theoretical treatment of signature curves, assuming that a reasonable representation of real processes is a time delay plus a single RC stage.

b. Optimum Controller Adjustments from Ultimate Gain and Frequency. Ziegler and Nichols also suggested setting controller adjustments on the basis of the ultimate proportional gain and the frequency of cycling at that gain as follows:

For proportional control only

$$K_p = \frac{K_0}{2} \tag{38}$$

corresponding to a gain margin of 2.0.

For proportional plus integral control

$$K_p = 0.45K_0$$
$$T_{int} = \frac{1}{1.2f_0} \tag{39}$$

For three-mode (Taylor Fulscope) control

$$K_p = 0.6K_0$$
$$T_{int} = \frac{1}{2f_0} \tag{40}$$
$$T_d = \frac{1}{8f_0}$$

where K_0 is the ultimate proportional gain and f_0 is the ultimate frequency in cycles per minute. These expressions are based on the following relations between the reaction curve constants and the ultimate gain and frequency:

$$T_D = \frac{1}{4f_0}$$
$$S = \frac{8f_0}{K} \tag{41}$$

The ultimate characteristics of a process control system are obtained readily by closing the loop through a proportional controller, and increasing the gain on the controller to the minimum proportional gain at which the system oscillates steadily. Johnson and Bay (J4) describe tests of both the ultimate gain approach and the reaction curve approach applied to pneumatic analog systems.

4. *Complex Systems*

The techniques for treating the synthesis of simple control systems are generally applicable to complex systems if the complexity arises from multiplicity of loops rather than intransigence of loop elements.

A common kind of complicated system is the *cascaded control system*, in which a master controller reacts to disturbances in the controlled variable by manipulating the set point of a secondary controller, which in turn manipulates a control agent affecting the process. Franks and Worley (F3) present a quantitative analysis of a cascade control system simulated on an analog computer.

Among other kinds of complex systems are those which incorporate high speed computers in the control loop to generate the command signals for the controllers on the basis of broad scale appraisals of operating performance, output requirements, and other pertinent variables. Such systems are not yet completely in being in the process industries, but studies of the possibilities are in progress both in industry and in academic laboratories.

IV. Present Status and New Directions

A. Summary of Present Status

Automatic control has long been a significant feature of the process industries, but only very recently has there been any attempt to treat process control problems quantitatively. It is now clear that the theoretical developments in the field of servomechanisms and governors are generally applicable to process control problems. The great simplification in treatment that results from frequency response characterization, makes possible a quantitative appraisal and design of process control systems.

Although most chemical engineers are aware that feedback control theory has important applicability to industrial process control problems, the number of engineers who make use of this applicability is small.

One limitation on using control theory in process control problems has been a lack of data on process characteristics, but now an increasing volume of effort is being applied to the dynamic characterization of processes. The fact that many typical processes of importance in chemical engineering are much harder to describe dynamically than the simple positioning operations in servomechanisms, may have been a deterrent to any earlier development of this area. Up to now most process characterization has focused on heat transfer processes, but every other type of process or operation has been studied to at least some extent. A number of process companies, particularly certain petroleum companies, have conducted extensive frequency response studies of their process units. Also a few process companies have undertaken dynamic studies of various kinds of instruments.

At the practical level, the status of some recent past trends in the automatic control of process plants has crystallized. Centralization of control and monitoring functions in compact and protected control centers has long been an important feature of process instrumentation. In some cases the centralization was overdone; today the practice is to mount as much instrumentation as possible at the points of application, leaving only critical items at the control center. By recording only really significant variables and by using miniaturized instruments, modern control centers are much less cluttered and confusing than their predecessors. Graphic panels, wherein the control center instruments are mounted on a flow diagram of the process, are widely used particularly for well-proved processes which are not likely to undergo appreciable revision after initial start-up.

A great variety of measuring devices for determining the chemical composition of process flow streams is now available, and many of these instruments are being incorporated in automatic control systems.

Although automatic computers are used widely in economic control of process plants and organizations, the incorporation of computers in the process control loop to obtain fully automatic control (automation) has been explored only in preliminary fashion.

At the theoretical level three different developments extend the possibilities of treating control problems. These developments are: (1) dynamic characterization of processes from normal operating records without need for disturbing process operations; (2) methods of attacking problems involving nonlinear systems; and (3) improved techniques for sampled-data systems. This last development is of particular importance in systems which utilize digital computation within the control loop.

B. Future Directions

The preceding sections have summarized briefly what is involved in automatic process control and how control problems may be attacked quantitatively. In the course of the presentation some soft spots have been manifest. For one thing it is clear that we do not know how to specify the kind of controlled behavior that is optimum for a given plant or process. General or approximate specifications can be given but they are not optimal. Actually, control quality specification can be made only on the basis of the disturbances affecting the process, and at the present time there is virtually no reliable knowledge about process disturbances and where they apply.

We are only beginning to characterize chemical engineering processes in ways useful for control system synthesis. Only very simple processes have been analyzed with success, and much more work is needed not only in analyzing complex processes but also in developing improved and simplified methods of analysis. Most of the important industrial processes involve distributed parameters, interacting components, and nonlinearities, all of which make the applicability of servomechanism techniques at best an approximation.

Because research and development engineers are not able to deal effectively with control system design, it is likely that many potentially profitable chemical processes never see the light of day. Processes which cannot be managed by conventional control instrumentation hardware are just not studied or developed, although a little imagination and an appreciation of what is involved in controlled loop behavior might make the processes commercially feasible.

The most important need for the immediate future is the continuing education on an increasingly broad scale of process engineers as to the practical possibilities of automatic control. Techniques developed by servomechanism engineers for handling control system analysis and design must be exploited insofar as they are applicable to process control systems, and new techniques immediately pertinent to process control problems must be developed by process-oriented chemical engineers. In particular the chemical engineer must offer improved methods of characterizing his processes and process systems. He must be quick to adapt new measuring techniques to process needs and quick to recognize possibilities in new instruments.

In the long run economic advantage is the only justification for automatic control in the process industries. Chemical engineers bear the primary responsibility for demonstrating such economic advantages, since they are the engineers most directly concerned with process performance. They must be alert to the possibilities of radically revamping processes and procedures, and in particular they must not be bound by preconceived notions. For example, if there is a potential economic advantage in harnessing certain unwieldy chemical reactions by means of carefully integrated control systems and reactors of special design, the chemical engineer should willingly explore the possibilities. Similarly, the whole problem of maintaining large inventories of products and stocks in process operations to provide stabilizing ballast against fluctuating demands and relatively inefficient plant control, should be reappraised. These and other problems cannot be assessed properly unless there is a clear understanding of the possibilities and limitations of automatic control.

Nomenclature

- a intermediate coefficient
- A constant; heat transfer area; area normal to process flow
- b intermediate coefficient
- B constant
- C capacitance
- C_p heat capacity at constant pressure
- d differential operator
- D diameter
- e base of natural logarithms
- f frequency, cycles per unit time
- $G(s)$ transfer function
- h film coefficient for heat transfer
- j $\sqrt{-1}$
- K gain
- L length
- $\mathcal{L}$ Laplace operator
- M closed loop magnitude ratio
- n number of stages
- R resistance
- s Laplacian complex variable
- S maximum slope of reaction curve
- t time
- T time constant
- u fluid velocity
- U overall heat transfer coefficient
- W mass
- x distance

y intercept of maximum slope of reaction curve

GREEK LETTERS

∂ (small delta) partial differential operator

ζ (small zeta) damping ratio

Θ (capital theta) process variable

ρ (small rho) density, mass per unit volume

T (capital tau) time constant, time

ψ (small psi) phase angle, radians

ω (small omega) frequency, radians per unit time

SUBSCRIPTS

a zero deviation

c controlled variable; controller

d derivative

D delay, dead time

f feed

fl fluid

i inlet, input; inside

int integral

L point where $x = L$

n natural

o outlet, output

p proportional control; constant pressure

r reference, set point

s steam

w wall

0 (number) point where $x = 0$; ultimate

REFERENCES

A1. Aikman, A. R., *in* "Frequency Response" (R. Oldenburger, ed.), p. 141. Macmillan, New York, 1956.

B1. Bilous, O., and Amundson, N. R., *A.I.Ch.E. Journal* **1,** 513 (1955).

B2. Bilous, O., and Amundson, N. R., *A.I.Ch.E. Journal* **2,** 117 (1956).

B3. Bradner, M., *Instruments* **22,** 618 (1949).

B4. Brown, G. S., and Campbell, D. P., "Principles of Servomechanisms," Wiley, New York, 1948.

C1. Campbell, D. P., *Ind. Eng. Chem.* **47,** 409 (1955).

C2. Carslaw, H. S., and Jaeger, J. C., "Conduction of Heat in Solids," Chapter XIV, Oxford Univ. Press, London and New York, 1947.

C3. Catheron, A. R., and Hainsworth, B. D., *Ind. Eng. Chem.* **48,** 1042 (1956).

C4. Ceaglske, N., "Automatic Process Control for Chemical Engineers," Wiley, New York, 1956.

C5. Chestnut, H., and Mayer, R. W., "Servomechanisms and Regulating System Design," Vol. 1. Wiley, New York, 1951.

C6. Churchill, R. V., "Modern Operational Mathematics in Engineering," McGraw-Hill, New York, 1944.

C7. Close, C. D., *Control Eng.* **2,** No. 9, 97 (1955).

C8. Cohen, G. H., and Coon, G. A., *Trans. Am. Soc. Mech. Engrs.* **75,** 827 (1953).

C9. Cohen, W. C., and Johnson, E. F., *Ind. Eng. Chem.* **48,** 1031 (1956).

D1. Draper, C. S., McKay, W., and Lees, S., "Instrument Engineering," Vols. 1 and 2. McGraw-Hill, New York, 1952, 1953.

E1. Eckman, D. P., "Industrial Instrumentation," Wiley, New York, 1950.

E2. Evans, W. R., "Control-System Dynamics," McGraw-Hill, New York, 1954.

F1. Farrington, G. H., "Fundamentals of Automatic Control," Wiley, 1951.

F2. Flügge-Lotz, I., "Discontinuous Automatic Control," Princeton Univ. Press, 1953.

F3. Franks, R. G., and Worley, C. W., *Ind. Eng. Chem.* **48,** 1074 (1956).

G1. Goodman, T. P., and Reswick, J. B., *Trans. Am. Soc. Mech. Engrs.* **78,** 259 (1956).

J1. Janssen, J. M. L., *in* "Frequency Response" (R. Oldenburger, ed.), p. 131. Macmillan, New York, 1956.
J2. Johnson, E. F., *Chem. Eng. Progr.* **51,** 353 (1955).
J3. Johnson, E. F., *Chem. Eng. Progr.* **52,** 64 (1956).
J4. Johnson, E. F., and Bay, T., *Ind. Eng. Chem.* **47,** 403 (1955).
K1. Kochenburger, Ralph, *Trans. Am. Inst. Elec. Engrs.* **69,** 270 (1950).
L1. Lees, S., and Hougen, J. O., *Ind. Eng. Chem.* **48,** 1064 (1956).
M1. MacColl, L. A., "Fundamental Theory of Servomechanisms," Appendix. VanNostrand, Princeton, New Jersey, 1945.
N1. Nyquist, H., *Bell System Technical J.* **11,** 126–147 (1932).
O1. Oldenbourg, R. C., *in* "Automatic and Manual Control" (A. Tustin, ed.), p. 435. Academic Press, New York, 1952.
O2. Oldenbourg, R. C., and Sartorius, H., "Dynamics of Automatic Control" (translated and edited by H. L. Mason), Am. Soc. Mech. Engrs., New York, 1948.
O3. Oldenburger, R., ed., "Frequency Response," Macmillan, New York, 1956.
P1. Perry, J. H., ed., "Chemical Engineers' Handbook," 3rd ed., Section 19. McGraw-Hill, New York, 1950.
R1. Ragazzini, J. R., and Zadeh, L. A., *Trans. Am. Inst. Elec. Engrs.* **71,** 225 (1952).
R2. Rose, A., and Williams, T. J., *Ind. Eng. Chem.* **47,** 2284 (1955).
R3. Rutherford, C. I., *Inst. Mech. Engrs. (London) Proc.* **162,** 334 (1950).
S1. Sartorius, H., *in* "Automatic and Manual Control" (A. Tustin, ed.), p. 421. Academic Press, New York, 1952.
S2. Schwent, G. V., McGregor, W. K., and Russell, D. W., *ISA Journal* **3,** 274 (1956).
T1. Takahashi, Y., *in* "Automatic and Manual Control" (A. Tustin, ed.), p. 275. Academic Press, New York, 1952.
T2. Trimmer, J. D., "Response of Physical Systems," Wiley, New York, 1950.
T3. Truxal, J. G., "Automatic Feedback Control System Synthesis," McGraw-Hill, New York, 1955.
T4. Tsien, H. S., "Engineering Cybernetics," McGraw-Hill, New York, 1954.
W1. Williams, T. J., Harnett, R. T., and Rose, A., *Ind. Eng. Chem.* **48,** 1008 (1956).
Y1. Young, A. J., "An Introduction to Process Control System Design," Instruments, Pittsburgh, 1955.
Z1. Ziegler, J. G., and Nichols, N. B., *Trans. Am. Soc. Mech. Engrs.* **64,** 759 (1942).