Quantitative Analysis of Cascade Control

R. G. FRANKS[1] AND C. W. WORLEY[2]

Minneapolis-Honeywell Regulator Co., Brown Instruments Division, Philadelphia 44, Pa.

THE recent trends in the processing industries toward increased process flows and reduced storage capacities occasioned by the increased value of product have placed exacting demands on the control of process variables. Since product quality must be maintained with this increased throughput, a high performance control system is required. The use of cascaded controls, in which the desired value of one controlling unit is dictated by the control action of another, is often indicated whenever higher than ordinary control performance is economically justified.

The cascade control principle is not a new concept, for in the last few years it has been the subject of several articles (*1, 3*). These articles have been largely of a descriptive nature and have usually been concerned with a particular application of the cascade control principle. Although such publications have increased the understanding of this control principle, there is a need for a more general treatment of cascade control systems. For instance, it often is apparent that a cascade control system will give performance superior to a straightforward, single loop control system, but until now there has been little or no work carried out in this field to determine quantitatively the relative advantage of one system over the other. With the introduction of feed-back control techniques to the industrial automatic control field and the invaluable aid of the analog computer, it became possible to conduct a study to determine quantitatively the improvement in controllability of the cascade system over the single controller. Herein are presented data resulting from such a study which may be used to determine the increase in performance afforded by a cascade control system for processes with various ratios of dominant process lags.

[1] Present address, E. I. du Pont de Nemours & Co., Wilmington, Del.

[2] Present address, Electronic Associates, Inc., Princeton Computation Center, Princeton, N. J.

Method

The results of this study are based on the simulation of the process and control system elements on an analog computer. The technique used consists of representing the characteristics of the process and instrumentation components in transfer function form. These dynamic characteristics are then simulated on the computer using well-known computer techniques (*4*) and connected into a system according to the signal flow characteristics of the control system. Disturbances are introduced into the system, and the resulting transient response of the controlled variable is recorded. The response of a single loop control system is compared with the response of a cascade control system for similar disturbances. By means of a special evaluation circuit the numerical advantage of the cascade system for various types of processes and for different kinds of disturbances is determined.

Successful simulation naturally depends on a knowledge of the process characteristics. These characteristics often are difficult to determine accurately—the degree of difficulty depending on the accuracy required. Experience has shown that, despite the

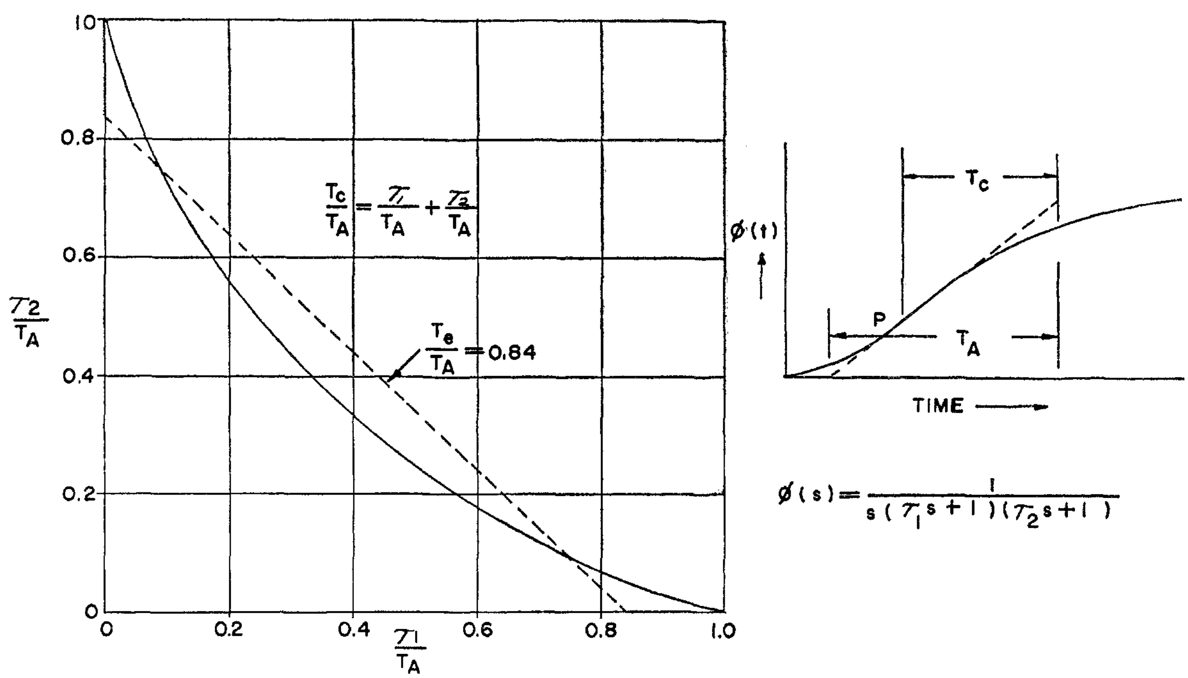


FIGURE 1. CURVE FOR DETERMINATION OF TWO TIME CONSTANTS FROM INDICIAL RESPONSE

complexity of most large physical systems, it is generally possible to approximate the characteristics of many processes to an acceptable degree.

In most physical systems there is usually one component whose time constant is large in comparison with the remaining time constants. This lag can be classified as a dominant lag and can be used to represent the dynamic characteristics of the series arrangement with acceptable errors. Generally, a ratio of time constants of 10:1 allows an approximation of the system characteristics by a dominant lag. The numerical values of the lags can be determined either by theoretical means, or, if a step transient test can be made to the system, an approximation can be made using a method developed by Oldenbourg (*5*). Oldenbourg's method for determining approximate time constants of multicapacity processes is based on the differential equations of a second-order system. Two series sequenced first-order systems have exactly the same form of equation as a second-order system. For example, the equation for two first-order systems in series is

$$[\tau_1\tau_2p^2 + (\tau_1 + \tau_2)p + 1]Y = X$$

and for the second-order system

$$[\tau_1\tau_2p^2 + (\tau_1 + \tau_2 + \tau_{1,2})p + 1]\,Y = X$$

When $\tau_{1,2} \ll \tau_1,\tau_2$ the equation for the second-order system becomes equal to the equation for the two first-order components in series. This similarity is useful in analyzing an experimentally found response of a multicapacity system.

The procedure for resolving the multicapacity system into two approximate independent first-order systems is as follows:

From the experimental response function the two times, T_A and T_c (Figure 1), are found and from them the time constants, τ_1 and τ_2, can be obtained. Here T_A is the projection on the asymptote of that segment of the tangent at the inflection point which is included between the time axis and asymptote, whereas T_c is the projection on the asymptote of that segment between inflection point and asymptote.

From an evaluation of the time response equation of the second-order system, times T_A and T_c will be related to the system time constants by

$$\frac{T_c}{T_A} = \frac{\tau_1}{T_A} + \frac{\tau_2}{T_A}$$

This equation can be solved using the curve shown in Figure 1. The ratio T_c/T_A is marked on each axis and a line joining them will intersect the curve at two points. The coordinates of any one point will give the values of τ_1/T_A, hence τ_2/T_A, and therefore τ_1 and τ_2. The other point is merely the inverse of these values as the system will respond the same way irrespective of the sequence of τ_1 and τ_2.

Process

In order to make this study as general as possible, a process with five passive first-order lags in series was selected as shown in Figure 2. These lags are grouped under P_1 and P_2, where P_1 is made up of a dominant lag of 10 seconds and two minor lags of 1 second; P_2 consists of a dominant lag τ_a and a minor lag of 0.1 τ_a where seven values of τ_a were selected as shown in Figure 2. The time scale of seconds was chosen for its suitability to the analog computer. As all the results are plotted against a ratio of time constants the time scale could be minutes or hours (which is more likely in industrial processes) as long as consistency is maintained. For convenience this ratio is expressed as P_2/P_1 where P_2 is the dominant time constant of the portion of the process appearing in the primary control loop and P_1 is the dominant time constant of that part of the process enclosed in the secondary controller loop.

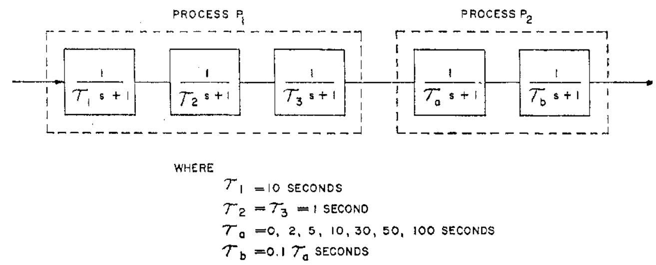


FIGURE 2. BLOCK DIAGRAM OF PROCESS USED IN CASCADE CONTROL STUDY

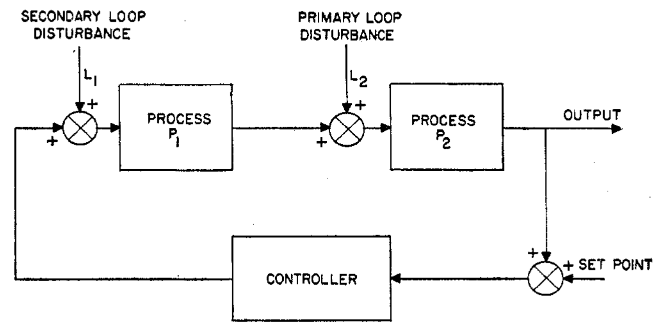


FIGURE 3a: BLOCK DIAGRAM OF SINGLE LOOP CONTROL SYSTEM

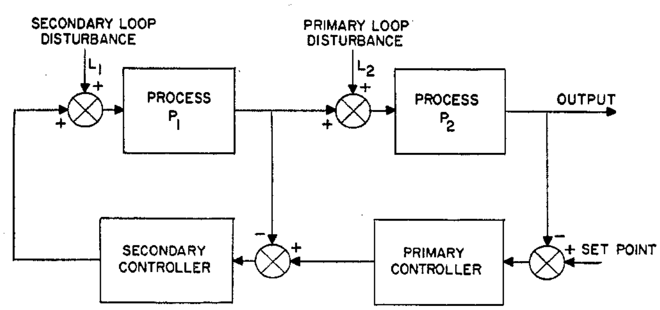


FIGURE 3b: BLOCK DIAGRAM OF CASCADE CONTROL SYSTEM

These seven processes then had a P_2/P_1 ratio—i.e., ratio of the dominant lags in P_1 and P_2—in the range 0 to 10. For each of these seven processes, the optimum controllability was determined under two conditions—single loop control as shown in Figure 3*a* and cascade control as shown in Figure 3*b*.

Controller

The single loop controller, the primary loop controller, and the secondary loop controller of the cascade control system used in this study were theoretical three-mode controllers with the rate section located in the process variable line ahead of the deviation section of the controller. The block diagram of the controller is shown in Figure 4 along with its theoretical transfer function.

Most pneumatic industrial controllers possess, in addition to their theoretical transfer functions, a second-order delay characteristic due to the pneumatic pilot relay and its tubing load. These delays were neglected in this study, as generally the majority of processes have considerably longer time constants and hence, by comparison, the controller delay becomes negligible. In actual operation the output of pneumatic controllers is limited to the standard 3 to 15 pounds per square inch range. This nonlinearity of controller action was simulated on the computer by limiting the output of the rate section and the two-mode controller to the voltage made analogous to the limits of actual controller output.

Control Criterion

Optimum controller settings were determined by the ultimate sensitivity method. The ultimate gain (maximum loop gain causing sustained oscillations) and period of oscillation were determined on the computer for all seven ratios of P_2/P_1. Using these values of ultimate gain and period, the approximate optimum controller settings were obtained by using the formula

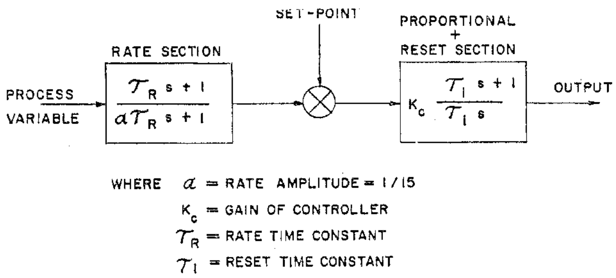


FIGURE 4. BLOCK DIAGRAM OF THEORETICAL 3-MODE CONTROLLER

developed by Pessen (*6*). For the cascaded control system, the secondary controller was "tuned" to P_1 using this technique and then optimized for a set-point disturbance. As P_1 remained constant for all seven cases these controller settings also remained constant. To determine the primary controller settings, the ultimate gain and period of oscillation of the secondary control loop in series with P_2 was determined on the computer, and again using Pessen's formula, the approximate controller settings were determined.

The next step was to optimize the controller settings for three disturbances, in each case starting from the nearly optimum settings obtained by the formula. At this point in the study, it was necessary to measure quantitatively the optimum control response of the system by choosing some factor that would be a measure of the response and would form a firm basis for comparison of response of various controller settings. For this purpose the integral of time × arithmetic error (ITAE No.) (*2*) was selected as being most suitable.

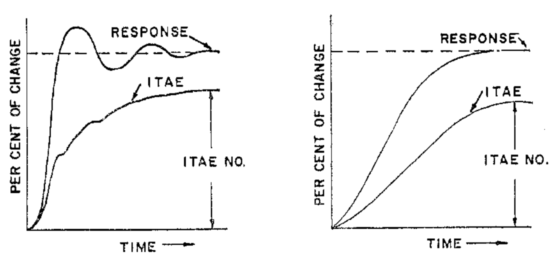


FIGURE 5. TRANSIENT AND ITAE RESPONSE FOR STEP DISTURBANCE TO SYSTEM

The integral of time × arithmetic error criterion is defined as

$$\text{ITAE} = \int_0^t t \cdot |\epsilon| \cdot dt$$

where t = time, and ϵ = arithmetic difference between actual output and desired output. The ITAE No. is a time weighted error and, if consistent units are used, the controllability of a system can be compared directly with that of another system. This control criterion brings several factors into account. The first and most important is the error itself which is the difference between the actual output and the desired output. As this error can oscillate to positive and negative quantities, it must be converted to the arithmetic value before it can be used. This error is multiplied by time (expressed in seconds, minutes, etc.) with the time origin coinciding with the start of the disturbance. The resulting product is continuously integrated until the function disappears which occurs when the error is finally reduced to zero.

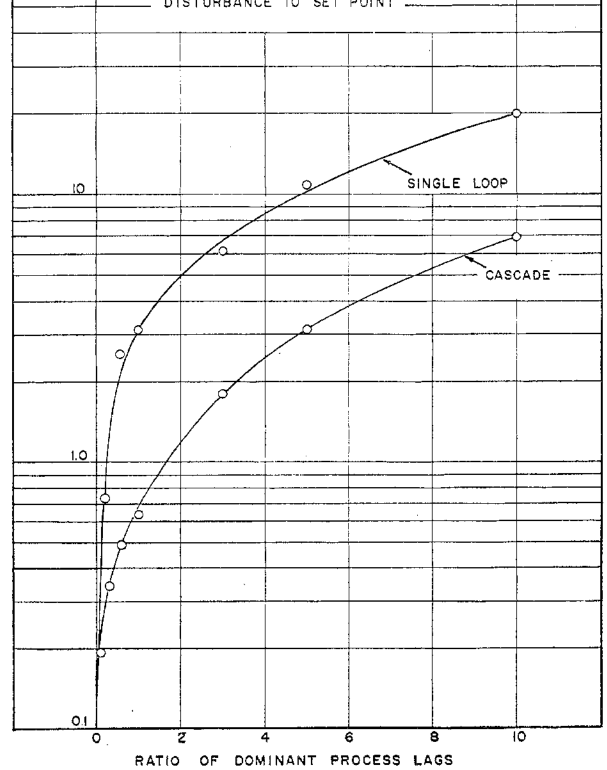
A simple example for an oscillatory and damped transient response to a step disturbance is shown in Figure 5. In the oscillatory example each time the output crosses the desired level, the slope of the ITAE curve is zero. The maximum value attained by the ITAE curve is called the ITAE No. and is the index representing the rapidity with which the system regains correspondence following a disturbance.

Clearly, to obtain the best control settings a number of runs had to be made until the smallest possible ITAE No. was obtained. This number then constituted the optimum ITAE value.

Computer Study

For all cases the process gain was unity and 0 to 100% measured variable from the process was sized as 0 to 100 volts on the computer. Thus a 0- to 100-volt input to the process from the controller produced a 0- to 100-volt output from the process. In the controller, the limits of output were set to 0 and 100 volts plus or minus depending on the sign of the signal at that point. By employing connections to a function switch, it was possible to switch from a single control loop to a cascaded control

FIGURE 6. I.T.A.E. NUMBERS FOR 5% STEP DISTURBANCE TO SET POINT



system which avoided any computer repatching when systems were changed.

The controllability of the control systems was investigated on the basis of their reaction to four disturbances as follows:

1. Start-up: 70% full scale change applied to the set point of the primary controller
2. Set-point change: 5% full scale step disturbance made to the set point
3. Secondary loop disturbance: ±20% full scale step change applied to the input of P_1
4. Primary loop disturbance: 5% full scale step change applied to the input of P_2

The location of these disturbances is shown in the block diagram of Figure 3*b*.

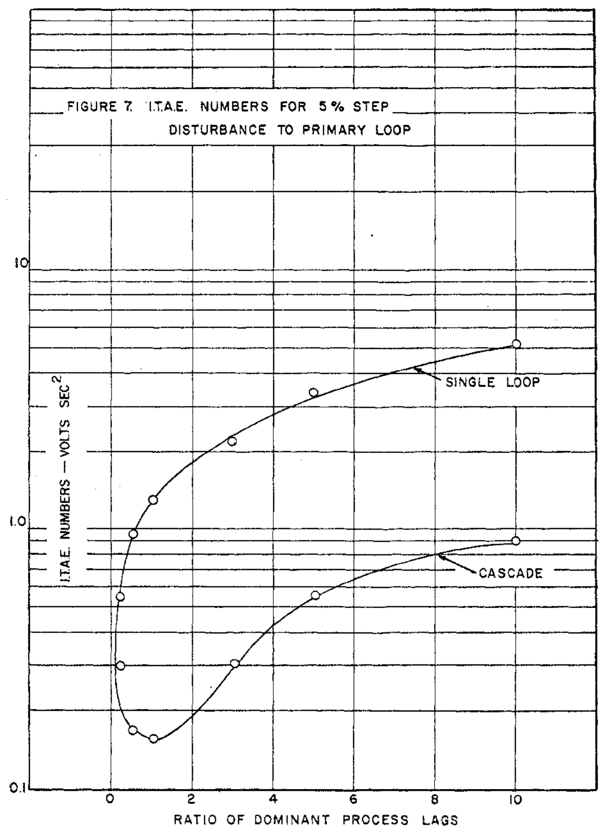


FIGURE 7. I.T.A.E. NUMBERS FOR 5% STEP DISTURBANCE TO PRIMARY LOOP

Disturbances 2, 3, and 4 were made while the system was operating at a steady-state level of 50% full scale. These disturbances were sized so as to keep the system operating in the linear region. This was checked by applying the disturbance and monitoring the controller output to ascertain whether the transient voltage limited at 0 or 100 volts. For start-up, the system operated during most of the transient period in the saturated state, so this series of tests differed from the other three mainly in that the operation was largely nonlinear.

Generally some 40 to 50 runs were necessary to obtain the optimum control conditions and hence the minimum ITAE No. for any one disturbance.

Results

Start-Up. A complete set of runs was made for all seven processes using both single loop control systems and cascaded control systems. The optimum ITAE Nos. obtained for all seven processes controlled by a single controller were approximately the same as the ITAE Nos. of the corresponding cascade controlled process. The reason for this is that in both cascaded and single controller systems, the input to the process is at its maximum value for most of the elapsed transient recovery time and the controller is ineffective until the output of the process approaches the desired level where it enters into the linear region and the controller can start operating. As 95% of the ITAE No. is formed during the nonlinear period, the difference between the two conditions of operation are small and, therefore, the results are inconclusive. It should be pointed out that the difference between cascaded and single loop control will become more apparent on start-up if the process gain is increased to higher values (in this study it was unity) which effectively increases the linear region of operation.

Set-Point Disturbance. As explained previously, the magnitude of the set-point disturbance to the primary controller was made small enough so as not to saturate either controller. Figure 6 shows the results of these trials as ITAE No. versus the ratio of P_2/P_1. When $P_2/P_1 = 10$—i.e., the time constants of the second half of the process are 10 times greater than the time constants of the first half—the ITAE No. for the single controller is three times greater than that for the cascaded controllers. This ratio increases to a maximum of 5 as the P_2/P_1 ratio decreases to 1, which means that quantitatively the cascaded system is always three to five times better than the single loop case, depending on where the primary element of the secondary loop controller is located. If a choice is available for locating this primary element, and set-point disturbances are likely to be frequent, then the optimum process P_2/P_1 division is 1:1.

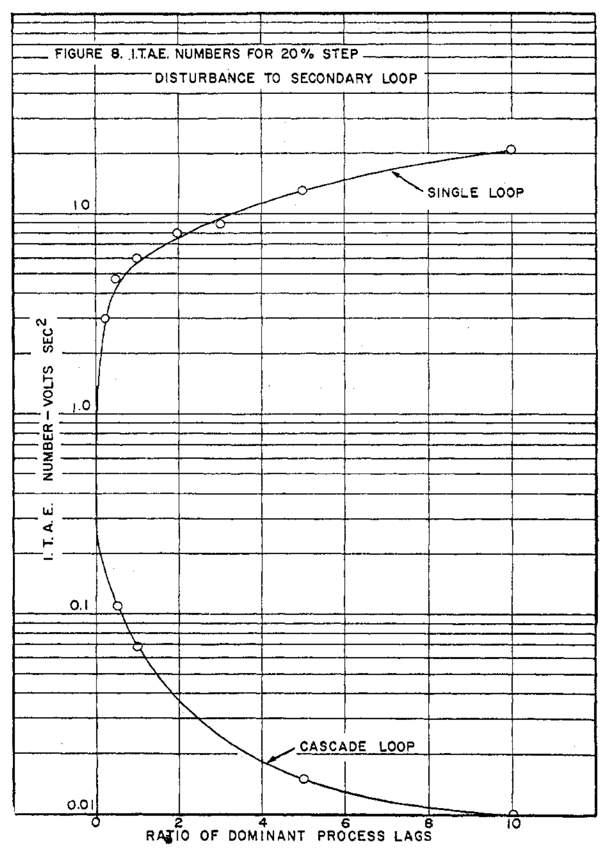


FIGURE 8. I.T.A.E. NUMBERS FOR 20% STEP DISTURBANCE TO SECONDARY LOOP

As the P_2/P_1 ratio approaches zero—i.e., $P_2 = 0$—a condition occurs in the cascade system where a complete controller is in parallel with the rate section of the secondary loop controller. With $P_2 = 0$ it was difficult to optimize the controller settings on the primary loop controller and in any case, no improvement over single loop control could be achieved. The advantage of cascade control over single loop control then diminished from a maximum of 5 at $P_2/P_1 = 1$ to 0 when $P_2/P_1 = 0$.

Primary Loop Disturbance. This is defined as a disturbance that enters the P_2 portion of the process and is detected initially by the primary loop controller. For this study it was assumed that this disturbance is not detected by the secondary loop controller—i.e., it does not reflect back in the process and affect the output of P_1. Using the techniques described, the controllers were optimized for this disturbance to give the smallest possible ITAE No. in both single and cascade control systems for all seven processes. The curves of ITAE No. versus P_2/P_1 ratio for this disturbance are shown in Figure 7.

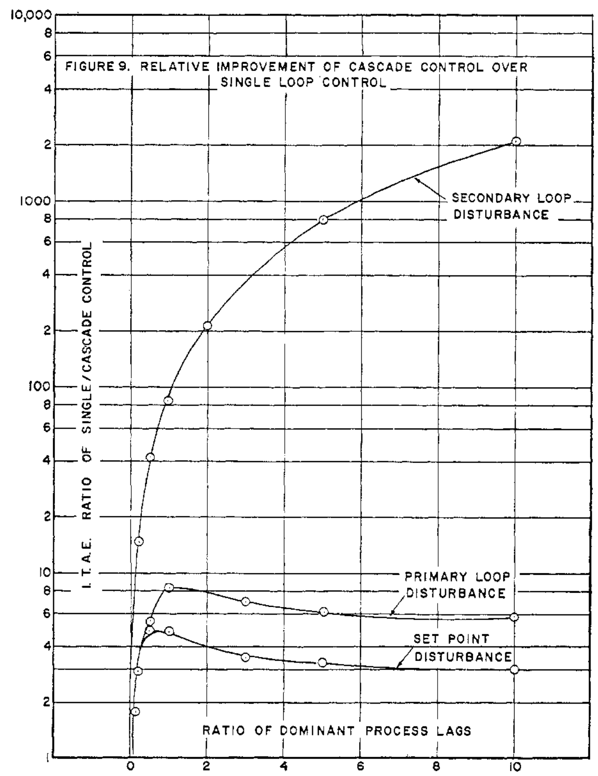


FIGURE 9. RELATIVE IMPROVEMENT OF CASCADE CONTROL OVER SINGLE LOOP CONTROL

For the single loop control system the ITAE No. increases as the P_2/P_1 ratio increases. For the cascade system, however, the ITAE No. at first decreases, reaching a minimum at $P_2/P_1 = 1$, then starts to increase. The relative improvement of the cascaded system over the single controller is a maximum at P_2/P_1 ratio of unity. When $P_2 = 0$, apparently a single controller is superior to a cascaded system. This is probably due to the low rate and gain settings necessary on the primary loop controller to prevent the system from becoming unstable. The secondary loop controller is already optimized to P_1 and with $P_2 = 0$, the primary loop controller merely becomes an additional control to P_1, thus increasing the gain of the secondary loop. The low gain settings on the primary loop controller cause it to behave sluggishly and thus causes the cascade system to give results inferior to that of a single controller.

Secondary Loop Disturbance. This is defined as a disturbance which is introduced at the entrance of the P_1 portion of the process. It is termed a secondary loop disturbance because whereas the L_2 and set-point disturbances occurred in the primary loop, this disturbance occurs within the secondary loop. This is important as it is with this type disturbance that the real value of cascade control becomes apparent. The ITAE No. measuring technique is the same as that described for the other disturbances to the system. The results are shown plotted in Figure 8. The single control systems are similar to the L_2 and set-point disturbances in that the ITAE No. increases as the P_2/P_1 ratio increases. On the other hand, for the cascade system the ITAE No. decreases appreciably as the P_2/P_1 ratio increases and reaches a minimum at $P_2/P_1 = 10$. The maximum improvement of the cascade system over the single controller system is at P_2/P_1 ratio of 10, where the relative improvement is 2000:1. It would seem then that for maximum benefit from a cascade controller to a disturbance in the secondary loop, it is best to locate the primary element as closely as possible to the course of the disturbance, thus reducing P_1 and increasing the P_2/P_1 ratio.

For convenience, the results of this study have been grouped into the single set of curves shown in Figure 9. These curves, which are plots of the ratio of ITAE Nos. for various values of P_2/P_1, show the relative improvement in controllability that can be obtained from the use of a cascaded control system. With some knowledge of the process characteristics these data may be used to quickly establish the advantage of using a cascade control system for the control of a particular process.

Example

As an example of how the results of this study may be used in designing process control systems, consider control of the chemical reactor process shown schematically in Figure 10. This process is already in operation with the product temperature being controlled directly by a single loop control system. The cascade control system shown in Figure 9 has been suggested as a means of improving control of the product temperature. The problem, therefore, becomes one of determining any improvement in controllability to be gained by using the cascaded system.

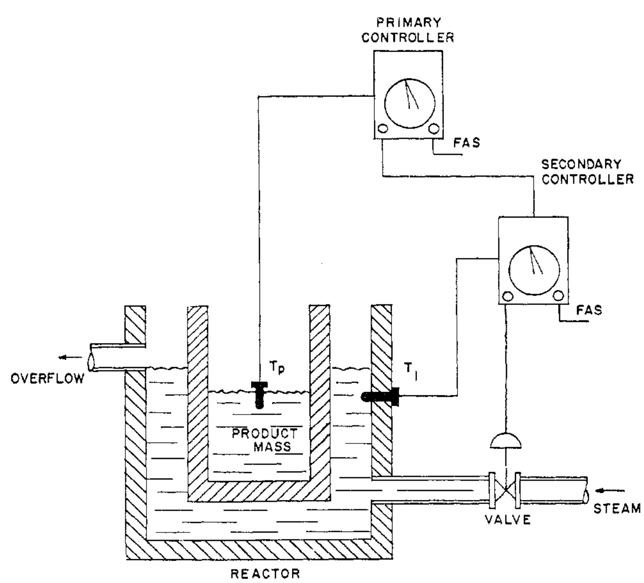


FIGURE 10. CASCADE CONTROL SYSTEM FOR CHEMICAL REACTOR

The time lags associated with the process could be obtained from a theoretical analysis, but since the process is already in operation it is convenient to determine them experimentally. The lags associated with the P_1 portion of the process can be determined by applying a step change to the steam valve and

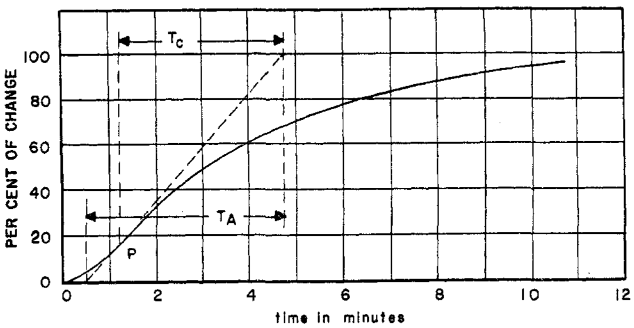


FIGURE 11a. TIME RESPONSE OF JACKET TEMPERATURE T_1 FOR A STEP CHANGE TO STEAM VALVE

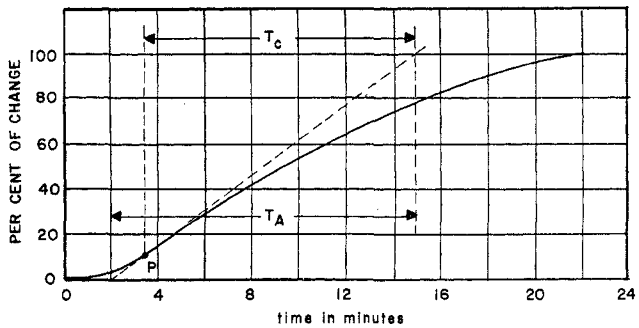


FIGURE 11b. TIME RESPONSE OF PRODUCT TEMPERATURE T_p FOR STEP CHANGE IN PRODUCT MASS

recording the resulting transient response of the jacket temperature. Then using the method described by Oldenbourg this response curve can be broken into approximate lags which can be used to represent that part of the process. The lags of the remaining part of the process (P_2) can be obtained from a transient curve of the product temperature to an immediate change in the mass being heated in the reactor. This change may be effected by the addition of additional charge needed for the chemical reaction.

The time response of the product temperature (T_p) and the jacket temperature (T_1) for step changes in the steam valve and the product mass, respectively, are shown in Figure 11. From the indicial response curve for the P_1 portion of the process, T_A is 4.3 minutes and T_c = 3.6 minutes which gives T_c/T_A ratio of 0.84. Using the curve included in Figure 1 the two time constants which can be used to represent this part of the process are τ_1 = 3.2 minutes and τ_2 = 0.4 minute. In a like manner, for the P_2 part of the process, T_A = 13 minutes and T_c = 11.5 minutes giving T_c/T_A = 0.9 minute. The time constants representing this part of the process are τ_3 = 10 minutes and τ_4 = 0.65 minute.

The dominant time constant of the P_1 portion of the process is equal to 3.2 minutes and 10 minutes for the P_2 part of the process. This gives a ratio of dominant time constants of $P_2/P_1 \approx 3$. From the curves of Figure 9 the relative improvement in controllability which will be gained by the use of the cascaded system is 3.6 to 1 for a set-point disturbance, 7.4 to 1 for a primary loop disturbance, and 400 to 1 for a secondary loop disturbance. The primary loop disturbance would correspond to a change in the mass of the product in the reactor, while the secondary loop disturbance would result from a change in steam flow or quality. One can conclude, therefore that the cascaded system would be much more effective in controlling the process should the steam supply be poorly regulated. The cascaded system also offers some advantage for secondary loop disturbances although the improvement is not as spectacular.

Conclusions

The results of this study show that the control of many processes may be improved by the use of a cascade control system. The amount of improvement depends on the nature of the process. A quantitative index shows how the improvement due to cascade control varies with the process characteristics.

Despite the utility of these data the relative improvement in controllability shown by its use should not be the sole justification for a cascaded system. Although the data obtained are as accurate as the techniques of simulation will allow, its use is based on certain approximations. These approximations may well overshadow the expected improvement in some marginal examples. Also, the service of the process could be such that a criterion of optimalization other than minimum error should be used. A performance specification calling for a limit on the overshoot of a process variable can be cited as an example. There are also economic considerations which must be taken into account. The process economics may be such that a minimum error cannot justify the added expense of additional instrumentation. The data, therefore, should not be regarded as the sole criterion for the selection of a cascade control system. Rather it is presented with the hope that it may be used as an effective tool for the design of process control systems.

Although this study was carried out for a reasonably wide range of processes there are certain areas that would benefit from further investigation. These investigations might include different types of control functions in either the primary or secondary controllers. The effect of dead time or other process peculiarities in the primary or secondary loops should be considered. The results of this study should offer encouragement for the solution of other basic industrial control problems.

One of the primary factors brought out by this study is the utility of the analog computer for solving industrial control problems. The analog computer makes it possible to set up and study complicated devices and systems. It provides quick answers to engineering problems and furnishes insight into and an understanding of system problems. Without its use the study reported in this paper could not have been made.

Literature Cited

(1) Bailey, W. R., "Application of a Variable-Ratio Pneumatic Relay to the Automatic Cascade Control of Flow," Instrument Soc. Am., No. 52-9-1.
(2) Graham, D., Lathrop, R. C., "The Synthesis of Optimum Transient Response," Am. Inst. Elec., Engrs., No. 53-249.
(3) Hartz, H. J., *Chemical Processing* (1954).
(4) Korn, G. A., Korn, T. M., "Electronic Analog Computers," McGraw-Hill, New York, 1952.
(5) Oldenbourg, R. C., Sartorius, H., "The Dynamics of Automatic Control," Am. Soc. Mech. Engrs., February 1948.
(6) Pessen, D. W., "Optimum Three-Mode Controller Settings for Automatic Start-Up," Am. Soc. Mech. Engrs., No. 52-A-58.

RECEIVED for review January 21, 1956. ACCEPTED April 6, 1956.

END OF ENGINEERING, DESIGN, AND PROCESS DEVELOPMENT SECTION