Introduction to Control Systems

K. Craig

Control system design is not a narrowly specialized mathematical exercise, but is a widely applied technology with close connections to, and important impacts on, overall system design. The ability to design and implement analog and digital control systems, with their associated analog and digital sensors, actuators, and electronics, is an essential skill of every engineer, as everything today needs controls! Control design is not just for specialists anymore!

The intelligent design and use of control systems requires that one have:

- Knowledge of the basic modes of control that have been devised and the characteristic performance features of each. This allows one to generate one or more design concepts that have potential for success.
- Familiarity with available hardware so that commercially-available components to implement the design concepts can be selected.
- Competence in modeling of physical systems with suitable equations, using judicious assumptions.
- Facility in the use of analytical, simulation, and experimental techniques for determination of system response and suggesting design changes.

An engineer must strive to develop insight into the problems of control and intuition about methods to solve them, emphasizing design in parallel with analysis techniques, showing the unity among several individual design techniques, and synthesizing them into a toolbox of problem-solving methods. The challenges to the engineer are many:

- Design as well as analysis techniques need to be mastered.
- Control is an active field of research and hence there is a steady influx of new concepts, ideas, and techniques that an engineer needs to evaluate for potential applications.
- Control is an interdisciplinary field that requires an interdisciplinary background.

All engineers must meet these challenges head on, as all engineers should be able to design and implement a control system as part of an overall design. This document provides a brief introduction to this very important area.

⇒ Role of Control Systems in Engineering System Design

Control systems play an important role in almost every area of engineering design and control concepts have a widespread and significant impact on many aspects of engineering. All types of engineered products and services depend more and more on associated control systems for their optimum functioning. These control systems are increasingly considered to be an integral part of the overall system rather than afterthought add-ons. The list of the many applications of control systems seems endless: energy production, materials production, vehicles and transport systems, construction equipment, manufacturing equipment, agricultural equipment, consumer goods, computers and peripherals, military applications, communications, measurement systems, medical equipment. The list seems to show that everything needs a control system. Although one does not have to be a controls specialist to design multidisciplinary engineering systems, one does have to have an understanding of the fundamentals of control system design and implementation.

In every control system, there is some device to be controlled that is called the *process, plant*, or *controlled system,* as shown in Figure 1. Process *inputs* are flows of energy and/or material that cause the process to react or respond. Mathematically, inputs are considered to be known or assumed and are classified into *manipulated inputs* (subject to our control) and *disturbance inputs* (undesirable and unavoidable effects beyond our control). Associated with the process are some *response variables* which we require to behave in some specified fashion. The need for controls can arise from a requirement for *command following*, *disturbance rejection*, or both. So one must be sufficiently clever in the control of the manipulated inputs to either cause a desired process response, or counteract the effects of disturbances, or both.

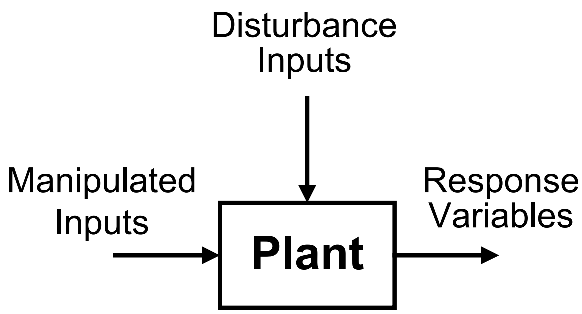

Figure 1. Process Input / Output Configuration

The design of every engineered device/system initially proceeds from specifications, a list of functions to be performed. Following the cardinal rule of good design, to strive for simplicity, one attempts to configure the device/system so that specifications may be met with minimum equipment and no, or only the most rudimentary, controls. If specifications cannot be met or, as seems to be inevitable, customers soon demand improved specifications, the device/system design must be refined. This refinement always runs into practical or theoretical limitations. Then the addition of suitable controls often allows significant further improvements to be made that would have been unavailable by other means. This evolutionary process of refinement of specifications and designs applies to all products and services. Indeed, everything needs controls and today the original invention most often includes control aspects that are vital to performance. Conservation of energy and materials, and continual increase in labor productivity are vital to the maintenance and improvement of world living standards. Since these goals invariably require improved performance of all technical processes, the increasingly important role of control in engineering design appears.

⇒ Classification of Control System Types

The broadest overall classification separates control systems into two fundamental types: *open-loop* and *closed-loop (feedback).* Figure 2 shows a functional block diagram of an open-loop control system.

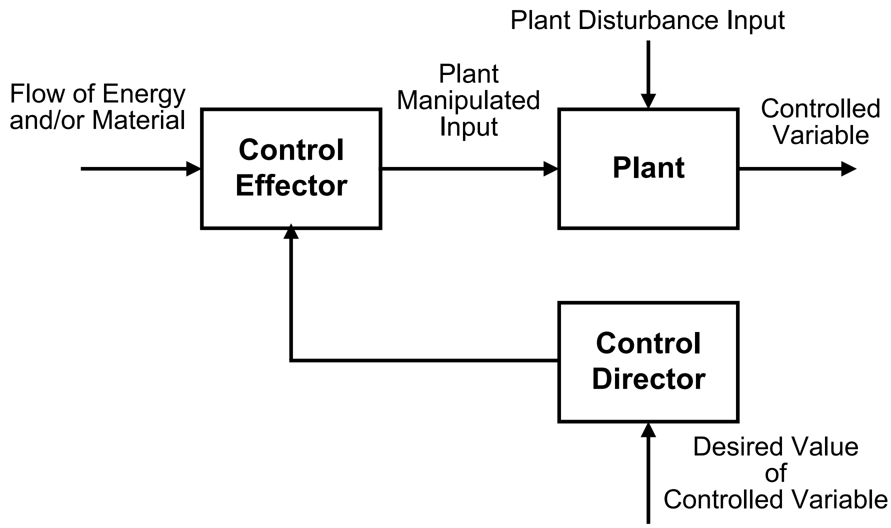

Figure 2. Basic Open-Loop Control System

In open-loop systems, the process response variable of interest, the *controlled variable*, is determined by the combined effects of the disturbance input and manipulated input. The manipulated input (flow of energy and/or material) is varied by the *control actuator* in response to signals from the *controller*. The controller receives information input as to the *desired value of the controlled variable* and translates this into a *control signal* for the control actuator by implementing the *control law* that is built into the controller. Open-loop systems of this basic type are often satisfactory if:

- disturbances are not too great
- changes in the desired value of the controlled variable are not too severe
- performance specifications are not too stringent

A refinement of the basic open-loop system is the *input-compensated (feedforward) open-loop system*. There are two types of open-loop systems in this category: *disturbance-compensated* and *command compensated*.

In a disturbance-compensated open-loop system, shown in Figure 3, the process manipulated input is derived not only from the desired value command but partially from a measurement of the disturbance. Implementation of such a scheme requires that we:

1. must be able to measure the disturbance
2. must be able to estimate the effect of the disturbance on the controlled variable, so we can compensate for it

Disturbance-compensation can be used by itself or in combination with feedback control.

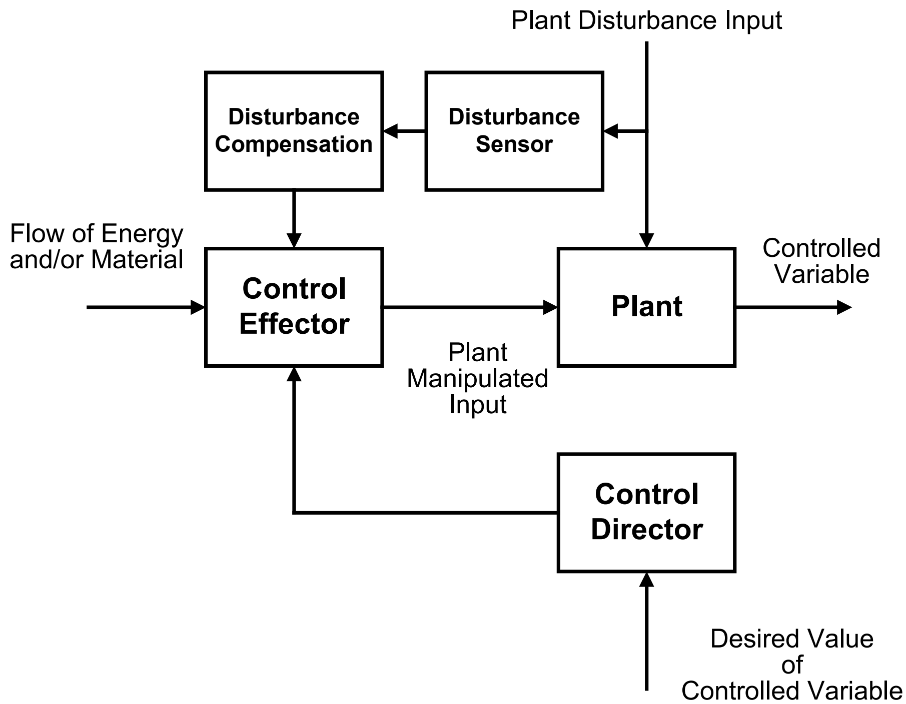

Figure 3. Disturbance-Compensated Open-Loop System

Although not as common as disturbance compensation, command-compensated open-loop systems (and combinations of these with feedback schemes) also exist. Here, based on knowledge of the process characteristics, the desired value command is augmented by the command compensator to produce improved performance. This is shown in Figure 4.

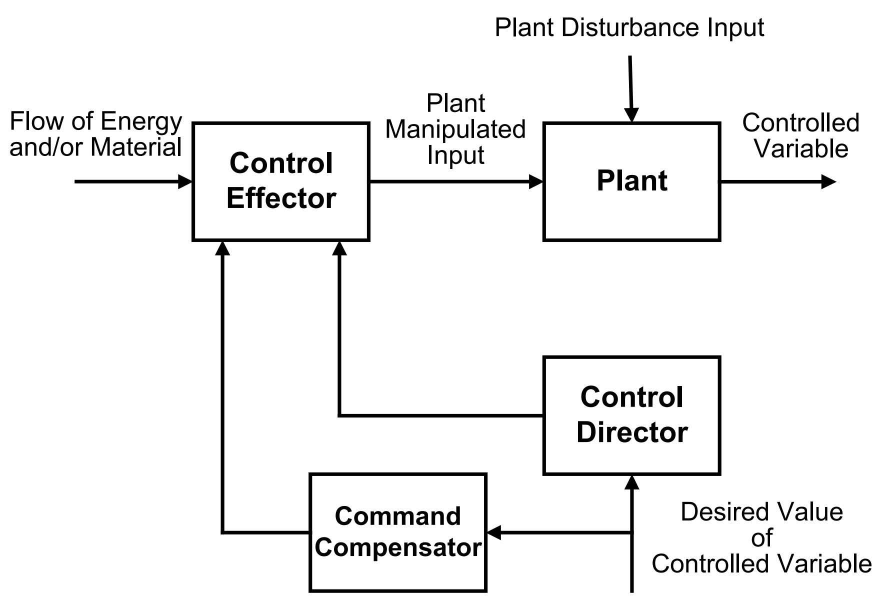

Figure 4. Command-Compensated Open-Loop System

Open-loop systems without disturbance or command compensation are generally the simplest, cheapest, and most reliable control schemes and should be considered first for any control task. If specifications cannot be met, disturbance and/or command compensation should be considered next. Design of these various types of open-loop systems does not generally require any specialized control theory, basic system dynamics being sufficient in most cases. When conscientious implementation of open-loop techniques by a knowledgeable designer fails to yield a workable system, the more powerful closed-loop (feedback) methods should be considered. Although the analytical design of feedback systems is currently based on a well-developed mathematical theory, the basic concept is quite obvious and reasonable. Figure 5 makes clear the operating principle of feedback systems and their basic advantages over open-loop systems.

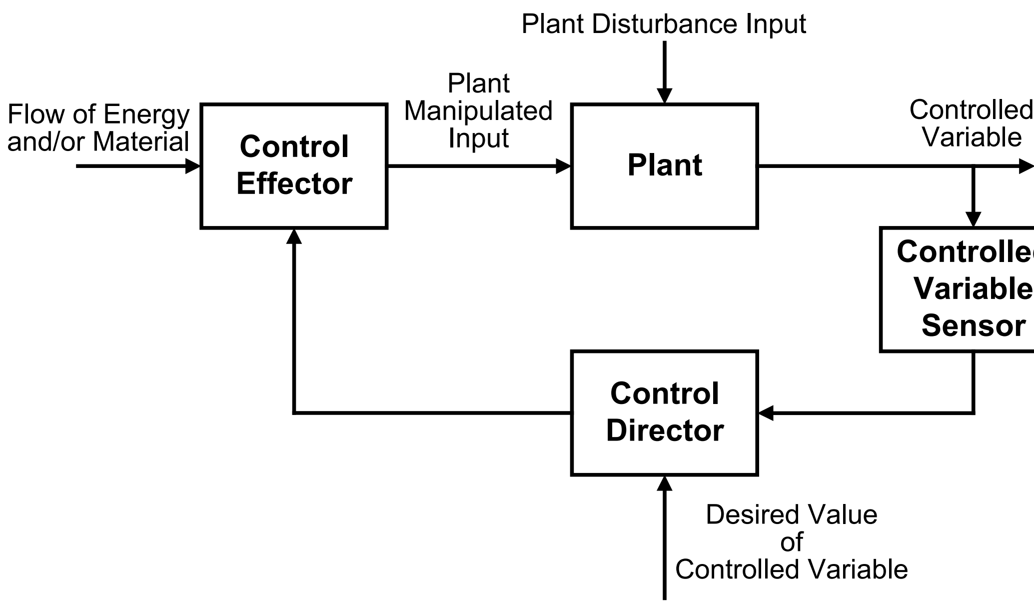

Figure 5. Basic Closed-Loop (Feedback) Control System

An open-loop system can be converted to a closed-loop by adding the functions of:
1. measurement of the controlled variable
2. comparison of the measured and desired values of the controlled variable

Errors between commanded and actual values of the controlled variable will tend to be corrected no matter what the source. This includes errors due to changing commands, process disturbances, disturbances to equipment other than the process, and changes in hardware parameter values. The one exception (which is often critical, however) is the controlled-variable sensor. If this sensor gives wrong information, the feedback system has no way of correcting for this. Feedback control depends vitally on accurate measurement.

The fact that open-loop systems never measure the controlled variable is the basis of their possible inadequacies. They fundamentally rely on conditions staying close to design values. When system parameters and/or disturbances depart from "normal" and cause the controlled variable to wander from the desired value, the open-loop system is unaware of such changes.

Even a disturbance-compensated open-loop system corrects only for the disturbance or disturbances measured; all others go uncorrected. Furthermore, changes due to wear, aging, environmental effects, and the like in the disturbance sensor/comparator, control actuator, and/or process cause the compensation to be imperfect. An associated problem, possible instability, accompanies the many benefits of feedback, however, this phenomenon is reasonably well understood and is controllable by proper design.

⇒ **Basic Benefits of Feedback Control**

Although a detailed exploration of the possibilities and problems of feedback control require a considerable amount of study, some essential characteristics can be illustrated quite easily. Consider first the open-loop control of a first-order process, shown in Figure 6. Note that the differential operator D in the first-order process mathematical model is defined as d/dt.

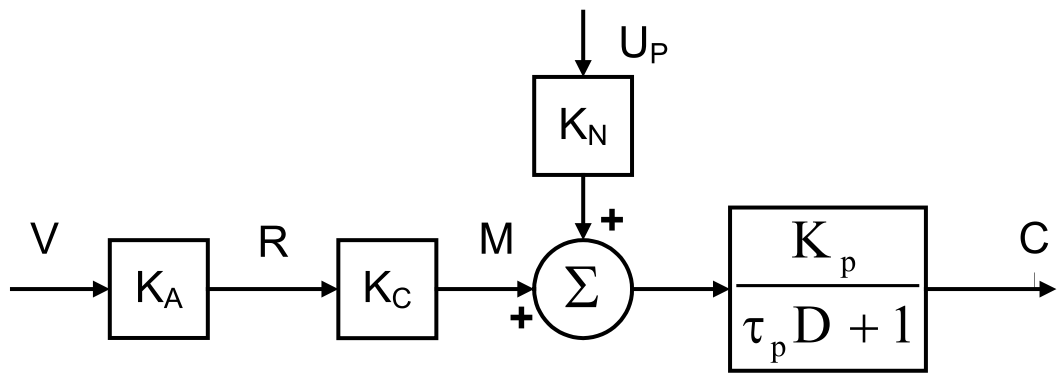

Figure 6. Open-Loop Control of a First-Order Process

Controlled variable C is related to its desired value V and process disturbance U_p by the differential equation:

$$\tau_p \frac{dC}{dt} + C = \left(K_A K_C K_p\right) V + \left(K_N K_p\right) U_P$$

where $\tau_P \equiv$ process time constant, $K_C \equiv$ controller sensitivity, $K_A \equiv$ sensitivity of the reference input element, $K_N K_P \equiv$ process sensitivity to U_P, $K_P \equiv$ process sensitivity to M.

Let's choose K_A such that $K_A K_C K_P = 1.0$, so that for any steady V, with no disturbance, C = V. With $U_P = 0$, solving the differential equation for a step input V_S of V gives:

$$C = K_A K_C K_P \left(1 - e^{\frac{-t}{\tau_P}}\right) V_S$$

A step U_{PS} with V = 0 gives:

$$C = K_N K_P \left(1 - e^{\frac{-t}{\tau_P}}\right) U_{PS}$$

We see that any change in $K_AK_CK_P$ from the design value of 1.0 results in a directly proportional error in *C*. A disturbance U_{PS} causes an additional error $K_NK_PU_{PS}$. While disturbance compensation (feedforward) schemes can reduce the error due to U_{PS}, errors due to changes in $K_AK_CK_P$ are not amenable to such improvement. Should faster response to V be desired, command compensation may be possible.

Now consider the response of a closed-loop (feedback) control system to a step command, as shown in Figure 7.

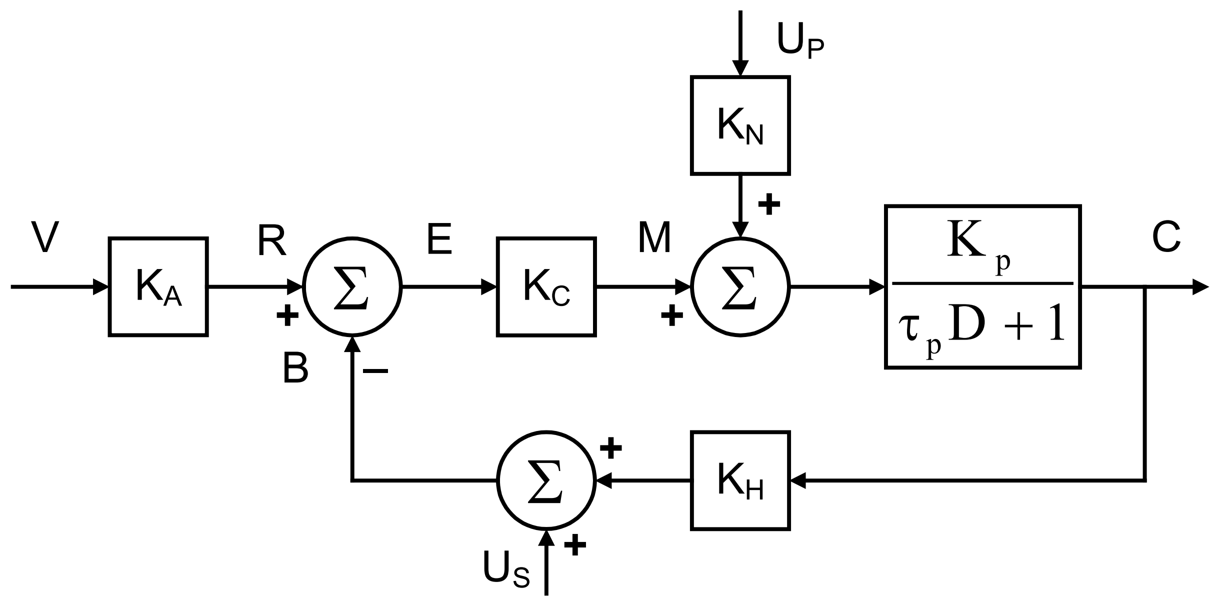

Figure 7. Closed-Loop (Feedback) Control of a First-Order Process

From Figure 7 we may write the following equations:

$$\left\{\left[K_AV-\left(U_S+K_HC\right)\right]K_C+U_PK_N\right\}\frac{K_P}{\tau_PD+1}=C$$

$$\left(K_AK_CV-K_CU_S-K_CK_HC+U_PK_N\right)K_P=\left(\tau_PD+1\right)C$$

$$K_AK_CK_PV-K_CK_PU_S+K_NK_PU_P-K_HK_CK_PC-C=\left(\tau_PD\right)C$$

$$K_AK_CK_PV-K_CK_PU_S+K_NK_PU_P-C\left(1+K_HK_CK_P\right)=\left(\tau_PD\right)C$$

As a result, the system differential equation is:

$$\left(\frac{\tau_P}{1+K_CK_PK_H}D+1\right)C=\frac{K_CK_PK_A}{1+K_CK_PK_H}V+\frac{K_NK_P}{1+K_CK_PK_H}U_P-\frac{K_PK_C}{1+K_CK_PK_H}U_S$$

System linearity allows separate consideration of the three inputs V, U_P, and U_S. Let us first take V as a step input V_S, and $U_P = U_S = 0$.

Therefore, solving the differential equation results in:

$$C = \frac{K_C K_P K_A}{1 + K_C K_P K_H}\left(1 - e^{\frac{-t}{\tau_{CL}}}\right)V_S$$

$$\tau_{CL} \equiv \frac{\tau_P}{1 + K_C K_P K_H} \equiv \text{ closed-loop system time constant}$$

In every feedback system, the static sensitivity (also called the steady-state gain) between signals E and B is the single most important design parameter. It is given the name loop gain and the symbol K (here $K \equiv K_C K_P K_H$). Generally, all aspects of control system performance (steady-state accuracy, speed of response, etc.) improve when K is made larger. Therefore large K is a basic design goal. There is, however, always an upper limit on K, beyond which system stability suffers.

The time constant τ_{CL} which governs the speed of response of the controlled variable C can be made much smaller (faster) than τ_P if K (here $K \equiv K_C K_P K_H$) is made large compared with 1.0. Thus if the process design has brought τ_P to its minimum feasible value, feedback control allows significant further improvements in the speed of response of C to V, without any changes in the process itself. This capability for overcoming apparent limitations in basic hardware performance is one of the major contributions of feedback. There is nothing magical about the speedup of C's response even though the process itself is as slow as before. The manipulated variable M simply over responds initially, as can be seen from its differential equation (with U_S = 0 and U_P = 0):

$$M\left(\frac{K_P}{\tau_P D + 1}\right) = C$$

But

$$\left(\frac{\tau_P}{1 + K_C K_P K_H} D + 1\right)C = \frac{K_C K_P K_A}{1 + K_C K_P K_H} V$$

$$\left(\tau_{CL} D + 1\right)C = \frac{K_C K_P K_A}{1 + K} V$$

Therefore

$$M\left(\frac{K_P}{\tau_P D + 1}\right) = \left(\frac{K_C K_P K_A}{1 + K}\right)\left(\frac{V}{\tau_{CL} D + 1}\right)$$

$$\left(\tau_{CL} D + 1\right)M = \left(\frac{K_C K_A}{1 + K}\right)\left(\tau_P D + 1\right)V$$

Feedback control achieves this increased speed of response by the initial over-response of process input M. Open-loop, feedforward, command compensation achieves a similar increased speed of response also by the initial over-response of process input M, but it does so by augmenting the command. However, they share the basic limitation on the degree of improvement possible, i.e., saturation nonlinearity caused by attempting excessive peaking in M.

Let's turn to steady-state behavior. With $U_P = 0$ and $U_S = 0$, we have seen:

$$C = \frac{K_C K_P K_A}{1 + K_C K_P K_H}\left(1 - e^{\frac{-t}{\tau_{CL}}}\right) V_S$$

If we choose $K_A = \frac{1 + K_C K_P K_H}{K_C K_P}$, then $C = V_S$ in the steady state. In actual practice K_A and K_H are normally made equal and $K >> 1$, so that $\frac{K_C K_P K_A}{1 + K_C K_P K_H} \approx 1$ and C is very nearly equal to V_S in the steady state. More importantly, if $K >> 1$, changes in K_C and/or K_P now have much less effect on accuracy than in an open-loop system. Thus if we can make the loop gain high enough, system accuracy becomes very insensitive to changes in hardware parameter values other than K_A and K_H.

Now take $V = 0$ and $U_S = 0$ and apply a step disturbance U_{PS}. The system differential equation reduces to:

$$\left(\frac{\tau_P}{1 + K_C K_P K_H} D + 1\right) C = \frac{K_N K_P}{1 + K_C K_P K_H} U_P$$

$$C = \frac{K_N K_P}{1 + K_C K_P K_H}\left(1 - e^{\frac{-t}{\tau_{CL}}}\right) U_{PS}$$

C responds exponentially, with time constant τ_{CL}, and levels off at a steady-state error of $\left(\frac{K_N K_P}{1 + K}\right) U_{PS}$ compared with $K_N K_P U_{PS}$ for an open-loop system. For a loop gain $K = 100$, the error is 101 times less for the closed-loop system than for the open-loop system.

Finally, take $V = 0$ and $U_P = 0$ and U_S = step input U_{SS}. The system differential equation reduces to:

$$\left(\frac{\tau_P}{1 + K_C K_P K_H} D + 1\right) C = -\frac{K_C K_P}{1 + K_C K_P K_H} U_S$$

$$C = -\frac{K_C K_P}{1 + K_C K_P K_H}\left(1 - e^{\frac{-t}{\tau_{CL}}}\right) U_{SS}$$

In the steady state

$$C = -\frac{K_C K_P}{1 + K_C K_P K_H} U_{SS} = -\frac{K_P K_C}{1 + K} U_{SS} = -\left(\frac{K}{1+K}\right)\frac{1}{K_H} U_{SS}$$

which does not go to zero for large K, whereas the error due to U_P did go to zero, but rather approaches $-U_{SS}/K_H$.

In summary, the basic benefits of feedback discovered in this example, but typical in general of feedback systems with high loop gain, are:

1. Cause the controlled variable to accurately follow the desired variable.
2. Greatly reduce the effect on the controlled variable of all external disturbances in the forward path. It is ineffective in reducing the effect of disturbances in the feedback path (e.g., those associated with the sensor), and disturbances outside the loop (e.g., those associated with the reference input element).
3. Are tolerant of variations (due to wear, aging, environmental effects, etc.) in hardware parameters of components in the forward path, but not those in the feedback path (e.g., sensor) or outside the loop (e.g., reference input element).
4. Can give a closed-loop response speed much greater than that of the components from which they are constructed.

⇒ **Accuracy-Stability Tradeoff in Feedback Systems**

All feedback systems can become unstable if improperly designed. Let's discuss qualitatively this instability phenomenon. In any real-world component there is some kind of lagging behavior between input and output. Instantaneous response is impossible in the real world as it requires a system to go from one energy level to another in zero time, implying power supplies of infinite power. Lagging behavior is characterized quantitatively by the time constants (τ's) of first-order components and natural frequencies (ω_n's) of second-order components. Fast response (i.e., small τ's and large ω_n's) of components is desirable in feedback control systems as it results in greater accuracy since large loop gain is allowed. In essence, *instability results from an improper balance between the strength of the corrective action (loop gain) and the system dynamic lags*.

Why does the combination of excessive loop gain with excessive lags always result in feedback system instability? Consider a feedback control system, with a fixed desired value and in equilibrium, with the controlled variable at the desired value. If a process disturbance occurs, the ensuing deviation of the controlled variable will cause a correction to be applied, but it will be delayed by the cumulative lags of the sensor, controller, actuator, and process. Eventually, however, the trend of the controlled variable caused by the disturbance will be reversed by the opposition of the process manipulated input, returning the controlled variable toward the desired value. Now, if the loop gain is high, a strong correction is applied and the controlled variable overshoots the desired value, causing a reversal in the algebraic sign of the system error (i.e., difference between the desired value and the controlled variable). Unfortunately, because of system lags, a reversal of correction does not occur immediately and the process manipulated input (acting on old information) is now actually driving the controlled variable in the same direction as it is already going, rather than opposing its excursions, leading to a larger deviation. Eventually the reversed error does cause a reversed correction but by then the controlled variable has also reversed and the correction is again in the wrong direction. The controlled variable is thus driven alternately in opposite directions and does not settle to an equilibrium condition. This oscillatory state is called instability and, except for certain classes of systems, is unacceptable as control system behavior.

Consider the following example, shown in Figure 8. The liquid level C in a tank of constant cross-sectional area A is manipulated by controlling the volume flowrate M of an incompressible fluid in and out of the tank by means of a three-position on/off controller. The transfer function K/D between M and C represents the conservation of volume relation for the system, i.e., A(dC/dt) = M, which with K ≡ 1/A and D ≡ d/dt, reduces to DC = KM, or $\frac{C}{M}(D) = \frac{K}{D}$. The pump that manipulates M is shut off (M = 0) if the error E between the desired tank level R and measured level B is less in absolute value than the error dead space $E_{DS}/2$. When the error exceeds these limits, the pump adds or removes liquid at a rate M_0. The liquid-level sensor is assumed to measure C perfectly but there is a data transmission delay of τ_{DT} seconds in sending this information to the controller. That is, signal B is identical in form to C but is delayed by τ_{DT} seconds, a behavior given the name dead time. Since instability can be triggered by either or both command and disturbance inputs, in this example we apply a step command input R_S and examine system response. Note that the loop gain (strength of the corrective action) in this system depends on both M_0 and K.

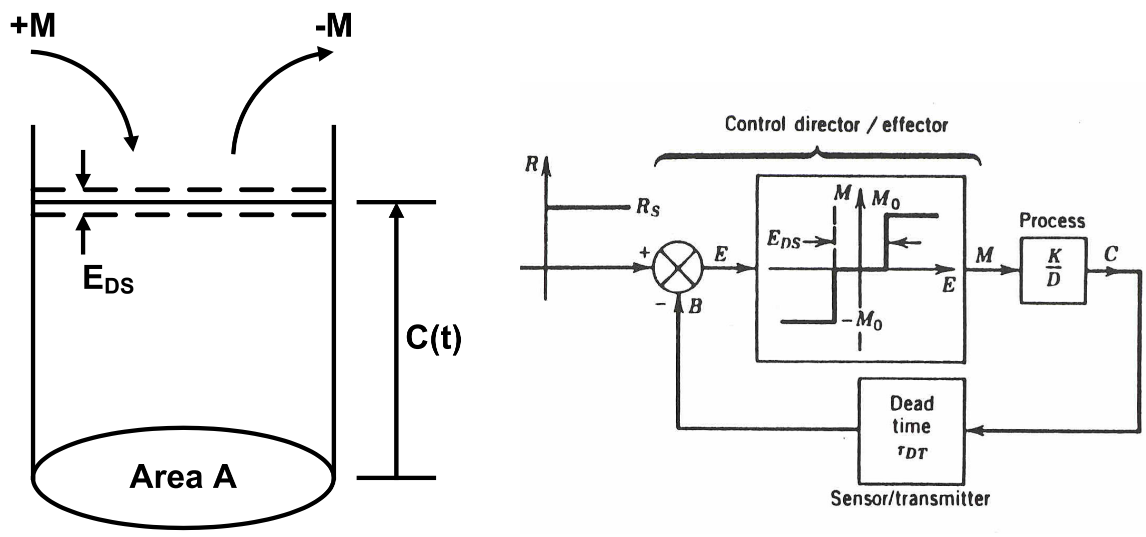

Figure 8. Tank Liquid-Level Feedback Control System

Shown in Figure 9 is a MatLab/Simulink block diagram for the tank liquid-level feedback control system.

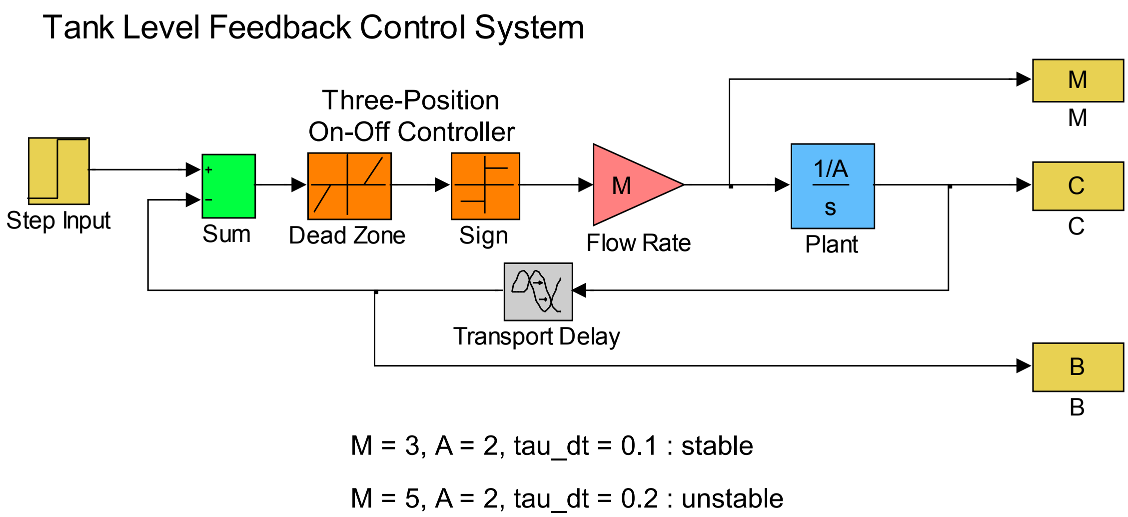

Figure 9. MatLab Simulink Block Diagram:
Tank Liquid-Level Feedback Control System

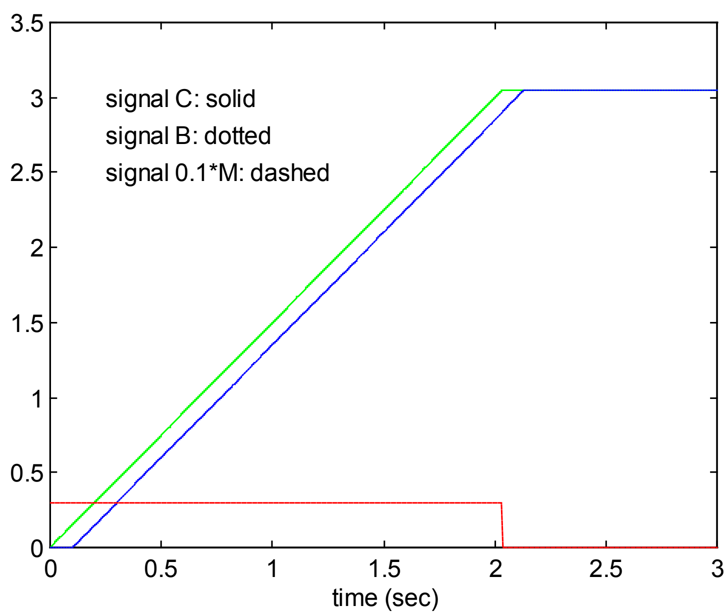

Figure 10. MatLab Simulation Results:
Stable Behavior of the Tank Liquid-Level Feedback Control System

In Figure 10, modest values of M_0 and K give a relatively slow, but stable, response of C. The simulation results are shown for a step input of 3 inches at time t = 0 seconds, a dead zone of ± 0.1 inches, a tank cross-sectional area A = 2.0 in^2 (K = 0.5 in^{-2}), a pump volume flowrate of M_0 = ± 3 in^3/sec, and a τ_{DT} = 0.1 seconds.

If specifications require a faster response, M_0 and/or K may be increased (or τ_{DT} may be decreased), but in Figure 11 the designer has gone too far with this, causing instability. This figure clearly shows how, because of time lags, correction M acts in a direction to increase, rather than reduce, the excursions of C during large parts of the cycle, a general condition for instability discussed earlier. In this case there is clearly an imbalance between the strength of the corrective action (M_0 and K) and the system dynamic lag (τ_{DT}) resulting in an unstable response. If we now reduce the strength of the corrective action and/or reduce the system dynamic lag so that there is a balance between the two, stable behavior will result. Shown in Figure 11 is the system response with M_0 = ± 5 in^3/sec and τ_{DT} = 0.2 seconds.

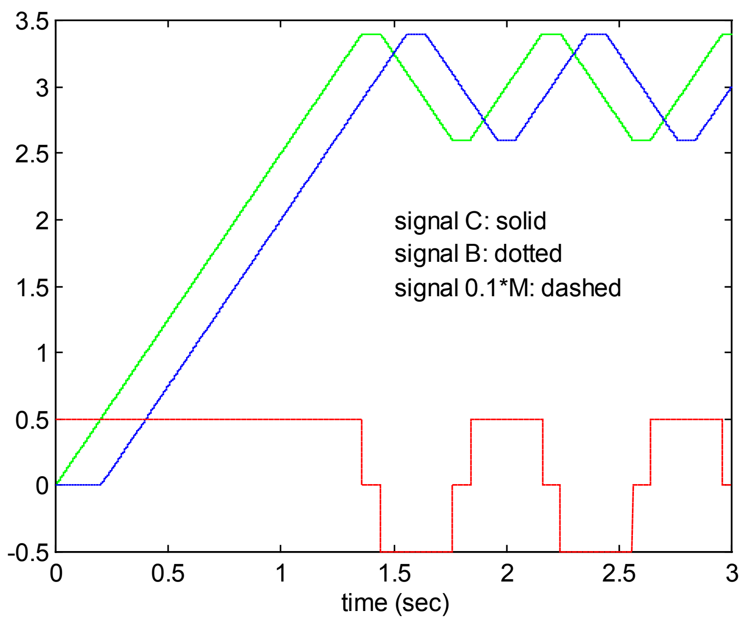

Figure 11. MatLab Simulation Results:
Unstable Behavior of the Tank Liquid-Level Feedback Control System