CHAPTER 28
BASIC CONTROL SYSTEMS DESIGN

William J. Palm III
Mechanical Engineering Department
University of Rhode Island
Kingston, Rhode Island

Revised from William J. Palm III, *Modeling, Analysis and Control of Dynamic Systems,* Wiley, 1983, by permission of the publisher.

Mechanical Engineers' Handbook, 2nd ed., Edited by Myer Kutz.
ISBN 0-471-13007-9

28.1 INTRODUCTION

The purpose of a *control system* is to produce a desired *output.* This output is usually specified by the command *input,* and is often a function of time. For simple applications in well-structured situations, *sequencing* devices like timers can be used as the control system. But most systems are not that easy to control, and the controller must have the capability of reacting to disturbances, changes in its environment, and new input commands. The key element that allows a control system to do this is *feedback,* which is the process by which a system's output is used to influence its behavior. Feedback in the form of the room-temperature measurement is used to control the furnace in a thermostatically controlled heating system. Figure 28.1 shows the *feedback loop* in the system's *block diagram,* which is a graphical representation of the system's control structure and logic. Another commonly found control system is the pressure regulator shown in Fig. 28.2.

Feedback has several useful properties. A system whose individual elements are nonlinear can often be modeled as a linear one over a wider range of its variables with the proper use of feedback. This is because feedback tends to keep the system near its reference operation condition. Systems that can maintain the output near its desired value despite changes in the environment are said to have good *disturbance rejection.* Often we do not have accurate values for some system parameter, or these values might change with age. Feedback can be used to minimize the effects of parameter changes and uncertainties. A system that has both good disturbance rejection and low sensitivity to parameter variation is *robust.* The application that resulted in the general understanding of the properties of feedback is shown in Fig. 28.3. The electronic amplifier gain A is large, but we are uncertain of its exact value. We use the resistors R_1 and R_2 to create a feedback loop around the amplifier, and pick R_1 and R_2 to create a feedback loop around the amplifier, and pick R_1 and R_2 so that $AR_2/R_1 >> 1$. Then the input–output relation becomes $e_o \approx R_1 e_i / R_2$, which is independent of A as long as A remains large. If R_1 and R_2 are known accurately, then the system gain is now reliable.

Figure 28.4 shows the block diagram of a *closed-loop* system, which is a system with feedback. An *open-loop* system, such as a timer, has no feedback. Figure 28.4 serves as a focus for outlining the prerequisites for this chapter. The reader should be familiar with the *transfer-function* concept based on the Laplace transform, the *pulse-transfer* function based on the z-transform, for digital control, and the differential equation modeling techniques needed to obtain them. It is also necessary to understand block-diagram algebra, characteristic roots, the final-value theorem, and their use in evaluating system response for common inputs like the step function. Also required are stability analysis techniques such as the Routh criterion, and transient performance specifications, such as the damping ratio ζ, natural frequency ω_n, dominant time constant τ, maximum overshoot, settling time, and bandwidth. The above material is reviewed in the previous chapter. Treatment in depth is given in Refs. 1, 2, and 3.

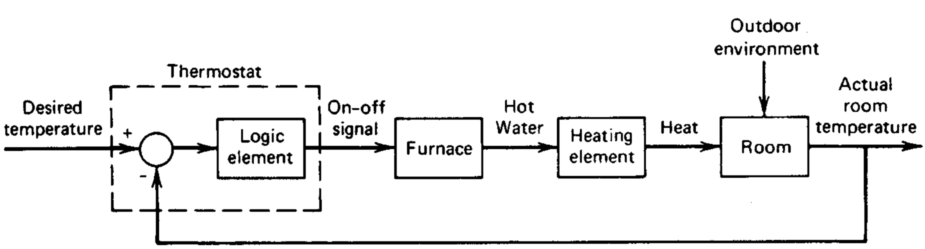


Fig. 28.1 Block diagram of the thermostat system for temperature control.[1]

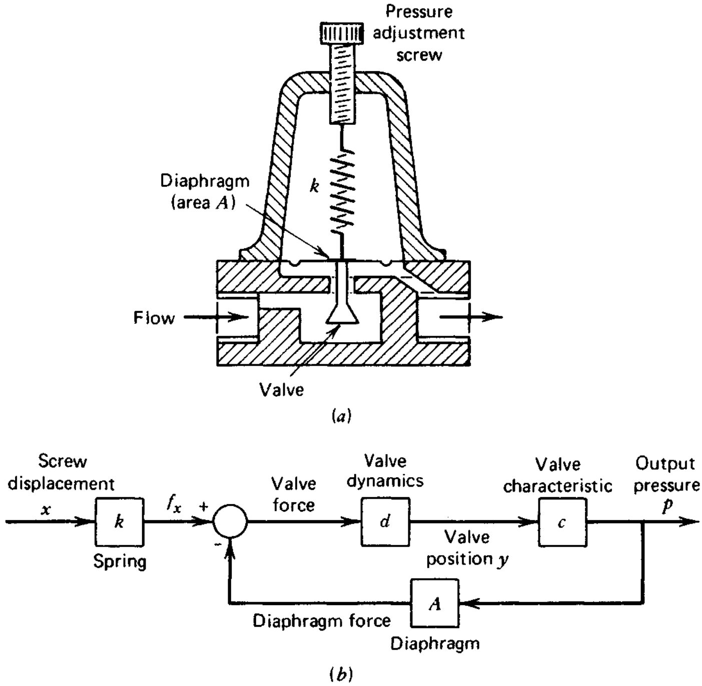


Fig. 28.2 Pressure regulator: (*a*) cutaway view; (*b*) block diagram.[1]

28.2 CONTROL SYSTEM STRUCTURE

The electromechanical position control system shown in Fig. 28.5 illustrates the structure of a typical control system. A load with an inertia I is to be positioned at some desired angle θ_r. A dc motor is provided for this purpose. The system contains viscous damping, and a disturbance torque T_d acts on the load, in addition to the motor torque T. Because of the disturbance, the angular position θ of the load will not necessarily equal the desired value θ_r. For this reason, a potentiometer, or some other sensor such as an encoder, is used to measure the displacement θ. The potentiometer voltage representing the controlled position θ is compared to the voltage generated by the command potentiometer. This device enables the operator to dial in the desired angle θ_r. The amplifier sees the difference e between the two potentiometer voltages. The basic function of the amplifier is to increase the small error voltage e up to the voltage level required by the motor and to supply enough current required by the motor to drive the load. In addition, the amplifier may shape the voltage signal in certain ways to improve the performance of the system.

The control system is seen to provide two basic functions: (1) to respond to a command input that specifies a new desired value for the controlled variable, and (2) to keep the controlled variable near the desired value in spite of disturbances. The presence of the feedback loop is vital to both

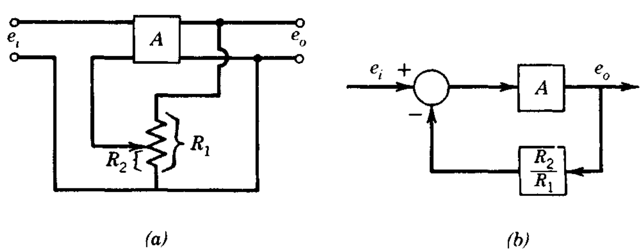


Fig. 28.3 A closed-loop system.

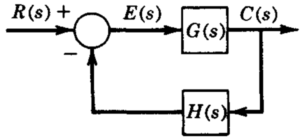


Fig. 28.4 Feedback compensation of an amplifier.

functions. A block diagram of this system is shown in Fig. 28.6. The power supplies required for the potentiometers and the amplifier are not shown in block diagrams of control system logic because they do not contribute to the control logic.

28.2.1 A Standard Diagram

The electromechanical positioning system fits the general structure of a control system (Fig. 28.7). This figure also gives some standard terminology. Not all systems can be forced into this format, but it serves as a reference for discussion.

The controller is generally thought of as a logic element that compares the command with the measurement of the output, and decides what should be done. The input and feedback elements are transducers for converting one type of signal into another type. This allows the error detector directly to compare two signals of the same type (e.g., two voltages). Not all functions show up as separate physical elements. The error detector in Fig. 28.5 is simply the input terminals of the amplifier.

The control logic elements produce the control signal, which is sent to the *final control elements.* These are the devices that develop enough torque, pressure, heat, and so on to influence the elements under control. Thus, the final control elements are the "muscle" of the system, while the control logic elements are the "brain." Here we are primarily concerned with the design of the logic to be used by this brain.

The object to be controlled is the *plant.* The *manipulated* variable is generated by the final control elements for this purpose. The disturbance input also acts on the plant. This is an input over which the designer has no influence, and perhaps for which little information is available as to the magnitude, functional form, or time of occurrence. The disturbance can be a random input, such as wind gust on a radar antenna, or deterministic, such as Coulomb friction effects. In the latter case, we can include the friction force in the system model by using a nominal value for the coefficient of friction. The disturbance input would then be the deviation of the friction force from this estimated value and would represent the uncertainty in our estimate.

Several control system classifications can be made with reference to Fig. 28.7. A *regulator* is a control system in which the controlled variable is to be kept constant in spite of disturbances. The command input for a regulator is its *set point.* A *follow-up system* is supposed to keep the control variable near a command value that is changing with time. An example of a follow-up system is a machine tool in which a cutting head must trace a specific path in order to shape the product properly. This is also an example of a *servomechanism,* which is a control system whose controlled variable is a mechanical position, velocity, or acceleration. A thermostat system is not a servomechanism, but a *process-control system,* where the controlled variable describes a thermodynamic process. Typically, such variables are temperature, pressure, flow rate, liquid level, chemical concentration, and so on.

28.2.2 Transfer Functions

A transfer function is defined for each input–output pair of the system. A specific transfer function is found by setting all other inputs to zero and reducing the block diagram. The *primary* or *command* transfer function for Fig. 28.7 is

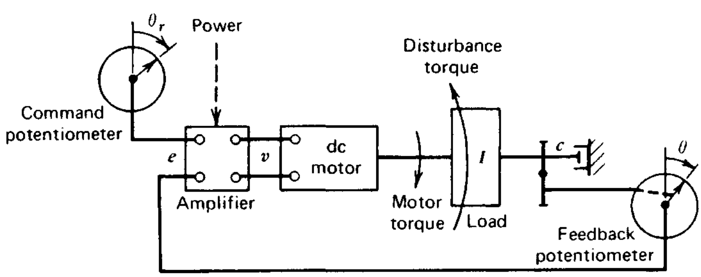


Fig. 28.5 Position-control system using a dc motor.[1]

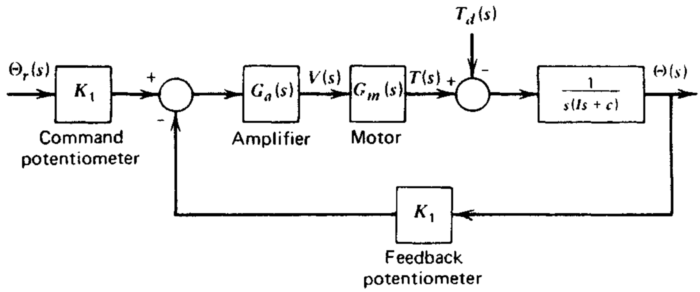


Fig. 28.6 Block diagram of the position-control system shown in Fig. 28.5.[1]

$$\frac{C(s)}{V(s)} = \frac{A(s)G_a(s)G_m(s)G_p(s)}{1 + G_a(s)G_m(s)G_p(s)H(s)} \tag{28.1}$$

The *disturbance* transfer function is

$$\frac{C(s)}{D(s)} = \frac{-Q(s)G_p(s)}{1 + G_a(s)G_m(s)G_p(s)H(s)} \tag{28.2}$$

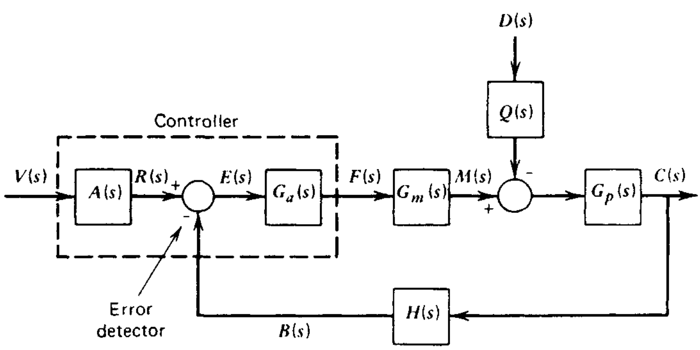
The transfer functions of a given system all have the same denominator.

28.2.3 System-Type Number and Error Coefficients

The error signal in Fig. 28.4 is related to the input as

$$E(s) = \frac{1}{1 + G(s)H(s)} R(s) \tag{28.3}$$

If the final value theorem can be applied, the steady-state error is



Elements

- $A(s)$ Input elements
- $G_a(s)$ Control logic elements
- $G_m(s)$ Final control elements
- $G_p(s)$ Plant elements
- $H(s)$ Feedback elements
- $Q(s)$ Disturbance elements

Signals

- $B(s)$ Feedback signal
- $C(s)$ Controlled variable or output
- $D(s)$ Disturbance input
- $E(s)$ Error or actuating signal
- $F(s)$ Control signal
- $M(s)$ Manipulated variable
- $R(s)$ Reference input
- $V(s)$ Command input

Fig. 28.7 Terminology and basic structure of a feedback-control system.[1]

$$e_{ss} = \lim_{s \to 0} \frac{sR(s)}{1 + G(s)H(s)} \tag{28.4}$$

The *static error coefficient* c_i is defined as

$$c_i = \lim_{s \to 0} s^i G(s)H(s) \tag{28.5}$$

A system is of *type n* if $G(s)H(s)$ can be written as $s^n F(s)$. Table 28.1 relates the steady-state error to the system type for three common inputs, and can be used to design systems for minimum error. The higher the system type, the better the system is able to follow a rapidly changing input. But higher-type systems are more difficult to stabilize, so a compromise must be made in the design. The coefficients c_0, c_1, and c_2 are called the *position, velocity,* and *acceleration error coefficients.*

28.3 TRANSDUCERS AND ERROR DETECTORS

The control system structure shown in Fig. 28.7 indicates a need for physical devices to perform several types of functions. Here we present a brief overview of some available transducers and error detectors. Actuators and devices used to implement the control logic are discussed in Sections 28.4 and 28.5.

28.3.1 Displacement and Velocity Transducers

A *transducer* is a device that converts one type of signal into another type. An example is the potentiometer, which converts displacement into voltage, as in Fig. 28.8. In addition to this conversion, the transducer can be used to make measurements. In such applications, the term *sensor* is more appropriate. Displacement can also be measured electrically with a *linear variable differential transformer* (LVDT) or a *synchro.* An LVDT measures the linear displacement of a movable magnetic core through a primary winding and two secondary windings (Fig. 28.9). An ac voltage is applied to the primary. The secondaries are connected together and also to a detector that measures the voltage and phase difference. A phase difference of 0° corresponds to a positive core displacement, while 180° indicates a negative displacement. The amount of displacement is indicated by the amplitude of the ac voltage in the secondary. The detector converts this information into a dc voltage e_o, such that $e_o = Kx$. The LVDT is sensitive to small displacements. Two of them can be wired together to form an error detector.

A synchro is a rotary differential transformer, with angular displacement as either the input or output. They are often used in paris (a *transmitter* and a *receiver*) where a remote indication of angular displacement is needed. When a transmitter is used with a synchro *control transformer,* two angular displacements can be measured and compared (Fig. 28.10). The output voltage e_0 is approximately linear with angular difference within $\pm 70°$, so that $e_0 = K(\theta_1 - \theta_2)$.

Displacement measurements can be used to obtain forces and accelerations. For example, the displacement of a calibrated spring indicates the applied force. The accelerometer is another example. Still another is the strain gage used for force measurement. It is based on the fact that the resistance of a fine wire changes as it is stretched. The change in resistance is detected by a circuit that can be calibrated to indicate the applied force. Sensors utilizing piezoelectric elements are also available.

Velocity measurements in control systems are most commonly obtained with a *tachometer.* This is essentially a dc generator (the reverse of a dc motor). The input is mechanical (a velocity). The output is a generated voltage proportional to the velocity. Translational velocity can be measured by converting it to angular velocity with gears, for example. Tachometers using ac signals are also available.

Table 28.1 Steady-State Error e_{ss} for Different System-Type Numbers

	System Type Number *n*			
R(s)	0	1	2	3
Step $1/s$	$\frac{1}{1 + C_0}$	0	0	0
Ramp $1/s^2$	∞	$\frac{1}{C_1}$	0	0
Parabola $1/s^3$	∞	∞	$\frac{1}{C_2}$	0

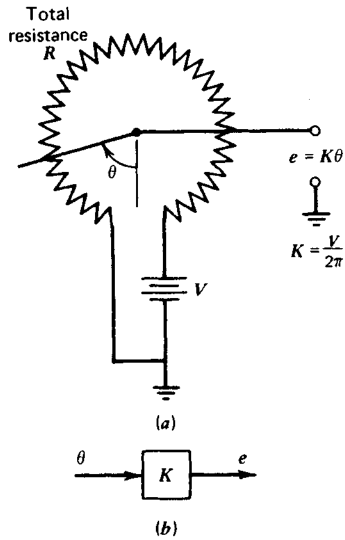


Fig. 28.8 Rotary potentiometer.[1]

Other velocity transducers include a magnetic pickup that generates a pulse every time a gear tooth passes. If the number of gear teeth is known, a pulse counter and timer can be used to compute the angular velocity. This principle is also employed in turbine flowmeters.

A similar principle is employed by *optical encoders,* which are especially suitable for digital control purposes. These devices use a rotating disk with alternating transparent and opaque elements whose passage is sensed by light beams and a photo-sensor array, which generates a binary (on–off) train of pulses. There are two basic types: the absolute encoder and the incremental encoder. By counting the number of pulses in a given time interval, the incremental encoder can measure the rotational speed of the disk. By using multiple tracks of elements, the absolute encoder can produce a binary digit that indicates the amount of rotation. Hence, it can be used as a position sensor.

Most encoders generate a train of TTL voltage level pulses for each channel. The incremental encoder output contains two channels that each produce N pulses every revolution. The encoder is mechanically constructed so that pulses from one channel are shifted relative to the other channel by a quarter of a pulse width. Thus, each pulse pair can be divided into four segments called *quadratures.* The encoder output consists of *4N quadrature counts per revolution.* The pulse shift also allows the

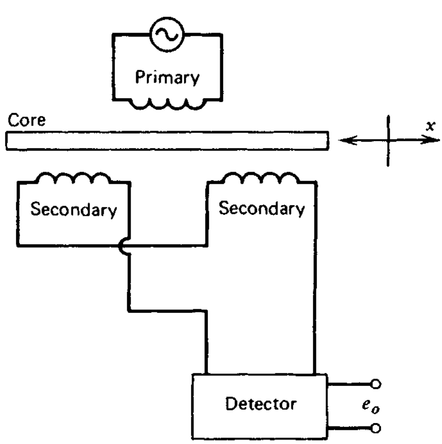


Fig. 28.9 Linear variable differential transformer (LVDT).[1]

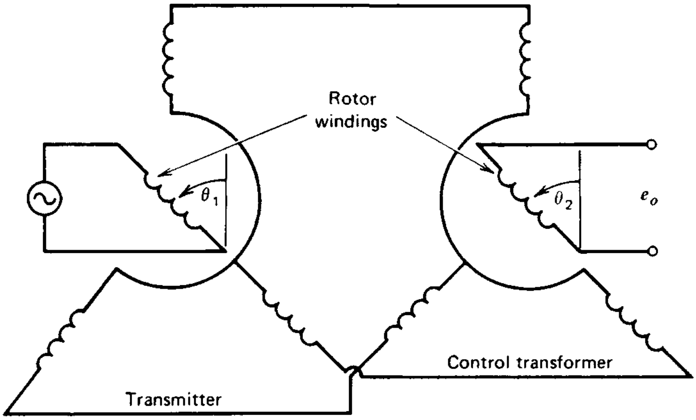


Fig. 28.10 Synchro transmitter-control transformer.[1]

direction of rotation to be determined by detecting which channel leads the other. The encoder might contain a third channel, known as the zero, index, or marker channel, that produces a pulse once per revolution. This is used for initialization.

The gain of such an incremental encoder is $4N/2\pi$. Thus, an encoder with 1000 pulses per channel per revolution has a gain of 636 counts per radian. If an absolute encoder produces a binary signal with n bits, the maximum number of positions it can represent is $2n$, and its gain is $2^n/2\pi$. Thus, a 16-bit absolute encoder has a gain of $2^{16}/2\pi = 10{,}435$ counts per radian.

28.3.2 Temperature Transducers

When two wires of dissimilar metals are joined together, a voltage is generated if the junctions are at different temperatures. If the reference junction is kept at a fixed, known temperature, the thermocouple can be calibrated to indicate the temperature at the other junction in terms of the voltage v. Electrical resistance changes with temperature. Platinum gives a linear relation between resistance and temperature, while nickel is less expensive and gives a large resistance change for a given temperature change. Seminconductors designed with this property are called *thermistors.* Different metals expand at different rates when the temperature is increased. This fact is used in the bimetallic strip transducer found in most home thermostats. Two dissimilar metals are bonded together to form the strip. As the temperature rises, the strip curls, breaking contact and shutting off the furnace. The temperature gap can be adjusted by changing the distance between the contacts. The motion also moves a pointer on the temperature scale of the thermostat. Finally, the pressure of a fluid inside a bulb will change as its temperature changes. If the bulb fluid is air, the device is suitable for use in pneumatic temperature controllers.

28.3.3 Flow Transducers

A flow rate q can be measured by introducing a flow restriction, such as an orifice plate, and measuring the pressure drop Δp across the restriction. The relation is $\Delta p = Rq^2$, where R can be found from calibration of the device. The pressure drop can be sensed by converting it into the motion of a diaphragm. Figure 28.11 illustrates a related technique. The Venturi-type flowmeter measures the static pressures in the constricted and unconstricted flow regions. Bernoulli's principle relates the pressure difference to the flow rate. This pressure difference produces the diaphragm displacement. Other types of flowmeters are available, such as turbine meters.

28.3.4 Error Detectors

The error detector is simply a device for finding the difference between two signals. This function is sometimes an integral feature of sensors, such as with the synchro transmitter–transformer combination. This concept is used with the diaphragm element shown in Fig. 28.11. A detector for voltage difference can be obtained, as with the position-control system shown in Fig. 28.5. An amplifier intended for this purpose is a *differential amplifier.* Its output is proportional to the difference between the two inputs. In order to detect differences in other types of signals, such as temperature, they are usually converted to a displacement or pressure. One of the detectors mentioned previously can then be used.

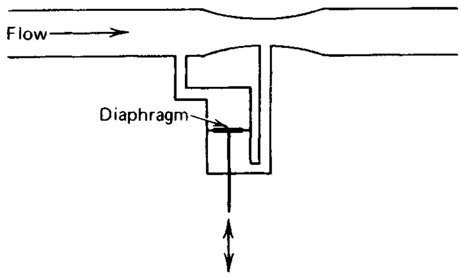


Fig. 28.11 Venturi-type flowmeter. The diaphragm displacement indicates the flow rate.[1]

28.3.5 Dynamic Response of Sensors

The usual transducer and detector models are static models, and as such imply that the components respond instantaneously to the variable being sensed. Of course, any real component has a dynamic response of some sort, and this response time must be considered in relation to the controlled process when a sensor is selected. If the controlled process has a time constant at least 10 times greater than that of the sensor, we often would be justified in using a static sensor model.

28.4 ACTUATORS

An *actuator* is the final control element that operates on the low-level control signal to produce a signal containing enough power to drive the plant for the intended purpose. The armature-controlled dc motor, the hydraulic servomotor, and the pneumatic diaphragm and piston are common examples of actuators.

28.4.1 Electromechanical Actuators

Figure 28.12 shows an electromechanical system consisting of an armature-controlled dc motor driving a load inertia. The rotating armature consists of a wire conductor wrapped around an iron core.

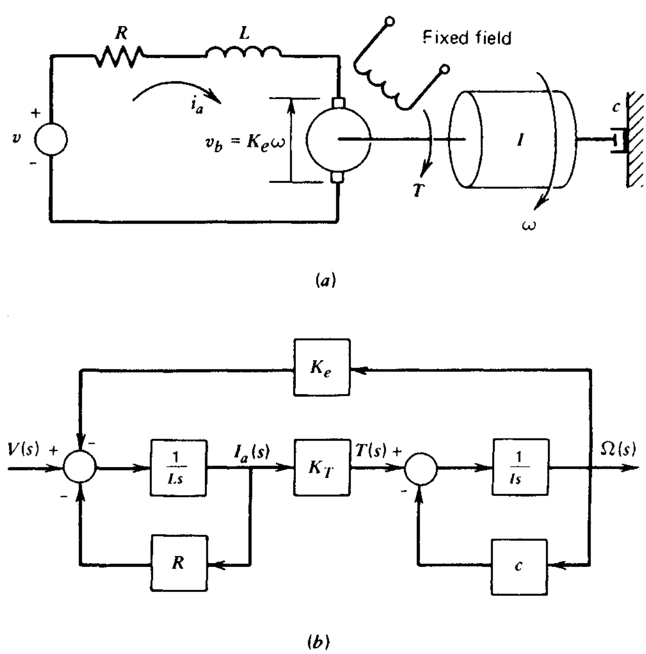


Fig. 28.12 Armature-controlled dc motor with a load, and the system's block diagram.[1]

This winding has an inductance L. The resistance R represents the lumped value of the armature resistance and any external resistance deliberately introduced to change the motor's behavior. The armature is surrounded by a magnetic field. The reaction of this field with the armature current produces a torque that causes the armature to rotate. If the armature voltage v is used to control the motor, the motor is said to be *armature-controlled*. In this case, the field is produced by an electromagnet supplied with a constant voltage or by a permanent magnet. This motor type produces a torque T that is proportional to the armature current i_a:

$$T = K_T i_a \tag{28.6}$$

The torque constant K_T depends on the strength of the field and other details of the motor's construction. The motion of a current-carrying conductor in a field produces a voltage in the conductor that opposes the current. This voltage is called the *back emf* (electromotive force). Its magnitude is proportional to the speed and is given by

$$e_b = K_e \omega \tag{28.7}$$

The transfer function for the armature-controlled dc motor is

$$\frac{\Omega(s)}{V(s)} = \frac{K_T}{LIs^2 + (RI + cL)s + cR + K_e K_T} \tag{28.8}$$

Another motor configuration is the *field-controlled* dc motor. In this case, the armature current is kept constant and the field voltage v is used to control the motor. The transfer function is

$$\frac{\Omega(s)}{V(s)} = \frac{K_T}{(Ls + R)(Is + c)} \tag{28.9}$$

where R and L are the resistance and inductance of the field circuit, and K_T is the torque constant. No back emf exists in this motor to act as a self-braking mechanism.

Two-phase ac motors can be used to provide a low-power, variable-speed actuator. This motor type can accept the ac signals directly from LVDTs and synchros without demodulation. However, it is difficult to design ac amplifier circuitry to do other than proportional action. For this reason, the ac motor is not found in control systems as often as dc motors. The transfer function for this type is of the form of Eq. (28.9).

An actuator especially suitable for digital systems is the *stepper motor*, a special dc motor that takes a train of electrical input pulses and converts each pulse into an angular displacement of a fixed amount. Motors are available with resolutions ranging from about 4 steps per revolution to more than 800 steps per revolution. For 36 steps per revolution, the motor will rotate by 10° for each pulse received. When not being pulsed, the motors lock in place. Thus, they are excellent for precise positioning applications, such as required with printers and computer tape drives. A disadvantage is that they are low-torque devices. If the input pulse frequency is not near the resonant frequency of the motor, we can take the output rotation to be directly related to the number of input pulses and use that description as the motor model.

28.4.2 Hydraulic Actuators

Machine tools are one application of the hydraulic system shown in Fig. 28.13. The applied force f is supplied by the servomotor. The mass m represents that of a cutting tool and the power piston, while k represents the combined effects of the elasticity naturally present in the structure and that introduced by the designer to achieve proper performance. A similar statement applies to the damping c. The valve displacement z is generated by another control system in order to move the tool through its prescribed motion. The spool valve shown in Fig. 28.13 had two *lands*. If the width of the land is greater than the port width, the valve is said to be *overlapped*. In this case, a dead zone exists in which a slight change in the displacement z produces no power piston motion. Such dead zones create control difficulties and are avoided by designing the valve to be *underlapped* (the land width is less the port width). For such valves there will be a small flow opening even when the valve is in the neutral position at $z = 0$. This gives it a higher sensitivity than an overlapped valve.

The variables z and $\Delta p = p_2 - p_1$ determine the volume flow rate, as

$$q = f(z, \Delta p)$$

For the reference equilibrium condition ($z = 0$, $\Delta p = 0$, $q = 0$), a linearization gives

$$q = C_1 z - C_2 \Delta p \tag{28.10}$$

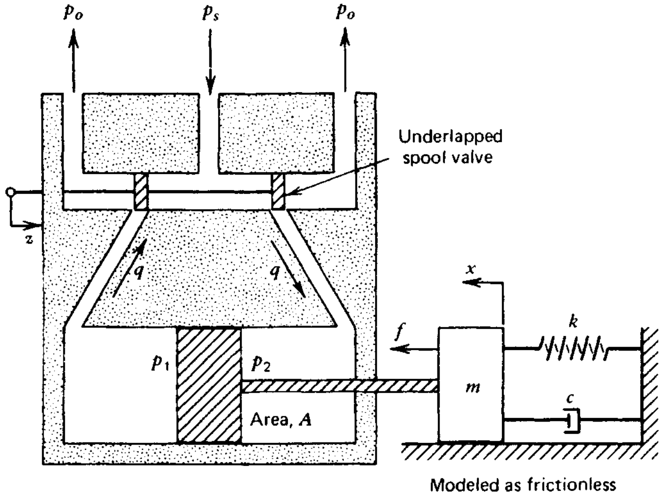


Fig. 28.13 Hydraulic servomotor with a load.[1]

The linearization constants are available from theoretical and experimental results.[4] The transfer function for the system is[1,2]

$$T(s) = \frac{X(s)}{Z(s)} = \frac{C_1}{\dfrac{C_2 m}{A} s^2 + \left(\dfrac{cC_2}{A} + A\right) s + \dfrac{C_2 k}{A}} \tag{28.11}$$

The development of the steam engine led to the requirement for a speed-control device to maintain constant speed in the presence of changes in load torque or steam pressure. In 1788, James Watt of Glasgow developed his now-famous flyball governor for this purpose (Fig. 28.14). Watt took the principle of sensing speed with the centrifugal pendulum of Thomas Mead and used it in a feedback loop on a steam engine. As the motor speed increases, the flyballs move outward and pull the slider

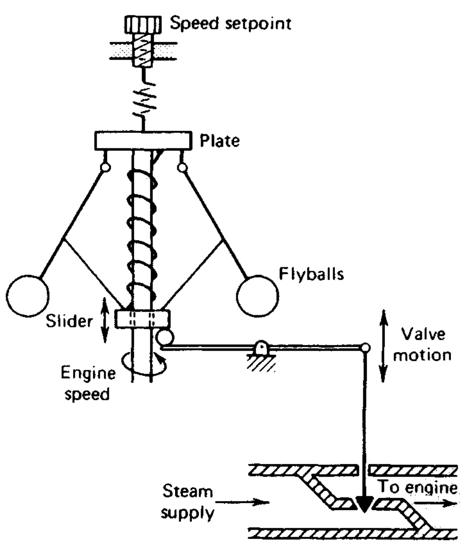


Fig. 28.14 James Watt's flyball governor for speed control of a steam engine.[1]

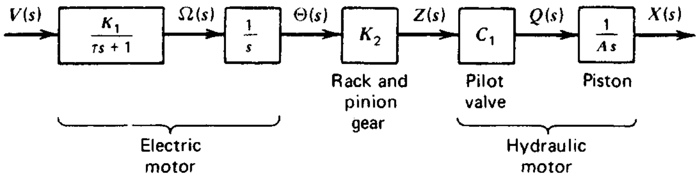


Fig. 28.15 Electrohydraulic system for translation.[1]

upward. The upward motion of the slider closes the steam valve, thus causing the engine to slow down. If the engine speed is too slow, the spring force overcomes that due to the flyballs, and the slider moves down to open the steam valve. The desired speed can be set by moving the plate to change the compression in the spring. The principle of the flyball governor is still used for speed-control applications. Typically, the pilot valve of a hydraulic servomotor is connected to the slider to provide the high forces required to move large supply valves.

Many hydraulic servomotors use multistage valves to obtain finer control and higher forces. A *two-stage valve* has a *slave value,* similar to the pilot valve, but situated between the pilot valve and the power piston.

Rotational motion can be obtained with a *hydraulic motor,* which is, in principle, a pump acting in reverse (fluid input and mechanical rotation output). Such motors can achieve higher torque levels than electric motors. A hydraulic pump driving a hydraulic motor constitutes a *hydraulic transmission.*

A popular actuator choice is the *electrohydraulic* system, which uses an electric actuator to control a hydraulic servomotor or transmission by moving the pilot valve or the swash-plate angle of the pump. Such systems combine the power of hydraulics with the advantages of electrical systems. Figure 28.15 shows a hydraulic motor whose pilot valve motion is caused by an armature-controlled dc motor. The transfer function between the motor voltage and the piston displacement is

$$\frac{X(s)}{V(s)} = \frac{K_1 K_2 C_1}{As^2(\tau s + 1)} \tag{28.12}$$

If the rotational inertia of the electric motor is small, then $\tau \approx 0$.

28.4.3 Pneumatic Actuators

Pneumatic actuators are commonly used because they are simple to maintain and use a readily available working medium. Compressed air supplies with the pressures required are commonly available in factories and laboratories. No flammable fluids or electrical sparks are present, so these devices are considered the safest to use with chemical processes. Their power output is less than that of hydraulic systems, but greater than that of electric motors.

A device for converting pneumatic pressure into displacement is the bellows shown in Fig. 28.16. The transfer function for a linearized model of the bellows is of the form

$$\frac{X(s)}{P(s)} = \frac{K}{\tau s + 1} \tag{28.13}$$

where x and p are deviations of the bellows displacement and input pressure from nominal values.

In many control applications, a device is needed to convert small displacements into relatively large pressure changes. The nozzle-flapper serves this purpose (Fig. 28.17*a*). The input displacement y moves the flapper, with little effort required. This changes the opening at the nozzle orifice. For a

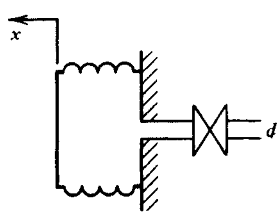


Fig. 28.16 Pneumatic bellows.[1]

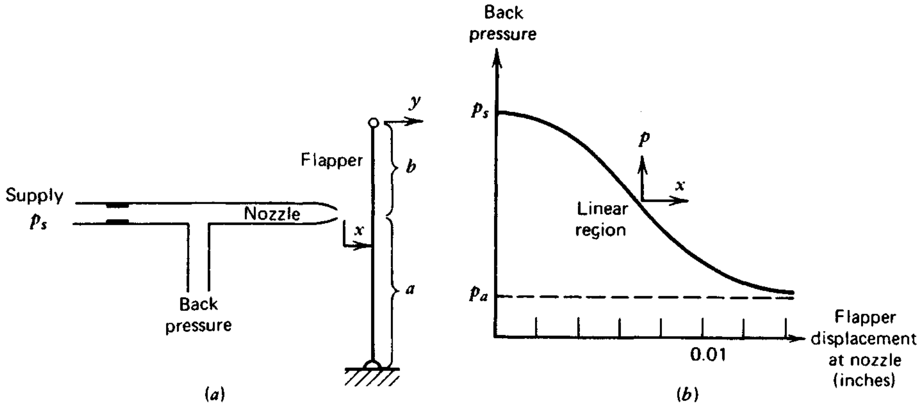


Fig. 28.17 Pneumatic nozzle-flapper amplifier and its characteristic curve.[1]

large enough opening, the nozzle back pressure is approximately the same as atmospheric pressure p_a. At the other extreme position with the flapper completely blocking the orifice, the back pressure equals the supply pressure p_s. This variation is shown in Fig. 28.17*b*. Typical supply pressures are between 30 and 100 psia. The orifice diameter is approximately 0.01 in. Flapper displacement is usually less than one orifice diameter.

The nozzle-flapper is operated in the linear portion of the back pressure curve. The linearized back pressure relation is

$$p = -K_f x \tag{28.14}$$

where $-K_f$ is the slope of the curve and is a very large number. From the geometry of similar triangles, we have

$$p = -\frac{aK_f}{a+b} y \tag{28.15}$$

In its operating region, the nozzle-flapper's back pressure is well below the supply pressure.

The output pressure from a pneumatic device can be used to drive a final control element like the pneumatic actuating valve shown in Fig. 28.18. The pneumatic pressure acts on the upper side of the diaphragm and is opposed by the return spring.

Formerly, many control systems utilized pneumatic devices to implement the control law in analog form. Although the overall, or higher-level, control algorithm is now usually implemented in digital form, pneumatic devices are still frequently used for final control corrections at the actuator level,

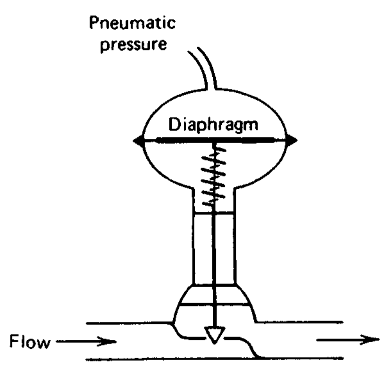


Fig. 28.18 Pneumatic flow-control valve.[1]

where the control action must eventually be supplied by a mechanical device. An example of this is the electro-pneumatic valve positioner used in Valtek valves, and illustrated in Fig. 28.19. The heart of the unit is a pilot valve capsule that moves up and down according to the pressure difference across its two supporting diaphragms. The capsule has a plunger at its top and at its bottom. Each plunger has an exhaust seat at one end and a supply seat at the other. When the capsule is in its equilibrium position, no air is supplied to or exhausted from the valve cylinder, so the valve does not move.

The process controller commands a change in the valve stem position by sending the 4–20 ma dc input signal to the positioner. Increasing this signal causes the electromagnetic actuator to rotate the lever counterclockwise about the pivot. This increases the air gap between the nozzle and flapper. This decreases the back pressure on top of the upper diaphragm and causes the capsule to move up. This motion lifts the upper plunger from its supply seat and allows the supply air to flow to the bottom of the valve cylinder. The lower plunger's exhaust seat is uncovered, thus decreasing the air pressure on top of the valve piston, and the valve stem moves upward. This motion causes the lever arm to rotate, increasing the tension in the feedback spring and decreasing the nozzle-flapper gap. The valve continues to move upward until the tension in the feedback spring counteracts the force produced by the electromagnetic actuator, thus returning the capsule to its equilibrium position.

A decrease in the dc input signal causes the opposite actions to occur, and the valve moves downward.

28.5 CONTROL LAWS

The control logic elements are designed to act on the error signal to produce the control signal. The algorithm that is used for this purpose is called the *control law*, the *control action*, or the *control algorithm*. A nonzero error signal results from either a change in command or a disturbance. The general function of the controller is to keep the controlled variable near its desired value when these occur. More specifically, the control objectives might be stated as follows:

1. Minimize the steady-state error.
2. Minimize the settling time.
3. Achieve other transient specifications, such as minimizing the overshoot.

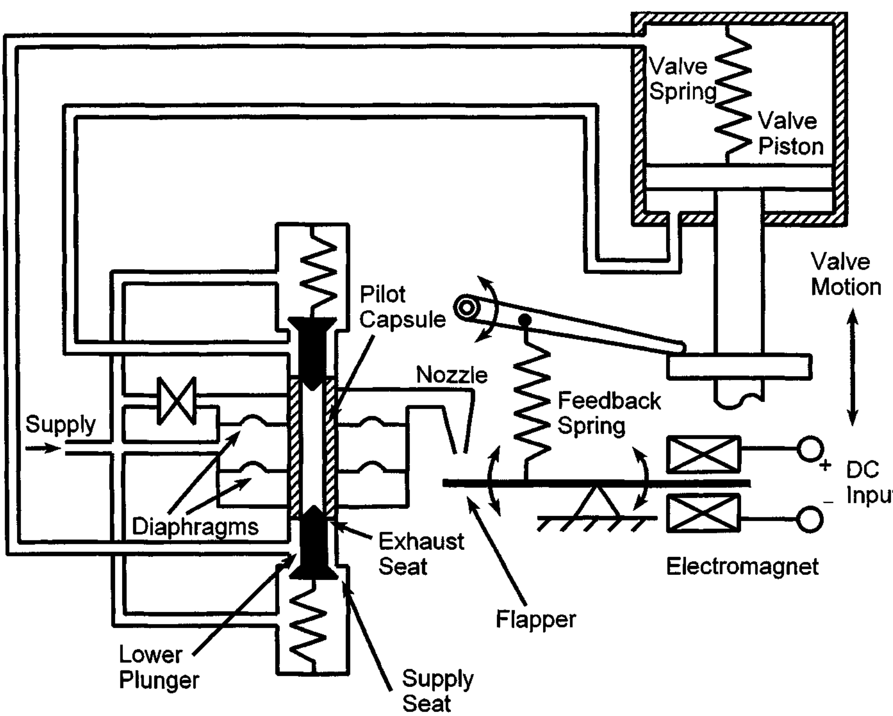


Fig. 28.19 An electro-pneumatic valve positioner.

In practice, the design specifications for a controller are more detailed. For example, the bandwidth might also be specified along with a safety margin for stability. We never know the numerical values of the system's parameters with true certainty, and some controller designs can be more sensitive to such parameter uncertainties than other designs. So a parameter sensitivity specification might also be included.

The following control laws form the basis of most control systems.

28.5.1 Proportional Control

Two-position control is the most familiar type, perhaps because of its use in home thermostats. The control output takes on one of two values. With the *on–off controller,* the controller output is either on or off (e.g., fully open or fully closed). Two-position control is acceptable for many applications in which the requirements are not too severe. However, many situations require finer control.

Consider a liquid-level system in which the input flowrate is controlled by a valve. We might try setting the control valve manually to achieve a flow rate that balances the system at the desired level. We might then added a controller that adjusts this setting in proportion to the deviation of the level from the desired value. This is *proportional control,* the algorithm in which the change in the control signal is proportional to the error. Block diagrams for controllers are often drawn in terms of the deviations from a zero-error equilibrium condition. Applying this convention to the general terminology of Fig. 28.6, we see that proportional control is described by

$$F(s) = K_P E(s)$$

where $F(s)$ is the deviation in the control signal and K_P is the *proportional gain.* If the total valve displacement is $y(t)$ and the manually created displacement is x, then

$$y(t) = K_P e(t) + x$$

The percent change in error needed to move the valve full scale is the *proportional band.* It is related to the gain as

$$K_P = \frac{100}{\text{band\%}}$$

The zero-error valve displacement x is the *manual reset.*

Proportional Control of a First-Order System

To investigate the behavior of proportional control, consider the speed-control system shown in Fig. 28.20; it is identical to the position controller shown in Fig. 28.6, except that a tachometer replaces the feedback potentiometer. We can combine the amplifier gains into one, denoted K_P. The system is thus seen to have proportional control. We assume the motor is field-controlled and has a negligible electrical time constant. The disturbance is a torque T_d, for example, resulting from friction. Choose the reference equilibrium condition to be $T_d = T = 0$ and $\omega_r = w = 0$. The block diagram is shown in Fig. 28.21. For a meaningful error signal to be generated, K_1 and K_2 should be chosen to be equal. With this simplification the diagram becomes that shown in Fig. 28.22, where $G(s) = K = K_1 K_P K_T/R$. A change in desired speed can be simulated by a unit step input for ω_r. For $\Omega_r(s) = 1/s$, the velocity approaches the steady-state value $\omega_{ss} = K/(c + K) < 1$. Thus, the final value is less than the desired value of 1, but it might be close enough if the damping c is small. The time required to

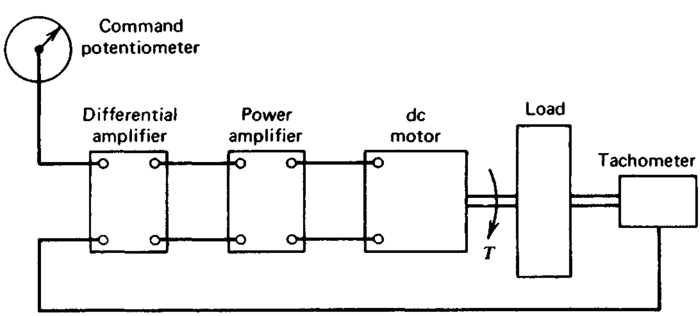


Fig. 28.20 Velocity-control system using a dc motor.[1]

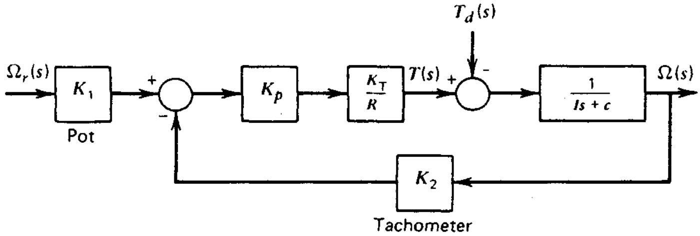


Fig. 28.21 Block diagram of the velocity-control system of Fig. 28.20.[1]

reach this value is approximately four time constants, or $4\tau = 4I/(c + K)$. A sudden change in load torque can also be modeled by a unit step function $T_d(s) = 1/s$. The steady-state response due solely to the disturbance is $-1/(c + K)$. If $(c + K)$ is large, this error will be small.

The performance of the proportional control law thus far can be summarized as follows. For a first-order plant with step function inputs:

1. The output never reaches its desired value if damping is present ($c \neq 0$), although it can be made arbitrarily close by choosing the gain K large enough. This is called *offset error.*
2. The output approaches its final value without oscillation. The time to reach this value is inversely proportional to K.
3. The output deviation due to the disturbance at steady state is inversely proportional to the gain K. This error is present even in the absence of damping ($c = 0$).

As the gain K is increased, the time constant becomes smaller and the response faster. Thus, the chief disadvantage of proportional control is that it results in steady-state errors and can only be used when the gain can be selected large enough to reduce the effect of the largest expected disturbance. Since proportional control gives zero error only for one load condition (the reference equilibrium), the operator must change the manual reset by hand (hence the name). An advantage to proportional control is that the control signal responds to the error instantaneously (in theory at least). It is used in applications requiring rapid action. Processes with time constants too small for the use of two-position control are likely candidates for proportional control. The results of this analysis can be applied to any type of first-order system (e.g., liquid-level, thermal, etc.) having the form in Fig. 28.22.

Proportional Control of a Second-Order System

Proportional control of a neutrally stable second-order plant is represented by the position controller of Fig. 28.6 if the amplifier transfer function is a constant $G_a(s) = K_a$. Let the motor transfer function be $G_m(s) = K_T/R$, as before. The modified block diagram is given in Fig. 28.23 with $G(s) = K = K_1 K_a K_T/R$. The closed-loop system is stable if I, c, and K are positive. For no damping ($c = 0$), the closed-loop system is neutrally stable. With no disturbance and a unit step command, $\Theta_r(s) = 1/s$, the steady-state output is $\omega_{ss} = 1$. The offset error is thus zero if the system is stable ($c > 0$, $K > 0$). The steady-state output deviation due to a unit step disturbance is $-1/K$. This deviation can be reduced by choosing K large. The transient behavior is indicated by the damping ratio, $\zeta = c/2\sqrt{IK}$.

For slight damping, the response to a step input will be very oscillatory and the overshoot large. The situation is aggravated if the gain K is made large to reduce the deviation due to the disturbance. We conclude, therefore, that proportional control of this type of second-order plant is not a good choice unless the damping constant c is large. We will see shortly how to improve the design.

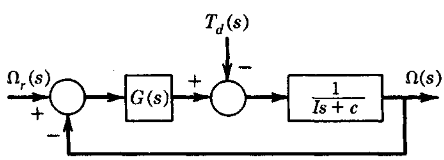


Fig. 28.22 Simplified form of Fig. 28.21 for the case $K_1 = K_2$.

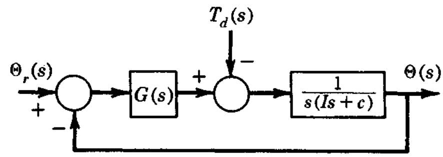


Fig. 28.23 Position servo.

28.5.2 Integral Control

The offset error that occurs with proportional control is a result of the system reaching an equilibrium in which the control signal no longer changes. This allows a constant error to exist. If the controller is modified to produce an increasing signal as long as the error is nonzero, the offset might be eliminated. This is the principle of *integral control.* In this mode the change in the control signal is proportional to the *integral* of the error. In the terminology of Fig. 28.7, this gives

$$F(s) = \frac{K_I}{s} E(s) \tag{28.16}$$

where $F(s)$ is the deviation in the control signal and K_I is the *integral gain.* In the time domain, the relation is

$$f(t) = K_I \int_0^t e(t)\, dt \tag{28.17}$$

if $f(0) = 0$. In this form, it can be seen that the integration cannot continue indefinitely because it would theoretically produce an infinite value of $f(t)$ if $e(t)$ does not change sign. This implies that special care must be taken to reinitialize a controller that uses integral action.

Integral Control of a First-Order System

Integral control of the velocity in the system of Fig. 28.20 has the block diagram shown in Fig. 28.22, where $G(s) = K/s$, $K = K_1 K_I K_T / R$. The integrating action of the amplifier is physically obtained by the techniques to be presented in Section 28.6, or by the digital methods presented in Section 28.10. The control system is stable if I, c, and K are positive. For a unit step command input, $\omega_{ss} = 1$; so the offset error is zero. For a unit step disturbance, the steady-state deviation is zero if the system is stable. Thus, the steady-state performance using integral control is excellent for this plant with step inputs. The damping ratio is $\zeta = c/2\sqrt{IK}$. For slight damping, the response will be oscillatory rather than exponential as with proportional control. Improved steady-state performance has thus been obtained at the expense of degraded transient performance. The conflict between steady-state and transient specifications is a common theme in control system design. As long as the system is underdamped, the time constant is $\tau = 2I/c$ and is not affected by the gain K, which only influences the oscillation frequency in this case. It might by physically possible to make K small enough so that $\zeta >> 1$, and the nonoscillatory feature of proportional control recovered, but the response would tend to be sluggish. Transient specifications for fast response generally require that $\zeta < 1$. The difficulty with using $\zeta < 1$ is that τ is fixed by c and I. If c and I are such that $\zeta < 1$, then τ is large if $I >> c$.

Integral Control of a Second-Order System

Proportional control of the position servomechanism in Fig. 28.23 gives a nonzero steady-state deviation due to the disturbance. Integral control [$G(s) = K/s$] applied to this system results in the command transfer function

$$\frac{\Theta(s)}{\Theta_r(s)} = \frac{K}{Is^3 + cs^2 + K} \tag{28.18}$$

With the Routh criterion, we immediately see that the system is not stable because of the missing s term. Integral control is useful in improving steady-state performance, but in general it does not improve and may even degrade transient performance. Improperly applied, it can produce an unstable control system. It is best used in conjunction with other control modes.

28.5.3 Proportional-Plus-Integral Control

Integral control raised the order of the system by one in the preceding examples, but did not give a characteristic equation with enough flexibility to achieve acceptable transient behavior. The instantaneous response of proportional control action might introduce enough variability into the coefficients of the characteristic equation to allow both steady-state and transient specifications to be satisfied. This is the basis for using *proportional-plus-integral control* (PI control). The algorithm for this two-mode control is

$$F(s) = K_P E(s) + \frac{K_I}{s} E(s) \tag{28.19}$$

The integral action provides an automatic, not manual, reset of the controller in the presence of a disturbance. For this reason, it is often called *reset action.*

The algorithm is sometimes expressed as

$$F(s) = K_P \left(1 + \frac{1}{T_I s}\right) E(s) \tag{28.20}$$

where T_I is the *reset time.* The reset time is the time required for the integral action signal to equal that of the proportional term, if a constant error exists (a hypothetical situation). The reciprocal of reset time is expressed as repeats per minute and is the frequency with which the integral action repeats the proportional correction signal.

The proportional control gain must be reduced when used with integral action. The integral term does not react instantaneously to a zero-error signal but continues to correct, which tends to cause oscillations if the designer does not take this effect into account.

PI Control of a First-Order System

PI action applied to the speed controller of Fig. 28.20 gives the diagram shown in Fig. 28.21 with $G(s) = K_P + K_I/s$. The gains K_P and K_I are related to the component gains, as before. The system is stable for positive values of K_P and K_I. For $\Omega_r(s) = 1/s$, $\omega_{ss} = 1$, and the offset error is zero, as with integral action only. Similarly, the deviation due to a unit step disturbance is zero at steady state. The damping ratio is $\zeta = (c + K_P)/2\sqrt{IK_I}$. The presence of K_P allows the damping ratio to be selected without fixing the value of the dominant time constant. For example, if the system is underdamped ($\zeta < 1$), the time constant is $\tau = 2I/(c + K_P)$. The gain K_P can be picked to obtain the desired time constant, while K_I used to set, the damping ratio. A similar flexibility exists if $\zeta = 1$. Complete description of the transient response requires that the numerator dynamics present in the transfer functions be accounted for.[1,2]

PI Control of a Second-Order System

Integral control for the position servomechanism of Fig. 28.23 resulted in a third-order system that is unstable. With proportional action, the diagram becomes that of Fig. 28.22, with $G(s) = K_P + K_I/s$. The steady-state performance is acceptable, as before, if the system is assumed to be stable. This is true if the Routh criterion is satisfied; that is, if I, c, K_P, and K_I are positive and $cK_P - IK_I > 0$. The difficulty here occurs when the damping is slight. For small c, the gain K_P must be large in order to satisfy the last condition, and this can be difficult to implement physically. Such a condition can also result in an unsatisfactory time constant. The root-locus method of Section 28.9 provides the tools for analyzing this design further.

28.5.4 Derivative Control

Integral action tends to produce a control signal even after the error has vanished, which suggests that the controller be made aware that the error is approaching zero. One way to accomplish this is to design the controller to react to the derivative of the error with *derivative control* action, which is

$$F(s) = K_D s E(s) \tag{28.21}$$

where K_D is the *derivative gain.* This algorithm is also called *rate action.* It is used to damp out oscillations. Since it depends only on the error rate, derivative control should never be used alone. When used with proportional action, the following PD-control algorithm results:

$$F(s) = (K_P + K_D s)E(s) = K_P(1 + T_D s)E(s) \tag{28.22}$$

where T_D is the *rate time* or *derivative time.* With integral action included, the *proportional-plus-integral-plus-derivative (PID) control law* is obtained.

$$F(s) = \left(K_P + \frac{K_I}{s} + K_D s\right) E(s) \tag{28.23}$$

This is called a *three-mode controller.*

PD Control of a Second-Order System

The presence of integral action reduces steady-state error, but tends to make the system less stable. There are applications of the position servomechanism in which a nonzero derivation resulting from the disturbance can be tolerated, but an improvement in transient response over the proportional control result is desired. Integral action would not be required, but rate action can be added to improve the transient response. Application of PD control to this system gives the block diagram of Fig. 28.23 with $G(s) = K_P + K_D s$.

The system is stable for positive values of K_D and K_P. The presence of rate action does not affect the steady-state response, and the steady-state results are identical to those with P control; namely, zero offset error and a deviation of $-1/K_P$, due to the disturbance. The damping ratio is $\zeta = (c + K_D)/2\sqrt{IK_P}$. For P control, $\zeta = c/2\sqrt{IK_P}$. Introduction of rate action allows the proportional gain K_P to be selected large to reduce the steady-state deviation, while K_D can be used to achieve an acceptable damping ratio. The rate action also helps to stabilize the system by adding damping (if $c = 0$ the system with P control is not stable).

The equivalent of derivative action can be obtained by using a tachometer to measure the angular velocity of the load. The block diagram is shown in Fig. 28.24. The gain of the amplifier–motor–potentiometer combination is K_1, and K_2 is the tachometer gain. The advantage of this system is that it does not require signal differentiation, which is difficult to implement if signal noise is present. The gains K_1 and K_2 can be chosen to yield the desired damping ratio and steady-state deviation, as was done with K_P and K_I.

28.5.5 PID Control

The position servomechanism design with PI control is not completely satisfactory because of the difficulties encountered when the damping c is small. This problem can be solved by the use of the full PID-control law, as shown in Fig. 28.23 with $G(s) = K_P + K_D s + K_I/s$.

A stable system results if all gains are positive and if $(c + K_D)K_P - IK_I > 0$. The presence of K_D relaxes somewhat the requirement that K_P be large to achieve stability. The steady-state errors are zero, and the transient response can be improved because three of the coefficients of the characteristic equation can be selected. To make further statements requires the root locus technique presented in Section 28.9.

Proportional, integral, and derivative actions and their various combinations are not the only control laws possible, but they are the most common. PID controllers will remain for some time the standard against which any new designs must compete.

The conclusions reached concerning the performance of the various control laws are strictly true only for the plant model forms considered. These are the first-order model without numerator dynamics and the second-order model with a root at $s = 0$ and no numerator zeros. The analysis of a control law for any other linear system follows the preceding pattern. The overall system transfer functions are obtained, and all of the linear system analysis techniques can be applied to predict the system's performance. If the performance is unsatisfactory, a new control law is tried and the process repeated. When this process fails to achieve an acceptable design, more systematic methods of altering the system's structure are needed; they are discussed in later sections. We have used step functions as the test signals because they are the most common and perhaps represent the severest test of system performance. Impulse, ramp, and sinusoidal test signals are also employed. The type to use should be made clear in the design specifications.

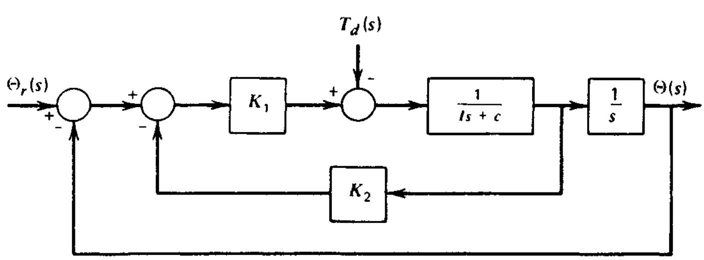


Fig. 28.24 Tachometer feedback arrangement to replace PD control for the position servo.[1]

28.6 CONTROLLER HARDWARE

The control law must be implemented by a physical device before the control engineer's task is complete. The earliest devices were purely kinematic and were mechanical elements such as gears, levers, and diaphragms that usually obtained their power from the controlled variable. Most controllers now are analog electronic, hydraulic, pneumatic, or digital electronic devices. We now consider the analog type. Digital controllers are covered starting in Section 28.10.

28.6.1 Feedback Compensation and Controller Design

Most controllers that implement versions of the PID algorithm are based on the following feedback principle. Consider the single-loop system shown in Fig. 28.1. If the open-loop transfer function is large enough that $|G(s)H(s)| >> 1$, the closed-loop transfer function is approximately given by

$$T(s) = \frac{G(s)}{1 + G(s)H(s)} \approx \frac{G(s)}{G(s)H(s)} = \frac{1}{H(s)} \tag{28.24}$$

The principle states that a power unit $G(s)$ can be used with a feedback element $H(s)$ to create a desired transfer function $T(s)$. The power unit must have a gain high enough that $|G(s)H(s)| >> 1$, and the feedback elements must be selected so that $H(s) = 1/T(s)$. This principle was used in Section 28.1 to explain the design of a feedback amplifier.

28.6.2 Electronic Controllers

The *operational amplifier* (*op amp*) is a high-gain amplifier with a high input impedance. A diagram of an op amp with feedback and input elements with impedances $T_f(s)$ and $T_i(s)$ is shown in Fig. 28.25. An approximate relation is

$$\frac{E_o(s)}{E_i(s)} = -\frac{T_f(s)}{T_i(s)}$$

The various control modes can be obtained by proper selection of the impedances. A proportional controller can be constructed with a *multiplier,* which uses two resistors, as shown in Fig. 28.26. An *inverter* is a multiplier circuit with $R_f = R_i$. It is sometimes needed because of the sign reversal property of the op amp. The multiplier circuit can be modified to act as an adder (Fig. 28.27).

PI control can be implemented with the circuit of Fig. 28.28. Figure. 28.29 shows a complete system using op amps for PI control. The inverter is needed to create an error detector. Many industrial controllers provide the operator with a choice of control modes, and the operator can switch from one mode to another when the process characteristics or control objectives change. When a switch occurs, it is necessary to provide any integrators with the proper initial voltages, or else undesirable transients will occur when the integrator is switched into the system. Commercially available controllers usually have built-in circuits for this purpose.

In theory, a differentiator can be created by interchanging the resistance and capacitance in the integrating op amp. The difficulty with this design is that no electrical signal is "pure." Contamination always exists as a result of voltage spikes, ripple, and other transients generally categorized as "noise." These high-frequency signals have large slopes compared with the more slowly varying primary signal, and thus they will dominate the output of the differentiator. In practice, this problem is solved by filtering out high-frequency signals, either with a low-pass filter inserted in cascade with the differentiator, or by using a redesigned differentiator such as the one shown in Fig. 28.30. For the ideal PD controller, $R_1 = 0$. The attenuation curve for the ideal controller breaks upward at $\omega = 1/R_2C$ with a slope of 20 db/decade. The curve for the practical controller does the same but then becomes flat for $\omega > (R_1 + R_2)/R_1R_2C$. This provides the required limiting effect at high frequencies.

PID control can be implemented by joining the PI and PD controllers in parallel, but this is expensive because of the number of op amps and power supplies required. Instead, the usual implementation is that shown in Fig. 28.31. The circuit limits the effect of frequencies above $\omega = 1/$

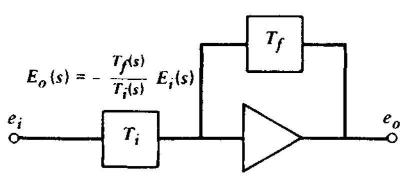


Fig. 28.25 Operational amplifier (op amp).[1]

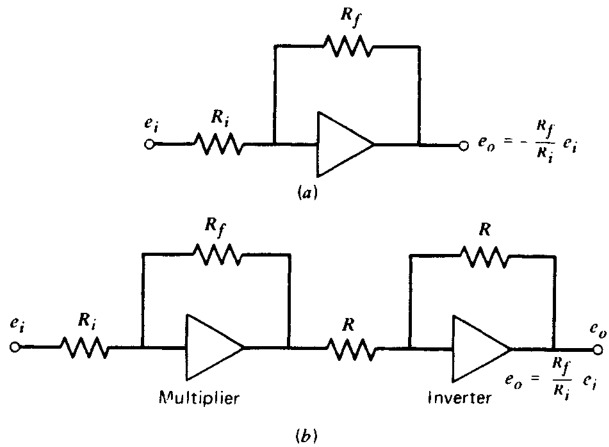


Fig. 28.26 Op-amp implementation of proportional control.[1]

$\beta R_1 C_1$. When $R_1 = 0$, ideal PID control results. This is sometimes called the *noninteractive* algorithm because the effect of each of the three modes is additive, and they do not interfere with one another. The form given for $R_1 \neq 0$ is the *real* or *interactive* algorithm. This name results from the fact that historically it was difficult to implement noninteractive PID control with mechanical or pneumatic devices.

28.6.3 Pneumatic Controllers

The nozzle-flapper introduced in Section 28.4 is a high-gain device that is difficult to use without modification. The gain K_f is known only imprecisely and is sensitive to changes induced by temperature and other environmental factors. Also, the linear region over which Eq. (28.14) applies is very small. However, the device can be made useful by compensating it with feedback elements, as was illustrated with the electropneumatic valve positioner shown in Fig. 28.19.

28.6.4 Hydraulic Controllers

The basic unit for synthesis of hydraulic controllers is the hydraulic servomotor. The nozzle-flapper concept is also used in hydraulic controllers.[4] A PI controller is shown in Fig. 28.32. It can be modified for P-action. Derivative action has not seen much use in hydraulic controllers. This action supplies damping to the system, but hydraulic systems are usually highly damped intrinsically because of the viscous working fluid. PI control is the algorithm most commonly implemented with hydraulics.

28.7 FURTHER CRITERIA FOR GAIN SELECTION

Once the form of the control law has been selected, the gains must be computed in light of the performance specifications. In the examples of the PID family of control laws in Section 28.5, the damping ratio, dominant time constant, and steady-state error were taken to be the primary indicators of system performance in the interest of simplicity. In practice, the criteria are usually more detailed. For example, the rise time and maximum overshoot, as well as the other transient response specifications of the previous chapter, may be encountered. Requirements can also be stated in terms of

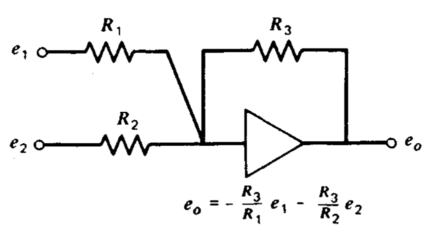


Fig. 28.27 Op-amp adder circuit.[1]

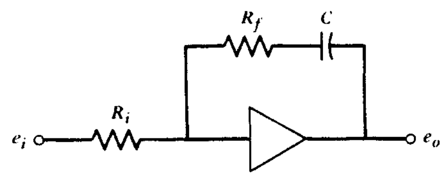


Fig. 28.28 Op-amp implementation of PI control.[1]

frequency response characteristics, such as bandwidth, resonant frequency, and peak amplitude. Whatever specific form they take, a complete set of specifications for control system performance generally should include the following considerations, for given forms of the command and disturbance inputs:

1. Equilibrium specifications
   - **(a)** Stability
   - **(b)** Steady-state error
2. Transient specifications
   - **(a)** Speed of response
   - **(b)** Form of response
3. Sensitivity specifications
   - **(a)** Sensitivity to parameter variations
   - **(b)** Sensitivity to model inaccuracies
   - **(c)** Noise rejection (bandwidth, etc.)

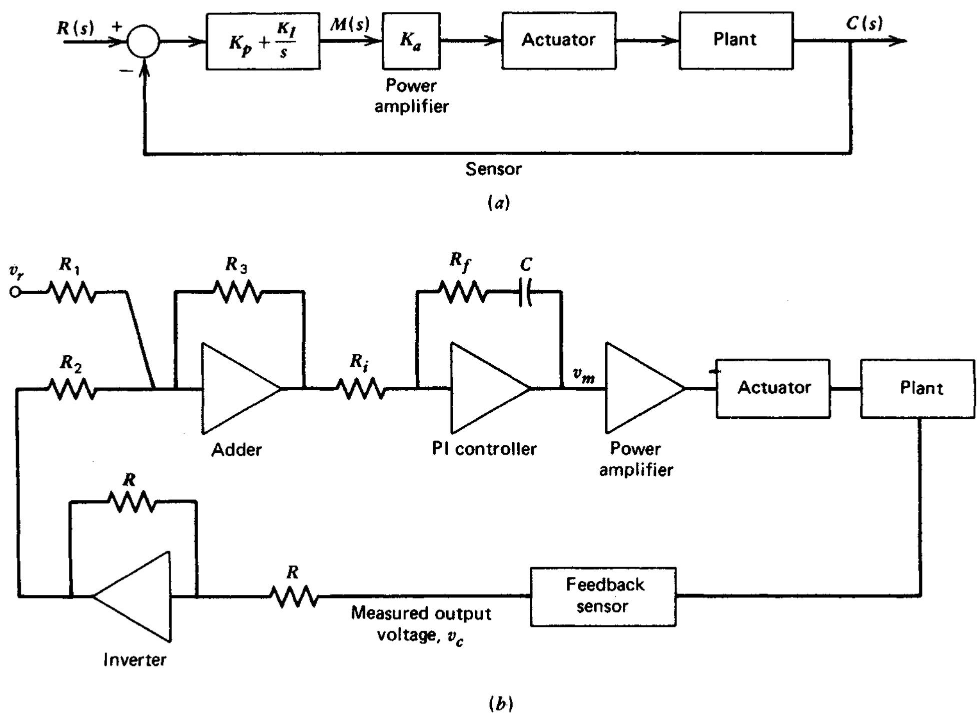
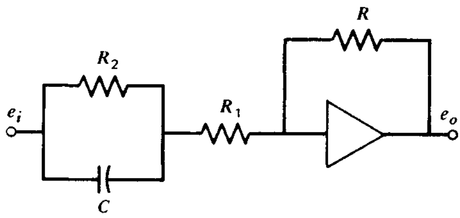


Fig. 28.29 Implementation of a PI-controller using op amps. (*a*) Diagram of the system. (*b*) Diagram showing how the op amps are connected.[2]



$$e_o = -K_p\left(e_i + T_D\frac{de_i}{dt}\right) - \alpha T_D\frac{de_o}{dt}$$

$$K_p = \frac{R}{R_1 + R_2} \qquad T_D = R_2C \qquad \alpha = \frac{R_1}{R_1 + R_2}$$

Fig. 28.30 Practical op-amp implementation of PD control.[1]

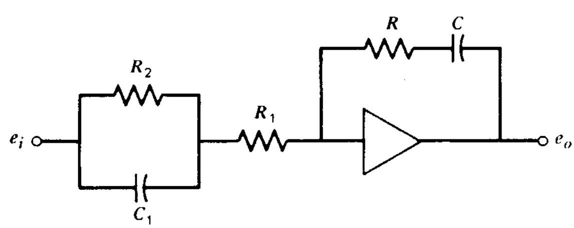
In addition to these performance stipulations, the usual engineering considerations of initial cost, weight, maintainability, and so on must be taken into account. The considerations are highly specific to the chosen hardware, and it is difficult to deal with such issues in a general way.

Two approaches exist for designing the controller. The proper one depends on the quality of the analytical description of the plant to be controlled. If an accurate model of the plant is easily developed, we can design a specialized controller for the particular application. The range of adjustment of controller gains in this case can usually be made small because the accurate plant model allows the gains to be precomputed with confidence. This technique reduces the cost of the controller and can often be applied to electromechanical systems.

The second approach is used when the plant is relatively difficult to model, which is often the case in process control. A standard controller with several control modes and wide ranges of gains is used, and the proper mode and gain settings are obtained by testing the controller on the process in the field. This approach should be considered when the cost of developing an accurate plant model might exceed the cost of controller tuning in the field. Of course, the plant must be available for testing for this approach to be feasible.

28.7.1 Performance Indices

The performance criteria encountered thus far require a set of conditions to be specified—for example, one for steady-state error, one for damping ratio, and one for the dominant time constant. If there



$$e_o = -\left(K_p e_i + K_I\int_0^t e_i\,dt + K_D\frac{de_i}{dt}\right) - \beta R_1C_1\frac{de_o}{dt}$$

$$\beta = \frac{R_2}{R_1 + R_2} \qquad K_p = \beta\frac{RC + R_2C_1}{R_2C}$$

$$K_I = \frac{\beta}{R_2C} \qquad K_D = \beta RC_1$$

Fig. 28.31 Practical op-amp implementation of PID control.[1]

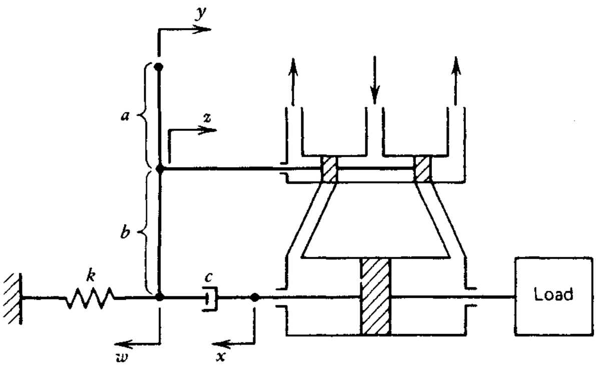


Fig. 28.32 Hydraulic implementation of PI control.[1]

are many such conditions, and if the system is of high order with several gains to be selected, the design process can get quite complicated because transient and steady-state criteria tend to drive the design in different directions. An alternative approach is to specify the system's desired performance by means of one analytical expression called a *performance index.* Powerful analytical and numerical methods are available that allow the gains to be systematically computed by minimizing (or maximizing) this index.

To be useful, a performance index must be selective. The index must have a sharply defined extremum in the vicinity of the gain values that give the desired performance. If the numerical value of the index does not change very much for large changes in the gains from their optimal values, the index will not be selective.

Any practical choice of a performance index must be easily computed, either analytically, numerically, or experimentally. Four common choices for an index are the following:

$$J = \int_0^\infty |e(t)|\, dt \qquad \text{(IAE Index)} \tag{28.25}$$

$$J = \int_0^\infty t|e(t)|\, dt \qquad \text{(ITAE Index)} \tag{28.26}$$

$$J = \int_0^\infty [e(t)]^2\, dt \qquad \text{(ISE Index)} \tag{28.27}$$

$$J = \int_0^\infty t[e(t)]^2\, dt \qquad \text{(ITSE Index)} \tag{28.28}$$

where $e(t)$ is the system error. This error usually is the difference between the desired and the actual values of the output. However, if $e(t)$ does not approach zero as $t \rightarrow \infty$, the preceding indices will not have finite values. In this case, $e(t)$ can be defined as $e(t) = c(\infty) - c(t)$, where $c(t)$ is the output variable. If the index is to be computed numerically or experimentally, the infinite upper limit can be replaced by a time t_f large enough that $e(t)$ is negligible for $t > t_f$.

The *integral absolute-error* (IAE) criterion (28.25) expresses mathematically that the designer is not concerned with the sign of the error, only its magnitude. In some applications, the IAE criterion describes the fuel consumption of the system. The index says nothing about the relative importance of an error occurring late in the response versus an error occurring early. Because of this, the index is not as selective as the *integral-of-time-multiplied absolute-error* (ITAE) criterion (28.26). Since the multiplier t is small in the early stages of the response, this index weights early errors less heavily than later errors. This makes sense physically. No system can respond instantaneously, and the index is lenient accordingly, while penalizing any design that allows a nonzero error to remain for a long time. Neither criterion allows highly underdamped or highly overdamped systems to be optimum. The ITAE criterion usually results in a system whose step response has a slight overshoot and well-damped oscillations.

The *integral squared-error* (ISE) and *integral-of-time-multiplied squared-error* (ITSE) criteria are analogous to the IAE and ITAE criteria, except that the square of the error is employed, for three reasons: (1) in some applications, the squared error represents the system's power consumption; (2) squaring the error weights large errors much more heavily than small errors; (3) the squared error is much easier to handle analytically. The derivative of a squared term is easier to compute than that of an absolute value and does not have a discontinuity at $e = 0$. These differences are important when the system is of high order with multiple error terms.

The closed-form solution for the response is not required to evaluate a performance index. For a given set of parameter values, the response and the resulting index value can be computed numerically. The optimum solution can be obtained using systematic computer search procedures; this makes this approach suitable for use with nonlinear systems.

28.7.2 Optimal Control Methods

Optimal control theory includes a number of algorithms for systematic design of a control law to minimize a performance index, such as the following generalization of the ISE index, called the *quadratic* index:

$$J = \int_0^\infty (\mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{u}^T\mathbf{R}\mathbf{u})\, dt \tag{28.29}$$

where $\mathbf{x}$ and $\mathbf{u}$ are the deviations of the state and control vectors from the desired reference values. For example, in a servomechanism, the state vector might consist of the position and velocity, and the control vector might be a scalar—the force or torque produced by the actuator. The matrices $\mathbf{Q}$ and $\mathbf{R}$ are chosen by the designer to provide relative weighting for the elements of $\mathbf{x}$ and $\mathbf{u}$. If the plant can be described by the linear state-variable model

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \tag{28.30}$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u} \tag{28.31}$$

where $\mathbf{y}$ is the vector of outputs—for example, position and velocity—then the solution of this *linear-quadratic* control problem is the linear control law:

$$\mathbf{u} = \mathbf{K}\mathbf{y} \tag{28.32}$$

where $\mathbf{K}$ is a matrix of gains that can be found by several algorithms.[1,5,6] A valid solution is guaranteed to yield a stable closed-loop system, a major benefit of this method.

Even if it is possible to formulate the control problem in this way, several practical difficulties arise. Some of the terms in (28.29) might be beyond the influence of the control vector $\mathbf{u}$; the system is then *uncontrollable.* Also, there might not be enough information in the output equation (28.31) to achieve control, and the system is then *unobservable.* Several tests are available to check controllability and observability. Not all of the necessary state variables might be available for feedback, or the feedback measurements might be noisy or biased. Algorithms known as *observers, state reconstructors, estimators,* and *digital filters* are available to compensate for the missing information. Another source of error is the uncertainty in the values of the coefficient matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$. Identification schemes can be used to compare the predicted and the actual system performance, and to adjust the coefficient values "on-line."

28.7.3 The Ziegler–Nichols Rules

The difficulty of obtaining accurate transfer function models for some processes has led to the development of empirically based rules of thumb for computing the optimum gain values for a controller. Commonly used guidelines are the *Ziegler–Nichols rules,* which have proved so helpful that they are still in use 50 years after their development. The rules actually consist of two separate methods. The first method requires the open-loop step response of the plant, while the second uses the results of experiments performed with the controller already installed. While primarily intended for use with systems for which no analytical model is available, the rules are also helpful even when a model can be developed.

Ziegler and Nichols developed their rules from experiments and analysis of various industrial processes. Using the IAE criterion with a unit step response, they found that controllers adjusted according to the following rules usually had a step response that was oscillatory but with enough damping so that the second overshoot was less than 25% of the first (peak) overshoot. This is the *quarter-decay* criterion and is sometimes used as a specification.

The first method is the *process-reaction* method and relies on the fact that many processes have an open-loop step response like that shown in Fig. 28.33. This is the *process signature* and is characterized by two parameters, R and L. R is the slope of a line tangent to the steepest part of the response curve, and L is the time at which this line intersects the time axis. First- and second-order linear systems do not yield positive values for L, and so the method cannot be applied to such

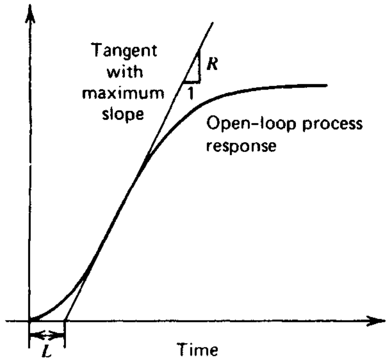


Fig. 28.33 Process signature for a unit step input.[1]

systems. However, third- and higher-order linear systems with sufficient damping do yield such a response. If so, the Zielger–Nichols rules recommend the controller settings given in Table 28.2.

The *ultimate-cycle* method uses experiments with the controller in place. All control modes except proportional are turned off, and the process is started with the proportional gain K_P set at a low value. The gain is slowly increased until the process begins to exhibit sustained oscillations. Denote the period of this oscillation by P_u and the corresponding *ultimate gain* by K_{Pu}. The Ziegler–Nichols recommendations are given in Table 28.2 in terms of these parameters. The proportional gain is lower for PI control than for P control, and is higher for PID control because I action increases the order of the system and thus tends to destabilize it; thus, a lower gain is needed. On the other hand, D action tends to stabilize the system; hence, the proportional gain can be increased without degrading the stability characteristics. Because the rules were developed for a typical case out of many types of processes, final tuning of the gains in the field is usually necessary.

28.7.4 Nonlinearities and Controller Performance

All physical systems have nonlinear characteristics of some sort, although they can often be modeled as linear systems provided the deviations from the linearization reference condition are not too great. Under certain conditions, however, the nonlinearities have significant effects on the system's performance. One such situation can occur during the start-up of a controller if the initial conditions are much different from the reference condition for linearization. The linearized model is then not accurate, and nonlinearities govern the behavior. If the nonlinearities are mild, there might not be much of a problem. Where the nonlinearities are severe, such as in process control, special consideration must be given to start-up. Usually, in such cases, the control signal sent to the final control elements is manually adjusted until the system variables are within the linear range of the controller.

Table 28.2 The Ziegler–Nichols Rules

Controller transfer function $G(s) = K_p\left(1 + \frac{1}{T_I s} + T_D s\right)$

Control Mode	Process-Reaction Method	Ultimate-Cycle Method
P control	$K_p = \frac{1}{RL}$	$K_p = 0.5K_{pu}$
PI control	$K_p = \frac{0.9}{RL}$	$K_p = 0.45K_{pu}$
	$T_I = 3.3L$	$T_I = 0.83P_u$
PID control	$K_p = \frac{1.2}{RL}$	$K_p = 0.6K_{pu}$
	$T_I = 2L$	$T_I = 0.5P_u$
	$T_D = 0.5L$	$T_D = 0.125P_u$

Then the system is switched into automatic mode. Digital computers are often used to replace the manual adjustment process because they can be readily coded to produce complicated functions for the start-up signals. Care must also be taken when switching from manual to automatic. For example, the integrators in electronic controllers must be provided with the proper initial conditions.

28.7.5 Reset Windup

In practice, all actuators and final control elements have a limited operating range. For example, a motor–amplifier combination can produce a torque proportional to the input voltage over only a limited range. No amplifier can supply an infinite current; there is a maximum current and thus a maximum torque that the system can produce. The final control elements are said to be *overdriven* when they are commanded by the controller to do something they cannot do. Since the limitations of the final control elements are ultimately due to the limited rate at which they can supply energy, it is important that all system performance specifications and controller designs be consistent with the energy-delivery capabilities of the elements to be used.

Controllers using integral action can exhibit the phenomenon called *reset windup* or *integrator buildup* when overdriven, if they are not properly designed. For a step change in set point, the proportional term responds instantly and saturates immediately if the set-point change is large enough. On the other hand, the integral term does not respond as fast, It integrates the error signal and saturates some time later if the error remains large for a long enough time. As the error decreases, the proportional term no longer causes saturation. However, the integral term continues to increase as long as the error has not changed sign, and thus the manipulated variable remains saturated. Even though the output is very near its desired value, the manipulated variable remains saturated until after the error has reversed sign. The result can be an undesirable overshoot in the response of the controlled variable.

Limits on the controller prevent the voltages from exceeding the value required to saturate the actuator, and thus protect the actuator, but they do not prevent the integral build-up that causes the overshoot. One way to prevent integrator build-up is to select the gains so that saturation will never occur. This requires knowledge of the maximum input magnitude that the system will encounter. General algorithms for doing this are not available; some methods for low-order systems are presented in Ref. 1, Chap. 7, and Ref. 2, Chap. 7. Integrator build-up is easier to prevent when using digital control; this is discussed in Section 28.10.

28.8 COMPENSATION AND ALTERNATIVE CONTROL STRUCTURES

A common design technique is to insert a *compensator* into the system when the PID control algorithm can be made to satisfy most but not all of the design specifications. A compensator is a device that alters the response of the controller so that the overall system will have satisfactory performance. The three categories of compensation techniques generally recognized are *series compensation, parallel* (or *feedback*) *compensation,* and *feedforward compensation.* The three structures are loosely illustrated in Fig. 28.34, where we assume the final control elements have a unity transfer function. The transfer function of the controller is $G_1(s)$. The feedback elements are represented by $H(s)$, and the compensator by $G_c(s)$. We assume that the plant is unalterable, as is usually the case in control system design. The choice of compensation structure depends on what type of specifications must be satisfied. The physical devices used as compensators are similar to the pneumatic, hydraulic, and electrical devices treated previously. Compensators can be implemented in software for digital control applications.

28.8.1 Series Compensation

The most commonly used series compensators are the *lead,* the *lag,* and the *lead-lag* compensators. Electrical implementations of these are shown in Fig. 28.35. Other physical implementations are available. Generally, the lead compensator improves the speed of response; the lag compensator decreases the steady-state error; and the lead-lag affects both. Graphical aids, such as the root locus and frequency response plots, are usually needed to design these compensators (Ref. 1, Chap. 8; Ref. 2, Chap. 9).

28.8.2 Feedback Compensation and Cascade Control

The use of a tachometer to obtain velocity feedback, as in Fig. 28.24, is a case of feedback compensation. The feedback-compensation principle of Fig. 28.3 is another. Another form is *cascade control,* in which another controller is inserted within the loop of the original control system (Fig. 28.36). The new controller can be used to achieve better control of variables within the forward path of the system. Its set point is manipulated by the first controller.

Cascade control is frequently used when the plant cannot be satisfactorily approximated with a model of second order or lower. This is because the difficulty of analysis and control increases rapidly with system order. The characteristic roots of a second-order system can easily be expressed in analytical form. This is not so for third order or higher, and few general design rules are available.