1.1 Introduction

Control systems are an integral part of modern society. Numerous applications are all around us: The rockets fire, and the space shuttle lifts off to earth orbit; in splashing cooling water, a metallic part is automatically machined; a self-guided vehicle delivering material to workstations in an aerospace assembly plant glides along the floor seeking its destination. These are just a few examples of the automatically controlled systems that we can create.

We are not the only creators of automatically controlled systems; these systems also exist in nature. Within our own bodies are numerous control systems, such as the pancreas, which regulates our blood sugar. In time of "fight or flight," our adrenaline increases along with our heart rate, causing more oxygen to be delivered to our cells. Our eyes follow a moving object to keep it in view; our hands grasp the object and place it precisely at a predetermined location.

Even the nonphysical world appears to be automatically regulated. Models have been suggested showing automatic control of student performance. The input to the model is the student's available study time, and the output is the grade. The model can be used to predict the time required for the grade to rise if a sudden increase in study time is available. Using this model, you can determine whether increased study is worth the effort during the last week of the term.

Control System Definition

A control system consists of *subsystems* and *processes* (or *plants)* assembled for the purpose of obtaining a desired *output* with desired *performance*, given a specified *input*. Figure 1.1 shows a control system in its simplest form, where the input represents a desired output.

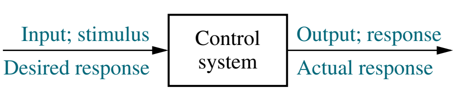


FIGURE 1.1 Simplified description of a control system

For example, consider an elevator. When the fourth-floor button is pressed on the first floor, the elevator rises to the fourth floor with a speed and floor-leveling accuracy designed for passenger comfort. The push of the fourth-floor button is an *input* that represents our desired *output*, shown as a step function in Figure 1.2. The *performance* of the elevator can be seen from the elevator response curve in the figure.

Two major measures of performance are apparent: (1) the transient response and (2) the steady-state error. In our example, passenger comfort and passenger patience are dependent upon the transient response. If this response is too fast, passenger comfort is sacrificed; if too slow, passenger patience is sacrificed. The steady-state error is another important performance specification since passenger safety and convenience would be sacrificed if the elevator did not properly level.

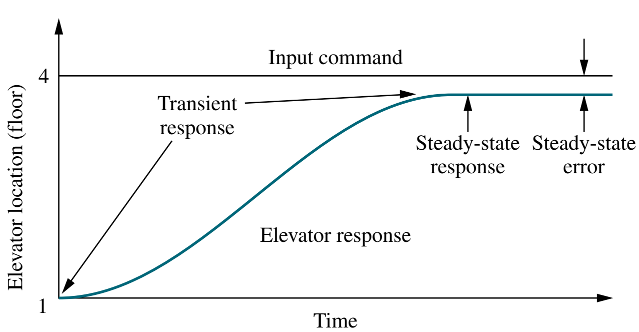
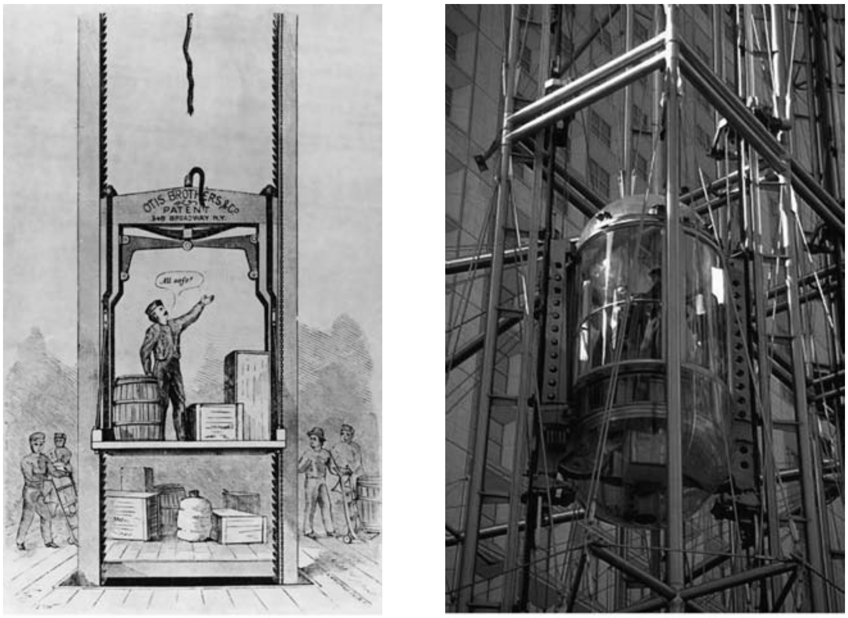


FIGURE 1.2 Elevator response



(*a*) (*b*)

FIGURE 1.3 **a.** Early elevators were controlled by hand ropes or an elevator operator. Here a rope is cut to demonstrate the safety brake, an innovation in early elevators (© Bettman/Corbis); **b.** One of two modern Duo-lift elevators makes its way up the Grande Arche in Paris. Two elevators are driven by one motor, with each car acting as a counterbalance to the other. Today, elevators are fully automatic, using control systems to regulate position and velocity.

Advantages of Control Systems

With control systems we can move large equipment with precision that would otherwise be impossible. We can point huge antennas toward the farthest reaches of the universe to pick up faint radio signals; controlling these antennas by hand would be impossible. Because of control systems, elevators carry us quickly to our destination, automatically stopping at the right floor (Figure 1.3). We alone could not provide the power required for the load and the speed; motors provide the power, and control systems regulate the position and speed.

We build control systems for four primary reasons:

1. Power amplification
2. Remote control
3. Convenience of input form
4. Compensation for disturbances

For example, a radar antenna, positioned by the low-power rotation of a knob at the input, requires a large amount of power for its output rotation. A control system can produce the needed power amplification, or power *gain*.

Robots designed by control system principles can compensate for human disabilities. Control systems are also useful in remote or dangerous locations. For example, a remote-controlled robot arm can be used to pick up material in a radioactive environment. Figure 1.4 shows a robot arm designed to work in contaminated environments.

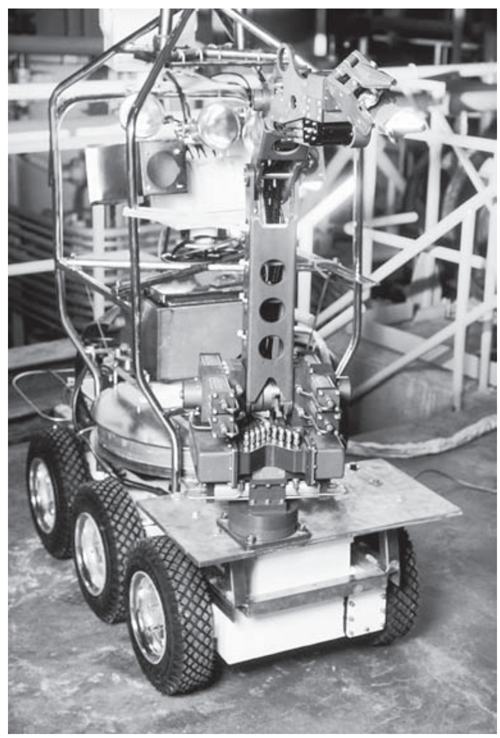


FIGURE 1.4 *Rover* was built to work in contaminated areas at Three Mile Island in Middleton, Pennsylvania, where a nuclear accident occurred in 1979. The remote-controlled robot's long arm can be seen at the front of the vehicle.

Control systems can also be used to provide convenience by changing the form of the input. For example, in a temperature control system, the input is a *position* on a thermostat. The output is *heat*. Thus, a convenient position input yields a desired thermal output.

Another advantage of a control system is the ability to compensate for disturbances. Typically, we control such variables as temperature in thermal systems, position and velocity in mechanical systems, and voltage, current, or frequency in electrical systems. The system must be able to yield the correct output even with a disturbance. For example, consider an antenna system that points in a commanded direction. If wind forces the antenna from its commanded position, or if noise enters internally, the system must be able to detect the disturbance and correct the antenna's position. Obviously, the system's input will not change to make the correction. Consequently, the system itself must measure the amount that the disturbance has repositioned the antenna and then return the antenna to the position commanded by the input.

1.2 A History of Control Systems

Feedback control systems are older than humanity. Numerous biological control systems were built into the earliest inhabitants of our planet. Let us now look at a brief history of human-designed control systems.[1]

Liquid-Level Control

The Greeks began engineering feedback systems around 300 B.C. A water clock invented by Ktesibios operated by having water trickle into a measuring container at a constant rate. The level of water in the measuring container could be used to tell time. For water to trickle at a constant rate, the supply tank had to be kept at a constant level. This was accomplished using a float valve similar to the water-level control in today's flush toilets.

Soon after Ktesibios, the idea of liquid-level control was applied to an oil lamp by Philon of Byzantium. The lamp consisted of two oil containers configured vertically. The lower pan was open at the top and was the fuel supply for the flame. The closed upper bowl was the fuel reservoir for the pan below. The containers were interconnected by two capillary tubes and another tube, called a *vertical riser*, which was inserted into the oil in the lower pan just below the surface. As the oil burned, the base of the vertical riser was exposed to air, which forced oil in the reservoir above to flow through the capillary tubes and into the pan. The transfer of fuel from the upper reservoir to the pan stopped when the previous oil level in the pan was reestablished, thus blocking the air from entering the vertical riser. Hence, the system kept the liquid level in the lower container constant.

Steam Pressure and Temperature Controls

Regulation of steam pressure began around 1681 with Denis Papin's invention of the safety valve. The concept was further elaborated on by weighting the valve top. If the upward pressure from the boiler exceeded the weight, steam was released, and the pressure decreased. If it did not exceed the weight, the valve did not open, and the pressure inside the boiler increased. Thus, the weight on the valve top set the internal pressure of the boiler.

Also in the seventeenth century, Cornelis Drebbel in Holland invented a purely mechanical temperature control system for hatching eggs. The device used a vial of alcohol and mercury with a floater inserted in it. The floater was connected to a damper that controlled a flame. A portion of the vial was inserted into the incubator to sense the heat generated by the fire. As the heat increased, the alcohol and mercury expanded, raising the floater, closing the damper, and reducing the flame. Lower temperature caused the float to descend, opening the damper and increasing the flame.

Speed Control

In 1745, speed control was applied to a windmill by Edmund Lee. Increasing winds pitched the blades farther back, so that less area was available. As the wind

---

[1] See *Bennett* (*1979*) and *Mayr* (*1970*) for definitive works on the history of control systems.

decreased, more blade area was available. William Cubitt improved on the idea in 1809 by dividing the windmill sail into movable louvers.

Also in the eighteenth century, James Watt invented the flyball speed governor to control the speed of steam engines. In this device, two spinning flyballs rise as rotational speed increases. A steam valve connected to the flyball mechanism closes with the ascending flyballs and opens with the descending flyballs, thus regulating the speed.

Stability, Stabilization, and Steering

Control systems theory as we know it today began to crystallize in the latter half of the nineteenth century. In 1868, James Clerk Maxwell published the stability criterion for a third-order system based on the coefficients of the differential equation. In 1874, Edward John Routh, using a suggestion from William Kingdon Clifford that was ignored earlier by Maxwell, was able to extend the stability criterion to fifth-order systems. In 1877, the topic for the Adams Prize was "The Criterion of Dynamical Stability." In response, Routh submitted a paper entitled *A Treatise on the Stability of a Given State of Motion* and won the prize. This paper contains what is now known as the Routh-Hurwitz criterion for stability, which we will study in Chapter 6. Alexandr Michailovich Lyapunov also contributed to the development and formulation of today's theories and practice of control system stability. A student of P. L. Chebyshev at the University of St. Petersburg in Russia, Lyapunov extended the work of Routh to nonlinear systems in his 1892 doctoral thesis, entitled *The General Problem of Stability of Motion*.

During the second half of the 1800s, the development of control systems focused on the steering and stabilizing of ships. In 1874, Henry Bessemer, using a gyro to sense a ship's motion and applying power generated by the ship's hydraulic system, moved the ship's saloon to keep it stable (whether this made a difference to the patrons is doubtful). Other efforts were made to stabilize platforms for guns as well as to stabilize entire ships, using pendulums to sense the motion.

Twentieth-Century Developments

It was not until the early 1900s that automatic steering of ships was achieved. In 1922, the Sperry Gyroscope Company installed an automatic steering system that used the elements of compensation and adaptive control to improve performance. However, much of the general theory used today to improve the performance of automatic control systems is attributed to Nicholas Minorsky, a Russian born in 1885. It was his theoretical development applied to the automatic steering of ships that led to what we call today proportional-plus-integral-plus-derivative (PID), or three-mode, controllers, which we will study in Chapters 9 and 11.

In the late 1920s and early 1930s, H. W. Bode and H. Nyquist at Bell Telephone Laboratories developed the analysis of feedback amplifiers. These contributions evolved into sinusoidal frequency analysis and design techniques currently used for feedback control system, and are presented in Chapters 10 and 11.

In 1948, Walter R. Evans, working in the aircraft industry, developed a graphical technique to plot the roots of a characteristic equation of a feedback system whose parameters changed over a particular range of values. This technique, now known as the root locus, takes its place with the work of Bode and Nyquist in forming the foundation of linear control systems analysis and design theory. We will study root locus in Chapters 8, 9, and 13.

Contemporary Applications

Today, control systems find widespread application in the guidance, navigation, and control of missiles and spacecraft, as well as planes and ships at sea. For example, modern ships use a combination of electrical, mechanical, and hydraulic components to develop rudder commands in response to desired heading commands. The rudder commands, in turn, result in a rudder angle that steers the ship.

We find control systems throughout the process control industry, regulating liquid levels in tanks, chemical concentrations in vats, as well as the thickness of fabricated material. For example, consider a thickness control system for a steel plate finishing mill. Steel enters the finishing mill and passes through rollers. In the finishing mill, X-rays measure the actual thickness and compare it to the desired thickness. Any difference is adjusted by a screw-down position control that changes the roll gap at the rollers through which the steel passes. This change in roll gap regulates the thickness.

Modern developments have seen widespread use of the digital computer as part of control systems. For example, computers in control systems are for industrial robots, spacecraft, and the process control industry. It is hard to visualize a modern control system that does not use a digital computer.

The space shuttle contains numerous control systems operated by an onboard computer on a time-shared basis. Without control systems, it would be impossible to guide the shuttle to and from earth's orbit or to adjust the orbit itself and support life on board. Navigation functions programmed into the shuttle's computers use data from the shuttle's hardware to estimate vehicle position and velocity. This information is fed to the guidance equations that calculate commands for the shuttle's flight control systems, which steer the spacecraft. In space, the flight control system gimbals (rotates) the orbital maneuvering system (OMS) engines into a position that provides thrust in the commanded direction to steer the spacecraft. Within the earth's atmosphere, the shuttle is steered by commands sent from the flight control system to the aerosurfaces, such as the elevons.

Within this large control system represented by navigation, guidance, and control are numerous subsystems to control the vehicle's functions. For example, the elevons require a control system to ensure that their position is indeed that which was commanded, since disturbances such as wind could rotate the elevons away from the commanded position. Similarly, in space, the gimbaling of the orbital maneuvering engines requires a similar control system to ensure that the rotating engine can accomplish its function with speed and accuracy. Control systems are also used to control and stabilize the vehicle during its descent from orbit. Numerous small jets that compose the reaction control system (RCS) are used initially in the exoatmosphere, where the aerosurfaces are ineffective. Control is passed to the aerosurfaces as the orbiter descends into the atmosphere.

Inside the shuttle, numerous control systems are required for power and life support. For example, the orbiter has three fuel-cell power plants that convert hydrogen and oxygen (reactants) into electricity and water for use by the crew. The fuel cells involve the use of control systems to regulate temperature and pressure. The reactant tanks are kept at constant pressure as the quantity of reactant diminishes. Sensors in the tanks send signals to the control systems to turn heaters on or off to keep the tank pressure constant (*Rockwell International*, *1984*).

Control systems are not limited to science and industry. For example, a home heating system is a simple control system consisting of a thermostat containing a bimetallic material that expands or contracts with changing temperature. This expansion or contraction moves a vial of mercury that acts as a switch, turning the heater on or off. The amount of expansion or contraction required to move the mercury switch is determined by the temperature setting.

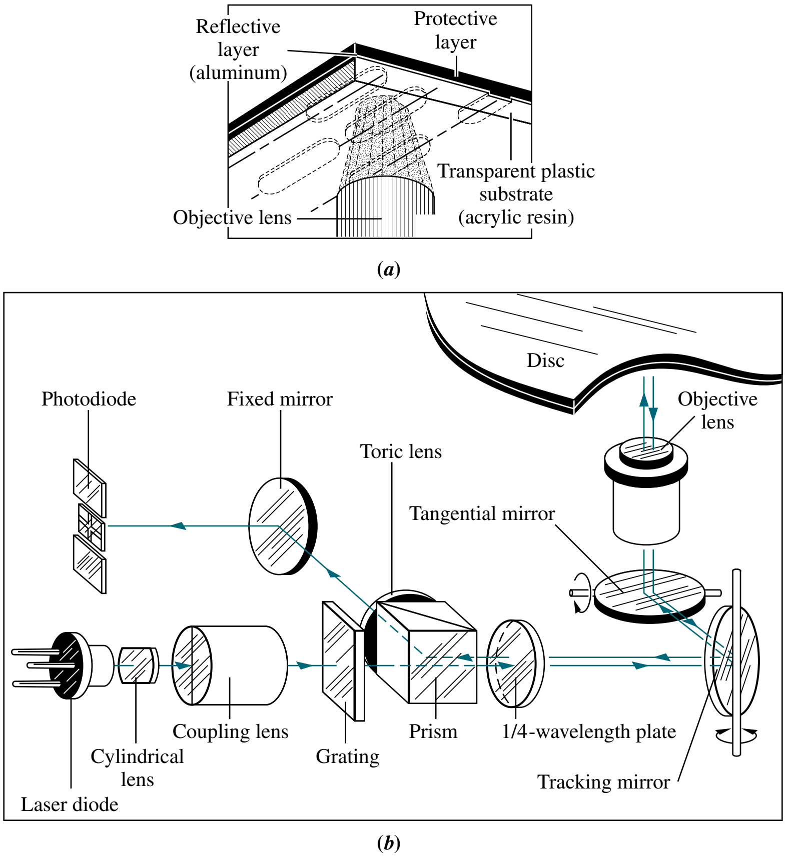


FIGURE 1.5 Optical playback system: **a.** objective lens reading pits on an optical disc; **b.** optical path for playback, showing tracking mirror rotated by a control system to keep the laser beam positioned on the pits (Pioneer Electronics (USA), Inc.)

Home entertainment systems also have built-in control systems. For example, in an optical disk recording system microscopic pits representing the information are burned into the disc by a laser during the recording process. During playback, a reflected laser beam focused on the pits changes intensity (Figure 1.5). The light intensity changes are converted to an electrical signal and processed as sound or picture. A control system keeps the laser beam positioned on the pits, which are cut as concentric circles.

There are countless other examples of control systems, from the everyday to the extraordinary. As you begin your study of control systems engineering, you will become more aware of the wide variety of applications.

1.3 System Configurations

In this section, we discuss two major configurations of control systems: open loop and closed loop. We can consider these configurations to be the internal architecture of the total system shown in Figure 1.1. Finally, we show how a digital computer forms part of a control system's configuration.

Open-Loop Systems

A generic *open-loop system* is shown in Figure 1.6(*a*). It starts with a subsystem called an *input transducer*, which converts the form of the input to that used by the *controller*. The controller drives a *process* or a *plant*. The input is sometimes called the *reference*, while the output can be called the *controlled variable*. Other signals, such as *disturbances*, are shown added to the controller and process outputs via *summing junctions*, which yield the algebraic sum of their input signals using associated signs. For example, the plant can be a furnace or air conditioning system, where the output variable is temperature. The controller in a heating system consists of fuel valves and the electrical system that operates the valves.

The distinguishing characteristic of an open-loop system is that it cannot compensate for any disturbances that add to the controller's driving signal (Disturbance 1 in Figure 1.6(*a*)). For example, if the controller is an electronic amplifier and Disturbance 1 is noise, then any additive amplifier noise at the first summing junction will also drive the process, corrupting the output with the effect of the noise. The output of an open-loop system is corrupted not only by signals that add to the controller's commands but also by disturbances at the output (Disturbance 2 in Figure 1.6(*a*)). The system cannot correct for these disturbances, either.

Open-loop systems, then, do not correct for disturbances and are simply commanded by the input. For example, toasters are open-loop systems, as anyone with burnt toast can attest. The controlled variable (output) of a toaster is the color of the toast. The device is designed with the assumption that the toast will be darker the longer it is subjected to heat. The toaster does not measure the color of the toast; it does not correct for the fact that the toast is rye, white, or sourdough, nor does it correct for the fact that toast comes in different thicknesses.

Other examples of open-loop systems are mechanical systems consisting of a mass, spring, and damper with a constant force positioning the mass. The greater the force, the greater the displacement. Again, the system position will change with a disturbance, such as an additional force, and the system will not detect or correct for the disturbance. Or assume that you calculate the amount of time you need to study

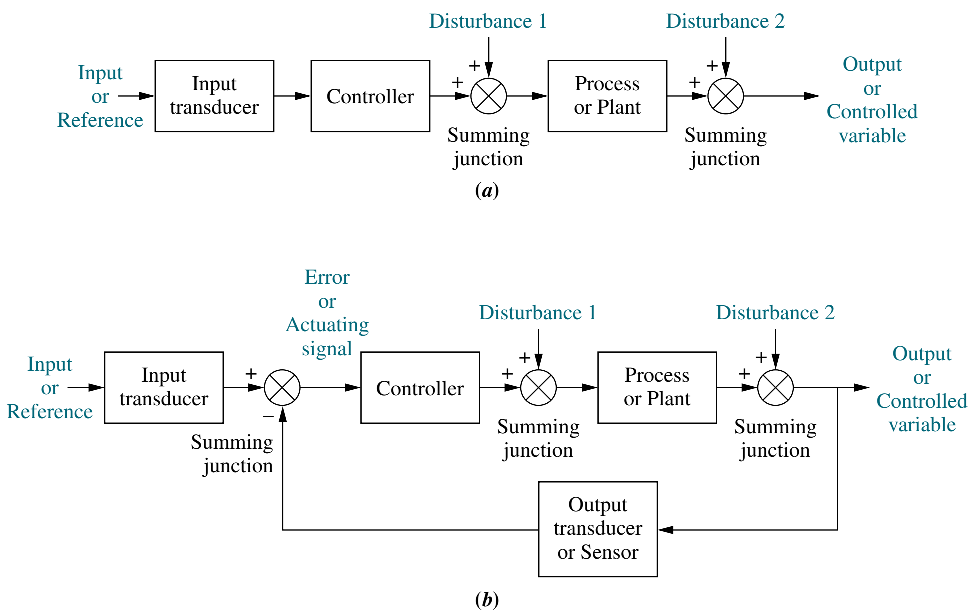


FIGURE 1.6 Block diagrams of control systems: **a.** open-loop system; **b.** closed-loop system

for an examination that covers three chapters in order to get an A. If the professor adds a fourth chapter—a disturbance—you are an open-loop system if you do not detect the disturbance and add study time to that previously calculated. The result of this oversight would be a lower grade than you expected.

Closed-Loop (Feedback Control) Systems

The disadvantages of open-loop systems, namely sensitivity to disturbances and inability to correct for these disturbances, may be overcome in *closed-loop systems*. The generic architecture of a closed-loop system is shown in Figure 1.6(*b*).

The input transducer converts the form of the input to the form used by the controller. An *output transducer*, or *sensor*, measures the output response and converts it into the form used by the controller. For example, if the controller uses electrical signals to operate the valves of a temperature control system, the input position and the output temperature are converted to electrical signals. The input position can be converted to a voltage by a *potentiometer*, a variable resistor, and the output temperature can be converted to a voltage by a *thermistor*, a device whose electrical resistance changes with temperature.

The first summing junction algebraically adds the signal from the input to the signal from the output, which arrives via the *feedback path*, the return path from the output to the summing junction. In Figure 1.6(*b*), the output signal is subtracted from the input signal. The result is generally called the *actuating signal*. However, in systems where both the input and output transducers have *unity gain* (that is, the transducer amplifies its input by 1), the actuating signal's value is equal to the actual difference between the input and the output. Under this condition, the actuating signal is called the *error*.

The closed-loop system compensates for disturbances by measuring the output response, feeding that measurement back through a feedback path, and comparing that response to the input at the summing junction. If there is any difference between the two responses, the system drives the plant, via the actuating signal, to make a correction. If there is no difference, the system does not drive the plant, since the plant's response is already the desired response.

Closed-loop systems, then, have the obvious advantage of greater accuracy than open-loop systems. They are less sensitive to noise, disturbances, and changes in the environment. Transient response and steady-state error can be controlled more conveniently and with greater flexibility in closed-loop systems, often by a simple adjustment of gain (amplification) in the loop and sometimes by redesigning the controller. We refer to the redesign as *compensating* the system and to the resulting hardware as a *compensator*. On the other hand, closed-loop systems are more complex and expensive than open-loop systems. A standard, open-loop toaster serves as an example: It is simple and inexpensive. A closed-loop toaster oven is more complex and more expensive since it has to measure both color (through light reflectivity) and humidity inside the toaster oven. Thus, the control systems engineer must consider the trade-off between the simplicity and low cost of an open-loop system and the accuracy and higher cost of a closed-loop system.

In summary, systems that perform the previously described measurement and correction are called closed-loop, or feedback control, systems. Systems that do not have this property of measurement and correction are called open-loop systems.

Computer-Controlled Systems

In many modern systems, the controller (or compensator) is a digital computer. The advantage of using a computer is that many loops can be controlled or compensated by the same computer through time sharing. Furthermore, any adjustments of the compensator parameters required to yield a desired response can be made by changes in software rather than hardware. The computer can also perform supervisory functions, such as scheduling many required applications. For example, the space shuttle main engine (SSME) controller, which contains two digital computers, alone controls numerous engine functions. It monitors engine sensors that provide pressures, temperatures, flow rates, turbopump speed, valve positions, and engine servo valve actuator positions. The controller further provides closed-loop control of thrust and propellant mixture ratio, sensor excitation, valve actuators, spark igniters, as well as other functions (*Rockwell International*, *1984*).

1.4 Analysis and Design Objectives

In Section 1.1 we briefly alluded to some control system performance specifications, such as transient response and steady-state error. We now expand upon the topic of performance and place it in perspective as we define our analysis and design objectives.

Analysis is the process by which a system's performance is determined. For example, we evaluate its transient response and steady-state error to determine if they meet the desired specifications. *Design* is the process by which a system's performance is created or changed. For example, if a system's transient response and steady-state error are analyzed and found not to meet the specifications, then we change parameters or add additional components to meet the specifications.

A control system is *dynamic:* It responds to an input by undergoing a transient response before reaching a steady-state response that generally resembles the input. We have already identified these two responses and cited a position control system (an elevator) as an example. In this section, we discuss three major objectives of systems analysis and design: producing the desired transient response, reducing steady-state error, and achieving stability. We also address some other design concerns, such as cost and the sensitivity of system performance to changes in parameters.

Transient Response

Transient response is important. In the case of an elevator, a slow transient response makes passengers impatient, whereas an excessively rapid response makes them uncomfortable. If the elevator oscillates about the arrival floor for more than a second, a disconcerting feeling can result. Transient response is also important for structural reasons: Too fast a transient response could cause permanent physical damage. In a computer, transient response contributes to the time required to read from or write to the computer's disk storage (see Figure 1.7). Since reading and writing cannot take place until the head stops, the speed of the read/write head's movement from one track on the disk to another influences the overall speed of the computer.

FIGURE 1.7 Computer hard disk drive, showing disks and read/write head

In this book, we establish quantitative definitions for transient response. We then analyze the system for its *existing* transient response. Finally, we adjust parameters or design components to yield a *desired* transient response—our first analysis and design objective.

Steady-State Response

Another analysis and design goal focuses on the steady-state response. As we have seen, this response resembles the input and is usually what remains after the transients have decayed to zero. For example, this response may be an elevator stopped near the fourth floor or the head of a disk drive finally stopped at the correct track. We are concerned about the accuracy of the steady-state response. An elevator must be level enough with the floor for the passengers to exit, and a read/write head not positioned over the commanded track results in computer errors. An antenna tracking a satellite must keep the satellite well within its beamwidth in order not to lose track. In this text we define steady-state errors quantitatively, analyze a system's steady-state error, and then design corrective action to reduce the steady-state error—our second analysis and design objective.

Stability

Discussion of transient response and steady-state error is moot if the system does not have *stability*. In order to explain stability, we start from the fact that the total response of a system is the sum of the *natural response* and the *forced response*. When you studied linear differential equations, you probably referred to these responses as the *homogeneous* and the *particular solutions*, respectively. Natural response describes the way the system dissipates or acquires energy. The form or nature of this response is dependent only on the system, not the input. On the other hand, the form or nature of the forced response is dependent on the input. Thus, for a *linear* system, we can write

$$\text{Total response} = \text{Natural response} + \text{Forced response} \qquad (1.1)^{2}$$

For a control system to be useful, the natural response must (1) eventually approach zero, thus leaving only the forced response, or (2) oscillate. In some systems, however, the natural response grows without bound rather than diminish to zero or oscillate. Eventually, the natural response is so much greater than the forced response that the system is no longer controlled. This condition, called *instability*, could lead to self-destruction of the physical device if limit stops are not part of the design. For example, the elevator would crash through the floor or exit through the ceiling; an aircraft would go into an uncontrollable roll; or an antenna commanded to point to a target would rotate, line up with the target, but then begin to oscillate about the target with *growing* oscillations and *increasing* velocity until the motor or amplifiers reached their output limits or until the antenna was damaged structurally. A time plot of an unstable system would show a transient response that grows without bound and without any evidence of a steady-state response.

Control systems must be designed to be stable. That is, their natural response must decay to zero as time approaches infinity, or oscillate. In many systems the transient response you see on a time response plot can be directly related to the natural response. Thus, if the natural response decays to zero as time approaches infinity, the transient response will also die out, leaving only the forced response. If the system is stable, the proper transient response and steady-state error characteristics can be designed. Stability is our third analysis and design objective.

---

[2] You may be confused by the words *transient* vs. *natural*, and *steady-state* vs. *forced*. If you look at Figure 1.2, you can see the transient and steady-state portions of the total response as indicated. The transient response is the sum of the natural and forced responses, while the natural response is large. If we plotted the natural response by itself, we would get a curve that is different from the transient portion of Figure 1.2. The steady-state response of Figure 1.2 is also the sum of the natural and forced responses, but the natural response is small. Thus, the transient and steady-state responses are what you actually see on the plot; the natural and forced responses are the underlying mathematical components of those responses.

Other Considerations

The three main objectives of control system analysis and design have already been enumerated. However, other important considerations must be taken into account. For example, factors affecting hardware selection, such as motor sizing to fulfill power requirements and choice of sensors for accuracy, must be considered early in the design.

Finances are another consideration. Control system designers cannot create designs without considering their economic impact. Such considerations as budget allocations and competitive pricing must guide the engineer. For example, if your product is one of a kind, you may be able to create a design that uses more expensive components without appreciably increasing total cost. However, if your design will be used for many copies, slight increases in cost per copy can translate into many more dollars for your company to propose during contract bidding and to outlay before sales.

Another consideration is *robust* design. System parameters considered constant during the design for transient response, steady-state errors, and stability change over time when the actual system is built. Thus, the performance of the system also changes over time and will not be consistent with your design. Unfortunately, the relationship between parameter changes and their effect on performance is not linear. In some cases, even in the same system, changes in parameter values can lead to small or large changes in performance, depending on the system's nominal operating point and the type of design used. Thus, the engineer wants to create a robust design so that the system will not be sensitive to parameter changes. We discuss the concept of system sensitivity to parameter changes in Chapters 7 and 8. This concept, then, can be used to test a design for robustness.

Case Study

Introduction to a Case Study

Now that our objectives are stated, how do we meet them? In this section we will look at an example of a feedback control system. The system introduced here will be used in subsequent chapters as a running case study to demonstrate the objectives of those chapters. A colored background like this will identify the case study section at the end of each chapter. Section 1.5, which follows this first case study, explores the design process that will help us build our system.

FIGURE 1.8 The search for extraterrestrial life is being carried out with radio antennas like the one pictured here. A radio antenna is an example of a system with position controls.

Antenna Azimuth: An Introduction to Position Control Systems

A position control system converts a position input command to a position output response. Position control systems find widespread applications in antennas, robot arms, and computer disk drives. The radio telescope antenna in Figure 1.8 is one example of a system that uses position control systems. In this section, we will look in detail at an antenna azimuth position control system that could be used to position a radio telescope antenna. We will see how the system works and how we can effect changes in its performance. The discussion here will be on a qualitative level, with the objective of getting an intuitive feeling for the systems with which we will be dealing.

An antenna azimuth position control system is shown in Figure 1.9(a), with a more detailed layout and schematic in Figures 1.9(b) and 1.9(c), respectively. Figure 1.9(d) shows *a functional block diagram* of the system. The functions are shown above the blocks, and the required hardware is indicated inside the blocks. Parts of Figure 1.9 are repeated on the front endpapers for future reference.

The purpose of this system is to have the azimuth angle output of the antenna, $\theta_o(t)$, follow the input angle of the potentiometer, $\theta_i(t)$. Let us look at Figure 1.9(d) and describe how this system works. The input command is an angular displacement. The potentiometer converts the angular displacement into a voltage.

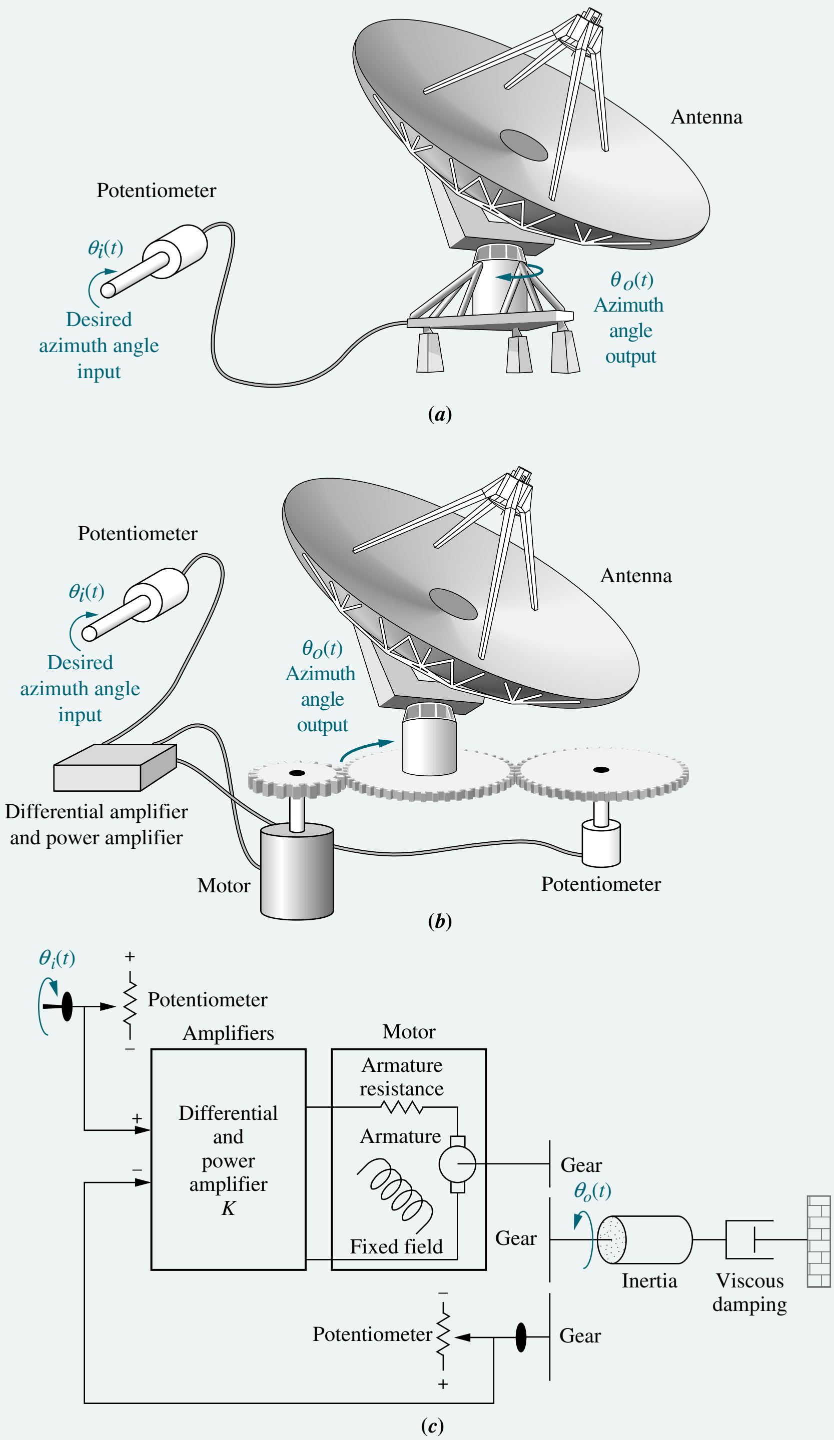
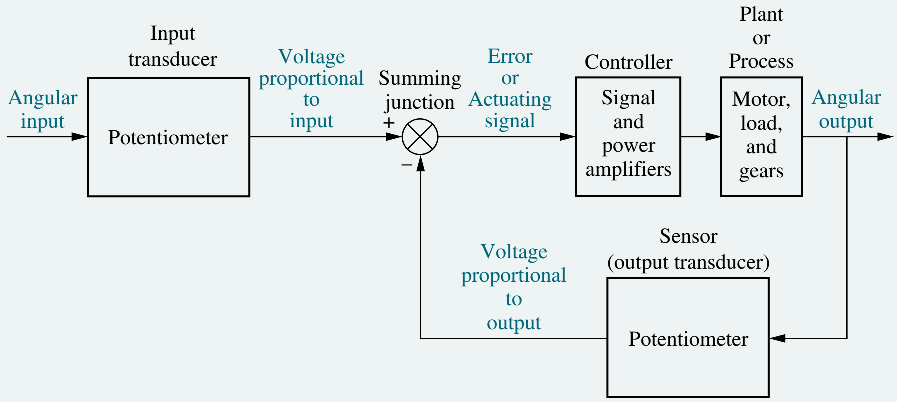


FIGURE 1.9 Antenna azimuth position control system: **a.** system concept; **b.** detailed layout; **c.** schematic; (*figure continues*)



(*d*)

FIGURE 1.9 (*Continued*) **d.** functional block diagram

Similarly, the output angular displacement is converted to a voltage by the potentiometer in the feedback path. The signal and power amplifiers boost the difference between the input and output voltages. This amplified actuating signal drives the plant.

The system normally operates to drive the error to zero. When the input and output match, the error will be zero, and the motor will not turn. Thus, the motor is driven only when the output and the input do not match. The greater the difference between the input and the output, the larger the motor input voltage, and the faster the motor will turn.

If we increase the gain of the signal amplifier, will there be an increase in the steady-state value of the output? If the gain is increased, then for a given actuating signal, the motor will be driven harder. However, the motor will still stop when the actuating signal reaches zero, that is, when the output matches the input. The difference in the response, however, will be in the transients. Since the motor is driven harder, it turns faster toward its final position. Also, because of the increased speed, increased momentum could cause the motor to overshoot the final value and be forced by the system to return to the commanded position. Thus, the possibility exists for a transient response that consists of *damped oscillations* (that is, a sinusoidal response whose amplitude diminishes with time) about the steady-state value if the gain is high. The responses for low gain and high gain are shown in Figure 1.10.

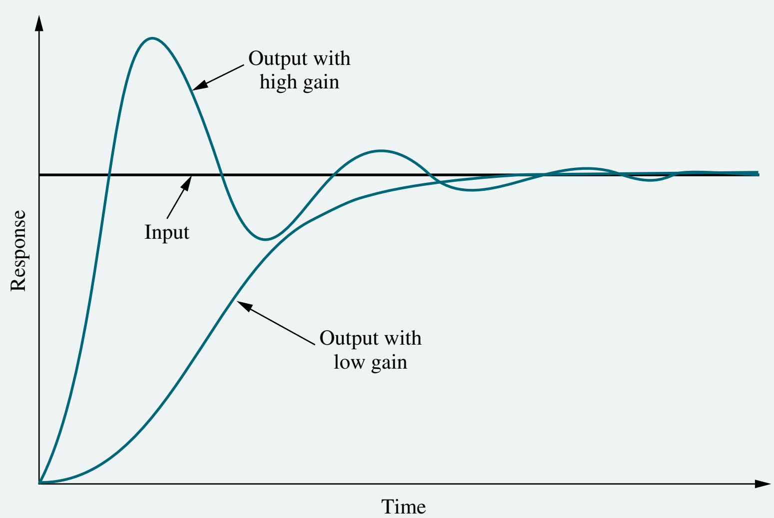


FIGURE 1.10 Response of a position control system, showing effect of high and low controller gain on the output response

We have discussed the transient response of the position control system. Let us now direct our attention to the steady-state position to see how closely the output matches the input after the transients disappear.

We define steady-state error as the difference between the input and the output after the transients have effectively disappeared. The definition holds equally well for step, ramp, and other types of inputs. Typically, the steady-state error decreases with an increase in gain and increases with a decrease in gain. Figure 1.10 shows zero error in the steady-state response; that is, after the transients have disappeared, the output position equals the commanded input position. In some systems, the steady-state error will not be zero; for these systems, a simple gain adjustment to regulate the transient response is either not effective or leads to a trade-off between the desired transient response and the desired steady-state accuracy.

To solve this problem, a controller with a dynamic response, such as an electrical filter, is used along with an amplifier. With this type of controller, it is possible to design both the required transient response and the required steady-state accuracy without the trade-off required by a simple setting of gain. However, the controller is now more complex. The filter in this case is called a compensator. Many systems also use dynamic elements in the feedback path along with the output transducer to improve system performance.

In summary, then, our design objectives and the system's performance revolve around the transient response, the steady-state error, and stability. Gain adjustments can affect performance and sometimes lead to trade-offs between the performance criteria. Compensators can often be designed to achieve performance specifications without the need for trade-offs. Now that we have stated our objectives and some of the methods available to meet those objectives, we describe the orderly progression that leads us to the final system design.

1.5 The Design Process

In this section, we establish an orderly sequence for the design of feedback control systems that will be followed as we progress through the rest of the book. Figure 1.11 shows the described process as well as the chapters in which the steps are discussed.

The antenna azimuth position control system discussed in the last section is representative of control systems that must be analyzed and designed. Inherent in

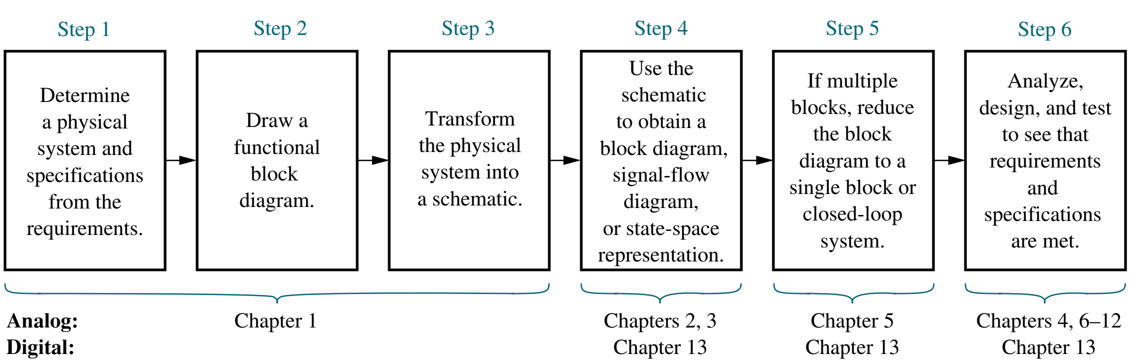


FIGURE 1.11 The control system design process

Figure 1.11 is feedback and communication during each phase. For example, if testing (Step 6) shows that requirements have not been met, the system must be redesigned and retested. Sometimes requirements are conflicting and the design cannot be attained. In these cases, the requirements have to be respecified and the design process repeated. Let us now elaborate on each block of Figure 1.11.

Step 1: Transform Requirements Into a Physical System

We begin by transforming the requirements into a physical system. For example, in the antenna azimuth position control system, the requirements would state the desire to position the antenna from a remote location and describe such features as weight and physical dimensions. Using the requirements, design specifications, such as desired transient response and steady-state accuracy, are determined. Perhaps an overall concept, such as Figure 1.9(*a*), would result.

Step 2: Draw a Functional Block Diagram

The designer now translates a qualitative description of the system into a functional block diagram that describes the component parts of the system (that is, function and/or hardware) and shows their interconnection. Figure 1.9(*d*) is an example of a functional block diagram for the antenna azimuth position control system. It indicates functions such as input transducer and controller, as well as possible hardware descriptions such as amplifiers and motors. At this point the designer may produce a detailed layout of the system, such as that shown in Figure 1.9(*b*), from which the next phase of the analysis and design sequence, developing a schematic diagram, can be launched.

Step 3: Create a Schematic

As we have seen, position control systems consist of electrical, mechanical, and electromechanical components. After producing the description of a physical system, the control systems engineer transforms the physical system into a schematic diagram. The control system designer can begin with the physical description, as contained in Figure 1.9(*a*), to derive a schematic. The engineer must make approximations about the system and neglect certain phenomena, or else the schematic will be unwieldy, making it difficult to extract a useful mathematical model during the next phase of the analysis and design sequence. The designer starts with a simple schematic representation and, at subsequent phases of the analysis and design sequence, checks the assumptions made about the physical system through analysis and computer simulation. If the schematic is too simple and does not adequately account for observed behavior, the control systems engineer adds phenomena to the schematic that were previously assumed negligible. A schematic diagram for the antenna azimuth position control system is shown in Figure 1.9(*c*).

When we draw the potentiometers, we make our first simplifying assumption by neglecting their friction or inertia. These mechanical characteristics yield a dynamic, rather than an instantaneous, response in the output voltage. We assume that these mechanical effects are negligible and that the voltage across a potentiometer changes instantaneously as the potentiometer shaft turns.

A differential amplifier and a power amplifier are used as the controller to yield gain and power amplification, respectively, to drive the motor. Again, we assume that the dynamics of the amplifiers are rapid compared to the response time of the motor; thus, we model them as a pure gain, K.

A dc motor and equivalent load produce the output angular displacement. The speed of the motor is proportional to the voltage applied to the motor's *armature circuit*. Both inductance and resistance are part of the armature circuit. In showing just the armature resistance in Figure 1.9(c), we assume the effect of the armature inductance is negligible for a dc motor.

The designer makes further assumptions about the load. The load consists of a rotating mass and bearing friction. Thus, the model consists of *inertia* and *viscous damping* whose resistive torque increases with speed, as in an automobile's shock absorber or a screen door damper.

The decisions made in developing the schematic stem from knowledge of the physical system, the physical laws governing the system's behavior, and *practical experience*. These decisions are not easy; however, as you acquire more design experience, you will gain the insight required for this difficult task.

Step 4: Develop a Mathematical Model (Block Diagram)

Once the schematic is drawn, the designer uses physical laws, such as Kirchhoff's laws for electrical networks and Newton's law for mechanical systems, along with simplifying assumptions, to model the system mathematically. These laws are

Kirchhoff's voltage law	The sum of voltages around a closed path equals zero.
Kirchhoff's current law	The sum of electric currents flowing from a node equals zero.
Newton's laws	The sum of forces on a body equals zero;[3] the sum of moments on a body equals zero.

Kirchhoff's and Newton's laws lead to mathematical models that describe the relationship between the input and output of dynamic systems. One such model is the *linear, time-invariant differential equation*, Eq. (1.2):

$$\boxed{\frac{d^m c(t)}{dt^n} + d_{n-1}\frac{d^{m-1}c(t)}{dt^{n-1}} + \cdots + d_0 c(t) = b_m \frac{d^m r(t)}{dt^m} + b_{m-1}\frac{d^{m-1}r(t)}{dt^{m-1}} + \cdots + b_0 r(t)} \qquad (1.2)^4$$

Many systems can be approximately described by this equation, which relates the output, $c(t)$, to the input, $r(t)$, by way of the system parameters, a_i and b_j. We assume the reader is familiar with differential equations. Problems and a bibliography are provided at the end of the chapter for you to review this subject.

Simplifying assumptions made in the process of obtaining a mathematical model usually leads to a low-order form of Eq. (1.2). Without the assumptions the system model could be of high order or described with nonlinear, time-varying, or partial differential equations. These equations complicate the design process and reduce the designer's insight. Of course, all assumptions must be checked and all simplifications justified through analysis or testing. If the assumptions for simplification cannot be justified, then the model cannot be simplified. We examine some of these simplifying assumptions in Chapter 2.

In addition to the differential equation, the *transfer function* is another way of mathematically modeling a system. The model is derived from the linear, time-invariant differential equation using what we call the *Laplace transform*. Although the transfer

---

[3] Alternately, $\sum$ forces $= Ma$. In this text the force, Ma, will be brought to the left-hand side of the equation to yield $\sum$ forces $= 0$ (D'Alembert's principle). We can then have a consistent analogy between force and voltage, and Kirchhoff's and Newton's laws (that is, $\sum$ forces $= 0$; $\sum$ voltages $= 0$).

[4] The right-hand side of Eq. (1.2) indicates differentiation of the input, $r(t)$. In physical systems, differentiation of the input introduces noise. In Chapters 3 and 5 we show implementations and interpretations of Eq. (1.2) that do not require differentiation of the input.

function can be used only for linear systems, it yields more intuitive information than the differential equation. We will be able to change system parameters and rapidly sense the effect of these changes on the system response. The transfer function is also useful in modeling the interconnection of subsystems by forming a block diagram similar to Figure 1.9(d) but with a mathematical function inside each block.

Still another model is the *state-space representation*. One advantage of state-space methods is that they can also be used for systems that cannot be described by linear differential equations. Further, state-space methods are used to model systems for simulation on the digital computer. Basically, this representation turns an nth-order differential equation into n simultaneous first-order differential equations. Let this description suffice for now; we describe this approach in more detail in Chapter 3.

Finally, we should mention that to produce the mathematical model for a system, we require knowledge of the parameter values, such as equivalent resistance, inductance, mass, and damping, which is often not easy to obtain. Analysis, measurements, or specifications from vendors are sources that the control systems engineer may use to obtain the parameters.

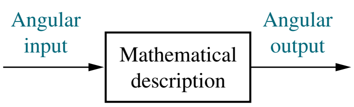


FIGURE 1.12 Equivalent block diagram for the antenna azimuth position control system

Step 5: Reduce the Block Diagram

Subsystem models are interconnected to form block diagrams of larger systems, as in Figure 1.9(d), where each block has a mathematical description. Notice that many signals, such as proportional voltages and error, are internal to the system. There are also two signals—angular input and angular output—that are external to the system. In order to evaluate system response in this example, we need to reduce this large system's block diagram to a single block with a mathematical description that represents the system from its input to its output, as shown in Figure 1.12. Once the block diagram is reduced, we are ready to analyze and design the system.

Step 6: Analyze and Design

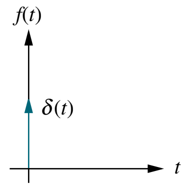
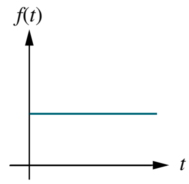
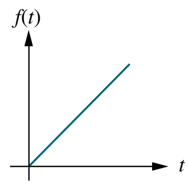
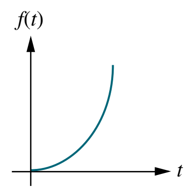
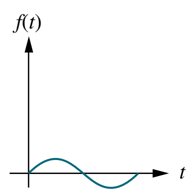
The next phase of the process, following block diagram reduction, is analysis and design. If you are interested only in the performance of an individual subsystem, you can skip the block diagram reduction and move immediately into analysis and design. In this phase, the engineer analyzes the system to see if the response specifications and performance requirements can be met by simple adjustments of system parameters. If specifications cannot be met, the designer then designs additional hardware in order to effect a desired performance.

Test input signals are used, both analytically and during testing, to verify the design. It is neither necessarily practical nor illuminating to choose complicated input signals to analyze a system's performance. Thus, the engineer usually selects standard test inputs. These inputs are impulses, steps, ramps, parabolas, and sinusoids, as shown in Table 1.1.

An *impulse* is infinite at $t = 0$ and zero elsewhere. The area under the unit impulse is 1. An approximation of this type of waveform is used to place initial energy into a system so that the response due to that initial energy is only the transient response of a system. From this response the designer can derive a mathematical model of the system.

A *step* input represents a *constant command*, such as position, velocity, or acceleration. Typically, the step input command is of the same form as the output. For example, if the system's output is position, as it is for the antenna azimuth position control system, the step input represents a desired position, and the output represents the actual position. If the system's output is velocity, as is the spindle speed for a video disc player, the step input represents a constant desired speed, and the output represents the actual speed. The designer uses step inputs because both the transient response and the steady-state response are clearly visible and can be evaluated.

TABLE 1.1 Test waveforms used in control systems

Input	Function	Description	Sketch	Use
Impulse	$\delta(t)$	$\delta(t) = \infty$ for $0- < t < 0+$ $= 0$ elsewhere $\int_{0-}^{0+} \delta(t)dt = 1$		Transient response Modeling
Step	$u(t)$	$u(t) = 1$ for $t > 0$ $= 0$ for $t < 0$		Transient response Steady-state error
Ramp	$tu(t)$	$tu(t) = t$ for $t \geq 0$ $= 0$ elsewhere		Steady-state error
Parabola	$\frac{1}{2}t^2u(t)$	$\frac{1}{2}t^2u(t) = \frac{1}{2}t^2$ for $t \geq 0$ $= 0$ elsewhere		Steady-state error
Sinusoid	$\sin \omega t$			Transient response Modeling Steady-state error

The *ramp* input represents a *linearly increasing command*. For example, if the system's output is position, the input ramp represents a linearly increasing position, such as that found when tracking a satellite moving across the sky at constant speed. If the system's output is velocity, the input ramp represents a linearly increasing velocity. The response to an input ramp test signal yields additional information about the steady-state error. The previous discussion can be extended to *parabolic* inputs, which are also used to evaluate a system's steady-state error.

Sinusoidal inputs can also be used to test a physical system to arrive at a mathematical model. We discuss the use of this waveform in detail in Chapters 10 and 11.

We conclude that one of the basic analysis and design requirements is to evaluate the time response of a system for a given input. Throughout the book you will learn numerous methods for accomplishing this goal.

The control systems engineer must take into consideration other characteristics about feedback control systems. For example, control system behavior is altered by fluctuations in component values or system parameters. These variations can be caused by temperature, pressure, or other environmental changes. Systems must be built so that expected fluctuations do not degrade performance beyond specified bounds. A *sensitivity* analysis can yield the percentage of change in a specification as a function of a change in a system parameter. One of the designer's goals, then, is to build a system with minimum sensitivity over an expected range of environmental changes.

In this section we looked at some control systems analysis and design considerations. We saw that the designer is concerned about transient response, steady-state error, stability, and sensitivity. The text pointed out that although the basis of evaluating system performance is the differential equation, other methods, such as transfer functions and state space, will be used. The advantages of these new techniques over differential equations will become apparent as we discuss them in later chapters.

1.6 Computer-Aided Design

Now that we have discussed the analysis and design sequence, let us discuss the use of the computer as a computational tool in this sequence. The computer plays an important role in the design of modern control systems. In the past, control system design was labor intensive. Many of the tools we use today were implemented through hand calculations or, at best, using plastic graphical aid tools. The process was slow, and the results not always accurate. Large mainframe computers were then used to simulate the designs.

Today we are fortunate to have computers and software that remove the drudgery from the task. At our own desktop computers, we can perform analysis, design, and simulation with one program. With the ability to simulate a design rapidly, we can easily make changes and immediately test a new design. We can play what-if games and try alternate solutions to see if they yield better results, such as reduced sensitivity to parameter changes. We can include nonlinearities and other effects and test our models for accuracy.

MATLAB

The computer is an integral part of modern control system design, and many computational tools are available for your use. In this book we use MATLAB and the MATLAB Control System Toolbox, which expands MATLAB to include control system–specific commands. In addition, presented are several MATLAB enhancements that give added functionality to MATLAB and the Control Systems Toolbox. Included are (1) Simulink, which uses a graphical user interface (GUI); (2) the LTI Viewer, which permits measurements to be made directly from time and frequency response curves; (3) the SISO Design Tool, a convenient and intuitive analysis and design tool; and (4) the Symbolic Math Toolbox, which saves labor when making symbolic calculations required in control system analysis and design. Some of these enhancements may require additional software available from The MathWorks, Inc.

MATLAB is presented as an alternate method of solving control system problems. You are encouraged to solve problems first by hand and then by MATLAB so that insight is not lost through mechanized use of computer programs. To this end, many examples throughout the book are solved by hand, followed by suggested use of MATLAB.

As an enticement to begin using MATLAB, simple program statements that you can try are suggested throughout the chapters at appropriate locations. Throughout the book, various icons appear in the margin to identify MATLAB references that direct you to the proper program in the proper appendix and tell you what you will learn. Selected end-of-chapter problems and Case Study Challenges to be solved using MATLAB have also been marked with appropriate icons. The following list itemizes the specific components of MATLAB used in this book, the icon used to identify each, and the appendix in which a description can be found:

MATLAB/Control System Toolbox tutorials and code are found in Appendix B and identified in the text with the MATLAB icon shown in the margin.

MATLAB
ML

Simulink tutorials and diagrams are found in Appendix C and identified in the text with the Simulink icon shown in the margin.

Simulink
SL

MATLAB GUI tools, tutorials, and examples are in Appendix E at www.wiley.com/college/nise and identified in the text with the GUI Tool icon shown in the margin. These tools consist of the LTI Viewer and the SISO Design Tool.

Gui Tool
GUIT

Symbolic Math Toolbox tutorials and code are found in Appendix F at www.wiley.com/college/nise and identified in the text with the Symbolic Math icon shown in the margin.

Symbolic Math
SM

MATLAB code itself is not platform specific. The same code runs on PCs and workstations that support MATLAB. Although there are differences in installing and managing MATLAB files, we do not address them in this book. Also, there are many more commands in MATLAB and the MATLAB toolboxes than are covered in the appendixes. Please explore the bibliographies at the end of the applicable appendixes to find out more about MATLAB file management and MATLAB instructions not covered in this textbook.

LabVIEW

LabVIEW is a programming environment presented as an alternative to MATLAB. This graphical alternative produces front panels of virtual instruments on your computer that are pictorial reproductions of hardware instruments, such as waveform generators or oscilloscopes. Underlying the front panels are block diagrams. The blocks contain underlying code for the controls and indicators on the front panel. Thus, a knowledge of coding is not required. Also, parameters can be easily passed or viewed from the front panel.

A LabVIEW tutorial is in Appendix D and all LabVIEW material is identified with the LabVIEW icon shown in the margin.

LabVIEW
LV

You are encouraged to use computational aids throughout this book. Those not using MATLAB or LabVIEW should consult Appendix H at www.wiley.com/college/nise for a discussion of other alternatives. Now that we have introduced control systems to you and established a need for computational aids to perform analysis and design, we will conclude with a discussion of your career as a control systems engineer and look at the opportunities and challenges that await you.

1.7 The Control Systems Engineer

Control systems engineering is an exciting field in which to apply your engineering talents, because it cuts across numerous disciplines and numerous functions within those disciplines. The control engineer can be found at the top level of large projects, engaged at the conceptual phase in determining or implementing