1 Introduction to Control Systems

1.1 CONTROL SYSTEMS

Introduction and Background

In a modern **control system**, electronic intelligence controls some physical process. Control systems are the "automatic" in such things as automatic pilot and automatic washer. Because the machine itself is making the routine decisions, the human operator is freed to do other things. In many cases, machine intelligence is better than direct human control because it can react faster or slower (keep track of long-term slow changes), respond more precisely, and maintain an accurate log of the system's performance. Control systems can be classified in several ways.

A **regulator system** automatically maintains a parameter at (or near) a specified value. An example of this is a home heating system maintaining a set temperature despite changing outside conditions.

A **follow-up system** causes an output to follow a set path that has been specified in advance. An example is an industrial robot moving parts from place to place.

An **event control system** controls a sequential series of events. An example is a washing machine cycling through a series of programmed steps.

The subject of control systems is really many subjects: electronics (both analog and digital), power-control devices, sensors, motors, mechanics, and *control system theory*, which ties together all these concepts. Many students find the subject of control systems to be interesting because it deals with applications of much of the theory to which they have already been exposed.

Every control system has (at least) a **controller** and an **actuator** (also called a final control element). Shown in the block diagram in Figure 1-1, the controller is the intelligence of the system and is usually electronic. The input to the controller is called the **set point,** which is a signal representing the desired system output. The actuator is an electromechanical device that takes the signal from the controller and converts it into some kind of physical action. Examples of typical actuators would be an electric motor, an electrically controlled valve, or a heating element. The last block in Figure 1-1 is labeled **process** and has an output labeled **controlled variable.** The process block represents the physical process being affected by the actuator, and the controlled variable is the measurable result of that process. For example, if the actuator is an electric heating element in a furnace, then the process is "heating the furnace," and the controlled variable is the temperature in the furnace. If the actuator is an electric motor that rotates an antenna, then the process is "rotating of the antenna," and the controlled variable is the angular position of the antenna.

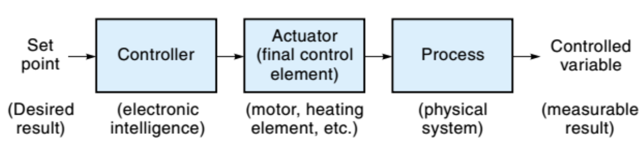

Figure 1-1 A block diagram of a control system.

1.1.1 Open-Loop Control Systems

Control systems can be broadly divided into two categories: open- and closed-loop systems. In an **open-loop control system,** the controller independently calculates exact voltage or current needed by the actuator to do the job and sends it. With this approach, however, *the controller never actually knows if the actuator did what it was supposed to* because there is no feedback. This system absolutely depends on the controller knowing the operating characteristics of the actuator. Open-loop control systems are appropriate in applications where the actions of the actuator on the process are very repeatable and reliable. Relays and stepper motors are devices with reliable characteristics and are usually open-loop operations. Actuators such as motors or flow valves are sometimes used in open-loop operations, but they must be calibrated and adjusted at regular intervals to ensure proper system operation.

Example 1-1

> **EXAMPLE 1.1**
>
> Figure 1.2 shows an open-loop control system. The actuator is a motor driving a robot arm. In this case, the process is the arm moving, and the controlled variable is the angular position of the arm. Earlier tests have shown that the motor rotates the arm at 5 degrees/second (deg/s) at the rated voltage. Assume that the controller is directed to move the arm from 0° to 30°. Knowing the characteristics of the process, the controller sends a 6-second power pulse to the motor. If the motor is acting properly, it will rotate exactly 30° in the 6 seconds and stop. On particularly cold days, however, the lubricant is more viscous (thicker), causing more internal friction, and the motor rotates only 25° in the 6 seconds; the result is a 5° error. The controller has no way of knowing of the error and does nothing to correct it.

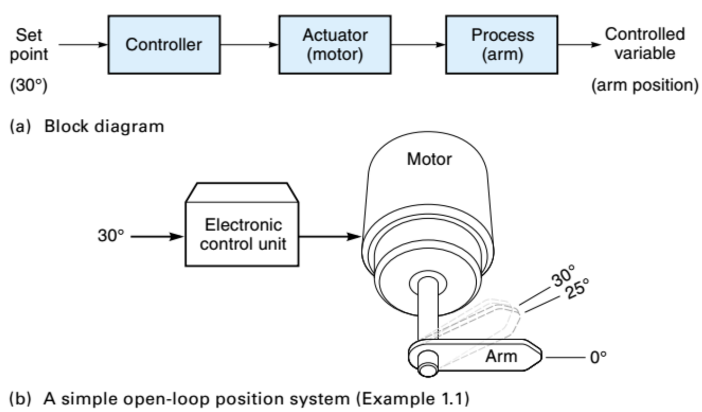

Figure 1-2 Open-loop control system.

1.1.2 Closed-Loop Control Systems

In a **closed-loop control system,** the output of the process (controlled variable) is constantly monitored by a **sensor,** as shown in Figure 1-3(a). The sensor samples the system output and converts this measurement into an electric signal that it passes back to the controller. Because the controller knows what the system is actually doing, it can make any adjustments necessary to keep the output where it belongs. The signal from the controller to the actuator is the **forward path**, and the signal from the sensor to the controller is the **feedback** (which "closes" the loop). In Figure 1-3(a), the feedback signal is subtracted from the set point at the **comparator** (just ahead of the controller). By subtracting the actual position (as reported by the sensor) from the desired position (as defined by the set point), we get the system **error.** The error signal represents the difference between "where you are" and "where you want to be." *The controller is always working to minimize this error signal.* A zero error means that the output is exactly what the set point says it should be.

Using a **control strategy,** which can be simple or complex, the controller minimizes the error. A simple control strategy would enable the controller to turn the actuator on or off—for example, a thermostat cycling a furnace on and off to maintain a certain temperature. A more complex control strategy would let the controller adjust the actuator force to meet the demand of the load, as described in Example 1.2

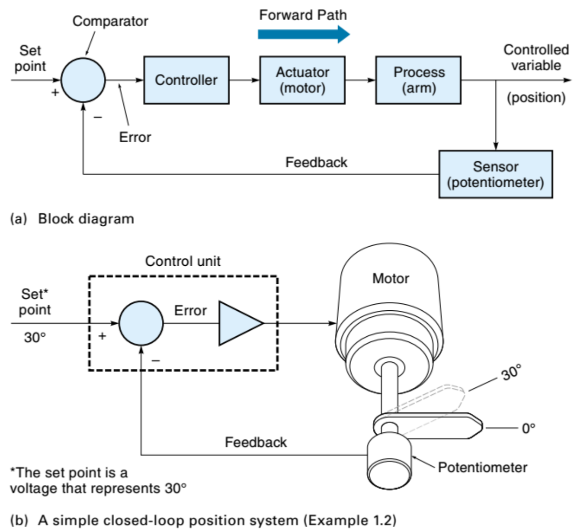

Figure 1-3 Closed-loop control system.

EXAMPLE 1.2

As an example of a closed-loop control system, consider again the robot arm resting at 0° [see Figure 1.3(b)]. This time a potentiometer (pot) has been connected directly to the motor shaft. As the shaft turns, the pot resistance changes. The resistance is converted to voltage and then fed back to the controller.

To command the arm to 30°, a set-point voltage corresponding to 30° is sent to the controller. Because the actual arm is still resting at 0°, the error signal "jumps up" to 30°. Immediately, the controller starts to drive the motor in a direction to reduce the error. As the arm approaches 30°, the controller slows the motor; when the arm finally reaches 30°, the motor stops. If at some later time, an external force moves the arm off the 30° mark, the error signal would reappear, and the motor would again drive the arm to the 30° position.

The self-correcting feature of closed-loop control makes it preferable over open-loop control in many applications, despite the additional hardware required. This is because closed-loop systems provide reliable, repeatable performance even when the system components themselves (in the forward path) are not absolutely repeatable or precisely known.

1.1.3 Transfer Functions

Physically, a control system is a collection of components and circuits connected together to perform a useful function. *Each component in the system converts energy from one form to another;* for example, we might think of a temperature sensor as converting degrees to volts or a motor as converting volts to revolutions per minute. To describe the performance of the entire control system, we must have some common language so that we can calculate the combined effects of the different components in the system. This need is behind the transfer function concept

A **transfer function** (TF) is a mathematical relationship between the input and output of a control system component. Specifically, the transfer function is defined as the output divided by the input, expressed as

$$\text{TF} = \frac{\text{output}}{\text{input}} \qquad \textbf{1-1}$$

the *steady-state values* for the transfer function, which is sometimes called simply the **gain,** expressed as

$$\text{TF}_{\text{steady-state}} = \text{gain} = \frac{\text{steady-state output}}{\text{steady-state input}} \qquad \textbf{1-2}$$

transfer functions can be used to analyze an entire system of components. One common situation involves a series of components where the output of one component becomes the input to the next and each component has its own transfer function. Figure 1-4(a) shows the block diagram for this situation. This diagram can be reduced into a single block that has a TF_{tot}, which is the product of all the individual transfer functions. This concept is illustrated in Figure 1-4(b) and stated in

$$\text{TF}_{\text{tot}} = \text{system gain} = \text{TF}_1 \times \text{TF}_2 \times \text{TF}_3 \times \ldots \qquad \text{1-3:}$$

$$TF_{tot} = \text{system gain} = TF_1 \times TF_2 \times TF_3 \times \ldots \quad \textbf{1-3}$$

Where TF_{tot} = total steady-state transfer function for the entire (open-loop) system
TF_1, TF_2, . . . = individual transfer functions These concepts are explained in Example 1.5.

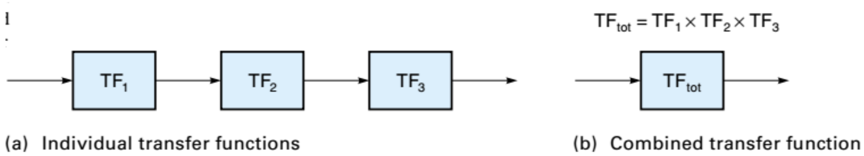

Figure 1-4 A series of transfer functions reduced to a single tranfer function.

EXAMPLE 1.3

A potentiometer is used as a position sensor [see Figure 1.3(b)]. The pot is configured in such a way that 0° of rotation yields 0 V and 300° yields 10 V. Find the transfer function of the pot.

SOLUTION

The transfer function is output divided by input. In this case, the input to the pot is "position in degrees," and output is volts:

$$TF = \frac{\text{output}}{\text{input}} = \frac{10\text{ V}}{300°} = 0.0333\text{ V/deg}$$

The transfer function of a component is an extremely useful number. It allows you to calculate the output of a component if you know the input. The procedure is simply to multiply the transfer function by the input, as shown in Example 1.4.

EXAMPLE 1.4

For a temperature-measuring sensor, the input is temperature, and the output is voltage. The sensor transfer function is given as 0.01 V/deg. Find the sensor-output voltage if the temperature is 600°F.

SOLUTION

If $TF = \frac{\text{output}}{\text{input}}$, then $\quad$ Output = input × TF

$$= \frac{600° \times 0.01\text{ V}}{\text{deg}} = 6\text{ V}$$

EXAMPLE 1.5

Consider the system shown in Figure 1.5. It consists of an electric motor driving a gear train, which is driving a winch. Each component has its own characteristics: The motor (under these conditions) turns at 100 rpm_m for each volt (V_m) supplied; the output shaft of the gear train rotates at one-half of the motor speed; the winch (with a 3-inch shaft circumference) converts the rotary motion (rpm_w) to linear speed. The individual transfer functions are given as follows:

Motor: $$TF_m = \frac{\text{output}}{\text{input}} = \frac{100\ \text{rpm}_m}{1\ V_m} = 100\ \text{rpm}_m/\text{V}$$

Gear train: $$TF_g = \frac{\text{output}}{\text{input}} = \frac{1\ \text{rpm}_w}{2\ \text{rpm}_m} = 0.5\ \text{rpm}_w/\text{rpm}_m$$

Winch: $$TF_w = \frac{\text{output}}{\text{input}} = \frac{3\ \text{in./min}}{1\ \text{rpm}_w} = 3\ \text{in./min/rpm}_w$$

Using Equation 1.3, we can calculate the system transfer function. If everything is correct, all units will cancel except for the desired set:

$$\begin{aligned} TF_{tot} &= TF_m \times TF_g \times TF_w \\ &= \frac{100\ \text{rpm}_m}{1\ V_m} \times \frac{0.5\ \text{rpm}_w}{1\ \text{rpm}_m} \times \frac{3\ \text{in./min}}{1\ \text{rpm}_w} \\ &= 150\ \text{in./min/V}_m \end{aligned}$$

We have shown that the transfer function of the complete system is 150 in./min/V_m. *Knowing this value, we can calculate the system output for any system input.* For example, if the input to the this system is 12 V (to the motor), the output speed of the winch is calculated as follows:

$$\text{Output} = \text{input} \times TF = \frac{12\ V \times 150\ \text{in./min}}{1\ V_m} = 1800\ \text{in./min}$$

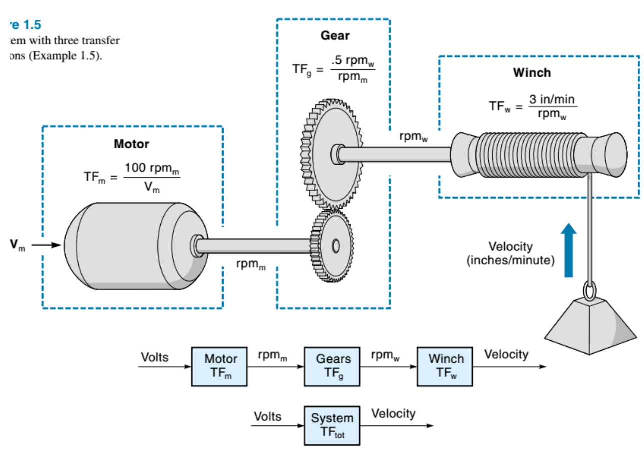

Figure 1-5 A system with three transfer functions (Example 1.5).

1.2 ANALOG AND DIGITAL CONTROL SYSTEMS

In an **analog control system,** the controller consists of traditional analog devices and circuits, that is, linear amplifiers. The first control systems were analog because it was the only available technology. In the analog control system, any change in either set point or feedback is sensed immediately, and the amplifiers adjust their output (to the actuator) accordingly.

In a **digital control system,** the controller uses a digital circuit. In most cases, this circuit is actually a computer, usually microprocessor- or microcontroller-based. The computer executes a program that repeats over-and-over (each repetition is called an **iteration** or **scan**). The program instructs the computer to read the set point and sensor data and then use these numbers to calculate the controller output (which is sent to the actuator). The program then loops back to the beginning and starts over again. The total time for one pass through the program may be less than 1 millisecond (ms). The digital system only "looks" at the inputs at a certain time in the scan and gives the updated output later. If an input changes just after the computer looked at it, that change will remain undetected until the next time through the scan. This is fundamentally different than the analog system, which is continuous and responds immediately to any changes. However, for most digital control systems, the scan time is so short compared with the response time of the process being controlled that, for all practical purposes, the controller response is instantaneous.

The physical world is basically an "analog place." Natural events take time to happen, and they usually move in a continuous fashion from one position to the next. Therefore, most control systems are controlling analog processes. This means that, in many cases, the digital control system must first convert real-world analog input data into digital form before it can be used. Similarly, the output from the digital controller must be converted from digital form back into analog form.

Figure 1-6 shows a block diagram of a digital closed-loop control system. Notice the two additional blocks: the digital-to-analog converter (DAC) and the analog-to-digital converter (ADC). Also note that the feedback line is shown going directly into the controller. This emphasizes the fact that the computer, not a separate subtraction circuit, makes the comparison between the set point and the feedback signal.

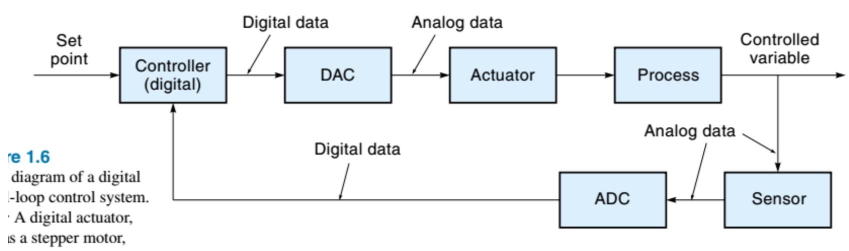

Figure 1-6 Block diagram of a digital closed-loop control system.

1.3 CLASSIFICATIONS OF CONTROL SYSTEMS

So far we have discussed control systems as being either open or closed loop, analog or digital. Yet we can classify control systems in other ways, which have to do with applications. Some of the most common applications are discussed next.

1.3.1 Process Control

Process control refers to a control system that oversees some industrial process so that a uniform, correct output is maintained. It does this by monitoring and adjusting the control parameters (such as temperature or flow rate) to ensure that the output product remains as it should.

The classic example of process control is a closed-loop system maintaining a specified temperature in an electric oven, as illustrated in Figure 1-7. In this case, the actuator is the heating element, the controlled variable is the temperature, and the sensor is a thermocouple (a device that converts temperature into voltage). The controller regulates power to the heating element in such a way as to keep the temperature (as reported by the thermocouple) at the value specified by the set point.

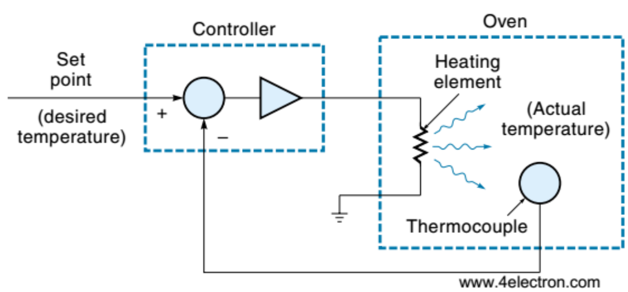

Figure 1-7 A closed-loop oven-heating system

Process control can be classified as being a batch process or a continuous process. In a **continuous process** there is a continuous flow of material or product. A **batch process** has a beginning and an end (which is usually performed over and over).

In a large plant such as a refinery, many processes are occurring simultaneously and must be coordinated because the output of one process is the input of another. In the early days of process control, separate independent controllers were used for each process, as shown in Figure 1-8(a). The problem with this approach was that, to change the overall flow of the product, each controller had to be readjusted manually. In the 1960s, a new system was developed in which all independent controllers were replaced by a single large computer. Illustrated in Figure 1-8(b), this system is called **direct digital control** (DDC). The advantage of this approach is that all local processes can be implemented, monitored, and adjusted from the same place. Also, because the computer can "see" the whole system, it is in a position to make adjustments to enhance total system performance. The drawback is that the whole plant is dependent on that one computer. If the computer goes off line to fix a problem in one process, the whole plant shuts down.

The advent of small microprocessor-based controllers has led to a new approach called **distributed computer control** (DCC), illustrated in Figure 1-8(c). In this system, each process has its own separate controller located at the site. These local controllers are interconnected via a local area network so that all controllers on the network can be monitored or reprogrammed from a single supervisory computer. Once programmed, each process is essentially operating independently. This makes for a more robust and safe system, because all the local processes will continue to function even if the supervisory computer or network goes down. For example, a local controller whose job it is to keep some material at a critical temperature will continue to function even if the supervisory computer is temporarily disabled. Increasingly, the components of a control system are being interconnected with the "business office" network in a factory, which allows the status of any process in the factory to be examined by any computer on anyone's desk. You might be able to sit down at a PC anywhere in the building and determine whether a particular photo sensor on an assembly line has a dirty lens or how much current a particular motor is drawing.

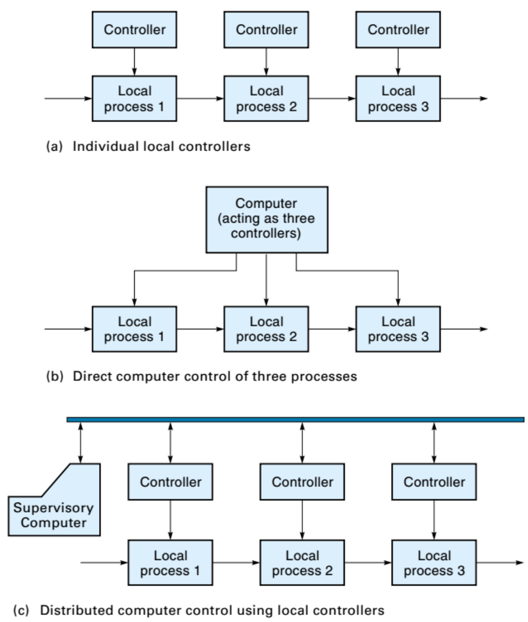

Figure 1-8 Approaches of multi-process control.

1.3.2 Sequentially Controlled Systems

A **sequentially controlled system** controls a process that is defined as a series of tasks to be performed—that is, a sequence of operations, one after the other. Each operation in the sequence is performed either for a certain amount of time, in which case it is **time-driven,** or until the task is finished (as indicated by, say, a limit switch), in which case it is **event-driven.** A time-driven sequence is *open-loop* because there is no feedback, whereas an event-driven task is *closed-loop* because a feedback signal is required to specify when the task is finished.

The classic example of a sequentially controlled system is the automatic washing machine. The first event in the wash cycle is to fill the tub. This is an event-driven task because the water is admitted until it gets to the proper level as indicated by a float and limit switch (closed loop). The next two tasks, wash and spin-drain, are each done for a specified period of time and are time-driven events (open loop). A timing diagram for a washing machine is shown in Figure 1-9. Another example of a sequentially controlled system is a traffic signal. The basic sequence may be time-driven: 45 seconds for green, 3 seconds for yellow, and 45 seconds for red. The presence or absence of traffic, as indicated by sensors in the roadbed, however, may alter the basic sequence, which is an event-driven control.

Many automated industrial processes could be classified as sequentially controlled systems. An example is a process where parts are loaded into trays, inserted into a furnace for 10 minutes, then removed and cooled for 10 minutes, and loaded into boxes in groups of six. In the past, most sequentially controlled systems used switches, relays, and electromechanical timers to implement the control logic. These tasks are now performed more and more by small computers known as **programmable logic controllers** (PLCs), which are less expensive, more reliable, and easily reprogrammed to meet changing needs—for example, to put eight items in a box instead of six.

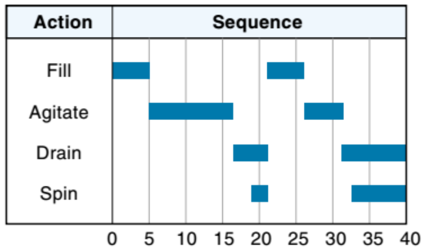

Figure 1-9 Timing diagram for an automatic washing machine

1.3.3 Motion Control

Motion control is a broad term used to describe an open-loop or closed-loop electromechanical system wherein things are moving. Such a system typically includes a motor, mechanical parts that move, and (in many cases) feedback sensor(s). Automatic assembling machines, industrial robots, and numerical control machines are examples.

1.3.4 Servomechanisms

Servomechanism is the traditional term applied to describe a closed-loop electromechanical control system that directs the precise movement of a physical object such as a radar antenna or robot arm. Typically, either the output position or the output velocity (or both) is controlled. An example of a servomechanism is the positioning system for a radar antenna, as shown in Figure 1-10. In this case, the controlled variable is the antenna position. The antenna is rotated with an electric motor connected to the controller located some distance away. The user selects a direction, and the controller directs the antenna to rotate to a specific position.

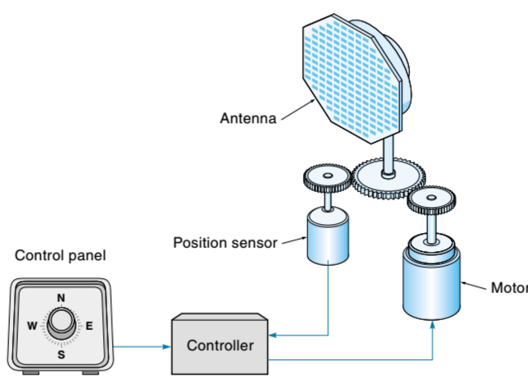

Figure 1-10 A servomechanism: a remote antenna positioning system

1.3.5 Numerical Control

Numerical control (NC) is the type of digital control used on machine tools such as lathes and milling machines. These machines can automatically cut and shape the workpiece without a human operator. Each machine has its own set of axes or parameters that must be controlled; as an example, consider the milling machine shown in Figure 1-11. The workpiece that is being formed is fastened to a movable table. The table can be moved (with electric motors) in three directions: *X*, *Y*, and *Z*. The cutting-tool speed is automatically controlled as well. To make a part, the table moves the workpiece past the cutting tool at a specified velocity and cutting depth. In this example, four parameters (*X*, *Y*, *Z*, and rpm) are continuously and independently controlled by the controller.

The controller takes as its input a series of numbers that completely describe how the part is to be made. These numbers include the physical dimensions and such details as cutting speeds and feed rates. NC machines have been used since the 1960s, and certain standards that are unique to this application have evolved. Traditionally, data from the part drawing were entered manually into a computer program. This program converted the input data into a series of numbers and instructions that the NC controller could understand, and either stored them on a floppy disk or tape, or sent the data directly to the machine tool. These data were read by the machine-tool controller as the part was being made. With the advent of **computer-aided design** (CAD), the job of manually programming the manufacturing instructions has been eliminated. Now it is possible for a special computer program (called a *postprocessor*) to read the CAD-generated drawing and then produce the necessary instructions for the NC machine to make the part.

This whole process—from CAD to finished part—is called **computer-aided manufacturing** (CAM). One big advantage of this process is that one machine tool can efficiently make many different parts, one after the other. This system tends to reduce the need for a large parts inventory. If the input tape (or software) is available, any needed part can be made in a short period of time. This is one example of **computer-integrated manufacturing** (CIM), a whole new way of doing things in the manufacturing industry. CIM involves using the computer in every step of the manufacturing operation—from the customer order, to ordering the raw materials, to machining the part, to routing it to its final destination.

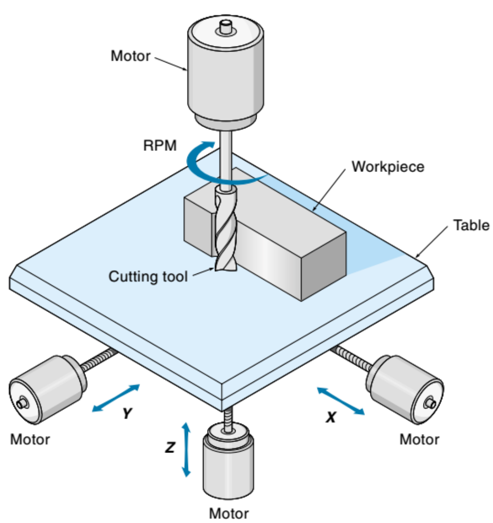

Figure 1-11 Basics of a numerical control milling machine

1.3.6 Robotics

Industrial **robots** are classic examples of position control systems. In most cases, the robot has a single arm with shoulder, elbow, and wrist joints, as well as some kind of hand known as an *end effector*. The end effector is either a gripper or other tool such as a paint spray gun. Robots are used to move parts from place to place, assemble parts, load and off-load NC machines, and perform such tasks as spray painting and welding.

Pick-and-place robots, the simplest type, pick up parts and place them somewhere else nearby. Instead of using sophisticated feedback control, they are often run openloop using mechanical stops or limit switches to determine how far in each direction to go (sometimes called a "bang-bang" system). An example is shown in Figure 1-12. This robot uses pneumatic cylinders to lift, rotate, and extend the arm. It can be programmed to repeat a simple sequence of operations.

Sophisticated robots use closed-loop position systems for all joints. An example is the industrial robot shown in Figure 1-13. It has six independently controlled axes (known as six degrees of freedom) allowing it to get to difficult-to-reach places. The robot comes with and is controlled by a dedicated computer-based controller. This unit is also capable of translating human instructions into the robot program during the "teaching" phase. The arm can move from point to point at a specified velocity and arrive within a few thousandths of an inch.

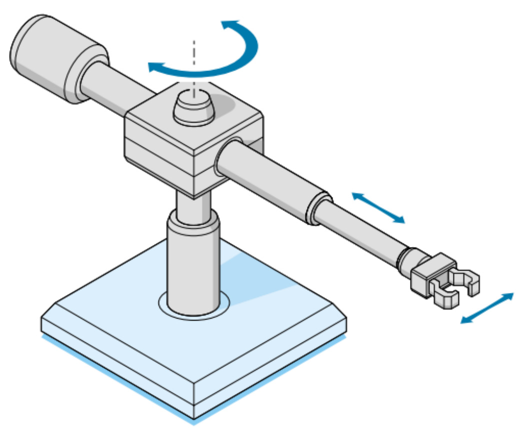

Figure 1-12 A pick-and-place robot.

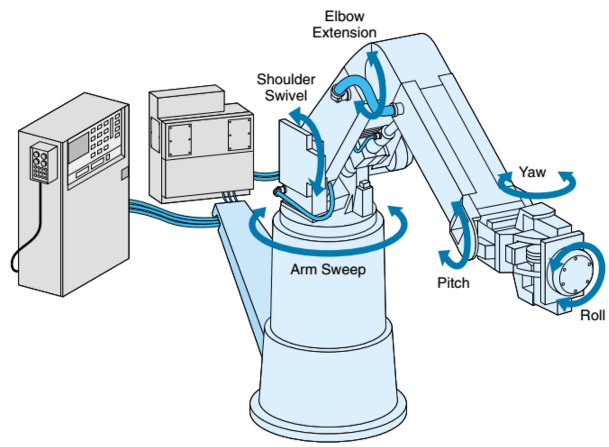

Figure 1-13 A large industrial robot

3 Operational Amplifiers

Introduction

An **operational amplifier (op-amp)** is a high-gain linear amplifier. Op-amps are usually packaged in IC form (one to four op-amps per IC). The op-amp approaches the ideal amplifier of the analog designer's dreams because it has such ideal characteristics:

1. Very high open-loop gain: A = 100,000+, but unpredictable
2. Very high input resistance: $R_{in} > 1$ MΩ
3. Low output resistance: R_{out} = 50-75 ohm

These characteristics make designing with op-amps relatively easy. the high open-loop gain makes it possible to create an amplifier with a very predictable stable gain of anywhere from 1 to 1000 or more. The significance of the very high input resistance (R_{in}) is that the op-amp draws very little input current. This means it will not load down whatever circuit or sensor is driving it. The op-amp's low output resistance (R_{out}) means it can drive a load without being loaded down itself. However, an op-amp is a signal amplifier, not a power amplifier. It is not designed to output large currents and so is not usually used to drive loads such as loudspeakers or motors directly.

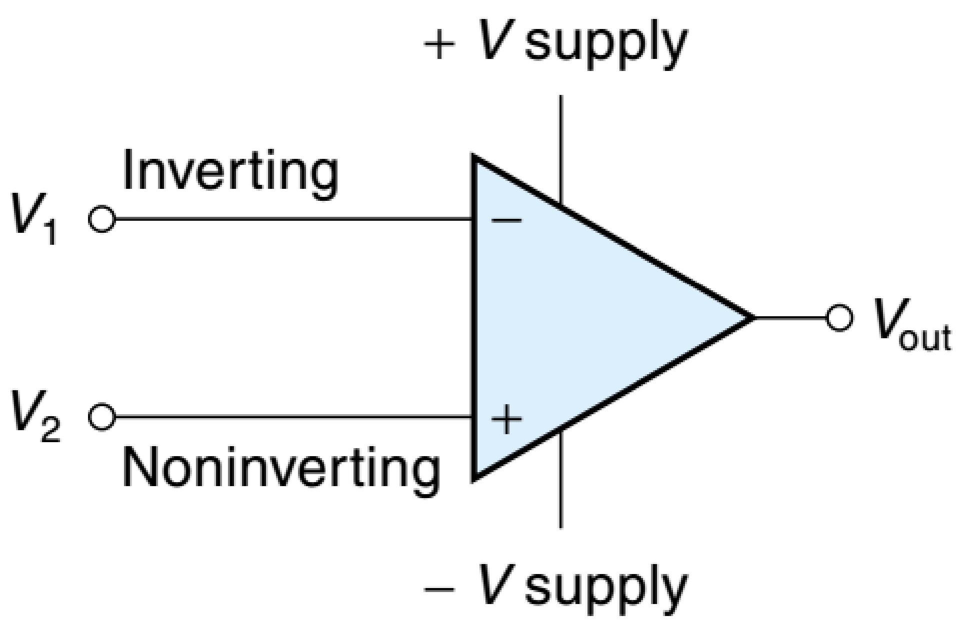

Figure 3-1 **The op-amp symbol**

Figure 3-1 shows the symbol for a typical op-amp. It has two inputs (V_1 and V_2) and one output (V_{out}). Also shown are the two power-supply inputs, which are typically +12 V and –12 V. The output voltage can swing to within about 80% of the supply voltages. Notice there is no ground connection at all. Most op-amps are actually differential amplifiers, which means they amplify the difference between V_1 and V_2. This is shown in Equation 3.1:

$$V_{out} = A(V_2 - V_1)$$

3-1

where
V_{out} = output voltage
A = open-loop gain
V_1 = inverting input
V_2= noninverting input

The **open-loop gain** (A) is the raw unmodified gain of the op-amp; it is high, typically 100,000 or more. V_2 is called the **noninverting input.** As the name implies, *the output is in phase with the noninverting input* (when the noninverting input goes positive, V_{out} goes positive; when the noninverting input goes negative, V_{out} goes negative). The noninverting input is identified by the + sign in the symbol of Figure 3-1. The other input to the op-amp is called the **inverting input.** *The output will be out of phase with the signal at the inverting input* (when the inverting input goes more positive, the output will go more negative, and vice versa). The inverting input is identified by the – sign in the symbol. Even though the op-amp has two separate inputs, there is just one input voltage, which is the difference between V_2 and V_1. This is illustrated in next example.

EXAMPLE 3.1

Figure 3.2 shows an op-amp with an open-loop gain of 100,000. Find the output for the following conditions:

a. V_1 and V_2 are both 4 μV.

b. V_1 is 2 μV, and V_2 is 4 μV.

c. V_1 is 6 μV, and V_2 is 3 μV.

SOLUTION

We will use Equation 3.1 to solve this problem.

a. Both V_1 and V_2 are 4 μV:

$$V_{out} = 100{,}000 \times (4\ \mu V - 4\ \mu V)$$
$$= 100{,}000 \times (0\ \mu V)$$
$$= 0\ V$$

This shows that the output of the op-amp is zero if the inputs are the same voltage, regardless of their actual value.

b. The noninverting input V_2 is 4 μV, and the inverting input V_1 is 2 μV:

$$V_{out} = 100{,}000 \times (4\ \mu V - 2\ \mu V)$$
$$= 100{,}000 \times (2\ \mu V)$$
$$= 0.2\ V$$

This result shows that the output is positive if the $(V_2 - V_1)$ quantity has a net positive value.

c. The inverting input is 6 μV, and the noninverting input is 3 μV:

$$V_{out} = 100{,}000 \times (3\ \mu V - 6\ \mu V)$$
$$= 100{,}000 \times (-3\ \mu V)$$
$$= -0.3\ V$$

This result is negative because the $(V_2 - V_1)$ quantity has a negative net value.

Note: The purpose of this example is only to show how an open-loop op-amp behaves. It would be very difficult to duplicate this in the lab because of the challenge of creating small, steady input voltages.

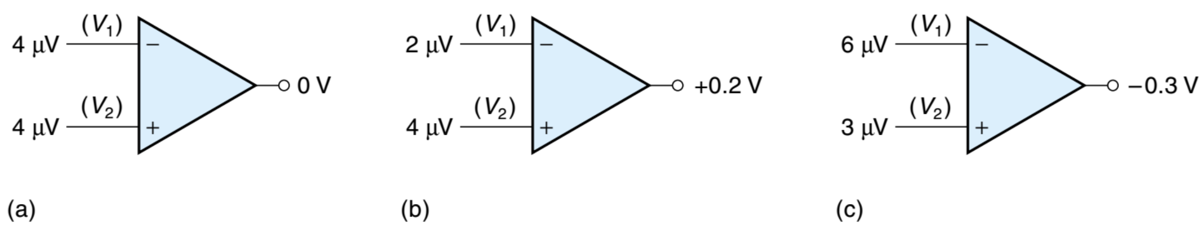

Figure 3-2 **Various input voltage combinations (*A* = 100,000)**

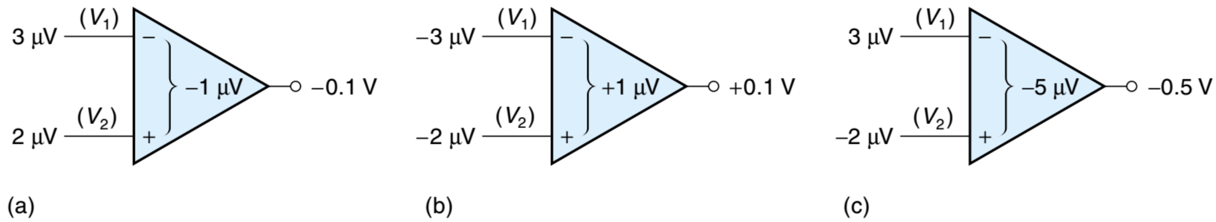

Figure 3-3 **Various input voltage combinations (*A* = 100,000).**

Consider the three op-amps in Figure 3-3. In Figure 3-3(a), both inputs are positive, yet the output is negative. The – input has the larger magnitude, so the quantity ($V_2 - V_1$) is negative (2 μV – 3 μV = –1 μV). From Equation 3.1 (the op-amp equation),

$$V_{out} = A(V_2 - V_1)$$

in case where only a single input is required. There are two possibilities: The output will be either in phase or out of phase with the input. To make a noninverting amplifier (where the output is in phase with the input), the – input is grounded, and the input signal is connected to the noninverting input (+), as shown in Figure 3-4(a). If we want an inverting amplifier, where the output is out of phase with the input, we connect the signal into the inverting input (–) and ground the + input, as shown in Figure 3-4(b)

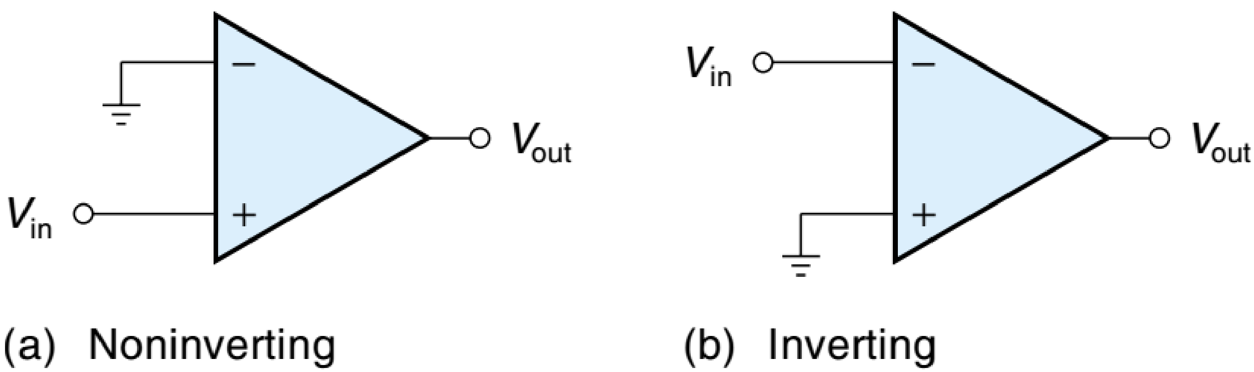

Figure 3-4 **Single-input, open-loop amplifiers**

All the amplifier circuits discussed so far are called *open loop* because they operate at open loop gain. However, Most op-amp circuits incorporate **negative feedback.** This means that a portion of the output signal is fed back and subtracted from the input. Negative feedback results in a very stable and predictable operation at the expense of lowered gain.

3.1.1 Assumption 1:

$V_1 = V_2$.

Explanation: the output voltage is equal to $A(V_2 - V_1)$ where A, being the open-loop gain, is a very high number. Thus, even a small difference between V_1 and V_2 will cause a very large output. However, the output has a practical upper limit established by the power supply; therefore, to keep the output from exceeding its limits, the difference between V_1 and V_2 must be very small.

This is illustrated in Figure 3-5. The power supply is +15 V and –15 V, which limits the output voltage swing to about +12 V and –12 V (being 80% of the supply). If the open-loop gain is 100,000, then the difference between V_1 and V_2 that would cause an output of 12 V is computed using Equation 3.1: $V_{out} = A(V_2 - V_1)$. Rearranging gives us

$$(V_2 - V_1) = \frac{V_{out}}{A} = \frac{12\ \text{V}}{100{,}000} = 0.00012\ \text{V}$$

Therefore, to keep the amplifier operating linearly with the output within its bounds, the difference between V_2 and V_1 must be less than 0.00012 V, which is a very small voltage. Hence, we say that V_1 *is virtually the same as* V_2.

3.1.2 Assumption 2:

Input current is zero.

Explanation : The input resistance of an op-amp is very high, typically 1 MΩ or more. It is so high that we can model the inputs as being open circuits as shown in Figure 3-6; of course, no current can flow into an open circuit.

3.1.3 Assumption 3:

Output resistance is zero.

Explanation: A low-output resistance means that the output voltage will not be pulled down even if the load draws a lot of current. This is the weakest of the three assumptions because the output resistance is typically between 50 and 75 Ω (however, it can be much lower with negative feedback). This assumption only holds if the load being driven is considerably higher than the output resistance of the op-amp, which is the case in most applications.

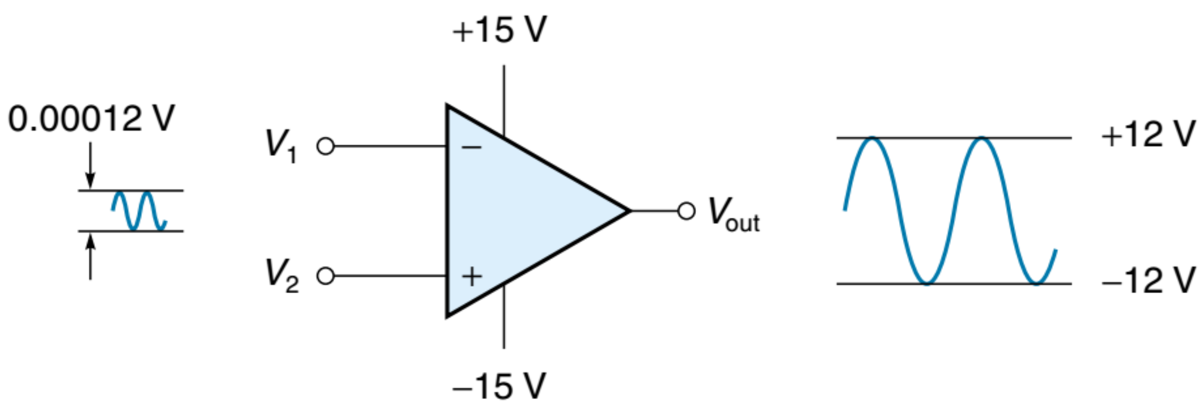

Figure 3-5 Op-amp inputs are always virtually the same voltage.

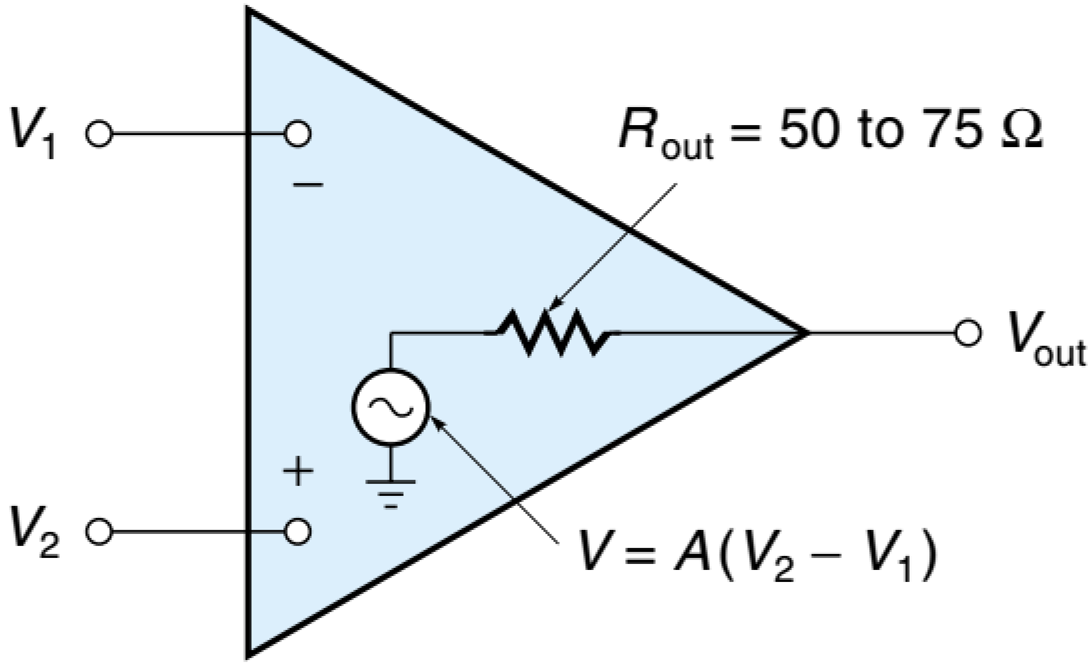

Figure 3-6 **Equivalent circuit model of an op-amp.**

Many useful signal-conditioning circuits can be built using op-amps. Some of the most common are presented in the pages that follow.

3.2 Voltage Follower

The **voltage follower,** which is very useful circuit, can boost the current of a signal without increasing the voltage. It can transform a high-impedance signal (easily loaded down) into a robust low-impedance signal. Figure 3-7 shows a voltage follower circuit. It has a voltage gain of 1, with a high R_{in} and a low R_{out}. Its operation can be explained as follows: We start with the basic op-amp equation:

$$V_{out} = A(V_2 - V_1)$$

In the circuit, V_{out} is connected to V_1; thus, $V_{out} = V_1$. Substituting in V_{out} and expanding Equation 3.1,

$$V_{out} = (AV_2) - (AV_{out})$$

Solving for V_{out}, we get

$$V_{out} = \frac{AV_2}{1 + A} \approx V_2$$

But because A is *much* greater than 1,

$$V_{out} = \frac{AV_2}{A} \approx V_2$$

We see that the output voltage V_{out} equals the input voltage V_2, meaning the overall gain is 1. Also notice that the actual input signal goes directly into the noninverting input, so it draws essentially no current.

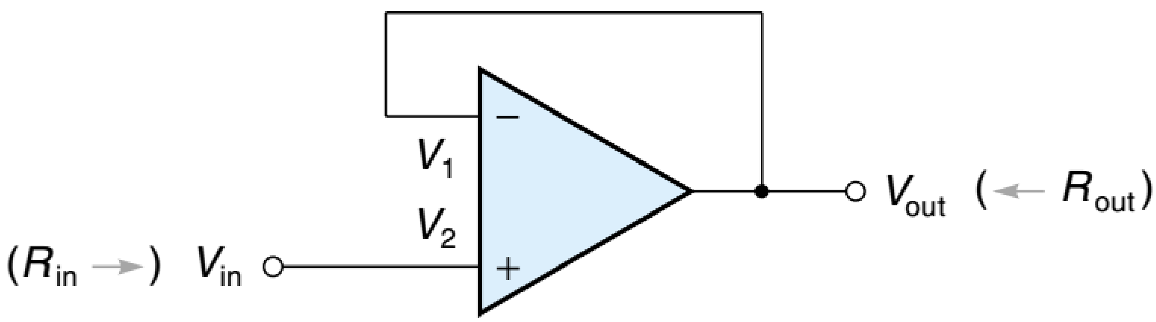

Figure 3-7 **A voltage follower circuit.**

A more intuitive way to explain the voltage follower circuit is as follows: The input V_2 is virtually the same voltage as V_1 (from Assumption 1). V_1 is connected to V_{out}, so it's as if V_2 were connected to V_{out}, hence the gain of 1. The voltage follower is a simple and very useful circuit. Consider the situation shown in Figure 3-8(a). In this case, a high-impedance sensor (10 kΩ), is connected directly to a controller with a 1 kΩ input resistance. The sensor generates 5 V internally, but this is reduced by the voltage drop across the 10 kΩ internal resistance. By redrawing the circuit [Figure 3-8(b)], we see that these two resistances form a voltage divider. The actual input voltage to the controller can be calculated as follows from the voltage-divider rule:

$$V_{in} = \frac{1\ \text{k}\Omega \times 5\ \text{V}}{1\ \text{k}\Omega + 10\ \text{k}\Omega} = 0.45\ \text{V}$$

This shows that only 0.45 V of the 5 V signal makes it to the controller. We could amplify the signal at the controller to make up for the attenuation, but that would amplify noise as well as the signal. A better solution is to insert a voltage follower near the sensor, as shown in Figure 3-8(c). Because the op-amp draws no signal current, there is no voltage drop across the 10 kΩ resistor, and the full 5 V enters the voltage follower and appears at its output. The 1 kΩ input resistance of the controller is so much higher than the output resistance of the op-amp that almost all of the 5 V will appear across the controller terminals.

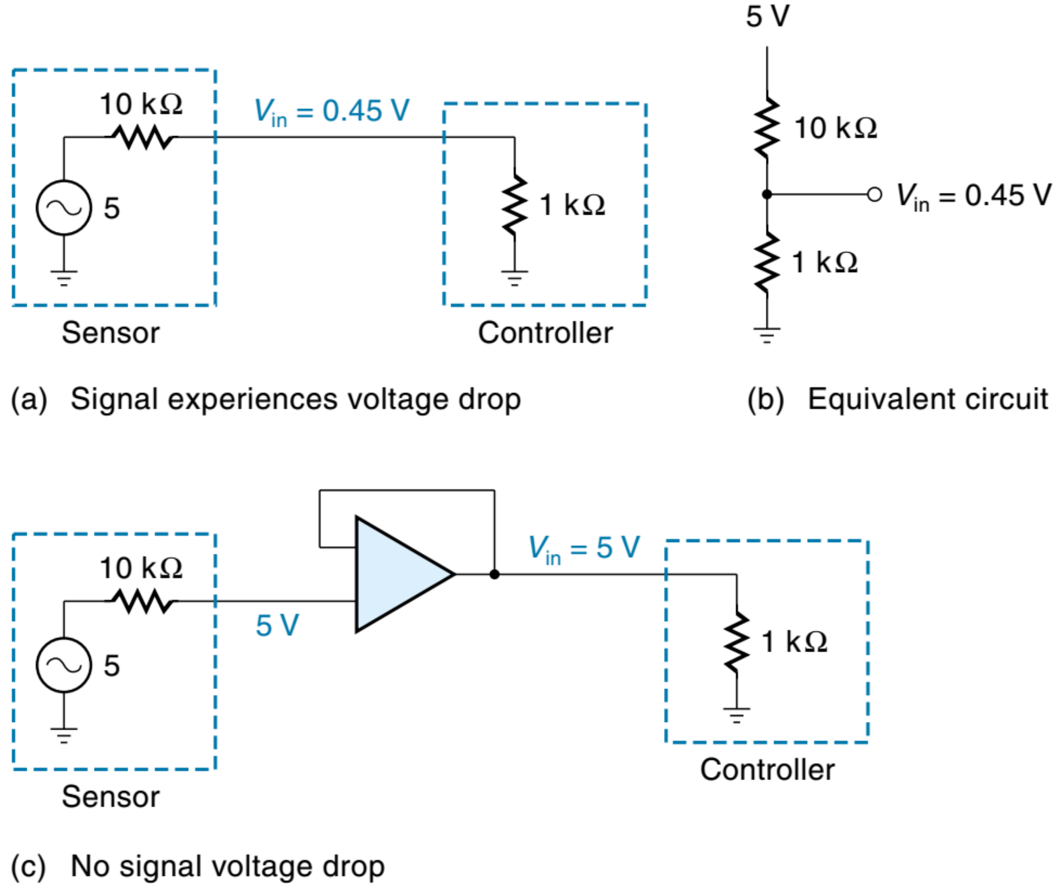

Figure 3-8 **Using a voltage follower to prevent load down.**

3.3 Inverting Amplifier

The **inverting amplifier** is probably the most common op-amp configuration. The circuit shown in Figure 3-9 requires just two resistors, R_i and R_f. R_i is the input resistor, and R_f is the feedback resistor that feeds part of the output signal back to the input. This is an inverting amplifier because *the input signal goes to the inverting input, which means the output is out of phase with the input*. The voltage gain is determined by the resistor values. An explanation of how the inverting amp works is as follows:

First, if the op-amp input draws no current, then all the signal current (I_{in}) must go through R_f—there is nowhere else for it to go. Therefore, $I_{in} = I_f$. By assumption, V_1 and V_2 are virtually the same voltage, and V_2 is grounded; thus, V_1 is at **virtual ground.** If V_1 is (almost) at ground, then the entire input signal voltage V_{in} is dropped across R_i. From Ohm's law,

$$I_{in} = \frac{V_{in}}{R_i}$$

As already noted, virtually all I_{in} goes through the feedback resistor R_f. The voltage across R_f is the difference between virtual ground and V_{out}. Thus, we can write Ohm's law equation for R_f:

$$I_{in} = I_f = \frac{0 - V_{out}}{R_f}$$

Combining the two previous equations,

$$\frac{V_{in}}{R_i} = \frac{0 - V_{out}}{R_f}$$

Solving for V_{out} and rearranging gives us

$$V_{out} = \frac{-V_{in} R_f}{R_i}$$

$$\frac{V_{out}}{V_{in}} = \frac{-R_f}{R_i}$$

However, V_{out}/V_{in} is the voltage gain, so

$$A_V = \frac{-R_f}{R_i} \qquad \textbf{3-2}$$

where
A_V = voltage gain of the inverting amp
R_f = value of the feedback resistor
R_i = value of the input resistor

$$A_V = \frac{-R_f}{R_i}$$

This result (Equation 3-2) shows us that the voltage gain of the inverting amp is simply the ratio of th and R_i. The minus sign reminds us that the output is inverted. The gain derived in

$$A_V = \frac{-R_f}{R_i}$$

Equation 3-2 is called the **closed-loop gain** and is always lower than the (open-loop) gain of the op-amp by itself Another important point is that the input impedance for the entire inverting amp is approximately R_i (*not* infinite as one might think). Figure 3-9 shows this: The right end of R_i is at virtual ground; therefore, the entire V_{in} is “dropped” across R_i.

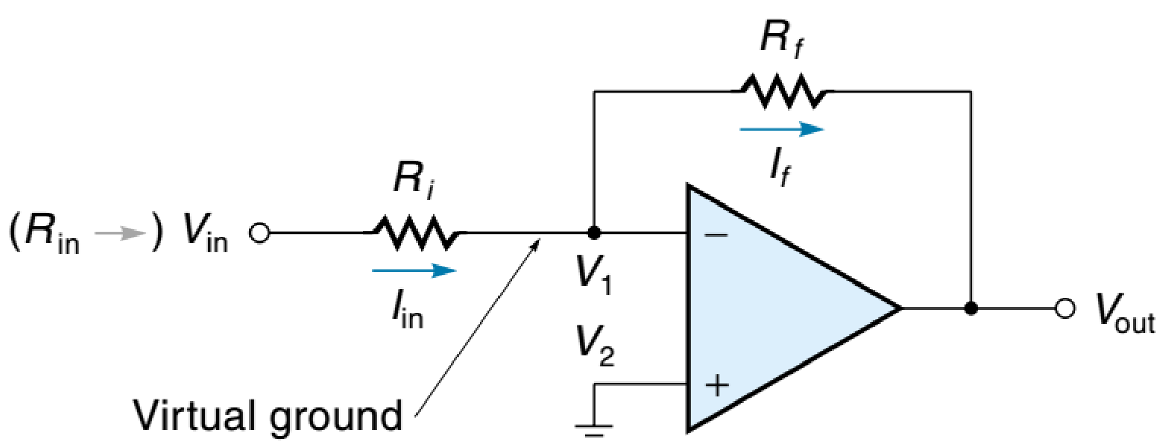

Figure 3-9 The inverting amplifier circuit.

EXAMPLE 3.2

An inverting amp is to have a gain of 10. The signal source is a sensor with an output impedance of 1 kΩ. Draw a circuit diagram of the completed amplifier.

SOLUTION

First, select a value for R_i. Because R_i essentially determines the amplifier's input resistance, *it should be at least ten times higher (if possible) than the signal source impedance to ensure maximum voltage transfer*. In this example, we select $R_i = 10$ kΩ. Next, rearrange Equation 3.2 to solve for R_f:

$$R_f = -AR_i$$
$$= -(-10) \times 10\text{ k}\Omega = 100\text{ k}\Omega$$

Figure 3.11 shows the completed circuit.

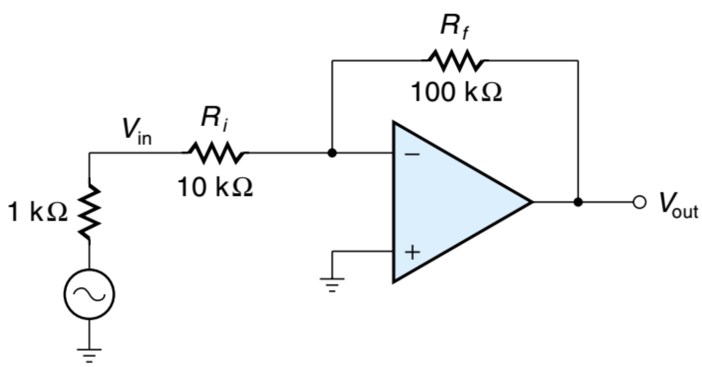

Figure 3-10 **An inverting amplifier circuit (Example 3.2).**

3.4 Noninverting Amplifier

In many situations we don't want the amplifier to invert the output. The circuit for the **noninverting amplifier** is shown in

Figure 3-11. It is similar to the inverting amp except the input signal V_{in} now goes directly to the noninverting input and R_i is grounded. Notice that the noninverting amp has an almost infinite input impedance (R_i) because V_{in} connects only to the opamp input.

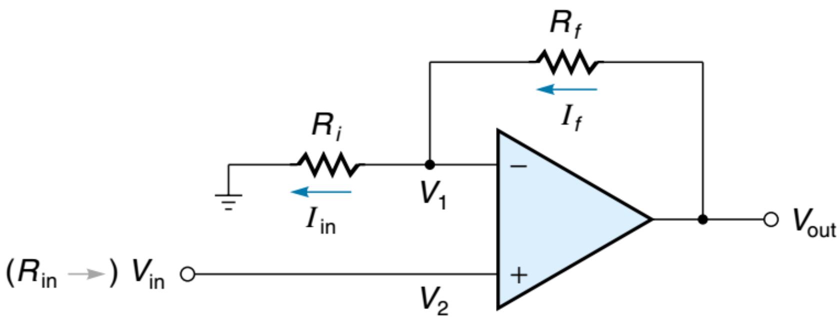

Figure 3-11 The noninverting amplifier circuit

An explanation of how the circuit works is as follows: If V_1 is virtually the same as V_2, then the voltage input (V_{in}) appears across R_i. Applying Ohm's law to R_i, we can calculate I_{in}

$$I_{in} = \frac{V_{in} - 0}{R_i}$$

The current in R_f can also be calculated using Ohm's law. We know the voltage across R_f is the difference between V_{in} and V_{out}. Therefore,

$$I_f = \frac{V_{out} - V_{in}}{R_f}$$

Because no current enters the inverting input of the op-amp, all current in R_f must go into R_i:

$$I_{in} = I_f$$

Combining these three equations gives us

$$I_{in} = I_f = \frac{V_{in} - 0}{R_i} = \frac{V_{out} - V_{in}}{R_f}$$

Solving for V_{out} and rearranging gives us

$$V_{out} - V_{in} = \frac{R_f V_{in}}{R_i}$$

$$V_{out} = \frac{R_f V_{in}}{R_i} + V_{in} = V_{in}\left(\frac{R_f}{R_i} + 1\right)$$

$$\frac{V_{out}}{V_{in}} = \frac{R_f}{R_i} + 1$$

V_{out}/V_{in} is the voltage gain, so the resulting equation for the gain of the noninverting amp is

$$A_V = \frac{R_f}{R_i} + 1 \quad \textbf{3-3}$$

Where A_V = voltage gain for the noninverting amp
R_f = value of the feedback resistor
R_i = value of the input resistor

EXAMPLE 3.3

Draw the circuit diagram of a noninverting amp with a gain of 20.

SOLUTION

Using Equation 3.3 and putting in a gain of 20,

$$A_V = 20 = \frac{R_f}{R_i} + 1$$

Rearranging gives us

$$\frac{R_f}{R_i} = 19 \qquad \text{or} \qquad R_f = 19 \times R_i$$

Now select R_i to be an appropriate value (as explained below) and solve for R_f. If we select R_i to be 2 kΩ, then

$$R_f = 19 \times R_i = 19 \times 2\ \text{k}\Omega = 38\ \text{k}\Omega$$

Figure 3.13 shows the completed circuit. The basis for selecting both R_i and R_f is that the current in these external resistors should be much larger than the small current that actually enters the op-amp (recall that the op-amp equation was based on the assumption that *no* current enters the op-amp). Therefore, both R_i and R_f should be at least ten times smaller than the op-amp input resistance—in this case, no more than 100 kΩ if possible.

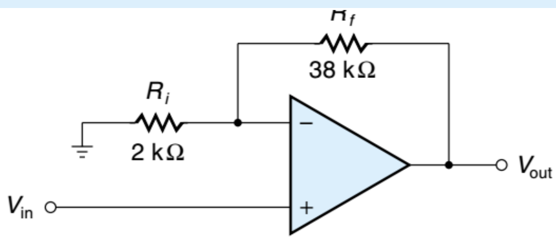

Figure 3-12 A noninverting amplifier circuit (Example 3.3).

3.5 Summing Amplifier

The summing amplifier has an output voltage that is the sum of any number of input voltages as depict in Figure 3-13 where the amplifier would add the input voltages of 1 V, 2 V, and 4 V and give an output of 7 V.

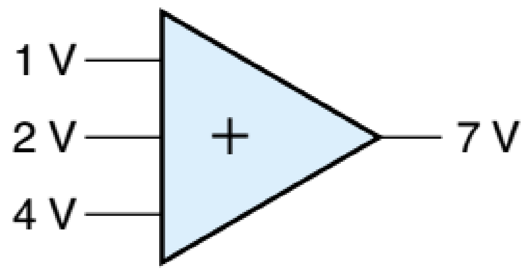

Figure 3-14 **Connecting the wires does not sum the voltage**

Figure 3.15 shows the summing amplifier circuit. An explanation of how it works is as follows: Because the op-amp input draws no current, all individual input currents must combine and go through the feedback resistor R_f

$$I_f = I_a + I_b + I_c \qquad \textbf{3-4}$$

Note that V_2 is grounded, so V_1 is at virtual ground. Therefore, the voltage across each of the four resistors is simply V_a, V_b, V_c, and V_{out}. Applying Ohm's law to express the current through each resistor, we can rewrite $I_f = I_a + I_b + I_c$ *3-4* as follows:

$$\frac{0 - V_{out}}{R_f} = \frac{V_a}{R_a} + \frac{V_b}{R_b} + \frac{V_c}{R_c}$$

Solving for V_{out},

$$V_{out} = -\left(\frac{R_f}{R_a}V_a + \frac{R_f}{R_b}V_b + \frac{R_f}{R_c}V_c\right)$$

If $R_a = R_b = R_c = R_i$, the output of the summing amp simplifies to:

$$V_{out} = -\frac{R_f}{R_i}(V_a + V_b + V_c) \qquad \textbf{3-5}$$

Where V_{out} = output voltage of the summing amp

R_f = value of the feedback resistor

R_i = value of all input resistors

V_a, V_b, V_c = input signal voltages

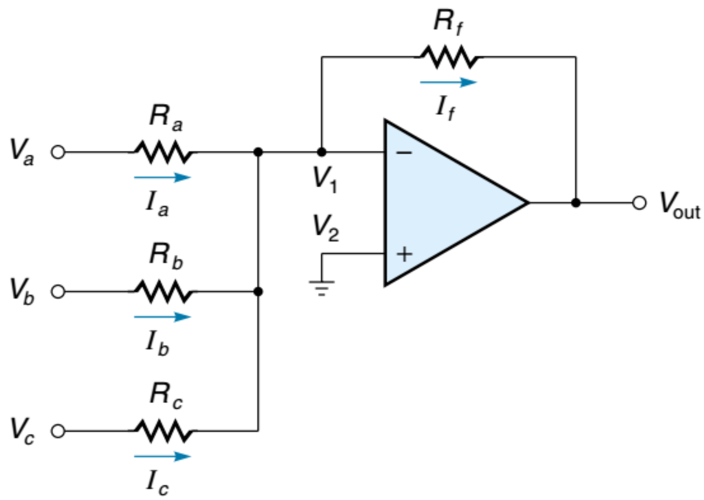

Figure 3-15 **The summing amplifier circuit.**

EXAMPLE 3.4

According to a comfort scale, the air conditioning in a building should come on when the sum of the temperature and humidity sensor voltages goes above 1 V. A threshold circuit in the air conditioner requires 5 V for turn-on. Design an interface circuit to connect the two sensors to the air conditioning unit.

SOLUTION

This circuit requires a summing amplifier with two inputs and a gain of 5. By specifying both input resistors to be the same (at 1 kΩ), we can use Equation 3.5, and our only calculation concerns the gain portion of the equation:

$$A = \frac{R_f}{R_i} = 5$$

Rearranging gives us

$$R_f = 5 \times R_i$$

When $R_f = 1$ kΩ,

$$R_f = 5 \times 1\ \text{k}\Omega = 5\ \text{k}\Omega$$

Figure 3.16 shows the completed circuit. Notice that an inverting amp with a gain of 1 ($R_i = R_f$) was added to make the final output positive.

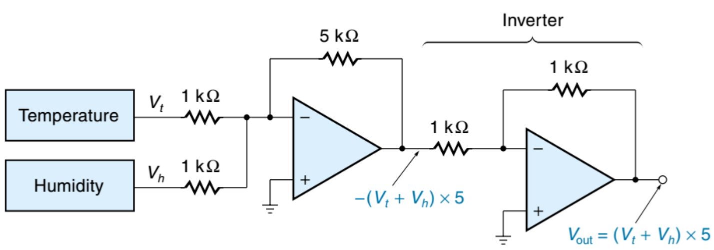

Figure 3-16 **A summing amplifier circuit (Example 3.4)**

3.6 Differential and Instrumentation Amplifiers

A **differential amplifier** amplifies the difference between two input voltages. In the circuits the input voltages have been referenced to ground, but the op-amp can be the basis of a practical differential amp as well. Figure 3-17 shows the circuit will amplify a **differential voltage,** which is the difference between the two voltage levels V_a and V_b, when neither is ground. The output of the amplifier (V_{out}) is a single voltage level referenced to ground, sometimes called a **single-ended voltage.** If $R_a = R_b$ and $R_f = R_g$, which is usually the case, then the equation for V_{out} is

$$V_{out} = \frac{R_f}{R_a}(V_b - V_a) \qquad \textbf{3-6}$$

Rearranging gives us

$$\frac{V_{out}}{(V_b - V_a)} = \frac{R_f}{R_a}$$

$V_{out}/(V_b - V_a)$ is output/input, which is a gain, so

$$A_V = \frac{R_f}{R_a} \qquad \textbf{3-7}$$

Where A_V = voltage gain of the differential amp
R_f = value of R_f and R_g
R_a = value of input resistors, R_a and R_b

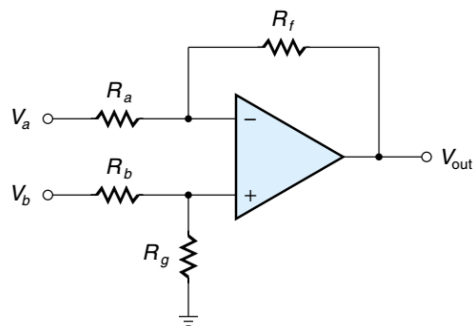

Figure 3-17 The differential amplifier circuit.

EXAMPLE 3.5

A differential amp is needed to amplify the voltage difference between two temperature sensors. The sensors have an internal resistance of 5 kΩ, and the maximum voltage difference between the sensors will be 2 V. Design the differential amp circuit to have an output of 12 V when the difference the inputs is 2 V.

SOLUTION

First calculate the gain required:

$$A_V = \frac{V_{\text{out}}}{V_{\text{in}}} = \frac{12\text{V}}{2\text{V}} = 6$$

By letting $R_a = R_b$ and $R_f = R_g$, we can use Equation 3.7. Noting that the sensor impedance is 5 kΩ, we would like the input resistance of the amp to be at least ten times 5 kΩ. Therefore, if we select $R_a = 50$ kΩ, then

$$A_V = \frac{R_f}{R_a} = 6$$

solving for R_f

$$R_f = R_a \times 6$$

$$= 50\ \text{k}\Omega \times 6 = 300\ \text{k}\Omega$$

Figure 3.18 shows the completed design.

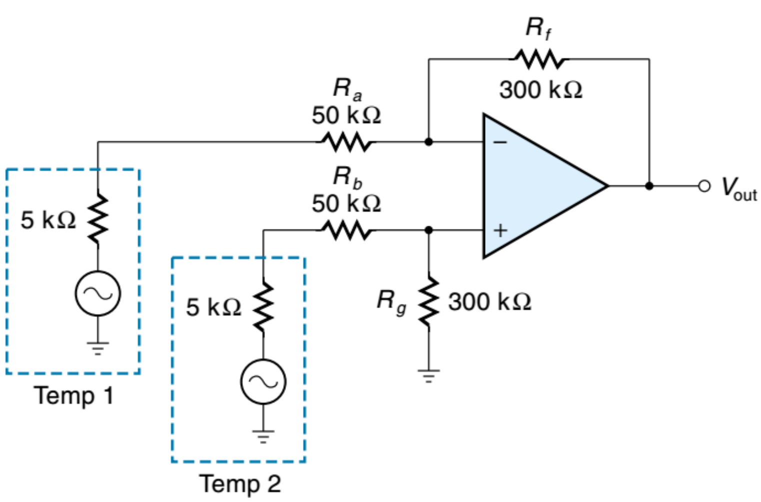

Figure 3-18 **A differential amplifier circuit (Example 3.5)**

As an example of such circuit is **common mode rejection** where the information is transmitted as a *differential signal*. The advantage of this system is that it reduces the effect of electrical noise. A noise signal would tend to couple onto both wires; for example, a positive voltage noise "spike" would cause a positive spike in *both* wires, which would be canceled out by the differential amplifier (because only voltage differences are amplified).

An instrumentation amplifier is a differential amp that has its inputs buffered with voltage followers, as shown in Figure 3-19. Voltage follower circuits on the inputs perform three desirable functions:

(1) They increase the input resistance so that the source (such as a sensor) will never be loaded down,

(2) they make both input resistances equal, and

(3) they isolate the gain-defining resistors (R_f, R_i, etc.) from the signal source. This last quality means that instrumentation amps can be prebuilt to have a specific gain.

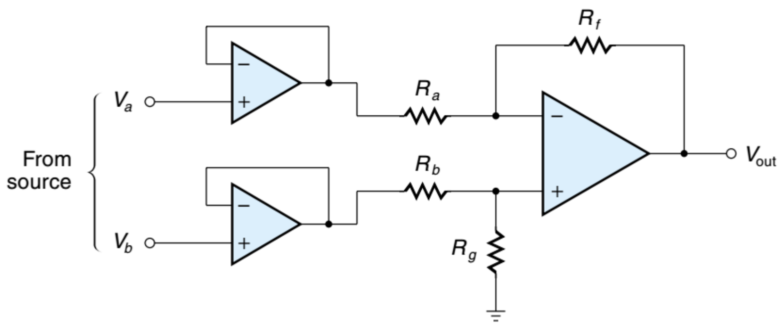

Figure 3-19 **The instrumentation amplifier circuit**

3.7 Integrators and Differentiators

Op-amp circuits can be designed to integrate or differentiate an incoming waveform.
These special-purpose circuits are likely to be found only inside an analog controller.
Figure 3-20 shows an **integrator** circuit. Notice that the feedback element is a capacitor. The integrator gives an output voltage (V_{out}) that is proportional to the total area under a curve traced out by the input voltage waveform (the horizontal axis being time), as specified in Equation 3.8:

$$V_{out} = -\frac{1}{RC} \times (\text{area under } V_{in} \cdot \textit{time} \text{ curve})$$

3-8

Where V_{out} = output voltage of the integrator

R, C = values of components
The integrator circuit works by converting V_{in} into a constant current source that forces the capacitor (*C*) to charge at a linear rate, thus building up the voltage across C.

The integrator concept can be explained by the sample waveforms shown in Figure 3-21 (which assumes $RC = 1$). Notice that the integrator input voltage (V_{in}) rises from 0 to 1 V in the first 10 s. The triangular area under that portion of the curve (*a*-*b*) is 5 V · s, so the output (V_{out}) of the integrator goes from 0 to –5 V during the same time. In other words, the output voltage ends up being the same magnitude as the area under the curve, in this case 5 (the minus sign appears because it is an inverting amp). From time *b* to *c*, V_{in} remains at 1 V, so the new area added is 10 V · s. Consequently, the magitude of V_{out} increases by 10 to become –15 V at time *c*. Then, V_{in} returns to 0 V. Because no new area is added between *c* and *d*, V_{out} remains at –15 V.

Integrators can be useful because they keep a record of what has gone on before. The simple integrator circuit shown is not practical, because any offset voltage will eventually cause the output to build up and saturate at the power supply voltage. One solution is to put a resistor (*Rf*) across the capacitor to provide some dc feedback. If the value of *R f* is at least 10 times greater than *R,* the circuit performance will usually not be adversely affected. Figure 3-22 shows a differentiator circuit. The differen tage that is proportional to the rate of change (slope) of the input voltage, as specified

$$V_{\text{out}} = -\frac{1}{RC} \times \frac{\Delta V_{\text{in}}}{\Delta t}$$

in Equation **3-9**:

$$V_{\text{out}} = -\frac{1}{RC} \times \frac{\Delta V_{\text{in}}}{\Delta t} \qquad \textbf{3-9}$$

Where V_{out} = output voltage of the differentiator *R, C* = values of components in Figure 3-22

$\Delta V_{in}/\Delta t$ = rate of change, or slope, of V_{in}

The differentiator concept is illustrated in Figure 3-23 (which assumes *RC* = 1). From time *a* to *b*, the input voltage (V_{in}) is 0 V, and because it is not changing, the output voltage (V_{out}) is 0 V. During the time period *b*-*c*, V_{in} increases at a constant rate of 1 V/s, so the V_{out} curve reflects this by staying at a constant –1 V (it is negative because the inverting input is used). From time *c* to *d*, the slope of V_{in} increases to 2 V/s, so V_{out} jumps to –2 V. After time *d*, V_{in} stays at 3 V, and because it is not changing, V_{out} is 0. Differentiators can tell us how fast a variable is changing. In practice, however, they suffer from the problem that even a small amount of noise in the input will be accentuated, giving a very "noisy" output.

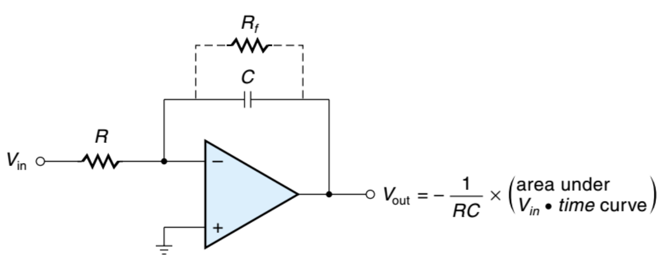

Figure 3-20 An integrator circuit

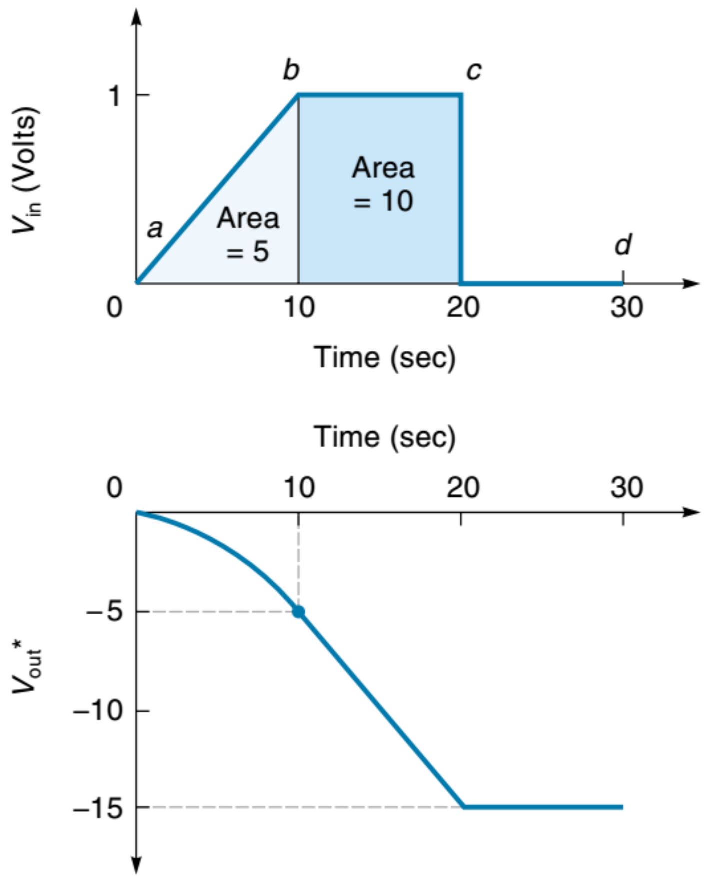

Figure 3-21 The voltage waveform of an integrator circuit. (*V_{out} = –V · s in this case because *RC* = 1.)

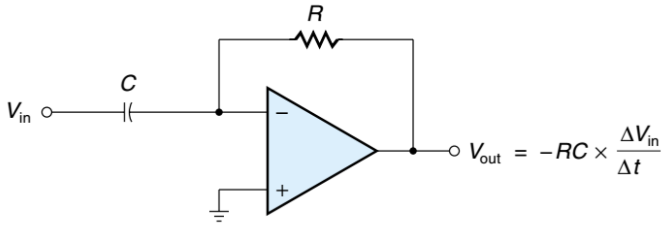

Figure 3-22 **A differentiator circuit**

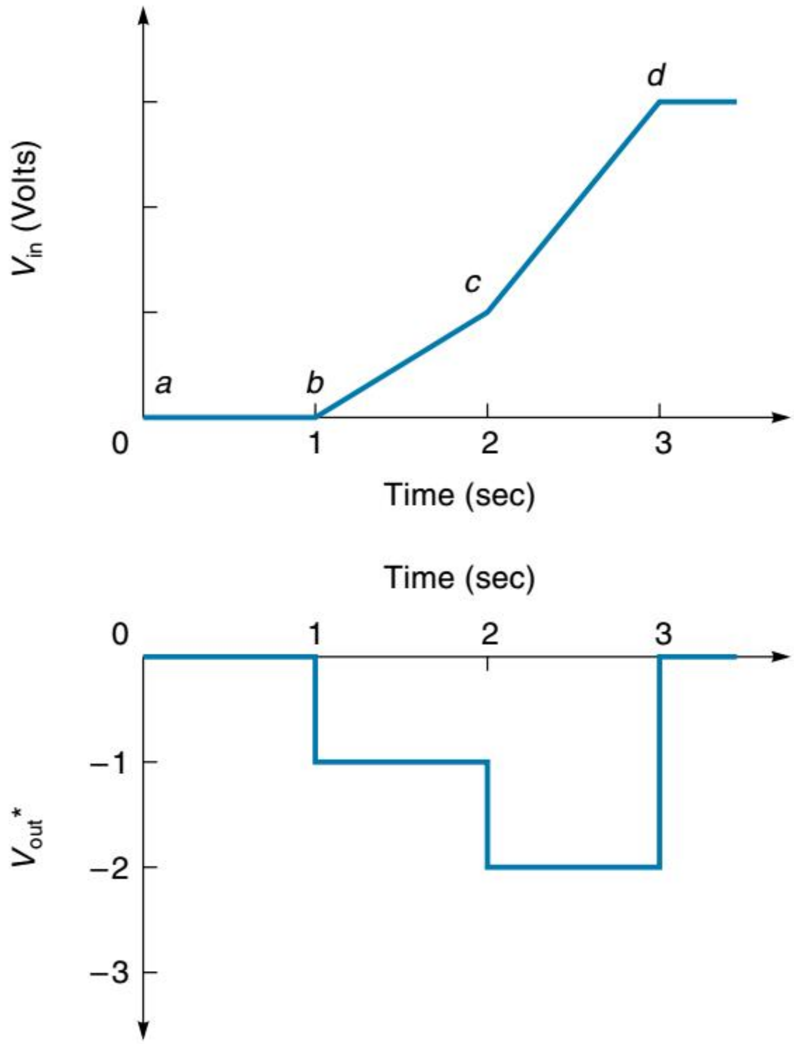

Figure 3-23 Voltage waveforms of a differentiator. (*V_{out} = -V_{in} this case because *RC* = 1).