Sensors

Your PIC32 interacts with the outside world through sensors and actuators. "Mechatronics" is the design of microprocessor-controlled electromechanical systems incorporating sensors and actuators. There is no clear distinction between mechatronics and robotics, but we typically think of robots as higher-level, more complex and general purpose devices, often with more sophisticated sensing and artificial intelligence than we associate with mechatronics. Robots often integrate multiple mechatronic subsystems.

In this chapter we focus on sensors. Sensors transduce physical properties of interest to a signal that a microcontroller can understand. Some sensors produce a simple digital or analog voltage signal, which may need *signal conditioning* circuitry before sending to the PIC32. Examples of signal conditioning include switch debouncing, voltage amplification, and low-pass filtering (see Appendix B). Other sensors, like rotary encoders, encode their signals in digital pulse trains, to be decoded either by the PIC32 itself or by an external circuit. Finally, some sensors incorporate their own microprocessor or ASIC (application-specific integrated circuit) and can communicate by one of the peripherals discussed earlier (e.g., UART, CAN, I^{2}C, or SPI).

An overview of sensors could be organized by the transduction principle involved (e.g., a voltage proportional to a magnetic field's strength due to the Hall effect or a current proportional to light intensity by the photoelectric effect). These transduction principles can be applied to measure many other quantities of interest; for example, sensors can be constructed to measure the rotation angle of a joint using either the Hall effect or photodiodes. To the mechatronics designer, the specific transduction principle employed is typically secondary to the sensor's capability of sensing the quantity of interest (e.g., the joint angle). Therefore, in this chapter, we roughly organize the presentation around typical quantities of interest: angle, angular velocity, acceleration, force, etc. We do not go into details of the physics of the transduction principles, but provide a broad overview of relatively inexpensive sensors common in mechatronics and how they can be interfaced to the PIC32.

Many of the sensors in this chapter can be purchased from vendors like Digikey or Mouser, or on convenient breakout boards from vendors like SparkFun, Pololu, or Adafruit.

Embedded Computing and Mechatronics with the PIC32 Microcontroller. http://dx.doi.org/10.1016/B978-0-12-420165-1.00021-4

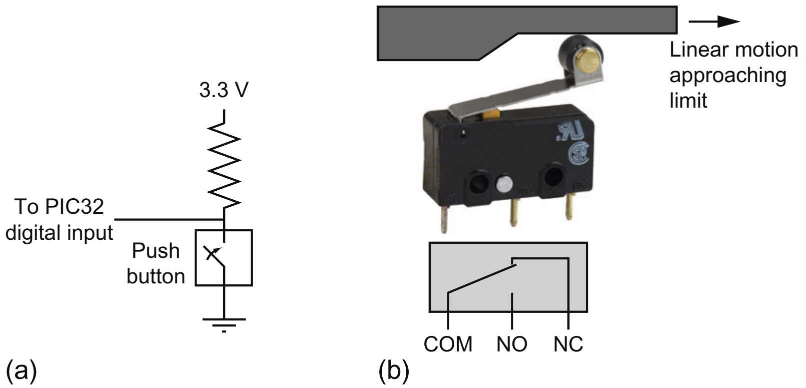

Figure 21.1
(a) A button interfacing to the PIC32. (b) A hinged roller-lever limit switch, and its use in a linear joint to detect the end of travel. (Image courtesy of Digi-Key Electronics, digikey.com.)

21.1 Contact: Buttons and Switches

Perhaps the simplest sensors are buttons and switches. Buttons and switches can be used to get information from a user (e.g., a keyboard or the USER button on the NU32) or to sense when a robot joint has reached the limit of its allowable travel (a *limit switch*).

A simple button interface to the PIC32 is shown in Figure 21.1(a). The button has two connections, one to a pull-up resistor and one to ground. When the button is unpressed, the internal switch is open circuit, and the digital input to the PIC32 is high (3.3 V). When the button is pressed, the switch is closed, pulling the digital input low (ground). This kind of button is called *normally open* (or NO for short). There are also buttons that are *normally closed* (NC), requiring the button to be pressed to open the internal switch.

One common application for a switch is as a limit switch. When a robot linear or rotary joint reaches its limit of travel, the limit switch depresses, sending a signal to the controller to stop driving the joint. The limit switch in Figure 21.1(b) has both an NO and an NC output.

The switch interface in Figure 21.1(a) shows an external pull-up resistor. The digital inputs on the PIC32 supporting Change Notification have internal pull-up resistors that can be used instead, eliminating the external resistor (see Chapter 7).

A common problem with mechanical switches such as those in Figure 21.1 is *bounce*—many on-off transitions in a brief period of time as the switch establishes or breaks contact. If bounce is a problem for the particular application, the designer should decide the shortest amount of time allowed between "real" transitions (as opposed to mechanical bounces), then design either a circuit or a software routine to *debounce* the switch. See, for example, the debouncing circuit in Appendix B.2.1.

A switch is often characterized by the number of *poles* and *throws*. The number of poles is the number of internal moving levers, and the number of throws is the number of different connections each lever can make contact with. Thus the switch in Figure 21.1(a) is a single-pole single-throw switch (or SPST for short), and the switch in Figure 21.1(b) is a single-pole double-throw (SPDT) switch. Other common configurations are DPST (two internal switches of the type in Figure 21.1(a), meaning four external connections) and DPDT (two internal switches of the type in Figure 21.1(b), meaning six external connections). Each of the two poles in DPST and DPDT switches is activated by the same external button or lever.

Mechanical switches are distinguished by their current rating. Switches with higher current ratings have larger contact surfaces between the throws and the poles.

Switches can be used in many different ways. For example, attaching a stiff wire to the end of a limit switch as in Figure 21.1(b) could allow you to use the wire as a binary "whisker" sensor.

21.2 Light

21.2.1 Types of Light Sensors

Light sensors include photocells (also called photoresistors), photodiodes, and phototransistors. Photodiodes and phototransistors are used not only to sense light levels directly, but as building blocks in many other types of sensors.

Photocell

A photocell is a resistor that changes resistance depending on the amount of light incident on it. A photocell operates on semiconductor photoconductivity: the energy of photons hitting the semiconductor frees electrons to flow, decreasing the resistance.

An example photocell is the Advanced Photonix PDV-P5002, shown in Figure 21.2. In the dark, this photocell has a resistance of approximately 500 kΩ, and in bright light the resistance drops to approximately 10 kΩ. The PDV-P5002 is sensitive to light in the wavelengths 400-700 nm, approximately the same wavelengths the human eye is responsive to. Figure 21.2 shows a simple circuit illustrating how it can be used as an ambient light sensor feeding either a digital or an analog input to the PIC32.

Photodiode

Photocells are easy to use, but their resistance changes relatively slowly. For example, the PDV-P5002 may take tens of milliseconds to fully change resistance in response to ambient light change. A much faster response can be obtained with a photodiode. As with a photocell, a photodiode operates by photons "kicking up" electrons that allow current to flow, but unlike a photocell, current can flow even without an externally imposed voltage due to the electric

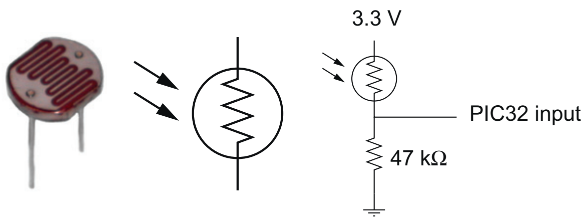

Figure 21.2
(Left) The PDV P5002 photocell. (Image courtesy of Advanced Photonix, Inc., advancedphotonix.com.) (Middle) Circuit symbol for a photocell. (Right) A simple light-level-detection circuit. In bright light, the photocell's resistance is around 10 kΩ, making an output of about 2.7 V. In darkness, the photocell's resistance is around 500 kΩ, making an output of about 0.3 V. The sensor output could go to a PIC32 digital or analog input.

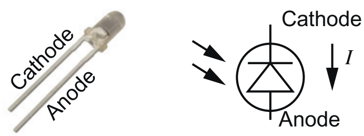

Figure 21.3
(Left) A photodiode. The cathode of a diode is the shorter leg, and the anode is the longer leg. (Right) The circuit symbol for a photodiode, and the direction that photocurrent flows when light hits the photodiode.

field in the diode. In response to a rapidly changing light source, this photocurrent can turn on and off in just a few nanoseconds, depending on the design of the circuit the photodiode is used in.

When light hits the photodiode, reverse photocurrent flows, from the cathode to the anode (Figure 21.3). This current is quite small; for the OPTEK Technology OP906, for example, the maximum current is on the order of tens of microamps. While it may be possible to simply pass this current through a large resistance to generate a measurable voltage, it is common to use an op amp or instrumentation amp circuit to create a sensor with a low-impedance output, a sufficient gain from light levels to voltage, and a fast switching time. It is also common to put a reverse bias voltage across the photodiode to reduce the diode's capacitance, allowing faster current switches. A drawback of the reverse bias voltage is the creation of a reverse *dark current*, in addition to the photocurrent. Thus the reverse bias voltage decreases switching time but reduces the sensitivity of the circuit.

When a photodiode is used without an imposed reverse bias, for maximum sensitivity, it is used in *photovoltaic* mode. When a photodiode is used with an imposed reverse bias, for maximum switching speed, it is used in *photoconductive* mode. This chapter does not cover amplifier circuit designs for these cases; see the references at the end of this chapter.

Some photodiodes come with light filters to adjust their sensitivity to different light wavelengths. The OP906 has no filter, and responds to light at wavelengths between approximately 500 nm and 1100 nm, with a peak response at 880 nm (infrared, invisible). The OP906 can be paired with a Fairchild QED123 LED, which emits IR light at 880 nm, for applications like photointerrupters and reflective object sensors (below).

Photodiodes also come with lenses to direct the incoming light, and the lens on the OP906 ensures that there is little response to light arriving at an angle more than 20 degrees off the central axis of the sensor. Other photodiodes have wider or even narrower viewing angles. Which is best for you depends on your application.

Phototransistor

A phototransistor is a type of bipolar junction transistor including a photodiode junction. An NPN phototransistor has a photodiode at its base-collector junction, and the photocurrent generated there acts as the base current I_B. Below saturation, the phototransistor implements the equations $I_C = \beta I_B$, where I_C is the collector current and β is the transistor's gain, and $I_E = I_C + I_B$, where I_E is the emitter current. (See Appendix B.3 for more on bipolar junction transistors.) Since a typical β is 100, a phototransistor has a higher gain from light to current than a photodiode.

For example, the OSRAM SFH 310 NPN phototransistor creates emitter currents of up to a few milliamps, as compared to the microamps of a photodiode. This higher current makes phototransistors much easier to interface to than photodiodes. See the example circuit in Figure 21.4. A drawback compared to a photodiode is the longer rise and fall times of the current, on the order of 10 µs for the SFH 310.

Like photodiodes, phototransistors may have filters to alter their sensitivity spectrum and lenses to control their viewing angle. The SFH 310 has a viewing angle of up to about 25 degrees off the central axis, and it is sensitive to light of wavelengths 450 nm to 1100 nm, which includes much of the visible spectrum (about 390 nm to 700 nm). Peak sensitivity of the SFH 310 is at 880 nm. The SFH 310 can be paired with the IR LED QED123, mentioned above, with its 880 nm wavelength. If a visible LED is preferred, you could use the Kingbright WP7113SRC/DU red LED at 640 nm (Figure 21.4). While this wavelength is below the SFH 310's 880 nm peak sensitivity, the response is still about 60% of peak.

21.2.2 Basic Applications

Photodiodes and phototransistors are often paired with LEDs to make a variety of different types of sensors. Two of the simplest are photointerrupters and reflective object sensors.

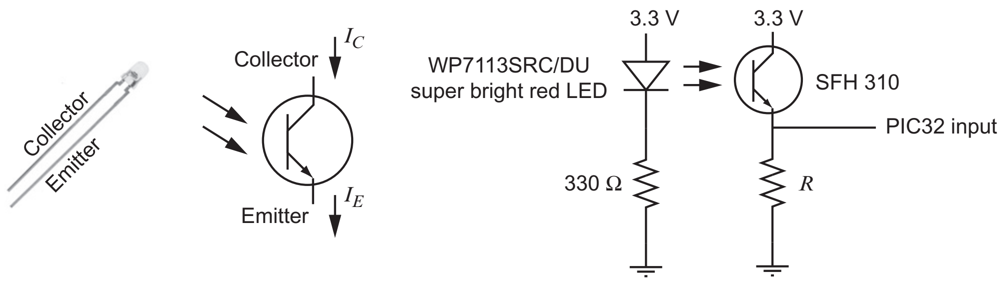

Figure 21.4
(Left) The SFH 310 NPN phototransistor. The shorter leg is the collector and the longer leg is the emitter. (Image courtesy of Digi-Key Electronics, digikey.com.) (Middle) The circuit symbol for an NPN phototransistor. (Right) A circuit with a WP7113SRC/DU red LED illuminating an SFH 310 phototransistor. The resistance R should be chosen to get the right voltage range at the input to the PIC32, which could be an analog or digital input, depending on the application. For a sufficiently large resistance R, the sensor's output voltage ranges from close to 0 V (no light on the phototransistor) to close to 3.15 V (transistor saturated, with 0.15 V drop from collector to emitter).

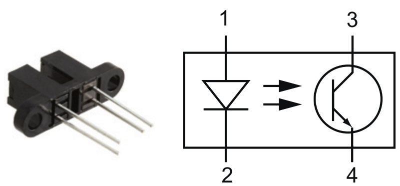

Figure 21.5
(Left) The OPB370T51 photointerrupter. (Image courtesy of Digi-Key Electronics, digikey.com.) (Right) The connections of the package's pins 1 to 4 to the LED and phototransistor.

Photointerrupter

A photointerrupter, or slotted optical switch, contains an LED and a phototransistor or photodiode in a single package. The two are aimed at each other across a small gap, as with the OPTEK Technology OPB370T51, which uses an infrared LED and a phototransistor (Figure 21.5). The LED is always powered, so if the gap is clear, current flows through the phototransistor. If the gap is blocked by an opaque object, blocking the LED light, current through the phototransistor drops. A complete circuit is similar to that illustrated in Figure 21.4(right). An optointerrupter can be used as a type of a limit switch or as a building block for an optical encoder (Section 21.3.2).

To obtain clean digital pulses to a PIC32 input, rather than slowly varying analog voltages as the gap transitions from unblocked to blocked and back, the phototransistor output in Figure 21.4(right) can be passed through a Schmitt trigger (Appendix B.2, Figure B.6).

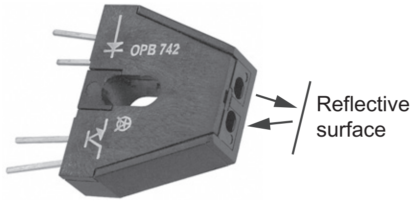

Figure 21.6
The OPB742 reflective object sensor uses an LED and a phototransistor to detect the presence of nearby reflective surfaces. (Image courtesy of Digi-Key Electronics, digikey.com.)

Reflective object sensor

A reflective object sensor is very similar to a photointerrupter, except instead of directly facing each other, the LED and phototransistor are pointed nearly parallel to each other, with a slight inward focus; see the OPTEK Technology OPB742 in Figure 21.6. The OPB742 is designed to detect reflective surfaces at distances between about 0.2 cm and 0.8 cm. When there is no reflective surface nearby, little current flows through the phototransistor; when there is a reflective surface within range, significant current flows through the phototransistor. As with the photointerrupter, a complete circuit is similar to that illustrated in Figure 21.4(right).

21.3 Angle of a Revolute Joint

There are many ways to measure the angle or angular velocity of a joint; here we mention a few of the most common.

21.3.1 Potentiometer

A potentiometer (Appendix B.2), or pot for short, is a variable resistor, typically with a rotating knob that determines the variable resistance (Figure 21.7). A pot has three output terminals: two at either end of the internal resistor, with a fixed resistance between them, and one at the *wiper*. As the knob rotates, the wiper slides over the resistive element, and the resistance between one end of the resistor and the wiper increases smoothly from zero to the max resistance of the pot, while the resistance between the wiper and the other end of the resistor drops smoothly from max resistance to zero. Thus, by putting a voltage across the two ends of the internal resistor, the pot's wiper provides a voltage proportional to the angle of the knob, which can be read by the PIC32's analog input (Figure 21.7(c)).

Pots come in many different styles, distinguished by total resistance across the resistive element; the type of knob or other attachment (e.g., the hollow-shaft pot in Figure 21.7(d)); the number of turns that the pot allows (from less than a single turn to multi-turn); the *taper* of the resistive element, which dictates how the resistance changes with rotation of the knob

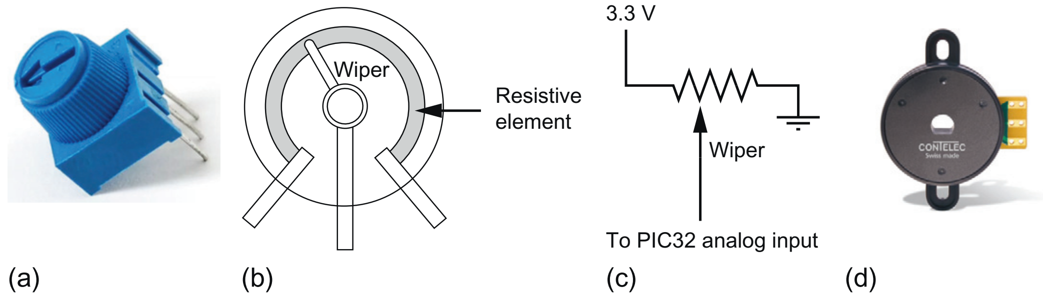

Figure 21.7
(a) A breadboardable 10 kΩ pot. (b) A representation of the wiper sliding over the resistive element as the knob is rotated. (c) The circuit symbol for a pot, and its use in a simple circuit that measures pot rotation as an analog value between 0 and 3.3 V. (d) The Contelec WAL305 hollow-shaft pot. The interior rotating element of this pot can be press-fitted on a motor shaft. (Image courtesy of Contelec AG, www.contelec.ch/en.)

(typically linear or logarithmic, where the latter is often used in audio applications); and the amount of power the resistive element can dissipate without damage. Since a typical pot has a sliding contact between the wiper and the resistive element, pots can only endure a limited number of cycles. More expensive pots have a longer lifetime and a more precise relationship between the rotation of the knob and the variable resistance.

Pots are relatively inexpensive and easy to use, but since they transmit their angle readings as analog voltages, their readings are subject to electrical noise. For applications where electrical noise is an issue, or where more precise angle estimates are needed, encoders, with their digital outputs, are more common choices.

21.3.2 Encoder

There are two major types of encoders: *incremental* and *absolute*.

Incremental encoder

An incremental encoder creates two pulse trains, A and B, as the encoder shaft rotates a *codewheel*. These pulse trains can be created by magnetic field sensors (Hall effect sensors) or light sensors (LEDs and phototransistors or photodiodes). The latter technique, used in *optical encoders*, is illustrated in Figure 21.8. The codewheel could be an opaque material with slots or a transparent material (glass or plastic) with opaque lines.

The relative phase of the A and B pulses determines whether the encoder is rotating clockwise or counterclockwise. A rising edge on B after a rising edge on A means the encoder is rotating one way, and a rising edge on B after a falling edge on A means the encoder is rotating the

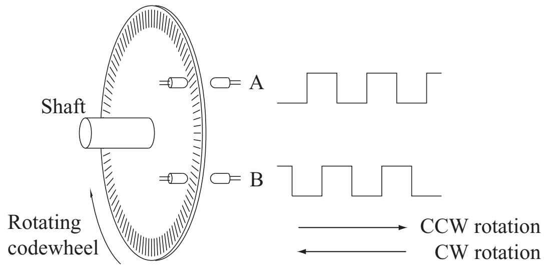

Figure 21.8
A rotating optical encoder creates 90 degree out-of-phase pulse trains on A and B using LEDs and phototransistors. Although two LEDs are shown in the image, it is common to use one LED and a mask to create the two light streams.

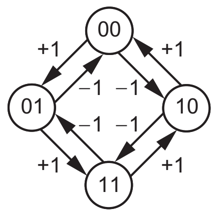

Figure 21.9
4x decoding of A/B quadrature encoder channels. Each node of the state machine shows the digital A/B signals as a two-bit number AB. When the signals change, the encoder count is either incremented or decremented according to the specific transition.

opposite direction. A rising edge on B followed by a falling edge on B (with no change in A) means that the encoder has undergone no net motion. The out-of-phase A and B pulse trains are known as *quadrature* signals.

In addition to determining the rotation direction, the pulses can be counted to determine how far the encoder has rotated. The encoder signals can be "decoded" at 1x, 2x, or 4x resolution, where 1x resolution means that a single count is generated for each full cycle of A and B (e.g., on the rising edge of A), 2x resolution means that two counts are generated for each full cycle (e.g., on the rising and falling edges of A), and 4x means that a count is generated for every rising and falling edge of A and B (four counts per cycle, Figure 21.9). Thus an encoder with "100 lines" or "100 pulses per revolution" can be used to generate up to 400 counts per revolution of the encoder. If the encoder is attached to a motor shaft, and the motor shaft is also attached to a 20:1 speed-reducing gearhead, then the encoder generates $400 \times 20 = 8000$ counts per revolution of the gearhead output shaft.

Some encoders offer a third output channel called the *index* channel, usually labeled I or Z. The index channel creates one pulse per revolution of the joint and can be used to determine when the joint is at a "home" position. Some encoders also offer differential outputs $\bar{A}$, $\bar{B}$, and $\bar{Z}$, which are always opposite A, B, and Z, respectively. This is for noise immunity in electrically noisy environments. For example, if a transient magnetic field induces a voltage change on all of the encoder lines, a *single-ended* reading (e.g., channel A only) may incorrectly interpret the voltage change as movement of the encoder. A differential reading of $A - \bar{A}$ would reject this noise, since it is common to both A and $\bar{A}$.

Some microcontrollers, but not the PIC32, are equipped with a "quad encoder interface" (QEI) peripheral that accepts A and B inputs directly and maintains the encoder count on an internal counter. On the PIC32, the A and B channels can be used with a change notification ISR that implements the state machine of Figure 21.9, provided the A and B lines do not change too quickly. A better solution is to use external encoder decoder circuitry to maintain the count, then query the count using SPI, I^2C, or parallel communication.

Absolute encoder

An incremental encoder can only tell you how far the joint has moved since the encoder was turned on. An absolute encoder can tell you where the joint is at any time, regardless of the position of the joint at power on. To provide absolute position information, an absolute encoder uses many more LED/phototransistor pairs, and each one provides a single bit of information on the joint's position. For example, an absolute encoder with 17 channels, like the Avago Technologies AEAT-9000-1GSH1, can distinguish the absolute orientation of a joint up to a resolution of $360°/(2^{17}) = 0.0027°$ (131,072 unique positions).

As the codewheel rotates, the binary count represented by the 17 channels increments according to *Gray code*, not the typical binary code, so that at each increment, only one of the 17 channels changes signal.[1] This removes the need for the infinite manufacturing precision needed to make two signals switch at exactly the same angle. Compare the following two three-bit sequences, for example (Figure 21.10):

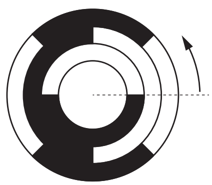

Figure 21.10
A 3-bit Gray code codewheel for an optical absolute encoder. If the innermost ring corresponds to the most significant bit, then as we proceed counterclockwise around the codewheel, the count is 000, 001, 011, 010, 110, 111, 101, and 100.

[1] The AEAT-9000-1GSH1 actually uses a 12-bit Gray code codewheel. The other five bits are obtained by advanced methods not discussed in this chapter.

State	0	1	2	3	4	5	6	7
Binary code	000	001	010	011	100	101	110	111
Gray code	000	001	011	010	110	111	101	100

Absolute encoders typically employ some type of serial communication to send their readings.

21.3.3 Magnetic Encoders

The angle of a rotational joint can be measured by the orientation of the magnetic field due to a magnet attached to the joint. An example is the Avago magnetic encoder sensor illustrated in Figure 21.16 and described in Section 21.6.

21.3.4 Resolver

A resolver consists of three coils: an input excitation coil on the rotor and "sine" and "cosine" measurement coils on the stator. The rotor coil is driven by a sinusoidal reference excitation voltage, $V_r(t) = V_{r,max} \sin \omega t$, where the precise voltage $V_{r,max}$ and frequency ω are not critical, but the frequency $\omega/2\pi$ is typically multiple kHz. The current through the excitation coil generates a magnetic field, which induces a current and therefore a voltage across the stator measurement coils. The sine and cosine coils are offset by 90 degrees relative to each other, so that the voltages induced across the measurement coils are given by

$$V_{\sin}(t) = V \sin \omega t \sin \theta,$$
$$V_{\cos}(t) = V \sin \omega t \cos \theta,$$

where θ is the angle of the rotor coil (see Figure 21.11).

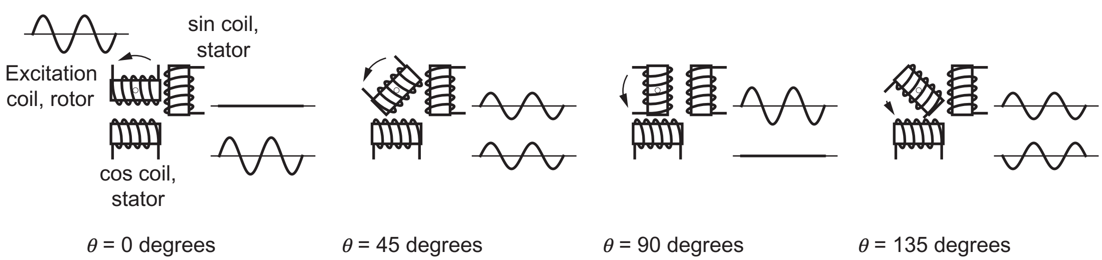

Figure 21.11
A resolver consists of an excitation coil on the rotor and sine and cosine pickup coils on the stator. The excitation coil is driven by a sinusoidal voltage as a function of time. The excitation coil induces sinusoidal voltages across the pickup coils. The amplitude and phase (zero or 180 degrees) of the pickup voltage sinusoids is a function of the angle of the rotor.

The angle of the resolver's rotor is decoded by a resolver-to-digital converter (RDC) chip, such as the Analog Devices AD2S90. The two pickup coil voltages are sent to the RDC, which decodes the angle and provides three kinds of outputs: (a) serial output simulating a 12-bit absolute encoder; (b) A and B outputs simulating a 1024-line incremental encoder (1024 A and B pulses per revolution, allowing 4x decoding for a resolution of 4096 counts per revolution); and (c) an analog voltage proportional to the angular velocity.

21.3.5 Tachometer

A tachometer refers to any device that produces a signal proportional to the speed of rotation of a joint. There are many different types of tachometers, some based on measuring the frequency of, or the time between, pulses generated by the rotating shaft. Any angle-measuring device can be used to simulate a tachometer by numerical differencing, taking angle measurements at times t and $t + \delta t$ and calculating $\dot{\theta}(t + \delta t) \approx (\theta(t + \delta t) - \theta(t))/\delta t$.

21.4 Position of a Prismatic Joint

Linear or *prismatic* joints are the second-most common type of joint, after rotary joints. Often prismatic joints are driven by rotary motors with a transmission that converts rotational motion to linear motion, such as a ball screw or a rack and pinion (Chapter 26). In this case, the linear motion can be indirectly measured by a rotational sensor (e.g., a pot or encoder) on the motor.

In other cases, it is necessary or desirable to measure the linear displacement directly. For these cases, potentiometers and encoders have direct linear analogs. Linear potentiometers are sometimes called *slide pots*, and are common on analog audio equipment. Linear incremental and absolute encoders simply take the codewheel and straighten it out into a line.

Just as a resolver employs an AC excitation signal and induction to determine the angle of a revolute joint, a *linear variable differential transformer* (LVDT) employs induction to determine the position of a prismatic joint (Figure 21.12). A stationary excitation coil, driven by a sinusoidal voltage in the kHz frequency range, is coupled to two stationary pickup coils by a ferromagnetic core that moves with the linear joint. The coupling between the excitation coil and each pickup coil changes with the position of the core. The voltages across the two pickup coils are differenced (hence the "D" in LVDT), and this differenced signal is used to determine the position of the core. When the core is centered, the difference between the two pickup voltages is zero. As the core moves away from the center position, the amplitude of the difference increases, and the phase of the differenced sinusoid, relative to the excitation sinusoid, is zero as the core moves in one direction and 180 degrees as the core moves in the other direction.

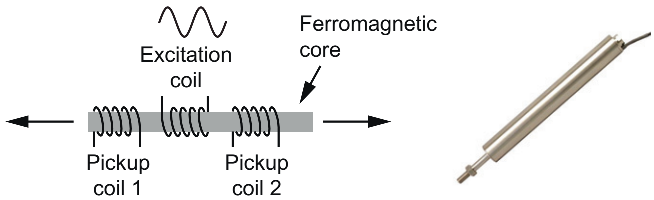

Figure 21.12
(Left) An LVDT consists of a movable core, a stationary excitation coil, and two stationary pickup coils. (Right) An Omega LD320 LVDT. (Image courtesy of Omega Engineering, Inc., omega.com.)

Figure 21.12 shows an Omega LD320 LVDT. An LVDT is typically interfaced to a microcontroller using an LVDT signal conditioning chip such as the Analog Devices AD698, which generates the excitation voltage and turns the differential pickup voltage into an analog voltage output proportional to the position of the LVDT core.

21.5 Acceleration and Angular Velocity: Gyros, Accelerometers, and IMUs

Gyroscopes (gyros) and accelerometers use sensing of inertial forces to measure the angular velocity of a body about one, two, or three axes (gyros) or linear acceleration of a body along one, two , or three axes (accelerometers). A three-axis gyro and a three-axis accelerometer can be used together to make an *inertial measurement unit* (IMU) that attempts to track the motion of a rigid body, based solely on inertial forces. The ability to do this is fundamentally limited by the fact that linear velocity relative to a "fixed" frame cannot be directly measured based on inertial forces—any reference frame translating at a constant velocity is an inertial frame, indistinguishable from other inertial frames.

Gyros and accelerometers have existed for many years, but the advent of microelectromechanical systems (MEMS) has brought these formerly expensive devices within reach of low-cost consumer applications. In this section we focus on MEMS gyros and accelerometers.

21.5.1 MEMS Accelerometer

A MEMS accelerometer measures acceleration (which includes the gravitational acceleration g) by measuring the deflection of a tiny mass m suspended on springs. For example, for a one-axis accelerometer that measures acceleration in a single direction, let $x = 0$ be the rest position of the mass when the acceleration is zero (Figure 21.13). Then the acceleration a of the accelerometer can be calculated from the deflection x using the equation $kx + ma = 0$, where k is the stiffness of the springs. The deflection is typically sensed by the change in capacitance between plates fixed to the mass and plates fixed to the body of the accelerometer.

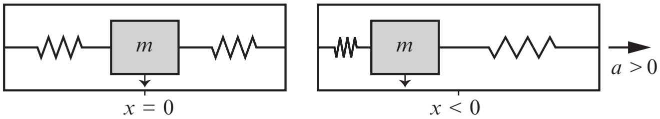

Figure 21.13
(Left) A MEMS accelerometer mass m at its home position $x = 0$ relative to the body of the accelerometer. (Right) An acceleration $a > 0$ leads to a deflection of the mass $x = -ma/k < 0$, where k is the total stiffness of the springs.

Accelerometers come in one-, two-, and three-axis (x, y, z) devices; different ranges of detectable accelerations; and different output types, including I^2C, SPI, and analog output. The Analog Devices ADXL362 is a three-axis accelerometer capable of measuring x-y-z accelerations in the range $\pm 2g$, $\pm 4g$, or $\pm 8g$, as selected by the user, and provides both analog and 12-bit resolution SPI output.

The STMicroelectronics LSM303D accelerometer, discussed in Chapters 12 (SPI) and 13 (I^2C), includes a three-axis accelerometer as well as a three-axis magnetometer. The magnetometer can be used to sense the Earth's magnetic field, and combined with the accelerometer to measure the gravity direction, allows the implementation of a tilt-compensated compass.

21.5.2 MEMS Gyro

Like an accelerometer, a MEMS gyro uses masses supported by springs and capacitive deflection sensors. Unlike an accelerometer, the masses in a gyro are forced to constantly vibrate. Because of this motion, when the gyro is rotated, the masses experience "Coriolis forces" that deflect the nominal vibration relative to a gyro-fixed frame (see Figure 21.14 for an explanation of Coriolis forces). These Coriolis effects are proportional to the rotation rate. The masses and capacitance sensors are physically configured so that the differential capacitance change is zero if deflection is caused by a linear acceleration of the gyro, ensuring that only deflections due to angular velocity are measured (Figure 21.15).

The STMicroelectronics L3GD20H is a three-axis gyro with a user-selectable full-scale range of up to ± 2000 degrees/s in each of the three axes. Data is transmitted by I^2C or SPI.

21.5.3 MEMS IMU

An IMU combines a gyro and an accelerometer, and optionally a magnetometer and/or a barometric pressure sensor. The barometer allows approximate altitude readings while the magnetometer allows sensing the orientation of the Earth's magnetic field.

The primary goal of an IMU is to track the 3D position and orientation of a rigid body in time without using any external references. For example, if the body begins at rest, the

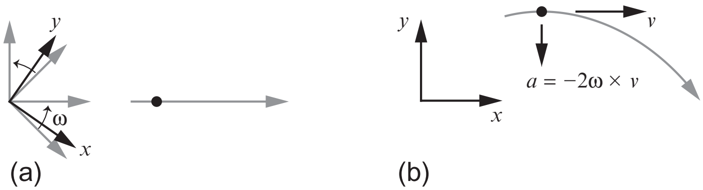

Figure 21.14
The Coriolis effect, illustrated by an *xyz* frame rotating with positive angular velocity about the *z*-axis out of the page. The constant angular velocity is written in vector form as $\omega = (0, 0, \omega_z)$. (a) Viewed in a stationary frame fixed to the page, a point mass *m* moves at a constant velocity to the right while the gray frame rotates. The frame's angle and the mass are shown in black at a specific instant in time. (b) Viewed by an observer in the non-inertial rotating frame, the velocity of the point mass does not appear constant. Instead, the mass follows a spiral trajectory. The mass is shown in black at the same instant as in (a). In the coordinates of the rotating frame, the acceleration of the point mass is $a = -2\omega \times v$, where v is the velocity vector of the mass as viewed from the rotating frame, and ω in the non-inertial frame is the same as in the inertial frame. The "Coriolis force" is simply $ma = -2m\omega \times v$. This is not a true force on the mass; the apparent acceleration of the mass is a result of observing its motion from a rotating frame.

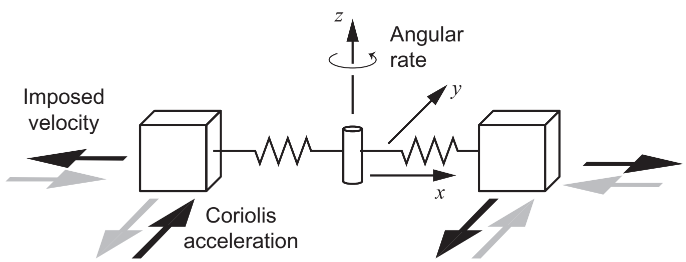

Figure 21.15
A schematic of a one-axis MEMS gyro. The two masses are oscillated opposite each other in the *x*-direction. When one is moving with $\dot{x} > 0$, the other is moving with $\dot{x} < 0$. If the gyro is rotated with a positive angular velocity about the *z*-axis, the masses experience equal and opposite "Coriolis forces" in the $\pm y$ direction. If the two masses are currently moving outward (black arrows), the masses experience accelerations as indicated by the corresponding black acceleration arrows; if the two masses are currently moving inward (gray arrows), the masses experience accelerations in the directions of the corresponding gray acceleration arrows.

accelerometer can use the gravitational field to determine the orientation of the body. There is no way for an IMU to determine the initial *x*-*y*-*z* position of the body,[2] so only relative motion can be estimated. When the body begins to move, it must experience either linear accelerations, angular velocities, or both. Angular velocity can be numerically integrated by

[2] The *z* position can be approximately measured by a barometer.

microcontroller software to estimate the orientation of the body, and linear accelerations can be integrated to get linear velocities and again to get the net position change from the initial position. Sophisticated software integrators exist; they typically employ extended Kalman filters.

Because of sensor errors and errors due to numerical integration, estimates of linear velocity and position tend to accumulate error over time. These errors can be mitigated by occasionally referencing an external reference such as GPS. IMUs allow systems that rely on GPS to continue functioning when GPS is briefly unavailable.

An IMU can consist of a PCB with separate accelerometer and gyro ICs; a single IC incorporating both an accelerometer and a gyro; or a complete integrated solution, including onboard estimation software. An example IMU is the STMicroelectronics ASM330LXH, which features a three-axis accelerometer with a user-selectable full scale between $\pm 2g$ and $\pm 16g$ and a three-axis gyro with a user-selectable full scale between ± 125 and ± 2000 degrees/s. It communicates by I^2C or SPI.

21.6 Magnetic Field Sensing: Hall Effect Sensors

The *Hall effect* is a manifestation of the *Lorentz force law*, which states that a moving charge carrier in a magnetic field experiences a force if the magnetic field flux lines are not aligned with the direction of motion. (The Lorentz force law is discussed in more detail in Chapter 25, as the basis for converting electrical energy to mechanical energy in DC motors.) Current flowing through a flat, stationary semiconductor plate is deflected by this force until there is a charge buildup at the edges of the plate that balances the effect of the Lorentz force. The charge buildup can be measured as a voltage across the plate, transverse to the direction of the current flow. The magnitude of this *Hall voltage* is a function of the strength of the magnetic field and its alignment relative to the current direction.

The Hall effect is used in a great variety of sensors, including 3D magnetic-field-sensing magnetometers (e.g., for a digital compass or for measuring orientation in a known, artificially created magnetic field), current sensors, rotary encoders, and angular position sensors for brushless DC motors (Chapter 29), and sensors that detect proximity to a magnet or ferromagnetic material. To sense whether a piece of metal is nearby, for example, a sensor can be designed consisting of a Hall sensor with a magnet fixed closed by. If the sensor comes close to a piece of metal, the magnet's magnetic field changes, changing the voltage read by the Hall sensor.

The Toshiba TCS20DPR is a digital Hall effect switch IC with three pins, for power (e.g., 3.3 V), GND, and a digital output that reads low when the magnetic flux density exceeds a limit B_{on} and high when the flux density drops below B_{off}, where $B_{\text{on}} > B_{\text{off}}$. If the flux is between B_{off} and B_{on}, the sensor reading does not change from its previous value. This *hysteresis* assures that the output only changes if the magnetic field has changed significantly.

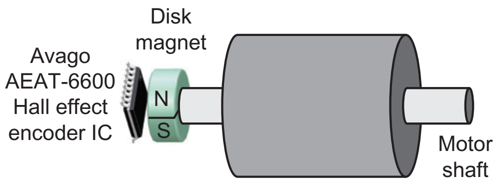

Figure 21.16
Many motors have a shaft extending on both sides of the motor: one side for a sensor element and the other side to connect to a gearhead or the load. Here, the orientation of a disk magnet mounted on one end of the shaft is sensed by the Avago AEAT-6600 magnetic encoder IC. (Image of AEAT-6600 courtesy of Avago Technologies, avagotech.com.)

The values of B_{off} and B_{on} are a few milliTeslas (mT). For reference, the strength of the Earth's magnetic field at the Earth's surface is a bit less than 0.1 mT, while a typical small magnet at a distance of a centimeter might have a field strength on the order of 100 mT. Thus the TCS20DPR could be used to test for the presence of a nearby magnet, perhaps on a moving joint.

Another application of the Hall effect is embodied in the Avago Technologies AEAT-6600 angular magnetic encoder IC (Figure 21.16). A diametrically magnetized two-pole disk magnet is mounted on a rotating shaft directly above the IC. The AEAT-6600 is capable of sensing the orientation of the magnet with 16-bit resolution, or 65,536 unique angles. The orientation can be communicated in several ways: as incremental encoder A/B/I signals; as three Hall sensor signals for brushless DC motors (Chapter 29); as a value encoded in the duty cycle of a PWM signal; or as a 16-bit position using asynchronous serial communication.

21.7 Distance

Inexpensive ultrasonic and infrared ranging sensors can sense the distance to nearby surfaces, from distances of a few centimeters to a few meters. The HC-SR04 ultrasonic sensor and the Sharp GP2Y0A60SZ are two such sensors (Figure 21.17).

The HC-SR04 has four pins: Vcc (powered by 5 V), GND, Trigger input, and Echo output. When the HC-SR04 receives a 10 μs high pulse on the Trigger input, its ultrasonic transmitter emits a brief burst of 40 kHz ultrasonic pulses. The time until the receiver hears the echo, together with the speed of sound (approximately 340 m/s in dry air at room temperature at sea level), is used to determine the distance to the nearby surface. A digital high pulse is created on the Echo output for a time proportional to the distance to the surface, where the distance in centimeters equals approximately the duration of the pulse in microseconds divided by 58.[3] The sensor works best when the reflecting surface is approximately parallel with, and directly in front of, the face of the HC-SR04, and large enough to generate a good reflection. The

3 A pulse of over 30 ms indicates that no reflection was detected.

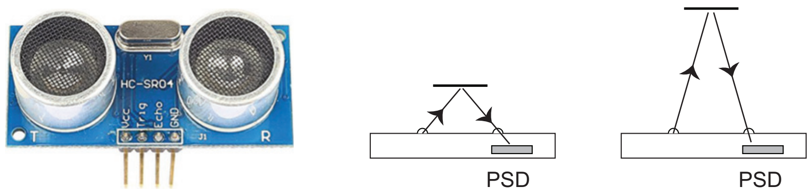

Figure 21.17
(Left) The HC-SR04 ultrasonic distance sensor. (Right) The Sharp GP2Y0A60SZ infrared distance sensor detects distance using a position-sensitive detector to triangulate the distance to the reflective surface.

HC-SR04 may also detect surfaces not parallel to the face of the HC-SR04, and up to 20 degrees off the central axis of the emitter/receiver. Under ideal conditions, the HC-SR04's distance readings can be accurate to up to about 1 cm.

The Sharp GP2Y0A60SZ consists of an infrared beam emitter and a position-sensitive detector (PSD). The emitted IR beam is reflected by the sensed surface and detected by the PSD. The PSD uses the photovoltaic effect, also used by photodiodes and phototransistors, to measure the linear position of the reflected spot of light (see Figure 21.17). The PSD output is returned as an analog voltage, updated at approximately 60 Hz, and distance is calculated by triangulation. The GP2Y0A60SZ is sensitive to ranges between 10 and 150 cm.

21.8 Force

A force sensor is used to sense the force transmitted through a body. For example, a robot arm may have a force-torque sensor between the arm and its gripper, to sense the mass of the object being grasped. Such a sensor actually senses the force in three axes (x, y, and z) as well as the torque about the three axes. Digital scales also employ force sensors.

Forces are commonly measured using *strain gauges*. A strain gauge is a resistor whose resistance changes as it is stretched or compressed. For example, consider a resistive, slightly flexible rod, with current flowing from one end to the other. As the rod is compressed, it becomes shorter and fatter, and resistance decreases. As it is stretched, it becomes longer and thinner, and resistance increases (Figure 21.18).

A typical strain gauge, such as the Vishay strain gauge in Figure 21.18, consists of a metallic foil resistor mounted on a flexible insulating backing. When glued to a (typically metallic) substrate, the change in resistance measures the *strain* (compression or stretching) of the underlying substrate. The stiffness properties of the substrate are then used to estimate the forces that the substrate is experiencing. The substrate is typically quite stiff, as flexibility is generally an undesired property in a force sensor, so strains tend to be quite small. To sense

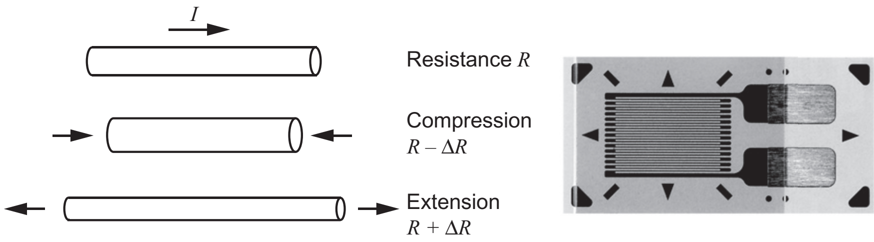

Figure 21.18
(Left) The principle behind a strain gauge, modeled here as a compressible/extensible rod: compression reduces the resistance and extension increases the resistance. (Right) A closeup of a Vishay metallic foil strain gauge. The long and thin resistor is patterned to accentuate the resistance change due to small strains. The large solder pads on the right are the ends of the resistor. (Image courtesy of Micro-Measurements, a brand of Vishay Precision Group, vpgsensors.com.)

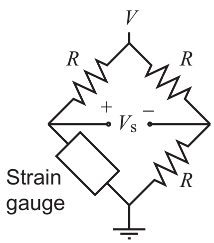

Figure 21.19
A Wheatstone bridge to sense the change in resistance of a strain gauge, whose nominal resistance is R. The small sensed voltage V_s is usually amplified by an instrumentation amplifier.

the small resistance change, a Wheatstone bridge (Figure 21.19) is often used in conjunction with an instrumentation amplifier (Appendix B.4). Because of the differential nature of the sensed voltage V_S, the reading is relatively immune to variations in supply voltage and effects of resistance changes due to temperature.

Working with strain gauges is a bit of an art. For example, forces in different directions will affect the reading of a single strain gauge. For this reason, the design of the shape of the substrate, and therefore its stiffness in different directions, is quite important. It is also common to use multiple strain gauges, for example two strain gauges at right angles to each other, to better identify the direction of the force. Finally, the output of a set of strain gauges must be carefully calibrated by applying known loads to the force sensor and fitting a mapping between sensor readings and applied forces.

Rather than gluing your own strain gauges, you are more likely to buy an integrated *load cell* including the substrate, strain gauge(s), and Wheatstone bridge(s). Load cells come in single-axis and multi-axis versions, up to a full six axes (forces and torques about three

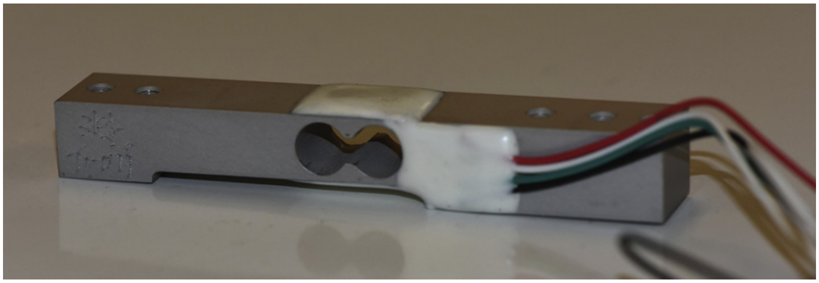

Figure 21.20
A one-axis load cell. The metal substrate is shaped to achieve the desired force sensitivity.

orthogonal axes). Professional load cells can cost several thousand dollars, but inexpensive one- and two-axis load cells are available from manufacturers of load cells for consumer digital scales (Figure 21.20). These load cells typically integrate the Wheatstone bridge but require you to supply your own instrumentation amplifier. Vendors include seeedstudio.com and elane.net.

A much more flexible force-sensitive resistor is the Flex Sensor from Spectra Symbol. Flex Sensors are flexible strips with the resistor printed on one side in conductive ink. As the strip is bent in the other direction, by up to 180 degrees, the resistance increases up to twice its original value.

21.9 Temperature

Thermistors are resistors whose resistances vary significantly with temperature, and they come in two types: NTC (negative temperature coefficient) and PTC (positive temperature coefficient). An example NTC thermistor is the TDK B57164K103J (Figure 21.21) which has a nominal resistance of 10 kΩ at 25° C, rising to 35.6 kΩ at 0° C and dropping to 549 Ω at 100° C.

While the resistance exhibited by the B57164K103J is nonlinear with respect to temperature in Celsius, the Analog Devices TMP37 is designed to produce an analog voltage proportional to the temperature in Celsius. The three pins of the TMP37 in Figure 21.21 are for power (e.g., 3.3 V), GND, and the output voltage, 20 mV/degree Celsius for the range 5-100° C, with a typical accuracy of $\pm 1°$ C.

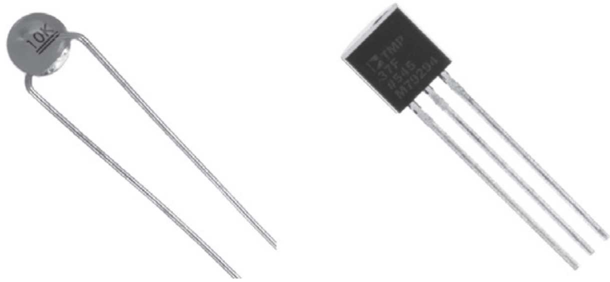

Figure 21.21
(Left) The TDK B57164K103J thermistor. (Right) The Analog Devices TMP37 temperature sensor. (Images courtesy of Digi-Key Electronics, digikey.com.)

The B57164K103J and the TMP37 each cost approximately 1 USD. To measure temperatures down to $-200°$ C and up to $1000°$ C and higher, a more expensive thermocouple and thermocouple amplifier can be used.

21.10 Current

Two common methods for measuring the current flowing through a wire are to (1) use a Hall effect sensor and (2) put a low-resistance resistor in series with the wire and measure the voltage across the resistor. We consider the latter case first.

21.10.1 Current-Sense Resistor and Amplifier

To measure current, a current-sensing resistor can be placed in series with it. Current flowing through this resistor creates a voltage drop across it, which is then measured. To have minimum effect on the current, the sensing resistance should be small. For good accuracy, the resistor should have a tight tolerance on its resistance, and the resistor's power rating should be high enough to allow it to survive the largest current that can flow through it. For example, a 15 mΩ resistor used on a wire that may carry up to 5 A should be rated for at least $(5\text{ A})^2\ 0.015\ \Omega = 0.375$ W to ensure that it will not burn up.

The voltage across a current-sense resistor is intended to be small, e.g., $5\text{ A} \times 0.015\ \Omega = 0.075$ V for 5 A through a 15 mΩ resistor. A specialized current-sense amplifier chip can be used to turn this small signal into a signal usable by a microcontroller. One such chip is the Maxim Integrated MAX9918 current-sense amplifier (Figure 21.22). The voltage across the current-sense resistor is registered by pins RS+ and RS−. The analog output voltage OUT is given by

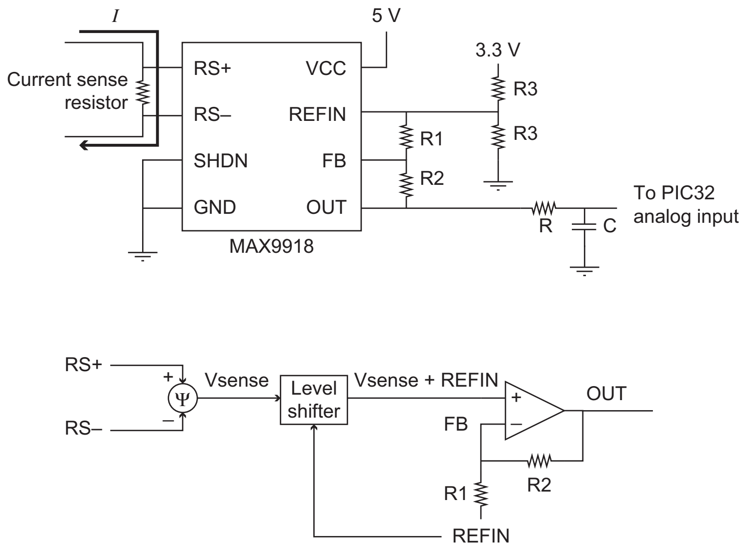

Figure 21.22

Top: Wiring the MAX9918 current-sense amplifier. Bottom: Effective internal circuit, showing how the R1 and R2 resistors are used to set the gain of a noninverting amplifier.

$$\text{OUT} = \text{G} * (\text{RS+} - \text{RS-}) + \text{REFIN}$$

where the gain G is set by the external feedback resistors R1 and R2 as G = 1 + (R2/R1). To implement this equation, the chip uses a noninverting instrumentation amplifier along with a level-shifting circuit (Figure 21.22).

The circuit shows REFIN as 1.65 V, so that zero current through the sense resistor reads as 1.65 V at OUT. This offset voltage allows OUT voltages less than 1.65 V to represent negative currents. The R3 voltage divider resistors feeding the reference voltage REFIN should be relatively small, perhaps a few hundred ohms, to prevent small currents in the feedback resistor network from affecting REFIN.[4] Lowering R3 further would waste power unnecessarily.

The gain G should be chosen so that the maximum expected current gives the maximum voltage at the output. For example, if the maximum expected current is 2 A, then the maximum expected voltage across the 15 mΩ resistor is ±0.03 V. To use the full resolution of the PIC32's ADC, this should map to ±1.65 V, meaning G = 1.65 V/0.03 V = 55. Then a current of 2 A reads as 3.3 V at OUT and a current of −2 A reads as 0 V. The feedback

[4] Ideally the voltage reference to REFIN would be from a lower-impedance source, like a buffered output, but here we are trying to keep the component count down.

resistors R1 and R2 should be relatively high resistance, so as not to load the REFIN voltage divider. See, for example, Exercise 13 of Chapter 27.

One common application of a current sensor is to sense the current flowing through a motor. Since motors are typically driven by a rapidly switching pulse-width modulated voltage (Chapter 27), the current through the motor may also be rapidly changing. We are less interested in this fast variation at the PWM frequency (typically tens of kHz), and more interested in the time-averaged current over several PWM cycles. One way to approximate this time-averaged current is by low-pass RC filtering the sensor output (Appendix B.2.2).[5] A good choice for the RC time constant would give a filter cutoff frequency $f_c = 1/(2\pi RC)$ of a few hundred Hz, approximately 100 times less than a typical PWM frequency. The filter, therefore, attenuates most of the variation due to the PWM without making current sensing overly sluggish.

21.10.2 Hall Effect Current Sensor

The Allegro ACS711 is a Hall effect-based current sensor. The current is passed into the IP+ pins of the IC and back out the IP− pins. The internal conduction path generates a magnetic field which is sensed by the internal Hall effect sensor. At 0 A current, the ACS711 output is half of the supply voltage (e.g., 3.3 V/2 = 1.65 V). The output voltage increases (decreases) proportionally with the positive (negative) current flowing into the ACS711 with a constant of proportionality of 45, 55, 90, or 110 mV/A, depending on the specific ACS711 model number. Like the output of the MAX9918, this analog output can be sent to a PIC32 analog input.

21.11 GPS

For well under 100 USD, you can buy a GPS receiver that listens to satellites in Medium Earth Orbit at an altitude of approximately 20,200 km. GPS satellites carry precise atomic clocks that are regularly synchronized with Earth-based clocks. GPS satellites continuously broadcast their time and their 3D location, as determined in communication with Earth-based stations. If a GPS receiver is able to receive transmissions from at least four satellites, then the difference between the satellites' transmission times and their relative arrival times, along with knowledge of the speed of light, can be used to calculate the receiver's latitude, longitude, and altitude.

Many GPS receivers are on the market. It is common to communicate with GPS receivers via UART or USB.

[5] An even better solution would be a high-impedance input active filter, using an op-amp.

21.12 Exercises

For any of the sensors described in this chapter, or any other sensor, get the data sheet for the sensor and understand the key sensor specifications; design and build a circuit to interface it to your PIC32; create a library consisting of a header file and a C file that gives the user access to the main sensor functions; write a demo program that uses that library; and calibrate or test your sensor under different conditions and report your results to show that the sensor works as expected.

Further Reading

ACS711 hall effect linear current sensor with overcurrent fault output for <100 V isolation appliations data sheet revision 3. (2015). Allegro Microsystems, LLC.

AD2S90 low cost, complete 12-bit resolver-to-digital converter data sheet revision D. (1999). Analog Devices.

ADXL362 micropower, 3-axis, ±2g/±4g/±8g digital output MEMS accelerometer data sheet. (2012). Analog Devices.

AEAT-6600-T16 10 to 16-bit programmable angular magnetic encoder IC data sheet. (2013). Avago Technologies.

AEAT-9000-1GSH1 ultra-precision 17-bit absolute single turn encoder data sheet. (2014). Avago Technologies.

ASM330LXH automotive inertial module: 3D accelerometer and 3D gyroscope data sheet. (2015). STMicroelectronics.

Flex sensor FS data sheet. (2014). Spectra Symbol.

GP2Y0A60SZ0F/GP2Y0A60SZLF distance measuring sensor unit data sheet. Sharp.

Global positioning system standard positioning service performance standard (4th ed.). (2008). Department of Defense, United States of America.

Hambley, A. R. (2000). *Electronics* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Horowitz, P., & Hill, W. (2015). *The art of electronics* (3rd ed.). New York, NY: Cambridge University Press.

Kingbright WP7113SRC/DU data sheet. (2015). Kingbright.

L3GD20H MEMS motion sensor: Three-axis digital output gyroscope data sheet. (2013). STMicroelectronics.

LSM303D ultra compact high performance e-compass 3D accelerometer and 3D magnetometer module. (2012). STMicroelectronics.

MAX9918/MAX9919/MAX9920 −20V to +75v input range, precision uni-/bidirectional, current-sense amplifiers. (2015). Maxim Integrated.

NTC thermistors for temperature measurement series B57164K data sheet. (2013). TDK.

Orozco, L. (2014). *Technical article MS-2624: Optimizing precision photodiode sensor circuit design.* Analog Devices.

PDV-P5002 CdS photoconductive photocells data sheet. (2006). Advanced Photonix, Inc.

PIN silicon photodiode type OP906 product bulletin. (1996). OPTEK Technology, Inc.

QED121, QED122, QED123 plastic infrared light emitting diode data sheet. (2008). Fairchild Semiconductor Corporation.

Reflective object sensor, OPB 740 series data sheet. (2011). OPTEK Technology, Inc.

Silicon NPN phototransistor SFH 310, SFH 310 FA data sheet. (2014). OSRAM Opto Semiconductors.

Slotted optical switch, series OPB370 data sheet. (2012). OPTEK Technology, Inc.

TCS20DPR digital output magnetic sensor data sheet. (2014). Toshiba.

Technical article SBOA061: Designing photodiode amplifier circuits with OPA128. (1994). Texas Instruments.

Technical article TA0343: Everything about STMicroelectronics' 3-axis digital MEMS gyroscopes. (2011). STMicroelectronics.

TMP35/TMP36/TMP37 low voltage temperature sensors data sheet revision F. (2010). Analog Devices.