Linear and Rotational Sensors

Kevin M. Lynch and Michael A. Peshkin
Northwestern University
June 14, 2001

By far the most common motions in mechanical systems are linear translation along a fixed axis and angular rotation about a fixed axis. More complex motions are usually accomplished by composing these simpler motions. In this chapter we provide a summary of some of the many technologies available for sensing linear and rotational motion along a single axis. We have arranged the sensing modalities according to the physical effect exploited to provide the measurement.

1 Contact

The simplest kind of displacement sensor is a mechanical switch which returns one bit of information: touching or not touching. A typical *microswitch* consists of a lever which, when depressed, creates a mechanical contact within the switch which closes an electrical connection (Figure 1). Microswitches may be used as bump sensors for mobile robots, often by attaching a compliant material to the lever (such as a whisker) to protect the robot body from impact with a rigid obstacle. Another popular application of the microswitch in robotics is as a *limit switch*, indicating that a joint has reached the limit of its allowable travel.

Figure 2 shows a typical configuration for a microswitch. The pull-up resistor keeps the signal at +V until the switch closes, sending the signal to ground. As the switch closes, however, a series of micro-impacts may lead to “bounce” in the signal. A “debouncing” circuit may be necessary to clean up the output signal (see section on signal conditioning).

Switches may be designated NO or NC for normally open or normally closed, where “normally” indicates the unactivated or unpressed state of the switch. A switch may also have multiple *poles* (P) and one or two *throws* (T) for each pole. A pole moves as the switch is activated, and the throws are the possible contact points for the pole. Thus a SPDT (single pole double throw) switch switches a single pole from contact with one throw to the other, and a DPST (double pole single throw) switch switches two poles from open to closed circuit (Figure 3).

2 Infrared

Infrared light can be used in a variety of ways to measure linear and rotational displacement. Typically, an infrared light-emitting diode (LED), or *photoemitter*, is used as a source, and an infrared sensitive device is used to detect the emitted light. The detector could be a *photoresistor* or *photocell*, a variable resistor which changes resistance depending on the strength of the incident light (possibly infrared or visible); a *photodiode*, which allows the flow of electrical current in one direction in the presence of infrared light, and otherwise acts as an open circuit; or a *phototransistor*. In a phototransistor, the incident infrared light acts as the base current for the transistor, allowing the flow of collector current proportional to the strength of the received infrared light (up to saturation of the transistor). Circuit symbols for the various elements are shown in Figure 4.

If the emitter and detector are facing each other, they can be used as a beam-breaker, to detect if something passes between. This is called a *photointerrupter* (Figure 5). If the emitter and detector are free to move along the line connecting them, the strength of the received signal can be used to measure the distance separating them. Infrared photodetectors may be sensitive to ambient light, however. To distinguish the photoemitter light from background light, the source can be modulated (i.e., switched on and off at a high frequency), and the detector circuitry designed to respond only to the modulated infrared.

An emitter and detector facing the same direction can be used to roughly measure the distance to a nearby surface by the strength of the returned light reflecting off the surface. This is called a *photoreflector* (Figure 6). Alternatively, such a sensor could be used to detect light absorbing or light reflecting surfaces at a constant distance, as in mobile robot line following. Light polarizing filters can also be used on the emitter and detector so that the detector only recognizes light reflected by a special “optically active” retroreflecting surface.

Photointerrupters and photoreflectors can be bought prepackaged or constructed separately from an infrared LED and a photodiode or phototransistor, after making certain the detector is sensitive to the wavelength produced by the LED.

Photoreflector units are also available with more advanced position sensitive detectors (PSD’s), which report the location of infrared light incident on the sensing surface

(Figure 7). The fixed location of the LED relative to the PSD, as well as the location of the image of the infrared light on the PSD, allows the use of triangulation to determine the distance to the target. Such distance measuring sensors are manufactured by Sharp and Hamamatsu.

2.1 Optical Encoders

An optical encoder uses photointerrupters to convert motion into an electrical pulse train. These electrical pulses “encode” the motion, and the pulses are counted or “decoded” by circuitry to produce the displacement measurement. The motion may be either linear or rotational, but we focus on more common rotary optical encoders.

There are two basic configurations for rotary optical encoders: the *incremental* encoder and the *absolute* encoder. In an incremental encoder, a disk (or codewheel) attached to a rotating shaft spins between two photointerrupters (Figure 8). The disk has a radial pattern of lines, deposited on a clear plastic or glass disk or cut out of an opaque disk, so that as the disk spins, the radial lines alternately pass and block the infrared light to the photodetectors. (Typically there is also a stationary mask, with the same pattern as the rotating codewheel, in the light path from the emitters to the detectors.) This results in pulse trains from each of the photodetectors at a frequency proportional to the angular velocity of the disk. These signals are labeled A and B, and they are 1/4 cycle out of phase with each other. The signals may come from photointerrupters aligned with two separate tracks of lines at different radii on the disk, or they may be generated by the same track, with the photointerrupters placed relative to each other to give out of phase pulse trains.

By counting the number of pulses, and knowing the number of radial lines in the disk, the rotation of the shaft can be measured. The direction of rotation is determined by the phase relationship of the A and B pulse trains, i.e., which signal leads the other. For example, a rising edge of A while B = 1 may indicate counterclockwise rotation, while a rising edge of A while B = 0 indicates clockwise rotation. The two out of phase signals are known as quadrature signals.

Incremental encoders commonly have a third output signal called the index signal, labeled I or Z. The index signal is derived from a separate track yielding a single pulse per revolution of the disk, providing a home signal for absolute orientation.

In practice, multiple photointerrupters can be replaced by a single source and a single array detecting device.

IC decoder chips are available to decode the pulse trains. The inputs to the chip are the A and B signals, and the outputs are one or more pulse trains to be fed into a counter chip. For example, the US Digital LS7083 outputs two pulse trains, one each for clockwise and counterclockwise rotation, which can be sent to the inputs of a 74193 counter chip (Figure 9). Standard decoding methods for the quadrature input are 1X, 2X, and 4X resolution. In 1X resolution, a single count is generated for the rising or falling edge of just one of the pulse trains, so that the total number of encoder counts for a single revolution of the disk is equal to the number of lines in the disk. In 4X resolution, a count is generated for each rising and falling edge of both pulse trains, resulting in four times the angular resolution. An encoder with 1000 lines on the code wheel being decoded at 4X resolution yields an angular resolution of $360^o/(4 \times 1000) = 0.09^o$.

While a *single-ended output* encoder provides the signals A, B, and possibly Z, a *differential output* encoder also provides the complementary outputs A', B', and Z'. Differential outputs, when used with a differential receiver, can increase the electrical noise immunity of the encoder.

A drawback of the incremental encoder is that there is no way to know the absolute position of the shaft at power-up without rotating it until the index pulse is received. Also, if pulses are momentarily garbled due to electrical noise, the estimate of the shaft rotation is lost until the index pulse is received. A solution to these problems is the absolute encoder. An absolute encoder uses k photointerrupters and k code tracks to produce a k-bit binary word uniquely representing 2^k different orientations of the disk, giving an angular resolution of $360^o/2^k$ (Figure 10). Unlike an incremental encoder, an absolute encoder always reports the absolute angle of the encoder.

The radial patterns on the tracks are arranged so that as the encoder rotates in one direction, the binary word increments or decrements according to a binary code. Although natural binary code is a possibility, the Gray code is a more common solution. With natural binary code, incrementing by one may change many or all of the bits, e.g., 7 to 8 in decimal is 0111 to 1000 in natural binary. With the Gray code, only one bit changes as the number increments or decrements, e.g., 7 to 8 in decimal is 0100 to 1100 in Gray code. The rotational uncertainty during a Gray code transition is only a single count, or $360^o/2^k$. With the natural binary code, an infinitesimal misalignment between the lines and the photointerrupters may cause the reading to briefly go from 0111 (7) to 1111 (15) during the transition to 1000 (8).

In general, incremental encoders provide higher resolution at a lower cost, and are the most common choice for many industrial and robotic applications.

3 Resistive

One of the simplest and least expensive ways to measure rotational or linear motion is using a variable resistor called a *potentiometer* or *rheostat*. We focus on rotary potentiometers, or "pots" for short, but the principle of operation is the same in the linear case.

A pot consists of three terminals (Figure 11a and 11b). Two end terminals, call them terminals 1 and 3, connect to either end of a length of resistive material, such as partially conductive plastic, ceramic, or a long thin wire. (For compactness, the long wire is wound around in loops to make a coil, leading to the name *wirewound* potentiometer.)

The other terminal, terminal 2, is connected to a *wiper* which slides over the material as the pot shaft rotates. The total resistance of the pot R_{13} is equal to the sum of the resistance R_{12} between terminal 1 and the wiper and the resistance R_{23} between the wiper and terminal 3. Typically the wiper can rotate from one end of the resistive material ($R_{13} = R_{12}$) to the other ($R_{13} = R_{23}$). If the full motion of the wiper is caused by one revolution of the shaft or less, the pot is called a *single-turn* pot. If the full motion is caused by multiple revolutions, it is called a *multi-turn* pot.

Typically a pot is used by connecting terminal 1 to a voltage V, terminal 3 to ground, and using the voltage at the wiper as a measure of the rotation. The voltage observed at the wiper is $V(R_{23}/R_{13})$ and is a linear function of the rotation of the shaft.

A remarkably simple absolute sensor for a wide range of distances is the string pot or draw-wire sensor (Figure 12). It consists of a string wrapped on a spool, with a potentiometer to monitor rotations of the spool. A return spring keeps the string taut. Lengths up to many meters may be measured, using sensors incorporating multi-turn pots. The same technique is similarly useful for short distances (a few centimeters) using compact single-turn pots and a small spool. Both tolerate misalignment or arc-like motion well. String pots are susceptible to damage to the string in exposed applications, but the sensor element is small and unobtrusive. Manufacturers include RDP Electronics, SpaceAge Control, Inc., and UniMeasure.

Another type of resistive sensor is the flexible bend sensor. Conductive ink between two electrical contacts on a flexible material changes resistance as the material bends and stretches. Used in a voltage divider with a fixed resistor, the analog voltage may be used as a measure of the bend. Such a sensor could be used to detect contact (like a whisker) or as a rough measure of the deformation of a surface to which it is attached.

4 Tilt (Gravity)

A *mercury switch* can be used to provide one bit of information about orientation relative to the gravity vector. A small drop of mercury enclosed in a glass bulb opens or closes the electrical connection between two leads depending on the orientation of the sensor. Several mercury switches at different orientations may be used to get a rough estimate of tilt. The signal from a mercury switch may “bounce” much like the signal from a mechanical contact switch (Figure 2).

An *inclinometer* can be used to measure the amount of tilt. One example is the *electrolytic tilt sensor*. Manufacturers include The Fredericks Company and Spectron Glass. Two-axis models have five parallel rod-like electrodes in a sealed capsule, partially filled with a conductive liquid. Four of the electrodes are at the corners of a square, with one in the middle. Tilting the sensor changes the distribution of current injected via the center electrode in favor of the electrodes which are more deeply immersed.

Tilt sensors may be obtained with liquids of varying viscosity, to minimize sloshing. Because a DC current through the liquid would cause electrolysis and eventually destroy the sensor, AC measurements of conductivity are used. As a result, the support electronics are not trivial.

The liquid conductivity is highly temperature dependent. The support electronics for the tilt sensor must rely on a ratio of conductivity between pairs of rods. Also, although the electrolytic tilt sensor operates over a wide temperature range, it is greatly disturbed by non-uniformities of temperature across the cell.

Another kind of simple inclinometer can be constructed from a rotary potentiometer with a pendulum bob attached. A problem with this solution is that friction may stop the bob’s motion when it is not vertical. A related idea is to use an absolute optical encoder with a pendulum bob. Complete sensors operating on this principle can be purchased with advanced options, such as magnetic damping to reduce overshoot and oscillation. An example is US Digital's 12-bit A2I absolute inclinometer.

Of course, gravity acting on a device is indistinguishable from acceleration. If the steady-state tilt of a device is the measurement of interest, simple signal conditioning should be used to ensure that the readings have settled.

Other more sophisticated tilt sensors include gyroscopes and microelectromechanical (MEMS) devices, which are not discussed here.

5 Capacitive

Capacitance can be used to measure proximity or linear motions on the order of millimeters. The capacitance C of a parallel plate capacitor is given by $C = \varepsilon_r \varepsilon_o A/d$, where ε_r is the relative permittivity of the dielectric between the plates, ε_o is the permittivity of free space, A is the area of overlap of the two plates, and d is the plate separation. As the plates translate in the direction normal to their planes, C is a nonlinear function of the distance d. As the plates translate relative to each other in their planes, C is a linear function of the area of overlap A. Used as proximity sensors, capacitive sensors can detect metallic or non-metallic objects, liquids, or any object with a dielectric constant greater than air.

One common sensing configuration has one plate of the capacitor inside a probe, sealed in an insulator. The external target object forms the other plate of the capacitor, and it must be grounded to the proximity sensor ground. As the sensor approaches the target, the capacitance increases, modifying the oscillation of a detector circuit including the capacitor. This altered oscillation may be used to signal proximity or to obtain a distance measurement.

Manufacturers of capacitive sensors include Cutler-Hammer and RDP Electronics.

6 AC Inductive

6.1 LVDT

The best known AC inductive sensor is the *linear variable differential transformer*, or LVDT. The LVDT is a tube with a plunger, the displacement of the plunger being the variable to be measured. (Figure 13) The tube is wrapped with at least two coils, an excitation coil and a pickup coil. An AC current (typically 1 KHz) is passed through the excitation coil, and an AC signal is detected from the pickup coil and compared in magnitude and in phase (0 or 180 degrees) to the excitation current. Support electronics are needed for the demodulation, which is called synchronous detection. The plunger carries a ferromagnetic slug, which enhances the magnetic coupling from the excitation coil to the pickup coil. Depending on the position of the slug within the pickup coil, the detected signal may be zero (when the ferrite slug is centered in the pickup coil), or increasing in amplitude and one or the other phase, depending on displacement of the slug.

LVDT's are a highly evolved technology and can be very accurate, in some cases to the micron level. They have displacement ranges of millimeters up to a meter. They do not tolerate misalignment or non-linear motion, as a string pot does.

6.2 Resolvers

A *resolver* provides a measure of shaft angle, typically with sine and cosine analog outputs. It uses an AC magnetic technique similar to the LVDT, and similar support electronics to provide synchronous detection. Resolvers are very rugged and for this reason are often preferred over optical encoders on motor shafts, although they are not as accurate and they have greater support electronics requirements. Some resolver drives have extra outputs as if they were incremental encoders, for compatibility. Additionally, resolvers provide an absolute measure of shaft angle. The resolver, like the LVDT, is a well established and evolved technology.

7 DC Magnetic

A magnetic field acting on moving electrons (e.g., a current in a semiconductor) produces a sideways force on the electrons, and this force can be detected as a voltage perpendicular to the current. The effect is small, even in semiconductors, but has become the basis of a class of very rugged, inexpensive, and versatile sensors.

7.1 Hall Effect Switches

Hall effect switches refers to devices which produce a binary output, depending on whether the magnetic field intensity exceeds a threshold or not. In their component form, these switches may be packaged as 3-terminal devices the size of a transistor package (TO-92) or surface mount, having only a power lead (3-24 volts), a ground lead, and an output lead. Typically the output is pulled to ground, or not, depending on the magnetic state. Hall effect switches are also available in environmental packages of all sorts.

The actuation threshold ranges from a few gauss (the Earth's magnetic field is 1/2 gauss) up to the hundreds of gauss levels typical of permanent magnets. Often there is a fair degree of unit-to-unit variability in threshold. Hall effect switches are hysteretic: their "turn on" threshold may be different than the "turn off" threshold. Sometimes hysteresis is used to make a switch latching, so that it stays in its last state (on or off) until the applied magnetic field is reversed. Non-latching Hall switches may be unipolar (responding only to one orientation of magnetic field) or bipolar (responding to a field of either polarity). Turn-on and turn-off times are in microseconds.

Hall switches have wide operating temperature ranges and are often used in automobile engine compartments. Another advantage is that they are not susceptible to most of the fouling mechanisms of optical or mechanical switches, such as liquids or dirt. While often the moving part that is detected is a magnet, it can also be arranged that a stationary "bias" magnet is intensified in its effect on the hall switch by the approach of a ferrous part, such as a gear tooth, thus allowing non-magnetized objects to be detected (Figure 14).

Typical applications are the detection of a moving part, replacing a mechanical limit switch. The Hall switch has no moving or exposed parts and is wear-free. Another common use is in indexing of rotational or translational motion. The Hall switch is installed to detect one position, and its output pulse is used as a reference for an incremental encoder which can count distance from that reference point. Hall switches are inexpensive and small so a number of them can be spaced at intervals of millimeters, forming a low-resolution linear or rotational encoder or multi-position switch. Such an encoder or switch has the ruggedness advantages of Hall switches.

7.2 Analog Hall Sensors

In a package the same small size as Hall switches, and costing little more, one can also get Hall devices that have an analog output proportional to magnetic field strength (Figure 15). Typically these have full-scale magnetic field sensitivity in the 100 gauss range. These are not useful as a compass in the Earth's sub-gauss magnetic field.

Hall sensors are useful as linear or rotational encoders. Two Hall sensors may be arranged at right angles to detect the sine and cosine of the angle of a rotating magnet, thus forming an absolute rotation sensor. Commercially available devices of this nature are called "Hall potentiometers" and have a variety of outputs (e.g. sine and cosine, or a linear ramp repeating with each revolution). In contrast to potentiometers with resistive strips and sliders, Hall pots allow continuous 360 degree rotation and experience no wear. All Hall effect devices are susceptible to external magnetic fields however.

Hall sensors are also excellent transducers of short linear or arc-like motions. The motion of a bar magnet past a Hall sensor exposes the sensor to a magnetic field which can be arranged to vary linearly with displacement, over a range of several millimeters up to several centimeters. (The bar magnet travels less than its own length.) Commercial implementations are known as throttle position sensors.

7.3 Tape-Based Sensors

There are a number of linear and rotational sensors, both incremental and absolute, which are similar to optical encoders but use magnetic patterns rather than optical ones. Linear applications are likely to require an exposed strip. In exposed applications, magnetic sensors have advantages in resistance to dirt, although the magnetic stripes must be protected from damage.

8 Ultrasonic

Ultrasonic (US) sensors use the time-of-flight of a pulse of ultrasonic sound through air or liquid to measure distance. Sensors are available with ranges from a few centimeters to ten meters. A great advantage of US sensors is that all of the sensor's electronic and transducer components are in one location, out of harm's way. The corresponding disadvantages are that US sensors tend to be indiscriminate: they may detect spurious targets, even very small ones, especially if these are near the transducer. Sensors are available with carefully shaped beams (down to 7 degrees) to minimize detection of spurious targets. Some include compensation for variation in air temperature, which affects sound velocity. US sensors can be used in surprising geometries. For instance, they can be used to detect the liquid level in a vertical pipe; back-reflection of sound pulses from the walls of a smooth pipe are minimal.

There is also an inexpensive and easily interfaced US sensor from Polaroid, derived from a ranging device for an instant camera, which is popular with experimenters.

Ultrasonic sensors typically have an analog output proportional to distance to target. Accuracies of the 1% level can be expected in a well controlled environment.

9 Magnetostrictive Time-of-Flight

A more accurate technique for using time-of-flight to infer distance is the *magnetostrictive wire transducer* (MTS). A moving magnet forms the "target" corresponding to the acoustic target in US sensors, and need not touch the magnetrostrictive wire which is the heart of the device. The magnet's field acting on the magnetrostrictive wire creates an ultrasonic pulse in the wire when a current pulse is passed through the wire. The time interval from the current pulse to the detection of the ultrasonic pulse at the end of the wire is used to determine the position of the magnet along the wire (Figure 16).

The magnetostrictive transducer does not have the inherent compactness and ruggedness of ultrasonic through air, but does achieve similarly large measurement lengths, up to several meters. Accuracy and stability are excellent, far better than ultrasonic. Some misalignment or non-linear motion is tolerated, because the target magnet does not need to be in very close proximity to the magnetostrictive wire.

10 Laser Interferometry

Laser interferometers are capable of measuring incremental linear motions with resolution on the order of nanometers. In an interferometer, collimated laser light passes through a beam-splitter, sending the light energy on two different paths. One path is directly reflected to the detector, such as an optical sensing array, giving a flight path of fixed length. The other path reflects back to the detector from a retroreflector (mirror) attached to the target to be measured. The two beams constructively or destructively interfere with each other at the detector, creating a pattern of light and dark fringes. The interference pattern can be interpreted to find the phase relationship between the two beams, which depends on the relative lengths of the two paths, and therefore the distance to the moving target. As the target moves, the pattern repeats when the length of the variable path changes by the wavelength of the laser. Thus the laser interferometer is inherently an incremental measuring device.

Laser interferometers are easily the most expensive sensors discussed in this chapter. They also have the highest resolution. Laser interferometers are very sensitive to mechanical misalignment and vibrations.

More information about sensors may be found in Sensors magazine (http://www.sensorsmag.com/).

References

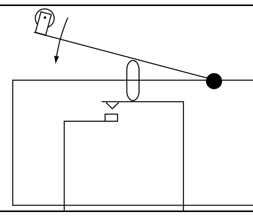
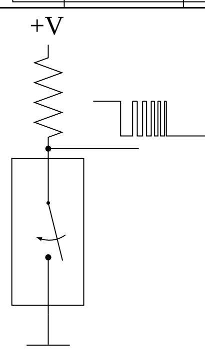
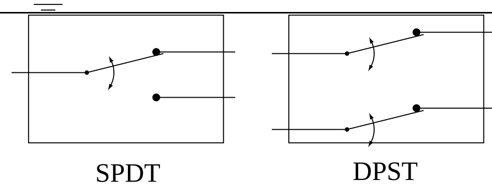
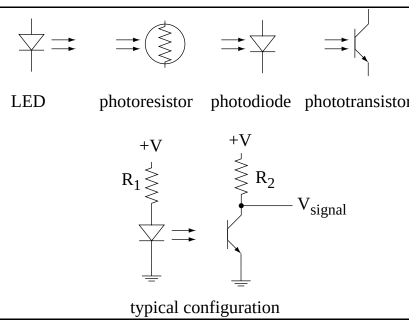
[1] Histand, M. B., and Alciatore, D. G. *Introduction to Mechatronics and Measurement Systems*, McGraw-Hill, Boston, MA, 1999.

[2] Bolton, W. *Mechatronics, Second Edition*, Addison Wesley Longman, New York, NY, 1999.

[3] Horowitz, P., and Hill, W. *The Art of Electronics, Second Edition*, Cambridge University Press, Cambridge, UK, 1998.

[4] Auslander, D. M., and Kempf, C. J. *Mechatronics: Mechanical System Interfacing*, Prentice-Hall, Inc., Upper Saddle River, NJ, 1996.

[5] Jones, J. L., Flynn, A. M., and Seiger, B. A. *Mobile Robots: Inspiration to Implementation, Second Edition*, A. K. Peters, Boston, MA, 1999.

<table>
<tr><td>Figure 1. A typical microswitch.</td><td></td></tr>
<tr><td>Figure 2. Signal bounce at a closing switch.</td><td>
</td></tr>
<tr><td>Figure 3. SPDT and DPST switch configurations.</td><td>
</td></tr>
<tr><td>Figure 4. Optoelectronic circuit symbols and a typical emitter/detector configuration.</td><td>
</td></tr>
</table>

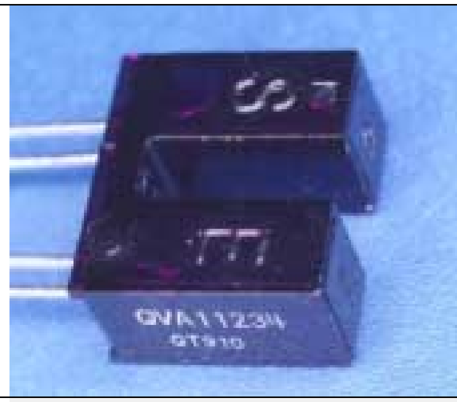


Figure 5. The Fairchild Semiconductor QVA11234 photointerrupter.

Figure 6. The Fairchild Semiconductor QRB1114 photoreflective sensor

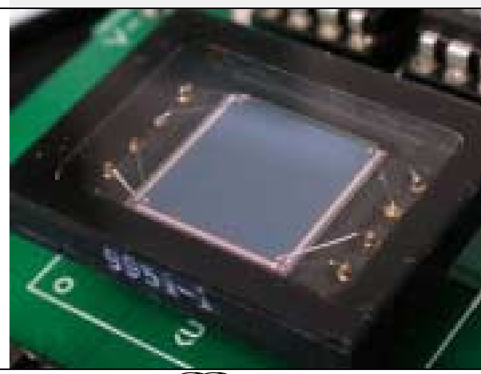


Figure 7. A position sensitive detector (PSD), UDT Sensors, Inc.

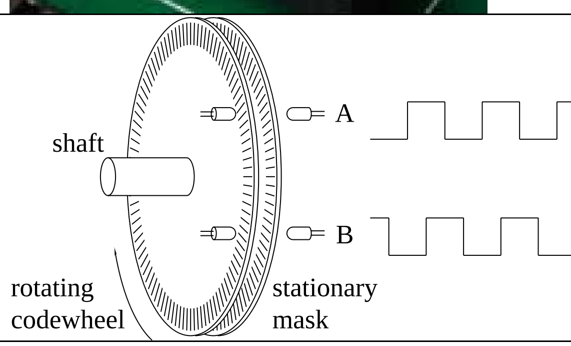


Figure 8. Schematic of an incremental encoder.

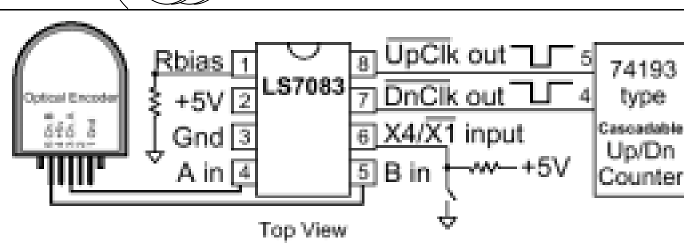
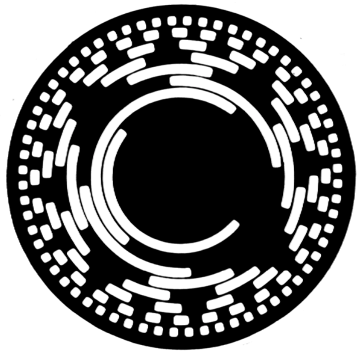
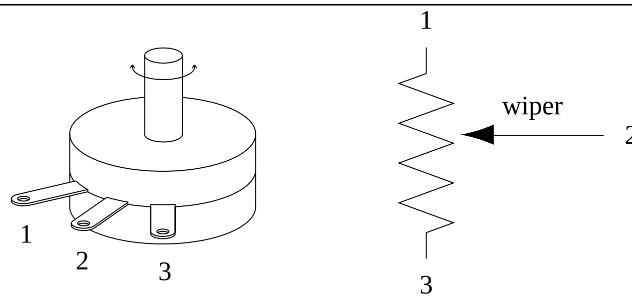
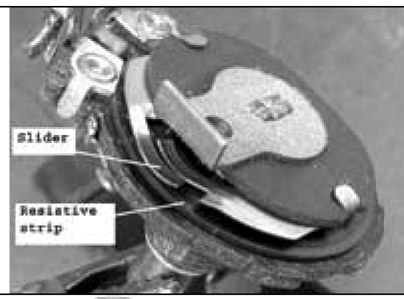
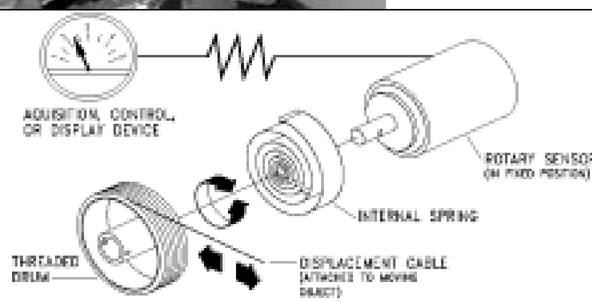
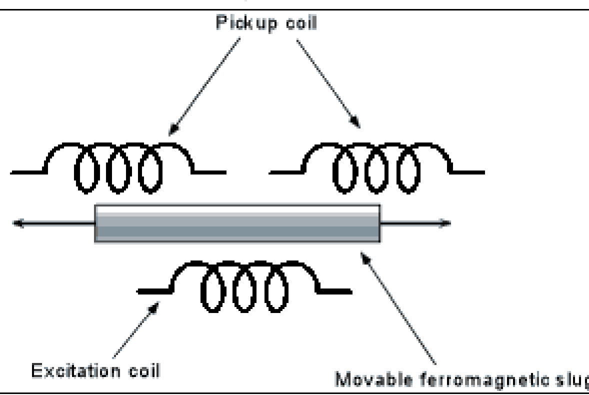
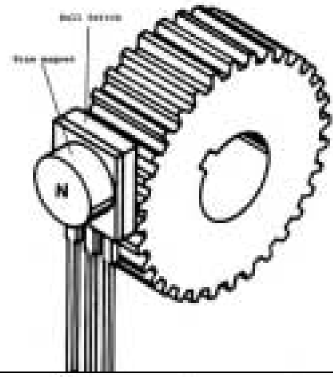
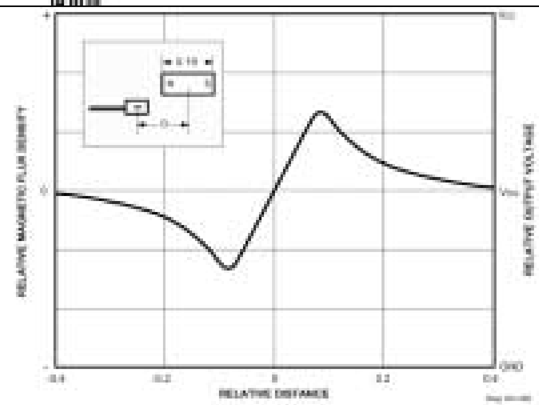
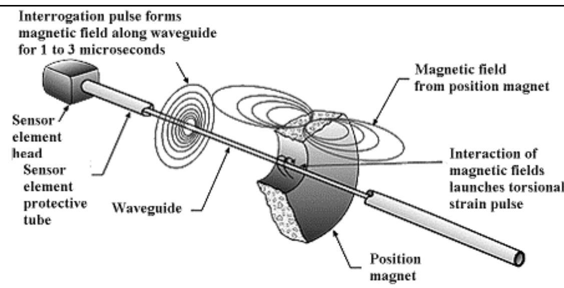


Figure 9. An optical encoder, US Digital LS7083 quadrature decoder chip, and counter (courtesy of US Digital, Inc.).

Figure 10. An 8-bit Gray code absolute encoder disk, courtesy of BEI Technologies Industrial Encoder Division.	
Figure 11a. As the shaft of the potentiometer rotates, the wiper moves from end of the resistive material to the other.	
Figure 11b. Inside of a typical potentiometer, showing wiper contacting resistive strip.	
Figure 12. A string pot, courtesy of Space Age Control, Inc.	
Figure 13. Operating principle of an LVDT.	

<table>
<tr><td>Figure 14. Detecting gear teeth in a ferrous material using a Hall switch and a bias magnet. Courtesy of Allegro Microsystems, Inc.</td><td>
</td></tr>
<tr><td>Figure 15. Output of an analog Hall sensor vs. position relative to a magnet. Courtesy of Allegro Microsystems, Inc.</td><td>
</td></tr>
<tr><td>Figure 16. Principle of operation of a magnetostrictive linear position sensor, courtesy of Temposonics, Inc.</td><td>
</td></tr>
</table>