Chapter 2 Sensors

2.1 Sensor Performance Specifications and Calibration

Sensor definition: A device that produces a signal related to the quantity being measured. A sensor is a system whose input is the quantity being measured and whose output is typically a voltage.

Sensor performance specifications can be broken into two classes, static and dynamic. Static performance specifications apply to steady-state measurement conditions (meaning the input is constant or changes slowly enough that the sensor can reach steady-state). Dynamic performance specifications apply when the quantity being measured varies over time.

Static Performance Specifications:

These include: range, full scale, resolution, repeatability, accuracy, sensitivity, hysteresis, linearity, and deadband.

Range: The limits between which the quantity being measured can vary.
Units: quantity being measured.
Example: a thermometer has a range of -20 to 100 °C.

Full scale: Difference between the maximum and minimum values that the sensor can measure.
Units: quantity being measured.

Resolution: Smallest change in the quantity being measured that is detectable by the sensor.
Units: quantity being measured.
Comments: This specification is easy to measure but does not truly indicate the quality of the sensor.

Repeatability: The ability of a sensor to produce the same output for repeated applications of the same input. Commonly defined as $\pm 3\sigma_y$, where σ_y is the standard deviation of the calibrated sensor output in measured quantity units. Note that this corresponds to the 99.7% confidence interval. Sometimes repeatability is called "precision".
Units: Either measured quantity, or % of full scale.
Comments: The repeatability specification measures the random errors of the sensor. These random errors are usually normally distributed and are due to electrical, thermal and mechanical noise.

Accuracy: The maximum extent the output of the sensor may be incorrect over its range. This includes both random and deterministic errors. The deterministic error may be reduced by proper calibration. In equation form we have:

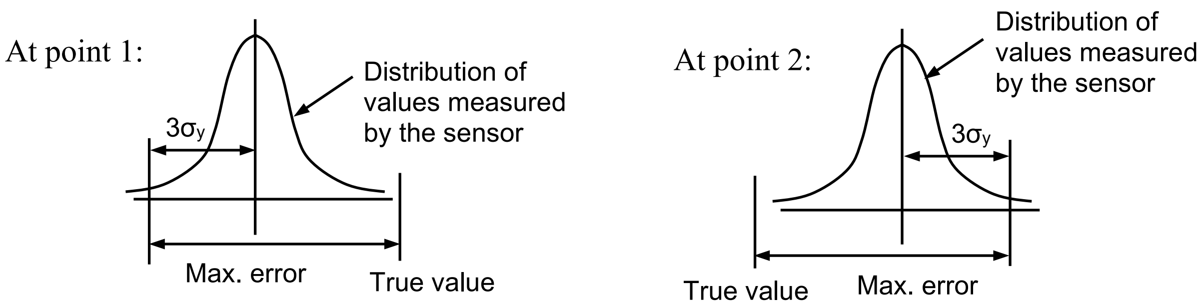
$$\text{Accuracy} = \pm\left(\max(\text{abs}(Y_{true} - Y_{sensor})) + 3\sigma_y\right) \tag{2.1}$$

where Y_{true} are the true values of the quantity being measured, Y_{sensor} are the mean values of the calibrated output of the sensor in measured quantity units, and σ_y is the standard deviation of the calibrated sensor output in measured quantity units.

Units: Either measured quantity, or % of full scale.

Comments: Accuracy should indicate the worst case error possible over the range. To find the accuracy, the distribution of the values measured by the sensor must be determined for a set of points spanning the range, and compared with the corresponding true values.

Example: At two points within the range we might see:



Sensitivity: Ratio of the magnitude of the sensor output to the magnitude of the quantity being measured.

Units: (sensor output)/(quantity being measured).

Example: the force sensor has a sensitivity of 15 mV/ N.

Cross-sensitivity: With directional sensors, the sensitivity of the sensor output to inputs orthogonal to the measurement direction.

Example: Say we want the sensor to measure the force in the X direction, the cross-sensitivities would be the ratios of the sensor output to the forces in the Y and Z directions.

Comments: We would like a sensor to have a large sensitivity and a small cross-sensitivity.

Deadband: The band of input values for which there is zero output.

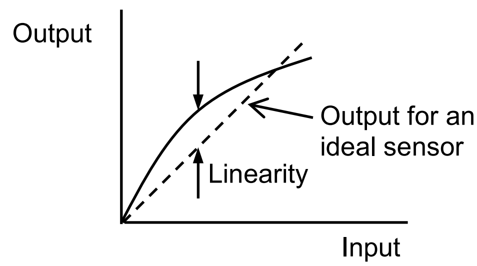
Units: quantity being measured.

Example: If the weight is increased in 1 lb increments and the smallest weight your bathroom scale responds to is 20 lbs, then its deadband is 19 lbs.

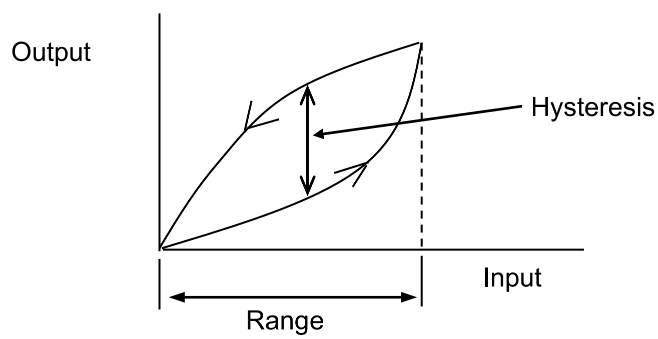
Linearity: This is typically defined as the maximum error between the mean values of the sensors outputs and the true values of the quantity being measured. It is related to accuracy but does not include the random component of the error. In equation form:

$$\text{Linearity} = \pm\left(\max\left(\text{abs}(Y_{true} - Y_{sensor})\right)\right) \tag{2.2}$$

Units: Either quantity being measured, or % of full scale.
Comments: Linearity indicates the worst case deterministic error of the sensor after calibration.
Example:



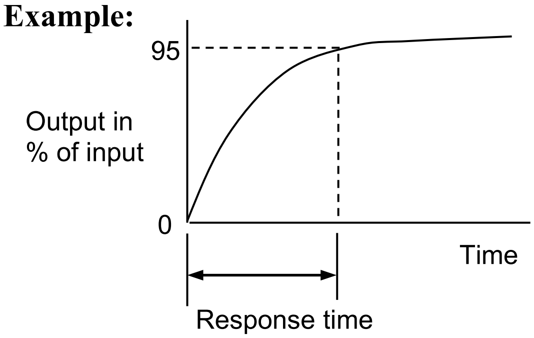
Hysteresis: The maximum difference between the calibrated sensor outputs for continuously increasing and continuously decreasing inputs.
Units: Either quantity being measured or % of full scale.
Example: For clarity the hysteresis has been greatly exaggerated in the figure below. The arrows on the curves indicate the input increasing and input decreasing directions.



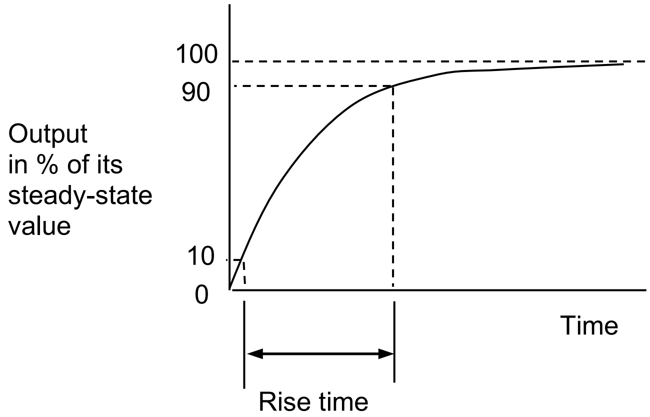
Dynamic Performance Specifications:

These include: response time, rise time, settling time, time constant, stability, bandwidth, lower frequency limit and upper frequency limit.

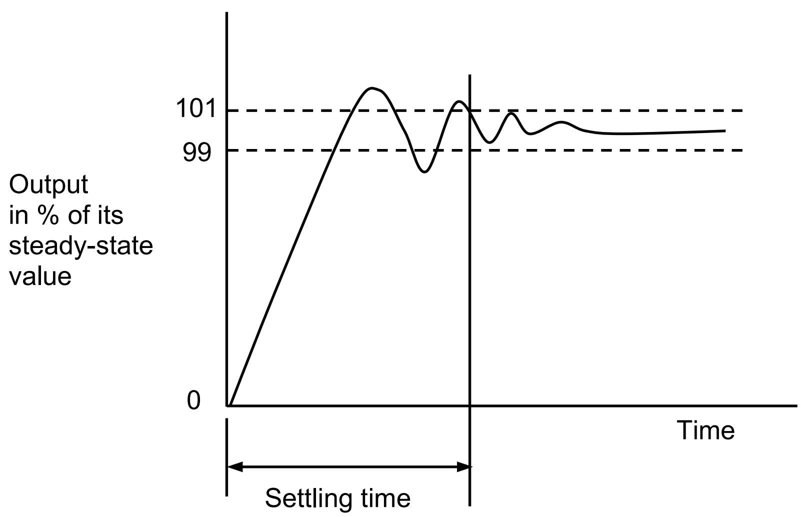
Response time: The time the output takes to reach 95% of the input when the input is a step.
Units: Seconds or milliseconds.
Example:



Rise time: Time required for the output to increase from 10% to 90% of its steady-state value for a step input.
Units: Seconds or milliseconds.
Example:



Settling time: The time required for the output to settle within ±1% of its steady-state value for a step input.
Units: Seconds or milliseconds.
Example:



Time constant: The time required for the output to reach 63.2% of its steady-state value for a step input. This specification assumes the sensor has a first order transfer function.
Units: Seconds or milliseconds.

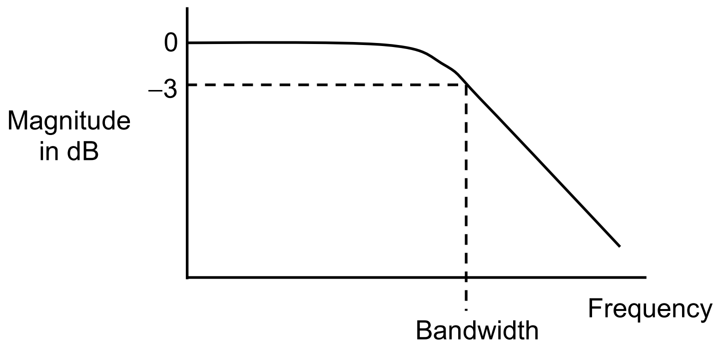
Stability: The ability of a sensor to produce the same output when measuring a constant input over an extended period of time. A change in the output under these conditions is termed "drift".
Units: (quantity being measured OR % of full scale) / (time unit).
Examples: 1% / hour, 10 mV / hour.

Bandwidth: The lowest frequency at which the magnitude of the sensor transfer function drops by 3 dB.

Units: Hz.

Comments: This is more useful specification than the time based specification since it applies to any dynamic input and not just a step. However, performing a frequency response test is much more difficult than a step input test, so this specification is less commonly seen.

Example:



Lower frequency limit: The frequency response of some sensors, such as those using piezoelectric crystals, drops off at low frequencies. The lower frequency limit is the lowest frequency at which the magnitude of the sensor transfer function equals –L dB.

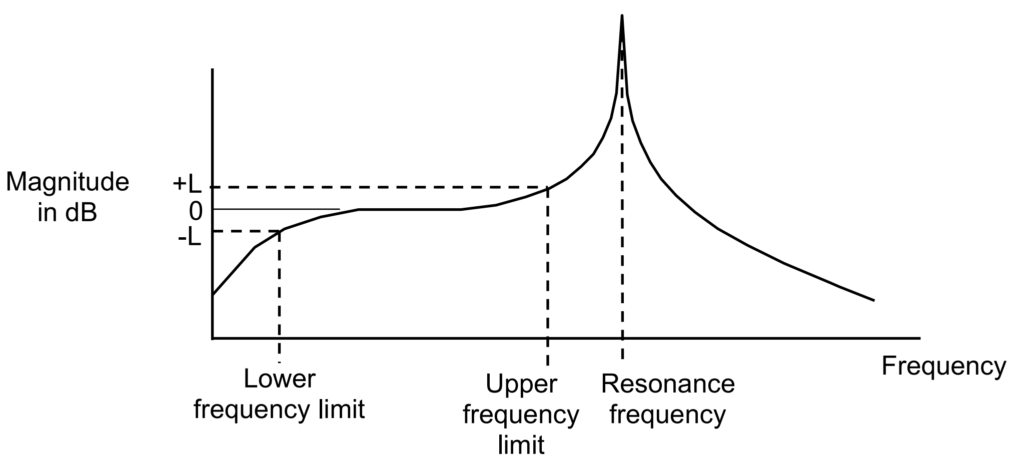
Upper frequency limit: With some sensors, the frequency response curve exhibits a peak due to resonance. This commonly occurs with accelerometers, for example. The upper frequency limit is the lowest frequency at which the magnitude of the sensor transfer function equals +L dB.

Units: Hz

Comments:

These are the least standardized of the specifications. The value L=3 dB is used, but lower values are also common.

Example:



Discussion Topic 1

When discussing specifications, the adjective “high” is unclear. For example, “high sensitivity” means we want a large magnitude for the sensitivity, while “high accuracy” means we want a small magnitude for the accuracy. So it better to talk in terms of large or small magnitudes.

So what specification magnitudes do we want for an ideal sensor?

range:
full scale:
resolution:
repeatability:
accuracy:
sensitivity:
cross-sensitivity:
hysteresis:
deadband:
linearity:
response time:
rise time:
settling time:
time constant:
bandwidth:
lower frequency limit:
upper frequency limit:
stability:

Discussion Topic 2

The relative importance of each sensor specification depends on the application.

For these applications, which specifications are important, and which are unimportant?

bathroom scale:

airbag sensor (accelerometer):

gas tank level sensor:

Sensor Calibration

Sensor calibration typically involves:

(1) Applying a constant input at the start of the range of the sensor

(2) Measuring the sensor output multiple times (at least 100).

(3) Calculating the mean value and standard deviation of the measured outputs.

(4) Increasing the input and then holding it constant.

(5) Repeat steps (2) and (3).

(6) Repeat steps (4) and (5) until at the input is at the end of the range of the sensor.

(7) Fitting a calibration line to the mean values of the measured outputs. Normally, linear regression is used. The input is the independent variable, x, and the sensor output (usually in Volts) is the dependent variable, y. If the line is of the form $y=Ax$ the slope is given by:

$$A = \frac{\sum xy}{\sum x^2} \tag{2.3}$$

A line of the form $y=Ax+B$ may also be used. In this case:

$$A = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{n\sum x^2 - \left(\sum x\right)^2} \tag{2.4}$$

$$B = \left(\sum y - A\sum x\right)/n \tag{2.5}$$

where n is the number of points.

(8) Implementing the calibration. For the $y=Ax$ line the calibrated output in measured quantity units is:

$$Y_{sensor} = \frac{Y_{volts}}{A} \tag{2.6}$$

For the $y=Ax+B$ line the calibrated output in measured quantity units is:

$$Y_{sensor} = \frac{\left(Y_{volts} - B\right)}{A} \tag{2.7}$$

(9) The calibrated sensor output and the maximum of the standard deviations found in step (3) may be used to calculate the various static specifications.

Example 2.1: Sensor Calibration & Specifications

(a) The data given in Table 2.0 was collected from a force sensor for calibration purposes. The tabulated values are the mean values. The standard deviation of the output was not affected by the input, and equaled 2 mV. We want to fit a calibration line by linear regression of the form $y=Ax$ and determine values for:

i) Accuracy
ii) Sensitivity
iii) Linearity
iv) Repeatability

Input (N)	**Output (mV)**
0	0.00
2.5	3.10
5	12.47
7.5	28.12
10	50.05
12.5	78.17
15	112.49
17.5	153.17
20	200.04

Table 2.0 Calibration data.

(b) Additional tests were performed with the input increased from 0 to 20 N and then decreased from 20 N back to 0. If the maximum deviation between these load increasing and load decreasing results was 19 mV, what is the value of this sensor's hysteresis?

(c) The mean output values from tests performed with small inputs are listed in Table 2.1. Based on this data, what is this sensor's deadband?

Input (N)	**Output (mV)**
0 to 2	0.00
2.1	0.00
2.2	0.00
2.3	0.00
2.4	2.91

Table 2.1 Test results for small inputs.

Solution

(a) Fitting a line of the form $y=Ax$ using equation (2.3) with the input as the independent variable and the sensor output as the dependent variable gives:

$$A = \frac{\sum xy}{\sum x^2} = \frac{10127 \text{ Nmv}}{1275 \text{ N}^2} = 7.94 \frac{mV}{N}$$

This calibration line is plotted with the raw data in Figure 2.1

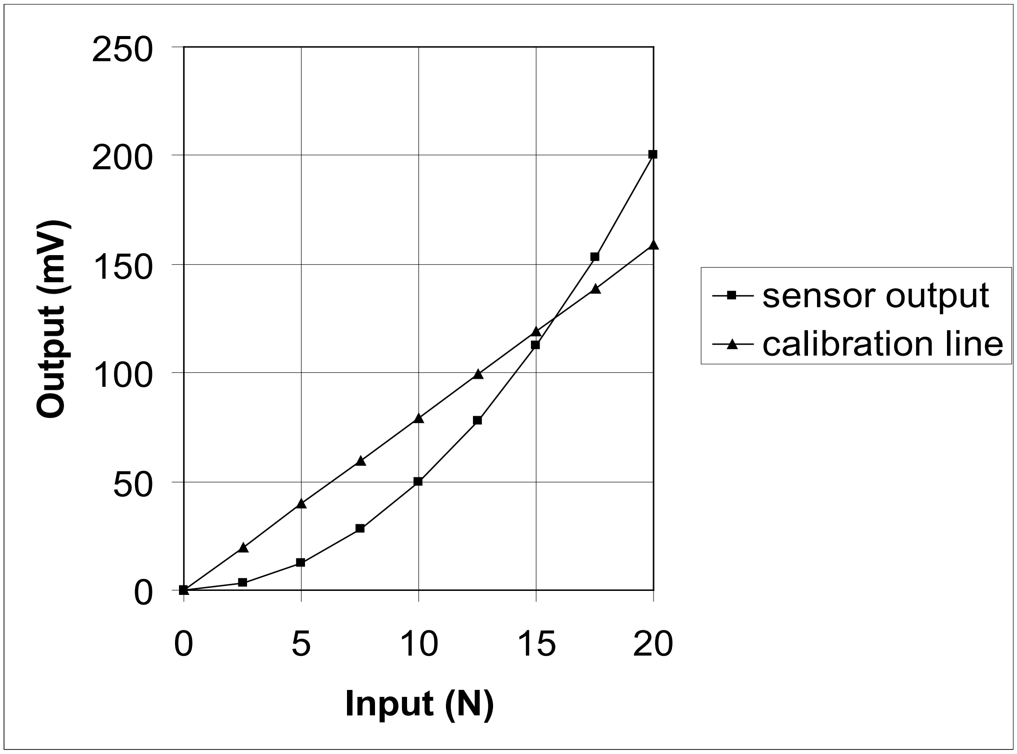


Figure 2.1

Next we use equation (2.6) to calculate the calibrated sensor output as follows:

$$Output(N) = \frac{Output(mV)}{A}$$

These results are listed in Table 2.2 and plotted in Figure 2.2, along with the values for an ideal sensor.

Input (N)	Output (N)	Ideal (N)	Abs. Error (N)
0	0.00	0	0.00
2.5	0.39	2.5	2.11
5	1.57	5	3.43
7.5	3.54	7.5	3.96
10	6.30	10	3.70
12.5	9.84	12.5	2.66
15	14.16	15	0.84
17.5	19.28	17.5	1.78
20	25.18	20	5.18

Table 2.2 Results of the calibration

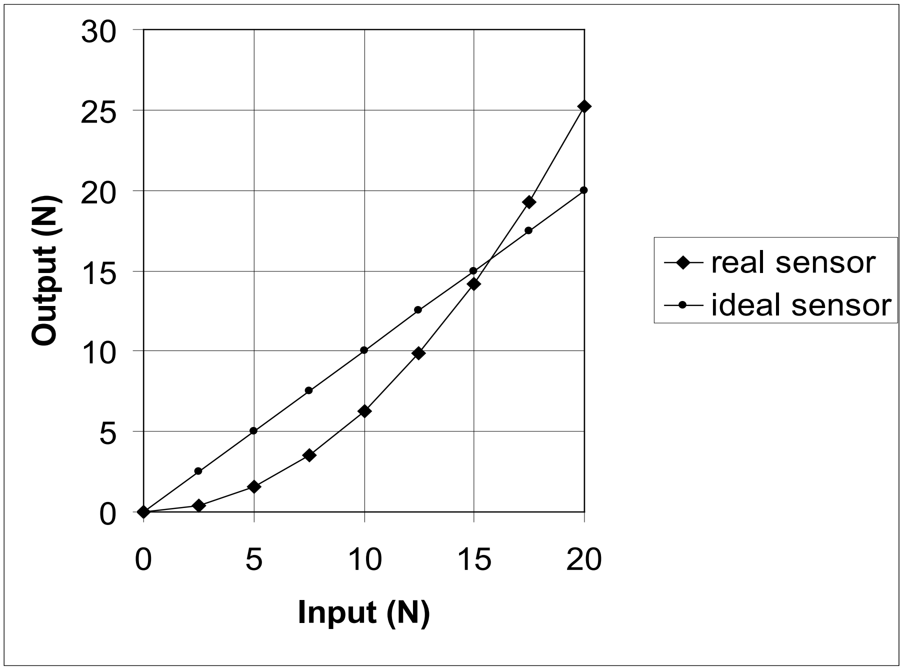


Figure 2.2

From the calibration results we find:

i) Accuracy $= \pm[\max(\text{abs}(F_{sensor} - F_{true})) + 3\sigma_{force}]$

$= \pm[5.18\text{ N} + 3\sigma_{millivolts}/A]$

$$= \pm[5.18\text{ N} + \frac{3(2mV)}{7.94mV/N}]$$

$= \pm 5.94$ N (or $\pm 30\%$ of full scale)

ii) Sensitivity = ratio of sensor output to quantity being measured = A = 7.94 mV/N

iii) Linearity = $\pm[\max(\text{abs}(F_{sensor} - F_{true})]$

$= \pm 5.18$ N

iv) Repeatability = $\pm 3\sigma_{force}$

$$= \pm 3\frac{\sigma_{mV}}{A}$$

$$= \pm\frac{3(2mV)}{7.94mV/N}$$

$= \pm 0.76$ N

(b) Hysteresis = $\dfrac{19\ mV}{7.94\ mV/N} = 2.39$ N

(c) Deadband = 2.3 N

2.2 Standard Industrial Sensors

In this section we will study sensors for the common measured quantities: displacement, velocity, acceleration, depth, proximity, force, pressure, flow, liquid level, and temperature. Samples of some of these sensors will be shown in class, and some will be demonstrated.

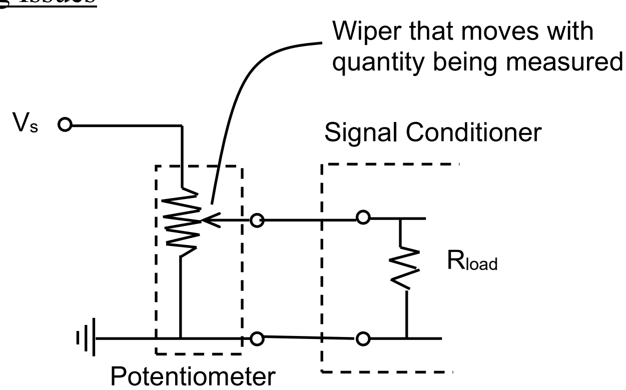
Displacement Sensors

1) Potentiometer: A resistance element with a sliding contact that can be moved over the length of the element.
   - The element can be wire-wound or conductive plastic.
   - A potentiometer can be rotary (single turn), rotary (multi-turn), or linear.

   - With a wire-wound resistance element:
     Disadvantage: limited resolution (due to wire diameter)
     Advantages: not temperature sensitive, linearity between ±0.1% and ±1%.

   - With a plastic resistance element:
     Advantages: unlimited resolution, linearity of ±0.05%
     Disadvantage: temperature sensitive.

   Interfacing Issues



To reduce loading errors we require that: $R_{load} >> R_p$
This issue will be discussed further in Section 2.3.

Summary

Advantages: low cost, large range.
Disadvantages: output affected by load resistance, analog output, and sensor is affected by dirt and wear.

2) Strain gauge: A wire, metal foil or semiconductor strip that can be stuck onto a surface to measure its strain. The change in resistance due to the strain can be converted into a voltage. The governing equation is:

$$\frac{\Delta R}{R} = G\varepsilon \tag{2.8}$$

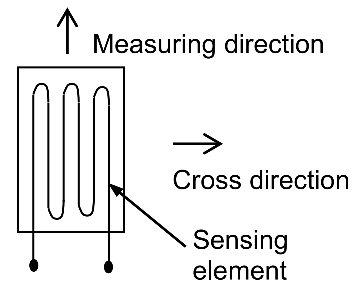
where ε is the strain and G is the "gauge factor". Note that ε is positive for tensile strain and negative for compressive strain.

Question: How is the gauge factor related to the sensitivity specification?

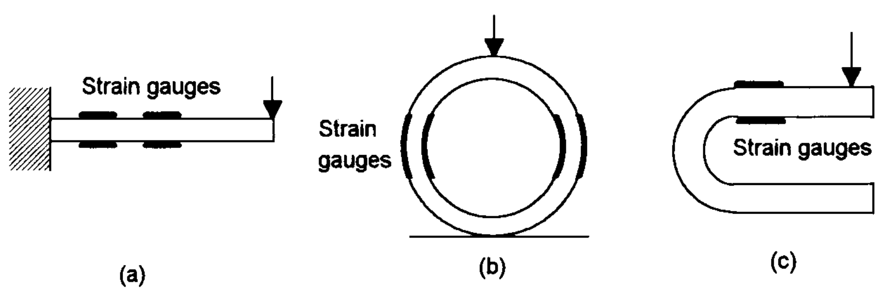
Design Issues

The basic design of a typical strain gauge is shown to the right.

Question: Can you explain the shape used for the sensing element?



- The range of a strain gauge is typically less than 1 mm. To increase the range they may be used with specially designed flexible mechanical elements. Examples are shown below. Strain gauges have a linearity of about ±1%.



Displacement sensors using strain gauges attached to flexible elements [1].
(a) Cantilever design.
(b) Ring shaped design.
(c) U shaped design.

Temperature Compensation

Since the change in resistance due to strain is very small, and resistance also changes with temperature, strain gauges are very sensitive to changes in temperature. To avoid inaccurate measurements, either the temperature of the sensor must be kept constant or a temperature compensation method must be used. The best approach is to mount a pair of sensors on opposite sides of the flexible element. For an example see figure (c) above. The temperatures of the two gauges will be the same if the flexible element is thin and made of thermally conductive material. When the element's tip is deflected in the direction shown the top strain gauge will be in tension (i.e., positive ΔR) while the bottom strain gauge will be in compression (i.e., negative ΔR). Due to their location the magnitudes of the strains will be identical. The resistance change of each gauge is converted to a corresponding voltage by a signal conditioner. The voltages for the top and bottom strain gauge voltage would be:

$$v_{top} = v_{strain} + v_{temperature} \text{ and}$$

$$v_{bottom} = -v_{strain} + v_{temperature}$$

where v_{strain} is the voltage due to the tensile strain and $v_{temperature}$ is the voltage due to the temperature change. The signal conditioner can then use a differential circuit to eliminate the temperature effect as follows:

$$v_{output} = \tfrac{1}{2}\left(v_{top} - v_{bottom}\right) = \tfrac{1}{2}\left(v_{strain} + v_{temperature} - (-v_{strain} + v_{temperature})\right) = v_{strain}$$

Summary

Advantage: no wear

Disadvantages: small range unless mechanical element is added, analog output, and temperature sensitive (unless temperature compensation is used with a pair of strain gauges).

3) Encoders

These devices are used extensively in the motion control systems of robots and machine tools. They produce a digital output for an angular (or linear) displacement.

There are two main types:

a) Incremental: output is a measure of a relative displacement.

b) Absolute: output is a measure of an absolute displacement.

They are typically optical, but magnetic (incremental and absolute), capacitive (incremental only) and inductive (incremental only) encoders are also available.

Incremental encoder

The basic working principle of an incremental optical encoder is illustrated in Figure 2.3. The device shown will produce a single bit digital output signal. Typically an LED is used as the light source and a phototransister is used as the light sensor. The fixed disk contains a single slot, while the rotating disk contains many equal width slots. As the rotating disk is rotated the light will be sequentially blocked and unblocked, producing an ON (or high) pulse for each angular increment. The disks are made from etched glass or metal foil, and resolutions up to 10,000 pulses/revolution are available. Normally the rotating disk is attached directly to the output shaft of the motor (*i.e.* before any mechanical elements such as a gearbox).

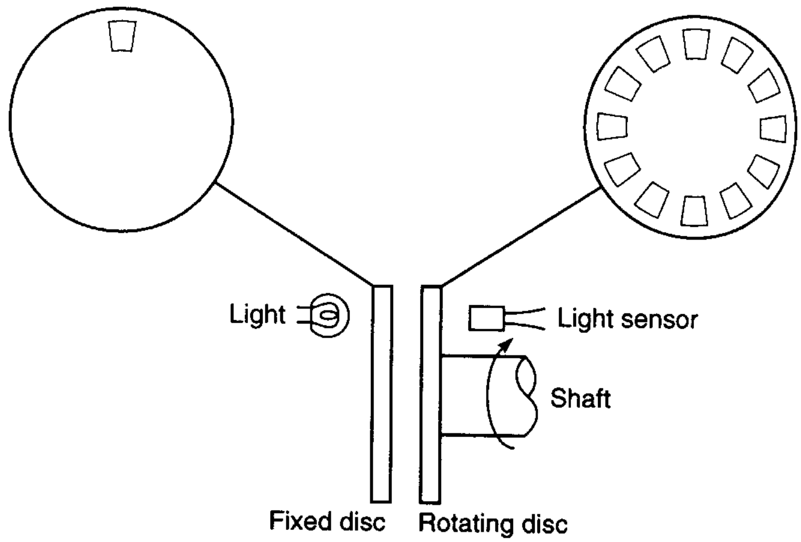


Figure 2.3 Basic working principle of a rotary incremental optical encoder [1].

The absolute displacement may be measured by counting the encoder output pulses using an electronic circuit. This requires first resetting the counting circuit to zero when the actuator is at its zero position. A single output encoder is not practical since there is no way to tell if the pulse was produced by a movement in the positive or negative direction (*i.e.* should we add or subtract from the count?). A typical incremental encoder has two outputs, named Channel A and Channel B. They are created by two slots in the fixed disk that are offset by half a slot width. This offset causes the signals to be 90° out of phase as shown in the top part of Figure 2.4. If the pulse from Channel A occurs before the pulse from Channel B then the rotation is clockwise. If Channel A lags behind Channel B then the direction is counterclockwise. Using more advanced circuitry, the rising and falling edges from both channels may be counted rather than just the pulses from one. This is shown in the bottom part of Figure 2.4. This feature improves the effective resolution by four and is known as "quadrature counting". For example, if a 500 pulse/rev encoder is used with quadrature counting the resolution will be 360°/2000 counts = 0.18°/count. In addition to rotary encoders, linear incremental encoders are available for use with linear actuators.

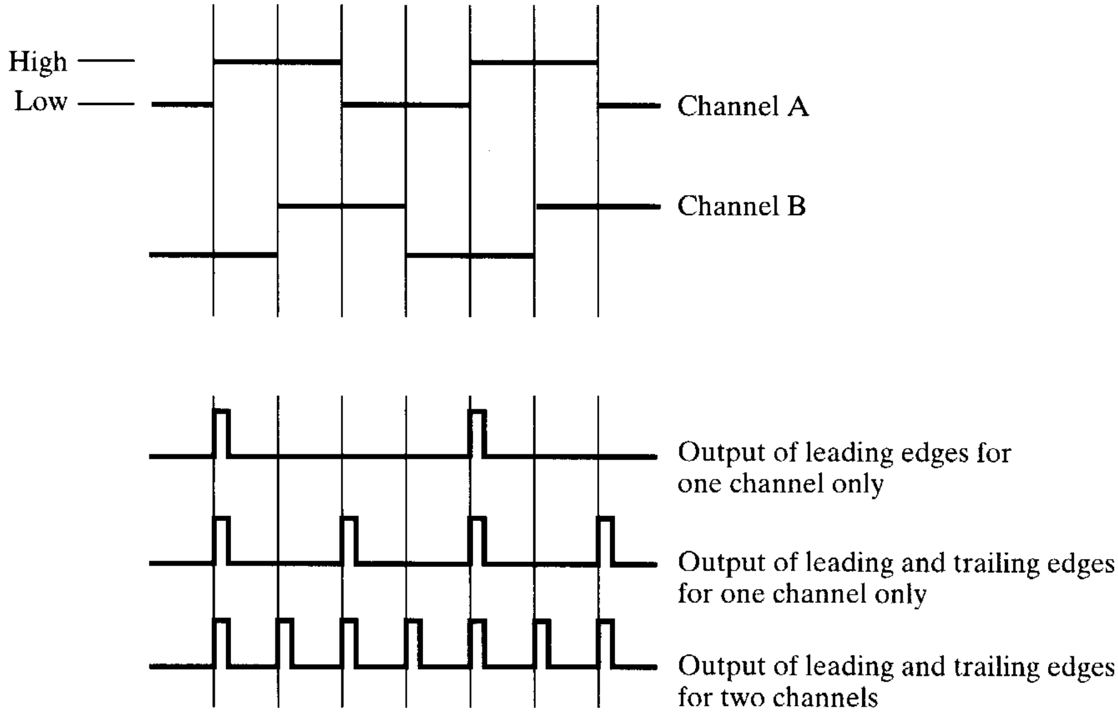


Figure 2.4 The use of quadrature counting to improve the effective resolution of an incremental encoder [2].

Absolute encoder

With an absolute encoder each position over a single revolution gives a unique output. The output is a binary number of several digits. This is accomplished by using several concentric tracks on the rotating disk, each with a different pattern. A simple example is shown in Figure 2.5. This 3 bit encoder has a resolution of $360°/(2^3) = 45°$. Absolute encoders are available with outputs up to 12 bits.

The main problem with an absolute encoder is that its range is limited to one revolution. Although a larger range is possible by using several geared disks it is much more common to use an incremental encoder with a counter. The cost is lower and the range is only limited by the counter. For example, with 1024 counts/rev and a 32 bit counter the range is:
$(2^{32}$ counts$)/(1024$ counts/rev$)$ = 4.2×10^6 revolutions!

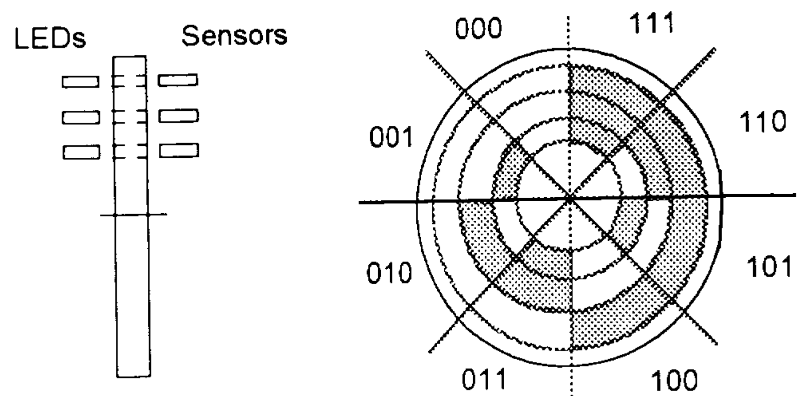


Figure 2.5 A 3 bit absolute optical encoder [1].

Summary

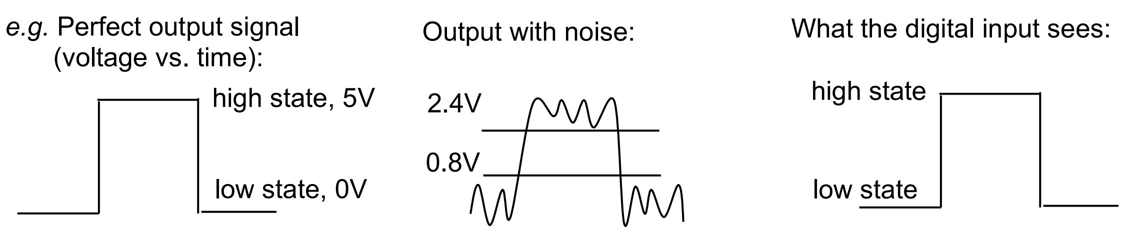
Optical encoders (of both types) are very common because of their advantages, namely: no wear, and their digital output is relatively insensitive to noise and easy to interface. They have the disadvantages of limited resolution, and susceptibility to large vibrations (particularly those employing glass disks).

More Info on Analog vs. Digital Sensors

Digital

Advantage: Noise immunity

For example, with TTL:
Low State<0.8V
High State>2.4V



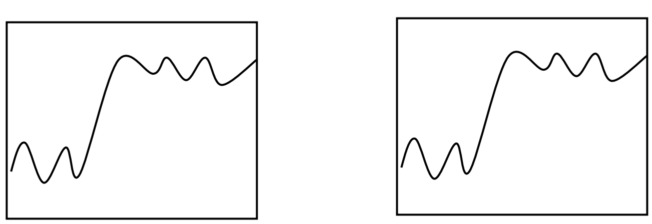
Disadvantages:

- Limited resolution, *e.g.* optical encoder, although this is not usually a problem.
- Few purely digital sensors exist.

Analog

Advantage: potentially infinite resolution.
Disadvantage: sensitive to noise, *e.g.*

Output with noise (voltage vs. time): What the analog input sees:



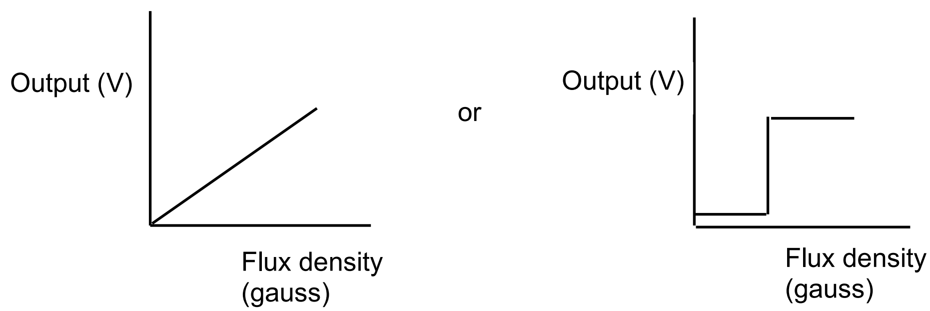
Final Comment on Analog vs. Digital Sensors

Today the solution is often to digitize the analog sensor output locally and transmit it digitally if the transmission distance is large (since longer cables increase noise like an antenna).

Getting back to displacement sensors...

4) Hall Effect Sensors

This sensor can be used to measure displacement indirectly by measuring the flux density of a magnet. The analog output can be linear or employ a threshold. Please see the figure below.



The advantages of hall effect sensors are: non-contact (no wear), and virtually immune to the working environment.

Example Applications
- The threshold type may be used to create a non-contact proximity switch.
- The linear type are being used extensively in automotive applications, *e.g.* sensing of accelerator angle, and sensing of fuel tank level (with a float).

Velocity sensors

Velocity sensors are used for motion control applications, such as robots and machine tools.

There are two common methods:

i) By differentiating displacement or position

Advantage: low cost
Disadvantages: Amplifies high frequency noise. For example, if the displacement voltage consists of a low frequency signal (ω_1) plus noise at a higher frequency (ω_2) then we have:

$$V_{disp.} = V_1 \sin(\omega_1 t) + V_2 \sin(\omega_2 t) \tag{2.9}$$

$$V_{vel.} = \omega_1 V_1 \cos(\omega_1 t) + \omega_2 V_2 \cos(\omega_2 t) \tag{2.10}$$

where $\omega_2 > \omega_1$. Comparing equation (2.10) to (2.9) we can see that differentiating the voltage has amplified the noise by ω_2 / ω_1 relative to the signal.

A second disadvantage is that numerical differentiation adds time delay (or phase lag). The reason is that the displacement signal is only known up to the current time. Therefore backward differencing must be used, and this introduces a time delay.

We will quantify the errors due to numerical differentiation in section 2.4.

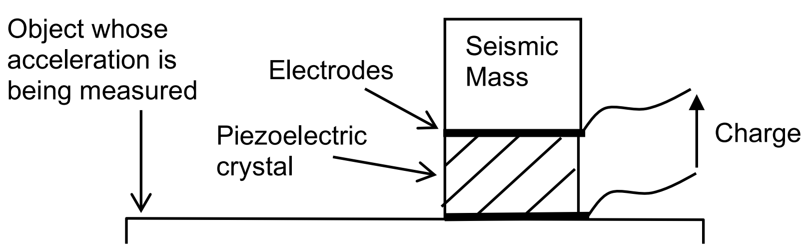
ii) Tachogenerator or Tachometer

This sensor produces an analog voltage that is proportion is the velocity. A DC tachogenerator consists of a wound armature, a permanent magnetic stator and a set of commutator brushes. This type has the disadvantage that the brushes create noise and are subject to wear. More recently brushless tachogenerators have been introduced. The commutation is performed using an electrical circuit.

Acceleration Sensors

Acceleration sensors are known as "accelerometers". They are used with automotive airbags to detect accidents, for detecting skids with anti-lock breaking systems, and for sensing vibrations. Vibration sensing may be used for the condition monitoring of machines, *e.g.* to sense when a bearing or gear is worn and should be replaced.

Accelerometers often consist of a very small mass attached to a piezoelectric crystal. The force produced by the acceleration of the mass squeezes the crystal and causes an electrical charge to be produced. Please see the figure below.



Another type uses a small mass (called a seismic mass) attached to a cantilever beam. The strain of the cantilever is used to measure the acceleration.

Design Issues

With both types of accelerometer the ratio of the seismic mass, *m*, to the stiffness of the crystal or beam, *k*, is an important design parameter. The larger the ratio *m/k* the larger the sensitivity of the sensor. However, when this ratio is increased the natural frequency of this mass-spring system will be decreased since:

$$f_n = \frac{1}{2\pi}\sqrt{\frac{k}{m}} \tag{2.11}$$

where f_n is the natural frequency in Hz. This will in turn limit the bandwidth. This issue will be discussed further in class.

Summary

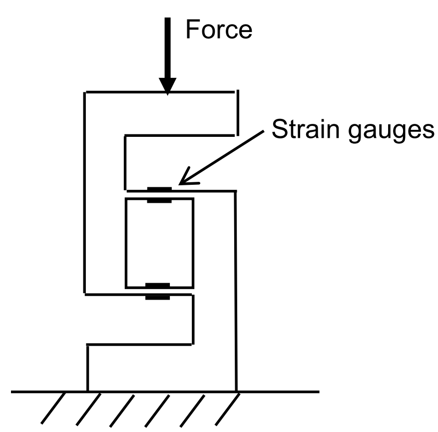
Accelerometers have the advantages:

- Simpler to use than displacement and velocity sensors because they only need to be attached at a single point (rather than two points) so alignment is much easier. This is the main reason they are used for vibration sensing.
- No wear.
- Multiple axis (X-Y-Z) and angular acceleration sensors are available.

Question: Given their advantages, why aren't acceleration sensors used with integration to measure velocity or displacement?

Force Sensors (also sometimes called Load Cells)

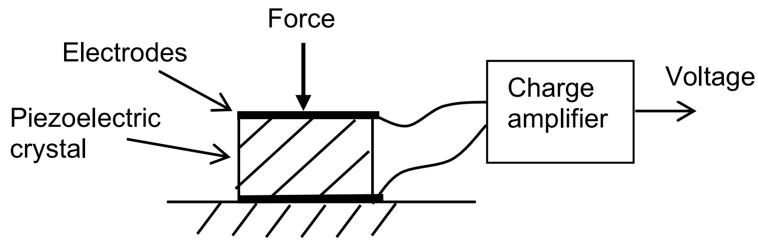
Force sensors have various applications such as weighing products during manufacture, and monitoring and controlling metal cutting and metal forming operations. There are two main types, using strain gauges or piezoelectric crystals. With both types the force is first converted to a very small displacement that is then sensed.



i) Strain Gauge Type: These use strain gauges to measure the strain produced when a force is applied to a specially designed flexible element. These elements can take the shapes shown on page 2-12. Another common design is the S shape shown to the right. An example will be shown in class.

Advantage: Low cost.
Disadvantages: Worse linearity and greater temperature sensitivity than the piezoelectric type.

ii) Piezoelectric type: The applied force squeezes a piezoelectric crystal causing it to output a charge. This charge is only emitted when the force changes. To measure constant forces a signal conditioner that amplifies and integrates this charge signal is necessary. This device is termed a "charge amplifier". Please see the figure below.



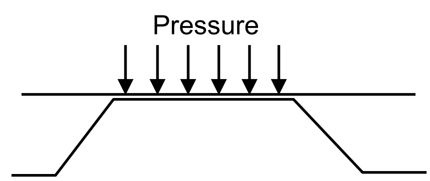
Advantages: Better linearity and temperature sensitivity than the strain gauge type.
Disadvantages: Higher cost, especially for constant loads. The charge amplifier must be carefully designed to prevent stability problems (drift).

Design Issues

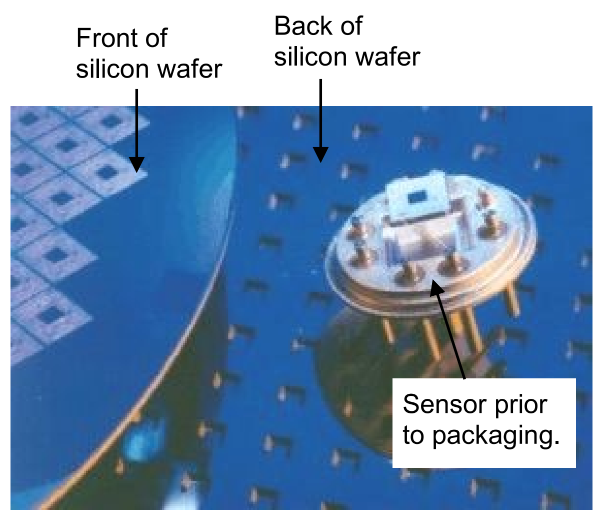
The nature of the piezoelectric crystal avoids problems with cross-sensitivity. The flexible elements used with the strain gauge type must be designed to avoid cross-sensitivity. The crystal or flexible element should be as stiff as possible to maximize the natural frequency. As with accelerometers there is a trade-off with sensitivity.

Pressure Sensors

These have traditionally been very expensive devices that have seen application mainly in industrial process monitoring and control applications. However, in the 1970s, government legislation forced the automotive industry to produce cleaner running and more fuel efficient automobiles. This lead to the sophisticated engine control systems we have today. One the sensors used measures the air pressure inside the intake manifold. The large automotive market, and the technology of micro-electromechanical systems (MEMS), has lead to pressure sensors that cost about $10 each. These typically employ a tiny diaphragm that incorporates strain gauges to measure its deflection when the pressure is applied. Please see the figures below. An example will be shown in class. Versions that measure absolute pressure, gauge pressure and differential pressure are available.



Cross-section of diaphragm etched from a silicon wafer. A pressure difference causes the diaphragm to deflect.

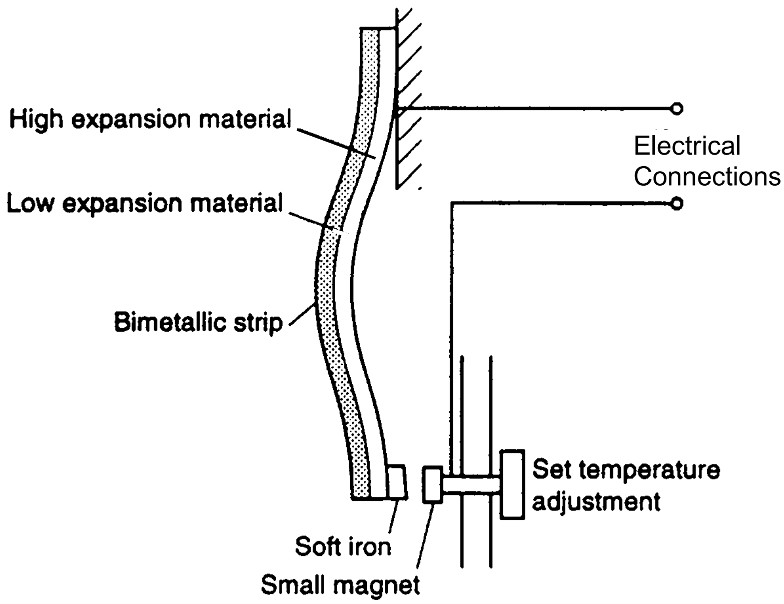


Several hundred MEMS pressure sensors can be created from a single silicon wafer.

Temperature Sensors

There are four common types: bimetallic strips, resistance temperature detectors, thermistors and thermocouples.

1) Bimetallic strips: Simply two metal strips bonded back to back. Each metal has different coefficient of thermal expansion. When the temperature changes the different expansions cause the strip to bend. This device is used as temperature controlled switch in some thermostats and toasters. An example of a bimetallic thermostat is shown in the figure below.



2) Resistance temperature detectors: This device consists of coil of wire. Since the resistance of metal is proportional to temperature, the change in temperature is detected by measuring the change in the coil resistance. This device has the advantages of simplicity and linearity and the disadvantage of small sensitivity.

3) Thermistors: A thermistor consists of a mixture of metal oxides. This mixture has been designed such that its resistance is very sensitive to temperature.

Advantages
- The oxide mixture can be easily formed into different shapes (catered to the application) such as disks or beads.
- Large sensitivity.
- The sensor can be very small, *e.g.* less than 1 mm in diameter.

Disadvantage: Very non-linear response so special calibration circuitry or software is required.

Question: What are the advantages of a very small temperature sensor?

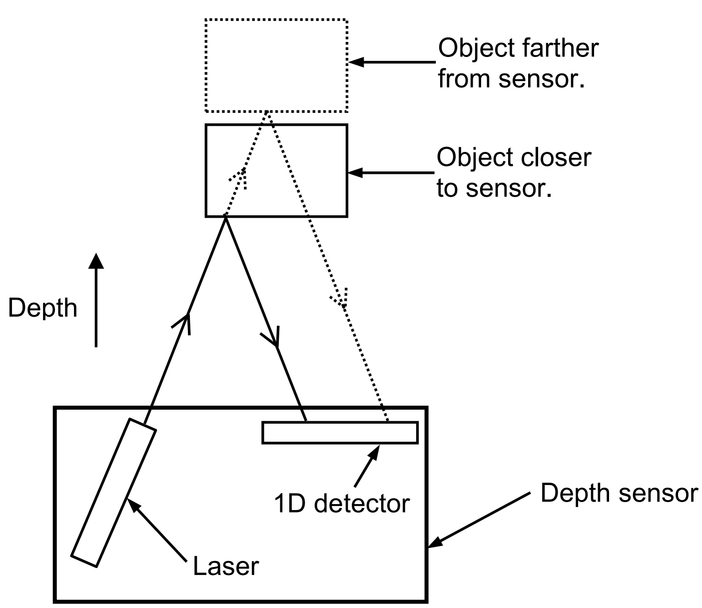
4) Thermocouples: Thermocouples are the traditional temperature sensor but are becoming less popular than thermistors and resistance temperature detectors. When two different metals are joined together a small voltage occurs across the junction. This voltage is a function of the metals involved and the temperature. This principle is used to produce thermocouples.

Advantage: Large range, excellent linearity.
Disadvantages: Very small sensitivity. The small output (a few mV) makes them very susceptible to noise. Special electronics are required to amplify and filter the sensor output, and to provide a temperature reference.

Depth Sensors

The non-contact sensing of the distance between an object and the sensor is known as "depth sensing". It is also called "range sensing" but we will avoid this confusing term (confusing since range is a property of any sensor). There are two main methods for depth sensing, named "triangulation" and "time-of-flight". Examples will be demonstrated in class.

With triangulation, a laser and a 1D detector are arranged to form a triangle. When the object moves towards or away from the sensor the light beam moves laterally along the detector. The depth can be determined from the geometry of the sensor and the detected beam position. This idea is illustrated in the figure below.



Design of a laser depth sensor based on the triangulation method.

In fact, triangulation also one of the ways humans and animals detect depth. Rather than the laser and 1D detector, we use the lateral position difference between the images from each eye, but the principle remains the same.

The time-of-flight method is based on the fact: distance traveled = speed × time. This assumes that the speed is constant and known. Ultrasonic depth sensors use this principle. A transmitter emits an ultrasonic pulse (normally between 30-100 kHz). The pulse then travels through the air and reflects off the object being measured. The reflected pulse is detected by a receiver. The product of the travel time and the speed of the ultrasonic pulse in air gives the depth measurement. The lidar sensors used with autonomous vehicles also use the time-of-flight method. They emit pulsed laser light and measure the travel time of the reflected light with picosecond accuracy.

Advantages of Ultrasonic Depth Sensors vs. Laser Triangulation Depth Sensors
- Greater range.
- Not affected by the optical properties of the object.
- The transmitter can be located close to the receiver without affecting the performance. In fact sometimes the same device is used as transmitter and receiver.
- Not affected by dirt.
- Lower cost.

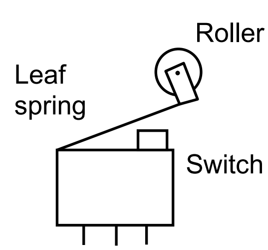
Disadvantages of Ultrasonic Depth Sensors vs. Laser Triangulation Depth Sensors
- Worse accuracy.
- Susceptible to indirect reflections. For example the ultrasonic pulse may reflect off the wall before reaching the receiver causing the depth to be overestimated.

Proximity Sensors

Proximity sensors are very common in automation and manufacturing applications. They produce a binary (on/off) output when an object is within a specified proximity. They are used to detect when or whether a location has been reached, for detecting the presence of parts on conveyors, and for safely limiting the motion range of robots and machine tools.

The common types are microswitches, inductive, and photoelectric.

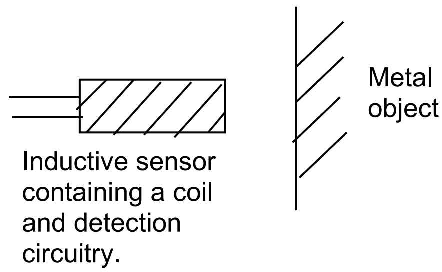
1) Microswitch: A small spring-loaded mechanical switch. An example is shown below.



Advantage: No power required.

Disadvantages:
- Object being sensed must come into contact and apply a small force to activate the switch.
- The dynamics of the spring and mass can cause false outputs
- Subject to wear.

2) Inductive: These sensors use the change in the magnetic circuit formed by a coil and the metal object to detect the presence of the object. Please see the figure below.

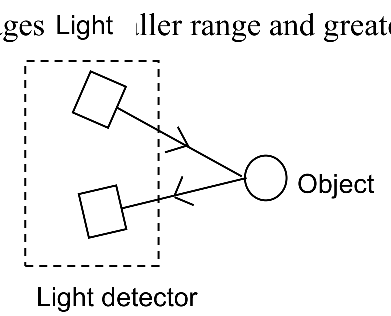


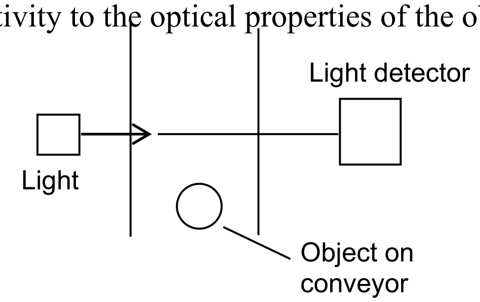
Advantage: Non-contact (no wear or dynamics problems).

Disadvantages
- Can detect the presence of metal objects only.
- Small range (typically less than 5 mm).
- Requires power.

Comments
Inductive sensors have replaced microswitches in many applications.

3) Photoelectric: These consist of a light (usually an LED) and a light detector (usually a phototransistor). There are two types: retro-reflective type or thru-beam type (see the figure below). The retro-reflective type has the advantage that the light and light detector are contained within the sensor and do not require separate mounting and alignment. It has the disadvantages [illegible]ller range and greater sensitivity to the optical properties of the object.





Retro-reflective proximity sensor. Thru-beam proximity sensor.

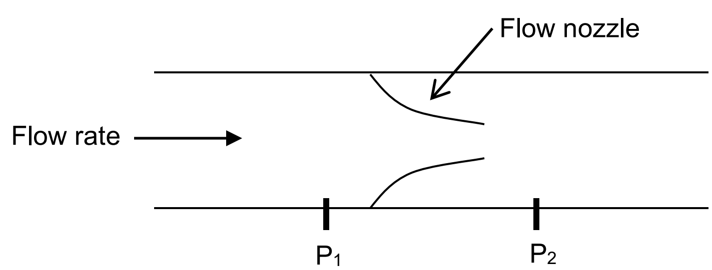
Advantages of Photoelectric Proximity Sensors
- Both types have a much larger range (typically about 1 m) than the other proximity sensors.

Disadvantages of Photoelectric Proximity Sensors
- Suitable for opaque objects only.
- Sensitive to dirt.
- Sensitive to ambient light.
- Requires power.

Flow Rate Sensors

Flow rate sensing is used in the monitoring and control of many chemical processing operations. Two of the more common types of flow rate sensors are the flow nozzle and the turbine flow meter.

1) Flow nozzle
   As shown in the figure below, a flow nozzle is inserted into a pipe. The nozzle restricts the flow causing a pressure drop. The pressure difference, $P_1–P_2$, is then measured using a differential pressure sensor. The pressure difference, along with the geometry of the nozzle and the properties of the fluid (or gas), are used to calculate the flow rate.



2) Turbine Flow Meter
   This consists of a propeller attached to a tachogenerator. The device is placed at the centre of the pipe to measure the flow rate.

Advantages of Flow Nozzle vs. Turbine Flow Meter
- Less expensive.
- Less susceptible to wear.

Disadvantage of Flow Nozzle vs. Turbine Flow Meter
- Less accurate.

Liquid Level Sensing

There are several methods for liquid level sensing. Most are based on the sensors described in the previous sections. The advantages and disadvantages of the methods are described in Example 2.2.

Example 2.2: Case Study on Liquid Level Sensing

You are an Engineer working in a brewery. You are to design a liquid level sensing system for a large tank of beer (density = 1100 kg/m^3). The tank's dimensions are 5 m x 5 m x 5 m, and its empty weight is 10 metric tons. You must measure its level within ± 3 cm. The tank is enclosed (but not air tight), and its level changes slowly.

Part 1: Brainstorming
List at least four different sensing methods that could be used to measure the liquid level.

Part 2: Design Calculations
Discuss the feasibility of the method for this application. If feasible, calculate the sensing range and accuracy required, plus any other required design parameters.

Solution

1) Distinct sensing methods:

a) Float with mechanism and displacement sensor
b) Float with depth sensor (laser triangulation or ultrasonic)
c) Ultrasonic depth sensor
d) Differential pressure sensor
e) Force (weight) sensor mounted under tank
f) Guided-wave radar sensor

2) Design discussions and calculations:

a) Float with mechanism and displacement sensor:

Since it is easier to measure rotation than linear displacement we should use a mechanism to convert the vertical movement of the float into rotation. The mechanism choices include:

- rigid link (possibly counterbalanced),
- lead screw (and nut),
- timing belt, and
- rack and pinion.

The rigid link is the simplest choice but only works for small, shallow tanks.

The lead screw design involves attaching the float to the nut and including a linear bearing that prevents the nut from rotating. Then when the float move the screw will rotate and this rotation is measured by an angular displacement sensor (typically an incremental or absolute encoder). This only works if the floatation force (or buoyancy force) is large enough to overcome the friction of the lead screw and nut. The friction can be reduced by using low friction materials, and a screw with a large lead.

The timing belt will have poor accuracy unless the belt is properly tensioned to remove any slack. Similarly, the rack must have high stiffness otherwise its deflections will result in poor accuracy.

Even without resorting to design calculations, for large tanks, like the one in this case study, the "float with mechanism and displacement sensor" sensing method is difficult to implement successfully so we won't pursue it further.

b) Float with depth sensor (laser triangulation or ultrasonic):

No calculations are required to convert the given specifications since the liquid level will be directly measured by the depth sensor. The only implementation issue is the need to mechanically constrain the float such that the laser light or ultrasound reflects directly back to the sensor.

c) Ultrasonic sensor:

Ultrasound can also be used without a float. The problem is that surface waves can produce incorrect measurements.

Note: With large tanks the float can be placed in an outer tube connected to bottom of the tank. This solves two problems. It constrains the motion of the float to the vertical direction and it eliminates the surface waves. This can be used with method (b), and without the float with method (c).

d) Differential pressure sensor:

From the given information, we require:

accuracy $\leq \pm 3$ cm

range: 0 to 5 m

We need to convert these values to differential pressure.

Define h to be the height of the liquid, and ρ to be its density.

Differential pressure of 3 cm of liquid $= \rho g h$

$$= 1100 \frac{kg}{m^3} \bullet 9.81 \frac{m}{s^2} \bullet 0.03m$$

$$= 323 \frac{kg}{ms^2} = 323 \ Pa$$

$\therefore$ differential pressure sensor accuracy $\leq \pm 323$ Pa

Maximum differential pressure = full tank pressure

$$= \rho g h$$

$$= 1100 \frac{kg}{m^3} \bullet 9.81 \frac{m}{s^2} \bullet 5m$$

$$= 53,900 \ Pa$$

When the tank is empty, differential pressure $= 0 \ Pa$

$\therefore$ differential pressure sensor range $= 0$ to $5.39 \times 10^4 \, Pa$

e) <u>Force sensor mounted under tank:</u>

Again, we must convert the given specifications.

Define A as the cross sectional area of the tank

Weight of liquid $= A\rho g h$

Weight of 3 cm of liquid $= A\rho g h$

$$= (5m \bullet 5m) \bullet 1100 \frac{kg}{m^3} \bullet 9.81 \frac{m}{s^2} \bullet 0.03m$$

$$= 8100 \frac{kg \bullet m}{s^2} = 8100 \, N$$

$\therefore$ force sensor accuracy $\leq \pm 8.1 \times 10^3$ N

Weight of tank $= mg = \left(10 \times 10^3 \, kg\right)\left(9.81 \frac{N}{kg}\right) = 9.81 \times 10^4$ N

Max weight = tank weight + liquid weight ,

= tank weight + $A\rho g h$, where h = 5 m

$$= 9.81 \times 10^4 \, N + (5m \times 5m) \bullet 1100 \frac{kg}{m^3} \bullet 9.81 \frac{N}{kg} \bullet 5m$$

$$= 1.45 \times 10^6 \, N$$

$\therefore$ force sensor range $= 9.81 \times 10^4$ N to 1.45×10^6 N

f) Guided-wave radar sensor:

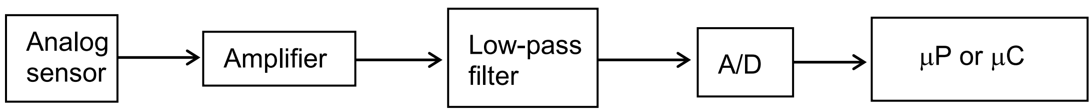
A special metal rod is inserted into the tank (top to bottom). A radar pulse is sent down the rod. At the top of the liquid some of this pulse is reflected back up the rod to a detector. The time-of-flight is used measure the liquid level. For more info please see: www.controleng.com

End of example 2.2

2.3 Sensor Interfacing

Typical Hardware

For the case of a sensor with an analog output the typical hardware is:



The amplifier, low-pass filter and analog to digital converter (abbreviated as A/D or ADC) form the signal conditioning system. The amplifier increases the voltage to a suitable value, and the low-pass filter is used to reduce the amplitude of any high frequency noise. The ADC is required to convert the analog voltage to a digital value for inputting to the μP or μC[1]. A typical A/D has input range of −10 V to +10 V and a 12 bit or 16 bit output. With this input range and a 16 bit output the resolution of the ADC is 20 V / 2^{16} = 0.3 mV. We will quantify the worst case error of an ADC in section 2.4. If the sensor has a digital output it may be possible to connect it directly to the μC or a digital to digital converter (D/D or DDC) may be required.

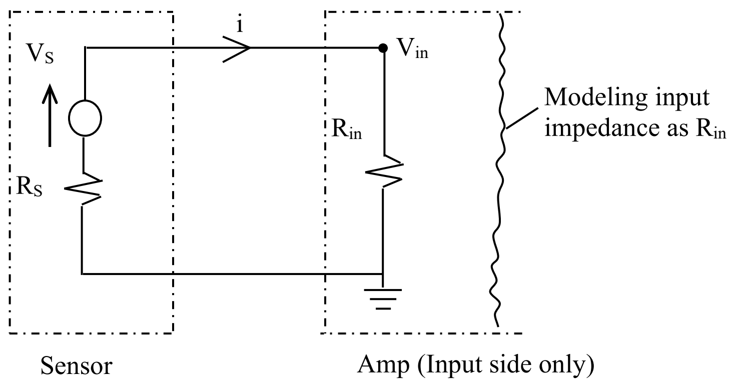
1. Note that some μC include an internal ADC, but they are usually of inferior quality compared to an external ADC.

Differential Inputs vs. Single-Ended Inputs

A signal may be transmitted as a voltage relative a common ground. An input that receives this signal is known as "single-ended". This has the advantage that if we want to send three signals, for example, then only four wires are required. However, if the signals are transmitted over a long distance then significant noise will be added to the signal reducing its quality and usefulness. Another option is to transmit two signals, the original and another with the sign inverted. These are then connected to a differential input that subtracts the inverted signal from the original one and then divides by two. Since the signals are together in the cable then will tend to pickup the same noise so subtracting then greatly reduces this common noise. For example, at an instant in time say our sensor outputs 0.3 V. If it has a differential output then the signals would be +0.3 V and −0.3 V. When the signals have reached the amplifier say 0.05 V of noise has been added to them. The result from the differential input is: $[+0.35 - (-0.25)]/2 = 0.3$ V. With a single-ended input the result would have been 0.35 V. Differential inputs can be used with analog or digital signals. In addition to noise reduction they prevent the problems that may arise from connecting ground signals together. Their disadvantages are the added costs due to more complex input circuitry and the extra wires needed (*e.g.* three signals would need six wires).

Loading Effects

If the hardware elements are not properly selected they can suffer from "loading effects". This applies both digital and analog signals, although (as with noise) digital signals are much less sensitive. Any connection of an output to an input can suffer from loading effects. The connection of a sensor output to an amplifier input will be used as an illustrative example.



Electrical model of a sensor output circuit and an amplifier input circuit.

First we must define the term "impedance". Impedance is the complex number relating the voltage to the current as a function of frequency, that is: impedance = $\frac{V(j\omega)}{I(j\omega)}$. We will model the input impedance of the amplifier as R_{in}. We will model the sensor as a voltage source with an output impedance equal to R_s. The voltage V_s is the open-circuit output of the sensor. We are interested in seeing the effect of the load created by connecting the sensor to the amplifier. We do this by applying Kirchhoff's 2nd law which states that in a loop:

$$\text{voltage supply} = \text{voltage drop} \tag{2.12}$$

For our loop (see figure above) we have:

$$V_S = iR_{in} + iR_S \tag{2.13}$$

And from Ohm's law $V_{in} = iR_{in}$ so that:

$$i = \frac{V_{in}}{R_{in}} \tag{2.14}$$

Substituting i from (2.14) into (2.13) gives:

$$V_S = V_{in} + \frac{V_{in}}{R_{in}} R_S \tag{2.15}$$

$$V_{in} = V_S \left(\frac{R_{in}}{R_{in} + R_S} \right) \tag{2.16}$$

Ideally we would like $V_{in} = V_S$. This means we want:

$$\frac{R_{in}}{R_{in} + R_S} \approx 1 \tag{2.17}$$

Therefore R_{in} should be large, and R_S should be small.

Conclusion

Each hardware element must have a small output impedance and a large input impedance for loading effects to be minimized.

2.4 Sensor Analysis

The specifications given with a sensor are not always sufficient to see if it is suitable for a particular application. In this subsection, methods for further analysis of the performance of a sensor alone and in combination with a signal conditioner are described.

2.4.1 Other Sources of Error and Propagation of Errors

The final measured value used by the control system (or displayed by the measurement system) will be affected by more than just errors due to the sensor's steady state behavior. Loading effects (covered in subsection 2.3), errors due to time varying inputs (covered in subsection 2.4.2), amplifier gain variations, analog to digital conversion errors and other sources of error can also be significant.

An analog to digital convertor (ADC) quantizes an analog input into a digital value. Its resolution is specified by the manufacturer in numbers of bits, e.g. 16-bit. However, the error due to quantization cannot be used to compute the ADC's accuracy. Luckily, manufacturers also specify the ADC's effective number of bits (ENOB). This can be related to accuracy using:

$$a_{ADC} = \pm \frac{V_{FS}}{2^{ENOB}} \tag{2.18}$$

where a_{ADC} is the ADC accuracy in Volts, and V_{FS} is ADC's full scale voltage (*i.e.* its maximum input minus its minimum input).

When uncertain quantities are multiplied and/or divided, the uncertainty in the result may be conservatively calculated by summing their relative uncertainties. For example, if X, Y, Z and P are uncertain quantities and:

$$R = \frac{XYZ}{P} \tag{2.19}$$

Then the absolute uncertainty in *R*, termed ΔR, may be calculated using:

$$\Delta R = |R|\left(\left|\frac{\Delta X}{X}\right| + \left|\frac{\Delta Y}{Y}\right| + \left|\frac{\Delta Z}{Z}\right| + \left|\frac{\Delta P}{P}\right|\right) \tag{2.20}$$

where $\Delta X/X$ is the relative uncertainty in *X*, $\Delta Y/Y$ is the relative uncertainty in *Y*, and so on.

Similarly, when uncertain quantities are added and/or subtracted, the uncertainty in the result is conservatively calculated by summing their absolute uncertainties. For example, if X, Y and Z are uncertain quantities and:

$$R = X + Y - Z \tag{2.21}$$

Then the absolute uncertainty in *R* may be calculated using:

$$\Delta R = |\Delta X| + |\Delta Y| + |\Delta Z| \tag{2.22}$$

This methodology will now be illustrated by an example.

Example 2.3

A temperature sensor with an analog output is connected to an amplifier. The amplifier is connected to an ADC. The output of the ADC is scaled to ºC and displayed. The sensor has a sensitivity of 0.01 V/ºC, a range of –50 ºC to 100ºC, and an accuracy of 0.1% of full scale. The amplifier has a gain of 5±0.01. The ADC has an input range of 0 to 5 V, a 10-bit resolution, and an effective number of bits equal to 8.9. Assuming that other sources of error (such as loading effects) are negligible, determine the temperature accuracy of this measurement system for a 100ºC input.

To start our answer, we need to define the equation relating the system output in ºC to the system input in ºC. From the given information we have:

$$T_{out} = K_{sensor} K_{amp} K_{ADC} K_{sf} T_{in}$$

where T_{in} (ºC) is the input temperature (*i.e.* the true temperature), T_{out} (ºC) is the output temperature (*i.e.* the measured temperature), K_{sensor} (V/ ºC) is the sensor sensitivity, K_{amp} (V/V) is the amplifier gain, K_{ADC} is the ADC gain (1/V) and K_{sf} (ºC /1) is the scalefactor used by the display's IC to obtain the temperature in ºC.

From the given information, in our T_{out} equation the uncertain quantities are K_{sensor}, K_{amp} and K_{ADC}. The uncertainty of K_{amp} is given. For the sensor and ADC we first need to determine their inputs and accuracies as follows:

T_{in} = 100 ºC

$$a_{sensor} = \pm(0.1\%)(\text{full scale}) = \pm(0.1\%)\left(100°\text{C} - (-50°\text{C})\right) = \pm 0.15°\text{C}$$

$V_{in} = K_{sensor} K_{amp} T_{in}$ = 5 V and

$$a_{ADC} = \pm\frac{V_{FS}}{2^{ENOB}} = \pm\frac{(5-0)\text{ V}}{2^{8.9}} = \pm 0.0105\text{ V}$$

Using (2.20) the uncertainty in the measured temperature is then:

$$\begin{aligned}
\Delta T_{out} &= \left|T^*_{out}\right|\left(\left|\frac{\Delta K_{sensor}}{K_{sensor}}\right| + \left|\frac{\Delta K_{amp}}{K_{amp}}\right| + \left|\frac{\Delta K_{ADC}}{K_{ADC}}\right|\right) \\
&= \left|T_{in}\right|\left(\left|\frac{a_{sensor}}{T_{in}}\right| + \left|\frac{\Delta K_{amp}}{K_{amp}}\right| + \left|\frac{a_{ADC}}{V_{in}}\right|\right) \\
&= \left(100°C\right)\left(\left|\frac{\pm 0.15°C}{100°C}\right| + \left|\frac{\pm 0.01}{5}\right| + \left|\frac{\pm 0.0105\text{ V}}{5\text{ V}}\right|\right) \\
&= 0.56°C
\end{aligned}$$

where T^*_{out} is the ideal output.

∴ Temperature accuracy of this system for 100 ºC input = ±0.56 ºC.

End of example 2.3.

2.4.2 Errors Due to Time Varying Inputs

Important specifications like accuracy only apply when the input is static, and the sensor's output has reached steady state. If the input is time varying then the error in the measured value will increase.

If no other information is available then it is reasonable to approximate the measurement system's transfer function as by this first order lag transfer function:

$$\frac{Y_{out}(s)}{Y_{true}(s)} = \frac{K_s}{\tau_s s + 1} \tag{2.23}$$

where K_s is the measurement system's steady state gain, and τ_s is the measurement system's dominant time constant.

Note that K_s is an uncertain quantity, with an ideal value of 1.

We will assume that the dominant time constant in (2.23) equals the sensor's time constant. The following equations can be used to obtain this time constant from either the sensor's rise time or its bandwidth specification:

$$\tau_s = 0.455 t_r \quad \text{and} \tag{2.24}$$

$$\tau_s = \frac{1}{\omega_b} = \frac{1}{2\pi f_b} \tag{2.25}$$

where t_r is the rise time, ω_b is the bandwidth in rad/s and f_b is the bandwidth in Hz.

We will analyze two common time varying inputs. The first is an input that changes suddenly and then remains constant. We can model this as a step input. The question is: How long should our mechatronic system wait before reading the measurement system's output? The answer comes from the measurement system's time response and its accuracy. For a step input applied at $t = 0$, the time response is:

$$y_{out}(t) = y_{out}(0)e^{-\frac{t}{\tau_s}} + K_s\left(1 - e^{-\frac{t}{\tau_s}}\right)y_{true} \tag{2.26}$$

Regarding how long to wait, a rule of thumb is to wait until the effect of the transient on y_{out} drops to $\frac{1}{10}$ of the accuracy. We will term the upper limit of the sensor's range y_{max}, and the corresponding lower limit y_{min}. To analyze the effect of the transient alone we set $K_s = 1$. Of course, we don't know y_{true}. The worst case transient will occur when either $y_{true} = y_{min}$ or $y_{true} = y_{max}$. Our rule of thumb then gives:

$$\begin{aligned}
\left|y_{true} - y_{out}(t)\right| &\le 0.1\left|a_y\right| \\
\left|y_{true} - \left(y_{out}(0)e^{-\frac{t}{\tau_s}} + y_{true}\left(1 - e^{-\frac{t}{\tau_s}}\right)\right)\right| &\le 0.1\left|a_y\right| \\
\left|\left(y_{true} - y_{out}(0)\right)e^{-\frac{t}{\tau_s}}\right| &\le 0.1\left|a_y\right| \\
\max\left(y_{max} - y_{out}(0), y_{out}(0) - y_{min}\right)e^{-\frac{t}{\tau_s}} &\le 0.1\left|a_y\right| \\
e^{-\frac{t}{\tau_s}} &\le \frac{0.1\left|a_y\right|}{\max\left(y_{max} - y_{out}(0), y_{out}(0) - y_{min}\right)} \\
-\frac{t}{\tau_s} &\le \ln\left(\frac{0.1\left|a_y\right|}{\max\left(y_{max} - y_{out}(0), y_{out}(0) - y_{min}\right)}\right)
\end{aligned} \tag{2.27}$$

with the final answer:

$$t \geq -\tau_s \ln\left(\frac{0.1|a_y|}{\max\left(y_{max} - y_{out}(0), y_{out}(0) - y_{min}\right)}\right) \tag{2.28}$$

where a_y is the measurement system's accuracy expressed in the units of the measured quantity.

If the measurement is required sooner, then the worst case error will include a significant transient component and a_y. Again, to analyze the effect of the transient alone we set $K_s = 1$, and the worst case transient will occur when either $y_{true} = y_{min}$ or $y_{true} = y_{max}$. Then from (2.26), for the transient component we have:

$$|y_{out}(t) - y_{true}| \leq \max\left(y_{max} - y_{out}(0), y_{out}(0) - y_{min}\right) e^{-\frac{t}{\tau_s}}$$

Including the measurement system's accuracy, the worst case error is:

$$\Delta y_{out}(t) = |a_y| + \max\left(y_{max} - y_{out}(0), y_{out}(0) - y_{min}\right) e^{-\frac{t}{\tau_s}} \tag{2.29}$$

where $\Delta y_{out}(t)$ is the worst case measurement error at time *t*.

The second common input is a function consisting of one or more sinusoids. We will assume that the sinusoids have an amplitude and phase that are constant enough that only the steady forced response of the sensor needs to be analyzed. For a single sinusoidal input $y_{true}(t) = A_{true} sin(\omega t)$, the steady forced response of (2.23) is given by:

$$\begin{aligned} y_{out}(t) &= A_{out}(\omega) \sin\left(\omega t + \phi(\omega)\right) \\ &= K_s M(\omega) A_{true} \sin\left(\omega t + \phi(\omega)\right) \end{aligned} \tag{2.30}$$

$$M(\omega) = \frac{1}{\sqrt{1 + \omega^2 \tau_s^2}} \text{ and} \tag{2.31}$$

$$\phi(\omega) = -\tan^{-1}(\omega \tau_s) \tag{2.32}$$

where *M* is the magnitude and ϕ is the phase shift in radians. The phase shift is the error in phase between the input and output sinusoids. If required, the time delay due to the phase shift is:

$$t_d = -\frac{\phi}{\omega} \tag{2.33}$$

Quantifying the worst case error in the measured amplitude takes further analysis. The errors introduced by $K_s \neq 1$ and $M \neq 1$ need to be combined. Applying (2.19) to (2.30) and simplifying gives:

$$\Delta A_{out}(\omega) = |a_y| + \left(1 - \frac{1}{\sqrt{1+\omega^2\tau_s^2}}\right) A_{out}(\omega) \tag{2.34}$$

Example 2.4

A measurement system consisting of an accelerometer and signal conditioner has an accuracy of 0.5 m/s^2 and a bandwidth of 750 Hz. If the measured acceleration amplitude is 100 m/s^2 at the bandwidth frequency, what is the worst case error in the measurement? How does this value change if the acceleration frequency is decreased to 75 Hz (*i.e.* $\frac{1}{10}$ of the bandwidth)?

Normally we would need to start by calculating the time constant using (2.25). However, since we are interested in finding the error at the bandwidth frequency we can substitute (2.25) and the given information into (2.34) and solve as follows:

$$\Delta A_{out}(\omega) = |a_y| + \left(1 - \frac{1}{\sqrt{1+\omega^2\tau_s^2}}\right) A_{out}(\omega) = \left|\pm 0.5 \text{ m/s}^2\right| + \left(1 - 1\Big/\sqrt{1+\omega_b^2 \frac{1}{\omega_b^2}}\right)(100 \text{ m/s}^2) = 29.8 \text{ m/s}^2$$

So the worst case error is almost 30% of the measured amplitude when the sinusoid's frequency equals the bandwidth!

Applying the same method when the acceleration frequency is 75 Hz or $\frac{1}{10}$ of the bandwidth gives:

$$\Delta A_{out}(\omega) = |a_y| + \left(1 - \frac{1}{\sqrt{1+\omega^2\tau_s^2}}\right) A_{out}(\omega) = \left|\pm 0.5 \text{ m/s}^2\right| + \left(1 - 1\Big/\sqrt{1+\left(\tfrac{1}{10}\omega_b\right)^2 \frac{1}{\omega_b^2}}\right)(100 \text{ m/s}^2) = 1.0 \text{ m/s}^2$$

The worst case error has been reduced to only 1%. This example shows the important influence of the input frequency on the measurement error.

End of example 2.4.

2.4.3 Errors Caused by Numerical Differentiation

Most of the controllers used for actuator position control employ velocity feedback. As mentioned in subsection 2.2, a popular low cost solution is to estimate the velocity by numerically differentiating the position measurements using backward differencing. In equation form:

$$v_{est}(kT) = \frac{p(kT) - p((k-1)T)}{T} \tag{2.35}$$

where v_{est} is the estimated velocity, p is the measured position, T is the controller's sampling period, k is an integer representing the current sample number, kT is the time of the current sample, and $(k-1)T$ is the time of the previous sample. The sampling period is assumed to be constant and known exactly. Equation (2.35) can be re-written as:

$$v_{est}(kT) = \left(\frac{p_{true}(kT) - p_{true}((k-1)T)}{T}\right) + \left(\frac{e_p(kT) - e_p((k-1)T)}{T}\right) \tag{2.36}$$

where p_{true} is the true position, and e_p is the measurement error. Each term on the right-hand side of (2.36) adds error to the velocity estimation.

The error from the backward difference of p_{true} is known as truncation error. The worst case truncation error equals $\frac{T}{2}\max\left(|a_{true}|\right)$, where a_{true} is the true acceleration. Note that $\max\left(|a_{true}|\right)$ can either be estimated using a dynamic model of the actuator and its mechanical load, or by using an accurate acceleration sensor.

The velocity error due to backward differencing of e_p is a form of amplified noise. By applying (2.22) to the e_p term in (2.36) the worst case error is: $\dfrac{\Delta p(kT) + \Delta p((k-1)T)}{T} = \dfrac{2\Delta p}{T}$. Note that $\Delta p(kT) = \Delta p((k-1)T) = \Delta p$ since the uncertainty is not time dependant.

We can combine the worst case errors from the two terms in (2.36) using (2.22). The worst case velocity error is then:

$$\Delta v_{est} = \frac{T}{2}\max\left(|a_{true}|\right) + \frac{2\Delta p}{T} \tag{2.37}$$

Examining (2.37), it becomes apparent that a smaller T will reduce the truncation error while simultaneously amplifying the position measurement noise. The truncation error is due to the phase lag that is unavoidable with backward differencing. If the velocity estimate is used as feedback then phase lag is undesirable since it reduces the phase margin of the control system. Amplified noise in the feedback can cause the actuator's position to vibrate which is again undesirable. So the choice of T depends on the relative importance of the velocity phase lag and noise to the particular application. If the errors from the two terms in (2.37) are equally important, then the T value that minimizes (2.37) should be used. The equation for the optimal T value is:

$$T_{opt} = \sqrt{\frac{4\Delta p}{\max\left(|a_{true}|\right)}} \tag{2.38}$$

Next, we will use (2.37) to analyze two common scenarios where (2.35) is used.

In the first scenario the position is sensed by an incremental encoder. If other error sources are negligible, then the uncertainty in each position measurement equals ½ of the encoder's position resolution. Substituting this information into (2.37) gives:

$$\begin{aligned}\Delta v_{est} &= \frac{T}{2}\max\left(\left|a_{true}\right|\right)+\frac{2\Delta p}{T} \\ &= \frac{T}{2}\max\left(\left|a_{true}\right|\right)+\frac{\text{encoder's position resolution}}{T}\end{aligned} \tag{2.39}$$

Example 2.5

An encoder produces 400 counts per revolution. Its position resolution[1] is then $360°/400 = 0.9°$. The controller's sampling period equals 0.001 s. The estimated maximum acceleration is 35,500°/s^2. Assuming quantization is the encoder's only source of error, from (2.39) the worst case velocity error is:

$$\begin{aligned}\Delta v_{est} &= \frac{T}{2}\max\left(\left|a_{true}\right|\right)+\frac{\text{encoder's position resolution}}{T} \\ &= (0.001s)(\ 35{,}500°/\text{s}^2)/2 + 0.9°/0.001\text{ s} \\ &= 18°/\text{s} + 900°/\text{s} \\ &= 918°/\text{s or } 153\text{ rpm}\end{aligned}$$

From the above calculation, the velocity error is almost completed caused by the amplified quantization error of the encoder. Figure 2.4.1 shows a computer simulation of the velocity estimated from the encoder measurements plotted on top of the true velocity. We can observe that the estimation error does not exceed 153 rpm as expected. It is also obvious that the error is predominantly high frequency and that the signal to noise ratio is very poor at low velocities.

Now let's try increasing the sampling period to $T = 0.02$ s. The new worst case velocity error is:

$$\begin{aligned}\Delta v_{est} &= \frac{T}{2}\max\left(\left|a_{true}\right|\right)+\frac{\text{encoder's position resolution}}{T} \\ &= (0.02s)(\ 35{,}500°/\text{s}^2)/2 + 0.9°/0.02\text{ s} \\ &= 355°/\text{s} + 45°/\text{s} \\ &= 400°/\text{s or } 67\text{ rpm}\end{aligned}$$

Now the velocity error is dominated by the truncation error. The corresponding velocity plots are shown in Figure 2.4.2. The estimated velocity is much smoother (i.e. less high frequency

[1] If quadrature counting is used then this value should be divided by four.

noise) and much closer to the true velocity. The disadvantage of the slower sampling is that the estimate has a significant phase lag relative to the true velocity.

Finally, let's try the optimal T. Equations (2.38) and (2.39) give:

$$T_{opt} = \sqrt{\frac{4\Delta p}{\max\left(\left|a_{true}\right|\right)}} = \sqrt{\frac{4(\frac{1}{2})(0.9°)}{35,500°/\text{s}^2}} = 0.0071\,\text{s} \text{ and}$$

$$\begin{aligned}\Delta v_{est} &= \frac{T}{2}\max\left(\left|a_{true}\right|\right) + \frac{\text{encoder's position resolution}}{T} \\ &= (0.0071\,\text{s})(35{,}500°/\text{s}^2)/2 + 0.9°/0.0071\text{ s} \\ &= 126°/\text{s} + 126°/\text{s} \\ &= 252°/\text{s or } 42\text{ rpm}\end{aligned}$$

The optimal T has made the worst case velocity error much smaller, and has also made the error contributions from the two sources equal. The improved performance is plotted in Figure 2.4.3

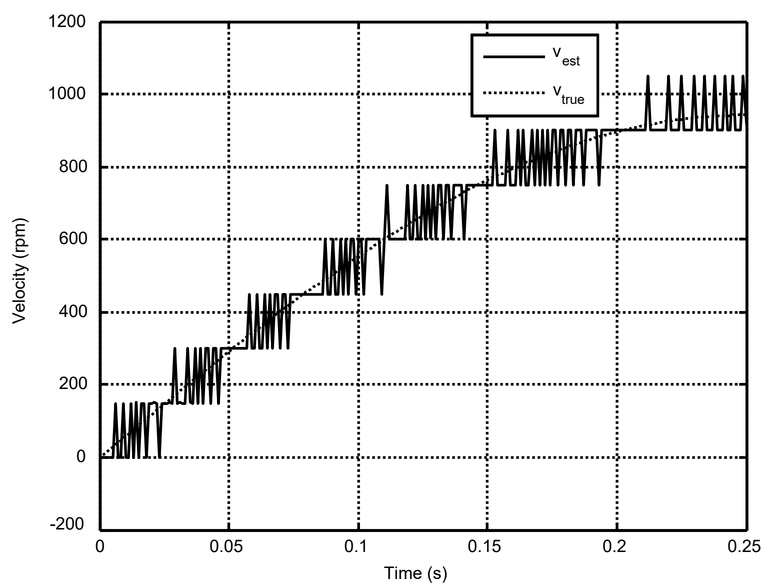


Figure 2.4.1 Estimated velocity and true velocity for example 2.4 with $T = 0.001$ s.

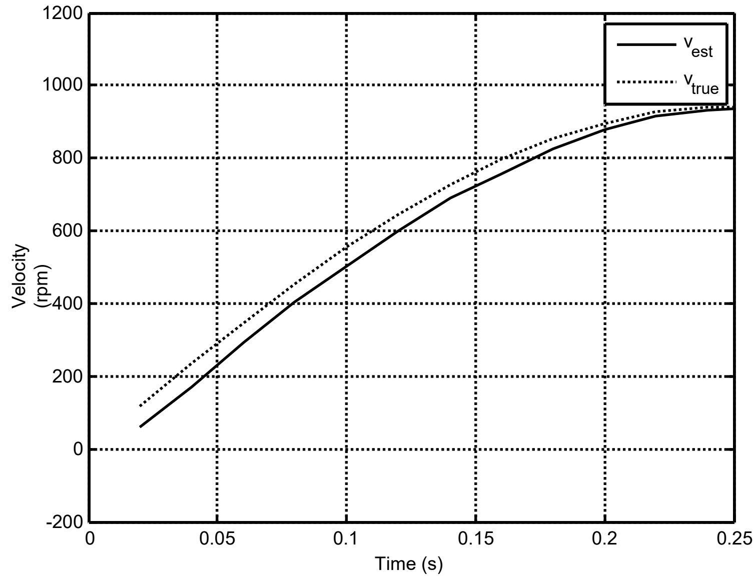


Figure 2.4.2 Estimated velocity and true velocity for example 2.4 with $T = 0.02$ s.

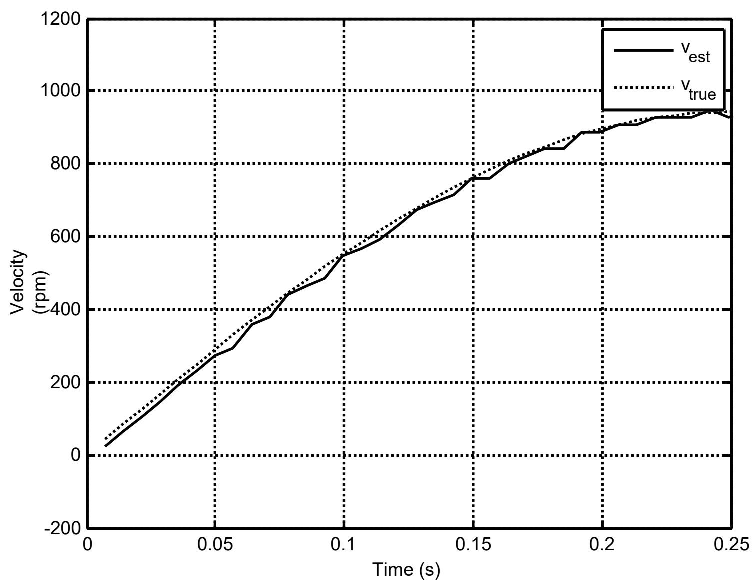


Figure 2.4.3 Estimated velocity and true velocity for example 2.4 with $T = T_{opt}$.

End of example 2.5.

In the second scenario a position sensor with an analog output is used. The output contains normally distributed random noise with a standard deviation of σ_p. This is assumed to be the only source of error. The uncertainty can then be approximated as $\Delta p = 3\sigma_p$. This corresponds to the 99.7% confidence interval. Substituting this information into (2.37) gives:

$$\Delta v_{est} = \frac{T}{2}\max\left(\left|a_{true}\right|\right) + \frac{6\sigma_p}{T} \tag{2.40}$$

Example 2.6

A multi-turn potentiometer is used to measure the position of a motor's shaft. The position measurement error is normally distributed with 0.5° standard deviation. The sampling period equals 0.002 s. The estimated maximum acceleration is 35,500°/s^2. From (2.40) the corresponding worst case velocity error is:

$$\begin{aligned}\Delta v_{est} &= \frac{T}{2}\max\left(\left|a_{true}\right|\right)+\frac{6\sigma_{\mathrm{p}}}{T}\\ &= (0.002\text{ s})(35{,}500°/\text{s}^2)/2 + 6(0.5°)/0.002\text{ s}\\ &= 36°/\text{s} + 1500°/\text{s}\\ &= 1536°/\text{s or } 256\text{ rpm}\end{aligned}$$

Figure 2.5.1 shows a computer simulation of the velocity estimated from the potentiometer measurements plotted on top of the true velocity. We can observe that the estimation error does not exceed 256 rpm as expected. As with the encoder, the signal to noise ratio is very poor at low velocities.

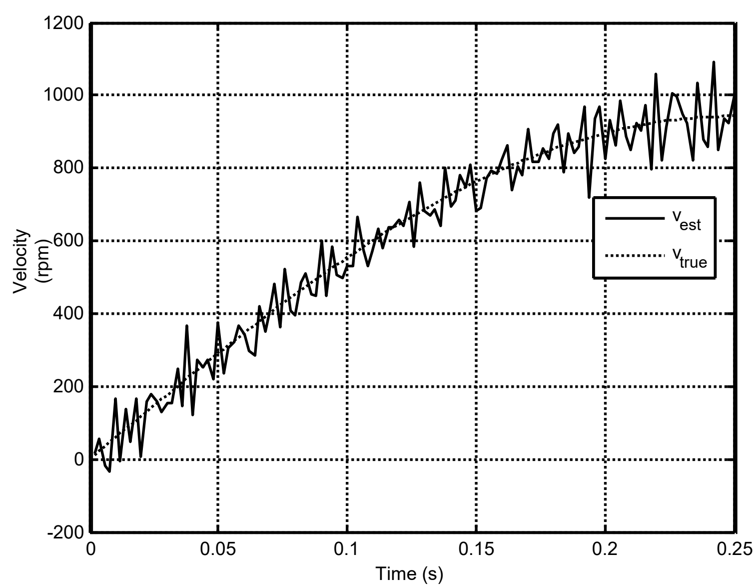


Figure 2.5.1 Estimated velocity and true velocity for example 2.5.

End of example 2.6.

2.5 Sensor Selection

The factors to consider when selecting a sensor include:

- Required performance in terms of standard specifications and those covered in section 2.4.
- Interfacing needs: signal conditioning requirements, power requirements, compatibility, etc.

- Physical properties: weight, size, etc.
- Environmental conditions: e.g. temperature, vibration, dirt.
- Quality Factors: reliability, durability, maintainability, lifespan, etc.
- Cost: expense, availability.
- Other application specific factors.

- With a mechatronic design, the sensor cannot be selected in isolation
    - We must consider actuator.
    - We must consider μC and interface electronics.