Chapter 2
TRANSDUCTION OF INFORMATION

- ***What is special about transducers, sensors and actuators?***
- ***How are self-generating and modulating sensors different?***
- ***How should sensor specifications be interpreted?***
- ***How are strain gauges used?***
- ***What are my options for resistive sensor read-out?***
- ***How are capacitive sensors read-out?***
- ***What are piezo-electric sensors like and how can these be used?***

2.1 Introduction

Information is embedded in a material (such as: text on paper, a hard disk or a CD-ROM) or is contained in one of the energy domains mentioned in Chapter 1 (such as: video signals in the electrical domain, biological signals, as processed in our nervous system, in the chemical domain (ion concentrations), etc.). Often the material or energy carrying the information is not suitable for a particular application or use and some means of transferring the information is required.

One example is a scanner which can transfer hard copy text into an electronic text file. The efforts made over the last few decades to scan complete libraries demonstrates the universal trend away from traditional paper-based documents towards electronic archives.

Electronic instruments for measuring non-electrical quantities serve a similar purpose, since the transfer of the non-electronic information into its electrical equivalent is required for further information processing. The operation of transferring information from one energy carrier to another is referred to as **transduction**. The component in which this is implemented is the **transducer**.

Transduction is the transfer of information from one energy carrier to another.

The term 'transducer' refers to a component that is able to transfer information from any input domain to any output domain and is, therefore, not limited to the transfer to or from the electrical domain. For this reason the term **sensor** is used for the **input transducer** in the case of the transduction of information from a non-electrical domain into the electrical domain. Similarly, the **actuator** is the **output transducer** in the case of a transduction from the electrical domain into the non-electrical domain.

In the literature a transducer that transfers from the electrical domain to the electrical domain (such as an electrode for ECG recording -see Section 1.2.2) is also sometimes referred to as a sensor. However, the signal processing remains within the electrical domain. Therefore, this approach is not adopted here in order to clearly distinguish between electrical measurement and non-electrical measurement.

Obviously, the information in the non-electrical domain is of limited quality. For instance, damping in the mechanical domain results in mechanical noise, which limits the detectivity even in an perfect measurement system. Practical sensor limitations may adversely affect the detection limit in several ways:

1. The transduction is non-selective (i.e. the sensor is also sensitive to a parasitic quantity).
2. The transducer interferes with the non-electrical quantity (e.g. source loading when using a bulky thermometer to measure the temperature of a small volume of water).
3. Uncertainty is added by the sensor (e.g. noise generated in a resistive sensor)
4. Uncertainty is added due to read-out (e.g. source loading at the sensor electrical output or injection of Electro-Magnetic Interference (EMI) via the power supply lines).

Some general considerations and sensor characteristics are presented in this chapter, along with a discussion on sensor specifications and circuits that are particularly suitable for the read-out of a certain sensor type. Due to the wide range of transduction principles available, it is not possible to provide an exhaustive overview of all the possible sensor types.

2.2 Sensor structures

As shown in Fig. 2.1a, a sensor includes three parts:

- a **non-electrical input port**, which is connected in some way to a non-electrical input parameter,
- an **electrical output port**, which can provide an electrical output signal or quantity and
- a **functional block representing the transduction effect**, which enables information in a selected input quantity in a non-electrical domain to be transferred to the electrical output quantity. Several quantities are of interest within a particular input domain (e.g. displacement, velocity, acceleration, force, momentum, etc., in the mechanical domain).

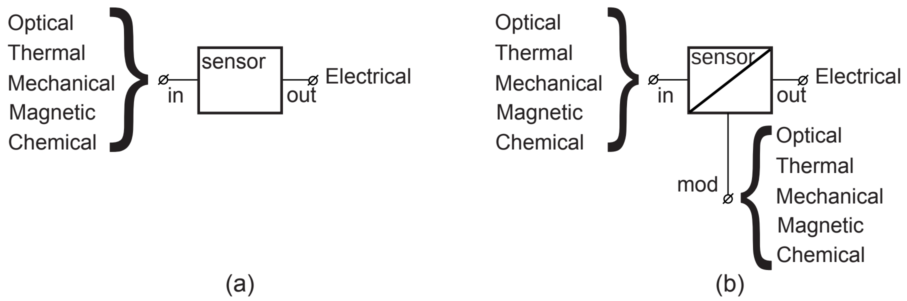


Figure 2.1, Functional block diagram of (a) the (self-generating) sensor and (b) the modulating sensor.

The **input port** should enable the extraction of information from the non-electrical signal domain without influencing the source of information (preventing loading of the source). The input port can often be presented as an impedance in the applicable signal domain.

The electrical quantity at the **output port** is usually either:

- a voltage or current source or
- an electrical component.

In the case of a voltage or current source as the output, a change in voltage or current level is generated to be proportional to the non-electrical input, whereas the electrical component (this is usually a resistor, a capacitor or an inductor) provides a change in its value that is proportional to the non-electrical input.

The sensor output port is not ideal. If the sensor provides an output voltage, this voltage source is usually associated with a source series impedance. Similarly, a sensor with a current output has a parallel impedance. Actually, sensors such as thermocouples are often highly non-ideal in the sense that the sensor has a voltage output that provides a low-level signal and has a relatively large series resistance. Another example is the reverse-biased photodiode, which has a current output with a relatively low source impedance. These characteristics have to be included in the detectivity analysis and the design of the read-out circuits.

The sensor with the resistor as the output port should have the resistance change be proportional to the input signal. However, there is often a parasitic component, such as a capacitance, in parallel. This parasitic component has to be taken into consideration in the read-out circuits used for converting the component value change into an electrical signal (for instance, it limits the maximum frequency to be used in the current source in the case of AC excitation).

The transduction effect used in the sensor should enable the **selective measurement** of the intended non-electrical quantity only and should do so unambiguously (only one value of the output parameter as a function of each input parameter value). The sensor should maintain this performance over the entire **dynamic range** of the input signal.

The dynamic range of a component is the ratio between the maximum and minimum signal levels that can be processed at a given inaccuracy (or SNR) specification.

The maximum signal level is usually set by distortion that is introduced when operating the electronic system close to the power supply level (refer to Section 1.6), whereas the minimum level is determined by the detection limit due to, for instance, noise.

The units of the sensor transfer function are expressed in terms of the units of the input and output quantities. For example in the case of a force sensor that supplies an output voltage, the units would be [V/N].

The direct implementation of the transfer function is shown in Fig. 2.1a and is referred to as a sensor based on a **self-generating effect**.

A self-generating sensor supplies an output signal that is proportional to the input signal, without the use of an intermediate signal of a different domain.

The energy involved in the transduction is derived from the input-signal domain, which may influence the input quantity (source loading). In a critical application this could give rise to a measurement error. Examples of self-generating transducers are thermocouples and photodiodes, as shown in Fig. 2.2a.

The alternative type of sensor is based on a **modulating effect** and is shown schematically in Fig. 2.1b.

The modulating sensor uses the input signal to modulate the energy transfer of an auxiliary quantity to the output.

An example of a modulating sensor is the optical linear encoder shown in Fig. 2.2b. The displacement (input= mechanical) interrupts the optical path between a light source (the auxiliary quantity is from the radiant domain) and a photodetector (output= electrical), thus modulating the opto-electrical conversion.

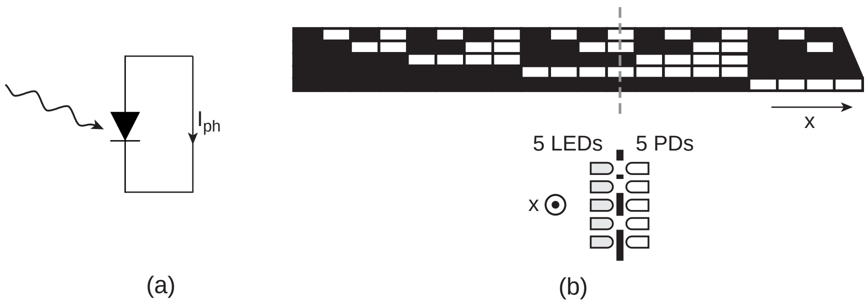


Figure 2.2, (a) The photodiode is a self-generating sensor and (b) the 5-bit linear encoder is an example of a modulating sensor with five Light Emitting Diodes (LEDs) as the light sources and five Photo Diodes (PDs) as the detectors.

The advantage of the modulating sensor is the feasibility of high conversion efficiency. As the energy required for the transduction is not supplied by the input quantity, a higher sensitivity is feasible. Moreover, loading of the input quantity can be avoided (compare the source loading of a non-contact optical encoder with the source loading of a mechanically coupled potentiometer). The disadvantage results from the uncertainty introduced by the additional quantity. In the optical encoder one defective light source affects the angular position measurement.

It should be noted that an auxiliary electrical component is required for the read-out of a resistive sensor. A current source should be connected to the resistor to generate a voltage proportional to the resistance, which can be read-out. Since the resistor and the current source are both electrical components and no third energy domain is involved, this is not a modulating sensor. The current source should formally be considered part of the read-out circuits, as is discussed in more detail in Section 2.4. The same argument basically also applies to the excitation voltage sources used in the read-out of capacitive sensors, as is elaborated in Section 2.6.

The transducer limitations listed in the previous section should be expressed in terms of sensor specifications. Sometimes limitations can be interchanged and optimum overall performance can be achieved by a balanced compromise and a clever design. Selecting, for instance, the modulating sensor type could reduce some of the source loading effects, yet, is usually at the expense of additional sources of uncertainty. Therefore, making design choices requires knowledge about practical transduction and practical limitations, such as:

- The input impedance in terms of lumped components in the non-electrical domain (for instance heat capacitance and heat conductance in the thermal domain to analyse source loading in temperature sensing with a thermometer in contact with the object)
- Cross-sensitivity to the non-selected parameter (often temperature)
- Offset and non-linearity in the transfer function
- The electrical output source impedance (to enable the analysis of the effect of read-out circuits, noise and EMI).

Transduction and cross-sensitivities are discussed in the next section. The most prevalent sensor limitations and their effect on overall performance are discussed in the following sections.

2.3 Describing transduction effects

2.3.1 The transduction matrix

The non-idealities in the transduction can be included in the transfer function by using a **transduction matrix** T, with (OUT)= (T)(IN)+(OS):

$$\begin{pmatrix} Ma \\ Me \\ Th \\ Opt \\ Ch \\ El \end{pmatrix} = \begin{pmatrix} t_{ma,ma} & t_{me,ma} & t_{th,ma} & t_{opt,ma} & t_{ch,ma} & t_{el,ma} \\ t_{ma,me} & t_{me,me} & t_{th,me} & t_{opt,me} & t_{ch,me} & t_{el,me} \\ t_{ma,th} & t_{me,th} & t_{th,th} & t_{opt,th} & t_{ch,th} & t_{el,th} \\ t_{ma,opt} & t_{me,opt} & t_{th,opt} & t_{opt,opt} & t_{ch,opt} & t_{el,opt} \\ t_{ma,ch} & t_{me,ch} & t_{th,ch} & t_{opt,ch} & t_{ch,ch} & t_{el,ch} \\ t_{ma,el} & t_{me,el} & t_{th,el} & t_{opt,el} & t_{ch,el} & t_{el,el} \end{pmatrix} \begin{pmatrix} Ma \\ Me \\ Th \\ Opt \\ Ch \\ El \end{pmatrix} + \begin{pmatrix} Ma_{os} \\ Me_{os} \\ Th_{os} \\ Opt_{os} \\ Ch_{os} \\ El_{os} \end{pmatrix} \tag{2-1}$$

This transduction matrix, T, describes all the applicable transduction effects from all six signal domains at the input (*IN*) to all six signal domains at the output (*OUT*). The subscript denotes the signal domains involved (*ma*= magnetic, *me*= mechanical, *th*= thermal, *opt*= optical, *ch*= chemical and *el*= electrical).

The transduction is described by the domain ranging from the first code in the subscript to the second (e.g. $t_{th,el}$ describes the thermo-electrical effect in a material), while the offset in the six domains is described by the column vector (*OS*) (e.g. Me_{os} is the elongation in a spring-based force-to-displacement transducer due to residual mechanical stress in the material). The value of an element on the diagonal of T is basically the gain factor in the respective signal domain.

The transduction matrix T of a sensor is a five-element row vector, since the electrical domain is the only output and is also not applicable at the input. Therefore, expression 2.1 reduces to:

$$El_{out} = t_{ma,el}Ma_{in} + t_{me,el}Me_{in} + t_{th,el}Th_{in} + t_{opt,el}Opt_{in} + t_{ch,el}Ch_{in} + El_{os} \qquad (2\text{-}2)$$

The intended transduction effect is the transfer function of the desired input parameter for the electrical output. All other transfer functions are **cross-sensitivities**.

The cross sensitivity of a sensor specifies the mapping of a parasitic (i.e. non-intended) non-electrical input onto the electrical output port.

A photodiode should only be sensitive to the intensity of incident light. Hence, $I_{ph}= R \times P_{opt}$, where I_{ph} denotes the photocurrent, P_{opt} the impinging optical power and R the **responsivity** of the diode. Insertion in Eqn. 2.2 yields: $El_{out}= I_{ph}$ and $t_{opt,el}= R$.

However, a practical device also reveals cross-sensitivities. The photodiode exhibits a parasitic sensitivity to temperature ($I_{ph}= F_T \times T$), where F_T [A/K] denotes a highly non-linear function of temperature, and T is the absolute temperature [K].

The current in the non-illuminated photodiode is often referred to as the **dark current**, I_d, and is basically the offset component. The magnetic-electrical, mechanical-electrical and chemical-electrical cross-effect in the photodetector (including the connection wires) can usually be disregarded ($t_{ma,el}= t_{me,el}= t_{ch,el}= 0$). Hence:

$$I_{ph} = t_{th,el}Th_{in} + t_{opt,el}Opt_{in} + El_{os} = F_T.T + R.P_{opt} + I_d \qquad (2\text{-}3)$$

2.3.2 Tandem transducers

For some non-electrical quantities no suitable direct transduction effect is available, which result in an electrical output signal. In this case one has to resort to a chain of two or more transduction effects. Such a sensor is generally referred to as a **tandem transducer**.

The tandem transducer contains two subsequent transduction effects to transfer the information from the input signal domain to the output signal domain by means of an intermediate signal domain that is different from either the input or the output.

A temperature sensor based on the capacitive measurement of the deflection of a **bi-metal** upon heating is a typical example of a tandem transducer. Two metals with different thermal expansion coefficients are glued together. A temperature increase (thermal) causes a deflection (mechanical). The resulting change in capacitance is the electrical output signal. This device is an infrared detector (bolometer) when using an absorbing coating, as shown in Fig. 2.3a.

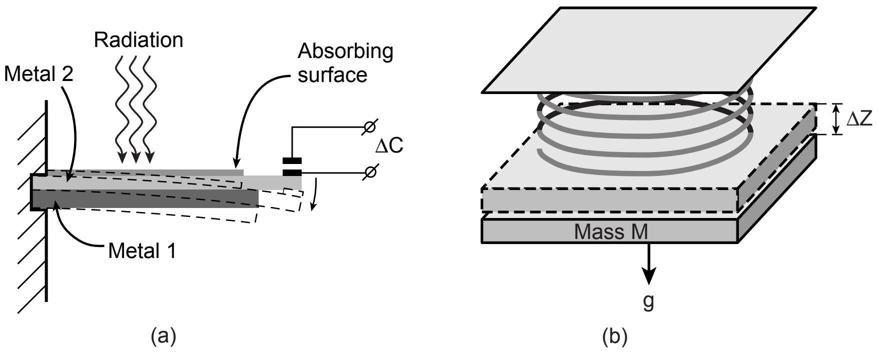


Figure 2.3, (a) Temperature sensing via capacitive displacement measurement using a bi-metal (tandem transducer) and (b) acceleration sensor via measurement of spring displacement due to force acting on the seismic mass (NOT a tandem transducer).

An essential characteristic of the tandem transducer is that each of the sub-transductions results in a change of signal domain. The sensor is actually a simple signal processor (modifier or circuit) when the input and output signals are within the same signal domain (e.g. an amplifier in that particular domain).

An example is a mechanical spring, which deflects when a force is applied. A force sensor results when measuring the spring deflection using a displacement sensor. If the force is due to an acceleration acting on a seismic mass, an accelerometer results, as depicted in Fig. 2.3b. It should be noted that this is not a tandem transducer. The conversion of acceleration into displacement is within the mechanical domain. The sensitivity is fully determined by the mechanical components, such as the mass and the stiffness of the spring.

Example 2.1
An accelerometer is based on a seismic mass attached to a spring (Fig. 2.2b). A displacement sensor is used with a measurement range in between –1 cm and 1 cm, which provides a linear output voltage, U_o, in the range between –2V and 2V. The moving part of the sensor (the slider) is mechanically coupled to a mass M= 0.1 kg at the free end of a spring with the spring constant $c_F = 1/k_F = 2\times10^{-3}$ m/N (k_F denotes the spring stiffness). Calculate the sensitivity of the accelerometer.

Solution:
$\partial U_o/\partial z$ = 2 V/1 cm = 200 V/m, $\partial U_o/\partial a_z = (\partial U_o/\partial z)\times(\partial z/\partial F)\times(\partial F/\partial a_z)$=
(4 V/2 cm)×c_F×M= 200 V/m× 2×10^{-3} m/N× 0.1 N/m/s= 40 mV/m/s^2.

2.4 Sensor with electronic signal conditioning

Electronic signal conditioning is added to the sensor output port for read-out. The general reasons for these front-end analog circuits are:

1. Reducing source loading
2. Providing gain to match the dynamic range of the sensor output signal to the available dynamic input range of the AD converter.
3. Providing frequency filtering to maximize the SNR.
4. Adding functionality not provided by the sensor.

These are discussed in the following sub-sections.

2.4.1 Source loading in voltage and current measurement

Measuring the voltage level at the output of a sensor in the case of a non-zero source impedance, Z_s, using a suitable read-out circuit or an instrument with a relatively low input impedance, Z_i, as shown in Fig. 2.4a, results in an error, which is due to loading of the source by the instrument.

The transfer function is described by:

$$u_m = \frac{Z_i}{Z_s + Z_i} u_g = \left(1 - \frac{Z_s}{Z_s + Z_i}\right) u_g \simeq \left(1 - \frac{Z_s}{Z_i}\right) u_g = u_g(1 - \varepsilon_1) \qquad (2\text{-}4)$$

where ε_1 denotes the multiplicative error due to source loading (= scale error).

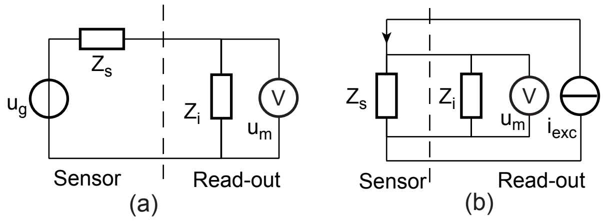


Figure 2.4, (a) Error due to voltage source loading and (b) error in the sensor impedance measurement using current excitation and voltage read-out.

A sensor impedance is often composed of the intended (resistive) part in parallel to a parasitic (capacitive) part. Also the input impedance of the voltage read-out is often composed of a resistive part in parallel to a capacitive part. This combination of a source impedance and a read-out input impedance results for $R_s \ll R_i$ in a frequency-dependent error, which is described by:

$$\varepsilon_1 = \frac{R_s}{R_s + R_i} \frac{1 + j\omega R_i C_i}{1 + j\omega R_i R_s (C_i + C_s)/(R_i + R_s)} \simeq \frac{R_s}{R_i} \frac{1 + j\omega R_i C_i}{1 + j\omega R_s C_i (C_i + C_s)} \tag{2-5}$$

Example 2.2
The input impedance of a Volt meter is specified as R_i= 10 MΩ and C_i= 10 pF. Measuring the voltage across a resistive sensor with R_s= 10 kΩ results in a DC error $\varepsilon_1 = R_s/R_i$= 0.1%. However, for $\omega > 1/R_s C_i$, the error increases to: ε_1=100%, which limits the operating frequency.

The source loading error also results in the readout of a sensor where the information is contained in an impedance Z_s, while current excitation and voltage measurement across the impedance is used for read-out. The current source has a finite source impedance, Z_{i1}, which is in parallel to the input impedance of the Volt meter, Z_{i2}, as is shown in Fig. 2.4b. This is basically the four-terminal resistance set-up that is usually implemented in multimeters. The transfer function results in:

$$u_m = \frac{Z_s Z_i}{Z_s + Z_i} i_{exc} = Z_s \times i_{exc} \left(1 - \frac{Z_s}{Z_s + Z_i} \right) = Z_s \times i_{exc} (1 - \varepsilon_1) \tag{2-6}$$

Similarly, read-out of a current source, i_s, with finite value for the source impedance, Z_s, while using a read-out circuit with non-zero input impedance, Z_i, as is indicated in Fig. 2.5a, results in an error. The transfer function for $Z_i \ll Z_s$ is described by:

$$i_m = i_s \frac{Z_s}{Z_s + Z_i} = i_s\left(1 - \frac{Z_i}{Z_s + Z_i}\right) \simeq i_s\left(1 - \frac{Z_i}{Z_s}\right) = i_s(1 - \varepsilon_2) \tag{2-7}$$

This expression also results in the read-out of the sensor impedance using voltage excitation and current read-out using an Ampère meter with a non-zero impedance series with the sensor impedance, as shown in Fig. 2.5b.

$$\frac{u_{exc}}{i_m} = \frac{u_{exc}}{u_{exc}/(Z_s + Z_i)} = Z_s + Z_i = Z_s\left(1 + \frac{Z_i}{Z_s}\right) \rightarrow i_m = \frac{u}{Z}(1 + \varepsilon_2) \tag{2-8}$$

It should be noted that this is in principle a deterministic error, which can be avoided if the impedances are known (specified) prior to the measurement and correction is applied (see Example 1.2). However, to simplify instrument use, usually this is deliberately not implemented and the loading is specified as an error. For this reason the polarity of this error is not relevant.

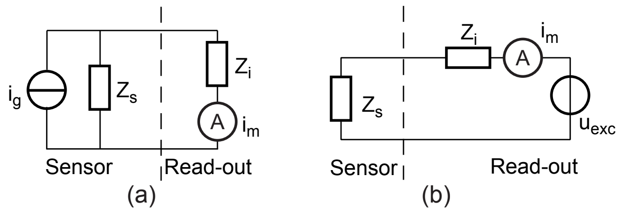


Figure 2.5, (a) Error due to current source loading and (b) error in sensor impedance read-out with voltage excitation.

2.4.2 Gain for covering ADC input dynamic range

The quantisation error, ε_q, in an Analog-to-Digital converter (ADC) is half the quantisation interval: $\varepsilon_q = (U_{i,max}-U_{i,min})/([2(2^n-1)]$, where n denotes the resolution in terms of the number of bits. This uncertainty is reduced by increasing the number of bits in the ADC, i.e. improving its resolution.

However, a poorly designed data-acquisition sub-system with an ADC input range exceeding well beyond the range of the sensor output, as shown in Fig. 2.6a, results in a large quantisation error irrespective of the resolution of the ADC. The function of an amplifier in signal conditioning is to match these signal ranges to yield minimum quantisation uncertainty at a specified resolution of the ADC, as is shown in Fig. 2.6b.

The **non-inverting voltage amplifier** shown in Fig. 2.7 is very suitable for the read-out of a resistive sensor, while providing gain. An excitation current, I_{exc}, is

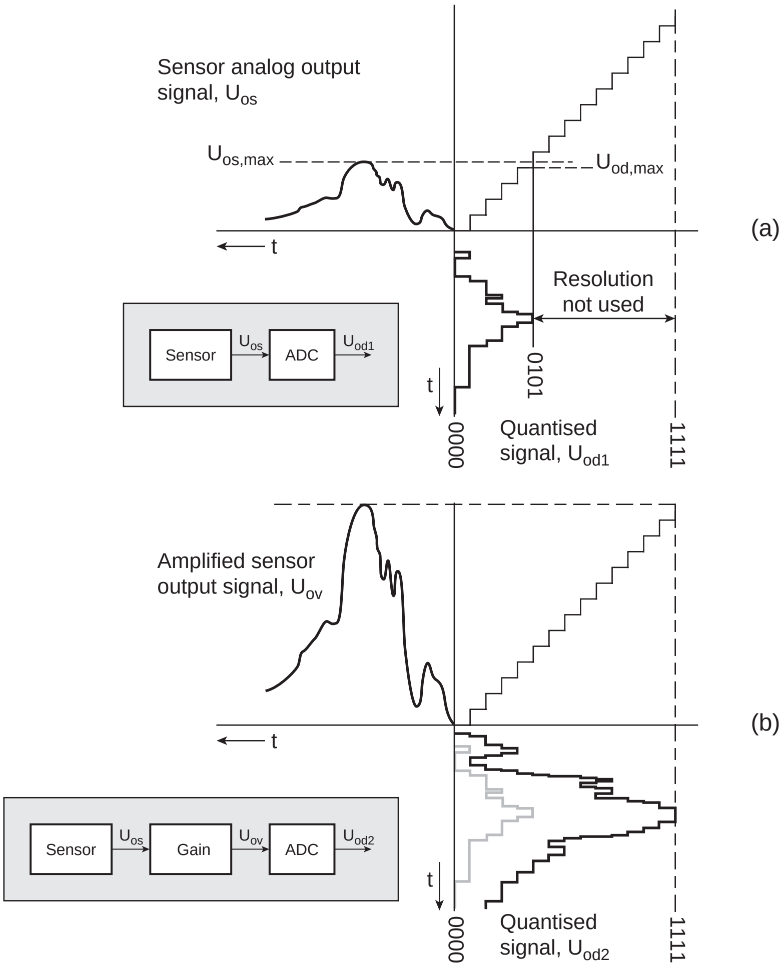


Figure 2.6, Quantisation of an input signal: (a) without gain and (b) with gain to match the sensor output signal to the input dynamic range of the ADC.

used to generate an input voltage $U_i = I_{exc}R_s$, with R_s the resistive sensor. U_i is connected to U_+ and U_o via a resistive voltage divider connected to U_-. When assuming ideal OPerational AMPlifier (OPAMP) characteristics, the transfer function results in:

$$\frac{U_o}{I_{exc}} = R_s \frac{R_1 + R_2}{R_1} \tag{2-9}$$

The desired gain, G, can be implemented using a proper choice of the feedback components and depends on the passive component values only.

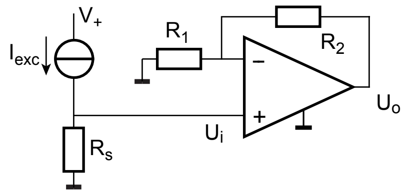


Figure 2.7, Non-inverting amplifier for resistive sensor read-out.

Another feature of the non-inverting amplifier is the very large value of the input impedance. Since the feedback ensures $U_+ = U_-$, no current flows into the non-inverting input terminal of the opamp. Consequently, U_i is not loaded by the input impedance of the read-out circuit. Alternative circuits for sensor signal conditioning are discussed in more detail in Chapter 6.

It should be noted that the implementation of gain in a measurement system for a non-electrical quantity is also possible using a non-electronic component. An example is the lever for force measurement in the mechanical domain.

2.4.3 Filtering for SNR

The benefits of frequency filtering on a sensor signal become evident from a simple analysis. Assume a sensor output signal with spectral power density $N_s(\omega)$ and bandwidth ω_s, containing thermal noise with spectral noise of spectral power density $N_n(\omega)$, as shown in Fig. 2.8.

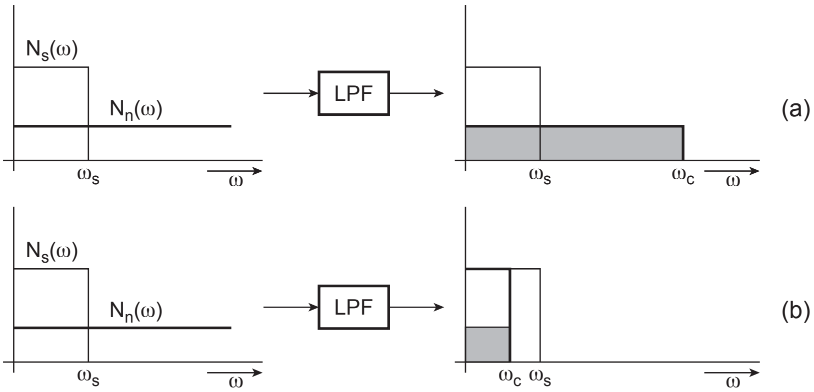


Figure 2.8, (a) Enhancing SNR by low-pass frequency filtering at (a) $\omega_c > \omega_s$ and (b) reducing information content by frequency filtering within the baseband at $\omega_c < \omega_s$.

The signal power is fixed and equal to: $P_s = N_s(\omega)\times\omega_s$, whereas the noise power increases with the filter bandwidth: $P_n = N_n(\omega)\times\omega_c$. Therefore, the signal-to-noise ratio $SNR= P_s/P_n= N_s(\omega)\times\omega_s/[N_n(\omega)\times\omega_c]$ increases linearly with decreasing ω_c. This only applies for $\omega_c> \omega_s$, since the signal content is affected if $\omega_c< \omega_s$, as is shown in Fig. 2.8b. Optimum signal conditioning is achieved at $\omega_c= \omega_s$.

2.4.4 Using circuits for adding functionality to the sensor

Adding circuits to the sensor can also be used to modify the sensor functionality electronically. This technique is especially suitable if the non-electrical quantity of interest has to be derived from another non-electrical quantity.

An example is in the mechanical domain, where the acceleration, a_z, is the time derivative of velocity, v_z, which in turn is the time derivative of displacement in the z-direction ($a_z = \partial v_z/\partial t= \partial^2 z/\partial t^2$). Therefore, an acceleration sensor can be constructed based on a displacement sensor and using two electronic time differentiators.

Alternatively, a displacement sensor that is based on an accelerometer and two electronic integrators can be implemented in the inertial navigation systems used in aviation. The detection limit of the overall quantity, of course, strongly depends on the specifications of the electronic circuits. The main problem of the electronic differentiator is the sensitivity to high-frequency noise. The performance of the electronic integrator is severely affected by offset.

Example 2.3
A displacement sensor is used to measure the amplitude of a sine-wave acceleration at frequency ω= 1 rad/s using two cascaded electronic differentiators. The displacement sensor generates an output signal equal to: $U_{os}= 1\times\sin(1\times t)+ 0.001\times\sin(100\pi t)$, with the second term originating from capacitive coupling between the sensor and the power lines (see Section 4.2.3).

Question:
Calculate the Signal-to-Noise Ratio (SNR) directly at the output of the displacement sensor and of the acceleration signal output, when using an ideal differentiator: $U_o = \partial U_i/\partial t$.

Solution:
$\partial U_{os}/\partial t = \cos(t)+ 0.001\times 100\pi\times\cos(100\pi t)$
$\partial^2 U_{os}/\partial t^2 = -\sin(t) - 0.001\times(100\pi)^2\times\sin(100\pi t)$,
$SNR = 1/[0.001\times(100\pi)^2] = 1/(10\pi^2)$.
This measurement set-up is, therefore, not practical. However, opamp circuits can be designed to act as a differentiator only within a restricted (low) frequency range.

2.5 Sensors with a resistive output port

Many frequently used sensors are basically a resistor with a resistance value that depends on the input quantity. These sensors are often referred to as **resistive sensors**. Typical specifications and techniques for the read-out of these resistive sensors are described in this section.

2.5.1 Sensor resolution

Resolution is the specification that is about the minimum change in the measurand that can be observed.

Resolution is specified EITHER as the minimally distinguishable increment at the output OR as the smallest possible increment at the input that results in a noticeable change in the output -whichever is larger.

Quite often the resolution specification is a characteristic feature of the display of the instrument. The resolution of an instrument with a numeric display can be readily determined by counting the number of digits. A display composed of five digits has a resolution equal to 10^{-5}. An analog display, such as a pointer indicator has in principle a very good resolution, as any value can in theory be indicated using an infinitely small angular displacement of the pointer.

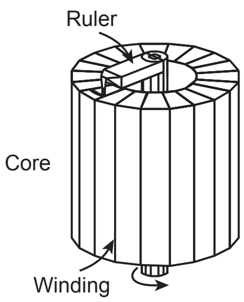


Figure 2.9, Resolution of a wire-wound potentiometer.

Resistors are based on a thin film on a substrate or on a wire wound around a core. The thin-film resistor can in principle be contacted by the ruler at any position. Therefore, it has a very good resolution (similar to the pointer indicator). The wire-wound resistor is usually more reliable and is considered the higher-quality component. However, the resolution of a wire-wound potentiometer used for angular displacement sensing and shown in Fig. 2.9 results directly from the construction. A wire is wound around a core and the ruler runs from wire to wire when rotated. The minimum change in resistance between one of the terminal contacts and the ruler contact is equal to the resistance of one winding. The relative resolution is, therefore, equal to this resistance per winding divided by the total resistance of the wire wound around the core. This is simply equal to the reciprocal of the number of windings: resolution= (number of windings)$^{-1}$.

When a slider moves over the thin-film resistor, the resistance changes continuously. Therefore, the thin-film resistor has better resolution. This does not necessarily make the thin-film resistor the more accurate component. Resolution and inaccuracy are, therefore, not interchangeable.

2.5.2 Sensor repeatability

Another important feature of the sensor is the **repeatability**, which is also specified as **precision** or **stability.** Usually two parts are identified: the **repeatability** over **time** and the **repeatability** with respect to the most significant cross-sensitivity: the **temperature**.

The repeatability of a system indicates the maximum difference between two subsequent measurements when performed in identical conditions.

The critical part of this definition is the term *identical conditions*. Two aspects of the measurement conditions are generally out of the control of the operator of an instrument:

- The time interval between two measurements, Δt, and the ageing of the instrument (= repeatability(t)).
- The change in ambient temperature between two measurements, ΔT, and the temperature sensitivity of the instrument (= repeatability(T)).
-

The ageing and temperature dependence combined, therefore, determine the repeatability: Repeatability= (repeatability(t)$\times\Delta t$ + repeatability(T)$\times\Delta T$). The units used are: [%/month] or [mV/year] for repeatability(t) and [%/K] or [µV/K] for repeatability(T).

Certain types of sensors feature an intrinsically high stability. The stability of a platinum resistive temperature sensor is good enough to enable the use of standardised tables to list the resistance of 100.00 Ω (the number of significant digits indicates the grade) element versus temperature, as is shown in Table 2.1, where S denotes the differential sensitivity, $\partial R/\partial T$, at temperature T. This sensor is generally referred to as the **Pt-100 probe**. The resistance at a particular temperature results from: $R(T)= R(0^oC)+ TCR(T)\times(T-0^oC)$.

Table 2.1

T [°C]	R(T) [Ω]	S [Ω/K]	T [°C]	R(T) [Ω]	S [Ω/K]	T [°C]	R(T) [Ω]	S [Ω/K]	T [°C]	R(T) [Ω]	S [Ω/K]
-200	18.49	0.44	-100	60.25	0.41	±0	100.00	0.39	100	138.50	0.38
-190	22.80	0.43	-90	64.30	0.40	10	103.90	0.39	110	142.29	0.37
-180	27.08	0.42	-80	68.33	0.40	20	107.79	0.39	120	146.06	0.38
-170	31.32	0.42	-70	72.33	0.40	30	111.67	0.39	130	149.82	0.38
-160	35.53	0.42	-60	76.33	0.40	40	115.54	0.39	140	153.58	0.37
-150	39.71	0.42	-50	80.31	0.39	50	119.40	0.38	150	157.31	0.38
-140	43.87	0.41	-40	84.27	0.40	60	123.24	0.38	160	161.04	0.38
-130	48.00	0.41	-30	88.22	0.40	70	127.07	0.38	170	164.76	0.37
-120	52.11	0.41	-20	92.16	0.39	80	130.89	0.38	180	168.46	0.37
-110	56.19	0.41	-10	96.09	0.39	90	134.70	0.38	190	172.16	0.37
200	175.84	0.37	300	212.02	0.35	400	247.04	0.34	500	280.90	0.33
210	179.51	0.37	310	215.57	0.36	410	250.48	0.34	510	284.22	0.33
220	183.17	0.36	320	219.12	0.35	420	253.90	0.34	520	287.53	0.33
230	186.82	0.36	330	222.65	0.35	430	257.32	0.34	530	290.83	0.33
240	190.45	0.36	340	226.17	0.35	440	260.72	0.34	540	294.11	0.33
250	194.07	0.37	350	229.67	0.35	450	264.11	0.34	550	297.39	0.33
260	197.69	0.36	360	233.17	0.35	460	267.49	0.34	560	300.65	0.33
270	201.29	0.36	370	236.65	0.35	470	270.86	0.34	570	303.91	0.32
280	204.88	0.35	380	240.13	0.34	480	274.22	0.33	580	307.15	0.32
290	208.45	0.36	390	243.59	0.34	490	277.56	0.34	590	310.38	0.32
600	313.59	0.33	700	345.13	0.31	800	375.51	0.30			
610	316.80	0.32	710	348.22	0.31	810	378.48	0.30			
620	319.99	0.32	720	351.30	0.31	820	381.45	0.29			
630	323.18	0.31	730	354.37	0.30	830	384.40	0.29			
640	326.35	0.31	740	357.42	0.31	840	387.34	0.29			
650	329.51	0.31	750	360.47	0.30	850	390.26	0.29			
660	332.66	0.31	760	363.50	0.30						
670	335.79	0.32	770	366.52	0.30						
680	338.92	0.31	780	369.53	0.30						
690	342.03	0.31	790	372.52	0.30						

Closer inspection of this table shows that the temperature sensitivity is not very linear. The relation between resistance and temperature can be described by $R(T)= R_o[1+\alpha(T-T_o)+ \beta(T-T_o)^2]$, with $R_o= 100\ \Omega$, $\alpha= 3.9\times10^{-3}\ K^{-1}$, $\beta= -5.8\times10^{-7}\ K^{-2}$, with T_o the calibration temperature at $T_o= 0\ ^oC= 273.16$ K. The sensitivity is equal to $S= \alpha R_o$, within the resolution of the table, only over the very limited temperature range between -20 ^{o}C and 40 ^{o}C. When assuming a constant differential sensitivity, the error at 200 ^{o}C is (100+200×0.39)-175.84= 2.16 Ω (which is equivalent to 2.16/0.39= 5.5 ^{o}C).

Example 2.3
Calculate the non-linearity error of the Pt-100 resistor at 500 ^oC.

Solution:
When assuming high linearity the temperature dependence is described by: $R(T)= R_o[1+\alpha(T-T_o)]$, thus: $R(500\ ^oC)= 100[1+0.0039\times 500]= 295\ \Omega$. Table 2.1 lists $R(500\ ^oC)= 280.9\ \Omega$, yielding a non-linearity error of: 14.1 Ω. Therefore, the relative error is equal to: $\varepsilon= 14.1/280.9= 5\%$.

2.5.3 Non-linearity and gain error

A wide range of potentiometric position sensors are commercially available, i.e. linear displacement measuring devices with an operating range from less than a cm to in excess of 1 metre and angular types from less than 360 degrees to multi-turn potentiometers. The resistance between the two stator connections can be any value between less than 1 Ω and more than 10 MΩ. The main advantage of the potentiometric displacement sensor is the very simple read-out using the ratio of the ruler voltage and an excitation voltage as a measure for the displacement. One problem is the mechanical coupling with the moving object, which may lead to loading of the source and thus to measurement errors.

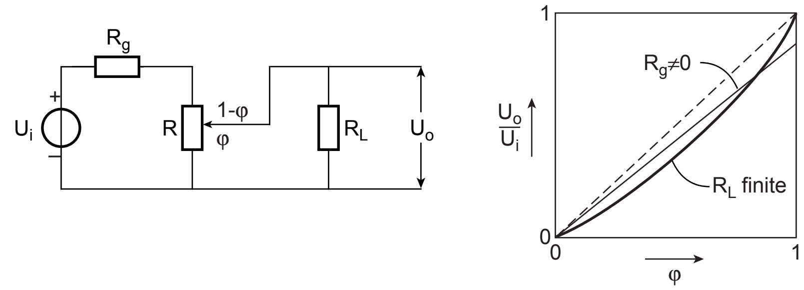


Figure 2.10, Errors due to source loading by the potentiometer and potentiometer loading due to R_L.

The potentiometric displacement sensor shown in Fig. 2.10 is also susceptible to measurement errors in the electrical domain. The first error is due to loading of a practical excitation source with a source resistance R_g. The second error is due to loading of the potentiometer by the read-out with input impedance R_L (Section 2.4.1) Considering R_g only yields:

$$\frac{U_o}{U_i} \simeq \frac{\varphi R}{R+R_g} = \varphi\left(1-\frac{R_g}{R+R_g}\right) \simeq \varphi\left(1-\frac{R_g}{R}\right) \qquad \text{(2-10)}$$

The result is a relative displacement-dependant (= scale) error equal to $\Delta\varphi/\varphi= R_g/R$. In deriving this expression it is assumed that $R_g \ll R$, while $R_L \rightarrow \infty$.

Similarly, loading by R_L, while $R \ll R_L$ and disregarding the effect of R_g (R_g= 0), results in a non-linearity in the transfer function expressed as:

$$\frac{U_o}{U_i} = \frac{\varphi R_L}{R_L + \varphi(1-\varphi)R} \simeq \varphi\left(1 - \varphi(1-\varphi)\frac{R}{R_L}\right) \tag{2-11}$$

This expression indicates a maximum non-linearity at φ= ½ equal to $\Delta\varphi/\varphi$= -$R/4R_L$.

2.5.4 Sensitivity and cross-sensitivities in strain gauges

Applying a force to a strip of resistive material results in the resistance value to change. Such a force sensitive resistor is generally referred to as a **strain gauge**.

A strain gauge (= piezo-resistor) uses the piezo-resistive effect. The resistance is strain-sensitive and is used for force measurement.

Strain gauges are suitable for the measurement of force, torque, bending, etc. The primary mechanical parameter is the force per unit cross-sectional area of a bar (= **stress,** σ). However, the measurand is the material deformation per unit length (= **strain,** ε), which requires prior knowledge of a well-defined material-dependent modulus of elasticity, E. At low-stress values the material is in the **elastic** regime. **Hooke's law** applies and the strain is linearly dependant on stress, E= σ/ε. Increasing stress beyond a certain threshold results in a regime where the material is **plastically deformed**. At even higher stress levels (the **yielding stress**), the material breaks. Operation is limited to the elastic regime with a wide safety margin from the threshold to plastic deformation, which implies the measurable strain in a construction remains below 1%.

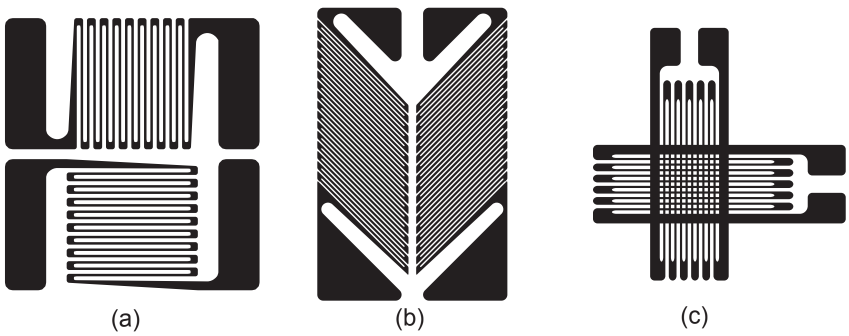


Figure 2.11, Different realisations of orthogonal pairs of strain gauges.

Wire- or film-based piezo-resistors are deposited on a thin and flexible substrate, which can be mounted on (glued to) a suitable part of a mechanical structure. A meander-shaped pattern, as shown in Fig. 2.11, is used to achieve an acceptable nominal resistance, despite the low value of the specific resistivity in the resistive material (e.g. metal).

Such a structure also enables selective strain measurement in only the direction of the meander. The top meander in Fig. 2.11a is sensitive to a vertical force component, while the bottom one is sensitive to the horizontal component. The corners are made wider than the resistive strips in the sensitive direction to minimise the sensitivity to strain in the non-intended direction. This property is specified as the **off-axis sensitivity**. The force vector can be measured using the two orthogonally placed meanders. Figure 2.11c shows a minimum-area solution with an electrically insulating foil in between the two strain gauges.

Applying a force results in a **tensile stress (i.e. resulting in an elongation)** or a **compressive stress (i.e. resulting in a compression)**.

It should be noted that the elongation or compression of the resistive layer is equal to that of the structure to which it is attached only if it can be assumed that:
(1) The adhesion between the strain gauge and structure is perfect and
(2) The structure on which the strain gauge is mounted has a much higher stiffness than that of the sensor.

A critical performance factor is the quality of the adhesion. Special glues are used to minimise the small relative displacement over time, which is referred to as **creep**. The thickness of the thin films used in strain gauges is usually small enough to make the sensor the more compliant part.

The strain gauge can be considered a strip of electrically resistive material of specific resistivity ρ, length L and cross-sectional area A, with a force applied at the ends. The associated change in resistance can be calculated via $R = \rho L/A$ as:

$$\frac{\partial R}{R} = \frac{\partial \rho}{\rho} + \frac{\partial L}{L} - \frac{\partial A}{A} = \frac{\partial \rho}{\rho} + 2\frac{\partial L}{L} \qquad (2\text{-}12)$$

The derivation of this expression is based on the assumption that the volume of any part in the strip of resistive film remains constant while strained (Poisson ratio ν=0). This assumption is within reasonable approximation valid for metal- or semiconductor-based strain gauges operating in the linear force-deflection range where Hooke's law applies. The assumption $\partial V/V = 0$ yields: $\partial L/L = -\partial A/A$.

In metals the specific resistivity, ρ, is not dependant on strain, and the relative change in resistance is expressed as:

$$\frac{\partial R}{R} = 2\frac{\partial L}{L} \qquad (2\text{-}13)$$

The **sensitivity (= gauge factor)** is equal to $k_\varepsilon = 2$.

In semiconductor materials also ρ changes significantly with strain. The resulting gauge factor is generally much larger as compared to metal-based strain gauges. However, the temperature sensitivity is also larger. Semiconductor (silicon)-based piezo resistors are formed in silicon using MEMS technologies and are used as part of an integrated sensor (see Chapter 8). Usually metal-film-based strain gauges are used to measure the mechanical loading of an existing structure.

Like most resistors, a strain gauge exhibits a sensitivity to temperature. Conventionally this sensitivity is expressed as a material-dependant **Temperature Coefficient of Resistivity (TCR)**, α, which is basically the differential temperature sensitivity normalised on the resistance: $\alpha = (1/R)\times(\partial R/\partial T)$ $[K]^{-1}$. Strain gauges are classified into two categories:

1. Metal film-based resistive and

2. Semiconductor-based.

The temperature dependence of the resistance is fundamentally material dependent. In a metal film, free-carries are abundant, whereas in a semiconductor, carrier mobility decreases with increasing temperature due to increasing crystal interaction.

Therefore, a thin-film metal strain gauge has a positive temperature coefficient of resistivity (TCR> 0). The gauge factor is relatively low, $k_\varepsilon = 2$. Constantan is a frequently used material in the fabrication of strain gauges with a temperature coefficient of resistivity $\alpha_{constantan} = 2\times10^{-5}$ K^{-1}.

The availability of free carriers is decisive in semiconductors. The number of charge carriers in the conduction band increases with temperature. Since these charge carriers participate in the conductivity, resistance decreases with increasing temperature; hence TCR< 0. In integrated silicon mechanical sensors, the piezo-resistivity in silicon is employed. A gauge factor in excess of $k_\varepsilon = 50$ is possible, depending on doping type and concentration. The absolute value of the TCR is significantly higher compared to that of a metal-film-based strain gauge.

The specification of temperature sensitivity in terms of a TCR makes the resistor value non-linearly dependent on temperature. The definition of the TCR results in:

$$\frac{\partial R(T)}{\partial T}=\alpha\times R(T) \tag{2-14}$$

This is a first-order differential equation and yields the solution:

$$R(T)=R(T_o)\exp\left(-\alpha\left(T-T_o\right)\right), \tag{2-15}$$

which can be linearised to: $R(T)= R_0[1+ \alpha(T-T_0)]$ for $\alpha(T-T_0)\ll 1$.

Since the measurable strain in a construction remains below 1%, while the range of strain values to be measured usually exceeds 10^3, strains of about 10^{-6} should be measurable (this is basically the detection limit). The unit micro-strain is often used (**1 μ*strain*= a deformation 10^{-6}**). At $k_\varepsilon=2$ the relative change in strain gauge resistance is in the same order of magnitude and thus also in the PPM (**Parts Per Million, 1 PPM= 10^{-6}**) range. In the case of an $R_0= 100\ \Omega$ strain gauge $\Delta R= 10^{-4}\ \Omega$ has to be reproducibly measured. Since this is an extremely small change in resistance, a special circuit is used, which is the **Wheatstone bridge** discussed in the next section.

2.5.5 Using the Wheatstone bridge for resistive sensor read-out

The **Wheatstone bridge** is basically a dual voltage divider, as shown in three different versions in Fig. 2.12. The transfer function from an excitation voltage at the input, U_{exc}, to the differential output voltage, U_{oW}, can be expressed as:

$$\frac{U_{oW}}{U_{exc}}=\frac{R_3}{R_2+R_3}-\frac{R_4}{R_1+R_4}=\frac{R_1R_3-R_2R_4}{\left(R_2+R_3\right)\left(R_1+R_4\right)} \tag{2-16}$$

If two resistors are available that have a value that increases with a measurand (e.g. a strain gauge that is elongated by a load), while simultaneously two other resistors are available with a value that decreases by the same amount (e.g. strain gauges compressed by the same load), then the most versatile version can be used: the full Wheatstone bridge (sometimes referred to as the differential Wheatstone bridge) that is shown in Fig. 2.12a. For $R_1= R_3= R_0+ \Delta R$ and $R_2= R_4= R_0- \Delta R$ Eqn. 2.16 is simplified to:

$$\frac{U_{oW}}{U_{exc}}=\frac{R_3}{R_2+R_3}-\frac{R_4}{R_1+R_4}=\frac{\left(R_o+\Delta R\right)-\left(R_o-\Delta R\right)}{\left(R_o+\Delta R\right)-\left(R_o-\Delta R\right)}=\frac{\Delta R}{R_o} \tag{2-17}$$

Note that the differential output voltage is read-out using a differential amplifier ($U_{id}= U_{oW}$). At this stage the read-out circuit is assumed to be sensitive only to the differential voltage, U_{id}, and not to the average value of the voltage levels

U_{oW+} and U_{oW-}, which is generally referred to as the **common-mode voltage**, U_{ic}. From the figure follows that $U_{ic}= (U_{oW+}+U_{oW-})/2= U_{exc}/2$ for $\Delta R \ll R_o$.

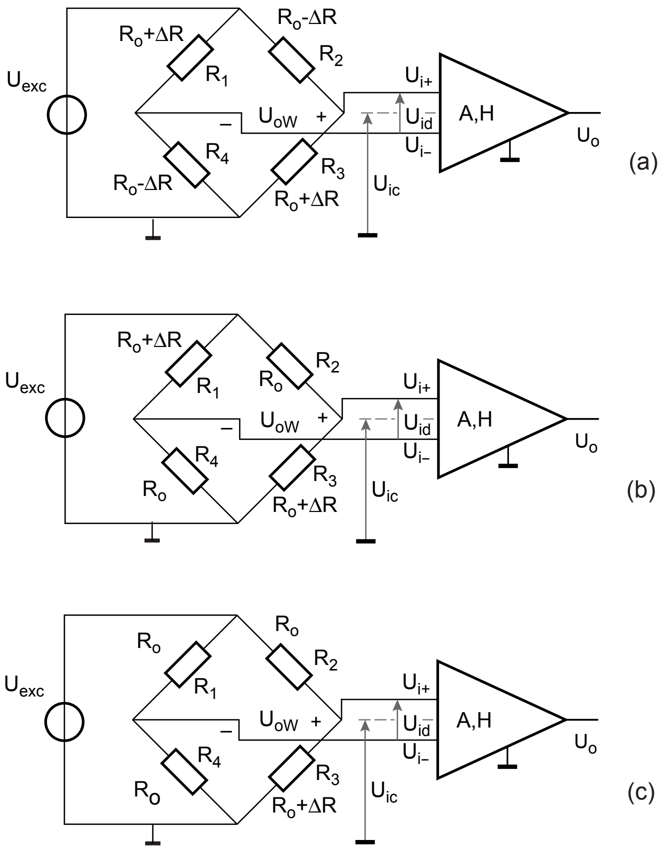


Figure 2.12, Wheatstone bridge types: (a) full bridge, (b) 1/2 bridge and (c) 1/4 bridge.

Practical differential amplifiers are not immune to the common-mode voltage. This effect is specified as the Common-mode rejection, *H*, which is discussed in detail in Chapter 4.

In many applications only resistors that either increase or decrease by the same amount are available. In addition, resistors are available that do not change in value. For example in mechanical constructions, strain gauge sensors can often only be attached in such a way that the resistance of all strain gauges increases (on the top surface of a downward bent beam) or decreases (on the bottom surface of a downward bent beam) with mechanical stress applied. A strain gauge on the sidewall would not change. In this case the half-bridge shown in Fig. 2.12b should be used.

The imbalance of the 1/2 Wheatstone bridge follows with R_1= R_3= R_o+ ΔR and R_2= R_4= R_o as:

$$\frac{U_{oW}}{U_{exc}} = \frac{(R_o + \Delta R)}{(R_o + \Delta R) + R_o} - \frac{R_o}{(R_o + \Delta R) + R_o} = \frac{\Delta R}{2R_o + \Delta R} \tag{2-18}$$

It can easily be verified that the bridge imbalance is non-linear with resistance change. However, for $\Delta R \ll R_o$ the transfer function can be linearised to: U_{oW}/U_{exc}= $\Delta R/2R_o$. Compared to the full Wheatstone bridge, a loss in gain by a factor of 2 remains.

In some applications, such as implantable sensors, space is very scarce and the sensor system only fits one resistor. In this case the quarter-bridge is used with one resistor in the sensor and three are placed externally, as shown in Fig. 2.12c.

$$\frac{U_{oW}}{U_{exc}} = \frac{(R_o + \Delta R)}{(R_o + \Delta R) + R_o} - \frac{R_o}{2R_o} = \frac{\Delta R}{2(2R_o + \Delta R)} \tag{2-19}$$

Also in this mode of operation the transfer function can be linearized. However U_{oW}/U_{exc}= $\Delta R/4R_o$ for $\Delta R \ll R_o$, so the loss in gain is by a factor of 4 as compared to the full Wheatstone bridge.

2.5.6 Applying strain gauges to the weighing bridge

When a bar with a strain gauge attached is pulled along the length of the sensor, the resistance of this strain gauge increases due to the elongation and reduced cross-sectional area of the resistive layer. It is important to note that the strain gauge is fundamentally sensitive to elongation, which is expressed in terms of the **strain**, ε= $\Delta L/L_o$, with ΔL as the elongation and L_o as the unstressed length.

The strain is the result of the mechanical stress (see also Section 2.5.4). Pulling a bar with force F results in **tensile stress**, σ= F/A_{struct}, with F as the force guided through the bar and A_{bar} as the cross-sectional area of the structure onto which the strain gauge is glued (conversely, pushing a bar results in **compressive stress**). Stress and strain are related by Hooke's law as: ε= σ/E, with E as the Young's modulus of the material used. The strain gauge resistance due to a force F applied results in:

$$R(\varepsilon) = R_o(1 + k_\varepsilon \varepsilon) = R_o\left(1 + k_\varepsilon \frac{\sigma}{E}\right) = R_o\left(1 + k_\varepsilon \frac{F}{A_{bar}E}\right) \tag{2-20}$$

with k_ε= $(\partial R/R)/(\partial L/L)$ as the gauge factor.

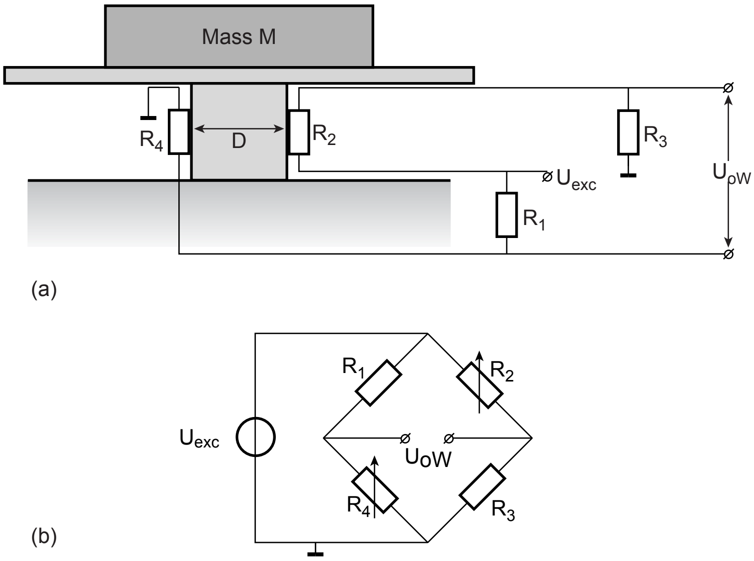


Figure 2.13, Read-out of a weighing bridge using strain gauges: (a) physical structure and (b) connection of strain gauges in 1/2 Wheatstone bridge.

Attaching two strain gauges to the column, while mounting two others on the column supporting a weighing platform and connecting these in a 1/2 Wheatstone bridge configuration, yields the weighing system shown schematically in Fig. 2.13. Strain gauges R_2 and R_4 are mounted on a column, which is at temperature T_{col} and subjected to stress σ, while R_1 and R_3 are external resistors at temperature T_{amb}. The resistor values can be expressed as:

$$\begin{aligned} R_1 = R_3 &= R_o\left[1+\alpha\left(T_{amb}-T_o\right)\right] \\ R_2 = R_4 &= R_o\left[1+k_\varepsilon \varepsilon\right]\times\left[1+\alpha\left(T_{col}-T_o\right)\right] \end{aligned} \tag{2-21}$$

with R_o as the resistance at calibration temperature T_o. Combining Eqns. 2.16 and 2.21 yields:

$$\begin{aligned} \frac{U_{oW}}{U_{exc}} &= \frac{1+\alpha\left(T_{amb}-T_o\right)-\left[1+k_\varepsilon\varepsilon\right]\times\left[1+\alpha\left(T_{con}-T_o\right)\right]}{1+\alpha\left(T_{amb}-T_o\right)+\left[1+k_\varepsilon\varepsilon\right]\times\left[1+\alpha\left(T_{con}-T_o\right)\right]} = \\ &\frac{\alpha\left(T_{amb}-T_{con}\right)-k_\varepsilon\varepsilon\times\left[1+\alpha\left(T_{con}-T_o\right)\right]}{2+\alpha\left(T_{amb}+T_{con}-2T_o\right)+k_\varepsilon\varepsilon\times\left[1+\alpha\left(T_{con}-T_o\right)\right]} \approx \\ &-\frac{\alpha}{2}\left(T_{con}-T_{amb}\right)-\frac{k_\varepsilon\varepsilon}{2} = -\left(\frac{\alpha}{2}\Delta T+\frac{k_\varepsilon}{2EA_{col}}F\right) \end{aligned} \tag{2-22}$$

The differential sensitivity to both the desired quantity (the force due to the mass on the platform) and the parasitic influence (the temperature difference between column and ambient) results in:

$$\frac{\partial\left(\frac{U_{oW}}{U_{exc}}\right)}{\partial F}=-\frac{k_{\varepsilon}}{2EA_{col}}\rightarrow\frac{\partial\left(\frac{U_{oW}}{U_{exc}}\right)}{\partial M}=-\frac{k_{\varepsilon}g}{2EA_{col}}\qquad\frac{\partial\left(\frac{U_{oW}}{U_{exc}}\right)}{\partial T}=-\frac{\alpha}{2}\tag{2-23}$$

This result indicates that the measurement system exhibits a linear sensitivity to the mass placed on the platform, as is desirable. However, it also reveals a significant sensitivity to the temperature difference between the bridge resistors, which is a cross-sensitivity (see Section 2.3.1). For this reason a special structure must be used to ensure that all bridge resistors are at an equal temperature. Moreover, the resistors should have the same TCR.

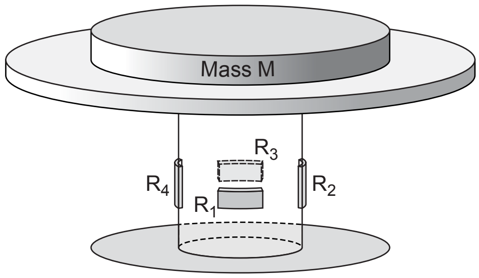


Figure 2.14, Weighing bridge with temperature compensation.

A suitable solution is to use four identical strain gauges all mounted on the column, as shown simplified in Fig. 2.14. A bridge imbalance with mechanical loading is made possible by placing two of the four strain gauges in the normal direction normal of the strain, so that these are not subjected to the mass-induced stress in the column. Therefore the read-out remains a 1/2 Wheatstone bridge. A metal column is a good thermal conductor, so the temperature gradient along the column can usually be kept very small, resulting in a very small temperature difference between the bridge resistors (< 0.1 K).

Example 2.4
The specifications of a weighing bridge based on strain gauges are as follows: gauge factor k_{ε}= 2, U_{exc}= 10 V, D_{col}= 10^{-2} m, E= 2×10^{11} N/m^2 and α= 2×10^{-5} K^{-1}.

The sensitivity results from Eqn. 2.23 and results in:
$\partial U_o/\partial M$ = $-2\times9.81\times10/[(\pi/2)\times10^{-4}\times2\times10^{11}]$ = -6.25 μV/kg.

The sensitivity to temperature is: $\partial U_o/\partial T$= -2×10^{-5}× 10/2= -100 μV/K.

A temperature difference of ΔT= 0.1 K, therefore, has the same effect as an equivalent mass of M= 1.6 kg. This mass sets the detection limit of the weighing systems, which makes it only suitable for the measurement of a relatively large mass.

So far only the effect of the TCR of the strain gauge used on the temperature equivalent detection limit has been discussed. It has been demonstrated that temperature effects can be reduced to negligible levels by placing the four strain gauges in a Wheatstone bridge in such a way that a net strain is measured, while the strain gauges are kept at the same temperature by using materials and a glue that ensures good thermal conductivity.

However, in a practical system other effects also have to be considered, which are:

- Temperature sensitivity of the strain, $\partial\varepsilon/\partial T$.
- Temperature sensitivity of the gauge factor, $\partial k_\varepsilon/\partial T$.

Assume a full Wheatstone bridge with strain gauges at exactly the same temperature. The bridge transfer function is: U_{oW}/U_{exc}= $\Delta R/R$= $k_\varepsilon\varepsilon$ (Eqn. 2.17). The temperature dependence of the bridge imbalance can be expressed as:

$$\partial\left(\frac{U_{oW}}{U_{exc}}\right)/\partial T=\frac{\partial(k_\varepsilon\varepsilon)}{\partial T}=k_\varepsilon\frac{\partial\varepsilon}{\partial T}+\varepsilon\frac{\partial k_\varepsilon}{\partial T} \qquad (2\text{-}24)$$

Note that $\partial\varepsilon/\partial T$ is the expansion coefficient of the material used for the mechanical construction, which is usually not the strain gauge material. Typical values for materials used are about 10^{-4} K^{-1}. Since k_ε= 2, the first term, $k_\sigma(\partial\varepsilon/\partial T)$, is in the order of magnitude of 10^{-4}. The maximum strain level in a typical measurement is also in the order of 10^{-4}. Since $\partial k_\varepsilon/\partial T \ll 1$, Eqn. 2.24 simplifies to:

$$\partial\left(\frac{U_{oW}}{U_{exc}}\right)/\partial T=k_\varepsilon\frac{\partial\varepsilon}{\partial T} \qquad (2\text{-}25)$$

Hence, the temperature cross-sensitivity of the strain measurement is proportional to the gauge factor and the thermal expansion coefficient of the material used. A construction material with a low thermal expansion coefficient is required, but there is little control over this parameter if the strain measurement is implemented in an existing construction.

Read-out of the Wheatstone bridge requires a differential amplifier. Practical differential amplifiers and their limitations are discussed in Chapter 4.

2.6 Capacitive and piezo-electric sensors and their read-out

2.6.1 Capacitive displacement sensors

Capacitive mechanical sensors are widely employed for the measurement of displacement. Especially the non-contact measurement option makes capacitive sensors very attractive. Capacitive displacement sensors are used for measuring either vertical or lateral displacements. In the case of a vertical displacement sensor, a linearity problem arises.

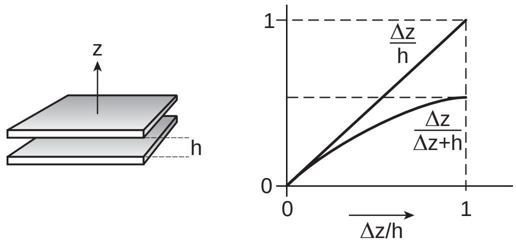


Figure 2.15, Non-linearity and non-conformity of a vertical capacitive displacement sensor.

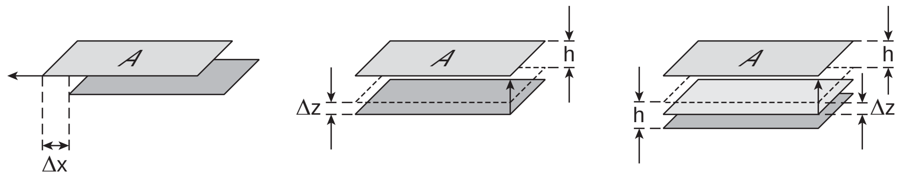
The specification of non-linearity is useful only if the transfer function of the system under consideration can be satisfactorily described with a straight line. In the basic capacitive vertical displacement sensor shown in Fig. 2.15 this is not the case. The area overlap, A, between the electrodes remains constant and the electrode spacing increases with positive vertical displacement, z. When assuming a valid **parallel-plate approximation** (stray fields can be disregarded: $z^2 \ll A$) and a nominal capacitance $C_0=C(h)$ at spacing h, the relative change in capacitance can be described by:

$$\Delta C = C(h+\Delta z) - C(h) = \frac{\varepsilon A}{h+\Delta z} - \frac{\varepsilon A}{h} = \frac{\varepsilon A}{h}\left(\frac{h}{h+\Delta z} - 1\right) = C_o\left(\frac{-\Delta z}{h+\Delta z}\right) \rightarrow$$
$$\frac{\Delta C}{C_o} = -\frac{\Delta z}{h+\Delta z} \tag{2-26}$$

This function is intrinsically non-linear. However, for $\Delta z \ll h$ it can be approximated by: $\Delta C/C_0= -\Delta z/h$. For $\Delta z/h= 0.25$ the error is equal to: 0.25-(0.25/1.25)= 0.05. This deviation from the straight line is predictable and, thus, does not specify a genuine uncertainty.

A more realistic approach is not to refer to a straight line, but rather to a use a specification in terms of conformity to a non-linear function, in which case:

$\Delta C/C_0 = -\Delta z/(h+\Delta z)$, as shown in Fig. 2.15. At an increasing Δz, the effect of fringe fields can no longer be ignored and deviation from the non-linear function results. Although stray field effects can also be predicted by calculations based on device geometry, these are usually included in the specifications as **non-conformity**.



(a) (b) (c)

Figure 2.16, Capacitive sensor for (a) lateral displacement, (b) vertical displacement and (c) linearised vertical displacement.

In a lateral capacitive displacement sensor the movable electrode remains at a constant vertical spacing with respect to the electrode at a fixed position (the stator), but lateral movement causes a change in area overlap, as shown in Fig. 2.16a.

The intrinsic non-linear sensitivity for vertical displacement (Fig. 2.16b) can be circumvented using the differential structure shown in Fig. 2.16c. The differential capacitance, ΔC, is defined as the capacitance between, on the one hand, the middle (moving) electrode and the upper stator and, on the other hand, that of the same moving middle electrode and the lower stator, and can be expressed as:

$$\Delta C = \varepsilon_o \frac{A}{h-\Delta z} - \varepsilon_o \frac{A}{h+\Delta z} = \frac{2\varepsilon_o A}{h^2 - \Delta z^2}\Delta z \approx \frac{2\varepsilon A}{h^2}\Delta z \rightarrow \frac{\Delta C}{C_o} \approx 2\frac{\Delta z}{h} \tag{2-27}$$

A linear response is achieved with a gain in sensitivity by a factor of two at the expense of a more complex sensor structure. As is discussed in the next section, a differential capacitive sensor structure also offers advantages for read-out. For this reason, also the practical lateral displacement sensor is based on the differential structure shown in Fig. 2.17a.

When assuming that the parallel-plate approximation is acceptable, then the differential capacitance, ΔC, due to the difference in overlap between the right-hand stator and the moving electrode, C_1, and the left-hand electrode and the moving electrode, C_2, can be described by:

$$\Delta C = C_1 - C_2 = \varepsilon_o d\frac{L+\Delta x}{h} - \varepsilon_o d\frac{L-\Delta x}{h} = \frac{2\varepsilon_o dL}{h}\frac{\Delta x}{L} \rightarrow \frac{\Delta C}{C_o} \approx 2\frac{\Delta x}{L} \tag{2-28}$$

This device is very suitable for non-contact displacement sensing, as is shown in Fig. 2.17b. An external top stator is added to the bottom stator. Moreover, a third set of electrodes is used to couple the voltage from the moving electrode capacitively back to the stator part of the structure via C_3= C_{3a}+ C_{3b}. The overlap between this electrode and the moving electrode is not dependant on lateral displacement, x, and C_3 remains constant.

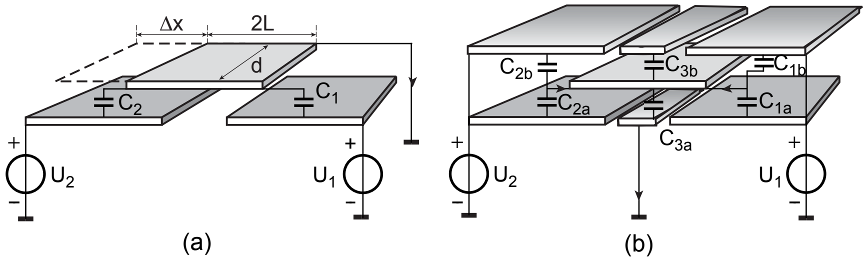


Figure 2.17, Differential capacitive sensor for lateral displacement: (a) basic structure and (b) structure without electrical contact with the moving part, but with compensated vertical displacement (stator electrode pairs are interconnected).

Since C_3 is relatively small, guarding electrodes are required on either side to avoid direct coupling of the excitation voltages. As an additional feature the sensitivity of the lateral displacement sensor to parasitic vertical displacements is strongly reduced by the coupling of the excitation voltages via both the top and bottom electrodes to result in: C_1= C_{1a}+ C_{1b} and C_2= C_{2a}+ C_{2b}. C_{1b} and C_{2b} increase in the case of an upward displacement, Δz, but C_{1a} and C_{2a} decrease by the same amount for $\Delta z \ll h$.

2.6.2 Read-out of the capacitive sensor

The non-inverting amplifier can in principle be used for the read-out of a capacitive sensor. A capacitive sensor is preferably AC operated at a high frequency to limit the impedance level. Selecting a frequency much higher than the frequency of the mains voltage minimises the effect of EMI. The circuit is shown in Fig. 2.18 with an output voltage U_o= $I_{exc}(R_1+R_2)/(j\omega R_1C_s)$. A transfer function proportional to $1/C_s$ results, which is convenient in the case the information is in the spacing between the plates of the capacitor.

The main problem with the voltage read-out of a capacitive sensor is the sensitivity to the parasitic capacitance, C_{par}, in parallel to C_s. The nominal value of a capacitive sensor is typically in the range 1-5 pF, which is in the same order of magnitude as the parasitic component. Moreover, the detection limit is typically < 1fF, which is less than the instability of C_{par}. The high value of the input

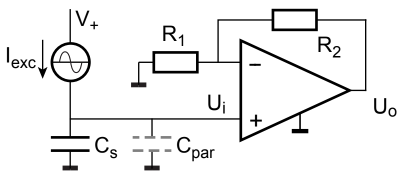


Figure 2.18, Non-inverting amplifier for read-out of a capacitive sensor.

impedance of the non-inverting amplifier is not helpful in reducing the sensitivity to this effect.

A more suitable and highly flexible approach for the read-out of a capacitive sensor is based on the charge amplifier shown in Fig. 2.19. The charge amplifier is basically used to read-out a non-ideal current source (the sensor capacitance with an AC excitation voltage that is connected in series) which is connected to a capacitive trans-impedance amplifier (current-to-voltage converter). The open-loop gain is assumed to be infinitely large. Therefore, the voltage at the inverting input is driven to be equal to that at the non-inverting input (inverting input is at **virtual ground** potential). As a consequence the input impedance is extremely small (more details on the input impedance of opamp-based circuits is presented in Chapter 6), which is a highly desirable property for the read-out of the non-ideal current source. Any parasitic capacitance is in parallel to the inverting input, and ground is short-circuited by the virtual ground.

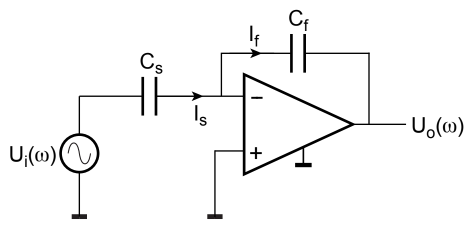


Figure 2.19, Charge amplifier.

The transfer function of the charge amplifier is described by:

$$
\begin{aligned}
&U_- = U_+ = 0 \\
&I_s(\omega) = U_i(\omega) j\omega C_\mathrm{s} \\
&I_\mathrm{f}(\omega) = -U_o(\omega) j\omega C_\mathrm{f} \qquad I_\mathrm{s}(\omega) = I_\mathrm{f}(\omega) \;\rightarrow\; \frac{U_o(\omega)}{U_i(\omega)} = -\frac{C_\mathrm{s}}{C_\mathrm{f}}
\end{aligned}
\tag{2-29}
$$

A characteristic of the charge amplifier is that it is not operated as an amplifier, but rather as an attenuator ($C_s/C_f \ll 1$) with the output signal $U_o = -C_s \times U_i / C_f$ as a measure for the capacitance, C_s. This circuit is not used to measure the amplitude of an unknown small signal, but rather to measure the value of a (sensor) capacitance, C_s, using a well-defined excitation signal for U_i.

Rearranging the expression for the transfer function also explains the reason why this circuit is referred to as a charge amplifier. The virtual ground enables the charge stored in C_s, $Q = U_i C_s$, to be transferred to C_f, $Q = U_o C_f$. Since the charge is balanced, unequal capacitors result in voltage gain or attenuation. However, it should be noted that the charge is transferred and not amplified.

A large resistor R_f is connected in parallel to C_f in order to limit the output DC level due to offset (see also Chapter 3). The huge advantage of the charge amplifier is the ability to operate the excitation source U_i at a frequency close to the unity-gain frequency, $\omega_T = A_o/\tau_v$, without cutting the open-loop gain. The advantages of operating the charge amplifier at a high frequency are:

- A low transducer impedance and
- A maximum spectral distance to the main source of interference (capacitive coupling of the mains voltage).

The limitations of the charge amplifier are discussed in more detail in Chapter 5.

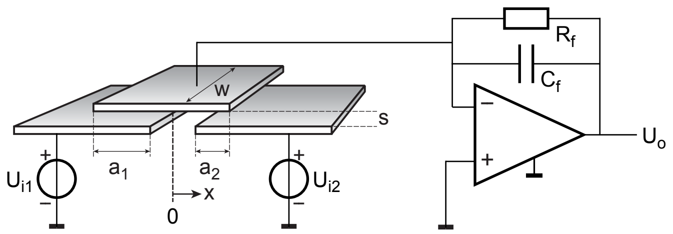


Figure 2.20, Differential capacitive displacement sensor with a charge amplifier for read-out (shown for a negative displacement $x = a_2 - a_1$).

A capacitive sensor usually exhibits a nominal capacitance at a zero non-electrical input, C_{so}. Applying a non-electrical input (e.g. a lateral displacement in a parallel-plate displacement sensor) results in a change in capacitance, ΔC_s. This change in capacitance, rather than the nominal value, is of importance. The nominal sensor capacitance can be compensated for using a differential sensor structure and two AC excitation voltages of opposite polarity (180 degrees out-of-phase), as shown in Fig. 2.20.

The component value of R_f is sufficiently high to ignore its effect at the operating frequency selected (see Section 5.6). The excitation voltage U_{i1} is connected to the left-hand stator of the sensor with capacitance to the moving part equal to: $C_s = C_{so} - \Delta C_s$. Similarly, U_{i2} to the right-hand stator electrode gives: $C_c = C_{so} + \Delta C_s$.

Since the inverting input is at (virtual) ground potential, currents are added at this node. The current through C_s is added to the current through C_c and for $U_{i1} = -U_{i2} = U_i$ the voltage at the output of the charge amplifier results in:

$$U_o = -\left(\frac{C_s}{C_f}U_{i1} + \frac{C_c}{C_f}U_{i2}\right) = -\left(\frac{C_{so} - \Delta C_s}{C_f}U_i + \frac{C_{so} + \Delta C_s}{C_f}(-)U_i\right) = 2\frac{\Delta C_s}{C_f}U_i \qquad (2\text{-}30)$$

Example 2.5
The left-hand electrode of the capacitive displacement sensor is connected to $U_{i1} = U_{exc}$ and the capacitance to the moving electrode is $C_{s1} = C_{so} - \Delta C_s$. The right-hand electrode is connected to $U_{i2} = -U_{exc}$ and $C_{s2} = C_{so} + \Delta C_s$. The area overlap at the left-hand electrode pair is $w \times a_1 = w \times (a - x)$ and at the right-hand electrode pair it is $w \times a_2 = w \times (a + x)$. The sensor dimensions are: $w = 50$ mm, $a = 50$ mm and $s = 2$ mm, while $\varepsilon_o = 8.85 \times 10^{-12}$ F/m. The dielectric constant in air is, $\varepsilon_r = 1$, thus $C_{so} = 11$pF. The transfer function from Eqn. 2.32 results in: $U_o/x = 2 \times [\varepsilon_o \varepsilon_r (w/s) U_{exc}]/C_f$.

Selecting $U_{exc} = 10$ V and $C_f = 40$ pF yields: $U_o/x = 110$ V/m. The range of displacement is limited to $x_{max} = a = 50$ mm, thus U_o is in between -5.5 V and 5.5 V.

2.6.3 Piezo-electric force sensors

Piezo-electric sensors resemble capacitive sensors. The difference is the solid-state dielectric layer in between the electrodes which is composed of molecules that are dipoles and can be polarised. Polarisation takes place when:

- the dielectric layer is subjected to a mechanical force or
- a voltage is applied across the electrodes.

The force sensitivity makes the structure suitable for force sensing, and the voltage sensitivity enables the fabrication of an actuator. The structure of a piezo-electric force sensor is shown simplified in Fig. 2.21. The internal structure of the dielectric layer is usually composed of long polar molecules without a preferential orientation. A mechanical load forces the long molecules into a preferred direction.

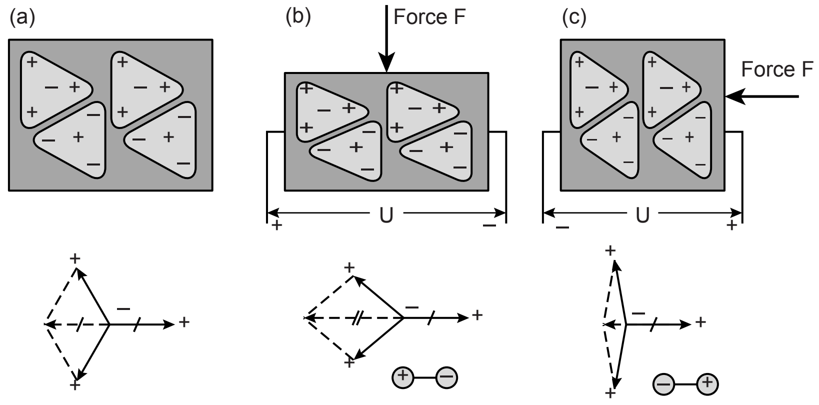


Figure 2.21, Principle of piezo-electricity: (a) non-loaded polar molecule, (b) shear piezo-electricity and (c) regular piezo-electricity.

The force applied results in a strain, which is revealed as a polarisation charge. Polarisation can be in the direction perpendicular to the force (Fig. 2.21b) or in the direction of the force (Fig. 2.21c). The polarisation charge can be measured as the voltage across the electrodes.

Polarisation in the direction of the force is due to regular piezo-electricity, whereas polarisation in the direction perpendicular to the force vector is due to shear piezo-electricity.

In a practical material the piezo-electric effect takes place in all three dimensions and must be described by a matrix. The advantage is that the same material can be used in different operating modes. The disadvantage is the parasitic sensitivity to a non-selected direction in the case of misalignment. Assume the piezo-electric crystal is positioned to measure the force through the structure in the direction normal to the polarisation. Any misalignment between the force vector and that direction results in a parasitic shear sensitivity. This effect should be taken into consideration in the analysis of the detection limit.

Applying a voltage has a similar effect as applying a force. The internal electric field in the dielectric layer causes an ordering of the molecules and thus results in a macroscopic deformation or a force acting on an external structure. This makes the effect suitable for the fabrication of actuators. Applying a voltage across the electrodes in the actuator mode leads in principle to the same deflection of the material as the strain in the case of an external mechanical force applied in the sensor mode. Also the resulting output voltage across the terminals in the sensor mode is in theory the same. For this reason piezo-electricity is referred to as a **reversible effect**.

Two approaches are available to measure the polarisation charge using the electrodes:

- Charge read-out by short-circuiting the electrodes and time-integration of the electrical current ($Q = I \times t$).
- Voltage read-out of the open-circuit voltage, which is a measure for the charge in the case of a constant capacitance ($Q = C \times U$).

Voltage read-out makes the system sensitive to parasitic capacitance, similar to the read-out of a capacitive sensor. The charge amplifier is, therefore, more suitable and can be used for charge read-out. The charge sensitivity of the sensor is defined as: $S_q = \partial Q/\partial F$ [C/N]. In the second approach a Volt meter is in principle connected to the terminals of the capacitor for read-out of the voltage. In this mode of operation the voltage sensitivity is the most practical performance parameter and is defined as: $S = \partial U/\partial F = \partial(Q/C)/\partial F = (1/C) \times (\partial Q/\partial F) = S_q/C$ [V/N].

The sensitivity of piezo-electric materials is in the range of 2-100 pC/N with a nominal sensor capacitance in the 10-50 pF range. Quartz is a natural piezo-electric material with a relatively low sensitivity. Synthetic materials in the form of foils are also available. Although the sensitivity is higher, these materials are more susceptible to a decay over time (ageing).

It should be noted that, like any capacitor, the dielectric layer is an imperfect insulator and thus should be modelled as a capacitor with a loss resistor in parallel, which causes a leakage current to erase any generated polarisation charge. Consequently, the piezo-electric sensor is not suitable for measuring static forces.

Example 2.6
1. A piezo-electric crystal with a charge sensitivity $S_q = 5$ pC/N in the normal direction is applied to measure acceleration. For this purpose a seismic mass $M = 0.1$ kg is attached to the force sensitive surface as shown in Fig. 2.22. Calculate the output voltage for a sensor capacitance $C_s = 20$ pF in the case of an acceleration $a = 20$ m/s^2.

Solution:
$F = M \times a = 0.1 \times 20 = 2$ N.
$U_{oa} = S_u F = (S_q/C_s) \times F = (5 \times 10^{-12}/2 \times 10^{-11})2 = 0.5$ V.

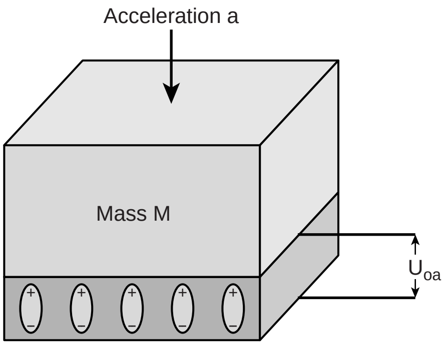


Figure 2.22, Acceleration sensor based on a piezo-electric crystal and a seismic mass.

2. Dielectric loss is accounted for by using a resistor, R_{loss}= 2 GΩ in parallel to C_s. If F can be regarded as a step function, what is the maximum measurement time, t_{meas}, for an error ε= 0.2%?

Solution:
$U_{oa}(t_{meas})= S_u F(1\text{-exp}(-t_{meas}/R_{loss}C_s)= 0.998\, S_u F$.
Hence, $t_{meas}= -R_{loss}C_s \ln(0.002)$= 248.6 ms.
Basically the sensor cannot be used for force measurement at frequencies below 4 rad/s= 0.64 Hz, which is not acceptable in many applications.

2.7 Exercises

2.1 A potentiometer is used for resistive angular displacement sensing with read-out as shown in Fig. 2.8. Calculate the maximum error δ in the measurement of φ for R_g= 10 Ω, R= 1 kΩ and $R_L \to \infty$ (non-ideal source). At which value of φ is the error maximum?

Solution:
For R_g= 10 Ω, R= 1 kΩ and $R_L \to \infty$, the transfer function is expressed as:

$$\frac{U_o}{U_i} = \frac{\varphi R}{R + R_g} \simeq \varphi\left(1 - \frac{R_g}{R}\right) = \varphi(1-\varepsilon) = \varphi\left(1 - \frac{10}{1000}\right) \to \varepsilon = 1\% \qquad (2\text{-}31)$$

The relative error is independent of φ.

2.2 Similar to question 2.1, calculate the maximum error δ in the measurement of φ for R_g= 0 Ω, R = 1 kΩ and R_L = 100 kΩ (output loading). At which value of φ is the error maximum?

Solution:
Output loading by R_L = 100 kΩ, while disregarding the effect of R_g,=0 yields:

$$\frac{U_o}{U_i} = \varphi\left(1-\varphi(1-\varphi)\frac{R}{R_L}\right) = \varphi\left(1-\varphi(1-\varphi)\frac{1k}{100k}\right) \tag{2-32}$$

The non-linearity is maximum for φ= ½ and is equal to $-R/4R_L$= -0.25%.

2.3 A piezo-electric acceleration sensor is composed of a piezo-electric element with a charge sensitivity S_q= 5 pC/N and a sensor capacitance C_s= 1 nF, with a 100 gram seismic mass attached (Fig. 2.22). Calculate the sensor output voltage if subjected to a (non-static) acceleration of 1 m/s^2.

Solution:
$F= M\times a$= 0.1× 1= 0.1 N.
$U_o= S_uF= (S_q/C_s)\times F= (5\times10^{-12}/10^{-9})0.1$= 500 μV.

2.4 Read-out of the piezo-electric accelerometer described in the previous question 2.3 involves a cable between sensor and a 1× voltage amplifier. The cable has a capacitance C_k= 40 pF between the inner conductor and shield. The amplifier has an input capacitance C_v= 80 pF. Calculate the error for voltage read-out if these capacitances had not been taken into consideration in the design of this data-acquisition system.

Solution:
$C_k + C_v$= 120 pF are in parallel to the sensor capacitance. Hence:
$U_o= S_uF= [S_q/(C_s+C_k+ C_v)]\times F= (5\times10^{-12}/1120\times10^{-12})0.1$= 446 μV.

Therefore, the absolute error amounts to 54 μV (relative error is 10.8%).
Note that the same answer results when taking the Thevenin equivalent of the source, while considering the capacitive voltage division.

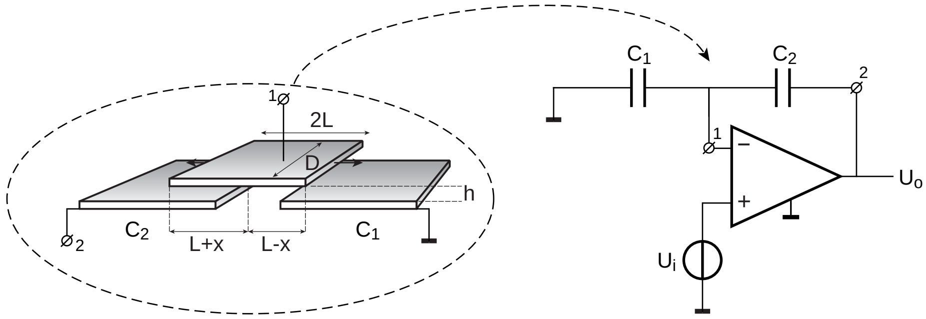


Figure 2.23, Differential capacitive sensor.

The capacitive differential sensor shown in Fig. 2.23 is read-out using the opamp circuit shown. The parallel plate approximation can be used for the sensor capacitors: $C_1= \varepsilon D(L\text{-}x)/h$ and $C_2= \varepsilon D(L+x)/h$, where ε denotes the dielectric constant, and D, L and h are the dimensions shown in the figure. The lateral displacement is x.

2.5 Derive the expression for the transfer function U_o/U_i.

Solution:
Non-inverting amplifier with capacitive impedances:

$$\frac{U_o}{U_i}=\frac{Z_1+Z_2}{Z_1}=\frac{C_2+C_1}{C_2}=\frac{(L+x)+(L-x)}{(L+x)}=\frac{2L}{L+x} \qquad (2\text{-}33)$$