CHAPTER 5

Mechanical Transduction Techniques

There are many examples of micromachined mechanical transducers and these will be reviewed in detail in the following chapters. The purpose of this chapter is to present some of the fundamental concepts and techniques that are used in the design of mechanical microsensors and actuators. The most sensing-important mechanisms include the following effects: piezoresistivity, piezoelectricity, variable capacitance, optical, and resonant techniques. We will also review the main actuation methods, including: electrostatic, piezoelectric, thermal, and magnetic. The final section of this chapter includes a review of so-called intelligent (or smart) sensors.

5.1 Piezoresistivity

Piezoresistivity derives its name from the Greek word *piezin*, meaning “to press.” It is an effect exhibited by various materials that exhibit a change in resistivity due to an applied pressure. The effect was first discovered by Lord Kelvin in 1856, who noted that the resistance of copper and iron wires increased when in tension. He also observed that iron wires showed a larger change in resistance than those made of copper. The first application of the piezoresistive effect did not appear until the 1930s, some 75 years after Lord Kelvin’s discovery. Rather than using metal wires, these so-called strain gauges are generally made from a thin metal foil mounted on a backing film, which can be glued onto a surface. A typical metal foil strain gauge is depicted in Figure 5.1.

![Metal foil sensing element](image)

Figure 5.1 Illustration of a metal foil strain gauge.
The sensitivity of a strain gauge is generally termed the gauge factor. This is a dimensionless quantity and is given by

\[ GF = \frac{\text{relative change in resistance}}{\text{applied strain}} = \frac{\Delta R / R}{\Delta L / L} = \frac{\Delta R / R}{\varepsilon} \]  

(5.1)

where \( R \) is the initial resistance of the strain gauge and \( \Delta R \) is the change in resistance. The term \( \Delta L / L \) is, by definition, the applied strain and is denoted as \( \varepsilon \) (dimensionless). For all elastic materials, there is a relationship between the stress \( \sigma (N/m^2) \) and the strain \( \varepsilon \); that is, they obey Hooke’s law and thus deform linearly with applied force. The constant of proportionality is the elastic modulus or Young’s modulus of the material and is given by

\[ \text{Young’s modulus, } E = \frac{\text{Stress}}{\text{Strain}} = \frac{\sigma}{\varepsilon} (N/m^2) \]  

(5.2)

The Young’s modulus of silicon is 190 GPa (1 Pa = 1 N/m\(^2\)), which is close to that of typical stainless steel (around 200 GPa). For a given material, the higher the value of Young’s modulus, the less it deforms for a given applied stress (i.e., it is stiffer).

When an elastic material is subjected to a force along its axis, it will also deform along the orthogonal axes. For example, if a rectangular block of material is stretched along its length, its width and thickness will decrease. In other words, a tensile strain along the length will result in compressive strains in the orthogonal directions. Typically, the axial and transverse strains will differ and the ratio between the two is known as Poisson’s ratio, \( \nu \). Most elastic materials have a Poisson’s ratio of around 0.3 (silicon is 0.22). The effect on a rectangular block is depicted in Figure 5.2. The strains along the length, width, and thickness are denoted by \( \varepsilon_l, \varepsilon_w, \) and \( \varepsilon_t \), respectively.

**Figure 5.2** Illustration of Poisson’s ratio on a rectangular, isotropic, elastic block. A longitudinal tensile strain results in deformation in the two orthogonal axes.
If it is assumed that the block is made of a resistive material, then its resistance, \( R \), is given by

\[
R = \frac{\rho l}{A}
\]  

(5.3)

where \( \rho \) is the bulk resistivity of the material (\( \Omega \text{cm} \)), \( l \) is the length, and \( A \) is the cross-sectional area (i.e., the product of width \( w \) and thickness \( t \)).

Hence,

\[
R = \frac{\rho l}{wt}
\]  

(5.4)

Differentiating the equation for resistance gives

\[
\frac{dR}{R} = \frac{d\rho}{\rho} + \frac{dl}{l} - \frac{dw}{w^2} - \frac{dt}{wt} \]

(5.5)

and hence

\[
\frac{dR}{R} = \frac{d\rho}{\rho} + \frac{dl}{l} - \frac{dw}{w} - \frac{dt}{t}
\]  

(5.6)

By definition, \( \varepsilon_l = \frac{dl}{l} \), so the following equations apply on the assumption that we are dealing with small changes, and hence \( dl = \Delta l \), \( dw = \Delta w \), and \( dt = \Delta t \):

\[
\frac{dw}{w} = \varepsilon_w = -\nu \varepsilon_i \quad \text{and} \quad \frac{dt}{t} = \varepsilon_t = -\nu \varepsilon_i
\]  

(5.7)

where \( \nu \) is Poisson’s ratio. Note the minus signs, indicating that the width and thickness both experience compression and hence shrink. It is worth noting that the above example illustrates a positive Poisson’s ratio.¹

Therefore, from (5.6) and (5.7) we have

\[
\frac{dR}{R} = \frac{d\rho}{\rho} + \varepsilon_i + \nu \varepsilon_i + \nu \varepsilon_i
\]  

(5.8)

From (5.1) the gauge factor is therefore

\[
GF = \frac{dR/R}{\varepsilon_i} = \frac{d\rho/\rho}{\varepsilon_i} + (1 + 2\nu)
\]  

(5.9)

Equation (5.9) indicates clearly that there are two distinct effects that contribute to the gauge factor. The first term is the piezoresistive effect \( (d\rho/\rho)/\varepsilon_i \) and the second is the geometric effect \( (1 + 2\nu) \). As Poisson’s ratio is usually between 0.2 and 0.3,

¹ Materials having a negative Poisson’s ratio do exist. That is to say, as you stretch them, the width and thickness actually increase. Examples of such materials include special foams and polymers such as polytetrafluoroethylene (PTFE).
the contribution to the gauge factor from the geometric effect is therefore between 1.4 and 1.6. Sensors that exhibit a change in resistance as a result of an applied strain are generally termed strain gauges. Those in which the piezoresistive effect dominates are often referred to as piezoresistors. As Table 5.1 shows, different materials can have widely differing gauge factors.

So for a metal foil strain gauge or thin metal film, the geometric effect dominates the piezoresistive effect; whereas for a semiconductor the converse is true.

Semiconductor strain gauges possess a very high gauge factor. P-type silicon has a gauge factor up to +200, and n-type silicon has a negative gauge factor down to −125. A negative polarity of gauge factor indicates that the resistance decreases with increasing applied strain. In addition to exhibiting high strain sensitivity, semiconductor strain gauges are also very sensitive to temperature. Compensation methods must therefore be adopted when using semiconductor strain gauges.

A detailed account of the piezoresistive effect in silicon can be found in Middelhoek and Audet [1]; only a brief account will be given in this text. Essentially, the effective mobilities of majority charge carriers are affected by the applied stress. With p-type materials, the mobility of holes decreases and so the resistivity increases. For n-type materials, the effective mobility of the electrons increases and hence the resistivity decreases with applied stress. The effect is highly dependent on the orientation. If the geometric effect in semiconductor strain gauges is neglected, then the fractional change in resistivity is given by

\[ \frac{d\rho}{\rho} = \pi_l \sigma_l + \pi_t \sigma_t \]  

(5.10)

where \( \pi_l \) and \( \pi_t \) are the longitudinal and transverse piezoresistive coefficients and \( \sigma_l \) and \( \sigma_t \) are the corresponding stresses. The longitudinal direction is defined as that parallel to the current flow in the piezoresistor, while the transverse is orthogonal to it. The two coefficients are dependent on the crystal orientation and doping (p-type or n-type) and concentration. The temperature coefficient of piezoresistivity is around 0.25 %/°C in both directions.

Polysilicon and amorphous silicon are also piezoresistive, but because they comprise crystallites, the net result is the average over all orientations. The temperature coefficient of resistance (TCR), however, is significantly lower than that of single crystal silicon and is generally less than 0.05 %/°C. By carefully choosing the doping levels, it is possible to reduce the TCR further.

Thin metal films behave in a similar manner to metal foil strain gauges and hence it not surprising that their gauge factors are very similar. Such films can be deposited directly onto the desired substrate (steel, ceramic, silicon) and therefore become an integral part of the system, thus removing the need for adhesives as with

<table>
<thead>
<tr>
<th>Material</th>
<th>Gauge Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metal foil strain gauge</td>
<td>2–5</td>
</tr>
<tr>
<td>Thin-film metal</td>
<td>2</td>
</tr>
<tr>
<td>Single crystal silicon</td>
<td>−125 to +200</td>
</tr>
<tr>
<td>Polysilicon</td>
<td>±30</td>
</tr>
<tr>
<td>Thick-film resistors</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 5.1 Gauge Factors of Different Materials
metal foil strain gauges. The adhesives can contribute to a phenomenon called creep, whereby the gauge can effectively slip and therefore produce false readings as the adhesive softens with increasing temperature or over long periods of time.

Thick-film resistors, often used in hybrid circuits, have also been shown to be piezoresistive. Their gauge factor is around 10, and therefore, they offer a sensitivity between that of a semiconductor and foil strain gauge. The TCR is around 100 parts per million (ppm) per degree Celsius and matching between adjacent resistors is often less than 10 ppm/°C, making them well suited for use as active elements in Wheatstone bridge circuits, which reduce the overall temperature sensitivity.

An associated effect that has been observed in semiconductors is the so-called piezojunction effect, whereby a shift in the I-V characteristic of a $p$-$n$ junction is observed as a result of an applied stress. Although this is an interesting physical effect, it has found little use in commercial micromachined devices.

5.2 Piezoelectricity

Certain classes of crystal exhibit the property of producing an electric charge when subjected to an applied mechanical force (direct effect). They also deform in response to an externally applied electric field (inverse effect). This is an unusual effect as the material can act as both a sensor and actuator. It was first discovered in quartz by Jacques and Pierre Curie in 1880. The physical origin of piezoelectricity arises because of charge asymmetry within the crystal structure. Such crystals are often termed noncentrosymmetric, and because of the lack of symmetry, they have anisotropic characteristics. Owing to its symmetric, cubic crystal structure, silicon is not, therefore, piezoelectric. Some crystals such as quartz and Rochelle salt are naturally occurring piezoelectrics, while others like the ceramic materials barium titanate, lead zirconate titanate (PZT), and the polymer material polyvinylidene fluoride (PVDF) are ferroelectric. Ferroelectric materials are those that exhibit spontaneous polarization upon the application of an applied electric field. In other words, ferroelectrics must be poled (polarized) in order to make them exhibit piezoelectric behavior. They are analogous to ferromagnetic materials in many respects. Figure 5.3 shows how an applied force gives rise to an electric charge (and hence voltage) across the faces of a slab of piezoelectric material.

![Figure 5.3](image.png)

Figure 5.3 An illustration of the piezoelectric effect. The applied force results in the generation of a voltage across the electrodes.
If a ferroelectric material is exposed to a temperature exceeding the Curie point, it will lose its piezoelectric properties. Hence, there is a limit beyond which they cannot be used as sensors (or actuators). The Curie point of PZT type 5H is around 195°C, and its maximum operating temperature is generally lower than this value. In addition to this, the piezoelectric coefficients of the material also vary with temperature, and this is referred to as the pyroelectric effect. This can be exploited in its own right, and pyroelectric sensors based on modified PZT are often used as the basis of infrared sensor arrays.

Owing to the anisotropic nature of piezoelectric materials, a system of identifying each axis is required in order to specify its parameters. By convention, the direction of polarization is taken as the 3-axis, with the 1- and 2-axes being perpendicular. For example, the material shown in Figure 5.3 has the electrodes across the thickness of the material, and hence, this is the 3-axis. An important piezoelectric parameter is the charge coefficient \( d_{ij} \) \((C/N)\). This relates the amount of charge generated on the surfaces of the material on the \( i \)-axis to the force applied on the \( j \)-axis. In the example given, the force applied and the charge generated are both across the thickness of the material, and hence, this charge coefficient is denoted as \( d_{33} \). If a force, \( F_3 \), is applied to the piezoelectric sample, then the charge generated is given by

\[
Q_3 = d_{33} F_3 \tag{5.11}
\]

and so the voltage produced from a rectangular block of area \( A \), thickness \( t \), and relative permittivity \( \varepsilon \) is

\[
V_3 = \frac{Q_3}{C} = \frac{d_{33} F_3 t}{\varepsilon_0 \varepsilon_r A} \tag{5.12}
\]

where \( \varepsilon_0 \) is the permittivity of free space. For a 10 × 10-mm slab of PZT 5H (\( d_{33} = 600 \) pC/N, \( \varepsilon_r = 3,000 \)) of thickness 1 mm, an applied force of 100N will produce an open circuit voltage of 22.6V. Strictly, the value of the relative permittivity is also dependent upon the direction in which it is used and the boundary conditions imposed upon the material. The nomenclature becomes a little cumbersome, however, and for the purpose of this text it should be assumed that the value quoted is for the direction in which the piezoelectric is being used.

Another important piezoelectric constant is the voltage coefficient denoted as \( g_{ij} \). It is related to the \( d \) coefficient as shown here:

\[
g_{ij} = \frac{d_{ij}}{\varepsilon_0 \varepsilon_r} \tag{5.13}
\]

Owing to the inverse piezoelectric effect, an applied electric field will result in a deformation of the material. This gives rise to two definitions of the \( d \) and \( g \) coefficients:

\[
d = \frac{\text{strain developed}}{\text{applied electric field}} \quad (\text{m/V}) = \frac{\text{charge density}}{\text{applied mechanical stress}} \quad (\text{C/N}) \tag{5.14}
\]
and
\[
g = \frac{\text{open circuit electric field}}{\text{applied mechanical stress}} \cdot \frac{V \cdot m}{N} = \frac{\text{strain developed}}{\text{applied charge density}} \cdot \frac{m}{C} \quad (5.15)
\]

Table 5.2 shows some properties of various types of piezoelectric material. A search through the literature will reveal a wide variation in some of these values. In general, manufacturers of bulk piezoelectric materials quote a relatively wide tolerance (20%) on the values of the piezoelectric properties. Measurement of the properties of films deposited onto substrates is notoriously difficult, as the boundary conditions can grossly affect the measured value. Additionally, some materials, such as PZT, are available in a variety of compositions (4D, 5H, 5A, 7A) each exhibiting vastly different figures for their piezoelectric coefficients. The figures quoted in the table are only intended as a rough comparison.

Quartz is a widely used piezoelectric material that has found common use in watches and as a resonant element in crystal oscillators. There are no available methods to deposit it as a thin-film over a silicon substrate. PVDF is a carbon-based polymer material that is readily available in a light, flexible sheet form of typical thickness 9 to 800 µm. It is possible to spin-on films of PVDF onto substrates, but this must be polarized (poled) after processing in order to obtain piezoelectric behavior. Barium titanate and PZT are two examples of piezo ceramic materials and each of these can be deposited onto silicon using a variety of methods including sputtering, screen-printing, and sol-gel deposition. PZT is generally characterized by its relatively high value of \(d_{33}\) and is thus a desirable choice of piezoelectric material. Both zinc oxide and lithium niobate can be deposited as polycrystalline thin-films, but consistent data about their properties is not readily available.

In general, because of the relatively high voltages required for piezoelectric actuators to generate displacements in the micron range, they are not often used. For subnanometer movement, however, they provide an excellent method of actuation. Their high sensitivity to small displacements means that they offer many advantages as micromachined sensors. Devices such as surface acoustic wave sensors (SAWS) and resonant sensors utilize both modes of operation, meaning that only a single material is required for both the sensing and actuating mechanism.

An approximate electrical equivalent circuit of a piezoelectric material is depicted in Figure 5.4. Electrical engineers will recognize the circuit as a series-parallel resonant system. A plot of impedance against frequency is also shown.

The impedance exhibits both resonant and antiresonant peaks at distinct frequencies.

<table>
<thead>
<tr>
<th>Material</th>
<th>Form</th>
<th>(d_{33}) (pC/N)</th>
<th>Relative Permittivity ((\varepsilon_r))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quartz</td>
<td>Single crystal</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>PVDF</td>
<td>Polymer</td>
<td>20</td>
<td>12</td>
</tr>
<tr>
<td>Barium titanate</td>
<td>Ceramic</td>
<td>190</td>
<td>2,000</td>
</tr>
<tr>
<td>PZT</td>
<td>Ceramic</td>
<td>300–600</td>
<td>400–3,000</td>
</tr>
<tr>
<td>Zinc oxide</td>
<td>Single crystal</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Lithium niobate</td>
<td>Single crystal</td>
<td>6–16</td>
<td>30</td>
</tr>
</tbody>
</table>
The mechanical resonance of the device is represented by the series inductor, capacitor and resistor \((L, C_1, R)\) and these are the equivalent of mass, spring, and damper, respectively. Since the piezoelectric material is a dielectric with electrodes, it will have a shunt capacitance associated with it \((C_2)\). The series resonant circuit is responsible for the resonant peak \((f_r)\), and the parallel circuit gives rise to the anti-resonant behavior \((f_a)\). The circuit behaves like a simple capacitor at frequencies below \(f_r\) and like an inductor between \(f_r\) and \(f_a\). After \(f_a\) the impedance decreases with frequency, indicating typical capacitor behavior again. The two resonant frequencies are

\[
f_r = \frac{1}{2\pi \sqrt{LC_1}} \quad \text{and} \quad f_a = \frac{1}{2\pi \sqrt{\frac{C_1 + C_2}{LC_1C_2}}}
\]

\((5.16)\)

### 5.3 Capacitive Techniques

The physical structures of capacitive sensors are relatively simple. The technique nevertheless provides a precise way of sensing the movement of an object. Essentially the devices comprise a set of one (or more) fixed electrode and one (or more) moving electrode. They are generally characterized by the inherent nonlinearity and temperature cross-sensitivity, but the ability to integrate signal conditioning circuitry close to the sensor allows highly sensitive, compensated devices to be produced. Figure 5.5 illustrates three configurations for a simple parallel plate capacitor structure.

![Figure 5.4](image1.png)  
*Figure 5.4* The equivalent electrical circuit of a piezoelectric material.

![Figure 5.5](image2.png)  
*Figure 5.5* Examples of simple capacitance displacement sensors: (a) moving plate, (b) variable area, and (c) moving dielectric.
For a simple parallel plate capacitor structure, ignoring fringing fields, the capacitance is given by

\[ C = \frac{\varepsilon_0 \varepsilon_r A}{d} \text{ (F)} \]  

(5.17)

where \( \varepsilon_0 \) is the permittivity of free space, \( \varepsilon_r \) is the relative permittivity of the material between the plates, \( A \) is the area of overlap between the electrodes, and \( d \) is the separation between the electrodes. The equation shows that the capacitance can be varied by changing one or more of the other variables. Figure 5.5(a) shows the simple case where the lower electrode is fixed and the upper electrode moves. In this case the separation, \( d \), is changing and hence the capacitance varies in a nonlinear manner. Figure 5.5(b) depicts a device where the separation is fixed and the area of overlap is varied. In this configuration, there is a linear relationship between the capacitance and area of overlap. Figure 5.5(c) shows a structure that has both a fixed electrode distance and area of overlap. The movement is applied to a dielectric material (of permittivity \( \varepsilon_2 \)) sandwiched between two electrodes. A common problem to all of these devices is that temperature will affect all three sensing parameters (\( d \), \( A \), and \( \varepsilon_r \)), resulting in changes in the signal output. This effect must be compensated for in some manner, whether by additional signal conditioning circuitry or, preferably, by geometric design.

Figure 5.6 shows a differential capacitance sensor, which is similar in nature to a moving plate capacitor sensor except that there is an additional fixed electrode. Any temperature effects are common to both capacitors and will therefore be cancelled out, as the output signal is a function of the difference between the upper and lower capacitors. If we assume that the outer two electrodes (\( X \) and \( Z \)) are fixed and the inner electrode (\( Y \)) is free to move in a parallel direction towards \( X \), then the gap between plates \( X \) and \( Y \) will decrease and that between \( Y \) and \( Z \) will increase. If the nominal gap distance is \( d \) and the center electrode is moved by a distance \( x \), then the relationship between the differential output voltage and the deflection is given by

\[ (V_2 - V_1) = V_s \frac{x}{d} \]  

(5.18)

Figure 5.6  A differential capacitance sensor.
where $V_s$ is the supply voltage. So this arrangement provides a linear relationship that is preserved over a range of $|x| < d$ and is capable of detecting displacement of a few picometers.

Capacitor structures are relatively straightforward to fabricate, and membrane-type devices are often used as the basis for pressure sensors and microphones. More elaborate structures, such as interdigitated capacitors, are also used, and the effects of the fringing fields cannot always be ignored. With such devices, the simple parallel plate capacitor equation only provides a crude estimate of the expected capacitance change.

Capacitive techniques are inherently less noisy than those based on piezoresistance owing to the lack of thermal (Johnson) noise. With micromachined devices, however, the values of capacitance are extremely small (in the range of femto- to attofarads), and the additional noise from the interface electronic circuits often exceeds that of a resistance-based system.

There are a variety of techniques for measuring capacitance changes including charge amplifiers (often used with piezoelectric devices), charge balance techniques, ac bridge impedance measurements, and various oscillator configurations. There are also a variety of commercially available ICs that can be used to measure capacitance changes of a few femtofarads in stray capacitances up to several hundred picofarads [2].

### 5.4 Optical Techniques

Optical sensing techniques primarily rely on modulating the properties of an optical frequency electromagnetic wave. In the case of optical sensors, the measurand directly modulates the properties of the electromagnetic wave. In the case of microsensors, which use optical interfacing, the miniaturized sensor interacts with the measurand. The microsensor then modulates a property of the optical signal in order to provide an indication of the measurand.

The following properties of the electromagnetic wave can be altered:

1. Intensity;
2. Phase;
3. Wavelength;
4. Spatial position;
5. Frequency;
6. Polarization.

The basic principles of each of these techniques will now be reviewed in turn.

#### 5.4.1 Intensity

The primary advantage of intensity modulation is that intensity variations are simply detected because all optical detectors (e.g., photodiodes, phototransistors) directly respond to intensity variations. Therefore, if the microsensor can be arranged to vary the intensity of an optical signal, these variations can then be simply observed using a
photodetector. A simple arrangement is for the microsensor to move in response to the measurand and for this movement to be arranged to block the path of the light beam incident on a photodetector. Figure 5.7 illustrates a simple transmissive arrangement, although reflected light is also used in some arrangements.

The optical source is shown as a light emitting diode (LED) since a coherent source is not required for intensity-based sensors. Alternative optical sources could be a laser, the output of an optical fiber, or simply an incandescent lamp.

The major difficulty with intensity-based systems is variations in intensity caused by factors not related to the measurand. For example, the output of an optical source can vary with time and temperature. For this reason intensity-based sensors often incorporate some form of reference measurement of the optical source intensity and a ratio taken between the optical intensity before and after modulation by the microsensor. This problem often negates the simplicity of intensity-based sensors. Variations in the sensitivity of the optical detector can also cause difficulties and complications.

A qualitative estimate of the resolution of intensity-based sensors can be obtained by estimation of the optical beam size. The minimum beam size is of the order of the wavelength of the optical source, so this gives an indication of the displacement required to give a 100% modulation of intensity.

### 5.4.2 Phase

As photodetectors do not respond directly to phase variation, it is necessary to convert a variation in phase to an intensity variation for measurement at the photodiode. This is usually achieved by using an interferometer to combine one or more optical beams that have interacted with the microsensor with one or more optical beams that are unaffected by the microsensor. A coherent source such as a laser diode is therefore typically used in phase-based optical sensing. The interaction with the microsensor has the effect of altering the optical path length of that optical beam and hence its phase. This can simply be achieved by reflecting the optical beam off the microsensor and the microsensor moving in response to the measurand so as to vary the optical path length.

A major advantage of phase-based systems is that subwavelength phase variations can be resolved, which equates to submicron displacement of the microsensor. Difficulties can be caused by the fact that the output of the interferometer is periodic; therefore, care has to be taken to establish the start point and the position relative to that. This can lead to complexity in the reference electrodes and errors in initializing the system.
5.4.3 Wavelength

Wavelength-based sensing relies on the source spectrum being modulated by interaction with the microsensor. Normally a source with a broad spectrum is used. The light returned from the microsensor is split into spectral segments and incident on a photodetector for measurement of its intensity. By a prior knowledge of the potential modulation mechanism present with the microsensor, one can identify the measurand and its magnitude. A good example of a wavelength-based sensor is one based on the gas absorption, which is highly wavelength specific according to the quantity of gas present.

The advantage of wavelength-based sensors is that they can be made insensitive to intensity variation since these affect the whole spectrum in the same way. Therefore, the measurement of a nonabsorbed wavelength can be used to reference the absorbed wavelength, therefore compensating for intensity variations. In addition, wavelength-based sensors often lend themselves to the measurement of multiple parameters since the light spectrum can be divided according to the particular wavelength corresponding to the measurand of interest.

5.4.4 Spatial Position

Figure 5.8 illustrates the principle of the modulation of special position by means of the movement of a microsensor. This technique is often known as triangulation.

This technique is simple to implement and has the advantage of immunity to source intensity variations. Its resolution is less then phase-based techniques.

5.4.5 Frequency

If optical radiation at a frequency $f$ is incident upon a body moving a velocity $v$, then the radiation reflected from the moving body appears to have a frequency $f_1$, where

$$ f_1 = \frac{f}{1 - \frac{v}{c}} \approx f \left(1 + \frac{v}{c}\right) $$  \hspace{1cm} (5.19)

![Figure 5.8](image)  

**Figure 5.8** An example of a spatial position measurement system.
This Doppler frequency shift from a moving target can therefore be used as the basis of a detection technique of the velocity of the target. Laser Doppler velocimetry is a well-established field of research. Frequency variation is converted into intensity variation by interferometry by combining a nonfrequency-shifted reference beam with the shifted beam.

### 5.4.6 Polarization

Linear polarization is defined by the direction of the electric vector of the electromagnetic wave. Circular polarized light is defined by the direction of rotation of the electric field vector when viewed looking towards the source. Any polarization can be resolved into two orthogonal modes, and sensing can be achieved by altering the optical path length traversed by one mode with respect to the other. In practice this is normally achieved by a relative modification of the refractive index. A polarized light source such as a laser is required and the photodetector must be made polarization sensitive by including a polarizer.

Polarization-based interrogation of microsensors has not been widely investigated owing to the limited sensitivity available, as it is a differential technique. In addition, the method is susceptible to intensity changes in the source.

### 5.5 Resonant Techniques

A resonator is a mechanical structure designed to vibrate at a particular resonant frequency. Resonators can be fabricated from a range of single crystal materials with micron-sized dimensions using various micromachining processes. The resonant frequencies of such microresonators are extremely stable, enabling them to be used as a time base (the quartz tuning fork in watches, for example) or as the sensing element of a resonant sensor [3, 4]. The performance benefits of a well-designed resonant sensor compared with piezoresistive and capacitive techniques are shown in Table 5.3 [5]. The fabrication of such devices is, however, more complex and the requirement for packaging such devices more demanding.

A block diagram of a typical resonant sensor is shown in Figure 5.9 [6]. A resonant sensor is designed such that the resonator’s natural frequency is a function of the measurand. The measurand typically alters the stiffness, mass, or shape of the resonator, hence causing a change in its resonant frequency. The other components of a resonant sensor are the vibration drive and detection mechanisms. The drive mechanism excites the vibrations in the structure while the detection mechanism senses these vibrations. The frequency of the detected vibration forms the output of

<table>
<thead>
<tr>
<th>Feature</th>
<th>Resonant</th>
<th>Piezoresistive</th>
<th>Capacitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output form</td>
<td>Frequency</td>
<td>Voltage</td>
<td>Voltage</td>
</tr>
<tr>
<td>Resolution</td>
<td>1 part in $10^8$</td>
<td>1 part in $10^5$</td>
<td>1 part in $10^7$–$10^9$</td>
</tr>
<tr>
<td>Accuracy</td>
<td>100–1000 ppm</td>
<td>500–10,000 ppm</td>
<td>100–10,000 ppm</td>
</tr>
<tr>
<td>Power consumption</td>
<td>0.1–10 mW</td>
<td>$=10$ mW</td>
<td>$&lt;0.1$ mW</td>
</tr>
<tr>
<td>Temperature cross-sensitivity</td>
<td>$-30 \times 10^{-6} ^\circ C$</td>
<td>$-1,600 \times 10^{-6} ^\circ C$</td>
<td>$4 \times 10^{-6} ^\circ C$</td>
</tr>
</tbody>
</table>

Source: [5].
the sensor and this signal is also fed back to the drive mechanism via an amplifier maintaining the structure at resonance over the entire measurand range.

In mechanical sensing applications, the most common mechanism for coupling the resonator to the measurand is to apply a strain across the structure. When used in such a manner the resonator effectively becomes resonant strain gauge. Coupling to the measurand is achieved by mounting the resonator in a suitable location on a specifically designed sensing structure that deflects due to the application of the measurand. The resonator output can be used to monitor the deflection of the sensing structure and thereby provide an indication of the magnitude of the measurand. When used as a resonant strain gauge, the applied strain effectively increases the stiffness of the resonator, which results in an increase in its natural frequency. This principle is commonly applied in force sensors, pressure transducers, and accelerometers (see Chapters 6 through 8 for detailed examples).

Coupling the measurand to the mass of the resonator can be achieved by surrounding the structure by a liquid or gas, by coating the resonator in a chemically sensitive material, or by depositing material onto the resonator. The presence of the surrounding liquids or gases increases the effective inertia of the resonator and lowers its resonant frequency. Density sensors and level sensors are examples of mass coupled resonant sensors. Coating the resonator in a chemically sensitive material is used in gas sensors. The sensitive material absorbs molecules of a particular gas, adding to the mass of the film and thereby reducing the frequency of the resonator.

The shape coupling effect is similar to the strain effect except changes in the measurand alter the geometry of the resonator, which leads to a shift in the resonant frequency. This is the least commonly used coupling mechanism.

### 5.5.1 Vibration Excitation and Detection Mechanisms

The piezoelectric nature of GaAs and quartz materials enables straightforward excitation and detection of resonant modes of vibrations [7]. Suitable electrode materials must be deposited and patterned on the surface of the resonator. The location and geometry of the electrodes should be carefully designed to maximize the electrical to mechanical coupling with the desired mode of operation (drive efficiency). Maximizing this coupling will promote the excitation of the desired mode and maximize the corresponding vibration detection signal.
The excitation and detection of resonance in silicon microresonators are not so straightforward because silicon is not intrinsically piezoelectric. Other mechanisms must therefore be fabricated on or adjacent to the resonator structure. There are many suitable mechanisms and these are all based on the sensing and actuating principles described in this chapter. For example, the resonators vibrations can be electrostatically excited and detected using implanted piezoresistors. Since the implanted piezoresistors could be used directly to measure the strain in the sensing structure, the added complexity of a resonant approach is only justifiable in high-performance sensing applications.

The various excitation and detection mechanisms used with silicon resonators are summarized in Table 5.4. Many of the mechanisms listed can be used to both excite and detect a resonator’s vibrations, either simultaneously or in conjunction with another mechanism. Devices where a single element combines the excitation and detection of the vibrations in the structure are termed one-port resonators. Those that use separate elements are termed two-port resonators.

The suitability of these mechanisms for driving or detecting a resonator’s vibrations depends upon a number of factors: the magnitude of the drive forces generated, the coupling factor (or drive efficiency), sensitivity of the detection mechanism, the effects of the chosen mechanism upon the performance and behavior of the resonator, and practical considerations pertaining to the fabrication of the resonator and the sensors final environment.

### 5.5.2 Resonator Design Characteristics

#### 5.5.2.1 Q-Factor

As a structure approaches resonance, the amplitude of its vibration will increase, its resonant frequency being defined as the point of maximum amplitude. The magnitude of this amplitude will ultimately be limited by the damping effects acting on the system. The level of damping present in a system can be defined by its quality factor ($Q$-factor). The $Q$-factor is a ratio of the total energy stored in the system ($E_M$) to the energy lost per cycle ($E_C$) due to the damping effects present:

$$ Q = 2\pi \frac{E_M}{E_C} $$

A high $Q$-factor indicates a pronounced resonance easily distinguishable from nonresonant vibrations, as illustrated in Figure 5.10. Increasing the sharpness of the resonance enables the resonant frequency to be more clearly defined and will improve the performance and resolution of the resonator. It will also simplify the operating electronics since the magnitude of the signal from the vibration detection

<table>
<thead>
<tr>
<th>Excitation Mechanism</th>
<th>Detection Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Piezoelectric</td>
<td>Piezoelectric</td>
</tr>
<tr>
<td>Magnetic</td>
<td>Magnetic</td>
</tr>
<tr>
<td>Electrothermal</td>
<td>Piezoresistive</td>
</tr>
<tr>
<td>Optothermal</td>
<td>Optical</td>
</tr>
</tbody>
</table>

*Source: [8].*
mechanism will be greater than that of a low-

A high $Q$ means little energy is required to maintain the resonance at constant amplitude, thereby broadening the range of possible drive mechanisms to include weaker techniques. A high $Q$-factor also implies the resonant structure is well isolated from its surroundings, and therefore, the influence of external factors (e.g., vibrations) will be minimized.

The $Q$-factor can also be calculated from Figure 5.10 using

$$Q = \frac{f_0}{\Delta f}$$

where resonant frequency $f_0$ corresponds with $a_{\text{max}}$, the maximum amplitude, and $\Delta f$ is the difference between frequencies $f_1$ and $f_2$. Frequencies $f_1$ and $f_2$ correspond to amplitudes of vibration 3 dB lower than $a_{\text{max}}$.

The $Q$-factor is limited by the various mechanisms by which energy is lost from the resonator. These damping mechanisms arise from three sources:

1. The energy lost to a surrounding fluid ($1/Q_a$);
2. The energy coupled through the resonator’s supports to a surrounding solid ($1/Q_s$);
3. The energy dissipated internally within the resonator’s material ($1/Q_i$).

Minimizing these effects will maximize the $Q$-factor as shown here:

$$\frac{1}{Q} = \frac{1}{Q_a} + \frac{1}{Q_s} + \frac{1}{Q_i}$$

Energy losses associated with $1/Q_a$ are potentially the largest, and therefore the most important, of the loss mechanisms. These losses occur due to the interactions
of the oscillating resonator with the surrounding gas. There are several distinguishable loss mechanisms and associated effects. The magnitude of each depends primarily upon the nature of the gas, surrounding gas pressure, size and shape of the resonator, the direction of its vibrations, and its proximity to adjacent surfaces. Gas damping effects can be negated completely by operating the resonator in a suitable vacuum, and this is used in most micromechanical resonator applications.

Molecular damping occurs at low pressures of between 1 and 100 Pa when the surrounding gas molecules act independently of one another [9]. The damping effect arises from the collisions between the molecules and the resonator’s surface as it vibrates. This causes the resonator and molecules to exchange momentum according to their relative velocities. The magnitude of the loss is directly proportional to the surrounding fluid pressure, and also close proximity of the oscillating structure to adjacent surfaces will exaggerate the damping effects. Viscous damping predominates at pressures above 100 Pa where the molecules can no longer be assumed to act independently and the surrounding gas must be considered as a viscous fluid. Viscous drag occurs as the fluid travels over the surface of the resonator. The formation of boundary layer around the resonator can also result in the vibrations forming a transverse wave, which travels into the fluid medium. Other damping mechanisms associated with surrounding fluids are acoustic radiation and squeezed film damping.

Structural damping, \( \frac{1}{Q_s} \), is associated with the energy coupled from the resonator through its supports to the surrounding structure and must be minimized by careful design of the resonant structure. Minimizing the energy lost from the resonator to its surroundings can be achieved by designing a balanced resonant structure, supporting the resonator at its nodes, or by employing a decoupling system between the resonator and its support.

The coupling mechanism between the resonator and its support can be illustrated by observing a fixed-fixed beam vibrating in its fundamental mode. Following Newton’s second law that every action has an equal and opposite reaction, the reaction to the beam’s vibrations is provided by its supports. The reaction causes the supports to deflect and as a result energy is lost from the resonator.

The degree of coupling of a fixed-fixed beam can be reduced by operating it in a higher-order mode. For example, the second mode in the plane of vibrations shown above will possess a node halfway along the length of the beam. The beam will vibrate in antiphase either side of the node, and the reactions from each half of the beam will cancel out at the node. There will inevitably still be a reaction at each support, but the magnitude of each reaction will be less than for mode 1. The use of such higher order modes is limited by their reduced sensitivity to applied stresses and the fact there will always be a certain degree of coupling.

Balanced resonator designs operate on the principle of providing the reaction to the structure’s vibrations within the resonator. Multiple-beam style resonators, for example, incorporate this inherent dynamic moment cancellation when operated in a balanced mode of vibration. Examples of such structures are the double-ended tuning fork (DETF), which consists of two beams aligned alongside each other, and the triple-beam tuning fork (TBTF), which consists of three beams aligned alongside each other, the center tine being twice the width of the outer tines. Figure 5.11 shows these structures and their optimum modes of operation.
1/Q\(_s\) is of fundamental importance since it not only affects the Q-factor of the resonator, but provides a key determinant of resonator performance. A dynamically balanced resonator design that minimizes 1/Q\(_s\) provides many benefits [10]:

- High resonator Q-factor and therefore good resolution of frequency;
- A high degree of immunity to environmental vibrations;
- Immunity to interference from surrounding structural resonances;
- Improved long-term performance since the influence of the surrounding structure on the resonator is minimized.

The Q-factor of a resonator is ultimately limited by the energy loss mechanisms within the resonator material. This is illustrated by the fact that even if the external damping mechanisms 1/Q\(_a\) and 1/Q\(_s\) are removed, the amplitude of its vibrations will still decay with time. There are several internal loss mechanisms by which vibrations can be attenuated. These include the movement of dislocations and scattering by impurities, phonon interaction, and the thermoelastic effect.

### 5.5.2.2 Nonlinear Behavior and Hysteresis

Nonlinear behavior becomes apparent at higher vibration amplitudes when the resonator’s restoring force becomes a nonlinear function of its displacement. This effect is present in all resonant structures. In the case of a flexurally vibrating fixed-fixed beam, the transverse deflection results in a stretching of its neutral axis. A tensile force is effectively applied and the resonant frequency increases. This is known as the hard spring effect. The magnitude of this effect depends upon the boundary conditions of the beam. If the beam is not clamped firmly, the nonlinear relationship can exhibit the soft spring effect whereby the resonant frequency falls with increasing amplitude. The nature of the effect and its magnitude also depends upon the geometry of the resonator.

The equation of motion for an oscillating force applied to an undamped structure is given by (5.23) where \(m\) is the mass of the system, \(F\) is the applied driving force, \(\omega\) is the frequency, \(y\) is the displacement, and \(s(y)\) is the nonlinear function [11].

\[
m\ddot{y} + s(y) = F_0 \cos \omega t \tag{5.23}
\]
In many practical cases $s(y)$ can be represented by (5.24), the nonlinear relationship being represented by the cubic term.

$$s(y) = s_1 y + s_3 y^3$$  \hspace{1cm} (5.24)

Placing (5.24) in (5.23), dividing through by $m$, and simplifying gives

$$s + s_1 / m (y + s_3 / s_1 y^3) = F_0 \cos \omega$$

where $s/m$ equals $\omega_{or}^2 (\omega_{or}$ representing the resonant frequency for small amplitudes of vibration) and $s_3/s_1$ is denoted by $\beta$. The restoring force acting on the system is therefore represented by

$$R = -\omega_{or}^2 (y + \beta y^3)$$  \hspace{1cm} (5.26)

If $\beta$ is equal to zero, the restoring force is a linear function of displacement; if $\beta$ is positive, the system experiences the hard spring nonlinearity; a negative $\beta$ corresponds to the soft spring effect. The hard and soft nonlinear effects are shown in Figure 5.12. As the amplitude of vibration increases and the nonlinear effect becomes apparent, the resonant frequency exhibits a quadratic dependence upon the amplitude, as shown in

$$\omega_r = \omega_{or} (1 + \frac{1}{\beta} \frac{\beta y^2}{\omega_{or}^2})$$  \hspace{1cm} (5.27)

The variable $\beta$ can be found by applying (5.27) to an experimental analysis of the resonant frequency and maximum amplitude for a range of drive levels.

The amplitude of vibration is dependent upon the energy supplied by the resonator’s drive mechanism and the $Q$-factor of the resonator. Driving the resonator too hard or a high $Q$-factor that results in excessive amplitudes at minimum practical drive levels can result in undesirable nonlinear behavior. Nonlinearities are undesirable since they can adversely affect the accuracy of a resonant sensor. If a resonator is driven in a nonlinear region, then changes in amplitude—due, for

![Figure 5.12](image)

**Figure 5.12** The hard and soft nonlinear effects.
example, to amplifier drift—will cause a shift in the resonant frequency indistinguishable from shifts due to the measurand. The analysis of a resonator’s nonlinear characteristics is therefore important when determining a suitable drive mechanism and its associated operating variables.

A nonlinear system can exhibit hysteresis if the amplitude of vibration increases beyond a critical value. Hysteresis occurs when the amplitude has three possible values at a given frequency. This critical value can be determined by applying

$$y_0^3 > \frac{8h}{3\omega_0^2\beta}$$

(5.28)

where $h$ is the damping coefficient and can be found by measuring the $Q$-factor of the resonator at small amplitudes and applying

$$Q = \frac{\omega_0}{2h}$$

(5.29)

### 5.6 Actuation Techniques

In Chapter 1 we defined an actuator as a device that responds to the electrical signals within the transduction system. Specifically, a mechanical actuator is one that translates a signal from the electrical domain into the mechanical domain. In the ideal case, we would like the conversion to be 100% efficient. Of course, any real system cannot achieve a figure anywhere near this, owing to internal and external losses. Typical micromechanical actuators offer an efficiency between 5% and 35%. Other factors such as ease of fabrication, robustness, resistance to external effects (i.e., temperature, humidity), and range of motion, result in a series of trade-offs for selecting the appropriate mechanism.

For the purpose of this text, four fundamental approaches for actuator design will be discussed. Other techniques such as chemical and biological actuation are not covered here.

#### 5.6.1 Electrostatic

Electrostatic actuators are based on the fundamental principle that two plates of opposite charge will attract each other. They are quite extensive as they are relatively straightforward to fabricate. They do, however, have a nonlinear force-to-voltage relationship. Consider a simple, parallel plate capacitor arrangement again, having a gap separation, $g$, and area of overlap, $A$, as shown in Figure 5.13. Ignoring fringing effects, the energy stored at a given voltage, $V$, is

![Figure 5.13](image-url)
and the force between the plates is given by

\[ F = \frac{dW}{dg} = \frac{\varepsilon_0 \varepsilon_r A V^2}{2g^2} \]  

(5.31)

It is therefore clear that the force is a nonlinear function of both the applied voltage and the gap separation. Use of closed loop control techniques can linearize the response.

An alternative type of electrostatic actuator is the so-called comb-drive, which is comprised of many interdigitated electrodes (fingers) that are actuated by applying a voltage between them. The geometry is such that the thickness of the fingers is small in comparison to their lengths and widths. The attractive forces are therefore mainly due to the fringing fields rather than the parallel plate fields, as seen in the simple structure above. The movement generated is in the lateral direction, as shown in Figure 5.14, and because the capacitance is varied by changing the area of overlap and the gap remains fixed, the displacement varies as the square of the voltage.

The fixed electrode is rigidly supported to the substrate, and the movable electrode must be held in place by anchoring at a suitable point away from the active fingers. Additional parasitic capacitances such as those between the fingers and the substrate and the asymmetry of the fringing fields can lead to out-of-plane forces, which can be minimized with more sophisticated designs.

Electrostatic actuation techniques have also been used to developed rotary motor structures. With these devices, a central rotor having surrounding capacitive plates is made to rotate by the application of voltages of the correct phase to induce rotation. Such devices have been shown to have a limited lifetime and require lubrication to prevent the rotor from seizing. The practical use has therefore been limited, but they are, nevertheless, the subject of intensive research.

Figure 5.14 An illustration of the electrostatic comb-drive actuator.
Another interesting type of electrostatic actuator is the so-called scratch drive actuator (SDA) as described by Akiyama and Katsufusa [12]. The device comprises a flexible, electrode plate and a small bushing at one end. It is depicted in Figure 5.15, which also illustrates the principle of operation. The free end of the electrode in the actual device is usually supported by a thin beam, but this is not shown in the figure. When a voltage is applied between the electrode plate and the buried electrode layer on the substrate, the plate buckles down and so causes the bushing to “scratch” along the insulator, thereby resulting in a small forward movement. When the voltage is removed, the plate returns to its original shape, thereby resulting in a net movement of the plate. The cycle can be repeated for stepwise linear motion.

![Figure 5.15](image-url)

**Figure 5.15** Illustration of the principle of operation of the electrostatic scratch drive actuator as described by Akiyama and Katsufusa. (After: [12].)

Another interesting type of electrostatic actuator is the so-called scratch drive actuator (SDA) as described by Akiyama and Katsufusa [12]. The device comprises a flexible, electrode plate and a small bushing at one end. It is depicted in Figure 5.15, which also illustrates the principle of operation. The free end of the electrode in the actual device is usually supported by a thin beam, but this is not shown in the figure. When a voltage is applied between the electrode plate and the buried electrode layer on the substrate, the plate buckles down and so causes the bushing to “scratch” along the insulator, thereby resulting in a small forward movement. When the voltage is removed, the plate returns to its original shape, thereby resulting in a net movement of the plate. The cycle can be repeated for stepwise linear motion.

![Figure 5.16](image-url)

**Figure 5.16** An example of a simple cantilever beam with a deposited piezoelectric layer: (a) the structure with no applied voltage; and (b) how the tip of the beam moves upon the application of an applied voltage.
5.6.2 Piezoelectric

As we have already seen, piezoelectric devices can be used for both sensor and actuator applications. An applied voltage across the electrodes of a piezoelectric material will result in a deformation that is proportional to the magnitude of the voltage (strictly electric field). The displacement across a bulk sample of PZT with an actuation voltage of several hundred volts, for example, is only a small fraction of a micron. When such a system is scaled down to that of a typical MEMS actuator, a displacement of several orders of magnitude less is obtained! For this reason, some form of mechanical amplification is needed in order to generate useful displacements. Such a device can be fabricated by depositing a piezoelectric film onto a substrate in the form of a cantilever beam as shown in Figure 5.16. This type of structure is referred to as a piezoelectric unimorph. The deflection at the free end of the beam is greater than that produced in the film itself.

Piezoelectric actuators are often used in micropumps (see Chapter 9) as a way of deflecting a thin membrane, which in turn alters the volume within a chamber below. Such a structure is depicted in Figure 5.17. The device comprises two silicon wafers bonded together. The lower wafer comprises an inlet and outlet port, which have been fabricated using bulk micromachining techniques. The upper wafer has been etched to form the pump chamber. The shape of the ports gives rise to a preferential direction for the fluid flow, although there is a degree of flow in the reverse direction during pumping. So the ports behave in a similar manner to valves. An alternative structure comprises cantilever-type flaps across the ports, but these often suffer from stiction during pumping. When a voltage is applied to the piezoelectric material, this results in a deformation of the thin membrane and hence changes the volume within the chamber. This is depicted in Figure 5.17(b). Typical flow rates are in the range of nanoliters to microliters per minute, depending on the dimension of the micropump.

5.6.3 Thermal

Thermal actuation techniques tend to consume more power than electrostatic or piezoelectric methods, but the forces generated are also greater. One of the basic
approaches is to exploit the difference in linear expansion coefficients of two materials bonded together. Such structures are often referred to as thermal bimorphs and are analogous to the familiar bimetallic strips often used in thermostats. One layer expands by a different amount to the other, resulting in thermal stresses at the interface leading to bending of the structure. The amount of bending depends on the difference in thermal coefficients of expansion and also on the temperature. An illustration of a thermal bimorph is shown in Figure 5.18. If an electric current is passed through the aluminum layer, it heats up (Joule heating), thereby causing the free end of the beam to move. These devices are relatively straightforward to fabricate and in addition to consuming relatively large amounts of power, they also have a low bandwidth because of the thermal time constant of the overall structure (i.e., beam and support).

An example of a commercial device based on thermal actuation is the so-called fluistor from Redwood Microsystems in California. This device is comprised of a cavity with a sealed fluid that can be heated and thus expanded. The heat is applied to the fluid via a thin-film resistive element. If one section of the cavity, such as a wall, is made more compliant than the other sections, then it will deform under pressure, thereby generating a mechanical force. The cavity is formed by bulk micromachining in silicon and is sealed using a Pyrex wafer, containing the heating element, anodically bonded to the silicon. Strictly, this is a thermopneumatic actuator and the commercial device is often used as a microvalve in applications such as medical instrumentation, gas mixers, and process control equipment. Such actuators may require up to 2W of power to operate.

Another thermal effect that can be exploited in thermal actuators is the shape-memory effect, which is a property of a special class of metal alloys known as shape-memory alloys. When these materials are heated beyond a critical transition temperature, they return to a predetermined shape. The SMA material has a temperature-dependent crystal structure such that, at temperatures below the transition point, it possesses a low yield strength crystallography referred to as a Martensite. In this state, the alloy is relatively soft and easy to deform into different shapes.

Figure 5.18 A simple thermal bimorph actuator (a) before and (b) after the application of electric current.
It will retain this shape until the temperature exceeds the phase transition temperature, at which point the material reverts to its parent structure known as Austenite. One of the most widely used SMA materials is an alloy of nickel and titanium called Nitinol. This has excellent electrical and mechanical properties and a long fatigue life. In its bulk form, it is capable of producing up to 5% strain. The transition temperature of Nitinol can be tailored between -100°C and +100°C by controlling the impurity concentration. The material has been used in MEMS by sputter depositing TiNi thin-film layers [13].

5.6.4 Magnetic

If a current-carrying element is placed within a magnetic field, an electromagnetic force (Lorentz force) will occur in a direction perpendicular to the current and magnetic field. The magnitude of the force is proportional to the current, length of the element, and the magnetic field. The availability of permanent magnetic materials, which are compatible with MEMS processing, is very limited, and thus it is common for the magnetic field to be generated externally. Discrete magnetic actuators often comprise coils, but such structures are not currently achievable with conventional MEMS processing and planar coils must be used.

Another approach that can be used as the basis of a magnetic actuator is the magnetostrictive effect. Magnetostriction is defined as the dimensional change of a magnetic material caused by a change in its magnetic state. Like the piezoelectric effect, it is reversible, and an applied stress results in a change of magnetic state. All magnetic materials exhibit varying degrees of magnetostriction. J. P. Joule discovered the effect in 1847 by observing the change in length of an iron bar when it was magnetized. A popular modern-day magnetostrictive material is Terfenol-D, an alloy of terbium, dysprosium, and iron. The magnetostriction of Terfenol-D is several orders of magnitude greater than that of iron, nickel, or cobalt and gives rise to strains in the region of $2 \times 10^{-3}$. Bulk Terfenol-D produces much larger strains than those achievable with piezoelectric materials. Research has been undertaken to investigate the feasibility of depositing thin and thick-films of magnetostrictive material onto substrates such as silicon, glass, and alumina; the magnetostriction achievable, however, is inferior to that of the bulk material.

Figure 5.19 shows an example of a magnetic actuator as described by Judy et al. [14]. The device comprises a 7-µm-thick layer of Permalloy, which was electroplated onto a polysilicon cantilever. The root of the beam is thin and narrow and acts as a spring, thereby allowing the tip to deflect over a wide angular range. The magnetic field is applied externally to the device, and this causes a deflection of the actuator in the direction of the plane of the substrate. The device is made using polysilicon surface micromachining techniques. Deflections exceeding 90° were achieved with this configuration.

5.7 Smart Sensors

Advances in the area of microelectronics in recent years have had a major effect on many aspects of measurement science. In particular, the distinction between the
sensor and the instrument may not be apparent. Many of today’s commercial
devices have some form of electronic processing within the main sensor housing;
perhaps simple electronic filtering or more sophisticated digital signal processing.
The terms intelligent and smart sensor have been used, almost interchangeably, over
the past 20 years or so to refer to sensors having additional functionality provided
by the integration of microprocessors, microcontrollers, or application specific inte-
grated circuits (ASICs) with the sensing element itself. The interested reader is
encouraged to read the texts by Brignell and White [15], Gardner et al. [16], and
Frank [17], for a deeper insight into the field of smart sensor technologies. For con-
sistency in this text, we will adopt the term smart sensor to refer to a microsensor
with integrated microelectronic circuitry.

Smart sensors offer a number of advantages for sensor system designers. The
integration of sensor and electronics allows it to be treated as a module, or
black-box, where the internal complexities of the sensor are kept remote from the
host system. Smart sensors may also have additional integrated sensors to monitor,
say, localized temperature changes. This is sometimes referred to as the sensor-
within-a-sensor approach and is an important feature of smart sensor technology.
An example of a smart sensor system is depicted in Figure 5.20.

Many physical realizations of smart sensors may contain some or all of these ele-
ments. Each of the main subsystems will now be described in more detail.

The sensing element is the primary source of information into the system. Exam-
pies of typical sensing techniques have already been outlined in this chapter. The
smart sensor may also have the ability to stimulate the sensing element to provide a
self-test facility, whereby a reference voltage, for example, can be applied to the
sensor in order to monitor its response. Some primary sensors, such as those based
on piezoelectrics, convert energy directly from one domain into another and there-
fore do not require a power supply. Others, such as resistive-based sensors, may
need stable dc sources, which may benefit from additional functionality like pulsed
excitation for power-saving reasons. So excitation control is another distinguishing
feature found in smart sensors.
Amplification is usually a fundamental requirement, as most sensors tend to produce signal levels that are significantly lower than those used in the digital processor. Resistive sensors in a bridge configuration often require an instrumentation amplifier; piezoelectric devices may need a charge amplifier. If possible, it is advantageous to have the gain as close as possible to the sensing element. In situations where a high gain is required, there can often be implications for handling any adverse effects such as noise. In terms of chip layout, the sharp transients associated with digital signals need to be kept well away from the front-end analog circuitry.

Examples of analog processing include antialiasing filters for the conversion stage. In situations where real-time processing power is limited, there may also be benefits in implementing analog filters.

Data conversion is the transition region between the continuous (real-world) signals and the discrete signals associated with the digital processor. Typically, this stage comprises an analog-to-digital converter (ADC). Inputs from other sensors (monitoring) can be fed into the data conversion subsystem and may be used to implement compensation, say for temperature. Note that such signals may also require amplification before data conversion. Resonant sensors, whose signals are in the frequency domain, do not need a data conversion stage as their outputs can often be fed directly into the digital system.

The digital processing element mainly concerns the software processes within the smart sensor. These may be simple routines such as those required for implementing sensor compensation (linearization, cross-sensitivity, offset), or they may be more sophisticated techniques such as pattern recognition methods (such as neural networks) for sensor array devices.

The data communications element deals with the routines necessary for passing and receiving data and control signals to the sensor bus. It is often the case that the smart sensor is a single device within a multisensor system. Individual sensors
can communicate with each other in addition to the host system. There are many examples of commercial protocols that are used in smart sensor systems, but we will not go into detail here. It is sufficient to be aware that the smart sensor will often have to deal with situations such as requests for data, calibration signals, error checking, and message identification. Of course, it is feasible in some applications that the data communications may simply be a unit that provides an analog voltage or current signal.

The control processor often takes the form of a microprocessor. It is generally the central component within the smart sensor and is connected to most of the other elements, as we have already seen. The software routines are implemented within the processor and these will be stored within the memory unit. The control processor may also issue requests for self-test routines or set the gain of the amplifier.

References