Mass and weight are often used interchangeably; however, they are different. Mass is a quantitative measure of inertia of a body at rest. As a physical quantity, mass is the product of density and volume.

Weight or weight force is the force with which a body is attracted toward the Earth. Weight force is determined by the product of the mass and the acceleration of gravity.

\[ M = V \times D \quad (20.1) \]

where

- \( M \) = mass
- \( V \) = volume
- \( D \) = density

Note: In most books, the symbol for density is the Greek letter \( \rho \).

\[ W = M \times G \quad (20.2) \]

where

- \( W \) = weight
- \( G \) = gravity

The embodiment of units of mass are called weights; this increases the confusion between mass and weight. In the International System of Units (SI), the modernized metric measurement system, the unit for mass is called the kilogram and the unit for force is called the newton. In the United States, the customary system the unit for mass is called the slug and the unit for force is called the pound. When using the U.S. customary units of measure, people are using the unit pound to designate the mass of an object because, in the United States, the pound has been defined in terms of the kilogram since 1893.

### 20.1 Weighing Instruments

Weighing is one of the oldest known measurements, dating back to before written history. The equal arm balance was probably the first instrument used for weighing. It is a simple device in which two pans are suspended from a beam equal distance from a central pivot point. The standard is placed in one pan and the object to be measured is placed in the second pan; when the beam is level, the unknown is equal to the standard. This method of measurement is still in wide use throughout the world today. Figure 20.1 shows an Ainsworth equal arm balance.
A balance is a measuring instrument used to determine the mass of an object by measuring the force exerted by the object on its support within the gravitational field of the Earth. One places a standard of known value on one pan of the balance. One then adds the unknown material to the second pan, until the gravitational force on the unknown material equals the gravitational force on the standard. This can be expressed mathematically as:

\[ S \times G = U \times G \]  

where

- \( S \) = mass of the standard
- \( G \) = gravity
- \( U \) = mass of the unknown

Given the short distance between pans, one assumes that the gravitational forces acting on them are equal. Another assumption is that the two arms of the balance are equal.

Since the gravitational force is equal, it can be removed from the equation and the standard and the unknown are said to be equal. This leads to one of the characteristics of the equal arm balance as well as of other weighing devices, the requirement to have a set of standards that allows for every possible measured value. The balance can only indicate if objects are equal and has a limited capability to determine how much difference there is between two objects.
Probably, the first attempt to produce direct reading balances was the introduction of the single pan substitution balance. A substitution balance is, in principle, similar to an equal arm balance. The object to be measured is placed in the weighing pan; inside the balance are a series of calibrated weights that can be added to the standard side of the balance, through the use of dials and levers. The standard weights can be added and subtracted through the use of the balance's mechanical system, to equal a large variety of weighing loads. Very small differences between the standard weights and the load are read out on an optical scale.

The spring scale is probably the least expensive device for making mass measurements. The force of gravity is once again used as the reference. The scale is placed so that the unknown object is suspended by the spring and the force of gravity can freely act on the object. The elasticity of the spring is assumed to be linear and the force required to stretch it is marked on the scale. When the force of gravity and the elastic force of the spring balance, the force is read from the scale, which has been calibrated in units of mass. Capacity can be increased by increasing the strength of the spring. However, as the strength of the spring increases, the resolution of the scale decreases. This decrease in resolution limits spring scales to relatively small loads of no more than a few kilograms. There are two kinds of springs used: spiral and cantilevered springs.

The torsion balance is a precise adaptation of the spring concept used to determine the mass indirectly. The vertical force produced by the load produces a torque on a wire or beam. This torque produces an angular deflection. As long as the balance is operated in the linear range, the angular deflection is proportional to the torque. Therefore, the angular deflection is also proportional to the applied load. When the torsion spring constant is calibrated, the angular deflection can be read as a mass unit. Unlike the crude spring scales, it is possible to make torsion balances capable of measuring in the microgram region. The torsion element could be a band, a wire, or a string.

The beam balance is probably the next step in accuracy and cost. The beam balance uses the same principle of operation as the equal arm balance. However, it normally has only one pan and the beam is offset. A set of sliding weights are mounted on the beam. As the weights slide out the beam, they gain a mechanical advantage due to the inequality of the distance from the pivot point of the balance. The weights move out along the beam until the balance is in equilibrium. Along the beam, there are notched positions that are marked to correspond to the force applied by the sliding weights. By adding up the forces indicated by the position of each weight, the mass of the unknown material is determined. Beam balances and scales are available in a wide range of accuracy's load capacities. Beam balances are available to measure in the milligram range and large beam scales are made with capacities to weigh trucks and trains. Once again, the disadvantage is that as load increases the resolution decreases. Figure 20.2 shows an example of a two pan beam balance.

The next progression in cost and accuracy is the strain gage load cell. A strain gage is an electrically resistive wire element that changes resistance when the length of the wire element changes. The gage is bonded to a steel cylinder that will shorten when compressed or lengthen when stretched. Because the gage is bonded to the cylinder, the length of the wire will lengthen or contract with the cylinder. The electrical resistance is proportional to the length of the wire element of the gage. By measuring the resistance of the strain gage, it is possible to determine the load on the load cell. The electric resistance is converted into a mass unit readout by the electric circuitry in the readout device.

The force restorative load cell is the heart of an electronic balance, shown in Figure 20.3. The force restorative load cell uses the principle of the equal arm balance. However, in most cases, the fulcrum is offset so it is no longer an equal arm balance, as one side is designed to have a mechanical advantage. This side of the balance is attached to an electric coil. The coil is suspended in a magnetic field. The other side is still connected to a weighing pan. Attached to the beam is a null indicating device, consisting of a photodiode and a light-emitting diode (LED) that are used to determine when the balance is in equilibrium. When a load is placed on the weighing pan, the balance goes out of equilibrium. The LED photodiode circuit detects that the balance is no longer in equilibrium, and the electric current in the coil is increased to bring the balance back to equilibrium. The electric current is then measured across a precision sense resistor and converted into a mass unit reading and displayed on the digital readout.
A variation of the latter is the new generation of industrial scales, laboratory balances, and mass comparators. Mass comparators are no longer called balances because they always perform a comparison between known masses (standards) and unknown masses. These weighing devices are employing the electromagnetic force compensation principle in conjunction with joint flexures elements replacing the
traditional knife-edge joints. Some of the advantages include a computer interfacing capability and a maintenance-free feature because there are no moving parts.

Another measuring method used in the weighing technology is the vibrating cord. A wire or cord of known length, which vibrates transversely, is tensioned by the force $F$. The vibration frequency changes in direct proportion to the load $F$. The piezoelectric effect is also used in weighing technology. Such weighing devices consist of the presence of an electric voltage at the surface of a crystal when the crystal is under load. Balances employing the gyroscopic effect are also used. This measuring device uses the output signal of a gyrodynamic cell similar to the frequency. Balances wherein the weight force of the load changes the reference distance of the capacitive or inductive converters are also known. As well, balances using the radioactivity changes of a body as a function of its mass under certain conditions exist.

20.2 Weighing Techniques

When relatively low orders of accuracy are required, reading mass or weight values directly from the weighing instrument are adequate. Except for the equal arm balance and some torsion balances, most modern weighing instruments have direct readout capability. For most commercial transactions and for simple scientific experiments, this direct readout will provide acceptable results.

In the case of equal arm balances, the balance will have a pointer and a scale. When relatively low accuracy is needed, the pointer and scale are used to indicate when the balance is close to equilibrium. The same is true when using a torsion balance. However, the equal arm balances of smaller (e.g., 30 g) or larger (e.g., 900 kg) capacity are also used for high-accuracy applications. Only the new generation of electronic balances are equal or better in terms of accuracy and benefit from other features.

Weighing is a deceptively simple process. Most people have been making and using weighing measurements for most of their lives. We have all gone to the market and purchased food that is priced by weight. We have weighed ourselves many times, and most of us have made weight or mass measurements in school. What could be simpler? One places an object on the weighing pan or platform and reads the result.

Unfortunately, the weighing process is very susceptible to error. There are errors caused by imperfections in the weighing instrument; errors caused by biases in the standards used; errors caused by the weighing method; errors caused by the operator; and errors caused by environmental factors. In the case of the equal arm balance, any difference between the lengths of the arms will result in a bias in the measurement result. Nearly all weighing devices will have some degree of error caused by small amounts of nonlinearity in the device. All standards have some amount of bias and uncertainty. Mass is the only base quantity in the International System of Units (SI) defined in relation with a physical artifact. The international prototype of the kilogram is kept at Sevres in France, under the custody of the International Bureau of Weights and Measures. All weighing measurements originate from this international standard. The international prototype of the kilogram is, by international agreement, exact; however, over the last century, it has changed in value. What one does not know is the exact magnitude or direction of the change. Finally, environmental factors such as temperature, barometric pressure, and humidity can affect the weighing process.

There are many weighing techniques used to reduce the errors in the weighing process. The simplest technique is substitution weighing. The substitution technique is used to eliminate some of the errors introduced by the weighing device. The single-substitution technique is one where a known standard and an unknown object are both weighed on the same device. The weighing device is only used to determine the difference between the standard and the unknown. First, the standard is weighed and the weighing device’s indication is noted. (In the case of an equal arm balance, tare weights are added to the second pan to bring the balance to equilibrium.) The standard is then removed from the weighing device and the unknown is placed in the same position. Again, the weighing device’s indication is noted. The first noted indication is subtracted from the second indication. This gives the difference between the standard and the unknown. The difference is then added to the known value of the standard to calculate the value of the unknown object. A variation of this technique is to use a small weight of known value.
to offset the weighing device by a small amount. The amount of offset is then divided by the known value of the small weight to calibrate the readout of the weighing device. The weighing results of this measurement is calculated as follows:

$$U = S + \left( O_2 - O_1 \right) \left( \frac{SW}{O_3 - O_1} \right)$$  \hspace{1cm} (20.4)

where

- \( U \) = value of the unknown
- \( S \) = known value of the standard
- \( SW \) = small sensitivity weight used to calibrate the scale divisions
- \( O_1 \) = first observation (standard)
- \( O_2 \) = second observation (unknown)
- \( O_3 \) = third observation (unknown + SW)

These techniques remove most of the errors introduced by the weighing device, and are adequate when results to a few tenths of a gram are considered acceptable.

If results better than a few tenths of a gram are required, environmental factors begin to cause significant errors in the weighing process. Differences in density between the standard and the unknown object and air density combine together to cause significant errors in the weighing process.

It is the buoyant force that generates the confusion in weighing. What is called the "true mass" of an object is the mass determined in vacuum. The terms "true mass" and "mass in vacuum" are referring to the same notion of inertial mass or mass in the Newtonian sense. In practical life, the measurements are performed in the surrounding air environment. Therefore, the objects participating in the measurement process adhere to the Archimedean principle being lifted with a force equal to the weight of the displaced volume of air. Applying the buoyancy correction to the measurement requires the introduction of the term “apparent mass.” The “apparent mass” of an object is defined in terms of “normal temperature” and “normal air density,” conventionally chosen as 20°C and 1.2 mg cm⁻³, respectively. Because of these conventional values, the “apparent mass” is also called the “conventional mass.” The reference material is either brass (8.4 g cm⁻³) or stainless steel (8.0 g cm⁻³), for which one obtains an “apparent mass versus brass” and an “apparent mass versus stainless steel,” respectively. The latter is preferred for reporting the “apparent mass” of an object.

Calibration reports from the National Institute of Standards and Technology will report mass in three ways: True Mass, Apparent Mass versus Brass, and Apparent Mass versus Stainless Steel. Conventional mass is defined as the mass of an object with a density of 8.0 g cm⁻³, at 20°C, in air with a density of 1.2 mg cm⁻³. However, most scientific weighings are of materials with densities that are different from 8.0 g cm⁻³. This results in significant measurement errors.

As an example, use the case of a chemist weighing 1 liter of water. The chemist will first weigh a mass standard, a 1 kg weight made of stainless steel; then the chemist will weigh the water. The 1 kg mass standard made of 8.0 g cm⁻³ stainless steel will have a volume of 125 cm³. The same mass of water will have a volume approximately equal to 1000 cm³ (Volume = Mass/Density). The mass standard will displace 125 cm³ of air, which will exert a buoyant force of 150 mg (125 cm³ × 1.2 mg cm⁻³). However, the water will displace 1000 cm³ air, which will exert a buoyant force of 1200 mg (1000 cm³ × 1.2 mg cm⁻³). Thus, the chemist has introduced a significant error into the measurement by not taking the differing densities and air buoyancy into consideration.

Using 1.2 mg cm⁻³ for the density of air is adequate for measurements made close to sea level; it must be noted that air density decreases with altitude. For example, the air density in Denver, CO, is approximately 0.98 mg cm⁻³. Therefore, to make accurate mass measurements, one must measure the air density at the time of the measurement if environmental errors in the measurement are to be reduced.

Air density can be calculated to an acceptable value using the following equations:

$$\rho_A \equiv 0.0034848 \left( t + 273.15 \right) \left( P - 0.0037960 \times U \times e \right)$$  \hspace{1cm} (20.5)
where $\rho_a = \text{air density in mg cm}^{-3}$

$\rho = \text{density of the standard}$

$\rho_u = \text{density of the unknown}$

$\rho_{sw} = \text{density of the sensitivity weight}$

$O_1 = \text{first observation (standard)}$

$O_2 = \text{second observation (unknown)}$

$O_3 = \text{third observation (unknown + SW)}$

$CM = \text{conventional mass}$

$M_u = \text{mass of the unknown in a vacuum}$

$M_s = \text{mass of the standard in a vacuum}$

$M_{sw} = \text{mass of the sensitivity weight}$

$\rho_a = \text{air density}$

$\rho_s = \text{density of the standard}$

$\rho_u = \text{density of the unknown}$

$\rho_{sw} = \text{density of the sensitivity weight}$

$\rho_t = \text{temperature in °C}$

$\rho_{atm} = \text{barometric pressure in pascals}$

$\rho_s = \text{density of the standard}$

$\rho_u = \text{density of the unknown}$

$\rho_{sw} = \text{density of the sensitivity weight}$

$O_1 = \text{first observation (standard)}$

$O_2 = \text{second observation (unknown)}$

$O_3 = \text{third observation (unknown + SW)}$

$CM = M_u \left(1 - 0.0012/\rho_u\right) / 0.99985$ (20.8)

To apply an air buoyancy correction to the single substitution technique, use the following formulae:

$M_u = \left(M_s \left(1 - \rho_A/\rho_s\right) + (O_2 - O_1) \left(M_{sw} \left(1 - \rho_A/\rho_{sw}\right)/\left(O_3 - O_2\right)\right)\right)/\left(1 - \rho_A/\rho_u\right)$ (20.7)

$CM = M_u \left(1 - 0.0012/\rho_u\right) / 0.99985$ (20.8)

When very precise measurements are needed, the double-substitution technique coupled with an air buoyancy correction will provide acceptable results for nearly all scientific applications. The double-substitution technique is similar to the single-substitution technique using the sensitivity weight. The main advantage of this technique over single substitution is that any drift in the weighing device is accounted for in the technique. Because of the precision of this weighing technique, it is only appropriate to use it on precision balances or mass comparators. As in the case of single substitution, one places the standard on the balance pan and takes a reading. The standard is then removed and the unknown object is placed on the balance pan and a second reading is taken. The third step is to add the small sensitivity weight to the pan with the unknown object and take a third reading. Then remove the unknown object and return the standard to the pan with the sensitivity weight and take a fourth reading. The mass is calculated using the following formulae:

$M_u = \left(M_s \left(1 - \rho_A/\rho_s\right) + (O_2 - O_1) \left(M_{sw} \left(1 - \rho_A/\rho_{sw}\right)/\left(O_3 - O_2\right)\right)\right)/\left(1 - \rho_A/\rho_u\right)$ (20.9)
where $M_u =$ mass of the unknown (in a vacuum)
$M_s =$ mass of the standard (in a vacuum)
$M_{sw} =$ mass of the sensitivity weight
$\rho_A =$ air density
$\rho_s =$ density of the standard
$\rho_u =$ density of the unknown
$\rho_{sw} =$ density of the sensitivity weight
$O_1 =$ first observation (standard)
$O_2 =$ second observation (unknown)
$O_3 =$ third observation (unknown + sensitivity weight)
$O_4 =$ fourth observation (standard + sensitivity weight)

<br><br>

$$CM = M_u \left( 1 - \frac{0.0012}{\rho_u} \right) / 0.99985$$  \hspace{1cm} (20.10)

where CM = conventional mass
$M_u =$ mass of the unknown in a vacuum
$\rho_u =$ density of the unknown

To achieve the highest levels of accuracy, advanced weighing designs have been developed. These advanced weighing designs incorporate redundant weighing, drift compensation, statistical checks, and multiple standards. The simplest of these designs is the three-in-one design. It uses two standards to calibrate one unknown weight. In its simplest form, one would perform three double substitutions. The first compares the first standard and the unknown weight; the second double substitution compares the first standard against the second standard, which is called the check standard; and the third and final comparison compares the second (or check standard) against the unknown weight. These comparisons would then result in the following:

$O_1 =$ reading with standard on the balance
$O_2 =$ reading with unknown on the balance
$O_3 =$ reading with unknown and sensitivity weight on the balance
$O_4 =$ reading with standard and sensitivity weight on the balance
$O_5 =$ reading with check standard on the balance
$O_6 =$ reading with check standard and sensitivity weight on the balance
$O_9 =$ reading with check standard on the balance
$O_{10} =$ reading with unknown on the balance
$O_{11} =$ reading with unknown and sensitivity weight on the balance
$O_{12} =$ reading with check standard and sensitivity weight on the balance

The measured differences are calculated using the following formulae:

$$a = \left[ \left( O_1 - O_2 + O_4 - O_3 \right) / 2 \right] \times \left[ M_{sw} \left( 1 - \rho_A / \rho_{sw} \right) \right] / \left( O_3 - O_2 \right)$$  \hspace{1cm} (20.11)

$$b = \left[ \left( O_3 - O_4 + O_6 - O_5 \right) / 2 \right] \times \left[ M_{sw} \left( 1 - \rho_A / \rho_{sw} \right) \right] / \left( O_4 - O_6 \right)$$  \hspace{1cm} (20.12)

$$c = \left[ \left( O_5 - O_9 + O_{12} - O_{11} \right) / 2 \right] \times \left[ M_{sw} \left( 1 - \rho_A / \rho_{sw} \right) \right] / \left( O_{11} - O_9 \right)$$  \hspace{1cm} (20.13)
where $a = \text{difference between standard and unknown}$
$b = \text{difference between standard and check standard}$
$c = \text{difference between check standard and unknown}$

$M_{sw} = \text{mass of sensitivity weight}$
$\rho_A = \text{air density calculated using Equations 20.5 and 20.6}$
$\rho_{sw} = \text{density of sensitivity weight}$

The least-squares measured difference is computed for the unknown from:

$$d_u = \left(-2a - b - c\right)/3 \quad (20.14)$$

Using the least-squares measured difference, the mass of the unknown is computed as:

$$U = \left[S\left(1 - \rho_A/\rho_s\right) + d_u\right]/\left[1 - \rho_A/\rho_U\right] \quad (20.15)$$

where $U = \text{mass of unknown}$

$S = \text{mass of the standard}$

$d_u = \text{least-squares measured difference of the unknown}$

$\rho_A = \text{air density calculated using Equations 20.5 and 20.6}$

$\rho_s = \text{density of the standard}$

$\rho_U = \text{density of the unknown}$

The conventional mass of the unknown is now calculated as:

$$CU = U \left(1 - 0.0012/\rho_U\right)/0.99985 \quad (20.16)$$

where $CU = \text{conventional mass}$

$U = \text{mass of unknown}$

$\rho_U = \text{density of unknown}$

The least-squares measured difference is now computed for the check standard as:

$$d_{CS} = \left(-a - 2b - c\right)/3 \quad (20.17)$$

Using the least-squares measured difference, the mass of the check standard is computed from:

$$CS = \left[S\left(1 - \rho_A/\rho_s\right) + d_{CS}\right]/\left[1 - \rho_A/\rho_{CS}\right] \quad (20.18)$$

where $CS = \text{mass of check standard}$

$s = \text{mass of the standard}$

$d_{CS} = \text{least-squares measured difference of the check standard}$

$\rho_A = \text{air density calculated using Equations 20.5 and 20.6}$

$\rho_s = \text{density of the standard}$

$\rho_{CS} = \text{density of unknown}$

The mass of the check standard must lie within the control limits for the check standard. If it is out of the control limits, the measurement must be repeated.

The short-term standard deviation of the process is now computed:

$$\text{Short-term standard deviation} = 0.577\left(a - b + c\right) \quad (20.19)$$
The short-term standard deviation is divided by the historical pooled short-time standard deviation to calculate the $F$-statistic:

$$F\text{-statistic} = \frac{\text{short-term standard deviation}}{\text{pooled short-time standard deviation}}$$

The $F$-statistic must be less than the value obtained from the student $t$-variant at the 99% confidence level for the number of degrees of freedom of the historical pooled standard deviation. If this test fails, the measurement is considered to be out of statistical control and must be repeated.

By measuring a check standard and by computing the short-term standard deviation of the process and comparing them to historical results, one obtains a high level of confidence in the computed value of the unknown.

There are many different weighing designs that are valid; the three-in-one (three equal weights) and four equal weights are the ones that can be easily calculated without the use of a computer. Primary calibration laboratories — private and government — are using these multiple intercomparisons, state-of-the-art mass calibration methods under the Mass Measurement Assurance Program using the Mass Code computer program provided by the National Institute of Standards and Technology. A full discussion of these designs can be found in the *National Bureau of Standards Technical Note 952*.

References

Density Measurement

21.1 Solid Density
21.2 Fluid Density

Pycnometric Densitometers • Buoyancy Type Densitometers • Hydrometers • Hydrostatic Weighing Densitometers • Balance-Type Densitometers • Column-Type Densitometers • Vibrating Element Densitometers • Radioactive Densitometers • Refractometer and Index of Refraction Densitometers • Coriolis Densitometers • Absorption-Type Densitometers

Density is a significant part of measurement and instrumentation. Density measurements are made for at least two important reasons: (1) for the determination of mass and volume of products, and (2) the quality of the product. In many industrial applications, density measurement ascertains the value of the product.

Density is defined as the mass of a given volume of a substance under fixed conditions. However, ultimate care must be exercised in measurements because density varies with pressure and temperature. The variation is much greater in gases.

In many modern applications, the densities of products are obtained by sampling techniques. In measurements, there are two basic concepts: static density measurements and dynamic (on-line) density measurements. Within each concept, there are many different methods employed. These methods are based on different physical principles. In many cases, the application itself and the characteristics of the process determine the best suitable method to be used. Generally, static methods are well developed, lower in cost, and more accurate. Dynamic samplers are expensive, highly automated, and use microprocessor-based signal processing devices. Nevertheless, nowadays, many static methods are also computerized, offering easy to use, flexible, and self-calibrating features.

There is no single universally applicable density measurement technique. Different methods must be employed for different products and materials. In many cases, density is normalized under reference conditions.

The density of a substance is determined by dividing the density of that substance by the density of a standard substance obtained under the same conditions. This dimensionless ratio is called the specific gravity (SG), also termed the relative density. The specific gravities of liquid and gases under reference conditions are given by:

\[
\text{Liquid SG} = \frac{\text{density of liquid}}{\text{density of water}} \tag{21.1}
\]

\[
\text{Gas SG} = \frac{\text{density of gas}}{\text{density of air}} \tag{21.2}
\]
Commonly accepted sets of conditions are normal temperature and pressure (NTP) and standard temperature and pressure (STP). NTP is usually taken as the temperature of 0.00°C and a pressure of 760 mm Hg. The NTP is accepted as 15.00 or 15.56°C and 101.325 kPa.

21.1 Solid Density

If the uniformity is maintained, the determination of density of solids is a simple task. Once the volume of the solid and its mass are known, the density can be found using the basic ratio: density = mass/volume (kg m⁻³).

However, in many applications, solids have different constituents and are made up from different materials having different ratios. Their volumes can also change often. In these cases, dynamic methods are employed, such as radioactive absorption types, ultrasonic, and other techniques. Some of these methods are described below.

21.2 Fluid Density

The measurement of densities of fluids is much more complex than for solids. For fluid densities, many different techniques are available. This is mainly due to complexities in processes, variations of fluid densities during the processes, and diverse characteristics of the process and the fluids themselves. Some of these methods are custom designed and applicable to special cases only. Others are very similar in principles and technology, and applicable to many different type of fluids. At present, apart from conventional methods, there are many novel and unusual techniques undergoing extensive development and research. For example, densitometers based on electromagnetic principles [1] can be given as part of an intelligent instrumentation system.

Fluids can be divided to liquids and gases. Extra care and further considerations are necessary in gas density measurements. For example, perfect gases contain an equal number of molecules under the same conditions and volumes. Therefore, molecular weights can be used in density measurements.

Depending on the application, fluid densities can be measured both in static and dynamic forms. In general, static density measurements of fluids are well developed, precise, and have greater resolution than most dynamic techniques. Pycnometers and buoyancy are examples of static techniques that can be adapted to cover small density ranges with a high resolution and precision. Nowadays, many manufacturers offer dynamic instruments previously known to be static. Also, many static density measurement devices are computerized and come with appropriate hardware and software. In general, static-type measurements are employed in laboratory conditions, and dynamic methods are employed for real-time measurements where the properties of a fluid vary from time to time.

In this chapter, the discussion will concentrate on the commonly applied, modern density measuring devices. These devices include:

1. Pycnometric densitometers
2. Buoyancy-type densitometers
3. Hydrometers
4. Hydrostatic weighing densitometers
5. Balance-type densitometers
6. Column-type densitometers
7. Vibrating element densitometers
8. Radioactive densitometers
9. Refractometer and index of reflection densitometers
10. Coriolis densitometers
11. Absorption-type densitometers
Pycnometric Densitometers

Pycnometers are static devices. They are manufactured as fixed volume vessels that can be filled with the sample liquid. The density of the fluid is measured by weighing the sample. The simplest version consists of a vessel in the shape of a bottle with a long stopper containing a capillary hole, as shown in Figure 21.1. The capillary is used to determine the exact volume of the liquid, thus giving high resolution when filling the pycnometer. The bottle is first weighed empty, and then with distilled-aerated water to determine the volume of the bottle. The bottle is then filled with the process fluid and weighed again. The density is determined by dividing the mass by the volume. The specific gravity of the liquid is found by the ratio of the fluid mass to water mass. When pycnometers are used, for good precision, ultimate care must be exercised during the measurements; that is, the bottle must be cleaned after each measurement, the temperature must be kept constant, and precision balances must be used. In some cases, to ensure filling of the pycnometer, twin capillary tubes are used. The two capillaries, made of glass, are positioned such that the fluid can be driven into the vessel under vacuum conditions. Accurate filling to graduation marks on the capillary is then made.

The pycnometers have to be lightweight, strong enough to contain samples, and they need to be nonmagnetic for accurate weighing to eliminate possible ambient magnetic effects. Very high-resolution balances must be used to detect small differences in weights of gases and liquids. Although many pycnometers are made of glass, they are also made of metals to give enough strength for the density measurements of gases and liquids at extremely high pressures. In many cases, metal pycnometers are necessary for taking samples from the line of some rugged processes.

Pycnometers have advantages and disadvantages. Advantages are that if used correctly, they are accurate; and they can be used for both density and specific gravity measurements. The disadvantages include:

1. Great care must be exercised for accurate results.
2. The sample must be taken off-line, with consequent time lag in results. This creates problems of relating samples to the materials that exist in the actual process.
3. High-precision pycnometers are expensive. They require precision weighing scales and controlled laboratory conditions. Specialized techniques must be employed to take samples in high-pressure processes and hostile conditions, such as offshore installations.
4. Their good performances might depend on the skill of operator.

Buoyancy-Type Densitometers

The buoyancy method basically uses Archimedes principle. A suspended sinker, with a known mass and volume attached to a fine wire, is totally immersed in the sample liquid. A precision force balance is used to measure the force to support the sinker. Once the mass, volume, and supporting weight of the sinker
are known, the density of the liquid can be calculated. However, some corrections need to be made for
surface tension on the suspension wire, the cubicle expansion coefficient of the sinker, and the temper-
ature of process. Buoyancy-type densitometers give accurate results and are used for the calibration of
the other liquid density transducers.

One advanced version of the buoyancy technique is the magnetic suspension system. The sinker is
fully enclosed in a pressure vessel, thus eliminating surface tension errors. Their uses can also be extended
to applications such as the specific gravity measurements under low vapor pressures and density mea-
measurements of hazardous fluids.

Hydrometers

Hydrometers are the most commonly used devices for measurement of the density of liquids. They are
so commonly used that their specifications and procedure of use are described by national and interna-
tional standards, such as ISO 387. The buoyancy principle is used as the main technique of operation.
The volume of fixed mass is converted to a linear distance by a sealed bulb-shaped glass tube containing
a long stem measurement scale, shown in Figure 21.2. The bulb is ballasted with a lead shot and pitch,
the mass of which depends on the density range of the liquid to be measured. The bulb is simply placed
into the liquid and the density is read from the scale. The scale is graduated in density units such as
kg m⁻³. However, many alternative scales are offered by manufacturers, such as specific gravity, API gravity,
Brix, Brine, etc. Hydrometers can be calibrated for different ranges for surface tensions and temperatures.
Temperature corrections can be made for set temperature such as 15°C, 20°C, or 25°C. ISO 387 covers
a density range of 600 kg m⁻³ to 2000 kg m⁻³. Hydrometers have a number of advantages and disadvan-
tages. The advantages include:

1. Relatively low cost and easy to use
2. Good resolution for small range
3. Traceable to national and international standards

The disadvantages include:

1. They have small span; therefore, a number of meters are required to cover a significant range.
2. They are made from glass and fragile. Metal and plastic versions are not as accurate.
3. The fluid needs to be an off-line sample, not representing the exact conditions of the process.
   There are pressure hydrometers for low vapor pressure hydrocarbons, but this adds a need for
   accurately determining the pressure too.
4. If good precision is required, they are difficult to use, needing surface tension and temperature
   corrections. Further corrections could be required for opaque fluids.
Hydrostatic Weighing Densitometers

The most common device using a hydrostatic weighing method consists of a U-tube that is pivoted on flexible end couplings. A typical example is shown in Figure 21.3. The total weight of the tube changes, depending on the density of fluid flowing through it. The change in the weight needs to be measured accurately, and there are a number of methods employed to do this. The most common commercial meters use a force balance system. The connectors are stainless steel bellows. In some cases, rubber or PTFE are used, depending on the process fluid characteristics and the accuracy required. There are temperature and pressure limitations due to bellows, and the structure of the system may lead to a reading offset. The meter must be securely mounted on a horizontal plane for optimal accuracy.

The advantages of this method include:

1. They give continuous reading and can be calibrated accurately.
2. They are rugged and can be used for two-phase liquids such as slurries, sugar solutions, powders, etc.

The disadvantages of these meters include:

1. They must be installed horizontally on a solid base. These meters are not flexible enough to adapt to any process; thus, the process must be designed around it.
2. They are bulky and cumbersome to use.
3. They are unsuitable for gas density measurements.

Balance-Type Densitometers

Balance-type densitometers are suitable for liquid and gas density measurements. Manufacturers offer many different types; four of the most commonly used ones are discussed below.

Balanced-Flow Vessel

A fixed volume vessel as shown in Figure 21.4 is employed for the measurements. While the liquid is flowing continuously through the vessel, it is weighed automatically by a sensitive scale — a spring balance system or a pneumatic force balance transmitter. Because the volume and the weight of the liquid are known, the density or specific gravity can easily be calculated and scaled in respective units. In the design process, extra care must be exercised for the flexible end connections.
Chain Balanced Float

In this system, a self-centering, fixed-volume, submerged plummet is used for density measurements, as illustrated in Figure 21.5. The plummet is located entirely under the liquid surface. At balance, the plummet operates without friction and is not affected by surface contamination. Under steady-state conditions, the plummet assumes a stable position. The effective weight of the chain on the plummet varies, depending on the position of the plummet, which in turn is a function of density. The plummet contains a metallic transformer core that transmits changes in the position to be measured by a pickup coil. The voltage differential, a function of plummet displacement, is calibrated as a measure of variations in specific gravity. A resistance thermometer bridge is used for the compensation of temperature effects on density.

Gas Specific Gravity Balance

A tall column of gas is weighed by the floating bottom of the vessel. This weight is translated into the motion of an indicating pointer, which moves over a scale graduated in units of density or specific gravity. This method can be employed for any gas density measurement.
Buoyancy Gas Balance

In this instrument, a displacer is mounted on a balance beam in a vessel, as shown in Figure 21.6. The displacer is balanced for air, and the manometer reading is noted at the exact balance pressure. The air is then displaced by gas of the same pressure. The difference in the reading of the balance beam gives the SG of the gas. The pressures are read on the manometer.

Column-Type Densitometers

There are number of different versions of column methods. As a typical example, a reference column method is illustrated in Figure 21.7. A known head of sample liquid and water from the respective bubbler pipes are used. A differential pressure measuring device compares the pressure differences, proportional to relative densities of the liquid and the water. By varying the depth of immersion of the pipes, a wide
Both columns must be maintained at the same temperature to avoid the necessity for corrections of temperature effects.

A simpler and more widely used method of density measurement is achieved by the installation of two bubbler tubes as illustrated in Figure 21.8. The tubes are located in the sample fluid such that the end of one tube is higher than that of the other. The pressure required to bubble air into the fluid from both tubes is equal to the pressure of the fluid at the end of the bubbler tubes. The outlet of one tube higher than the other and the distances of the openings of the tubes are fixed; hence, the difference in the pressure is the same as the weight of a column of liquid between the ends. Therefore, the differential pressure measurement is equivalent to the weight of the constant volume of the liquid, and calibrations can be made that have a direct relationship to the density of the liquid. This method is accurate to within 0.1% to 1% specific gravity. It must be used with liquids that do not crystallize or settle in the measuring chamber during measurements.

Another version is the range suppression type, which has an additional constant pressure drop chamber as shown in Figure 21.9. This chamber is in series with the low-pressure side to give advantages in scaling and accurate readings of densities.
Vibrating Element Densitometers

If a body containing or surrounded by a fluid is set to resonance at its natural frequency, then the frequency of oscillation of the body will vary as the fluid properties and conditions change. The natural frequency is directly proportional to the stiffness of the body and inversely proportional to the combined mass of the body and the fluid. It is also dependent on the shape, size, and elasticity of the material, induced stress, mass, and mass distribution of the body. Basically, the vibration of the body can be equated to motion of a mass attached to a mechanical spring. Hence, an expression for the frequency can be written as:

\[
\text{Resonant frequency} = \sqrt{\frac{K}{M + k\rho}}
\]  

(21.3)

where \(K\) is the system stiffness, \(M\) is the transducer mass, \(k\) is the system constant, and \(\rho\) is the fluid density.

A factor common to all types of vibrating element densitometers is the problem of setting the element in vibration and maintaining its natural resonance. There are two drives for the purpose.

Magnetic Drives

Magnetic drives of the vibrating element and the pickup sensors of vibrations are usually achieved using small coil assemblies. Signals picked up by the sensors are amplified and fed back as a drive to maintain the disturbing forces on the vibrating body of the meter.

In order to achieve steady drives, the vibrating element sensor can be made from nonmagnetic materials. In this case, small magnetic armatures are attached.

The main advantage of magnetic drive and pickup systems is they are noncontact methods. They use conventional copper windings and they are reliable within the temperature range of –200 to +200°C.

Piezoelectric Drives

A wide range of piezoelectric materials are available to meet the requirements of driving vibrating elements. These materials demonstrate good temperature characteristics as do magnetic drive types. They also have the advantage of being low in cost. They have high impedance, making the signal conditioning circuitry relatively easy. They do not require magnetic sensors.

The piezoelectric drives are mechanically fixed on the vibrating element by adhesives. Therefore, attention must be paid to the careful placement of the mount in order to reduce the strain experienced by the piezo element due to thermal and pressure stresses while the instrument is in service.

A number of different types of densitometers have been developed that utilize this phenomenon. The three main commercial types are introduced here.

Vibrating Tube Densitometers

These devices are suitable for highly viscous liquids or slurry applications. The mode of operation of vibration tube meters is based on the transverse vibration of tubes as shown in Figure 21.10. The tube and the driving mechanisms are constrained to vibrate on a single plane. As the liquid moves inside the tube, the density of the entire mass of the liquid is measured. The tube length is approximately 20 times greater than the tube diameter.

A major design problem with the vibrating tube method is the conflict to limit the vibrating element to a finite length and accurately fix the nodes. Special attention must be paid to avoid any exchange of vibrational energy outside the sensory tube. The single tube has the disadvantage of presenting obstruction to the flow, thus experiencing some pressure losses. The twin tube, on the other hand, offers very small blockage (Figure 21.11) and can easily be inspected and cleaned. Its compact size is another distinct advantage. In some densitometers, the twin tube is designed to achieve a good dynamic balance, with the two tubes vibrating in antiphase. Their nodes are fixed at the ends, demonstrating maximum sensitivity to installation defects, clamping, and mass loading.
The main design problems of the vibrating tube sensors are in minimizing the influence of end padding and overcoming the effects of pressure and temperature. Bellows are used at both ends of the sensor tubes to isolate the sensors from external vibrations. Bellows also minimize the end loadings due to differential expansions and installation stresses.

The fluid runs through the tubes; therefore, no pressure balance is required. Nevertheless, in some applications, the pressure stresses the tubes, resulting in stiffness changes. Some manufacturers modify the tubes to minimize the pressure effects. In these cases, corrections are necessary only when high accuracy is mandatory. The changes in the Young’s modulus with temperature can be reduced to near-zero using Ni-span-C materials whenever corrosive properties of fluids permit. Usually, manufacturers provide pressure and temperature correction coefficients for their products.

It is customary to calibrate each vibration element densitometer against others as a transfer of standards. Often, the buoyancy method is used for calibration purposes. The temperature and pressure coefficients are normally found by exercising the transducer over a range of temperatures and pressures on a liquid with well-known properties. Prior to calibration, the vibration tube densitometers are subjected to a programmed burn-in cycle to stabilize them against temperatures and pressures.

**Vibrating Cylinder Densitometers**

A thin-walled cylinder, with a 3:1 length:diameter ratio, is fixed with stiff ends. The thickness of the cylinder wall varies from 25 μm to 300 μm, depending on the density range and type of fluid used. The cylinder can be excited to vibrate in a hoop mode by magnetic drives mounted either in or outside the cylinder.

For good magnetic properties, the cylinder is made of corrosion-resistant magnetic materials. Steel such as FV520 is often used for this purpose. Such materials have good corrosion-resistance characteristics; unfortunately, due to their poor thermoeelastic properties, they need extensive temperature corrections.
Nickel-iron alloys such as Ni-span-C are often used to avoid temperature effects. Once correctly treated, the Ni-span-C alloy has near-zero Young’s modulus properties. Because the cylinder is completely immersed in the fluid, there are no pressure coefficients.

The change in the resonant frequency is determined by the local mass loading of the fluid in contact with the cylinder. The curve of frequency against density is nonlinear and has a parabolic shape, thus requiring linearization to obtain practical outputs. The resonant frequency range varies from 2 kHz to 5 kHz, depending on the density range of the instrument. The cylinders need precision manufacturing and thus are very expensive to construct. Each meter needs to be calibrated individually for different temperatures and densities to suit specific applications. In the case of gas density applications, gases with well-known properties (e.g., pure argon or nitrogen) are used for calibrations. In this case, the meters are subjected to a gas environment with controlled temperature and pressure. The calibration curves are achieved by repetitions to suit the requirements of individual customers for particular applications. In the case of liquids, the meters are calibrated with liquids of known density, or they are calibrated against another standard (e.g., pycnometer or buoyancy type densitometers).

Vibration cylinder-type densitometers have zero pressure coefficients and they are ideal for liquefied gas products or refined liquids. Due to relatively small clearances between cylinder and housing, they require regular cleaning. They are not suitable for liquids or slurries with high viscous properties.

**Tuning Fork Densitometers**

These densitometers make use of the natural frequency of low-mass tuning forks, shown in Figure 21.12. In some cases, the fluid is taken into a small chamber in which the electromechanically driven forks are situated. In other cases, the fork is inserted directly into the liquid. Calibration is necessary in each application.

The advantages of vibrating element meters include:

1. They are suitable for both liquids and gases with reasonable accuracy.
2. They can be designed for real-time measurements.
3. They can easily be interfaced because they operate on frequencies and are inherently digital.
4. They are relatively robust and easy to install.
5. Programmable and computerized versions are available. Programmed versions make all the corrections automatically. They provide the output of live density, density at reference conditions, relative density, specific gravity, concentration, solid contents, etc.

The disadvantages include:

1. They do not relate directly to primary measurements; therefore, they must be calibrated.
2. They all have problems in measuring multiphase liquids.

**Radioactive Densitometers**

As radioactive isotopes decay, they emit radiation in the form of particles or waves. This physical phenomenon can be used for the purposes of density measurement. For example, γ rays are passed...
through the samples and their rate of arrivals are measured using ion- or scintillation-based detection [2]. Generally, γ-ray mass absorption rate is independent of material composition; hence they can be programmed for a wide range materials. Densitometers based on radiation methods can provide accuracy up to +0.0001 g mL⁻¹. Many of these devices have self-diagnostic capabilities and are able to compensate for drift caused by source decay, thus pinpointing any signaling problems.

If γ rays of intensity $J_0$ penetrate a material of a density $\rho$ and thickness $d$ then the intensity of the radiation after passing through the material can be expressed by:

$$J = J_0 \exp\left(n \rho d\right)$$

(21.4)

where $n$ is the mass absorption coefficient.

The accuracy of the density measurement depends on the accuracy of the measurement of the intensity of the radiation and the path length $d$. A longer path length through the material gives a stronger detection signal.

For accurate operations, there are many arrangements for relative locations of transmitters and detectors, some of which are illustrated in Figures 21.13 and 21.14. Generally, the source is mounted in a lead container clamped onto the pipe or the container wall. In many applications, the detector is also clamped onto the wall.

**FIGURE 21.13** Fixing radioactive densitometer on an enlarged pipe. The pipe is enlarged to give longer beam length through the liquid, and hence better attenuation of the radioactive energy.

**FIGURE 21.14** Fixing radioactive densitometer on an elongated pipe. Elongated path yields a longer path length of the radioactive energy through the liquid, hence, a stronger attenuation.
The advantages of using radioactive methods include:

1. The sensor does not touch the sample; hence, there is no blockage to the path of the liquid.
2. Multiphase liquids can be measured.
3. They come in programmable forms and are easy to interface.
4. They are most suitable in difficult applications, such as mining and heavy process industries.

The disadvantages include:

1. A radioactive source is needed; hence, there is difficulty in handling.
2. For reasonable accuracy, a minimum path length is required.
3. There could be long time constants, making them unsuitable in some applications.
4. They are suitable only for solid and liquid density measurements.

**Refractometer and Index of Refraction Densitometers**

Refractometers are essentially optical instruments operating on the principles of refraction of light traveling in liquid media. Depending on the characteristics of the samples, measurement of refractive index can be made in a variety of ways (e.g., critical angle, collimation, and displacement techniques). Usually, an in-line sensing head is employed, whereby a sensing window (commonly known as a prism) is wetted by the product to be measured. In some versions, the sensing probes must be installed inside the pipelines or in tanks and vessels. They are most effective in reaction-type process applications where blending and mixing of liquids take place. For example, refractometers can measure dissolved soluble solids accurately.

Infrared diodes, lasers, and other lights may be used as sources. However, this measurement technique is not recommended in applications in processes containing suspended solids, high turbidity, entrained air, heavy colors, poor transparency and opacity, or extremely high flow rates. The readings are automatically corrected for variations in process temperature. The processing circuitry can include signal outputs adjustable in both frequency and duration.

Another version of a refractometer is the index of refraction type densitometer. For example, in the case of position-sensitive detectors, the index of refraction of liquid under test is determined by measuring the lateral displacement of a laser beam. When the laser beam impinges on the cell at an angle of incidence, as in Figure 21.15, the axis of the emerging beam is displaced by the cell wall and by the inner liquid. The lateral displacement can accurately be determined by position-sensitive detectors. For maximum sensitivity, the devices need to be calibrated with the help of interferometers.

Refractometers are often used for the control of adulteration of liquids of common use (e.g., edible oils, wines, and gasoline). They also find application in pulp and paper, food and beverage, sugar, dairy, and other chemical industries.

**Coriolis Densitometers**

The Coriolis density metering systems are similar to vibrating tube methods, but with slight variations in the design. They are comprised of a sensor and a signal-processing transmitter. Each sensor consists of one or two flow tubes enclosed in a sensor housing. They are manufactured in various sizes and shapes [3]. The sensor tubes are securely anchored at the fluid inlet and outlet points and force is vibrated at the free end, as shown in Figure 21.16. The sensor operates by applying Newton’s second law of motion \( F = ma \).

Inside the housing, the tubes are vibrated in their natural frequencies using drive coils and a feedback circuit. This resonant frequency of the assembly is a function of the geometry of the element, material of construction, and mass of the tube assembly. The tube mass comprises two parts: the mass of the tube itself and the mass of the fluid inside the tube. The mass of the tube is fixed for a given sensor. The mass of fluid in the tube is equal to the fluid density multiplied by volume. Because the tube volume is constant, the frequency of oscillation can be related directly to the fluid density. Therefore, for a given geometry...
of tube and the material of the construction, the density of the fluid can be determined by measuring
the resonant frequency of vibration. Temperature sensors are used for overcoming the effects of changes
in modulus of elasticity of the tube. The fluid density is calculated using a linear relationship between
the density and the tube period and calibration constants.

Special peripherals, based on microprocessors, are offered by various manufacturers for a variety of
measurements. However, all density peripherals employ the natural frequency of the sensor coupled with
the sensor temperature to calculate on-line density of process fluid. Optional communication, interfacing
facilities, and appropriate software are also offered.

Absorption-Type Densitometers

Absorption techniques are also used for density measurements in specific applications. X-rays, visible
light, UV light, and sonic absorptions are typical examples of this method. Essentially, attenuation and

![Index of refraction-type densitometer. The angle of refraction of the beam depends on the shape,
size, and thickness of the container, and the density of fluid in the container. Because the container has the fixed
characteristics, the position of the beam can be related to density of the fluid. Accurate measurement of the position
of the beam is necessary.](image1)

![Coriolis densitometer. Vibration of the tube is detected and related to the mass and flow rate of the
fluid. Further calibrations and calculations must be made to determine the densities.](image2)
phase shift of a generated beam going through the sample is sensed and related to the density of the sample. Most absorption-type densitometers are custom designed for applications having particular characteristics. Two typical examples are: (1) UV absorption or X-ray absorptions are used for determining the local densities of mercury deposits in arc discharge lamps, and (2) ultrasonic density sensors are used in connection with difficult density measurements (e.g., density measurement of slurries). The lime slurry, for example, is a very difficult material to handle. It has a strong tendency to settle out and coat all equipment with which it comes in contact. An ultrasonic density control sensor can fully be emerged into an agitated slurry, thus avoiding the problems of coating and clogging. Inasmuch as the attenuation of the ultrasonic beam is proportional to the suspended solids, the resultant electronic signal is proportional to the specific gravity of the slurry. Such devices can give accuracy up to 0.01%. The ultrasonic device measures the percentage of the suspended solids in the slurry by providing a close approximation of the specific gravity.

References


Appendix

List of Manufacturers

ABB K-Flow Inc.
Drawer M Box 849
Millville, NJ 08332
Tel: (800) 825-3569

American Density Materials Inc.
Rd. 2, Box 38E
Belvidere, J 07823
Tel: (908) 475-2373

Anton Paar U.S.A.
13, Maple Leaf Ct.
Ashland, VA 23005
Tel: (800) 221-0174

Arco Instrument Company, Inc.
1745 Production Circle
Riverside, CA 92509
Tel: (909) 788-2823
Fax: (909) 788-2409

Cambridge Applied Systems, Inc.
196 Boston Avenue
Medford, MA 02155
Tel: (617) 393-6500
Dynatron
Automation Products, Inc.
3032 Max Roy Street
Houston, TX 77008
Tel: (800) 231-2062
Fax: (713) 869-7332

Kay-Ray/Sensall, Fisher-Rosemount
1400 Business Center Dr.
Mount Prospect, IL 60056
Tel: (708) 803-5100
Fax: (708) 803-5466

McGee Engineering Co., Inc.
Tujunga Canyon Blvd.
Tujunga, CA 91042
Tel: (800) 353-6675

Porous Materials, Inc.
Cornell Business & Technology Park
Ithaca, NY 14850
Tel: (800) 825-5764

Princo Instruments Inc
1020 Industrial Hwy., Dept L
Southampton, PA 18966-4095
Tel: (800) 496-5345

Quantachrome Corp.
1900-T Corporate Drive
Boynton Beach, FL 33426
Tel: (800) 966-1238

Tricor Systems, Inc.
400-T River Ridge Rd.
Elgin, IL 60123
Tel: (800) 575-0161

X-rite, Inc.
3100-T 44th St. S.W
Grandville, MI 49418
Tel: (800) 545-0694
This chapter begins with a review of the fundamental definitions of strain and ways it can be measured. This is followed by a review of the many types of strain sensors and their application, and sources for strain sensors and signal conditioners. Next, a more detailed look is taken at operating principles of various strain measurement techniques and the associated signal conditioning.

### 22.1 Fundamental Definitions of Strain

Stress and strain are defined in many elementary textbooks about the mechanics of deformable bodies [1, 2]. The terms stress and strain are used to describe loads on and deformations of solid materials. The simplest types of solids to describe are homogeneous and isotropic. Homogeneous means the material properties are the same at different locations and isotropic means the material properties are independent of direction in the material. An annealed steel bar is homogeneous and isotropic, whereas a human femur is not homogeneous because the marrow has very different properties from the bone, and it is not isotropic because its properties are different along the length and along the cross-section.

The concepts of stress and strain are introduced in the context of a long homogeneous isotropic bar subjected to a tensile load (Figure 22.1). The stress \( \sigma \), is the applied force \( F \), divided by the cross-sectional area \( A \). The resulting strain \( \varepsilon \), is the length change \( \Delta L \), divided by the initial length \( L \). The bar elongates in the direction the force is pulling (longitudinal strain \( \varepsilon_L \)) and contracts in the direction perpendicular to the force (transverse strain \( \varepsilon_T \)).

When the strain is not too large, many solid materials behave like linear springs; that is, the displacement is proportional to the applied force. If the same force is applied to a thicker piece of material, the spring is stiffer and the displacement is smaller. This leads to a relation between force and displacement that depends on the dimensions of the material. Material properties, such as the density and specific heat, must be defined in a manner that is independent of the shape and size of the specimen. Elastic material properties are defined in terms of stress and strain. In the linear range of material response, the stress is proportional to the strain (Figure 22.2). The ratio of stress to strain for the bar under tension is an elastic constant called the Young’s modulus \( E \). The negative ratio of the transverse strain to longitudinal strain is the Poisson’s ratio \( \nu \).

Forces can be applied to a material in a manner that will cause distortion rather than elongation (Figure 22.3). A force applied tangent to a surface divided by the cross-sectional area is described as a shear stress \( \tau \). The distortion can be measured by the angle change produced. This is the shear strain \( \gamma \).
when the angle change is small. When the relation between shear stress and shear strain is linear, the ratio of the shear stress to shear strain is the shear modulus $G$.

Temperature change also induces strain. This is thermal expansion. In most materials, thermal strain increases with temperature. Over a limited temperature range, the relationship between thermal strain
and temperature is linear. In this case, the strain divided by the temperature change is the thermal expansion coefficient $\alpha$. In isotropic materials, thermal expansion only produces elongation strain, no shear strain.

Principal directions in a material are directions that undergo elongation but no shear. On any particular surface of a solid, there are always at least two principal directions in which the strain is purely elongation. This is seen if two squares are drawn on the bar under uniform tension (Figure 22.4). When the bar is stretched, the square aligned with the load is elongated, whereas the square at 45° is distorted (angles have changed) and elongated. If the principal directions are known, as with the bar under tension, then strain gages can be applied in these directions. If the principal directions are not known, such as near a hole or notch, in an anisotropic specimen, or in a structure with complicated geometry, then additional strain gages are needed to fully characterize the strain state.

The elastic and thermal properties can be combined to give Hooke’s law, Equations 22.1 to 22.6.
Several types of sensors are used to measure strain. These include piezoresistive gages (foil or wire strain gages and semiconductor strain gages), piezoelectric gages (polyvinylidene fluoride (PVDF) film and quartz), fiber optic gages, birefringent films and materials, and Moiré grids. Each type of sensor requires its own specialized signal conditioning. Selection of the best strain sensor for a given measurement is based on many factors, including specimen geometry, temperature, strain rate, frequency, magnitude, as well as cost, complexity, accuracy, spatial resolution, time resolution, sensitivity to transverse strain, sensitivity to temperature, and complexity of signal conditioning. Table 22.1 describes typical characteristics of several sensors. Table 22.2 lists some manufacturers and the approximate cost of the sensors and associated signal conditioning electronics.

The data in Table 22.1 are to be taken as illustrative and by no means complete. The sensor description section describes only the type of sensor, not the many sizes and shapes. The longitudinal strain sensitivity is given as sensor output per unit longitudinal strain in the sensor direction. If the signal conditioning is included, the sensitivities can all be given in volts out per unit strain \[ \frac{\sigma}{E} \] but this is a function of amplification and the quality of the signal conditioner. The temperature sensitivity is given as output change due to a temperature change. In many cases, higher strain resolution can be achieved, but resolving smaller strain is more difficult and may require vibration and thermal isolation. For the Moiré technique, the strain resolution is a function of the length of the viewing area. This technique can resolve a displacement of 100 nm \((1/4\) fringe order). This is divided by the viewing length to obtain the strain resolution. The spatial resolution corresponds to the gage length for most of the sensor types. The measurable strain range listed is the upper limit for the various sensors. Accuracy and reliability are usually reduced when sensors are used at the upper limit of their capability.

Manufacturers of the various sensors provide technical information that includes details of using the sensors, complete calibration or characterization data, and details of signal conditioning. The extensive technical notes and technical tips provided by Measurements Group, Inc. address such issues as thermal effects \[ 5 \], transverse sensitivity corrections \[ 6 \], soldering techniques \[ 7 \], Rosettes \[ 8 \], and gage fatigue \[ 9 \]. Strain gage catalogs include information about gage materials, sizes, and selection. Manufacturers of other sensors provide similar information.
<table>
<thead>
<tr>
<th>Description</th>
<th>Longitudinal strain sensitivity</th>
<th>Transverse strain sensitivity</th>
<th>Temperature sensitivity</th>
<th>Strain resolution</th>
<th>Spatial resolution</th>
<th>Time resolution</th>
<th>Measurable strain range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Piezoresistive constantan foil</td>
<td>$\frac{\Delta R}{R} \Delta \varepsilon = 2.1$</td>
<td>$\frac{\Delta R}{R} \Delta \varepsilon = \frac{&lt;0.02}{\varepsilon}$</td>
<td>$\frac{\Delta R}{R} \Delta \varepsilon = 2 \times 10^{-8}/^\circ C$</td>
<td>$&lt;1 \mu$strain$^a$</td>
<td>5–100 mm$^b$</td>
<td>$&lt;1 \mu s$</td>
<td>0–3%</td>
</tr>
<tr>
<td>Annealed constantan foil</td>
<td>$\frac{\Delta R}{R} \Delta \varepsilon = 2.1$</td>
<td>$\frac{\Delta R}{R} \Delta \varepsilon = \frac{&lt;0.02}{\varepsilon}$</td>
<td>$\frac{\Delta R}{R} \Delta \varepsilon = 2 \times 10^{-8}/^\circ C$</td>
<td>$&lt;11 \mu$strain</td>
<td>5–100 mm</td>
<td>$&lt;1 \mu s$</td>
<td>0–10%</td>
</tr>
<tr>
<td>Piezoresistive semiconductor</td>
<td>$\frac{\Delta R}{R} \Delta \varepsilon = 150$</td>
<td>$\frac{\Delta R}{R} \Delta \varepsilon = ???$</td>
<td>$\frac{\Delta R}{R} \Delta \varepsilon = 1.7 \times 10^{-8}/^\circ C$</td>
<td>$&lt;0.1 \mu$strain</td>
<td>1–15 mm</td>
<td>$&lt;1 \mu s$</td>
<td>0–1%</td>
</tr>
<tr>
<td>Piezoelectric PVDF</td>
<td>$\frac{\Delta Q}{A} \frac{\Delta \varepsilon}{\Delta \varepsilon} = 120 \text{ nC/m}^2/\mu \varepsilon$</td>
<td>$\frac{\Delta Q}{A} \frac{\Delta \varepsilon}{\Delta \varepsilon} = 60 \text{ nC/m}^2/\mu \varepsilon$</td>
<td>$\Delta Q/A/\Delta T = -27 \text{ } \mu \varepsilon /^\circ C$</td>
<td>1–10 $\mu$strain</td>
<td>Gage size</td>
<td>$&lt;1 \mu s$</td>
<td>0–30%</td>
</tr>
<tr>
<td>Piezoelectric quartz</td>
<td>$\frac{\Delta Q}{A} \frac{\Delta \varepsilon}{\Delta \varepsilon} = 150 \text{ nC/m}^2/\mu \varepsilon$</td>
<td>$\Delta Q/A/\Delta \varepsilon = 0$</td>
<td>$\Delta Q/A/\Delta T = -0.01 \text{ } \mu \varepsilon /^\circ C$</td>
<td>$&lt;0.01 \mu$strain</td>
<td>Gage size</td>
<td>$&lt;10 \mu s$</td>
<td>0–1%</td>
</tr>
<tr>
<td>Fiber optic Fabry–Perot</td>
<td>2 to 1000 $\mu$strain/V</td>
<td>Near zero</td>
<td></td>
<td>$&lt;1 \mu$strain</td>
<td>2–10 mm</td>
<td>$&lt;20 \mu s$</td>
<td></td>
</tr>
<tr>
<td>Birefringent Film</td>
<td>$K = 0.15-0.002$</td>
<td></td>
<td></td>
<td></td>
<td>0.5 mm$^c$</td>
<td>$&lt;5 \mu s$</td>
<td>0.05–5%</td>
</tr>
<tr>
<td>Moiré</td>
<td>1 fringe order/417 nm displ.</td>
<td>1 fringe order/417 nm displ.</td>
<td>Not defined</td>
<td>41.7 $\mu$ over 10 mm</td>
<td>full field$^d$</td>
<td>Limited by signal conditioning</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ With good signal conditioning.

$^b$ Equal to grid area.

$^c$ Gage response is within 100 ns. Most signal conditioning limits response time to far less than this.

$^d$ Annealed foil has a low yield stress and a large strain to failure. It also has hysteresis in the unload and a zero shift under cyclic load.

$^e$ This technique measures a difference in principal strains, $\varepsilon_2 - \varepsilon_1 = N\lambda/2k$

$^f$ Approximately the film thickness.

$^g$ The spatial strain resolution depends on the strain level. This is a displacement measurement technique.
### 22.2 Principles of Operation of Strain Sensors

**Piezoresistive Foil Gages**

Piezoresistive foil and wire gages comprise a thin insulating substrate (usually polyimide film), a foil or wire grid (usually constantan) bonded to the substrate, lead wires to connect the grid to a resistance measuring circuit, and often an insulating encapsulation (another sheet of polyimide film) (Figure 22.5). The grid is laid out in a single direction so that strain will stretch the legs of the grid in the length direction. The gages are designed so that strain in the width or transverse direction separates the legs of the grid without straining them. This makes the gage sensitive to strain only along its length. There is always some sensitivity to transverse strain, and almost no sensitivity to shear strain. In most cases, the transverse sensitivity can be neglected.

When piezoresistive foil or wire strain gages are bonded to a specimen and the specimen is strained, the gage strains as well. The resistance change is related to the strain by a gage factor, Equation 22.7.

---

### TABLE 22.2 Sources and Prices of Strain Sensors

<table>
<thead>
<tr>
<th>Supplier</th>
<th>Address</th>
<th>Sensor Types</th>
<th>Sensor Cost</th>
<th>Signal Conditioning</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Micro Measurements</td>
<td>P.O. Box 27777 Raleigh, NC 27611</td>
<td>Piezoresistive foil, Birefringent film</td>
<td>From $5.00</td>
<td>Wheatstone bridge, Polariscopic</td>
<td>From $500 $5000 to 10,000</td>
</tr>
<tr>
<td>Texas Measurements</td>
<td>P.O. Box 2618 College Station, TX 77841</td>
<td>Piezoresistive foil, and wire Load cells</td>
<td>From $5.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Omega Engineering</td>
<td>P.O. Box 4047 Stamford, CT 06907-0047</td>
<td>Piezoresistive foil</td>
<td>From $5.00</td>
<td>Strain meter, Wheatstone bridge</td>
<td>From $550 $2700</td>
</tr>
<tr>
<td>Dynasen, Inc.</td>
<td>20 Arnold Pl. Goleta, CA 93117</td>
<td>Piezoresistive foil, Specialty gages for shock wave measurements, Piezoelectric PVDF (calibrated)</td>
<td>From $55.00 $55.00</td>
<td>2-Channel pulsed Wheatstone bridge, Passive charge integrator</td>
<td>$5000 $250.00</td>
</tr>
<tr>
<td>Entran Sensors and Electronics</td>
<td>Entran Devices, Inc. 10 Washington Ave. Fairfield, NJ 07004-3877</td>
<td>Piezoresistive semiconductor</td>
<td>From $15.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amp Inc.</td>
<td>P.O. Box 799 Valley Forge, PA 19482</td>
<td>Piezoelectric PVDF (not calibrated)</td>
<td>From $5.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kistler Instrument Corp.</td>
<td>Amherst, NY 14228-2171</td>
<td>Piezoelectric quartz</td>
<td>Electrometer</td>
<td></td>
<td>From $3500.00</td>
</tr>
<tr>
<td>F&amp;S Inc.</td>
<td>Fiber and Sensor Technologies P.O. Box 11704, Blacksburg, VA 24062-1704</td>
<td>Fabry–Perot strain sensors</td>
<td>Charge amplifier</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Photomechanics, Inc.</td>
<td>512 Princeton Dr. Vestal, NY 13850-2912</td>
<td>Moiré interferometer</td>
<td>$60,000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

© 1999 by CRC Press LLC
where \( \frac{\Delta R}{R} = \text{Relative resistance change} \)

\( G = \text{Gage factor} \)

\( \varepsilon = \text{Strain} \)

These gages respond to the average strain over the area covered by the grid [10]. The resistance change is also sensitive to temperature. If the temperature changes during the measurement period, a correction must be made to distinguish the strain response from the thermal response. The gage response to longitudinal strain, transverse strain, and temperature change is given by Equation 22.8.

\[
\frac{\Delta R}{R} = G_L \varepsilon_L + G_T \varepsilon_T + G_T \Delta T
\]  

(22.8)

where \( G_L, G_T, \) and \( G_T \) are the longitudinal, transverse, and temperature sensitivity, respectively. Micromeasurements, Inc. uses a different notation. Their gage data is provided as \( G_L = F_G, G_T = K_T F_G, \) and \( G_T = \beta_g \).

When a strain gage is bonded to a specimen and the temperature changes, the strain used in Equation 22.8 is the total strain, thermal plus stress induced, as given by Equation 22.7.

The temperature contribution to gage output must be removed if the gages are used in tests where the temperature changes. A scheme referred to as self-temperature compensation (STC) can be used. This is accomplished by selecting a piezoresistive material whose thermal output can be canceled by the strain induced by thermal expansion of the test specimen. Gage manufacturers specify STC numbers that match the thermal expansion coefficients of common specimen materials.

Strain of piezoresistive materials produces a relative resistance change. The resistance change is the result of changes in resistivity and dimensional changes. Consider a single leg of the grid of a strain gage with a rectangular cross-section (Figure 22.6). The resistance is given by Equation 22.9.

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

(22.9)

where \( R = \text{Resistance} \)

\( \rho = \text{Resistivity} \)

\( L = \text{Length} \)

\( A = \text{Area of the cross-section} \)
A small change in resistance is given by the first-order terms of a Taylor’s series expansion, Equation 22.10.

\[ \Delta R = \frac{\partial R}{\partial \rho} \Delta \rho + \frac{\partial R}{\partial L} \Delta L + \frac{\partial R}{\partial A} \Delta A \]  \hspace{1cm} (22.10)

Differentiating Equation 22.9 to obtain each term of Equation 22.10 and then dividing by the initial resistance leads to Equation 22.11.

\[ \frac{\Delta R}{R_0} = \frac{\Delta \rho}{\rho_0} + \frac{\Delta L}{L_0} - \frac{\Delta A}{A_0} \]  \hspace{1cm} (22.11)

The relative resistance change is due to a change in resistivity, a change in length strain, and a change in area strain.

The strain gage is a composite material. The metal in a strain gage is like a metal fiber in a polymer matrix. When the polymer matrix is deformed, the metal is dragged along in the length direction; but in the width and thickness directions, the strain is not passed to the metal. This results in a stress state called uniaxial stress. This state was discussed in the examples above. The mathematical details involve an inclusion problem [11, 12]. Accepting that the stress state is uniaxial, the relationship between the area change and the length change in Equation 22.11 is found from the Poisson’s ratio. The area strain is the sum of the width and thickness strain, Equation 22.12.

\[ \frac{\Delta A}{A_0} = \frac{\Delta w}{w_0} + \frac{\Delta t}{t_0} \]  \hspace{1cm} (22.12)

The definition of the Poisson’s ratio gives Equation 22.13.

\[ \frac{\Delta A}{A_0} = -2\nu \frac{\Delta L}{L_0} \]  \hspace{1cm} (22.13)

where \( \Delta w/w \) = width strain and \( \Delta t/t \) = thickness strain. Substitution of Equation 22.13 into Equation 22.11 gives Equation 22.14 for the relative resistance change.

\[ \frac{\Delta R}{R_0} = \frac{\Delta \rho}{\rho_0} + \frac{\Delta L}{L_0} \left( 1 + 2\nu \right) \]  \hspace{1cm} (22.14)

The relative resistivity changes in response to stress. The resistivity is a second-order tensor [13], and the contribution to the overall resistance change can be found in terms of strain using the elastic
The results lead to an elastic gage factor just over 2 for constantan gages. If the strain is large, the foil or wire in the gage will experience plastic deformation. When the deformation is plastic, the resistivity change is negligible and the dimensional change dominates. In this case, Poisson’s ratio is 0.5 and the gage factor is 2. This effect is utilized in manufacturing gages for measuring strains larger than 1.5%. In this case, annealed foil is used. The annealed foil undergoes plastic deformation without failure. These gages are capable of measuring strain in excess of 10%. When metals undergo plastic deformation, they do not unload to the initial strain. This shows up as hysteresis in the gage response, that is, on unload, the resistance does not return to its initial value.

Foil and wire strain gages can be obtained in several configurations. They can be constructed with different backing materials, and left open faced or fully encapsulated. Backing materials include polyimide and glass fiber-reinforced phenolic resin. Gages can be obtained with solder tabs for attaching lead wires, or with lead wires attached. They come in many sizes, and in multiple gage configurations called rosettes.

Strain gages are mounted to test specimens with adhesives using a procedure that is suitable for bonding most types of strain sensors. This is accomplished in a step-by-step procedure that starts with surface preparation. An overview of the procedure is briefly described. To successfully mount strain gages, the surface is first degreased. The surface is abraded with a fine emery cloth or 400 grit paper to remove any loose paint, rust, or deposits. Gage layout lines are drawn (not scribed) on the surface in a cross pattern with pen or pencil, one line in the grid direction and one in the transverse direction. The surface is then cleaned with isopropyl alcohol. This can be done with an ultrasonic cleaner or with wipes. If wiped, the paper or gauze wipe should be folded and a small amount of alcohol applied. The surface should be wiped with one pass and the wipe discarded. This should be repeated, wiping in the other direction. The final step is to neutralize the surface, bringing the alkalinity to a pH of 7 to 7.5. A surface neutralizer is available from most adhesive suppliers. The final step is to apply the gage.

Gage application is accomplished with cellophane tape, quick-set glue, and a catalyst. The gage is placed on a clean glass or plastic surface with bonding side down, using tweezers. (Never touch the gage. Oils from skin prevent proper adhesion.) The gage is then taped down with a 100 mm piece of cellophane tape. The tape is then peeled up with the gage attached. The gage can now be taped onto its desired location on the test specimen. Once the gage has been properly aligned, the tape is peeled back from one side, lifting the gage from the surface. The tape should remain adhered to the surface about 1 cm from the gage. Note that one side of the tape is still attached to the specimen so that the gage can be easily returned to its desired position. A thin coating of catalyst is applied to the exposed gage surface. A drop of glue is placed at the joint of the tape and the specimen. Holding the tape at about a 30° angle from the surface, the tape can be slowly wiped down onto the surface. This moves the glue line forward. After the glue line has passed the gage, the gage should be pressed in place and held for approximately 1 min. The tape can now be peeled back to expose the gage, and lead wires can be attached.

The relative resistance change of piezoresistive gages is usually measured using a Wheatstone bridge. This allows a small change of resistance to be measured relative to an initial zero value, rather than relative to a large resistance value, with a corresponding increase in sensitivity and resolution. The Wheatstone bridge is a combination of four resistors and a voltage source (Figure 22.7). One to four of the resistors in the bridge can be strain gages. The output of the bridge is the difference between the voltage at points B and D. Paths ABC and ADC are voltage dividers so that \( V_B \) and \( V_D \) are given by Equations 22.15a and b.

\[
V_B = V_{in} \frac{R_2}{R_1 + R_2} \tag{22.15a}
\]

\[
V_D = V_{in} \frac{R_3}{R_3 + R_4} \tag{22.15b}
\]
The bridge output, Equation 22.16, is zero when the balance condition, Equation 22.17, is met.

\[ V_0 = V_b - V_d \]  \hspace{1cm} (22.16)

\[ R_1 R_3 = R_2 R_4 \]  \hspace{1cm} (22.17)

Wheatstone bridge signal conditioners are constructed with a way to “balance” the bridge by adjusting the ratio of the resistances so that the bridge output is initially zero.

The balance condition is no longer met if the resistance values undergo small changes \( \Delta R_1, \Delta R_2, \Delta R_3, \Delta R_4 \). If the values \( R_1 + \Delta R_1 \), etc. are substituted into Equation 22.15, the results substituted into Equation 22.16, condition (22.17) used, and the higher order terms neglected, the result is Equation 22.18 for the bridge output.

\[ V_{out} = V_{in} \left( \frac{R_2}{R_1 + R_2} \right) \left( \frac{\Delta R_1}{R_1} + \frac{\Delta R_2}{R_2} - \frac{\Delta R_3}{R_3} - \frac{\Delta R_4}{R_4} \right) \]  \hspace{1cm} (22.18)

The Wheatstone bridge can be used to directly cancel the effect of thermal drift. If \( R_1 \) is a strain gage bonded to a specimen and \( R_2 \) is a strain gage held onto a specimen with heat sink compound (a thermally conductive grease available at any electronics store), then \( R_1 \) will respond to strain plus temperature, and \( R_2 \) will only respond to temperature. Since the bridge subtracts the output of \( R_1 \) from that of \( R_2 \), the temperature effect cancels.

The sensitivity of a measuring system is the output per unit change in the quantity to be measured. If the resistance change is from a strain gage, the sensitivity of the Wheatstone bridge system is proportional to the input voltage. Increasing the voltage increases the sensitivity. There is a practical limitation to increasing the voltage to large values. The power dissipated (heat) in the gage is \( P = IR \), where \( I \) is the current through the gage, can be found from the input voltage and the bridge resistances. This heat must go somewhere or the temperature of the gage will continuously rise and the resistance will change due to heating. If the gage is mounted on a good thermal conductor, more power can be conducted away than if the gage is mounted on an insulator. The specimen must act as a heat sink.

Heat sinking ability is proportional to the thermal conductivity of the specimen material. A reasonable temperature gradient to allow the gage to induce in a material is 40°C per meter (about 1°C per 25 mm).
For thick specimens (thickness several times the largest gage dimension), this can be conducted away to the grips or convected to the surrounding atmosphere. If the four bridge resistances are approximately equal, the power to the gage in terms of the bridge voltage is given by Equation 22.19.

\[ P_g = \frac{V_{in}^2}{4R} \]  

The power per unit grid area, \( A_g \), or power density to the gage can be equated to the thermal conductivity of the specimen and the allowable temperature gradient in the specimen by Equation 22.20.

\[ \frac{P_g}{A_g} = \frac{V_{in}^2}{4RA_g} = KV T \]  

Thermal conductivities of most materials can be found in tables or on the Web. Some typical values are Al: \( K = 204 \text{ W m}^{-1} \text{ °C}^{-1} \), steel: \( K = 20 \text{ to } 70 \text{ W m}^{-1} \text{ °C}^{-1} \), glass: \( K = 0.78 \text{ W m}^{-1} \text{ °C}^{-1} \) [17].

The acceptable bridge voltage can be calculated from Equation 22.21.

\[ V_{in} = \sqrt{KV T 4RA_g} \]  

A sample calculation shows that for a 0.010 m \( \times \) 0.010 m 120 \( \Omega \) grid bonded to a thick piece of aluminum with a thermal conductivity of 204 W m\(^{-1}\) °C\(^{-1}\) and an acceptable temperature gradient of 40°C per meter, the maximum bridge voltage is 19 V. If thin specimens are used, the allowable temperature gradient will be smaller. If smaller gages are used for better spatial resolution, the bridge excitation voltage must be reduced with a corresponding reduction in sensitivity.

A considerably higher bridge voltage can be used if the bridge voltage is pulsed for a short duration. This dissipates substantially less energy in the gage and thus increases the sensitivity by a factor of 10 to 100. Wheatstone bridge pulse power supplies with variable pulse width from 10 \( \mu \)s and excitation of 350 V are commercially available [18].

The strain measurement required is often in a complex loading situation where the directions of principal strain are not known. In this case, three strain gages must be bonded to the test specimen at three angles. This is called a strain rosette. The angle between the rosette and the principal directions, as well as the magnitude of the principal strains, can be determined from the output of the rosette gages. This is most easily accomplished using the construct of a Mohr’s circle (Figure 22.8).

A common rosette is the 0–45–90° pattern. The rosette is bonded to the specimen with relative rotations of 0°, 45°, and 90°. These will be referred to as the \( x \), \( x' \), and \( y \) directions. The principal directions are labeled the 1 and 2 directions. The unknown angle between the \( x \) direction and the 1 direction is labeled \( \theta \). The Mohr’s circle is drawn with the elongational strain on the horizontal axis and the shear strain on the vertical axis. The center of the circle is labeled \( C \) and the radius \( R \). The principal directions correspond to zero shear strain. The principal values are given by Equations 22.22 and 22.23.

\[ \varepsilon_1 = C + R \]  

\[ \varepsilon_2 = C - R \]  

A rotation through an angle \( 2\theta \) on the Mohr’s circle corresponds to a rotation of the rosette of \( \theta \) relative to the principal directions. The center of the circle is given by Equation 22.24 and the output of the strain gages is given by Equations 22.25 to 22.27.
Dividing Equation 22.25 by Equation 22.26 leads to $\theta$ and then to $R$, Equations 22.28 and 22.29.

$$C = \frac{\varepsilon_{xx} + \varepsilon_{yy}}{2} \quad (22.24)$$

$$\varepsilon_{xx} - C = R \cos 2\theta \quad (22.25)$$

$$\varepsilon_{x'x'} - C = -R \sin 2\theta \quad (22.26)$$

$$\varepsilon_{yy} - C = -R \cos 2\theta \quad (22.27)$$

The principal directions and principal strain values have been found from the output of the three rosette gages.

**Piezoresistive Semiconducting Gages**

Piezoresistive semiconductor strain gages, like piezoresistive foil and wire gages, undergo a resistance change in response to strain, but with nearly an order of magnitude larger gage factor [19]. The coupling between resistance change and temperature is very large, so these gages have to be temperature compensated. The change of resistance with a small temperature change can be an order of magnitude larger than that induced by strain. Semiconductor strain gages are typically used to manufacture transducers such as load cells. They are fragile and require great care in their application.

© 1999 by CRC Press LLC
Piezoelectric Gages

Piezoelectric strain gages are, effectively, parallel plate capacitors whose dielectric changes polarization in response to strain [14]. When the polarization changes, a charge proportional to the strain is produced on the electrodes. PVDF film strain gages are inexpensive, but not very accurate and subject to depoling by moderate temperature. They make good sensors for dynamic measurements such as frequency and logarithm decrement, but not for quantitative measurements of strain. When used for quasistatic measurements, the charge tends to drain through the measuring instrument. This causes the signal to decay with a time constant dependent on the input impedance of the measuring instrument. Quartz gages are very accurate, but also lose charge through the measuring instrument. Time constants can be relatively long (seconds to hours) with electrometers or charge amplifiers.

The PVDF gage consists of a thin piezoelectric film with metal electrodes (Figure 22.9). Lead wires connect the electrodes to a charge measuring circuit. Gages can be obtained with the electrodes encapsulated between insulating layers of polyimide.

The gage output can be described in terms of a net dipole moment per unit volume. If the net dipole moment is the total charge, $Q$, on the electrodes multiplied by spacing, $d$, between the electrodes, then the polarization is given by Equation 22.30.

$$P = \frac{Qd}{V} \quad (22.30)$$

From Equation 22.30, it is seen that the polarization $P$ (approximately equal to the electric displacement $D$) is the charge per unit electrode area (Figure 22.10).

A Taylor’s series expansion of Equation 22.30 gives Equation 22.31.

$$\Delta P = \frac{\partial P}{\partial V} \Delta V + \frac{\partial P}{\partial (Qd)} \Delta Qd \quad (22.31)$$

Which, after differentiating Equation 22.30 and substituting becomes Equation 22.31.

$$\Delta P = \frac{\Delta Qd}{V_0} - \frac{\Delta V}{V_0} P_0 \quad (22.32)$$
For PVDF film, the second term in Equation 22.32 dominates. The output is proportional to the remanent polarization $P_r$. The remanent polarization slowly decays with time, has a strong dependence on temperature, and decays rapidly at temperatures around 50°C. This makes accuracy a problem. If the sensors are kept at low temperature, accuracy can be maintained within $\pm 3\%$.

Strain sensors can also be constructed from piezoelectric ceramics like lead zirconate titanate (PZT) or barium titanate. Ceramics are brittle and can be depoled by strain so should only be used at strains less than 200 microstrain. PZT loses some of its polarization with time and thus has accuracy problems, but remains polar to temperatures of 150°C or higher. “Hard” PZT (usually iron doped) is the best composition for polarization stability and low hysteresis. Quartz has the best accuracy. It is not polar, but polarization is induced by strain. Quartz has excellent resolution and accuracy over a broad temperature range but is limited to low strain levels. It is also brittle, so is limited to small strain.

Two circuits are commonly used for piezoelectric signal conditioning: the electrometer and the charge amplifier (Figure 22.11). In the electrometer circuit, the piezoelectric sensor is connected to a capacitor with a capacitance value $C_s$ at least 1000 times that of the sensor $C_g$. There is always some resistance in
the cable that connects the sensor to the capacitor. The circuit is simply two capacitors in parallel connected by a resistance. The charge equilibrates with a time constant given by \( R_{\text{g}}C_{\text{g}} \). This time constant limits the fastest risetime that can be resolved to about 50 ns, effectively instantaneous for most applications. The charge is measured by measuring the voltage on the capacitor, then using Equation 22.33.

\[
Q = CV
\] (22.33)

The difficulty is that measuring devices drain the charge, causing a time decay with a time constant \( RC \). This causes the signal to be lost rapidly if conventional op amps are used. FET input op amps have a very high input impedance and can extend this time constant to many hours. The charge amplifier is another circuit used to measure charge. This is usually an FET input op amp with a capacitor feedback. This does not really amplify charge, but produces a voltage proportional to the input charge. Again, the time constant can be many hours, allowing use of piezoelectric sensors for near static measurements.

High input impedance electrometer and charge amplifier signal conditioners for near static measurements are commercially available [20] as well as low-cost capacitive terminators for high-frequency (high kilohertz to megahertz) measurements [18]. An advantage of piezoelectric sensors is that they are active sensors that do not require any external energy source.

**Fiber Optic Strain Gages**

Fiber optic strain gages are miniature interferometers [21, 22]. Many commercially available sensors are based on the Fabry–Perot interferometer. The Fabry–Perot interferometer measures the change in the size of a very small cavity.

Fabry–Perot strain sensors (Figure 22.12) comprise a laser light source, single-mode optical fibers, a coupler (the fiber optic equivalent of a beam splitter), a cavity that senses strain, and a photodetector. Light leaves the laser diode. It passes down the fiber, through the coupler, and to the cavity. The end of the fiber is the equivalent of a partially silvered mirror. Some of the light is reflected back up the fiber and some is transmitted. The transmitted light crosses the cavity and then is reflected from the opposite end back into the fiber where it recombines with the first reflected beam. The two beams have a phase difference related to twice the cavity length. The recombined beam passes through the coupler to the photodetector. If the two reflected beams are in phase, there will be constructive interference. If the two
beams are out of phase, there will be destructive interference. The cavity is bonded to a specimen. When the specimen is strained, the cavity stretches. This results in a phase change of the cavity beam, causing a cycling between constructive and destructive interference. For a 1.3 μm light source, each peak in output corresponds to a 650 nm gap displacement. The gap displacement divided by the gap length gives the strain. The output is continuous between peaks so that a 3 mm gage can resolve 1 μstrain.

Birefringent Film Strain Sensing

Birefringent film strain sensors give a full field measurement of strain. A nice demonstration of this effect can be achieved with two sheets of inexpensive Polaroid film, a 6 mm thick, 25 mm × 200 mm bar of Plexiglas (polymethylmethacrylate or PMMA), and an overhead projector. Place the two Polaroid sheets at 90° to one another so that the light is blocked. Place the PMMA between the Polaroid sheets. Apply a bending moment to the bar and color fringes will appear. Birefringent materials have a different speed of light in different directions. This means that if light is polarized in a particular direction and passed through a birefringent specimen, if the fast direction is aligned with the electric field vector, the light passes through faster than if the slow direction is aligned with the electric field vector. This effect can be used to produce optical interference. In some materials, birefringence is induced by strain. The fast and slow directions correspond to the directions of principal strain, and the amount of birefringence corresponds to the magnitude of the strain. One component of the electric field vector travels through the specimen faster than the other. They emerge with a phase difference. This changes the relative amplitude and thus rotates the polarization of the light. If there is no birefringence, no light passes through the second polarizer. As the birefringence increases with strain, light passes through. As it further increases, the polarization rotation will be a full 180° and again no light will pass through. This produces a fringe that corresponds to a constant difference in principal strains. The difference in principal strains is given by Equation 22.34.

\[ \varepsilon_2 - \varepsilon_1 = \frac{N\lambda}{tK} \]  

(22.34)

where \( \varepsilon_1, \varepsilon_2 = \) Principal strains
\( N = \) Fringe order
\( \lambda = \) Wavelength
\( t = \) Specimen thickness
\( K = \) Strain-optical coefficient of the photoelastic material

A similar technique can be used with a birefringent plastic film with a silvered backing laminated to the surface of a specimen. Polarized light is passed through the film; it reflects from the backing, passes back through the film, and through the second polarizer. In this case, because light passes twice through the film, the equation governing the difference in principal strains is Equation 22.35.

\[ \varepsilon_2 - \varepsilon_1 = \frac{N\lambda}{2tK} \]  

(22.35)

If the polarizers align with principal strain directions, no birefringence is observed. Rotation of both polarizers allows the principal directions to be found at various locations on the test specimen. If a full view of the fringes is desired, quarter wave plates are used (Figure 22.13). In this arrangement, light is passed through the first polarizer, resulting in plane polarization; through the quarter wave plate, resulting in circular polarization; through the test specimen, resulting in phase changes; through the second quarter wave plate to return to plane polarization; and then through the final polarizer.

The optical systems for viewing birefringence are commercially available as "Polariscopes" [23]. Optical components to construct custom systems are available from many optical components suppliers.
Moiré Strain Sensing

Moiré interference is another technique that gives a full field measurement, but it measures displacement rather than strain. The strain field must be computed from the displacement field. This technique is based on the interference obtained when two transparent plates are covered with equally spaced stripes. If the plates are held over one another, they can be aligned so that no light will pass through or so that all light will pass through. If one of the plates is stretched, the spacing of the lines is wider on the stretched plate. Now, if one plate is placed over the other, in some regions light will pass through and in some regions it will not (Figure 22.14). The dark and light bands produced give information about the displacement field.

Moiré is defined as a series of broad dark and light patterns formed by the superposition of two regular gratings [24]. The dark or light regions are called fringes. Examples of pure extension and pure rotation are shown. In both cases, some of the light that would emerge from the first grating is obstructed by the superimposed grating. At the centers of the dark fringes, the bar of one grating covers the space of the other and no light comes through. The emergent intensity, I, is zero. Proceeding from there toward the next dark fringe, the amount of obstruction diminishes linearly and the amount of light increases linearly until the bar of one grating falls above the bar of the other. There, the maximum amount of light passes through the gratings.

Both geometric interference and optical interference are used. This discussion is restricted to geometric interference. Geometric moiré takes advantage of the interference of two gratings to determine displacements and rotations in the plane of view. In-plane moiré is typically conducted with two gratings, one applied to the specimen (specimen grating) and the other put in contact with the specimen grating (reference grating). When the specimen is strained, interference patterns or fringes occur. N is the moiré fringe order. Each fringe corresponds to an increase or decrease of specimen displacement by one grating pitch. The relationship between displacement and fringes is \( \delta = gN \), where \( \delta \) is component of the displacement perpendicular to the reference grating lines, \( g \) is reference grating pitch, and \( N \) is the fringe order.

For convenience, a zero-order fringe is designated assuming the displacement there is zero. With the reference grating at 0° and 90°, the fringe orders \( N_x \) and \( N_y \) are obtained. The displacements in \( x \), \( y \) directions are then obtained from Equations 22.36 and 22.37.

\[
\begin{align*}
    u_x(x, y) &= gN_x(x, y) \quad (22.36) \\
    u_y(x, y) &= gN_y(x, y) \quad (22.37)
\end{align*}
\]

Differentiation of Equations 22.36 and 22.37 gives the strains, Equations 22.38 through 22.40.
In most cases, the sensitivity of geometric moiré is not adequate for determination of strain distributions. Strain analysis should be conducted with high-sensitivity measurement of displacement using moiré interferometry [24, 25]. Moiré interferometers are commercially available [26]. Out-of-plane measurement can be conducted with one grating (the reference grating). The reference grating is made to interfere with either its reflection or its shadow [27, 28].

\[
\varepsilon_x = \frac{\partial u_x}{\partial x} = g \frac{\partial N_x}{\partial x} 
\]

\[
\varepsilon_{xy} = \frac{1}{2} \left( \frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} \right) = \frac{1}{2} \left( g \frac{\partial N_x}{\partial x} + g \frac{\partial N_y}{\partial y} \right) 
\]

\[
\varepsilon_y = \frac{\partial u_y}{\partial y} = g \frac{\partial N_y}{\partial y}
\]
References

9. TN-08-1, Fatigue of Strain Gages, Measurements Group, Inc., Raleigh, NC 27611.
18. Dynasen, Inc. 20 Arnold Pl., Goleta, CA 93117.
23. Photoelastic Division, Measurements Group, Inc., P.O. Box 27777, Raleigh, NC 27611.
27. Photomechanics, Inc. 512 Princeton Dr. Vestal, NY, 13850-2912.
Force, which is a vector quantity, can be defined as an action that will cause an acceleration or a certain reaction of a body. This chapter will outline the methods that can be employed to determine the magnitude of these forces.

### 23.1 General Considerations

The determination or measurement of forces must yield to the following considerations: if the forces acting on a body do not produce any acceleration, they must form a system of forces in equilibrium. The system is then considered to be in static equilibrium. The forces experienced by a body can be classified into two categories: internal, where the individual particles of a body act on each other, and external otherwise. If a body is supported by other bodies while subject to the action of forces, deformations and/or displacements will be produced at the points of support or contact. The internal forces will be distributed throughout the body until equilibrium is established, and then the body is said to be in a state of tension, compression, or shear. In considering a body at a definite section, it is evident that all the internal forces act in pairs, the two forces being equal and opposite, whereas the external forces act singly.

### 23.2 Hooke’s Law

The basis for force measurement results from the physical behavior of a body under external forces. Therefore, it is useful to review briefly the mechanical behavior of materials. When a metal is loaded in uniaxial tension, uniaxial compression, or simple shear (Figure 23.1), it will behave elastically until a critical value of normal stress ($S$) or shear stress ($\tau$) is reached, and then it will deform plastically [1]. In the elastic region, the atoms are temporarily displaced but return to their equilibrium positions when the load is removed. Stress ($S$ or $\tau$) and strain ($e$ or $\gamma$) in the elastic region are defined as indicated in Figure 23.2.

\[
\nu = \frac{e_2}{e_1} \tag{23.1}
\]
Poisson’s ratio \((v)\) is the ratio of transverse \((e_2)\) to direct \((e_1)\) strain in tension or compression. In the elastic region, \(v\) is between 1/4 and 1/3 for metals. The relation between stress and strain in the elastic region is given by Hooke’s law:

\[
S = \frac{p}{A_0} \quad e = \frac{\Delta l}{l_0}
\]

\[
\tau = G\gamma
\]

where \(E\) and \(G\) are the Young’s and shear modulus of elasticity, respectively. A small change in specific volume \((\Delta Vol/Vol)\) can be related to the elastic deformation, which is shown to be as follows for an isotropic material (same properties in all directions).

\[
\frac{\Delta Vol}{Vol} = e_i \left(1 - 2v\right)
\]

The bulk modulus \((K = \text{reciprocal of compressibility})\) is defined as follows:

\[
K = \Delta p \left(\frac{\Delta Vol}{Vol}\right)
\]
where $\Delta p$ is the pressure acting at a particular point. For an elastic solid loaded in uniaxial compression (S):

$$K = S\left(\frac{\Delta Vol}{Vol}\right) = \frac{S}{\epsilon_1(1-2v)} = \frac{E}{1-2v}$$  \hspace{1cm} (23.6)

Thus, an elastic solid is compressible as long as $v$ is less than 1/2, which is normally the case for metals. Hooke’s law (Equation 23.2) for uniaxial tension can be generalized for a three-dimensional elastic condition.

The theory of elasticity is well established and is used as a basis for force measuring techniques. Note that the measurement of forces in separate engineering applications is very application specific, and care must be taken in the selection of the measuring techniques outlined below.

### Basic Methods of Force Measurement

An unknown force may be measured by the following means:

1. Balancing the unknown force against a standard mass through a system of levers.
2. Measuring the acceleration of a known mass.
3. Equalizing it to a magnetic force generated by the interaction of a current-carrying coil and a magnet.
4. Distributing the force on a specific area to generate pressure, and then measuring the pressure.
5. Converting the applied force into the deformation of an elastic element.

The aforementioned methods used for measuring forces yield a variety of designs of measuring equipment. The challenge involved with the task of measuring force resides primarily in sensor design. The basics of sensor design can be resolved into two problems:

1. Primary geometric, or physical constraints, governed by the application of the force sensor device.
2. The means by which the force can be converted into a workable signal form (such as electronic signals or graduated displacements).

The remaining sections will discuss the types of devices used for force to signal conversion and finally illustrate some examples of applications of these devices for measuring forces.

### 23.3 Force Sensors

Force sensors are required for a basic understanding of the response of a system. For example, cutting forces generated by a machining process can be monitored to detect a tool failure or to diagnose the causes of this failure in controlling the process parameters, and in evaluating the quality of the surface produced. Force sensors are used to monitor impact forces in the automotive industry. Robotic handling and assembly tasks are controlled by detecting the forces generated at the end effector. Direct measurement of forces is useful in controlling many mechanical systems.

Some types of force sensors are based on measuring a deflection caused by the force. Relatively high deflections (typically, several micrometers) would be necessary for this technique to be feasible. The excellent elastic properties of helical springs make it possible to apply them successfully as force sensors that transform the load to be measured into a deflection. The relation between force and deflection in the elastic region is demonstrated by Hooke’s law. Force sensors that employ strain gage elements or piezoelectric (quartz) crystals with built-in microelectronics are common. Both impulsive forces and slowly varying forces can be monitored using these sensors.

Of the available force measuring techniques, a general subgroup can be defined as that of load cells. Load cells are comprised generally of a rigid outer structure, some medium that is used for measuring
the applied force, and the measuring gage. Load cells are used for sensing large, static or slowly varying forces with little deflection and are a relatively accurate means of sensing forces. Typical accuracies are of the order of 0.1% of the full-scale readings. Various strategies can be employed for measuring forces that are strongly dependent on the design of the load cell. For example, Figure 23.3 illustrates different types of load cells that can be employed in sensing large forces for relatively little cost. The hydraulic load cell employs a very stiff outer structure with an internal cavity filled with a fluid. Application of a load increases the oil pressure, which can be read off an accurate gage.

Other sensing techniques can be utilized to monitor forces, such as piezoelectric transducers for quicker response of varying loads, pneumatic methods, strain gages, etc. The proper sensing technique needs special consideration based on the conditions required for monitoring.

**Strain Gage Load Cell**

The strain gage load cell consists of a structure that elastically deforms when subjected to a force and a strain gage network that produces an electrical signal proportional to this deformation. Examples of this are beam and ring types of load cells.

**Strain Gages**

Strain gages use a length of gage wire to produce the desired resistance (which is usually about 120 Ω) in the form of a flat coil. This coil is then cemented (bonded) between two thin insulating sheets of paper or plastic. Such a gage cannot be used directly to measure deflection. It has to be first fixed properly to a member to be strained. After bonding the gage to the member, they are baked at about 195°F (90°C) to remove moisture. Coating the unit with wax or resin will provide some mechanical protection. The resistance between the member under test and the gage itself must be at least 50 MΩ. The total area of all conductors must remain small so that the cement can easily transmit the force necessary to deform the wire. As the member is stressed, the resulting strain deforms the strain gage and the cross-sectional area diminishes. This causes an increase in resistivity of the gage that is easily determined. In order to measure very small strains, it is necessary to measure small changes of the resistance per unit resistance (ΔR/R). The change in the resistance of a bonded strain gage is usually less than 0.5%. A wide variety of gage sizes and grid shapes are available, and typical examples are shown in Figure 23.4.

The use of strain gages to measure force requires careful consideration with respect to rigidity and environment. By virtue of their design, strain gages of shorter length generally possess higher response frequencies (examples: 660 kHz for a gage of 0.2 mm and 20 kHz for a gage of 60 mm in length). The environmental considerations focus mainly on the temperature of the gage. It is well known that resistance is a function of temperature and, thus, strain gages are susceptible to variations in temperature. Thus, if it is known that the temperature of the gage will vary due to any influence, temperature compensation is required in order to ensure that the force measurement is accurate.

A Wheatstone bridge (Figure 23.5) is usually used to measure this small order of magnitude. In Figure 23.5, no current will flow through the galvanometer (G) if the four resistances satisfy a certain
condition. In order to demonstrate how a Wheatstone bridge operates [3], a voltage scale has been drawn at points C and D of Figure 23.5. Assume that $R_1$ is a bonded gage and that initially Equation 23.7 is satisfied. If $R_1$ is now stretched so that its resistance increases by one unit (+$\Delta R$), the voltage at point D will be increased from zero to plus one unit of voltage (+$\Delta V$), and there will be a voltage difference of one unit between C and D that will give rise to a current through C. If $R_4$ is also a bonded gage, and at the same time that $R_1$ changes by +$\Delta R$, $R_4$ changes by –$\Delta R$, the voltage at D will move to +2$\Delta V$. Also, if at the same time, $R_2$ changes by –$\Delta R$, and $R_3$ changes by +$\Delta R$, then the voltage of point C will move to –2$\Delta V$, and the voltage difference between C and D will now be 4$\Delta V$. It is then apparent that although a single gage can be used, the sensitivity can be increased fourfold if two gages are used in tension while two others are used in compression.

$$\frac{R_1}{R_4} = \frac{R_2}{R_3}$$  \hspace{1cm} (23.7)

The grid configuration of the metal-foil resistance strain gages is formed by a photo-etching process. The shortest gage available is 0.20 mm; the longest is 102 mm. Standard gage resistance are 120 $\Omega$ and 350 $\Omega$. A strain gage exhibits a resistance change $\Delta R/R$ that is related to the strain in the direction of the grid lines by the expression in Equation 23.8 (where $S_g$ is the gage factor or calibration constant for the gage).

$$\frac{\Delta R}{R} = S_g \varepsilon$$  \hspace{1cm} (23.8)
Beam-type load cells are commonly employed for measuring low-level loads [3]. A simple cantilever beam (see Figure 23.6(a)) with four strain gages, two on the top surface and two on the bottom surface (all oriented along the axis of the beam) is used as the elastic member (sensor) for the load cell. The gages are wired into a Wheatstone bridge as shown in Figure 23.6(b). The load $P$ produces a moment $M = Px$ at the gage location ($x$) that results in the following strains:

$$
\varepsilon_1 = -\varepsilon_2 = \varepsilon_3 = -\varepsilon_4 = \frac{6M}{Ebh^2} = \frac{6Px}{Ebh^2}
$$

(23.9)

where $b$ is the width of the cross-section of the beam and $h$ is the height of the cross-section of the beam. Thus, the response of the strain gages is obtained from Equation 23.10.

$$
\frac{\Delta R_1}{R_1} = \frac{\Delta R_2}{R_2} = \frac{\Delta R_3}{R_3} = \frac{\Delta R_4}{R_4} = \frac{6S_i Px}{Ebh^2}
$$

(23.10)

The output voltage $E_o$ from the Wheatstone bridge, resulting from application of the load $P$, is obtained from Equation 23.11. If the four strain gages on the beam are assumed to be identical, then Equation 23.11 holds.

$$
E_o = \frac{6S_i Px E_i}{Ebh^2}
$$

(23.11)

The range and sensitivity of a beam-type load cell depends on the shape of the cross-section of the beam, the location of the point of application of the load, and the fatigue strength of the material from which the beam is fabricated.

Ring-Type Load Cell

Ring-type load cells incorporate a proving ring (see Figure 23.7) as the elastic element. The ring element can be designed to cover a very wide range of loads by varying the diameter $D$, the thickness $t$, or the depth $w$ of the ring. Either strain gages or a linear variable-differential transformer (LVDT) can be used as the sensor.

The load $P$ is linearly proportional to the output voltage $E_o$. The sensitivity of the ring-type load cell with an LVDT sensor depends on the geometry of the ring ($R$, $t$, and $w$), the material from which the ring is fabricated ($E$), and the characteristics of the LVDT ($S$ and $E_i$). The range of a ring-type load cell is controlled by the strength of the material used in fabricating the ring.
Piezoelectric Methods

A piezoelectric material exhibits a phenomenon known as the piezoelectric effect. This effect states that when asymmetrical, elastic crystals are deformed by a force, an electrical potential will be developed within the distorted crystal lattice. This effect is reversible. That is, if a potential is applied between the surfaces of the crystal, it will change its physical dimensions [4]. Elements exhibiting piezoelectric qualities are sometimes known as electrorestrictive elements.

The magnitude and polarity of the induced surface charges are proportional to the magnitude and direction of the applied force [4]:

\[ Q = dF \]  

(23.12)

where \( d \) is the charge sensitivity (a constant for a given crystal) of the crystal in \( C/N \). The force \( F \) causes a thickness variation \( \Delta t \) meters of the crystal:

\[ F = \frac{aY}{t} \Delta t \]  

(23.13)

where \( a \) is area of crystal, \( t \) is thickness of crystal, and \( Y \) is Young’s modulus.

\[ Y = \frac{\text{stress}}{\text{strain}} = \frac{Ft}{a\Delta t} \]  

(23.14)

The charge at the electrodes gives rise to a voltage \( E_0 = Q/C \), where \( C \) is capacitance in farads between the electrodes and \( C = \varepsilon a/t \) where \( \varepsilon \) is the absolute permittivity.
The voltage sensitivity $E_u = g = d/\varepsilon$ in volt m/N can be obtained as:

$$E_u = \frac{dF}{C} = \frac{d}{\varepsilon} F$$

(23.15)

The voltage sensitivity $g = d/\varepsilon$ in volt m/N can be obtained as:

$$E_u = g = \frac{tF}{a} = gtP$$

(23.16)

The piezoelectric materials used are quartz, tourmaline, Rochelle salt, ammonium dihydrogen phosphate (ADP), lithium sulfate, barium titanate, and lead zirconate titanate (PZT) [4]. Quartz and other earthly piezoelectric crystals are naturally polarized. However, synthetic piezoelectric materials, such as barium titanate ceramic, are made by baking small crystallites under pressure and then placing the resultant material in a strong dc electric field [4]. After that, the crystal is polarized, along the axis on which the force will be applied, to exhibit piezoelectric properties. Artificial piezoelectric elements are free from the limitations imposed by the crystal structure and can be molded into any size and shape. The direction of polarization is designated during their production process.

The different modes of operation of a piezoelectric device for a simple plate are shown in Figure 23.8 [4]. By adhering two crystals together so that their electrical axes are perpendicular, bending moments or torque can be applied to the piezoelectric transducer and a voltage output can be produced (Figure 23.9) [4]. The range of forces that can be measured using piezoelectric transducers are from 1 to 200 kN and at a ratio of $2 \times 10^5$. 

---

FIGURE 23.8  Modes of operation for a simple plate as a piezoelectric device [4].

FIGURE 23.9  Curvature of “twister” and “bender” piezoelectric transducers when voltage applied [4].

$E_u = \frac{dF}{C} = \frac{d}{\varepsilon} F = gtP$
Piezoelectric crystals can also be used in measuring an instantaneous change in the force (dynamic forces). A thin plate of quartz can be used as an electronic oscillator. The frequency of these oscillations will be dominated by the natural frequency of the thin plate. Any distortion in the shape of the plate caused by an external force, alters the oscillation frequency. Hence, a dynamic force can be measured by the change in frequency of the oscillator.

**Resistive Method**

The resistive method employs the fact that when the multiple contact area between semiconducting particles (usually carbon) and the distance between the particles are changed, the total resistance is altered. The design of such transducers yields a very small displacement when a force is applied. A transducer might consist of 2 to 60 thin carbon disks mounted between a fixed and a movable electrode. When a force is applied to the movable electrode and the carbon disks move together by 5 to 250 μm per interface, the transfer function of their resistance against the applied force is approximately hyperbolic, that is, highly nonlinear. The device is also subject to large hysteresis and drift together with a high transverse sensitivity.

In order to reduce hysteresis and drift, rings are used instead of disks. The rings are mounted on an insulated rigid core and prestressed. This almost completely eliminates any transverse sensitivity error. The core’s resonant frequency is high and can occur at a frequency as high as 10 kHz. The possible measuring range of such a transducer is from 0.1 kg to 10 kg. The accuracy and linear sensitivity of this transducer is very poor.

**Inductive Method**

The inductive method utilizes the fact that a change in mechanical stress of a ferromagnetic material causes its permeability to alter. The changes in magnetic flux are converted into induced voltages in the pickup coils as the movement takes place. This phenomenon is known as the Villari effect or magnetostriction. It is known to be particularly strong in nickel–iron alloys.

Transducers utilizing the Villari effect consist of a coil wound on a core of magnetostrictive material. The force to be measured is applied on this core, stressing it and causing a change in its permeability and inductance. This change can be monitored and used for determining the force.

The applicable range for this type of transducer is a function of the cross-sectional area of the core. The accuracy of the device is determined by a calibration process. This transducer has poor linearity and is subject to hysteresis. The permeability of a magnetostrictive material increases when it is subjected to pure torsion, regardless of direction. A flat frequency response is obtained over a wide range from 150 Hz to 15,000 Hz.

**Piezotransistor Method**

Devices that utilize anisotropic stress effects are described as piezotransistors. In this effect, if the upper surface of a p–n diode is subjected to a localized stress, a significant reversible change occurs in the current across the junction. These transistors are usually silicon nonplanar type, with an emitter base junction. This junction is mechanically connected to a diaphragm positioned on the upper surface of a typical TO-type can [4]. When a pressure or a force is applied to the diaphragm, an electronic charge is produced. It is advisable to use these force-measuring devices at a constant temperature by virtue of the fact that semiconducting materials also change their electric properties with temperature variations. The attractive characteristic of piezotransistors is that they can withstand a 500% overload.

**Multicomponent Dynamometers Using Quartz Crystals As Sensing Elements**

The Piezoelectric Effects in Quartz.

For force measurements, the direct piezoelectric effect is utilized. The direct longitudinal effect measures compressive force; the direct shear effect measures shear force in one direction. For example, if a disk of crystalline quartz (SiO₂) cut normally to the crystallographic x-axis is loaded by a compression force, it will yield an electric charge, nominally 2.26 pC/N. If a disk of crystalline quartz is cut normally to the
crystallographic y-axis, it will yield an electric charge (4.52 pC/N) if loaded by a shear force in one specific
direction. Forces applied in the other directions will not generate any output [5].

A charge amplifier is used to convert the charge yielded by a quartz crystal element into a proportional
voltage. The range of a charge amplifier with respect to its conversion factor is determined by a feedback
capacitor. Adjustment to mechanical units is obtained by additional operational amplifiers with variable
gain.

The Design of Quartz Multicomponent Dynamometers.
The main element for designing multicomponent dynamometers is the three-component force transducer
(Figure 23.10). It contains a pair of X-cut quartz disks for the normal force component and a pair of
Y-cut quartz disks (shear-sensitive) for each shear force component.

Three-component dynamometers can be used for measuring cutting forces during machining. Four
three-component force transducers sandwiched between a base plate and a top plate are shown in
Figure 23.10. The force transducer is subjected to a preload as shear forces are transmitted by friction.
The four force transducers experience a drastic change in their load, depending on the type and position
of force application. An overhanging introduction of the force develops a tensile force for some trans-
ducers, thus reducing the preload. Bending of the dynamometer top plate causes bending and shearing
stresses. The measuring ranges of a dynamometer depend not only on the individual forces, but also on
the individual bending stresses.

Measuring Signals Transmitted by Telemetry.
Figure 23.11 shows the newly designed force measuring system RCD (rotating cutting force dynamom-
eter). A ring-shaped sensor (1) is fitted in a steep angle taper socket (2) and a base ring (3) allowing
sensing of the three force components \( F_x, F_y \) and \( F_z \) at the cutting edge as well as the moment \( M_z \). The
physical operating principle of this measuring cell is based on the piezoelectric effect in quartz plates. The quartz plates incorporated in the sensor are aligned so that the maximum cross-sensitivity between the force components is 1%. As a result of the rigid design of the sensor, the resonant frequencies of the force measuring system range from 1200 Hz to 3000 Hz and the measuring ranges cover a maximum of 10 kN [6].

Force-proportional charges produced at the surfaces of the quartz plates are converted into voltages by four miniature charge amplifiers (7) in hybrid construction. These signals are then filtered by specific electrical circuitry to prevent aliasing effects, and digitized with 8 bit resolution using a high sampling rate (pulse-code modulation). The digitized signals are transmitted by a telemetric unit consisting of a receiver and transmitter module, an antenna at the top of the rotating force measuring system (8), as well as a fixed antenna (9) on the splash cover of the two-axis milling head (10). The electrical components, charge amplifier, and transmitter module are mounted on the circumference of the force measuring system [6].

The cutting forces and the moment measured are digitized with the force measuring system described above. They are modulated on an FM carrier and transmitted by the rotating transmitter to the stationary receiver. The signals transmitted are fed to an external measured-variable conditioning unit.

Measuring Dynamic Forces.

Any mechanical system can be considered in the first approximation as a weakly damped oscillator consisting of a spring and a mass. If a mechanical system has more than one resonant frequency, the lowest one must be taken into consideration. As long as the test frequency remains below 10% of the resonant frequency of the reference transducer (used for calibration), the difference between the dynamic sensitivity obtained from static calibration will be less than 1%. The above considerations assume a sinusoidal force signal. The static calibration of a reference transducer is also valid for dynamic calibration purposes if the test frequency is much lower (at least 10 times lower) than the resonant frequency of the system.

**Capacitive Force Transducer**

A transducer that uses capacitance variation can be used to measure force. The force is directed onto a membrane whose elastic deflection is detected by a capacitance variation. A highly sensitive force transducer can be constructed because the capacitive transducer senses very small deflections accurately. An electronic circuit converts the capacitance variations into dc-voltage variations [7].

The capacitance sensor illustrated in Figure 23.12 consists of two metal plates separated by an air gap. The capacitance C between terminals is given by the expression:

\[ C = \varepsilon \varepsilon_r \frac{A}{h} \]  

(23.17)
where 
\[ C = \text{Capacitance in farads (F)} \]
\[ \varepsilon_0 = \text{Dielectric constant of free space} \]
\[ \varepsilon_r = \text{Relative dielectric constant of the insulator} \]
\[ A = \text{Overlapping area for the two plates} \]
\[ h = \text{Thickness of the gap between the two plates} \]

The sensitivity of capacitance-type sensors is inherently low. Theoretically, decreasing the gap \( h \) should increase the sensitivity; however, there are practical electrical and mechanical conditions that preclude high sensitivities. One of the main advantages of the capacitive transducer is that moving of one of its plate relative to the other requires an extremely small force to be applied. A second advantage is stability and the sensitivity of the sensor is not influenced by pressure or temperature of the environment.

**Force Sensing Resistors (Conductive Polymers)**

Force sensing resistors (FSRs) utilize the fact that certain polymer thick-film devices exhibit decreasing resistance with the increase of an applied force. A force sensing resistor is made up of two parts. The first is a resistive material applied to a film. The second is a set of digitating contacts applied to another film. Figure 23.13 shows this configuration. The resistive material completes the electrical circuit between the two sets of conductors on the other film. When a force is applied to this sensor, a better connection is made between the contacts; hence, the conductivity is increased. Over a wide range of forces, it turns out that the conductivity is approximately a linear function of force. Figure 23.14 shows the resistance of the sensor as a function of force. It is important to note that there are three possible regions for the sensor to operate. The first abrupt transition occurs somewhere in the vicinity of 10 g of force. In this
region, the resistance changes very rapidly. This behavior is useful when one is designing switches using force sensing resistors.

FSRs should not be used for accurate measurements of force because sensor parts may exhibit 15% to 25% variation in resistance between each other. However, FSRs exhibit little hysteresis and are considered far less costly than other sensing devices. Compared to piezofilm, the FSR is far less sensitive to vibration and heat.

**Magnetoresistive Force Sensors**

The principle of magnetoresistive force sensors is based on the fact that metals, when cooled to low temperatures, show a change of resistivity when subjected to an applied magnetic field. Bismuth, in particular, is quite sensitive in this respect. In practice, these devices are severely limited because of their high sensitivity to ambient temperature changes.

**Magnetoelastic Force Sensors**

Magnetoelastic transducer devices operate based on the Joule effect; that is, a ferromagnetic material is dimensionally altered when subjected to a magnetic field. The principle of operation is as follows: Initially, a current pulse is applied to the conductor within the waveguide. This sets up a magnetic field circumference-wise around the waveguide over its entire length. There is another magnetic field generated by the permanent magnet that exists only where the magnet is located. This field has a longitudinal component. These two fields join vectorally to form a helical field near the magnet which, in turn, causes the waveguide to experience a minute torsional strain or twist only at the location of the magnet. This twist effect is known as the Wiedemann effect [8].

Magnetoelastic force transducers have a high frequency response (on the order of 20 kHz). Some of the materials that exhibit magnetoelastic include Monel metal, Permalloy, Cekas, Alfer, and a number of nickel–iron alloys. Disadvantages of these transducers include: (1) the fact that excessive stress and aging may cause permanent changes, (2) zero drift and sensitivity changes due to temperature sensitivity, and (3) hysteresis errors.

**Torsional Balances**

Balancing devices that utilize the deflection of a spring may also be used to determine forces. Torsional balances are equal arm scale force measuring devices. They are comprised of horizontal steel bands instead of pivots and bearings. The principle of operation is based on force application on one of the arms that will deflect the torsional spring (within its design limits) in proportion to the applied force. This type of instrument is susceptible to hysteresis and temperature errors and therefore is not used for precise measurements.

**Tactile Sensors**

Tactile sensors are usually interpreted as a touch sensing technique. Tactile sensors cannot be considered as simple touch sensors, where very few discrete force measurements are made. In tactile sensing, a force “distribution” is measured using a closely spaced array of force sensors.

Tactile sensing is important in both grasping and object identification operations. Grasping an object must be done in a stable manner so that the object is not allowed to slip or damaged. Object identification includes recognizing the shape, location, and orientation of a product, as well as identifying surface properties and defects. Ideally, these tasks would require two types of sensing [9]:

1. Continuous sensing of contact forces
2. Sensing of the surface deformation profile

These two types of data are generally related through stress–strain relations of the tactile sensor. As a result, almost continuous variable sensing of tactile forces (the sensing of the tactile deflection profile) is achieved.
Tactile Sensor Requirements.

Significant advances in tactile sensing are taking place in the robotics area. Applications include automated inspection of surface profiles, material handling or parts transfer, parts assembly, and parts identification and gaging in manufacturing applications and fine-manipulation tasks. Some of these applications may need only simple touch (force–torque) sensing if the parts being grasped are properly oriented and if adequate information about the process is already available.

Naturally, the main design objective for tactile sensing devices has been to mimic the capabilities of human fingers [9]. Typical specifications for an industrial tactile sensor include:

1. Spatial resolution of about 2 mm
2. Force resolution (sensitivity) of about 2 g
3. Maximum touch force of about 1 kg
4. Low response time of 5 ms
5. Low hysteresis
6. Durability under extremely difficult working conditions
7. Insensitivity to change in environmental conditions (temperature, dust, humidity, vibration, etc.)
8. Ability to monitor slip

Tactile Array Sensor.

Tactile array sensors (Figure 23.15) consist of a regular pattern of sensing elements to measure the distribution of pressure across the finger tip of a Robot. The $8 \times 8$ array of elements at 2 mm spacing in each direction, provides 64 force sensitive elements. Table 23.1 outlines some of the characteristics of early tactile array sensors. The sensor is composed of two crossed layers of copper strips separated by strips of thin silicone rubber. The sensor forms a thin, compliant layer that can be easily attached to a variety of finger-tip shapes and sizes. The entire array is sampled by computer.

A typical tactile sensor array can consist of several sensing elements. Each element or taxel (Figure 23.16) is used to sense the forces present. Since tactile sensors are implemented in applications where sensitivity providing semblance to human touch is desired, an elastomer is utilized to mimic the human skin. The elastomer is generally a conductive material whose electrical conductivity changes locally when pressure is applied. The sensor itself consists of three layers: a protective covering, a sheet of conductive elastomer, and a printed circuit board. The printed circuit board consists of two rows of two “bullseyes,” each with conductive inner and outer rings that compromise the taxels of the sensor. The outer rings are connected together and to a column-select transistor. The inner rings are connected to diodes (D) in Figure 23.16. Once the column in the array is selected, the current flows through the diodes, through the elastomer, and thence through a transistor to ground. As such, it is generally not possible to excite just one taxel because the pressure applied causes a local deformation in neighboring taxels. This situation is called crosstalk and is eliminated by the diodes [10].

Tactile array sensor signals are used to provide information about the contact kinematics. Several feature parameters, such as contact location, object shape, and the pressure distribution, can be obtained.
The general layout of a sensor array system can be seen in Figure 23.17. An example of this is a contact and force sensing finger. This tactile finger has four contact sensors made of piezoelectric polymer strips on the surface of the fingertip that provide dynamic contact information. A strain gage force sensor provides static grasp force information.
References


Further Information


24.1 Fundamental Concepts
Angular Displacement, Velocity, and Acceleration • Force, Torque, and Equilibrium • Stress, Rigidity, and Strain • Work, Energy, and Power

24.2 Arrangements of Apparatus for Torque and Power Measurement

24.3 Torque Transducer Technologies
Surface Strain • Twist Angle • Stress

24.4 Torque Transducer Construction, Operation, and Application
Mechanical Considerations • Electrical Considerations • Costs and Options

24.5 Apparatus for Power Measurement
Absorption Dynamometers • Driving and Universal Dynamometers • Measurement Accuracy • Costs

Torque, speed, and power are the defining mechanical variables associated with the functional performance of rotating machinery. The ability to accurately measure these quantities is essential for determining a machine's efficiency and for establishing operating regimes that are both safe and conducive to long and reliable services. On-line measurements of these quantities enable real-time control, help to ensure consistency in product quality, and can provide early indications of impending problems. Torque and power measurements are used in testing advanced designs of new machines and in the development of new machine components. Torque measurements also provide a well-established basis for controlling and verifying the tightness of many types of threaded fasteners. This chapter describes the basic concepts as well as the various methods and apparatus in current use for the measurement of torque and power; the measurement of speed, or more precisely, angular velocity, is discussed elsewhere in this handbook [1].

24.1 Fundamental Concepts

Angular Displacement, Velocity, and Acceleration

The concept of rotational motion is readily formalized: all points within a rotating rigid body move in parallel or coincident planes while remaining at fixed distances from a line called the axis. In a perfectly rigid body, all points also remain at fixed distances from each other. Rotation is perceived as a change in the angular position of a reference point on the body, i.e., as its angular displacement, $\Delta \theta$, over some time interval, $\Delta t$. The motion of that point, and therefore of the whole body, is characterized by its
clockwise (CW) or counterclockwise (CCW) direction and by its angular velocity, \( \omega = \frac{\Delta \theta}{\Delta t} \). If during a time interval \( \Delta t \), the velocity changes by \( \Delta \omega \), the body is undergoing an angular acceleration, \( \alpha = \frac{\Delta \omega}{\Delta t} \).

With angles measured in radians, and time in seconds, units of \( \omega \) become radians per second (rad s\(^{-1}\)) and of \( \alpha \), radians per second per second (rad s\(^{-2}\)). Angular velocity is often referred to as rotational speed and measured in numbers of complete revolutions per minute (rpm) or per second (rps).

### Force, Torque, and Equilibrium

Rotational motion, as with motion in general, is controlled by forces in accordance with Newton's laws. Because a force directly affects only that component of motion in its line of action, forces or components of forces acting in any plane that includes the axis produce no tendency for rotation about that axis. Rotation can be initiated, altered in velocity, or terminated only by a tangential force \( F_t \) acting at a finite radial distance \( l \) from the axis. The effectiveness of such forces increases with both \( F_t \) and \( l \); hence, their product, called a moment, is the activating quantity for rotational motion. A moment about the rotational axis constitutes a torque.

Figure 24.1(a) shows a force \( F \) acting at an angle \( \beta \) to the tangent at a point \( P \), distant \( l \) (the moment arm) from the axis. The torque \( T \) is found from the tangential component of \( F \) as:

\[
T = F_l = (F \cos \beta)l
\]  

(24.1)

The combined effect, known as the resultant, of any number of torques acting at different locations along a body is found from their algebraic sum, wherein torques tending to cause rotation in CW and CCW directions are assigned opposite signs. Forces, hence torques, arise from physical contact with other solid bodies, motional interaction with fluids, or via gravitational (including inertial), electric, or magnetic force fields. The source of each such torque is subjected to an equal, but oppositely directed, reaction torque. With force measured in newtons and distance in meters, Equation 24.1 shows the unit of torque to be a Newton meter (N·m).

A nonzero resultant torque will cause the body to undergo a proportional angular acceleration, found, by application of Newton's second law, from:

\[
T = I \alpha
\]

(24.2)

where \( I \), having units of kilogram meter\(^2\) (kg m\(^2\)), is the moment of inertia of the body around the axis (i.e., its polar moment of inertia). Equation 24.2 is applicable to any body regardless of its state of motion.
When \( \alpha = 0 \), Equation 24.2 shows that \( T_r \) is also zero; the body is said to be in equilibrium. For a body to be in equilibrium, there must be either more than one applied torque, or none at all.

**Stress, Rigidity, and Strain**

Any portion of a rigid body in equilibrium is also in equilibrium; hence, as a condition for equilibrium of the portion, any torques applied thereto from external sources must be balanced by equal and directionally opposite internal torques from adjoining portions of the body. Internal torques are transmitted between adjoining portions by the collective action of stresses over their common cross-sections. In a solid body having a round cross-section (e.g., a typical shaft), the shear stress \( \tau \) varies linearly from zero at the axis to a maximum value at the surface. The shear stress, \( \tau_m \), at the surface of a shaft of diameter, \( d \), transmitting a torque, \( T \), is found from:

\[
\tau_m = \frac{16T}{\pi d^3}
\]  

(24.3)

Real materials are not perfectly rigid but have instead a modulus of rigidity, \( G \), which expresses the finite ratio between \( \tau \) and shear strain, \( \gamma \). The maximum strain in a solid round shaft therefore also exists at its surface and can be found from:

\[
\gamma_m = \frac{\tau_m}{G} = \frac{16T}{\pi d^3 G}
\]  

(24.4)

Figure 24.1(b) shows the manifestation of shear strain as an angular displacement between axially separated cross-sections. Over the length \( L \), the solid round shaft shown will be twisted by the torque through an angle \( \phi \) found from:

\[
\phi = \frac{32LT}{\pi d^3 G}
\]  

(24.5)

**Work, Energy, and Power**

If during the time of application of a torque, \( T \), the body rotates through some angle \( \theta \), mechanical work:

\[
W = T\theta
\]  

(24.6)

is performed. If the torque acts in the same CW or CCW sense as the displacement, the work is said to be done on the body, or else it is done by the body. Work done on the body causes it to accelerate, thereby appearing as an increase in kinetic energy (\( KE = \frac{1}{2}I\omega^2 \)). Work done by the body causes deceleration with a corresponding decrease in kinetic energy. If the body is not accelerating, any work done on it at one location must be done by it at another location. Work and energy are each measured in units called a joule (J). Equation 24.6 shows that 1 J is equivalent to 1 N·m rad, which, since a radian is a dimensionless ratio, \( = 1 \text{ N·m} \). To avoid confusion with torque, it is preferable to quantify mechanical work in units of m·N, or better yet, in J.

The rate at which work is performed is termed power, \( P \). If a torque \( T \) acts over a small interval of time \( \Delta t \), during which there is an angular displacement \( \Delta \theta \), work equal to \( T\Delta \theta \) is performed at the rate \( T\Delta \theta/\Delta t \). Replacing \( \Delta \theta/\Delta t \) by \( \omega \), power is found simply as:

\[
P = T\omega
\]  

(24.7)
The unit of power follows from its definition and is given the special name watt (W). 1 W = 1 J s\(^{-1}\) = 1 m·N s\(^{-1}\). Historically, power has also been measured in horsepower (Hp), where 1 Hp = 746 W. Rotating bodies effectively transmit power between locations where torques from external sources are applied.

24.2 Arrangements of Apparatus for Torque and Power Measurement

Equations 24.1 through 24.7 express the physical bases for torque and power measurement. Figure 24.2 illustrates a generalized measurement arrangement. The actual apparatus used is selected to fulfill the specific measurement purposes. In general, a driving torque originating within a device at one location (B in Figure 24.2), is resisted by an opposing torque developed by a different device at another location (F). The driving torque (from, e.g., an electric motor, a gasoline engine, a steam turbine, muscular effort, etc.) is coupled through connecting members C, transmitting region D, and additional couplings E, to the driven device (an electric generator, a pump, a machine tool, mated threaded fasteners, etc.) within which the resisting torque is met at F. The torque at B or F is the quantity to be measured. These torques may be indirectly determined from a correlated physical quantity, e.g., an electrical current or fluid pressure associated with the operation of the driving or driven device, or more directly by measuring either the reaction torque at A or G, or the transmitted torque through D. It follows from the cause-and-effect relationship between torque and rotational motion that most interest in transmitted torque will involve rotating bodies.

To the extent that the frames of the driving and driven devices and their mountings to the “Earth” are perfectly rigid, the reaction at A will at every instant equal the torque at B, as will the reaction at G equal the torque at F. Under equilibrium conditions, these equalities are independent of the compliance of any member. Also under equilibrium conditions, and except for usually minor parasitic torques (due, e.g., to bearing friction and air drag over rapidly moving surfaces), the driving torque at B will equal the resisting torque at F.

Reaction torque at A or G is often determined, using Equation 24.1, from measurements of the forces acting at known distances fixed by the apparatus. Transmitted torque is determined from measurements, on a suitable member within region D, of \(\tau_m\), \(\gamma_m\), or \(\phi\) and applying Equations 24.3, 24.4, or 24.5 (or analogous expressions for members having other than solid round cross-sections [2]). Calibration, the measurement of the stress, strain, or twist angle resulting from the application of a known torque, makes it unnecessary to know any details about the member within D. When \(\alpha \neq 0\), and is measurable, \(T\) may also be determined from Equation 24.2. Requiring only noninvasive, observational measurements, this method is especially useful for determining transitory torques; for example those associated with firing events in multicylinder internal combustion engines [3].

Equations 24.6 and 24.7 are applicable only during rotation because, in the absence of motion, no work is done and power transfer is zero. Equation 24.6 can be used to determine average torque from calorimetric...
measurements of the heat generated (equal to the mechanical work \( W \)) during a totalized number of revolutions \( (\equiv \theta / 2\pi) \). Equation 24.7 is routinely applied in power measurement, wherein \( T \) is determined by methods based on Equations 24.1, 24.3, 24.4, or 24.5, and \( \omega \) is measured by any suitable means [4].

\( F, T, \) and \( \phi \) are sometimes measured by simple mechanical methods. For example, a “torque wrench” is often used for the controlled tightening of threaded fasteners. In these devices, torque is indicated by the position of a needle moving over a calibrated scale in response to the elastic deflection of a spring member, in the simplest case, the bending of the wrench handle [5]. More generally, instruments, variously called sensors or transducers, are used to convert the desired (torque or speed related) quantity into a linearly proportional electrical signal. (Force sensors are also known as load cells.) The determination of \( P \) most usually requires multiplication of the two signals from separate sensors of \( T \) and \( \omega \). A transducer, wherein the amplitude of a single signal proportional to the power being transmitted along a shaft, has also been described [6].

24.3 Torque Transducer Technologies

Various physical interactions serve to convert \( F, \tau, \gamma, \) or \( \phi \) into proportional electrical signals. Each requires that some axial portion of the shaft be dedicated to the torque sensing function. Figure 24.3 shows typical features of sensing regions for four sensing technologies in present use.

Surface Strain

Figure 24.3(a) illustrates a sensing region configured to convert surface strain \( (\gamma_m) \) into an electric signal proportional to the transmitted torque. Surface strain became the key basis for measuring both force and torque following the invention of bonded wire strain gages by E. E. Simmons, Jr. and Arthur C. Ruge in 1938 [7]. A modern strain gage consists simply of an elongated electrical conductor, generally formed in a serpentine pattern in a very thin foil or film, bonded to a thin insulating carrier. The carrier is attached, usually with an adhesive, to the surface of the load carrying member. Strain is sensed as a change in gage resistance. These changes are generally too small to be accurately measured directly and so it is common to employ two to four gages arranged in a Wheatstone bridge circuit. Independence from axial and bending loads as well as from temperature variations are obtained by using a four-gage bridge comprised of two diametrically opposite pairs of matched strain gages, each aligned along a principal strain direction. In round shafts (and other shapes used to transmit torque), tensile and compressive principal strains occur at 45° angles to the axis. Limiting strains, as determined from Equation 24.4 (with \( \tau_m \) equal to the shear proportional limit of the shaft material), rarely exceed a few parts in 10³. Typical practice is to increase the compliance of the sensing region (e.g., by reducing its diameter or with hollow or specially shaped sections) in order to attain the limiting strain at the highest value of the torque to be measured. This maximizes the measurement sensitivity.

Twist Angle

If the shaft is slender enough (e.g., \( L > 5 d \)) \( \phi \), at limiting values of \( \tau_m \) for typical shaft materials, can exceed 1°, enough to be resolved with sufficient accuracy for practical torque measurements \( (\phi \text{ at } \tau_m \text{ can be found by manipulating Equations 24.3, 24.4, and 24.5}) \). Figure 24.3(b) shows a common arrangement wherein torque is determined from the difference in tooth-space phasing between two identical “toothed” wheels attached at opposite ends of a compliant “torsion bar.” The phase displacement of the periodic electrical signals from the two “pickups” is proportional to the peripheral displacement of salient features on the two wheels, and hence to the twist angle of the torsion bar and thus to the torque. These features are chosen to be sensible by any of a variety of noncontacting magnetic, optical, or capacitive techniques. With more elaborate pickups, the relative angular position of the two wheels appears as the amplitude of a single electrical signal, thus providing for the measurement of torque even on a stationary shaft (e.g., [13–15]). In still other constructions, a shaft-mounted variable displacement transformer or a related type of electric device is used to provide speed independent output signals proportional to \( \phi \).
Stress

In addition to elastic strain, the stresses by which torque is transmitted are manifested by changes in the magnetic properties of ferromagnetic shaft materials. This “magnetoelectric interaction” [8] provides an inherently noncontacting basis for measuring torque. Two types of magnetoelectric (sometimes called magnetostrictive) torque transducers are in present use: Type 1 derive output signals from torque-induced variations in magnetic circuit permeances; Type 2 create a magnetic field in response to torque. Type 1 transducers typically employ “branch,” “cross,” or “solenoidal” constructions [9]. In branch and cross designs, torque is detected as an imbalance in the permeabilities along orthogonal 45° helical paths (the principal stress directions) on the shaft surface or on the surface of an ad hoc material attached to the shaft. In solenoidal constructions torque is detected by differences in the axial permeabilities of two adjacent surface regions, preened with symmetrical magnetic “easy” axes (typically along the 45° principal stress directions). While branch and cross type sensors are readily miniaturized [10], local variations in magnetic properties of typical shaft surfaces limit their accuracy. Solenoidal designs, illustrated in Figure 24.3(c), avoid this pitfall by effectively averaging these variations. Type 2 transducers are generally constructed with a ring of magnetoelectrically active material rigidly attached to the shaft. The ring is magnetized during manufacture of the transducer, usually with each axial half polarized in an

---

**FIGURE 24.3** Four techniques in present use for measuring transmitted torque. (a) Torsional strain in the shaft alters the electrical resistance for four strain gages (two not seen) connected in a Wheatstone bridge circuit. In the embodiment shown, electrical connections are made to the bridge through slip rings and brushes. (b) Twist of the torsion section causes angular displacement of the surface features on the toothed wheels. This creates a phase difference in the signals from the two pickups. (c) The permeabilities of the two grooved regions of the shaft change oppositely with torsional stress. This is sensed as a difference in the output voltages of the two sense windings. (d) Torsional stress causes the initially circumferential magnetizations in the ring (solid arrows) to tilt (dashed arrows). These helical magnetizations cause magnetic poles to appear at the domain wall and ring ends. The resulting magnetic field is sensed by the field sensor.
opposite circumferential direction as indicated by the solid arrows in Figure 24.3(d) [11]. When torque is applied, the magnetizations tilt into helical directions (dashed arrows), causing magnetic poles to develop at the central domain wall and (of opposite polarity) at the ring end faces. Torque is determined from the output signal of one or more magnetic field sensors (e.g., Hall effect, magnetoresistive, or flux gate devices) mounted so as to sense the intensity and polarity of the magnetic field that arises in the space near the ring.

24.4 Torque Transducer Construction, Operation, and Application

Although a torque sensing region can be created directly on a desired shaft, it is more usual to install a preassembled modular torque transducer into the driveline. Transducers of this type are available with capacities from 0.001 N·m to 200,000 N·m. Operating principle descriptions and detailed installation and operating instructions can be found in the catalogs and literature of the various manufacturers [12–20]. Tradenames often identify specific type of transducers; for example, Torquemeters [13] refers to a family of noncontact strain gage models; Torkducer® [18] identifies a line of Type 1 magnetoelastic transducers; Torqstar™ [12] identifies a line of Type 2 magnetoelastic transducers; Torquetronic [16] is a class of transducers using wrap-around twist angle sensors; and TorXimitor™ [20] identifies optoelectronic based, noncontact, strain gage transducers. Many of these devices show generic similarities transcending their specific sensing technology as well as their range. Figure 24.4 illustrates many of these common features.

Mechanical Considerations

Maximum operating speeds vary widely; upper limits depend on the size, operating principle, type of bearings, lubrication, and dynamic balance of the rotating assembly. Ball bearings, lubricated by grease, oil, or oil mist, are typical. Parasitic torques associated with bearing lubricants and seals limit the accuracy of low-end torque measurements. (Minute capacity units have no bearings [15]). Forced lubrication can
allow operation up to 80,000 rpm [16]. High-speed operation requires careful consideration of the effects of centrifugal stresses on the sensed quantity as well as of critical (vibration inducing) speed ranges. Torsional oscillations associated with resonances of the shaft elasticity (characterized by its spring constant) with the rotational inertia of coupled masses can corrupt the measurement, damage the transducer by dynamic excursions above its rated overload torque, and even be physically dangerous.

Housings either float on the shaft bearings or are rigidly mounted. Free floating housings are restrained from rotating by such "soft" means as a cable, spring, or compliant bracket, or by an eccentric external feature simply resting against a fixed surface. In free floating installations, the axes of the driving and driven shafts must be carefully aligned. Torsionally rigid "flexible" couplings at each shaft end are used to accommodate small angular and/or radial misalignments. Alternatively, the use of dual flexible couplings at one end will allow direct coupling of the other end. Rigidly mounted housings are equipped with mounting feet or lugs similar to those found on the frame of electric motors. Free-floating models are sometimes rigidly mounted using adapter plates fastened to the housing. Rigid mountings are preferred when it is difficult or impractical to align the driving and driven shafts, as for example when driving or driven machines are changed often. Rigidly mounted housings require the use of dual flexible couplings at both shaft ends.

Modular transducers designed for zero or limited rotation applications have no need for bearings. To ensure that all of the torque applied at the ends is sensed, it is important in such "reaction"-type torque transducers to limit attachment of the housing to the shaft to only one side of the sensing region. Whether rotating or stationary, the external shaft ends generally include such torque coupling details as flats, keyways, splines, tapers, flanges, male/female squares drives, etc.

**Electrical Considerations**

By their very nature, transducers require some electrical input power or excitation. The "raw" output signal of the actual sensing device also generally requires "conditioning" into a level and format appropriate for display on a digital or analog meter or to meet the input requirements of data acquisition equipment. Excitation and signal conditioning are supplied by electronic circuits designed to match the characteristics of the specific sensing technology. For example, strain gage bridges are typically powered with 10 V to 20 V (dc or ac) and have outputs in the range of 1.5 mV to 3.0 mV per volt of excitation at the rated load. Raising these millivolt signals to more usable levels requires amplifiers having gains of 100 or more. With ac excitation, oscillators, demodulators (or rectifiers) are also needed. Circuit elements of these types are normal when inductive elements are used either as a necessary part of the sensor or simply to implement noncontact constructions.

Strain gages, differential transformers, and related sensing technologies require that electrical components be mounted on the torqued member. Bringing electrical power to and output signals from these components on rotating shafts require special methods. The most direct and common approach is to use conductive means wherein brushes (typically of silver graphite) bear against (silver) slip rings. Useful life is extended by providing means to lift the brushes off the rotating rings when measurements are not being made. Several "noncontacting" methods are also used. For example, power can be supplied via inductive coupling between stationary and rotating transformer windings [12–15], by the illumination of shaft mounted photovoltaic cells [20], or even by batteries strapped to the shaft [21] (limited by centrifugal force to relatively low speeds). Output signals are coupled off the shaft through rotary transformers, by frequency-modulated (infrared) LEDs [19, 20], or by radio-frequency (FM) telemetry [21]. Where shaft rotation is limited to no more than a few full rotations, as in steering gear, valve actuators or oscillating mechanisms, hard wiring both power and signal circuits is often suitable. Flexible cabling minimizes incidental torques and makes for a long and reliable service life. All such wiring considerations are avoided when noncontact technologies or constructions are used.
Costs and Options

Prices of torque transducers reflect the wide range of available capacities, performance ratings, types, styles, optional features, and accessories. In general, prices of any one type increase with increasing capacity. Reaction types cost about half of similarly rated rotating units. A typical foot-mounted, 565 N·m capacity, strain gage transducer with either slip rings or rotary transformers and integral speed sensor, specified nonlinearity and hysteresis each within ±0.1%, costs about $4000 (1997). Compatible instrumentation providing transducer excitation, conditioning, and analog output with digital display of torque and speed costs about $2000. A comparable magnetoelastic transducer with ±0.5% accuracy costs about $1300. High-capacity transducers for extreme speed service with appropriate lubrication options can cost more than $50,000. Type 2 magnetoelastic transducers, mass produced for automotive power steering applications, cost approximately $10.

24.5 Apparatus for Power Measurement

Rotating machinery exists in specific types without limit and can operate at power levels from fractions of a watt to some tens of megawatts, a range spanning more than 10⁸. Apparatus for power measurement exists in a similarly wide range of types and sizes. Mechanical power flows from a driver to a load. This power can be determined directly by application of Equation 24.7, simply by measuring, in addition to \( \omega \), the output torque of the driver or the input torque to the load, whichever is the device under test (DUT). When the DUT is a driver, measurements are usually required over its full service range of speed and torque. The test apparatus therefore must act as a controllable load and be able to absorb the delivered power. Similarly, when the DUT is a pump or fan or other type of load, or one whose function is simply to alter speed and torque (e.g., a gear box), the test apparatus must include a driver capable of supplying power over the DUT’s full rated range of torque and speed. Mechanical power can also be determined indirectly by conversion into (or from) another form of energy (e.g., heat or electricity) and measuring the relevant calorimetric or electrical quantities. In view of the wide range of readily available methods and apparatus for accurately measuring both torque and speed, indirect methods need only be considered when special circumstances make direct methods difficult.

Dynamometer is the special name given to the power-measuring apparatus that includes absorbing or/and driving means and wherein torque is determined by the reaction forces on a stationary part (the stator). An effective dynamometer is conveniently assembled by mounting the DUT in such a manner as to allow measurement of the reaction torque on its frame. Figure 24.5 shows a device designed to facilitate such measurements. Commercial models (Torque Table® [12]) rated to support DUTs weighing 222 N to 4900 N are available with torque capacities from 1.3 N·m 226 to N·m. “Torque tubes” [4] or other DUT mounting arrangements are also used. Other than for possible rotational/elastic resonances, these systems have no speed limitations. More generally, and especially for large machinery, dynamometers include a specialized driving or absorbing machine. Such dynamometers are classified according to their function as absorbing or driving (sometimes motoring). A universal dynamometer can function as either a driver or an absorber.

Absorption Dynamometers

Absorption dynamometers, often called brakes because their operation depends on the creation of a controllable drag torque, convert mechanical work into heat. A drag torque, as distinguished from an active torque, can act only to restrain and not to initiate rotational motion. Temperature rise within a dynamometer is controlled by carrying away the heat energy, usually by transfer to a moving fluid, typically air or water. Drag torque is created by inherently dissipative processes such as: friction between rubbing surfaces, shear or turbulence of viscous liquids, the flow of electric current, or magnetic hysteresis. Gaspard Riche de Prony (1755–1839), in 1821 [22], invented a highly useful form of a friction brake to meet the needs for testing the steam engines that were then becoming prevalent. Brakes of this
The power that would normally be delivered by the shaft of the driving engine to the driven load is (for measurement purposes) converted instead into heat via the work done by the frictional forces between the friction blocks and the flywheel rim. Adjusting the tightness of the friction blocks changes the frictional force and thus the power absorbed.

**Figure 24.5** Support system for measuring the reaction torque of a rotating machine. The axis of the machine must be accurately set on the “center of rotation.” The holes and keyway in the table facilitate machine mounting and alignment. Holes in the front upright provide for attaching a lever arm from which calibrating weights may be hung [4, 11].

**Figure 24.6** A classical prony brake. This brake embodies the defining features of all absorbing dynamometers: conversion of mechanical work into heat and determination of power from measured values of reaction torque and rotational velocity.

type are often used for instructional purposes, for they embody the general principles and major operating considerations for all types of absorption dynamometers. Figure 24.6 shows the basic form and constructional features of a prony brake. The power that would normally be delivered by the shaft of the driving engine to the driven load is (for measurement purposes) converted instead into heat via the work done by the frictional forces between the friction blocks and the flywheel rim. Adjusting the tightness of the
clamping bolts varies the frictional drag torque as required. Heat is removed from the inside surface of the rim by arrangements (not shown) utilizing either a continuous flow or evaporation of water. There is no need to know the magnitude of the frictional forces nor even the radius of the flywheel (facts recognized by Prony), because, while the drag torque tends to rotate the clamped-on apparatus, it is held stationary by the equal but opposite reaction torque $F_r$. $F_r$ at the end of the torque arm of radius $r$ (a fixed dimension of the apparatus) is monitored by a scale or load cell. The power is found from Equations 24.1 and 24.7 as $P = F_r \omega = F_r \frac{2\pi N}{60}$ where $N$ is in rpm.

Uneven retarding forces associated with fluctuating coefficients of friction generally make rubbing friction a poor way to generate drag torque. Nevertheless, because they can be easily constructed, ad hoc variations of prony brakes, often using only bare ropes or wooden cleats connected by ropes or straps, find use in the laboratory or wherever undemanding or infrequent power measurements are to be made. More sophisticated prony brake constructions are used in standalone dynamometers with self-contained cooling water tanks in sizes up to 746 kW (1000 Hp) for operation up to 3600 rpm with torques to 5400 N·m [23]. Available in stationary and mobile models, they find use in testing large electric motors as well as engines and transmissions on agricultural vehicles. Prony brakes allow full drag torque to be imposed down to zero speed.

William Froude (1810–1879) [24] invented a water brake (1877) that does not depend on rubbing friction. Drag torque within a Froude brake is developed between the rotor and the stator by the momentum imparted by the rotor to water contained within the brake casing. Rotor rotation forces the water to circulate between cup-like pockets cast into facing surfaces of both rotor and stator. The rotor is supported in the stator by bearings that also fix its axial position. Labyrinth-type seals prevent water leakage while minimizing frictional drag and wear. The stator casing is supported in the dynamometer frame in cradle fashion by trunnion bearings. The torque that prevents rotation of the stator is measured by reaction forces in much the same manner as with the prony brake. Drag torque is adjusted by a valve, controlling either the back pressure in the water outlet piping [25] or the inlet flow rate [26] or sometimes (to allow very rapid torque changes) with two valves controlling both [27]. In any case, the absorbed energy is carried away by the continuous water flow. Other types of cradle-mounted water brakes, while externally similar, have substantially different internal constructions and depend on other principles for developing the drag torque (e.g., smooth rotors develop viscous drag by shearing and turbulence). Nevertheless, all hydraulic dynamometers purposefully function as inefficient centrifugal pumps. Regardless of internal design and valve settings, maximum drag torque is low at low speeds (zero at standstill) but can rise rapidly, typically varying with the square of rotational speed. The irreducible presence of some water, as well as windage, places a speed-dependent lower limit on the controllable drag torque. In any one design, wear and vibration caused by cavitation place upper limits on the speed and power level. Hydraulic dynamometers are available in a wide range of capacities between 300 kW and 25,000 kW, with some portable units having capacities as low as 75 kW [26]. The largest ever built [27], absorbing up to about 75,000 kW (100,000 Hp), has been used to test propulsion systems for nuclear submarines. Maximum speeds match the operating speeds of the prime movers that they are built to test and therefore generally decrease with increasing capacity. High-speed gas turbine and aerospace engine test equipment can operate as high as 30,000 rpm [25].

In 1855, Jean B. L. Foucault (1819–1868) [22] demonstrated the conversion of mechanical work into heat by rotating a copper disk between the poles of an electromagnet. This simple means of developing drag torque, based on eddy currents, has, since circa 1935, been widely exploited in dynamometers. Figure 24.7 shows the essential features of this type of brake. Rotation of a toothed or spoked steel rotor through a spatially uniform magnetic field, created by direct current through coils in the stator, induces locally circulating (eddy) currents in electrically conductive (copper) portions of the stator. Electromagnetic forces between the rotor, which is magnetized by the uniform field, and the field arising from the eddy currents, create the drag torque. This torque, and hence the mechanical input power, are controlled by adjusting the excitation current in the stator coils. Electric input power is less than 1% of the rated capacity. The dynamometer is effectively an internally short-circuited generator because the power associated with the resistive losses from the generated eddy currents is dissipated within the machine.
Being heated by the flow of these currents, the stator must be cooled, sometimes (in smaller capacity machines) by air supplied by blowers [23], but more often by the continuous flow of water [25, 27, 28]. In dry gap eddy current brakes (the type shown in Figure 24.7), water flow is limited to passages within the stator. Larger machines are often of the water in gap type, wherein water also circulates around the rotor [28]. Water in contact with the moving rotor effectively acts as in a water brake, adding a nonelectromagnetic component to the total drag torque, thereby placing a lower limit to the controllable torque. Windage limits the minimum value of controllable torque in dry gap types. Since drag torque is developed by the motion of the rotor, it is zero at standstill for any value of excitation current. Initially rising rapidly, approximately linearly, with speed, torque eventually approaches a current limited saturation value. As in other cradled machines, the torque required to prevent rotation of the stator is measured by the reaction force acting at a fixed known distance from the rotation axis. Standard model eddy current brakes have capacities from less than 1 kW [23, 27] to more than 2000 kW [27, 28], with maximum speeds from 12,000 rpm in the smaller capacity units to 3600 rpm in the largest units. Special units with capacities of 3000 Hp (2238 kW) at speeds to 25,000 rpm have been built [28].

Hysteresis brakes [29] develop drag torque via magnetic attractive/repulsive forces between the magnetic poles established in a reticulated stator structure by a current through the field coil, and those created in a "drag cup" rotor by the stator field gradients. Rotation of the special steel rotor, through the spatial field pattern established by the stator, results in a cyclical reversal of the polarity of its local magnetizations. The energy associated with these reversals (proportional to the area of the hysteresis loop of the rotor material) is converted into heat within the drag cup. Temperature rise is controlled by forced air cooling from a blower or compressed air source. As with eddy current brakes, the drag torque of these devices is controlled by the excitation current. In contrast with eddy current brakes, rated drag torque is available down to zero speed. (Eddy current effects typically add only 1% to the drag torque for each 1000 rpm). As a result of their smooth surfaced rotating parts, hysteresis brakes exhibit low parasitic torques and hence cover a dynamic range as high as 200 to 1. Standard models are available having continuous power capacities up to 6 kW (12 kW with two brakes in tandem cooled by two blowers). Intermittent capacities per unit (for 5 min or less) are 7 kW. Some low-capacity units are convection cooled; the smallest has a continuous rating of just 7 W (35 W for 5 min). Maximum speeds range from 30,000 rpm for the smallest to 10,000 rpm for the largest units. Torque is measured by a strain gage bridge on a moment arm supporting the machine stator.
Driving and Universal Dynamometers

Electric generators, both ac and dc, offer another means for developing a controllable drag torque and they are readily adapted for dynamometer service by cradle mounting their stator structures. Moreover, electric machines of these types can also operate in a motoring mode wherein they can deliver controllable active torque. When configured to operate selectively in either driving or absorbing modes, the machine serves as a universal dynamometer. With dc machines in the absorbing mode, the generated power is typically dissipated in a convection-cooled resistor bank. Air cooling the machine with blowers is usually adequate, since most of the mechanical power input is dissipated externally. Nevertheless, all of the mechanical input power is accounted for by the product of the reaction torque and the rotational speed. In the motoring mode, torque and speed are controlled by adjustment of both field and armature currents. Modern ac machines utilize regenerative input power converters to allow braking power to be returned to the utility power line. In the motoring mode, speed is controlled by high-power, solid-state, adjustable frequency inverters. Internal construction is that of a simple three-phase induction motor, having neither brushes, slip rings, nor commutators. The absence of rotor windings allows for higher speed operation than dc machines. Universal dynamometers are "four-quadrant" machines, a term denoting their ability to produce torque in the same or opposite direction as their rotational velocity. This unique ability allows the effective drag torque to be reduced to zero at any speed. Universal dynamometers [25, 28] are available in a relatively limited range of capacities (56 to 450 kW), with commensurate torque (110 to 1900 N·m) and speed (4500 to 13,500 rpm) ranges, reflecting their principal application in automotive engine development. Special dynamometers for testing transmissions and other vehicular drive train components insert the DUT between a diesel engine or electric motor prime mover and a hydraulic or eddy current brake [30].

Measurement Accuracy

Accuracy of power measurement (see discussion in [4]) is generally limited by the torque measurement (±0.25% to ±1%) since rotational speed can be measured with almost any desired accuracy. Torque errors can arise from the application of extraneous (i.e., not indicated) torques from hose and cable connections, from windage of external parts, and from miscalibration of the load cell. Undetected friction in the trunnion bearings of cradled dynamometers can compromise the torque measurement accuracy. Ideally, well-lubricated antifriction bearings make no significant contribution to the restraining torque. In practice, however, the unchanging contact region of the balls or other rolling elements on the bearing races makes them prone to brinelling (a form of denting) from forces arising from vibration, unsupported weight of attached devices, or even inadvertently during the alignment of connected machinery. The problem can be alleviated by periodic rotation of the (primarily outer) bearing races. In some bearing-in-bearing constructions, the central races are continuously rotated at low speeds by an electric motor while still others avoid the problem by supporting the stator on hydrostatic oil lift bearings [28].

Costs

The wide range of torque, speed, and power levels, together with the variation in sophistication of associated instrumentation, is reflected in the very wide range of dynamometer prices. Suspension systems of the type illustrated in Figure 24.5 (for which the user must supply the rotating machine) cost $4000 to $6000, increasing with capacity [12]. A 100 Hp (74.6 kW) portable water brake equipped with a strain gage load cell and a digital readout instrument for torque, speed, and power costs $4500, or $8950 with more sophisticated data acquisition equipment [26]. Stationary (and some transportable [23]) hydraulic dynamometers cost from $113/kW in the smaller sizes [25], down to $35/kW for the very largest [27]. Transportation, installation, and instrumentation can add significantly to these costs. Eddy current dynamometers cost from as little as $57/kW to nearly $700/kW, depending on the rated capacity, type of control system, and instrumentation [24, 25, 28]. Hysteresis brakes with integral speed sensors cost
from $3300 to $14,000 according to capacity [29]. Compatible controllers, from manual to fully pro-
grammable for PC test control and data acquisition via an IEEE-488 interface, vary in price from $500 to
$4200. The flexibility and high performance of ac universal dynamometers is reflected in their compar-
atively high prices of $670 to $2200/kW [25, 28].

References


3. S. J. Citron, On-line engine torque and torque fluctuation measurement for engine control utilizing
4. Supplement to ASME Performance Test Codes, Measurement of Shaft Power, ANSI/ASME PTC

5. See, for example, the catalog of torque wrench products of Consolidated Devices, Inc., 19220 San

Jose Ave., City of Industry, CA 91748.
6. I. J. Garshelis, C. R. Conto, and W. S. Fiegel, A single transducer for non-contact measurement of

the power, torque and speed of a rotating shaft, SAE Paper No. 950536, 1995.

(This book covers all phases of strain gage technology.)
8. B. D. Cullity, Introduction to Magnetic Materials, Reading, MA: Addison-Wesley, 1972, Section 8.5,

266–274.
9. W. J. Fleming, Magnetostrictive torque sensors—comparison of branch, cross and solenoidal


11. I. J. Garshelis, Circularly magnetized non-contact torque sensor and method for measuring torque

12. Lebow® Products, Siebe, plc., 1278 Maplelawn Road, Troy, MI 48099, Transducer Design Funda-

mentals/Product Listings, Load Cell and Torque Sensor Handbook No. 710, 1997, also: Torqstar™

and Torque Table®.
13. S. Himmelstein & Co., 2490 Pembroke, Hoffman Estates, IL 60195, MCRT® Non-Contact Strain

Gage Torquemeters and Choosing the Right Torque Sensor.
14. Teledyne Brown Engineering, 513 Mill Street, Marion, MA 02738-0288.
15. Staiger, Mohilo & Co. GmbH, Baumwesenstrasse 5, D-7060 Schorndorf, Germany (In the U.S.:

Schlenker Enterprises Ltd., 5143 Electric Ave., Hillside, IL 60162), Torque Measurement.
16. Torquemeters Ltd., Ravensthorpe, Northampton, NN6 8EH, England (In the U.S.: Torquetronics

Inc., P.O. Box 100, Allegheny, NY 14707), Power Measurement.
17. Vibrac Corporation, 16 Columbia Drive, Amherst, NH 03031, Torque Measuring Transducer.
18. GSE, Inc., 2360 Research Drive, Farmington Hills, MI 48335-2621, Torkducer®.
19. Sensor Developments Inc., P.O. Box 290, Lake Orion, MI 48361-0290, 1996 Catalog.
20. Bently Nevada Corporation, P.O. Box 157, Minden, NV 89423, TorXimitor™.

1975.
23. AW Dynamometer, Inc., P.O. Box 428, Colfax, IL 61728, Traction dynamometers: Portable and

stationary dynamometers for motors, engines, vehicle power take-offs.

30. Hicklin Engineering, 3001 NW 104th St., Des Moines, IA 50322, Transdyne™ (transmission test systems, brake and towed chassis dynamometers).
Robots in industrial settings perform repetitive tasks, such as machine loading, parts assembly, painting, and welding. Only in rare instances can these autonomous manipulators modify their actions based on sensory information. Although, thus far, a vast majority of research work in the area of robot sensing has concentrated on computer vision, contact sensing is an equally important feature for robots and has received some attention as well. Without tactile-perception capability, a robot cannot be expected to effectively grasp objects. In this context, robotic tactile sensing is the focus of this chapter.

### 25.1 Sensing Classification

Robotic sensing can be classified as either of the noncontact or contact type [1]. Noncontact sensing involves interaction between the robot and its environment by some physical phenomenon, such as acoustic or electromagnetic waves, that interact without contact. The most important types of robotic sensors of the noncontact type are vision and proximity sensors.

Contact sensing, on the other hand, implies measurement of the general interaction that takes place when the robot’s end effector is brought into contact with an object. Contact sensing is further classified into force and tactile sensing.

Force sensing is defined as the measurement of the global mechanical effects of contact, while tactile sensing implies the detection of a wide range of local parameters affected by contact. Most significant among those contact-based effects are contact stresses, slippage, heat transfer, and hardness.

The properties of a grasped object that can be derived from tactile sensing can be classified into geometric and dynamometric types [2]. Among the geometric properties are presence, location in relation to the end-effector, shape and dimensions, and surface conditions [3–7]. Among the dynamometric parameters associated with grasping are: force distribution, slippage, elasticity and hardness, and friction [8–12].

Tactile sensing requires sophisticated transducers; yet the availability of these transducers alone is not a sufficient condition for successful tactile sensing. It is also necessary to accurately control the modalities through which the tactile sensor interacts with the explored objects (including contact forces, as well as end-effector position and orientation) [13–15]. This leads to active tactile sensing, which requires a high degree of complexity in the acquisition and processing of the tactile data [16].
25.2 Mechanical Effects of Contact

Tactile sensing normally involves a rigid object indenting the compliant cover layer of a tactile sensor array [17], Figure 25.1. The indentation of a compliant layer due to contact can be analyzed from two conceptually different points of view [1]. The first one is the measurement of the actual contact stresses (force distribution) in the layer, which is usually relevant to controlling manipulation tasks. The second one is the deflection profile of the layer, which is usually important for recognizing geometrical object features. Depending on the approach adopted, different processing and control algorithms must be utilized.

There exists a definite relationship between the local shape of a contacting body and a set of subsurface strains (or displacements); however, this relationship is quite complex. Thus, it requires the use of the Theory of Elasticity and Contact Mechanics to model sensor–object interaction [18], and the use of Finite Element Analysis (FEA) as a practical tool for obtaining a more representative model of the sensor [19].

In general, the study of tactile sensors comprises two steps: (1) the forward analysis, related to the acquisition of data from the sensor (changes on the stress or strains, induced by the indentation of an object on the compliant surface of the transducer); and, (2) the inverse problem, normally related to the recovery of force distribution or, in some cases, the recovery of the indentor's shape.

Simplified Theory for Tactile Sensing

For simplicity, the general two-dimensional tactile problem is reduced herein to a one-dimensional one. Figure 25.2 shows a one-dimensional transducer that consists of a compliant, homogeneous, isotropic, and linear layer subjected to a normal stress $q(y)$ created by the indentation of an object.

For modeling purposes, it is assumed that the compliant layer is an elastic half-space. This simplification yields closed-form equations for the analysis and avoids the formation of a more complex problem, in which the effect of the boundary conditions at $x_{\text{min}}$ and $x_{\text{max}}$ must be taken into account. It has been proven that the modeling of the sensor by an elastic half-space represents a reasonable approximation to the real case [18]. Under these conditions, it can be shown that the normal strain, at a depth $y = d$, due to the normal stress $q(y)$ is given by [20]:

\begin{equation}
\varepsilon(y) = \frac{q(y)}{E} \left( 1 - \frac{y}{d} \right)
\end{equation}
where $\varepsilon_z$ is the strain at $x$ and $z = d$ due to the normal stress on the surface, and

$$E \text{ and } v \text{ are, respectively, the modulus of elasticity and the Poisson's coefficient of the compliant layer.}
$$

In obtaining Equation 25.2, it is assumed that the analysis is performed under planar strain conditions. It should be noted that a similar analysis can be performed for tangential contact stresses or strains.

The normal displacement at the surface, $w$, is given by:

$$w(x) = \int_{-\infty}^{\infty} q(x - x_0) k(x_0) \, dx_0$$

(25.3)

where

$$k(x) = \frac{-2(1 - v^2)}{\pi E} \log \left| \frac{x}{x_a} \right|$$

(25.4)

The singularity at $x = 0$ is expected due to the singularity of stress at that point. Note that, $k(x)$ is the deformation of the surface when a singular load of 1 N is applied at $x = 0$. The constant $x_a$ should be chosen such that at $x = x_a$, the deformation is zero. In this case, zero deformation should occur at $x \to \infty$ (note that it has been assumed that the sensor is modeled by an elastic half space), namely $x_a \to \infty$. This problem is associated with the two-dimensional deformation of an elastic half-space. To eliminate this difficulty, the boundary conditions of the transducer must be taken into account (i.e., a finite transducer must be analyzed), which requires, in general, the use of FEA.

FIGURE 25.2 Ideal one-dimensional transducer subjected to a normal stress.
Since measurements of strain (or stress) are usually done by a discrete number of sensing elements, Equation 25.2 must be discretized (Figure 25.3). Correspondingly, the force distribution must be reconstructed at discrete positions as shown in Figure 25.3. Let \( \Delta x_q \) be the distance between points, where the force distribution must be reconstructed from strain (or stress) measurements carried out by strain (or stress) sensing elements uniformly distributed at intervals \( \Delta x_p \), at \( z = d \). Also assume, even though it is not necessary, that \( \Delta x_q = \Delta x_p = \Delta x \) and that the forces are applied at positions immediately above the sensor elements. One can now define the strain (stress)-sample vector, \( \zeta \), whose components are given by
\[
\zeta_i = \varepsilon_x(x_i), \quad i = 1, 2, \ldots, n
\]
and the force distribution vector, \( F \), whose components are given by
\[
f_i = q_v(x_j), \quad j = 1, 2, \ldots, n
\]
Then, the discrete form of Equation 25.1 is given by:
\[
(25.5) \quad \zeta = TF
\]
where the elements of the matrix \( T \) are given by \( T_{ij} = k_v(x_i - x_j) \), \( i = 1, 2, \ldots, n \) and \( j = 1, 2, \ldots, n \) [23]. A similar relation to Equation 25.5 can be obtained discretizing Equation 25.3. In the general case, where \( \Delta x_q \neq \Delta x_p \), \( T \) is not square. Furthermore, in the general case, the vector \( F \) comprises both vertical and tangential components.

Equations 25.1 and 25.3 represent the regular forward problem, while Equation 25.5 represents the discretized version of the forward problem. The inverse problem, in most cases, consists of recovering the applied force profile from the measurements of strain, stress, or deflection. (Note that the surface displacement can also be used to recover the indentor’s shape.)

In [20], it was shown that the inverse problem is ill-posed because the operators \( h \) and \( k \) of Equations 25.1 and 25.3, respectively, are ill-conditioned. Consequently, the inverse problem is susceptible to noise. To solve this problem, regularization techniques must be utilized [20].

It has been proven that, in order to avoid aliasing in determining the continuous strain (stress) at a depth \( d \) using a discretized transducer, the elements have to be separated by one tenth of the compliant layer’s thickness. However, good results were obtained, without much aliasing, by separating the sensing elements by a distance equal to the sensor’s depth [18].

**Requirements for Tactile Sensors**

In 1980, Harmon conducted a survey to determine general specifications for tactile sensors [21]. Those specifications have been used subsequently as guidelines by many tactile sensor designers:
1. Spatial resolution of 1 to 2 mm
2. Array sizes of $5 \times 10$ to $10 \times 20$ points
3. Sensitivity of $0.5 \times 10^{-2}$ to $1 \times 10^{-2}$ N for each force-sensing element (tactel)
4. Dynamic range of 1000:1
5. Stable behavior and with no hysteresis
6. Sampling rate of 100 Hz to 1 kHz
7. Monotonic response, though not necessarily linear
8. Compliant interface, rugged and inexpensive

While properties (5), (7), and (8) above should apply to any practical sensor, the others are merely suggestions, particularly with respect to the number of array elements and spatial resolution.

Developments on tactile sensing following [21] have identified additional desirable qualities; namely, reliability, modularity, speed, and the availability of multisensor support [16].

25.3 Technologies for Tactile Sensing

The technologies associated with tactile sensing are quite diverse: extensive surveys of the state-of-the-art of robotic-tactile-transduction technologies have been presented in [2, 3, 16, 17]. Some of these technologies will be briefly discussed.

Resistive

The transduction method that has received the most attention in tactile sensor design is concerned with the change in resistance of a conductive material under applied pressure. A basic configuration of a resistive transducer is shown in Figure 25.4. Each resistor, whose value changes with the magnitude of the force, represents a resistive cell of the transducer. Different materials have been utilized to manufacture the basic cell.

Conductive elastomers were among the first resistive materials used for the development of tactile sensors. They are insulating, natural or silicone-based rubbers made conductive by adding particles of conductive or semiconductive materials (e.g., silver or carbon). The changes in resistivity of the elastomers...
under pressure are produced basically by two different physical mechanisms. In the first approach, the change in resistivity of the elastomer under pressure is associated with deformation that alters the particle density within it. Two typical designs of this kind are given in [22, 23]. In the second approach, while the bulk resistance of the elastomer changes slightly when it is compressed, the design allows the increase of the area of contact between the elastomer and an electrode, and correspondingly a change in the contact resistance. A typical design of this kind is given in [24]. In [25], a newer tactile sensor is reported with both three-axis force sensing and slippage sensing functions. In the former case, the pressure sensing function is achieved utilizing arrays of pressure transducers that measure a change in contact resistance between a specially treated polyimide film and a resistive substrate.

Piezoresistive elements have also been used in several tactile sensors. This technology is specifically attractive at present because, with micromachining, the piezoresistive elements can be integrated together with the signal-processing circuits in a single chip [26]. A 32 × 32-element silicon pressure sensor array incorporating CMOS processing circuits for the detection of a high-resolution pressure distribution was reported in [8]. The sensor array consists of an x–y-matrix-organized array of pressure cells with a cell spacing of 250 μm. CMOS processing circuits are formed around the array on the same chip. Fabrication of the sensor array was carried out using a 3 mm CMOS process combined with silicon micromachining techniques. The associated diaphragm size is 50 μm × 50 μm. The overall sensor-array chip size is 10 mm × 10 mm.

In Figure 25.4, a circuit topology, to scan a 3 × 3 array of piezoresistive elements, is shown. The basic idea was originally proposed in [24] and adapted on several occasions by different researchers. Using this method, the changes in resistance are converted into voltages at the output. With the connections as shown in Figure 25.4, the resistance \( R_{21} \) can be determined from:

\[
V_0 = \frac{R_i}{R_{21}} V_{cc}
\]

where \( V_0 \) is the output voltage, \( V_{cc} \) is the bias voltage, and \( R_i \) is the feedback resistance of the output amplifier stage.

One problem with the configuration shown in Figure 25.4 is the difficulty in detecting small changes in resistance due to the internal resistance of the multiplexer as well changes in the voltage of power source, which have a great influence at the output. Other methods utilized to scan resistive transducer arrays are summarized in [3].

When piezoresistors and circuits are fabricated on the same silicon substrate, the sensor array can be equipped with a complex switching circuit, next to the sensing elements, that allows a better resolution in the measurements [9].

**Capacitive**

Tactile sensors within this category are concerned with measuring capacitance, which varies under applied load. The capacitance of a parallel-plate capacitor depends on the separation of the plates and their areas. A sensor using an elastomeric separator between the plates provides compliance such that the capacitance will vary according to the applied normal load, Figure 25.5(a).

Figure 25.5(b) shows the basic configuration of a capacitive tactile sensor. The intersections of rows and columns of conductor strips form capacitors. Each individual capacitance can be determined by measuring the corresponding output voltage at the selected row and column. To reduce cross-talk and electromagnetic interference, the rows and columns that are not connected are grounded. Figure 25.5(c) shows an equivalent circuit when the sensor is configured to measure the capacitance formed at the intersection of row \( i \) and row \( j \), \( C_{ij} \). \( R_i \) is the input resistance of the detector and \( C_d \) represents the effects of the stray capacitances, including the detector-amplifier input capacitance, the stray capacitance due
to the unselected rows and columns, and the capacitance contributed by the cable that connects the transducer to the detector. Since the stray capacitance due to the unselected rows and columns changes with the applied forces, the stray capacitance due to the cable is designed to be predominant [18].

The magnitude of voltage at the input of the detector, $|V_d|$ is given by:

$$|V_d| = \frac{C_d R_d \omega}{\sqrt{1 + \left(\omega R_d (C_{ij} + C_d)\right)^2}} |V|$$  \hspace{1cm} (25.7)

Assuming that $C_d \gg C_{ij}$ and $\omega$ is sufficiently large,

$$|V_d| \cong \frac{C_{ij}}{C_d} |V|$$  \hspace{1cm} (25.8)

When a load is applied to the transducer, the capacitor is deformed as shown in Figure 25.5(a). For modeling purposes, it is assumed that the plate capacitor is only under compression. When no load is applied, the capacitance due to the element in the $i$th row and the $j$th column, $C_{ij}^0$, is given by:

$$C_{ij}^0 = \varepsilon \frac{wl}{h_0}$$  \hspace{1cm} (25.9)

where $\varepsilon$ is the permittivity of the dielectric, $w$ and $l$ are the width and the length of the plate capacitor, respectively, and $h_0$ is the distance between plates when no load is applied. The voltage at the input of the detector for this particular case is indicated by $V_{d0}$; then from Equation 25.8, one obtains:

$$|V_{d0}| \cong \frac{C_{ij}^0}{C_d} |V|$$  \hspace{1cm} (25.10)
When a load is applied, the capacitor is under compression and the capacitance is given by:

\[ C_0 = \varepsilon \frac{w l}{h_0 - \Delta h} \quad (25.11) \]

The strain in this case is given by:

\[ \varepsilon = \frac{\Delta h}{h_0} \quad (25.12) \]

where \( \Delta h \) is the displacement of the top metal plate and \( \Delta h \sim h_0 \). The strain can be measured by:

\[ \frac{V_d - V_{d0}}{V_d} = \frac{C_0 - C_0^0}{C_0} = 1 - \frac{C_0^0}{C_0} = 1 - \frac{h_0 - \Delta h}{h_0} = \frac{\Delta h}{h_0} \equiv \varepsilon \quad (25.13) \]

Consequently, the strain at each tactel can be determined by measuring the magnitudes of \( V_d \) and \( V_{d0} \) for each element.

Note that the presence of a tangential force would offset the plates tangentially and change the effective area of the capacitor plates. An ideal capacitive pressure sensor can quantify basic aspects of touch by sensing normal forces, and can detect slippage by measuring tangential forces. However, distinguishing between the two forces at the output of a single sensing element is a difficult task and requires a more complex transducer than the one presented in Figure 25.5(a) [27].

Micromachined, silicon-based capacitive devices are especially attractive due to their potential for high accuracy and low drift. A sensor with 1024 elements and a spatial resolution of 0.5 mm was reported in [28]. Several possible structures for implementing capacitive high-density tactile transducers in silicon have been reported in [29]. A cylindrical finger-shaped transducer was reported in [18].

The advantages of capacitive transducers include: wide dynamic range, linear response, and robustness. Their major disadvantages are susceptibility to noise, sensitivity to temperature, and the fact that capacitance decreases with physical size, ultimately limiting the spatial resolution. Research is progressing toward the development of electronic processing circuits for the measurement of small capacitances using charge amplifiers [30], and the development of new capacitive structures [29].

**Piezoelectric**

A material is called piezoelectric, if, when subjected to a stress or deformation, it produces electricity. Longitudinal piezoelectric effect occurs when the electricity is produced in the same direction of the stress, Figure 25.6. In Figure 25.6(a), a normal stress \( \sigma (= F/A) \) is applied along the Direction 3 and the charges are generated on the surfaces perpendicular to Direction 3. A transversal piezoelectric effect occurs when the electricity is produced in the direction perpendicular to the stress.

The voltage \( V \) generated across the electrodes by the stress \( \sigma \) is given by:

\[ V = d_{13} \frac{h}{\varepsilon} \sigma \quad (25.14) \]

where \( d_{13} = \) Piezoelectric constant associated with the longitudinal piezoelectric effect
\( \varepsilon = \) Permittivity
\( h = \) Thickness of the piezoelectric material

---

© 1999 by CRC Press LLC
Since piezoelectric materials are insulators, the transducer shown in Figure 25.6(a), can be considered as a capacitor, from an electrical point of view. Consequently,

$$V = \frac{Q}{C} = \frac{Q}{\varepsilon A}$$

(25.15)

where \( Q \) = Charge induced by the stress \( \sigma \)
\( C \) = Capacitance of the parallel capacitor
\( A \) = Area of each electrode

A comparison of Equations 25.14 and 25.15 leads to:

$$Q = d_{33} A \sigma$$

(25.16)

It is concluded that the force applied to the photoelastic material can be determined by finding the charge \( Q \). Charge amplifiers are usually utilized for determining \( Q \). The basic configuration of a charge amplifier is shown in Figure 25.6(b). The charge generated in the transducer is transferred to the capacitor \( C_f \) and the output voltage, \( V_o \), is given by:

$$V_o = -\frac{Q}{C_i}$$

(25.17)

The circuit must periodically discharge the feedback capacitor \( C_i \) to avoid saturation of the amplifier by stray charges generated by the offset voltages and currents of the operational amplifier. This is achieved by a switch as shown in Figure 25.6(b) or by a resistor parallel to \( C_i \).

The piezoelectric material most widely used in the implementation of tactile transducers is PVF2. It shows the largest piezoelectric effect of any known material. Its flexibility, small size, sensitivity, and large electrical output offer many advantages for sensor applications in general, and tactile sensors in particular. Examples of tactile sensors implemented with this technology can be found in [1, 31].

The major advantages of the piezoelectric technology are its wide dynamic range and durability. Unfortunately, the response of available materials does not extend down to dc and therefore steady loads cannot be measured directly. Also, the PVF2 material produces a charge output that is prone to electrical interference and is temperature dependent.
The possibility of measuring transient phenomenon using piezoelectric material has recently encouraged some researchers to use the piezoelectric effect for detecting vibrations that indicate incipient slip, occurrence of contact, local change in skin curvature, and estimating friction and hardness of the object [7, 10, 11]. If the piezoelectric transducer shown in Figure 25.6(a) is connected to an FET-input operational amplifier configured as a current-to-voltage converter as shown in Figure 25.7, the output voltage is given by:

\[ V_o = \frac{dQ}{dt} R_f = AR_1 d_{33} \frac{d\sigma}{dt} \]

(25.18)

where \( R_f \) is the feedback resistor. Correspondingly, the circuit configuration provides the mean to measure of changes in the contact stress. A detailed explanation of the behavior of this sensor can be found in [7].

**Optical**

Recent developments in fiber optic technology and solid-state cameras have led to numerous novel tactile sensor designs [32, 33]. Some of these designs employ flexible membranes incorporating a reflecting surface, Figure 25.8. Light is introduced into the sensor via a fiber optic cable. A wide cone of light propagates out of the fiber, reflects back from the membrane, and is collected by a second fiber. When an external force is applied onto the elastomer, it shortens the distance between the reflective side of the

![FIGURE 25.7 Current-to-voltage converter.](image)

![FIGURE 25.8 (a) Reflective transducer. (b) Light-intensity as a function of the distance h.](image)
membrane and the fibers, $h$. Consequently, the light gathered by the receiving fiber changes as a function of $h$, Figure 25.8(b). To recover univocally the distance from the light intensity, a monotonic function is needed. This can be achieved by designing the transducer such it operates for $h > h_{\text{min}}$, where $h_{\text{min}}$ is indicated in Figure 25.8(b). (The region $h > h_{\text{min}}$ is preferred to the $h < h_{\text{min}}$ for dynamic range reasons.)

Another optical effect that can be used is that of frustrated total internal reflection [5, 34]. With this technique, an elastic rubber membrane covers, without touching, a glass plate (waveguide); light entering the side edge of the glass is totally reflected by the top and bottom surfaces and propagates along it, Figure 25.9.

The condition for total internal reflection occurs when:

$$n_2 \sin \alpha \leq n_1$$  \hspace{1cm} (25.19)

where $n_1$ = Index of refraction of the medium surrounding the waveguide (in this case air, $n_i \equiv 1$)

$n_2$ = Index of refraction of the waveguide

$\alpha$ = Angle of incidence at the interface glass-air

Objects in contact with the elastic membrane deform it and induce contact between the bottom part of the membrane and the top surface of the waveguide, disrupting the total internal reflection. Consequently, the light in the waveguide is scattered at the contact location. Light that escapes through the bottom surface of the waveguide can be detected by an array of photodiodes, a solid-state sensor, or, alternatively, transported away from the transducer by fibers [3]. The detected imaged is stored in a computer for further analysis. A rubber membrane with a flat surface yields a high-resolution binary (contact or noncontact) image [5]. If the rubber sheet is molded with a textured surface (Figure 25.9), then an output proportional to the area of contact is obtained and, consequently, the applied forces can be detected [3]. Shear forces can also be detected using special designs [35]. Sensors based on frustrated internal reflection can be molded into a finger shape [5] and are capable of forming very high-resolution tactile images. Such sensors are commercially available. An improved miniaturized version of a similar sensor was proposed in [34].

Other types of optical transducers use “occluder” devices. One of the few commercially available tactile sensors uses this kind of transducer [36]. In one of the two available designs, the transducer’s surface is made of a compliant material, which has on its underside a grid of elongated pins. When force is applied to the compliant surface, the pins on the underside undergo a mechanical motion normal to the surface,
blocking the light path of a photoemitter–detector pair. The amount of movement determines the amount of light reaching the photoreceiver. Correspondingly, the more force applied, the less amount of light is collected by the photoreceiver, Figure 25.10. The major problems with this specific device are associated with creep, hysteresis, and temperature variation. This scheme also requires individual calibration of each photoemitter–photodetector pair.

Fibers have also been used directly as transducers in the design of tactile sensors. Their use is based on two properties of fiber optic cables: (1) if a fiber is subjected to a significant amount of bending, then the angle of incidence at the fiber wall can be reduced sufficiently for light to leave the core [37]; and (2) if two fibers pass close to one another and both have roughened surfaces, then light can pass between the fibers. Light coupling between adjacent fibers is a function of their separation [3].

An example of an optical fiber tactile sensor, whose sensing mechanism is based on the increase of light attenuation due to the microbend in the optical fibers, is shown in Figure 25.11 [37]. The transducer consists of a four-layer, two-dimensional fiber optic array constructed by using two layers of optical fibers as a corrugation structure, through which microbends are induced in two orthogonal layers of active fibers. Each active fiber uses an LED as the emitter and a PIN photodiode as a detector. When an object is forced into contact with the transducer, a light distribution is detected at each detector. This light distribution is related to the applied force and the shape of the object. Using complex algorithms and active sensing (moving the object in relation to the transducer), the object position, orientation, size, and contour information can be retrieved [37]. However, the recovery of the applied force profiles was not reported in [37].

---

**FIGURE 25.10** Principle of operation of an occluder transducer.

**FIGURE 25.11** A four-layer tactile transducer.
Photoelastic

An emerging technology in optical tactile sensing is the development of photoelastic transducers. When a light ray propagates into an optically anisotropic medium, it splits into two rays that are linearly polarized at right angles to each other and propagate at different velocities. This splitting of a ray into two rays that have mutually perpendicular polarizations results from a physical property of crystalline material that is called optical birefringence or simply birefringence. The direction in which light propagates with the higher velocity is called the fast axis; and the one in which it propagates more slowly is called the slow axis. Some optically isotropic materials — such as glass, celluloid, bakelite, and transparent plastics in general — become birefringent when they are subjected to a stress field. The birefringent effect lasts only during the application of loads. Thus, this phenomenon is called temporary or artificial birefringence or, more commonly, the photoelastic phenomenon.

Figure 25.12(a) shows a photoelastic transducer proposed in [38]. It consists of a fully supported two-layer beam with a mirrored surface sandwiched in between. Normal line forces are applied to the top surface of the beam at discrete tactels, separated by equal distances, s, along the beam. The upper compliant layer is for the protection of the mirror, while the lower one is the photoelastic layer. Circularly polarized monochromatic light, incident along the z-axis, illuminates the bottom surface of the transducer. The light propagates parallel to the z-axis, passes through the photoelastic layer, and then reflects back from the mirror. If no force is applied to the transducer, the returning light is circularly polarized because unstressed photoelastic material is isotropic. If force is applied, stresses are induced in the photoelastic layer, making the material birefringent. This introduces a certain phase difference between the components of the electric field associated with the light-wave propagation. The two directions of polarization are in the plane perpendicular to the direction of propagation (in this case, the x–y plane). As a consequence of this effect, the output light is elliptically polarized, creating a phase difference distribution, p, between the input light and the output light at each point in the x–y plane. The phase difference distribution carries the information of the force distribution applied to the transducer.

A polariscope is a practical method to observe the spatial variation on light intensity (fringes) due to the effect of induced phase difference distribution. Polarisopes can be either linear or circular, depending on the required polarization of the light. They can also be characterized as a reflective or a transparent type, depending on whether the photoelastic transducer reflect or transmits the light.

A circular, reflective polariscope, shown in Figure 25.12(b), is utilized to illuminate the transducer shown in Figure 25.12(a). The input light is linearly polarized and is directed toward the photoelastic transducer by a beam splitter. Before reaching the transducer, the light is circularly polarized by a quarter-wave plate. The output light is elliptically polarized when a force is applied. This light is directed toward a detector passing through the quarter-wave plate, the beam splitter, and an analyzer. Finally, it is detected by a camera linked to a frame grabber connected to a PC, for further data processing. The light that illuminates the camera consists of a set of fringes from where the force distribution applied to the transducer must be recovered. A technique for the recovery of the forces from the fringes is described in [38]. A model of the transducer using FEA is reported in [39].

One of the earlier applications of photoelasticity to tactile sensing dates back to the development phase of the Utah/MIT dexterous hand [40]. The researchers proposed the use of the photoelastic phenomenon as a transduction method for the recovery of the force profile applied to the fingers of the hand. They limited their application to the development of a single-touch transducer, although they claimed that an array of such devices could be implemented. However, the construction of a large array of their devices would be difficult. To overcome this difficulty, another research group proposed a different transducer [41]. Although an analytical model was developed for the sensor, a systematic method for recovering the two-dimensional force profile from the light intensity distribution was not reported. Thus, the sensor was used mainly for the study of the forward analysis, namely, observing the light intensity distribution for different touching objects brought into contact with the sensor. This sensor could eventually be used for determining some simple geometric properties of a touching object.
FIGURE 25.12  (a) Photoelastic transducer. (b) Circular reflective polariscope.
A tactile sensor reported in [42] is capable of detecting slippage. The output light intensity (the fringe pattern) is captured by a camera interfaced to a PC. When an object moves across the surface of the transducer, the light intensity distribution changes. A direct analysis of the fringes is used to detect movement of the grasped object; a special technique was reported to optimize the comparison process for detecting differences between two fringe patterns occurring due to the slippage of the object in contact with the sensor [42]. It is important to note that such an analysis of the fringes does not require the recovery of the applied force profile.

Photoelasticity offers several attractive properties for the development of tactile sensors: good linearity, compatibility with vision-base sensing technologies, and high spatial resolution associated with the latter, that could lead to the development of high-resolution tactile imagers needed for object recognition and fine manipulation. Also, photoelastic sensors are compatible with fiber optic technology that allows remote location of electronic processing devices and avoidance of interference problems.

Other technologies for tactile sensing include acoustic, magnetic, and microcavity vacuum sensors [43, 44].

References
