which could be used by students, researchers interested in modern instrumentation (applied physicists and engineers), sensor designers, application engineers and technicians whose job is to understand, select and/or design sensors for practical systems.

While this volume covers a broad range of sensors and detectors, an emphasis is on the devices which are less known, whose technology is still on a rise, and whose use permits the measurement of variables which were previously inaccessible. It is the author’s intention to present a comprehensive and up-to-date account of the theory (physical principles), design, and practical implementations of various, especially the newest, sensors for scientific, industrial, and consumer applications.

This book is a second, updated and expanded edition of a popular manual on sensors for science in industry. The first edition has been used quite extensively not only as a desktop reference by sensor designers and users, but as a textbook for related college courses as well. Comments and suggestions from users prompted me to implement several changes in this new edition. Recent ideas and developments have been added, while less important and less essential designs were dropped. Specifically, sections covering surface acoustic waves, temperature, and chemical sensors were revised and updated and a new chapter (No. 18) on manufacturing methods and materials has been included.

This book contains 18 chapters and an Appendix—however, it is actually divided into three major parts. The first consists of four chapters (1–3, and 18) which provide a general framework of a sensor’s overall characteristics, physical principles, and effects which form a foundation for their practical designs. Chapter 4 describes some useful electronic interfaces between sensors and peripheral processing devices. Here, an emphasis is made on those circuits which are suitable for integration with sensors. Certainly, the “meat” of this book is contained in the third part (Chapters 5–17) which is organized by the type of variables which are measured. The topics included in this part of the book reflect the author’s own preferences and interpretations. Some may find a description of a particular sensor either too detailed or too broad, or too brief. In most cases, the author made an attempt to strike a balance between a detailed description and simplicity of coverage.

Quite often, the best sensor is the simplest one. It is, therefore, appropriate to repeat after the great American inventor Charles F. Kettering: “Inventing is a combination of brains and materials. The more brains you use, the less materials you need.”

Jacob Freuden
San Diego, California
April, 1996

CHAPTER 1

DATA ACQUISITION

"It's as large as life, and twice as natural!"
—LEWIS CARROLL, "THROUGH THE LOOKING GLASS"

1.1 SENSORS, SIGNALS, AND SYSTEMS

A sensor is often defined as a “device that receives and responds to a signal or stimulus.” This definition is broad. In fact, it is so broad that it covers almost everything from a human eye to a trigger in a pistol. Consider the level control system shown in Fig. 1-1 [1]. The operator adjusts the level of fluid in the tank by manipulating its valve. Variations in the inlet flow rate, temperature changes (these would alter the fluids’ viscosity and consequently the flow rate through the valve) and similar disturbances must be compensated for by the operator. Without control, the tank is likely to flood, or run dry. To act appropriately, the operator must obtain information about the level of fluid in the tank on a timely basis. In this example, the information is perceived by the sensor which consists of two main parts: the sight tube on the tank and the operator’s eye which generates an electric response in the optic nerve. The sight tube by itself is not a sensor and, in this particular control system, the eye is not a sensor either. Only the combination of these two components makes a narrow purpose sensor (detector) which is selectively sensitive to fluid level. If a sight tube is designed properly, it will very quickly reflect variations in the level and, it is said, that the sensor has a fast speed response. If internal diameter of the tube is too small for a given fluid viscosity, the level in the tube may lag behind the level in the tank. Then, we have to consider a phase characteristic of such a sensor. In some cases, the lag may be quite acceptable, while in other cases, a better sight tube design would be required. Hence, the sensor’s performance must be assessed only as a part of a data acquisition system.

This world is divided into natural and man-made objects. The natural sensors, like those found in living organisms, usually respond with signals, having electrochemical character. That is, their physical nature is based on ion transport, like in the nerve fibers (such as an optic nerve in the fluid tank operator). In man-made devices, information is also transmitted and processed in electrical form, however, through the transport of electrons. Sensors which are used in artificial systems must speak the same language as the devices with which they are interfaced. This language is electrical in its nature and a man-made sensor should be capable of
responding with signals where information is carried by displacement of electrons, rather than ions. Thus, it should be possible to connect a sensor to an electronic system through electrical wires, rather than through an electrochemical solution or a nerve fiber. Hence, in this book, we use a somewhat narrower definition of sensors, which may be phrased as “a sensor is a device that receives a signal or stimulus and responds with an electrical signal.” The term stimulus is used throughout this book and needs to be clearly understood. The stimulus is the quantity, property, or condition that is sensed and converted into electrical signal. Some texts (for instance, [2]) use a different term measure which has the same meaning, however with the stress on quantitative characteristic of sensing.

The purpose of a sensor is to respond to some kind of an input physical property (stimulus) and to convert it into an electrical signal which is compatible with electronic circuits. We may say that a sensor is a translator of a generally nonelectrical value into an electrical value. When we say “electrical” we mean a signal, which can be channeled, amplified, and modified by electronic devices. The sensor’s output signal may be in a form of voltage, current, or charge. These may be further described in terms of amplitude, frequency, and phase. This set of characteristics is called the output signal format. Therefore, a sensor has input properties (of any kind) and electrical output properties.

The term sensor should be distinguished from transducer. The latter is a converter of one type of energy into another, while the former converts any type of energy into electrical. An example of a transducer is a loudspeaker which converts electrical signal into variable magnetic field and, subsequently, into acoustic waves. This is nothing to do with perception or sensing. Transducers may be used as actuators in various systems. Also, they may be parts of sensors. For example, a chemical sensor may have a part which converts chemical energy into thermal (transducer) and another part, a thermopile, which converts heat into electrical signal. The combination of the two makes a chemical sensor—a device which produces an electrical signal in response to a chemical reaction.

A sensor does not function by itself—it is always a part of a larger system which may incorporate many other detectors, signal conditioners, signal processors, memory devices, data recorders, and actuators. The sensor’s place in a device is either intrinsic or extrinsic. It may be positioned at the input of a device to perceive the outside effects and to signal the system about variations in the outside stimuli. Also, it may be an internal part of a device which monitors the devices’ own state to cause the appropriate performance. A sensor is always a part of some kind of a data acquisition system. Often such a system may be a part of a larger control system which includes various feedback mechanisms. To select an appropriate sensor, a system designer must address the question: “What is the simplest way to sense the stimulus without degradation of the overall system performance?”

All sensors may be of two kinds: passive and active. The passive sensors directly generate an electric signal in response to an external stimulus. That is, the input stimulus energy is converted by the sensor into output energy without the need for an additional power source. The examples are a thermocouple, a pyroelectric detector, and a piezoelectric sensor. The active sensors require external power for their operation, which is called an excitation signal. That signal is modified by the sensor to produce the output signal. The active sensors sometimes are called parametric because their own properties change in response to an external effect and these properties can be subsequently converted into electrical signals. For example, a thermistor is a temperature sensitive resistor. It does not generate any signal, but by passing an electric current through it (excitation signal), its resistance can be measured by detecting variations in current and/or voltage across the thermistor. These variations (presented in ohms) directly relate to temperature.

To illustrate a place of sensors in a larger system, Fig. 1.2 shows a block diagram of a data acquisition and control device. An object can be anything: car, space ship, animal or human, liquid, or gas. Any material object may become a subject of some kind of a measurement. Data are collected from an object by a number of sensors. Some of them (2, 3, and 4) are positioned directly on or inside the object. Sensor 1 perceives the object without a physical contact and, therefore, is called a noncontact sensor. Examples of such a sensor is a radiation detector and a television camera. Sensor 5 serves a different purpose. It monitors internal conditions of a data acquisi-
A sensor is an input device that converts a physical quantity into an electrical signal. The type of sensor used depends on the physical quantity to be measured. For example, a temperature sensor is used to measure temperature, a pressure sensor to measure pressure, and a light sensor to measure light intensity.

In a sensor system, the sensor is connected to a data acquisition system (DAS) through a multiplexer. The DAS is responsible for converting the analog signal from the sensor into a digital signal that can be processed by a computer. The data acquisition system is connected to a computer through a data acquisition card (DAC) or a data acquisition interface (DAI).

The DAS consists of several components, including a multiplexer, an analog-to-digital converter (ADC), and a computer. The multiplexer is used to select which sensor is connected to the ADC at any given time. The ADC is used to convert the analog signal from the sensor into a digital signal that can be processed by the computer. The computer is used to store the digital data in a database and to analyze the data using various software tools.

There are different types of sensors, each designed to measure a specific physical quantity. For example, a thermistor is a type of temperature sensor that changes its electrical resistance in response to changes in temperature. A photodiode is a type of light sensor that converts light into an electrical signal.

In summary, a sensor system is a combination of a sensor, a data acquisition system, and a computer. The sensor converts a physical quantity into an electrical signal, the data acquisition system converts the signal into a digital format, and the computer processes the data and stores it in a database.
TABLE 1.1. Specifications.

<table>
<thead>
<tr>
<th>Specification</th>
<th>Stimulus range (spar)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>Resolution</td>
</tr>
<tr>
<td>Stability (short and long term)</td>
<td>Selectivity</td>
</tr>
<tr>
<td>Accuracy</td>
<td>Environmental conditions</td>
</tr>
<tr>
<td>Speed of response</td>
<td>Linearity</td>
</tr>
<tr>
<td>Overload characteristics</td>
<td>Dead band</td>
</tr>
<tr>
<td>Hysteresis</td>
<td>Output format</td>
</tr>
<tr>
<td>Operating life</td>
<td>Other</td>
</tr>
<tr>
<td>Cost, size, weight</td>
<td></td>
</tr>
</tbody>
</table>

what conversion mechanism is employed, what material it is fabricated from, and what is its field of application. Tables 1.1–1.6, adapted from [3], represent such a classification scheme which is pretty much broad and representative. If we take, for example a surface acoustic-wave oscillator accelerometer, the table entries might be as follows:

Stimulus: acceleration
Specifications: sensitivity in frequent shift per g of acceleration, short and long-term stability in Hz per unit time, etc.
Detection means: mechanical
Conversion phenomenon: elastoelectric
Material: inorganic insulator
Field: automotive, marine, space and scientific measurement

1.3 UNITS OF MEASUREMENTS

In this book, we use base units which have been established in The 14th General Conference on Weights and Measures (1971). The base measurement system is known as SI which stands for French “Le Système International d’Unités” (Table 1.7). All other physical quantities are derivatives of these base units. Some of them are listed in Table A.3.

Often it is not convenient to use base or derivative units directly—in practice quantities may be either too large or too small. For convenience in the engineering work, multiples and submultiples of the units are generally employed. They can be obtained by multiplying a unit by a factor from Table A.2. When pronounced, in all cases the first syllable is accented. For example, 1 ampere (A) may be multiplied by factor of $10^{-3}$ to obtain a smaller unit: 1 milliampere (mA) which is one thousandth of an ampere.

Sometimes, two other systems of units are used. They are the Gaussian System and the British System, which in the U.S. its modification is called the U.S. customary system. The United States is the only developed country where SI is still not in common use. However, with the end of communism and the increase of

TABLE 1.2. Sensor material.

<table>
<thead>
<tr>
<th>Organic</th>
<th>Insulator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inorganic</td>
<td>Liquid gas or plasma</td>
</tr>
<tr>
<td>Conductor</td>
<td>Other</td>
</tr>
<tr>
<td>Semiconductor</td>
<td>Biological substance</td>
</tr>
</tbody>
</table>

TABLE 1.3. Detection means used in sensors.

<table>
<thead>
<tr>
<th>Biological</th>
<th>Chemical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electric, magnetic or electromagnetic wave</td>
<td>Heat, temperature</td>
</tr>
<tr>
<td>Mechanical displacement or wave</td>
<td>Radioactivity, radiation</td>
</tr>
<tr>
<td>Other</td>
<td>Other</td>
</tr>
</tbody>
</table>

TABLE 1.4. Conversion phenomena.

<table>
<thead>
<tr>
<th>Physical</th>
<th>Biological</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thermoelectric</td>
<td>Biochemical reaction</td>
</tr>
<tr>
<td>Photoelectric</td>
<td>Physical transformation</td>
</tr>
<tr>
<td>Photomagnetic</td>
<td>Electrochemical process</td>
</tr>
<tr>
<td>Magnetoelastic</td>
<td>Spectroscopy</td>
</tr>
<tr>
<td>Electromagnetic</td>
<td>Other</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chemical</th>
<th>Biological</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical transformation</td>
<td>Biochemical reaction</td>
</tr>
<tr>
<td>Physical transformation</td>
<td>Physical transformation</td>
</tr>
<tr>
<td>Electrochemical process</td>
<td>Electrochemical process</td>
</tr>
<tr>
<td>Spectroscopy</td>
<td>Other</td>
</tr>
</tbody>
</table>

TABLE 1.5. Field of applications.

| Agriculture         | Automotive          |
| Civil engineering, construction | Domestic, appliances |
| Distribution, commerce, finance | Environmental, meteorology, security |
| Energy, power       | Information, telecommunications |
| Health, medicine    | Marine               |
| Manufacturing       | Recreation, toys    |
| Military            | Space                |
| Scientific measurement | Other               |
| Transportation (excluding automotive) | Other                |
TABLE 1.7. SI basic units.

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Name</th>
<th>Symbol</th>
<th>Defined by</th>
<th>(Year established)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>meter</td>
<td>m</td>
<td>... the length of the path traveled by light in vacuum in 1/299,792,458 of a second...</td>
<td>(1983).</td>
</tr>
<tr>
<td>Mass</td>
<td>kilogram</td>
<td>kg</td>
<td>... after a platinum-iridium prototype (1889).</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>second</td>
<td>s</td>
<td>... the duration of 9,192,631,770 periods of the radiation corresponding to the transition...</td>
<td>(1967).</td>
</tr>
<tr>
<td>Electric current</td>
<td>ampere</td>
<td>A</td>
<td>... the force equal to 10⁻⁹ newton per meter of length extended on two parallel conductors in...</td>
<td>(1948).</td>
</tr>
<tr>
<td>Thermodynamic temperature</td>
<td>kelvin</td>
<td>K</td>
<td>... the fraction 1/273.16 of the thermodynamic temperature of the triple point of water...</td>
<td>(1867).</td>
</tr>
<tr>
<td>Amount of substance</td>
<td>mole</td>
<td>mol</td>
<td>... the amount of substance which contains as many elementary entities as there are atoms in 0.012 kg of carbon 12</td>
<td>(1971).</td>
</tr>
<tr>
<td>Luminous intensity</td>
<td>candela</td>
<td>cd</td>
<td>... intensity in the perpendicular direction of a surface of 1/4050,000 m² of a blackbody at the temperature of freezing pure water under pressure of 101,325 newton per m²</td>
<td>(1967).</td>
</tr>
</tbody>
</table>

Plane angle | radian | rad | (supplemental unit) |
Solid angle | steradian | sr | (supplemental unit) |

References:

World integration, international cooperation gains strong momentum. Hence, it is unavoidable that America will convert to SI in the very near future. In this book, we will generally use SI, however, for the convenience of the reader, the U.S. customary system units will be used in places where U.S. manufacturers employ them for sensor specifications. For the conversion to SI from other systems the reader may use Table A.4 in the Appendix. To make a conversion, a non-SI value should be multiplied by a number given in the table. For instance, to convert acceleration of 35 ft/s² to SI, it must be multiplied by 0.3048:

55 ft/s²×0.3048=16.764 m/s².

Similarly, to convert electric charge of 1.7 faraday, it must be multiplied by 9.65×10⁹:

1.7 faraday×9.65×10⁹=1.64×10²⁹ C.

SI is often called the modernized metric system.

Nomenclature, abbreviations, and spellings in the conversion tables are in accordance with "Standard practice for use of the International System of Units (SI); the Modernized Metric System," Standard E380-94a. 01991 ASTM. (80 Rues Harbor Dr., West Conshohocken, PA 19428-2959).
CHAPTER 2

SENSOR CHARACTERISTICS

"O, what men dare do! What men may do! What men daily do, not knowing what they do."
—SHAKESPEARE, "MUCH ADO ABOUT NOTHING"

From the input to the output, a sensor may have several conversion steps before it produces an electrical signal. For instance, pressure inflicted on the fiber-optic sensor, first results in strain in the fiber, which, in turn, causes deflection in its refractive index, which, in turn, results in an overall change in optical transmission and modulation of photon density. Finally, photon flux is detected and converted into electric current. In this chapter, we discuss the overall sensor characteristics, regardless of its physical nature or steps which are required to make a conversion. We regard a sensor as a "black box" where we concern only with relationships between its output and input signals.

2.1 TRANSFER FUNCTION

An ideal or theoretical output-stimulus relationship exists for every sensor. If the sensor is ideally designed and fabricated with ideal materials by ideal workers using ideal tools, the output of such a sensor would always represent the true value of the stimulus. The ideal function may be stated in the form of a table of values, a graph, or a mathematical equation. An ideal (theoretical) output-stimulus relationship is characterized by the so-called transfer function. This function establishes dependence between the electrical signal \( S \) produced by the sensor, and the stimulus \( x \): \( S = f(x) \). That function may be a simple linear connection or a nonlinear dependence, for instance logarithmic, exponential, or power function. In many cases the relationship is two-dimensional, that is, the output versus one input stimulus. A two-dimensional linear relationship is represented by equation:

\[
S = ax + bx,
\]

where \( a \) is the intercept, that is, the output signal at zero input signal, and \( b \) is the slope, which is sometimes called sensitivity. \( S \) is one of the characteristics of the output electric signal which is used by the data acquisition devices as the sensor's output. It may be amplitude, frequency or phase, depending on the sensor properties.

Logarithmic function is:

\[
S = a + b \ln x,
\]

Exponential function is:

\[
S = ae^{bx},
\]

Power function is:

\[
S = ax^b + ax^b t,
\]

where \( k \) is a constant number.

A sensor may have such a transfer function that none of the above approximations fits sufficiently well. In that case, a higher order polynomial approximation is often employed.

For a nonlinear transfer function, sensitivity \( b \) is not a fixed number as for the linear relationship (Eq. 2.1). At any particular input value, \( x_i \), it can be defined as:

\[
b = \frac{dS(x_i)}{dx}.
\]

In many cases, a nonlinear sensor may be considered linear over a limited range. Over the extended range, a nonlinear transfer function may be modeled by several straight lines. This is called a piece-wise approximation. To determine whether a function can be represented by a linear model, the incremental variables are introduced for the input while observing the output. A difference between the actual response and a liner model is compared with the specified accuracy limits (see below).

A transfer function may have more than two dimensions when the sensor's output is influenced by more than one input stimulus. An example is the transfer function of a thermal radiation (infrared) sensor. The function connects two temperatures \( T \), absolute temperature of the object of measurement, and \( T \), absolute temperature of the sensor's surface and the output voltage \( V \):

\[
V = G(T - T)
\]

where \( G \) is a constant. Clearly, the relationship between the object's temperature and the output voltage (transfer function) is not only nonlinear (the 4-th order parabolic) but also depends on the sensor's surface temperature. To determine sensitivity of the sensor with respect to the object temperature, a partial derivative will be calculated as

\[
b = \frac{\partial V}{\partial T} = 4GT^4
\]

This function is generally known as Stefan-Boltzmann law.
2.2 SPAN (INPUT FULL SCALE)

A dynamic range of stimuli which may be converted by a sensor is called a span or an input full scale (FS). It represents the highest possible input value which can be applied to the sensor without causing unacceptably large inaccuracy. For the sensors with a very broad and nonlinear response characteristic, a dynamic range of the input stimuli is often expressed in decibels, which is a logarithmic measure of ratios of either power or force (voltage). It should be emphasized that decibels do not measure absolute values, but a ratio of values only. A decibel scale represents signal magnitudes by much smaller numbers, which in many cases is far more convenient. Being a nonlinear scale, it may represent low level signals with high resolution while compressing the high level numbers. In other words, the logarithmic scale for small objects works as a microscope and for the large objects as a telescope. By definition, decibels are equal to 20 times the log of the ratio of powers (Table 2.1):

\[ 1 \text{ dB} = 10 \log \frac{P_2}{P_1} \]

(2.5)

In a similar manner, decibels are equal to 20 times the log of the force, or current, or voltage:

\[ 1 \text{ dB} = 20 \log \frac{F_2}{F_1} \]

(2.9)

2.3 FULL SCALE OUTPUT

Full scale output (FSO) is the algebraic difference between the electrical output signals measured with maximum input stimulus and the lowest input stimulus applied. This must include all deviations from the ideal transfer function. For instance, the FSO output in Fig. 2.2A is represented by \( S_{FS} \).

2.4 ACCURACY

A very important characteristic of a sensor is accuracy which really means inaccuracy. Inaccuracy is measured as a highest deviation of a value represented by the sensor from the ideal or true value at its input. The true value is attributed to the object of measurement and accepted as having a specified uncertainty (see below).

The deviation can be described as a difference between the value which was converted by the sensor into voltage and then, without any error, converted back, and the actual input value. For example, a linear displacement sensor ideally should generate 1 mV per 1 mm displacement. That is, its ideal sensitivity is \( h = 1 \text{ mV/mm} \).

However, in the experiment, a displacement \( s = 10 \text{ mm} \) produced an output of \( V = 10.5 \text{ mV} \). Converting this number back into displacement without error (100 = 1 \text{ mm/mV} \), we would calculate that the displacement was \( s = 10.5 \text{ mm} \), that is, \( s = 0.5 \text{ mm} \) more than the actual. This extra 0.5 mm is an erroneous deviation in the measurement, or error. Therefore, in a 10-mm range the sensor's absolute inaccuracy is 0.5 mm, or in the relative terms inaccuracy is 0.5 mm/10 mm = 0.05 = 5%. If we repeat this experiment over and over again without any random error and every time we observe an error of 0.5 mm we may say that the sensor has a systematic inaccuracy of 0.5 mm over a 10-mm span.

Figure 2.2A shows an ideal or theoretical transfer function. In the real world, any sensor performs with some kind of imperfection. A possible real transfer function is represented by a thick line, which generally may be neither linear nor monotonic. A real function rarely coincides with the ideal. Because of material variations,
workmanship, design errors, manufacturing tolerances, and other limitations, it is possible to have a large family of real transfer functions, even when sensors are tested under the identical conditions. However, all runs of the real transfer functions must fall within the limits of a specified accuracy. These permissive limits differ from the ideal transfer function line by $\pm \Delta$. The real functions deviate from the ideal by $\pm \Delta$, where $\Delta = \Delta$. For example, let us consider a stimulus having value, $x$. Ideally, we would expect this value to correspond to point $z$ on the transfer function, resulting in the output value $y$. Instead, the real function will respond at point $Z$ producing output value $y'$. This output value corresponds to point $z'$ on the ideal transfer function, which, in turn, relates to a "would-be" input stimulus $x'$ whose value is smaller than $x$. Thus, in this instance, imperfection in the sensor’s transfer function leads to a measurement error—$\delta$.

The accuracy rating includes a combined effect of part-to-part variations, a hysteresis, a dead band, calibration and repeatability errors (see below). The specified accuracy limits generally are used in the worst case analysis to determine the worst possible performance of the system. Figure 2.2B shows that $\pm \Delta$ may move more closely the real transfer function, meaning better tolerances of the sensor’s accuracy. This can be accomplished by a multiple-point calibration. Thus, the specified accuracy limits are established not around the theoretical (ideal) transfer func-

![Figure 2.2](image)

**FIGURE 2.2.** Transfer function (A) and accuracy limits (B). Error is specified in terms of input value.

For example, a piezoresistive pressure sensor has a 100 kPa input full scale and 10 Ω full scale output. Its inaccuracy may be specified as ±0.5%, or ±500 Pa, or ±0.05 Ω. In modern sensors, specification of accuracy often is replaced by a more comprehensive value of uncertainty (see Sec. 2.19) because uncertainty is comprised of all disturbing effects both systematic and random and is not limited to inaccuracy of a transfer function.

### 2.5 CALIBRATION ERROR

**Calibration error** is a bias permitted by a manufacturer when a sensor is calibrated in the factory. This error is of a systematic nature, meaning that it is added to all possible real transfer functions. It shifts the accuracy of transduction for each stimulus point by a constant. This error is not necessarily uniform over the range and may change depending on the type of error in calibration. For example, let us consider a two-point calibration of a real linear transfer function (thick line in Fig. 2.3). To determine the slope and the intercept of the function, two stimuli, $x_1$ and $x_2$, are applied to the sensor. The sensor responds with two corresponding output signals $A_1$ and $A_2$. The first response was measured absolutely accurately, however, the higher signal was measured with error $-\Delta$. This results in errors in the slope and intercept calculation. A new intercept, $a_3$, will differ from the real intercept, $a$, by

$$\delta_a = a_3 - a = \frac{\Delta}{x_2 - x_1}$$

and the slope will be calculated with error:

$$\delta_s = \frac{\Delta}{x_2 - x_1}$$

In some applications it may be prohibitive because of a higher cost. Inaccuracy rating may be represented in a number of forms:

1. In terms of measured value ($A$);
2. In percent of input span (full scale);
3. In terms of output signal.

For example, a piezoresistive pressure sensor has a 100 kPa input full scale and 10 Ω full scale output. Its inaccuracy may be specified as ±0.5%, or ±500 Pa, or ±0.05 Ω.
2.6 HYSTERESIS

A **hysteresis error** is a deviation of the sensor's output at a specified point of the input signal when it is approached from the opposite directions (Fig. 2.4). For example, a displacement sensor when the object moves from left to right at a certain point produces voltage which differs by 20 mV from that when the object moves from right to left. If the sensitivity of the sensor is 10 mV/mm, the hysteresis error in terms of displacement units is 2 mm. Typical causes for hysteresis are friction and structural changes in the materials.

2.7 NONLINEARITY

**Nonlinearity error** is specified for sensors whose transfer function may be approximated by a straight line [Eq. (2.1)]. A nonlinearity is a maximum deviation \( (L) \) of a real transfer function from the approximation straight line. The term "linearity" actually means "nonlinearity." When more than one calibration run is made, the worst linearity seen during any one calibration cycle should be stated. Usually, it is specified either in % of span or in terms of measured value, for instance, in kPa or °C. "Linearity," when not accompanied by a statement explaining what sort of straight line it is referring to, is meaningless. There are several ways to specify a nonlinearity, depending how the line is superimposed on the transfer function. One way is to use **terminal points** (Fig. 2.5A), that is, to determine output values at the smallest and highest stimulus values and to draw a straight line through these two points (line 1). Here, near the terminal points, the nonlinearity error is the smallest and it is higher somewhere in between.

Another way to define the approximation line is to use a method of **least squares**...
This can be done in the following manner. Measure several output values \( S \) at input values \( x \) over a substantially broad range, preferably over an entire full scale. Use the following formulas for linear regression to determine intercept \( a \) and slope \( b \) of the best fit straight line:

\[
\begin{align*}
  a &= \frac{\sum xy - \frac{1}{n} \sum x \sum y}{\sum x^2 - \left(\frac{1}{n} \sum x\right)^2}, \\
  b &= \frac{\sum x y - \frac{1}{n} \sum x \sum y}{\sum x^2 - \left(\frac{1}{n} \sum x\right)^2},
\end{align*}
\]

where \( \Sigma \) is the summation of \( n \) numbers.

In some applications, higher accuracy may be desirable in a particular narrower section of the input range. For instance, a medical thermometer should have the best accuracy in a fever definition region which is between 37 and 38 °C. It may have a somewhat lower accuracy beyond these limits. Usually, such a sensor is calibrated in the region where the highest accuracy is desirable. Then, the approximation line may be drawn through the calibration point \( c \) (line 3 in Fig. 2.5A). As a result, nonlinearity has the smallest value near the calibration point and it increases toward the ends of the span. In this method, the line is often determined as tangent to the transfer function in point \( c \). If the actual transfer function is known, the slope of the line can be found from Eq. (2.5).

**Independent linearity** is referred to the so-called “best straight line” (Fig. 2.5B), which is a line midway between two parallel straight lines closest together and enveloping all output values on a real transfer function.

Depending on the specification method, approximation lines may have different intercepts and slopes. Therefore, nonlinearity measures may differ quite substantially from one another. A user should be aware that manufacturers often publish the smallest possible number to specify nonlinearity, without defining what method was used.

**2.6 SATURATION**

Almost any sensor has its operating limits. Even if it is considered linear, at some levels of the input stimuli, its output signal no longer will be responsive. Further increase in stimulus does not produce a desirable output. It is said that the sensor exhibits a span-end nonlinearity or saturation (Fig. 2.6).

**2.9 REPEATABILITY**

Repeatability (reproducibility) error is caused by the inability of a sensor to represent the same value under identical conditions. It is expressed as the maximum difference between output readings as determined by two calibrating cycles (Fig. 2.7A), unless otherwise specified. It is usually represented as percent of \( FS \):

\[
\delta_r = \frac{\Delta}{FS} \times 100\%.
\]

**2.10 DEAD BAND**

Dead band is the insensitivity of a sensor in a specific range of input signals (Fig. 2.7B). In that range, the output may remain near a certain value (often zero) over an entire dead band zone.

**2.11 RESOLUTION**

Resolution describes smallest increments of stimulus which can be sensed. When a stimulus continuously varies over the range, the output signals of some sensors will
not be perfectly smooth, even under the no-noise conditions. The output may change in small steps. This is typical for potentiometric transducers, occupancy infrared detectors with grid masks, and other sensors where the output signal change is enabled only upon a certain degree of stimulus variation. The magnitude of the input variation which results in the output smallest step is specified as resolution under specified conditions (if any). For instance, for the occupancy detector the resolution may be specified as follows: "resolution—minimum equivalent displacement of the object for 20 cm at 3 m distance." For wire-wound potentiometric angular sensors, resolution may be specified as "a minimum angle of 0.5°." Sometimes, it may be specified as percent of full scale (FS). For instance, for the angular sensor having 270° FS, the 0.5° resolution may be specified as 0.181% of FS. It should be noted, that the step size may vary over the range, hence, the resolution may be specified as typical, average, or "worst." The resolution of digital output format sensors is given by the number of bins in the data word. For instance, the resolution may be specified as "8-bit resolution." When there are no measurable steps in the output signal, it is said that the sensor has continuous or infinitesimal resolution (sometimes erroneously referred to as "infinite resolution").

2.12 SPECIAL PROPERTIES

Special input properties may be needed to specify for some sensors. For instance, light detectors are sensitive within a limited optical bandwidth. Therefore, it is appropriate to specify for them a spectral response.

2.13 OUTPUT IMPEDANCE

Output impedance \( Z_{out} \) is important to know to better interface a sensor with the electronic circuit. This impedance is connected either in parallel with the input impedance \( Z_{in} \) of the circuit (voltage connection) or in series (current connection). Figure 2.8 shows two connections. Output and input impedances generally should be represented in a complex form, as they may include active and reactive components. To minimize output signal distortions, the current generating sensor (B) should have output impedance as high as possible and the circuit's input impedance should be low. For the voltage connection (A), a sensor is preferable with lower \( Z_{out} \) and the circuit should have \( Z_{in} \) as high as practical.

2.14 EXCITATION

Excitation is the electrical signal needed for the active transducer operation. Excitation is specified as a range of voltage and/or current. For some transducers, the frequency of the excitation signal, and its stability must also be specified. Variations in the excitation may alter the transducer's transfer function and cause output error. An example of excitation signal specification is:

2.15 DYNAMIC CHARACTERISTICS

Under static conditions a sensor is fully described by its transfer function, span, calibration, etc. However, when an input stimulus varies, a sensor response generally does not follow with perfect fidelity. The reason is that both the sensor and its coupling with the source of stimulus can not always respond instantly. In other words, a sensor may be characterized with a time-dependent characteristic, which is called a dynamic characteristic. If a sensor does not respond instantly, it may indicate values of stimuli which are somewhat different from the real, that is, the sensor responds with a dynamic error. A difference between static and dynamic errors is that the latter is always time dependent. If a sensor is a part of a control system which has its own dynamic characteristics, the combination may cause oscillations.

Warm-up time is the time between applying to the sensor power or excitation signal and the moment when the sensor can operate within its specified accuracy. Many sensors have a negligibly short warm-up time. However, some detectors, especially those that operate in a thermally controlled environment (a thermostat) may require seconds and minutes of warm-up time before they are fully operational within the specified accuracy limits.

Frequency response is an important dynamic characteristic of a detector as it specifies how fast the sensor can react to a change in the input stimulus. The frequency response is expressed in Hz or rad/sec to specify the relative reduction in the output signal at certain frequency (Fig. 2.9A). A commonly used reduction number (frequency limit) is −3 dB. It shows at what frequency the output voltage (or current) drops by about 30%. Frequency response limit \( f_c \) is often called the upper cutoff frequency, as it is considered the highest frequency which a sensor can process.

The frequency response directly relates to a speed response, which is defined in
response when both upper and lower cutoff frequencies are limited. Eventually, the response never reaches its would-be steady state level, $S_{c}$. For the first order response, it can be expressed as a product of two exponential processes:

$$S = S_{c}(1 - e^{-\frac{t}{\tau}})e^{-\frac{t}{\tau}}.$$  \hspace{1cm} (2.16)

As a rule of thumb, a simple formula can be used to establish a connection between the cutoff frequency, $f_{c}$ (either upper and lower) and time constant in a first-order sensor:

$$f_{c} \approx \frac{0.159}{\tau}.$$  \hspace{1cm} (2.17)

Clearly, for a relatively narrow bandwidth sensor (when the upper and lower cutoff frequencies are close to one another), use of time constants becomes inappropriate, because it is almost impossible to separate two exponential slopes in measurements. However, for a broad-bandwidth sensor (when the upper cutoff frequency is much higher, say 50 times), both time constants can be measured quite accurately.

There is a large class of sensors which may respond to constant stimuli. Such sensors, is said, have a dc response, therefore $\tau_{c}=\infty$ and $f_{c}=0$. Figure 2.10 shows typical responses of sensors which are the result of various combinations of cutoff frequencies.

Phase shift at a specific frequency defines how the output signal lags behind in representing the stimulus change (Fig. 2.9A). The shift is measured in angular degrees or radians. If a sensor is a part of a feedback control system, it is very important to know its phase characteristic. Phase lag reduces the phase margin of the system and may result in the overall instability.

Resonant (natural) frequency is a number expressed in Hz or rad/sec which shows where the sensor’s output signal increases considerably. Many sensors

![Figure 2.9](image_url)

**Figure 2.9.** Frequency characteristic (A) and response of a first order sensor (B) with limited upper and lower cutoff frequencies. $\tau_{c}$ and $\tau_{i}$ are corresponding time constants.

units of input stimulus per unit of time. Which response, frequency or speed, to specify in any particular case, depends on the sensor type, its application, and a preference of a designer.

Another way to specify speed response is by time which is required by the sensor to reach 90% of a steady-state or maximim level upon exposure to a step stimulus. For the first-order response, it is very convenient to use a so-called time constant. Time constant, $\tau$, is a measure of the sensor’s inertia. In electrical terms, it is equal to a product of electrical capacitance and resistance, $\tau=CR$. In thermal terms, thermal capacity and thermal resistances should be used instead. Practically, time constant can be easily measured. A first order system response is:

$$S = S_{c}(1 - e^{-\frac{t}{\tau}}).$$  \hspace{1cm} (2.14)

where $S_{c}$ is steady-state output, $t$ is time, and $e$ is base of natural logarithm.

Substituting 1 = $\tau$, we get:

$$\frac{S}{S_{c}} = 1 - \frac{t}{\tau} = 0.6321.$$  \hspace{1cm} (2.15)

In other words, after an elapsed time equal to one time constant, the response reaches about 63% of its steady-state level. Similarly, it can be shown that after two time constants, the height will be 86.5% and after three time constants it will be 95%.

**Lower cutoff frequency** shows what is the lowest frequency of stimulus the sensor can process. There is a lot of similarities between definitions of the upper and the lower cutoff frequencies. They are defined in the same terms and the time constants have the same meanings. It should be emphasized that while the upper cutoff frequency shows how fast the sensor reacts, the lower cutoff frequency shows how slowly changing stimuli the sensor can process. Figure 2.9B depicts the sensor’s

![Figure 2.10](image_url)

**Figure 2.10.** Types of responses. A—unlimited upper and lower frequencies; B—first order limited upper cutoff frequency; C—first order limited lower cutoff frequency; D—first order limited both upper and lower cutoff frequencies; E—narrow bandwidth response (resonant).
behave as linear, first-order systems which do not resonate. However, if a dynamic transducer’s output conforms to the standard curve of a second-order response, the manufacturer will state the natural frequency and the damping ratio of the transducer. The resonant frequency may be related to mechanical, thermal, or electrical properties of the detector. Generally, the operating frequency range for the sensor should be selected well below (at least 60%) or above the resonant frequency. However, in some sensors, the resonant frequency is the operating point. For instance, in glass breakage detectors (used in security systems) the resonant makes the sensor selectively sensitive to a narrow bandwidth which is specific for the acoustic spectrum produced by shattered glass.

Damping is the progressive reduction or suppression of the oscillation in the sensor having higher than the first order response. When the sensor’s response is as fast as possible without overshoot, the response is said to be critically damped (Fig. 2.11). Underdamped response is when the overshoot occurs and the overdamped response is slower than the critical. The damping ratio is a number expressing the quotient of the actual damping of a second-order linear transfer by its critical damping. The second order transfer function must include a quadratic factor: $s^2 + 2\zeta\omega_n s + \omega_n^2$, where $\omega_n$ is natural frequency (rad/sec), $s$ is the complex variable, and $\zeta$ is damping ratio. For a critically damped detector $\zeta = 1$. The damping factor is defined as:

$$z = \frac{\sigma}{\omega_n} = \frac{\text{complex damping}}{\omega_n^2}$$

where $\sigma$ is real part of a complex variable. For an oscillating response, as shown in Fig. 2.10, a damping factor is a measure of damping, expressed (without sign) as the quotient of the greater by the lesser of pair of consecutive swings in opposite directions of the output signal, about an ultimately steady-state value. Hence, the damping factor can be measured as:

![Figure 2.11. Responses of sensors with different damping characteristics.](image)

2.16 ENVIRONMENTAL FACTORS

Storage conditions are nonoperating environmental limits to which a sensor may be subjected during a specified period without permanently altering its performance under normal operating conditions. Usually, storage conditions include the highest and the lowest storage temperatures and maximum relative humidities at these temperatures. Words “noncondensing” may be added to the relative humidity number. Depending on the sensor’s nature, some specific limitation for the storage may need to be considered. For instance, maximum pressure, presence of some gases, or contaminating fumes, etc.

Short and long-term stabilities (drift) are parts of the accuracy specification. The short-term stability is manifested as changes in the sensor’s performance within minutes, hours or even days. Eventually, it is another way to express repeatability (see above) as drift may be bi-directional. That is, the sensor’s output signal may increase or decrease, which, in other terms, may be described as ultra-low-frequency noise. The long-term stability may be related to aging of the sensor materials, which is an irreversible change in the material’s electrical, mechanical, chemical, or thermal properties. That is, the long-term drift is usually unidirectional. It happens over a relatively long time span, such as months and years. Long-term stability is one of the most important requirements for the sensors that are used for precision measurements. Aging greatly depends on environmental storage and operating conditions, how well the sensor components are isolated from the environment and what materials are used for their fabrication. For instance, glass coated metal-oxide thermistors exhibit much greater long-term stability as compared with epoxy coated. A powerful way to improve long-term stability is to pre-age the component at extreme conditions. The extreme conditions may be cycled from the lowest to the highest. For instance, a sensor may be periodically swung from freezing to hot temperatures. Such accelerated aging not only enhances stability of the sensor’s characteristics, but also improves the reliability (see below), as the pre-aging process reveals many hidden defects. For instance, epoxy-coated thermistors may be substantially improved if they are maintained at $+150^\circ$ for 1 month before they are calibrated and installed into a product.

Environmental conditions to which a sensor is subjected do not include variables which the sensor measures. For instance, an air pressure sensor usually is subjected not just to air pressure, but to other influences as well, such as temperatures of air and surrounding components, humidity, vibration, ionizing radiation, electromagnetic fields, gravitational forces, etc. All these factors may and usually do affect the sensor’s performance. Both static and dynamic variations in these conditions should be considered. Some environmental conditions are of a multiplicative nature, that is, they alter a transfer function of the sensor, for instance changing its gain. One example is resistive strain gauge whose sensitivity increases with temperature.
Environmental stability is quite broad and usually a very important requirement. Both the sensor designer and the application engineer should consider all possible external factors which may affect the sensor’s performance. A piezoelectric accelerometer may generate spurious signals if affected by a sudden change in ambient temperature, electrostatic discharge, formation of electrical charges (triboelectric effect), vibration of a connecting cable, electromagnetic interferences (EMI), etc. Even if a manufacturer does not specify such effects, an application engineer should simulate them during the prototype phase of the design process. If, indeed, the environmental factors degrade the sensor’s performance, additional corrective measures may be required (see Chapter 4). For instance, placing the sensor in a protective box, electrical shielding, using a thermal insulation, or a thermostat.

Temperature factors are very important for sensor performance, and they must be known and accounted for. The operating temperature range is the span of ambient temperatures given by their upper and lower extremes (e.g., −20 to +100 °C) within which the sensor maintains its specified accuracy. Many sensors change with temperature and their transfer functions may shift significantly. Special compensating elements are often incorporated either directly into the sensor or into signal conditioning circuits, to compensate for temperature errors. The simplest way of specifying tolerances of thermal effects is provided by the error-band concept, which is simply the error band that is applicable over the operating temperature band. A temperature band may be divided into sections while the error band is separately specified for each section. For example, a sensor may be specified to have an accuracy of ±1% in the range from 0 to 50 °C, ±2% from −20 to 0 °C, and from +50 to 100 °C, and ±3% beyond these ranges within operating limits which are specified from −40 to +150 °C.

Temperatures will also affect dynamic characteristics, particularly when they employ viscous damping. A relatively fast temperature change may cause the sensor to generate a spurious output signal. For instance, a dual pyroelectric sensor in a motion detector is insensitive to slow varying ambient temperature. However, when the temperature changes fast, the sensor will generate electric current which may be recognized by a processing circuit as a valid response to a stimulus, thus causing a false positive detection.

A self-heating error may be specified when an excitation signal is absorbed by a sensor and changes its temperature by such a degree that it may affect its accuracy. For instance, a thermistor temperature sensor requires passage of electric current, causing heat dissipation within the sensor’s body. Depending on its coupling with the environment, the sensor’s temperature may increase due to a self-heating effect. This will result in errors in temperature measurement. The coupling depends on the media where the sensor operates—a dry contact, liquid, air, etc. A worst coupling may be through still air. For thermistors, manufacturers often specify self-heating errors in air, stirred liquid, or other media.

A sensor’s temperature increase above its surroundings may be found from the formula:

\[ \Delta T = \frac{V^2}{(4\pi^2 + a)R} \]  

where \( \xi \) is the sensor’s mass density, \( c \) is specific heat, \( v \) is the volume of the sensor, \( a \) is the coefficient of thermal coupling between the sensor and the outside (thermal conductivity), \( R \) is the electrical resistance, and \( V \) is the effective voltage across the resistance. If a self-heating results in an error, Eq. (2.20) may be used as a design guidance. For instance, to increase \( a \), a thermistor detector should be well coupled to the object by increasing the contact area, applying thermally conductive grease, or using thermally conductive adhesives. Also, high resistance sensors and low measurement voltages are preferable.

2.17 RELIABILITY

Reliability is the ability of a sensor to perform a required function under stated conditions for a stated period. It is expressed in statistical terms as a probability that the device will function without failure over a specified time or a number of uses. It should be noted, that reliability is not a characteristic of drift or noise stability. It specifies a failure, that is, temporary or permanent, exceeding the limits of a sensor’s performance under normal operating conditions.

Reliability is an important requirement, however, it is rarely specified by the sensor manufacturers. Probably, the reason is the absence of a commonly accepted measure for the term. In the U.S., for many electronic devices, the procedure for predicting in-service reliability is the MTBF (mean-time-between-failure) calculation described in the MIL-HDBK-217 standard. Its basic approach is to arrive at a MTBF rate for a device by calculating the individual failure rates of the individual components used and by factoring in the kind of operation the device will see: its temperature, stress, environmental, and screening level (measure of quality). Unfortunately, MTBF reflects reliability only indirectly and is often hardly applicable to everyday use of the device. The qualification tests on sensors are performed at combinations of the worst possible conditions. One approach (suggested by MIL-STD-883) is 1000 hours, loaded at maximum temperature. This test does not qualify for such important impacts as fast temperature changes. The most appropriate method of testing would be accelerated life qualification. It is a procedure that emulates the sensor’s operation, providing real-world stresses, but compressing years into weeks. Three goals are behind the test: to establish MTBF; to identify first failure points that can then be strengthened by design changes; and to identify the overall system practical life time.

One possible way to compress time is to use the same profile as actual operating cycle, including maximum loading and power-on, power-off cycles, but expanded environmental highest and lowest ranges (temperature, humidity, and pressure). The highest and lowest limits should be substantially broader than normal operating conditions. Performance characteristics may be outside specifications, but must return to those when the device is brought back to the specified operating range. For
example, if a sensor is specified to operate up to 50 °C at the highest relative humidity (RH) of 85% at maximum supply voltage of +15 V, it may be cycled up to 100 °C at 95% RH and at +18 V power supply. To estimate number of test cycles (n), the following empirical formula (developed by Sandstrand Aerospace, Rockford, Ill. and Interpoint Corp., Redmond, WA) [1] may be useful:

\[ n = N \left( \frac{\Delta T_{\text{test}}}{\Delta T_{\text{spec}}} \right)^{1.5} \]  

(2.21)

where \( N \) is the estimated number of cycles per lifetime, \( \Delta T_{\text{test}} \) is the maximum specified temperature fluctuation and \( \Delta T_{\text{spec}} \) maximum cycled temperature fluctuation during the test. For instance, if the normal temperature is 25 °C, the maximum specified temperature is 50 °C, cycling is up to 100 °C, and over the lifetime (say, 10 years), the sensor is estimated to be subjected to 20,000 cycles, then the number of test cycles is calculated as:

\[ n = 20,000 \left( \frac{50 - 25}{100 - 25} \right)^{1.5} = 1283. \]

As a result, the accelerated life test requires about 1300 cycles instead of 20,000.

It should be noted, however, that the 2.5 factor was derived from a solder fatigue multiple, since that element is heavily influenced by cycling. Some sensors have no solder connections at all, and some might have even more sensitivity to cycling than solder, for instance, electrically conductive epoxy. Then, the factor should be selected somewhat smaller. As a result of the accelerated life test, the reliability may be expressed as a probability of failure. For instance, if 2 out of 100 sensors (with an estimated life time of 10 years) failed the accelerated life test, the reliability is specified as 98% over 10 years.

A sensor, depending on its application, may be subjected to some other environmental effects which potentially can alter its performance or uncover hidden defects. Among such additional tests are:

- High temperature/high humidity while being fully electrically powered. For instance, a sensor may be subjected to its maximum allowable temperature at 85–95% relative humidity (RH) and kept under these conditions during 500 hours. This test is very useful for detecting contaminations and evaluation of packaging integrity. Life of sensors, operating at normal room temperatures, is often accelerated at 85 °C and 85% RH, that sometimes is called an "85–85 test."

- Mechanical shocks and vibrations may be used to simulate adverse environmental conditions, especially in evaluation wire bonds, adhesion of epoxy, etc. A sensor may be dropped to generate high level accelerations (up to 3000 g of force). The drops should be made on different axes. Harmonic vibrations should be applied to the sensor over the range which includes its natural frequency, in the U.S. military standard #750, methods 2016 and 2056 are often used for mechanical tests.

Extreme storage conditions may be simulated, for instance at +100 and −40 °C while maintaining a sensor for at least 1000 hours under these conditions. This test simulates storage and shipping conditions and usually is performed on non-operating devices. The upper and lower temperature limits must be consistent with the sensor’s physical nature. For example, a TGS pyroelectric sensors manufactured in the past by Philips are characterized by a Curie temperature of +60 °C. Approaching and surpassing this temperature results in a permanent destruction of sensitivity. Hence, the temperature of such sensors should never exceed +50 °C, which must be clearly specified and marked on its packaging material.

Thermal shock or temperature cycling (TC) is subjecting a sensor to alternate extreme conditions. For example, it may be dwelled for 30 minutes at −40 °C, then rapidly moved to +100 °C for 30 minutes, and then back to cold. The method must specify total number of cycling, like 100 or 1000. This test helps to uncover die bond, wire bond, epoxy connections, and packaging integrity.

To simulate sea conditions, sensors may be subjected to a salt spray atmosphere for a specified time, for example 24 hours. This helps to uncover its resistance to corrosion and structural defects.

2.18 APPLICATION CHARACTERISTICS

Design, weight, and overall dimensions are geared to specific areas of applications. Price may be a secondary issue when the sensor’s reliability and accuracy are of paramount importance. If a sensor is intended for life support equipment, weapons, or spacecraft, a high price tag may be well justified to assure high accuracy and reliability. On the other hand, for a very broad range of consumer applications, the price of a sensor often becomes a corner stone of a design.

For instance, human body temperatures preferably should be taken from the ear canal by thermal radiation thermometers. During the first several years after introduction of these instruments in 1986, the applications of ear thermometry were limited only to hospitals. The reason was that thermopile sensors (which are relatively expensive) were employed in the ear thermometers for sensing thermal radiation. As a result, the price tag for such an instrument was in the range of $500 to $700. In 1991, a home model of a medical infrared thermometer RM-1 was introduced for $15 of that price. It was made possible only because a thermopile was replaced by a pyroelectric sensor—a much less expensive sensor.

2.19 UNCERTAINTY

As it was noted above, no matter how accurate the measurement is, it’s only an approximation or estimate of the true value of the specific quantity subject to
measurement, that is the stimulus or measurand. Thus, the result of measurement should be considered complete only when accompanied by a quantitative statement of its uncertainty.

When taking individual measurements (samples) under noisy conditions we expect that stimulus $s$ is represented by the sensor as having a somewhat different value $s'$, so that the error in measurement is expressed as

$$
\delta = s' - s.
$$

(2.22)

The difference between the error that is specified by Eq. (2.22) and uncertainty should always be clearly understood. An error can be compensated to a certain degree by correcting its systematic component. The result of such a correction can unknowingly be very close to the unknown true value of the stimulus and thus will have a very small error. Yet, in spite of a small error, the uncertainty of measurement may be very large so we cannot really trust that the error is indeed that small. In other words, an error is what we unknowingly get when we measure, while uncertainty is what we think how large that error might be.

The International Committee for Weight and Measures (CIPM) considers that uncertainty consists of many factors that can be grouped into two classes or types [2,3]

- **A**: those which are evaluated by statistical methods;
- **B**: those which are evaluated by other means.

This division is not clear-cut and the borderline between A and B is somewhat illusive. Generally, A components of uncertainty arise from random effects, while the B components arise from systematic effects.

Type A uncertainty is generally specified by a standard deviation $\sigma$, equal to the positive square root of the statistically estimated variance $\sigma^2$, and the associated number of degrees of freedom $v$. For such a component the standard uncertainty is $u = \sigma / \sqrt{v}$. Standard uncertainty represents each component of uncertainty that contributes to the uncertainty of the measurement result.

- **A** standard uncertainty may be based on an any valid statistical method for treating data. Examples are calculating standard deviation of the mean of a series of independent observations, using the method of least squares to fit a curve to data in order to estimate the parameters of the curve and their standard deviations. If the measurement situation is especially complicated, one should consider obtaining the guidance of a statistician.

- **B** standard uncertainty is usually based on scientific judgment using all the relevant information available, such as:
  - previous measurement data.
  - experience with or general knowledge of the behavior and property of relevant sensors, materials, and instruments.
  - manufacturer’s specifications.

For detailed guidance of assessing and specifying standard uncertainties one should consult specialized texts, for instance [4].

When both A and B uncertainties are evaluated, they should be combined to represent the **combined standard uncertainty**. This can be done by using a conventional method for combining standard deviations. This method is often called the **law of propagation of uncertainty** and in common parlance is known as "root-sum-of-squares" (square root of the sum-of-squares), or the "RSS" method of combining uncertainty components estimated as standard deviations:

$$
\sqrt{u^2_1 + u^2_2 + \ldots + u^2_n}
$$

(2.23)

where $n$ is a number of standard uncertainties in the uncertainty budget.

Table 2.2 shows an example of uncertainty budget for an electronic thermometer with a thermistor sensor which measures temperature of a water bath. While compiling such a table one shall be very careful not to oversee any standard uncertainty not only in a sensor, but also in the interface instrument, experimental setup, and the object of measurement. This shall be done for various environmental conditions which may include temperature, humidity, atmospheric pressure, power supply variations, transmitted noise, aging, and many other factors.

No matter how accurately any individual measurement is made, that is, how close the measured temperature is to the true temperature of an object, one never can be

---

**TABLE 2.2. Uncertainty budget for thermistor thermometer.**

<table>
<thead>
<tr>
<th>Source of uncertainty</th>
<th>Standard uncertainty</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calibration of sensor</td>
<td>$0.03^{\circ}C$</td>
<td>B</td>
</tr>
<tr>
<td>Measured errors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repeated observations</td>
<td>$0.02^{\circ}C$</td>
<td>A</td>
</tr>
<tr>
<td>Sensor noise</td>
<td>$0.01^{\circ}C$</td>
<td>A</td>
</tr>
<tr>
<td>Amplifier noise</td>
<td>$0.006^{\circ}C$</td>
<td>A</td>
</tr>
<tr>
<td>Sensor aging</td>
<td>$0.025^{\circ}C$</td>
<td>B</td>
</tr>
<tr>
<td>Thermal bias through connecting wires</td>
<td>$0.015^{\circ}C$</td>
<td>A</td>
</tr>
<tr>
<td>Dynamic error due to sensor’s inertia</td>
<td>$0.005^{\circ}C$</td>
<td>B</td>
</tr>
<tr>
<td>Time relationship instability of object measurement</td>
<td>$0.04^{\circ}C$</td>
<td>A</td>
</tr>
<tr>
<td>Transmitted noise</td>
<td>$0.01^{\circ}C$</td>
<td>A</td>
</tr>
<tr>
<td>Mist of transfer function</td>
<td>$0.02^{\circ}C$</td>
<td>B</td>
</tr>
<tr>
<td>Ambient drifts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voltage reference</td>
<td>$0.01^{\circ}C$</td>
<td>A</td>
</tr>
<tr>
<td>Bridge instruments</td>
<td>$0.01^{\circ}C$</td>
<td>A</td>
</tr>
<tr>
<td>Diodelectric absorption in A/D capacitor</td>
<td>$0.005^{\circ}C$</td>
<td>B</td>
</tr>
<tr>
<td>Digital resolution</td>
<td>$0.01^{\circ}C$</td>
<td>A</td>
</tr>
<tr>
<td><strong>Combined standard uncertainty</strong></td>
<td>$0.058^{\circ}C$</td>
<td></td>
</tr>
</tbody>
</table>
sure that it is indeed accurate. The combined standard uncertainty of 0.068 °C does not mean that error of measurement is no greater than 0.068 °C. That value is just a standard deviation and if an observer has enough patience he may find that individual errors may be much larger. The word “uncertainty” by its very nature, implies that the uncertainty of the result of a measurement is an estimate and generally does not have well-defined limits.

REFERENCES


CHAPTER 3

PHYSICAL PRINCIPLES OF SENSING

"The way we have to describe Nature is generally incomprehensible to us."
—RICHARD P. FEYNMAN.

"QED. THE STRANGE THEORY OF LIGHT AND MATTER"

"It should be possible to explain the laws of physics to a barmaid."
—ALBERT EINSTEIN

Since a sensor is a converter of generally nonelectrical effects into electrical signals, often several transformation steps are required before the electric output signal can be generated. These steps involve changes of types of energy where the final step must produce an electrical signal of a desirable format. There are several physical effects which cause generation of electrical signals in response to nonelectrical influences. Examples are thermoelectric (Seebeck) effect, piezoelectricity, and photovoltaic effect. However, many stimuli can not be directly converted into electricity, thus multiple conversion steps would be required. If, for instance, one wants to detect displacement of an opaque object, a fiber-optic sensor can be employed. A pilot (excitation) signal is generated by a photodiode, transmitted via an optical fiber to the object and reflected from its surface. The reflected photon flux enters the receiving optical fiber and propagates toward a photodiode where it produces an electric current representing the distance from the fiber-optic end to the object. We see that such a sensor involves transformation of electrical current into photons, propagation of photons through some refractive media, reflection, and conversion back into electric current. Therefore, such a sensing process includes two energy conversion steps and a manipulation of optical signal as well.

This chapter examines various physical effects which can be used for conversion of stimuli into electric signals. Also, here we describe mechanical, thermal, optical and other properties of materials and sensor elements which can be employed in the conversion steps and manipulation of nonelectrical signals. Circuits for manipulation of electrical signals are described in the next chapter.