SENSORS AND SIGNAL CONDITIONING

Second Edition

Ramon Pallàs-Areny / John G. Webster

SENSORS AND SIGNAL CONDITIONING

SENSORS AND SIGNAL CONDITIONING

Second Edition

RAMON PALLÀS-ARENY
Universitat Politècnica de Catalunya

JOHN G. WEBSTER
University of Wisconsin—Madison

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.
New York • Chichester • Weinheim • Brisbane • Singapore • Toronto

This book is printed on acid-free paper.⊗

Published simultaneously in Canada.

For ordering and customer service, call 1-800-CALL-WILEY.

Library of Congress Cataloging-in-Publication Data:
Pallàs-Areny, Ramon.
Sensors and signal conditioning/Ramon Pallàs-Areny, John G. Webster.—2nd ed.
p. cm.
"A Wiley-Interscience publication."
Includes bibliographical references.
ISBN 0-471-33232-1 (cloth: alk. paper)
1. Transducers. 2. Detectors. 3. Interface circuits. I. Webster, John G., 1932- II. Title.
TK7872.T6 P25 2000
621.3815—dc21 00-028293

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

CONTENTS

PREFACE

Sensors have been traditionally used for industrial process control, measurement, and automation, often involving temperature, pressure, flow, and level measurement. Nowadays, sensors enable a myriad of applications fostered by developments in digital electronics and involving the measurement of several physical and chemical quantities in automobiles, aircraft, medical products, office machines, personal computers, consumer electronics, home appliances, and pollution control.

Many of the new application areas for sensors do not pose any severe working conditions and are high-volume consumers. This makes those applications a target for semiconductor-based sensors, particularly sensors built by microfabrication techniques (microsensors), which can be manufactured in large scale. Annual sales of accelerometers and pressure sensors in the automotive industry, along with the annual sales of blood pressure sensors in the medical industry, amount to tens of millions units. Gas sensors, rate sensors, CMOS image sensors, and biosensors can similarly boom.

Classical sensors (or macrosensors) have not been superseded by the new microsensors. Many conventional sensors are still required for specialized applications, so there is no replacement for them in the foreseeable future. Nevertheless, the performance of several integrated circuits commonly used in signal conditioning has improved and allows the design of simpler circuits. Also, there are specific integrated circuits intended for conditioning the signals of common sensors such as thermocouples, RTDs, capacitive sensors, and LVDTs, and microcontrollers have become an inexpensive resource for low-cost, low-resolution analog-to-digital interfacing. Furthermore, the low cost of digital computing has moved part of the calculations and compensations closer to the sensor. The communication with a central controller is increasingly digital, and intelligent (or smart) sensors are being installed in new factories.

This second edition responds to this new scenario from the same point of view of the first edition: that of electronic engineering students or professionals interested in designing measurement systems using available sensors and integrated circuits. For each sensor we describe the working principle, advantages, limitations, types, equivalent circuit, and relevant applications. To clarify sensor types and materials, there is a new section on sensor materials and another on microsensor technology. Microsensors available for different applications are mentioned in the corresponding sections. Sensors are grouped depending on whether (a) they are variable resistors, inductors, capacitors, (b) they generate voltage, charge, or current, or (c) they are digital, semiconductor-junction based, or use some form of radiation. This approach simplifies the study of signal conditioners, which are instrumental in embedding sensors in any electronic system. Basic measurement methods and primary sensors for common physical quantities are described in an expanded section. Further information can be found in J. G. Webster (ed.), *The Measurement, Instrumentation, and Sensors Handbook*, CRC Press, 1999.

Some new sensors covered are giant magnetoresistive sensors, resistive gas sensors, liquid conductivity sensors, magnetostrictive sensors, SQUIDs, fluxgate magnetometers, Wiegand and pulse-wire sensors, position-sensitive detectors (PSDs), semiconductor-junction nuclear radiation detectors, CMOS image sensors, and biosensors. Several of these have moved from the research stage to the commercialization stage since the publication of the first edition. Velocity sensors, fiber-optic sensors, and chemical sensors, in general, receive expanded coverage because of their wider use.

Signal conditioners use new ICs with improved parameters, which often enable novel approaches to circuit design. Some new topics are error analysis of single-ended amplifiers, current feedback amplifiers, composite amplifiers, and IC current integrators. The section on noise now includes noise fundamentals, noise analysis of transimpedance and charge amplifiers, and noise and drift in resistors. Chapter 8, on digital and intelligent sensors, has been expanded by adding sections on variable oscillators including a sensor, direct microcomputer interfacing, sensor communications, and intelligent sensors.

Because the selection of the sensor influences the sensitivity, accuracy, and stability of the measurement system, we describe a broad range of sensors and list the actual specifications of several commercial sensors in tables elsewhere in the book. We have summarized several relevant specifications of common integrated circuits for signal conditioning in tables. New sections deal with basic statistical analysis of measurement results, and reliability. We give 68 worked-out examples and include a total of 103 end-of-chapter problems, many from actual design cases. The annotated solution to the problems is in an appendix at the end of the book. End-of-chapter references have been updated. For ease of reference, figures for examples or problems are respectively preceded by an E or a P. Line crossings in figures are not a connection, unless indicated by a dot.

In the study of any field, the knowledge of important dates adds perspective. Hence, this book names the discoverer and approximate date of the discovery of different physical laws applied in sensors. This may also help in preventing professionals from thinking that sensors are subsequent to the transistor (1947), the operational amplifier (1963), or the microprocessor (1971). Some sensors existed long before all of them. It is the work of electronic engineers to apply all the capabilities of integrated circuits in order that the information provided by sensors results in more economical, reliable, and efficient systems for the benefit of the humans, who certainly have limited perception but who have unmatched intelligence and creativity.

RAMON PALLÀS-ARENY
elerpa@eel.upc.es

JOHN G. WEBSTER
webster@engr.wisc.edu

Barcelona, Spain
Madison, Wisconsin
August, 2000

SENSORS AND SIGNAL CONDITIONING

1

INTRODUCTION TO SENSOR-BASED MEASUREMENT SYSTEMS

Measurements pervade our life. Industry, commerce, medicine, and science rely on measurements. Sensors enable measurements because they yield electric signals with embedded information about the measurand. Electronic circuits process those signals in order to extract that information. Hence, sensors are the basis of measurement systems. This chapter describes the basics of sensors, their static and dynamic characteristics, primary sensors for common quantities, and sensor materials and technology.

1.1 GENERAL CONCEPTS AND TERMINOLOGY

1.1.1 Measurement Systems

A system is a combination of two or more elements, subsystems, and parts necessary to carry out one or more functions. The function of a measurement system is the objective and empirical assignment of a number to a property or quality of an object or event in order to describe it. That is, the result of a measurement must be independent of the observer (objective) and experimentally based (empirical). Numerical quantities must fulfill the same relations fulfilled by the described properties. For example, if a given object has a property larger than the same property in another object, the numerical result when measuring the first object must exceed that when measuring the second object.

One objective of a measurement can be process monitoring: for example, ambient temperature measurement, gas and water volume measurement, and clinical monitoring. Another objective can be process control: for example, for temperature or level control in a tank. Another objective could be to assist

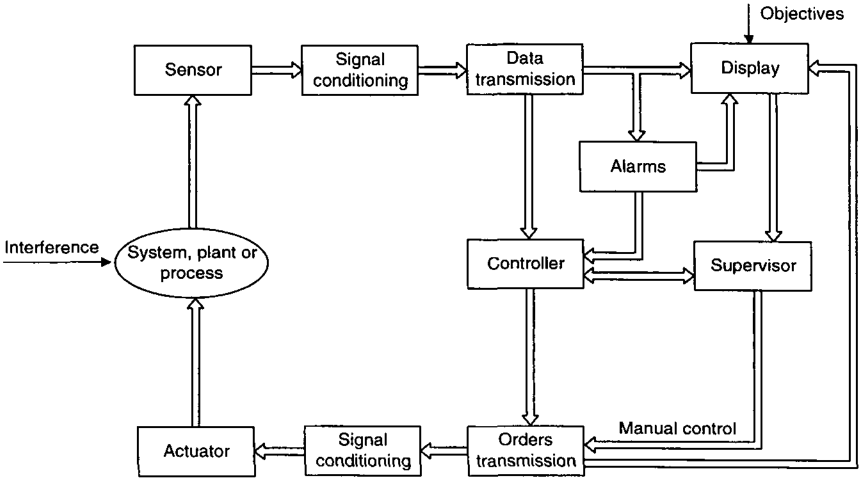

Figure 1.1 Functions and data flow in a measurement and control system. Sensors and actuators are transducers at the physical interface between electronic systems and processes or experiments.

experimental engineering: for example, to study temperature distribution inside an irregularly shaped object or to determine force distribution on a dummy driver in a car crash. Because of the nature of the desired information and its quantity, computer-aided design (CAD) does not yield complete data for these experiments. Thus measurements in prototypes are also necessary to verify the results of computer simulations.

Figure 1.1 shows the functions and data flow of a measurement and control system. In general, in addition to the acquisition of information carried out by a sensor, a measurement requires the processing of that information and the presentation of the result in order to make it perceptible to human senses. Any of these functions can be local or remote, but remote functions require information transmission. Modern measurement systems are not physically arranged according to the data flow in Figure 1.1 but are instead arranged according to their connection to the digital bus communicating different subsystems (Sections 8.6 and 8.7).

1.1.2 Transducers, Sensors, and Actuators

A *transducer* is a device that converts a signal from one physical form to a corresponding signal having a different physical form. Therefore, it is an energy converter. This means that the input signal always has energy or power; that is, signals consist of two component quantities whose product has energy or power dimension. But in measurement systems, one of the two components of the measured signal is usually so small that it is negligible, and thus only the remaining component is measured.

When measuring a force, for example, we assume that the displacement in the transducer is insignificant. That is, that there is no "loading" effect. Otherwise it might happen that the measured force is unable to deliver the needed energy to allow the movement. But there is always some power taken by the transducer, so we must ensure that the measured system is not perturbed by the measuring action.

Since there are six different kinds of signals—mechanical, thermal, magnetic, electric, chemical, and radiation (corpuscular and electromagnetic, including light)—any device converting signals of one kind to signals of a different kind is a transducer. The resulting signals can be of any useful physical form. Devices offering an electric output are called *sensors*. Most measurement systems use electric signals, and hence rely on sensors. Electronic measurement systems provide the following advantages:

1. Sensors can be designed for any nonelectric quantity, by selecting an appropriate material. Any variation in a nonelectric parameter implies a variation in an electric parameter because of the electronic structure of matter.
2. Energy does not need to be drained from the process being measured because sensor output signals can be amplified. Electronic amplifiers yield (low) power gains exceeding 10^{10} in a single stage. The energy of the amplifier output comes from its power supply. The amplifier input signal only controls (modulates) that energy.
3. There is a variety of integrated circuits available for electric signal conditioning or modification. Some sensors integrate these conditioners in a single package.
4. Many options exist for information display or recording by electronic means. These permit us to handle numerical data and text, graphics, and diagrams.
5. Signal transmission is more versatile for electric signals. Mechanical, hydraulic, or pneumatic signals may be appropriate in some circumstances, such as in environments where ionizing radiation or explosive atmospheres are present, but electric signals prevail.

Sensor and transducer are sometimes used as synonymous terms. However, sensor suggests the extension of our capacity to acquire information about physical quantities not perceived by human senses because of their subliminal nature or minuteness. Transducer implies that input and output quantities are not the same. A sensor may not be a transducer. The word *modifier* has been proposed for instances where input and output quantities are the same, but it has not been widely accepted.

The distinction between input-transducer (physical signal/electric signal) and output-transducer (electric signal/display or actuation) is seldom used at present. Nowadays, input transducers are termed *sensors*, or *detectors* for radiation, and output transducers are termed *actuators* or *effectors*. Sensors are intended to acquire information. Actuators are designed mainly for power conversion.

Sometimes, particularly when measuring mechanical quantities, a *primary sensor* converts the measurand into a measuring signal. Then a sensor would convert that signal into an electric signal. For example, a diaphragm is a primary sensor that stresses when subject to a pressure difference, and strain gages (Section 1.7.2 and Section 2.2) sense that stress. In this book we will designate as sensor the whole device, including the package and leads. We must realize, however, that we cannot directly perceive signals emerging from sensors unless they are further processed.

1.1.3 Signal Conditioning and Display

Signal conditioners are measuring system elements that start with an electric sensor output signal and then yield a signal suitable for transmission, display, or recording, or that better meet the requirements of a subsequent standard equipment or device. They normally consist of electronic circuits performing any of the following functions: amplification, level shifting, filtering, impedance matching, modulation, and demodulation. Some standards call the sensor plus signal conditioner subsystem a *transmitter*.

One of the stages of measuring systems is usually digital and the sensor output is analog. Analog-to-digital converters (ADCs) yield a digital code from an analog signal. ADCs have relatively low input impedance, and they require their input signal to be dc or slowly varying, with amplitude within specified margins, usually less than ± 10 V. Therefore, sensor output signals, which may have an amplitude in the millivolt range, must be conditioned before they can be applied to the ADC.

The display of measured results can be in an analog (optical, acoustic, or tactile) or in a digital (optical) form. The recording can be magnetic, electronic, or on paper, but the information to be recorded should always be in electrical form.

1.1.4 Interfaces, Data Domains, and Conversion

In measurement systems, the functions of signal sensing, conditioning, processing, and display are not always divided into physically distinct elements. Furthermore, the border between signal conditioning and processing may be indistinct. But generally there is a need for some signal processing of the sensor output signal before its end use. Some authors use the term *interface* to refer to signal-modifying elements that operate in the electrical domain, even when changing from one data domain to another, such as an ADC.

A *data domain* is the name of a quantity used to represent or transmit information. The concept of data domains and conversion between domains helps in describing sensors and electronic circuits associated with them [1]. Figure 1.2 shows some possible domains, most of which are electrical.

In the *analog domain* the information is carried by signal amplitude (i.e.,

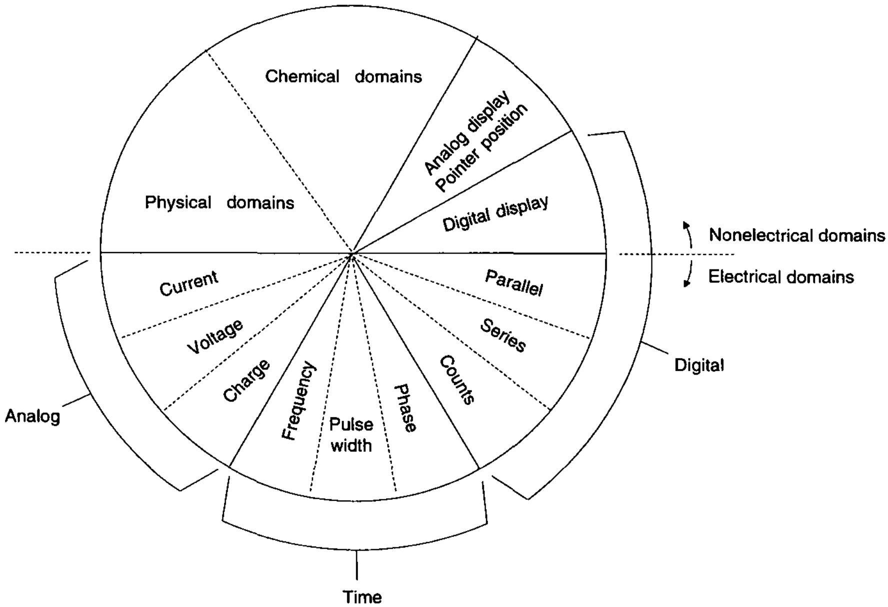

Figure 1.2 Data domains are quantities used to represent or transmit information [1]. (From H. V. Malmstadt, C. G. Enke, and S. R. Crouch, *Electronics and Instrumentation for Scientists*, copyright 1981. Reprinted by permission of Benjamin/Cummings, Menlo Park, CA.)

charge, voltage, current, or power). In the *time domain* the information is not carried by amplitude but by time relations (period or frequency, pulse width, or phase). In the *digital domain*, signals have only two values. The information can be carried by the number of pulses or by a coded serial or parallel word.

The analog domain is the most prone to electrical interference (Section 1.3.1). In the time domain, the coded variable cannot be measured—that is, converted to the numerical domain—in a continuous way. Rather, a cycle or pulse duration must elapse. In the digital domain, numbers are easily displayed.

The structure of a measurement system can be described then in terms of domain conversions and changes, depending on the direct or indirect nature of the measurement method.

Direct physical measurements yield quantitative information about a physical object or action by direct comparison with a reference quantity. This comparison is sometimes simply mechanical, as in a weighing scale.

In *indirect physical measurements* the quantity of interest is calculated by applying an equation that describes the law relating other quantities measured with a device, usually an electric one. For example, one measures the mechanical power transmitted by a shaft by multiplying the measured torque and speed of rotation, the electric resistance by dividing dc voltage by current, or the traveled distance by integrating the speed. Many measurements are indirect.

1.2 SENSOR CLASSIFICATION

A great number of sensors are available for different physical quantities. In order to study them, it is advisable first to classify sensors according to some criterion. White [10] provides additional criteria to those used here.

In considering the need for a power supply, sensors are classified as modulating or self-generating. In modulating (or active) sensors, most of the output signal power comes from an auxiliary power source. The input only controls the output. Conversely, in self-generating (or passive) sensors, output power comes from the input.

Modulating sensors usually require more wires than self-generating sensors, because wires different from the signal wires supply power. Moreover, the presence of an auxiliary power source can increase the danger of explosion in explosive atmospheres. Modulating sensors have the advantage that the power supply voltage can modify their overall sensitivity. Some authors use the terms *active* for self-generating and *passive* for modulating. To avoid confusion, we will not use these terms.

In considering output signals, we classify sensors as analog or digital. In *analog sensors* the output changes in a continuous way at a macroscopic level. The information is usually obtained from the amplitude, although sensors with output in the time domain are usually considered as analog. Sensors whose output is a variable frequency are called *quasi-digital* because it is very easy to obtain a digital output from them (by counting for a time).

The output of *digital sensors* takes the form of discrete steps or states. Digital sensors do not require an ADC, and their output is easier to transmit than that of analog sensors. Digital output is also more repeatable and reliable and often more accurate. But regrettably, digital sensors cannot measure many physical quantities.

In considering the operating mode, sensors are classified in terms of their function in a deflection or a null mode. In *deflection sensors* the measured quantity produces a physical effect that generates in some part of the instrument a similar but opposing effect that is related to some useful variable. For example, a dynamometer to measure force is a sensor where the force to be measured deflects a spring to the point where the force it exerts, proportional to its deformation, balances the applied force.

Null-type sensors attempt to prevent deflection from the null point by applying a known effect that opposes that produced by the quantity being measured. There is an imbalance detector and some means to restore balance. In a weighing scale, for example, the placement of a mass on a pan produces an imbalance indicated by a pointer. The user has to place one or more calibrated weights on the other pan until a balance is reached, which can be observed from the pointer's position.

Null measurements are usually more accurate because the opposing known effect can be calibrated against a high-precision standard or a reference quantity. The imbalance detector only measures near zero; therefore it can be very

TABLE 1.1 Sensor Classifications According to Different Exhaustive Criteria

Criterion	Classes	Examples
Power supply	Modulating	Thermistor
	Self-generating	Thermocouple
Output signal	Analog	Potentiometer
	Digital	Position encoder
Operation mode	Deflection	Deflection accelerometer
	Null	Servo-accelerometer

sensitive and does not require any calibration. Nevertheless, null measurements are slow; and despite attempts at automation using a servomechanism, their response time is usually not as short as that of deflection systems.

In considering the input–output relationship, sensors can be classified as zero, first, second, or higher order (Section 1.5). The order is related to the number of independent energy-storing elements present in the sensor, and this affects its accuracy and speed. Such classification is important when the sensor is part of a closed-loop control system because excessive delay may lead to oscillation [6].

Table 1.1 compares the classification criteria above and gives examples for each type in different measurement situations. In order to study these myriad devices, it is customary to classify them according to the measurand. Consequently we speak of sensors for temperature, pressure, flow, level, humidity and moisture, pH, chemical composition, odor, position, velocity, acceleration, force, torque, density, and so forth. This classification, however, can hardly be exhaustive because of the seemingly unlimited number of measurable quantities. Consider, for example, the variety of pollutants in the air or the number of different proteins inside the human body whose detection is of interest.

Electronic engineers prefer to classify sensors according to the variable electrical quantity—resistance, capacity, inductance—and then to add sensors generating voltage, charge, or current, and other sensors not included in the preceding groups, mainly *p–n* junctions and radiation-based sensors. This approach reduces the number of groups and enables the direct study of the associated signal conditioners. Table 1.2 summarizes the usual sensors and sensing methods for common quantities.

1.3 GENERAL INPUT–OUTPUT CONFIGURATION

1.3.1 Interfering and Modifying Inputs

In a measurement system the sensor is chosen to gather information about the measured quantity and to convert it to an electric signal. A priori it would be unreasonable to expect the sensor to be sensitive to only the quantity of interest

TABLE 1.2 Usual Sensors and Sensing Methods for Common Quantities

Sensor type	Acceleration Vibration	Flow Rate Point velocity	Force	Humidity Moisture
Resistive	Mass–spring + strain gage	Anemometer	Strain gage	Humistor
		Thermistor		
		Target + strain gage		
Capacitive	Mass–spring + variable capacitor		Capacitive strain gage	Dielectric-variation capacitor
Inductive and electromagnetic	Mass–spring + LVDT	Faraday's law	Load cell + LVDT	
		Rotameter + LVDT	Magnetostriction	
Self-generating	Mass–spring + piezoelectric sensor	Thermal transport + thermocouple	Piezoelectric sensor	
Digital		Impeller, turbine		SAW sensor
		Positive displacement		
		Vortex shedding		
PN junction				
Optic, fiber optic		Laser anemometry		Chilled mirror
Ultrasound		Doppler effect		
		Travel time		
		Vortex		
Other		Differential pressure		
		Variable area + level sensor (open channel)		
		Variable area + displacement		
		Coriolis effect + force		

Quantity				
Level	Position Distance Displacement	Pressure	Temperature	Velocity Speed
Float + potentiometer	Magnetoresistor	Bourdon tube + potentiometer	RTD	
LDR	Potentiometer	Diaphragm + strain gage	Thermistor	
Thermistor	Strain gage			
Variable capacitor	Differential capacitor	Diaphragm + variable capacitor		
Magnetostriction	Eddy currents	Diaphragm + LVDT		Eddy currents
Magnetoresistive	Hall effect	Diaphragm + variable reluctance		Hall effect
Float + LVDT	Inductosyn			Faraday's law
Eddy currents	LVDT			LVT
	Resolver, synchro			
	Magnetostriction			
		Piezoelectric sensor	Pyroelectric sensor	
			Thermocouple	
Vibrating rod	Position encoder	Bourdon tube + encoder	Quartz oscillator	Incremental encoder
Float + pulley		Bourdon tube or bellows + quartz resonator		
		Diaphragm + vibrating wire		
Photoelectric	Photoelectric sensor		Diode	
			Bipolar transistor	
			T/I converter	
		Diaphragm + light reflection		
Absorption	Travel time			Doppler effect
Travel time				
Differential pressure		Liquid-based manometer + level sensor		
Microwave radar				
Nuclear radiation				

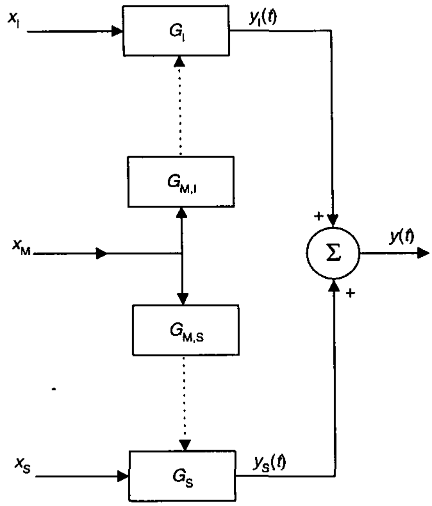

Figure 1.3 Effect of internal and external perturbations on measurement systems. x_S is the signal of interest. $y(t)$ is the system output. x_I is an interference or external perturbation. x_M is a modifying input. (From E. O. Doebelin, *Measurement Systems Application and Design*, 4th ed., copyright 1990. Reprinted by permission of McGraw-Hill, New York.)

and also to expect the output signal to be entirely due to the input signal. No measurement is ever obtained under ideal circumstances; therefore we must address real situations. We follow here the method proposed by Doebelin [2]. Figure 1.3 shows a general block diagram for classifying desired signal gains and interfering input gain for instruments. The desired signal x_S passes through the gain block G_S to the output y. Interfering inputs x_I represent quantities to which the instrument is unintentionally sensitive. These pass through the gain block G_I to the output y. Modifying inputs x_M are the quantities that through $G_{M,S}$ cause a change in G_S for the desired signal and through $G_{M,I}$ cause a change in G_I for interfering inputs. The gains G can be linear, nonlinear, varying, or random.

For example, to measure a force, it is common to use strain gages (Section 2.2). Strain gages operate on the basis of variation in the electric resistance of a conductor or semiconductor when stressed. Because temperature change also yields a resistance variation, we can regard any temperature variation as an interference or external disturbance x_I with gain G_I. At the same time, to measure resistance changes as a result of the stress, an electronic amplifier is required. Since any temperature change x_M through $G_{M,S}$ affects the amplifier gain G_S and therefore the output, it turns out that a temperature variation also acts as a modifying input x_M. If the same force is measured with a capacitive gage (Section 4.1), a temperature variation does not interfere but can still modify the amplifier gain.

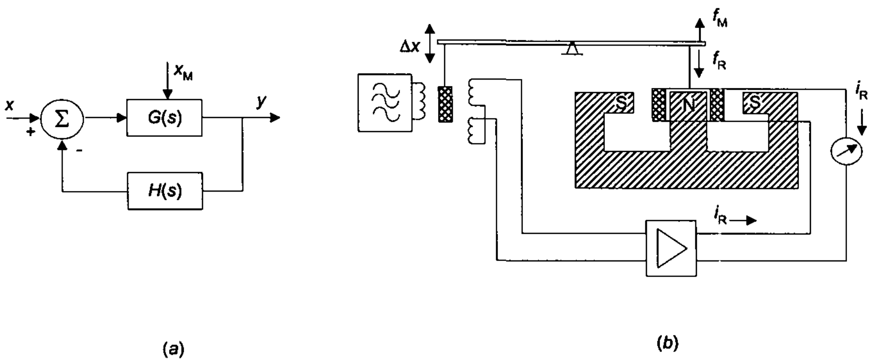

Figure 1.4 (*a*) Negative feedback method to reduce the effect of internal perturbations. Block H may be insensitive to those perturbations because it handles lower power than block G. (*b*) Force-to-current converter that relies on negative feedback and a balance sensor.

1.3.2 Compensation Techniques

The effects of interfering and modifying inputs can be reduced by changing the system design or by adding new elements to it. The best approach is to design systems insensitive to interference and that respond only to the desired signals. In the preceding example, it would have been best to use strain gages with a low temperature coefficient ($G_I = 0$). Thin, narrow, long magnetic sensors are only sensitive to magnetic fields parallel to their long dimension. In designing sensors for vector mechanical quantities, it would be best to obtain a unidirectional sensitivity and a low transverse sensitivity—that is, in directions perpendicular to the desired direction. In electronic circuits, low-drift components such as metal-film resistors and NP0 capacitors are less sensitive to temperature. Nevertheless, this method is not always possible for obvious practical reasons.

Negative feedback is a common method to reduce the effect of modifying inputs, and it is the method used in null measurement systems. Figure 1.4*a* shows the working principle. It assumes that the measurement system and the feedback are linear and can be described by their respective transfer functions $G(s)$ and $H(s)$. The input–output relation is

$$\frac{Y(s)}{X(s)} = \frac{G(s)}{1 + G(s)H(s)} \cong \frac{1}{H(s)} \tag{1.1}$$

where the approximation is valid when $G(s)H(s) \gg 1$. If the negative feedback is insensitive to the modifying input, and it has been designed so that the system remains stable, then the output signal is not affected by the modifying input.

The advantage of such a solution stems from the different physical characteristics of the elements described by $G(s)$ and $H(s)$. The probable insensitivity of H to a modifying input is a consequence of its lower power-handling capacity than G. This also results in higher accuracy and linearity for H. Moreover, negative feedback results in less energy extracted from the measured system because G is designed very large. The force-to-current converter in Figure 1.4*b* relies on negative feedback. The force to be measured, f_M, is compared with a restoring force f_R, generated by an internal moving-coil system. f_R is proportional to the current i_R in the coil, and i_R is proportional to the output voltage from the displacement sensor—here an LVDT (Section 4.2.3)—that senses the balance between f_M and f_R. If the amplifier gain is high enough, a very small input voltage from the sensor yields a current high enough to produce a force f_R able to balance f_M. Because i_R is proportional to f_R and $f_R \approx f_M$, we can determine f_M from i_R, regardless, for example, of the sensor linearity.

Filtering is a common method for interference reduction. A filter is any device that separates signals according to their frequency or another criterion. Filters are very effective when frequency spectra of signals and interference do not overlap. Filters can be placed at the input or at any intermediate stage. They can be electric, mechanical (e.g., to reduce vibrations), pneumatic, thermal (e.g., a high mass covering to reduce turbulence effects when measuring the average temperature of a flowing fluid), or electromagnetic. Filters placed at intermediate stages are usually electric.

Another common compensation technique for interfering and modifying inputs is the use of opposing inputs, often applied to compensate for temperature variations. If, for example, a gain that depends on a resistor having a positive temperature coefficient changes due to a temperature change, another resistor can be placed in series with the affected resistor. If the added resistor has a negative temperature coefficient, it is possible to keep the gain constant in spite of temperature changes. This method is also used for temperature compensation in strain gages, sensor-bridge supply, catalytic gas sensors, resistive gas sensors, and copper-wire coils (e.g., in electromagnetic relays, galvanometers, and tachometers), as well as to compensate vibration in piezoelectric sensors.

Finally, when the mathematical relationship between the interference and sensor output is known, interference can be compensated by digital calculation after measuring the magnitude of the interfering variable—for example, temperature in a pressure sensor. This method is common in smart sensors.

1.4 STATIC CHARACTERISTICS OF MEASUREMENT SYSTEMS

Because the sensor influences the characteristics of the whole measurement system, it is important to describe its behavior in a meaningful way. In most measurement systems the quantity to be measured changes so slowly that it is only necessary to know the static characteristics of sensors.

Nevertheless, the static characteristics influence also the dynamic behavior of the sensor—that is, its behavior when the measured quantity changes with time. However, the mathematical description of the joint consideration of static and dynamic characteristics is complex. As a result, static and dynamic behavior are studied separately. The concepts used to describe static characteristics are not exclusive to sensors. They are common to all measurement instruments.

1.4.1 Accuracy, Precision, and Sensitivity

Accuracy is the quality that characterizes the capacity of a measuring instrument for giving results close to the true value of the measured quantity. The "true," "exact," or "ideal" value is the value that would be obtained by a perfect measurement. It follows that true values are, by nature, indeterminate. The conventional true value of a quantity is "the value attributed to a particular quantity and accepted, sometimes by convention, as having an uncertainty appropriate for a given purpose" [3].

Sensor accuracy is determined through static calibration. It consists of keeping constant all sensor inputs, except the one to be studied. This input is changed very slowly, thus taking successive constant values along the measurement range. The successive sensor output results are then recorded. Their plot against input values forms the calibration curve. Obviously each value of the input quantity must be known. Measurement standards are such known quantities. Their values should be at least ten times more accurate than that of the sensor being calibrated.

Any discrepancy between the true value for the measured quantity and the instrument reading is called an *error*. The difference between measurement result and the true value is called *absolute error*. Sometimes it is given as a percentage of the maximal value that can be measured with the instrument (full-scale output, FSO) or with respect to the difference between the maximal and the minimal measurable values—that is, the measurement range or *span*. Therefore we have

$$\text{Absolute error} = \text{Result} - \text{True value}$$

The common practice, however, is to specify the error as a quotient between the absolute error and the true value for the measured quantity. This quotient is called the *relative error*. Relative error usually consists of two parts: one given as a percentage of the reading and another that is constant (see Problem 1.1). The constant part can be expressed as a percentage of the FSO, a threshold value, a number of counts in digital instruments, or a combination of these. Then,

$$\text{Relative error} = \frac{\text{Absolute error}}{\text{True value}}$$

Because true values are indeterminate, error calculations use a conventional true value.

Some sensors have a relative error specified only as a percentage of the FSO. If the measurement range includes small values, the full-scale specification implies that for them the measurement error is very large. Some sensors have a relative error specified as a percentage of the reading. If the measurement range includes small values, the percent-of-reading specification implies unbelievable low errors for small quantities.

The *Accuracy Class* concept facilitates the comparison of several sensors with respect to their accuracy. All the sensors belonging to the same class have the same measurement error when the applied input does not exceed their nominal range and work under some specified measurement conditions. That error value is called the *index of class*. It is defined as the percent measurement error, referred to a conventional value that is the measurement range or the FSO. For example, a class 0.2 displacement sensor whose end-of-scale displacement is 10 mm, in the specified reference conditions, has an error lower than 20 μm when measuring any displacement inside its measuring range.

The measured value and its error must be expressed with consistent numerical values. That is, the numerical result of the measurement must not have more figures than those that can be deemed reliable by considering the uncertainty of the result. For example, when measuring ambient temperature, $20\,^{\circ}\mathrm{C} \pm 1\,^{\circ}\mathrm{C}$ is a result correctly expressed, while $20\,^{\circ}\mathrm{C} \pm 0.1\,^{\circ}\mathrm{C}$, $20.5\,^{\circ}\mathrm{C} \pm 1\,^{\circ}\mathrm{C}$ and $20.5\,^{\circ}\mathrm{C} \pm 10\%$ are incorrect expressions because the measured value and the error have different uncertainty (see Problem 1.2).

Care must be taken also when converting units to avoid false gains of accuracy. For example, a 19.0 inch length (1 inch = 25.4 mm) should not be directly expressed as 482.6 mm, since the original figure suggests an uncertainty of tenths of an inch while the converted figure indicates an uncertainty of tenths of a millimeter. That is, the original result indicates that the length is between 485 mm and 480 mm, while the converted result would suggest that it is between 482.5 mm and 482.7 mm.

Precision is the quality that characterizes the capability of a measuring instrument of giving the same reading when repetitively measuring the same quantity under the same prescribed conditions (environmental, operator, etc.), without regard for the coincidence or discrepancy between the result and the true value. Precision implies an agreement between successive readings and a high number of significant figures in the result. Therefore, it is a necessary but not sufficient condition for accuracy. Figure 1.5 shows different possible situations.

The *repeatability* is the closeness of agreement between successive results obtained with the same method under the same conditions and in a short time interval. Quantitatively, the repeatability is the minimum value that exceeds, with a specified probability, the absolute value of the difference between two successive readings obtained under the specified conditions. If not stated, it is assumed that the probability level is 95%.

The *reproducibility* is also related to the degree of coincidence between successive readings when the same quantity is measured with a given method, but in this case with a long-term set of measurements or with measurements carried out by different people or performed with different instruments or in different laboratories. Quantitatively, the reproducibility is the minimal value that

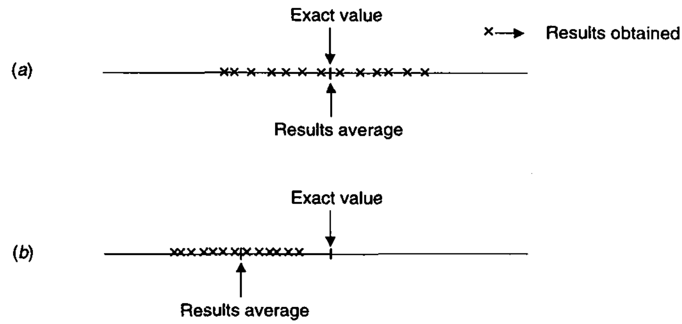

Figure 1.5 Measurement situations illustrating the difference between accuracy and precision. In case (*a*) there is a high accuracy and a low repeatability. In case (*b*) the repeatability is higher but there is a low accuracy.

exceeds, with a given probability, the absolute value of the difference between two single measurement results obtained under the above-mentioned conditions. If not stated, it is assumed that the probability level is 95%.

When a sensor output changes with time (for a constant input), it is sometimes said that there are instabilities and that the sensor drifts. In particular, some sensors have zero and scale factor drifts specified. The *zero drift* describes output variations when the input is zero. *Scale factor drift* describes sensitivity changes.

The *sensitivity* or scale factor is the slope of the calibration curve, whether it is constant or not along the measurement range. For a sensor in which output y is related to the input x by the equation $y = f(x)$, the sensitivity $S(x_a)$, at point x_a, is

$$S(x_a) = \left.\frac{dy}{dx}\right|_{x=x_a} \tag{1.2}$$

It is desirable in sensors to have a high and, if possible, constant sensitivity. For a sensor with response $y = kx + b$ the sensitivity is $S = k$ for the entire range of values for x where it applies. For a sensor with response $y = k^2x + b$ the sensitivity is $S = 2kx$, and it changes from one point to another over the measurement range.

1.4.2 Other Characteristics: Linearity and Resolution

Accuracy, precision, and sensitivity are the characteristics that sufficiently describe the static behavior of a sensor. But sometimes others are added or substituted when it is necessary to describe alternative behavior or behavior that is of particular interest for a given case; likewise, characteristics can be added that are complementary to describe the suitability of a measurement system for a specific application.

The *linearity* describes the closeness between the calibration curve and a specified straight line. Depending on which straight line is considered, several definitions apply.

Independent Linearity. The straight line is defined by the least squares criterion. With this system the maximal positive error and the minimal negative error are equal. This is the method that usually gives the "best" quality.

Zero-Based Linearity. The straight line is also defined by the least squares criterion but with the additional restriction of passing through zero.

Terminal-Based Linearity. The straight line is defined by the output corresponding to the lower input and the theoretical output when the higher input is applied.

End-Points Linearity. The straight line is defined by the real output when the input is the minimum of the measurement range and the output when the input is the maximum (FSO).

Theoretical Linearity. The straight line is defined by the theoretical predictions when designing the sensor.

Figure 1.6 shows these different straight lines for a sensor with a given calibration curve. In sum, the linearity of the calibration curve indicates to what

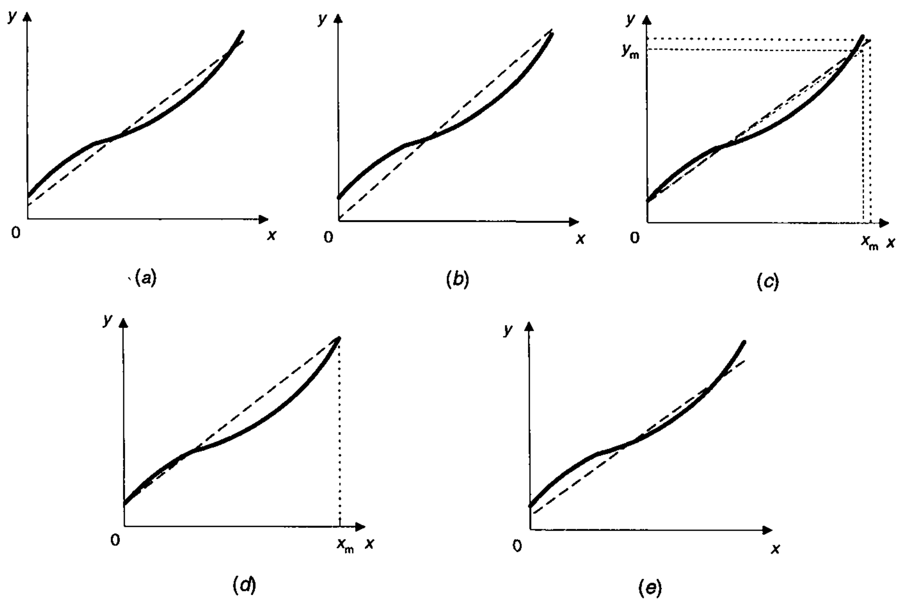

Figure 1.6 Different straight lines used as a reference to define linearity: (*a*) independent linearity (least squares method); (*b*) zero-based linearity (least squares adjusted to zero); (*c*) terminal-based linearity; (*d*) end-points-defined linearity; (*e*) theoretical linearity.

extent a sensor's sensitivity is constant. Nevertheless, for a sensor to be acceptable, it does not need to have a high linearity. The interest of linearity is that when sensitivity is constant we only need to divide the reading by a constant value (the sensitivity) in order to determine the input. In linear instruments the nonlinearity equals the inaccuracy.

Current measurement systems incorporate microprocessors so that there is more interest in repeatability than in linearity, because we can produce a look-up table giving input values corresponding to measured values. By using interpolation, it is possible to reduce the size of that table to a reasonable dimension.

The main factors that influence linearity are resolution, threshold, and hysteresis. The *resolution* (or discrimination) is the minimal change of the input necessary to produce a detectable change at the output. When the input increment is from zero, then it is called the *threshold.* When the input signal can display fast changes, the *noise floor* of the sensor determines the resolution. Noise is a random fluctuation of the sensor output unrelated to the measured quantity.

The *hysteresis* refers to the difference between two output values that correspond to the same input, depending on the direction (increasing or decreasing) of successive input values. That is, similarly to the magnetization in ferromagnetic materials (Section 1.8.2), it can happen that the output corresponding to a given input depends on whether the previous input was higher or lower than the present one.

1.4.3 Systematic Errors

The static calibration of a sensor allows us to detect and correct the so-called systematic errors. An error is said to be *systematic* when in the course of measuring the same value of a given quantity under the same conditions, it remains constant in absolute value and sign or varies according to a definite law when measurement conditions change. Because time is also a measurement condition, the measurements must be made in a short time interval. Systematic errors yield measurement bias.

Such errors are caused not only by the instrument, but also by the method, the user (in some cases), and a series of factors (climatic, mechanical, electrical, etc.) that never are ideal—that is, constant and known.

The presence of systematic errors can therefore be discovered by measuring the same quantity with two different devices, by using two different methods, by using the readings of two different operators, or by changing measurement conditions in a controlled way and observing their influence on results. To determine the consistency of the different results it is necessary to use statistical methods [4]. In any case, even in high-accuracy measurements, there is always some risk that a systematic error may remain undetected. The goal therefore is to have a very low risk for large errors to remain undetected.

Errors in indirect measurements propagate from each measured quantity to the estimated quantity, so that indirect measurements are usually less accurate than direct measurements (see Problem 1.3).

Example 1.1 In order to measure the drop in voltage across a resistor, we consider two alternative methods: (1) Use a voltmeter, whose accuracy is about 0.1 % of the reading. (2) Use an ammeter, whose accuracy is also about 0.1 % of the reading and apply Ohm's law. If the resistor has 0.1 % tolerance, which method is more accurate?

We first differentiate Ohm's law to obtain

$$dV = RdI + IdR$$

Dividing each term by V yields

$$\frac{dV}{V} = \frac{RdI + IdR}{V} = \frac{RdI + IdR}{IR} = \frac{dI}{I} + \frac{dR}{R}$$

For small variations, we can approximate differentials by increments to obtain

$$\frac{\Delta V}{V} = \frac{\Delta I}{I} + \frac{\Delta R}{R}$$

The relative uncertainty for the current and resistance add together. Therefore, the uncertainty in the voltage when measuring current is

$$\frac{\Delta V}{V} = \frac{0.1}{100} + \frac{0.1}{100} = 0.2\,\%$$

The uncertainty when measuring voltage directly is 0.1 %, hence lower.

1.4.4 Random Errors

Random errors are those that remain after eliminating the causes of systematic errors. They appear when the same value of the same quantity is measured repeatedly, using the same instrument and the same method. They have the following properties:

1. Positive and negative random errors with the same absolute value have the same occurrence probability.
2. Random errors are less probable as the absolute value increases.
3. When the number of measurements increases, the arithmetic mean of random errors in a sample (set of measurements) approaches zero.
4. For a given measurement method, random errors do not exceed a fixed value. Readings exceeding that value should be repeated and, if necessary, studied separately.

Random errors are also called *accidental* (or fortuitous) errors, thus meaning that they may be unavoidable. The absence of changes from one reading to another when measuring the same value of the same quantity several times does not necessarily imply an absence of random errors. It may happen, for example, that the instrument does not have a high enough resolution—that is, that its ability to detect small changes in the measured quantity is rather limited and therefore the user does not perceive them.

The presence of random errors implies that the result of measuring n times a measurand x is a set of values $\{x_1, x_2, \ldots, x_n\}$. If there is no systematic error, the best estimate of the actual value of the measurand is the average of the results:

$$\hat{x}_n = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{1.3}$$

Were n infinite, (1.3) would yield a conventional true value for x. When n is finite, however, each set of n measurements yields different x_i and a different average. These averages follow a Gaussian distribution whose variance is σ^2/n, where σ^2 is the variance of x. Then,

$$\text{Prob}\left[-k \le \frac{\hat{x}_n - x}{\sigma/\sqrt{n}} \le +k\right] = 1 - \alpha \tag{1.4}$$

where k and α can be obtained from tables for the unit normal (Gaussian) distribution. From (1.4), we obtain

$$\text{Prob}\left[\hat{x}_n - k\frac{\sigma}{\sqrt{n}} \le x \le \hat{x}_n + k\frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha \tag{1.5}$$

which yields the (*confidence*) *interval* with a probability $1 - \alpha$ of including the true value x. The term $\pm k\sigma/\sqrt{n}$ is also called the *uncertainty* (see Problems 1.4 and 1.5).

Example 1.2 Determine the confidence interval that has 50% probability of including the true value of a quantity when the average from n measurements is $\hat{x}_n$ and the variance is σ^2.

For $k = 0.67$ the tail area of unit normal distribution is 0.2514, and for $k = 0.68$ the tail area is 0.2483. We need the value for a tail area of $(1 - 0.5)/2 = 0.25$ because we are looking for a two-sided interval. By interpolating, we obtain

$$k = 0.67 + \frac{0.68 - 0.67}{0.2483 - 0.2514}(0.25 - 0.2514) = 0.67 + 0.0045 = 0.6745$$

Hence, the interval $\hat{x}_n \pm 0.6745\sigma/\sqrt{n}$ has 50% probability of including the true value. $0.6745\sigma/\sqrt{n}$ is sometimes termed *probable error*; but it is not an "error," nor is it "probable."

Often, however, the variance of the population of (infinite) possible results for the measurand is unknown. If we estimate that variance from a sample of n results by

$$s_n^2 = \frac{\sum_{i=1}^{n}(x_i^2 - \hat{x}_n)^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} \tag{1.6}$$

it is not possible to directly substitute s_n for σ in (1.4). Nevertheless, when the distribution of possible results is Gaussian and $n > 31$, (1.4) holds true even if s_n replaces σ. For $n < 31$, $(\hat{x}_n - x)/(s_n/\sqrt{n})$ follows a Student t distribution instead. Therefore,

$$\text{Prob}\left[-t_{1-\alpha/2}(n-1) \le \frac{\hat{x}_n - x}{s_n/\sqrt{n}} \le +t_{1-\alpha/2}(n-1)\right] = 1 - \alpha \tag{1.7a}$$

where $t_{1-\alpha/2}(n-1)$ is the probability point of the t distribution with $n-1$ degrees of freedom, corresponding to a tail area probability α. The confidence interval follows from

$$\text{Prob}\left[\hat{x}_n - t_{1-\alpha/2}(n-1)\frac{s_n}{\sqrt{n}} \le x \le \hat{x}_n + t_{1-\alpha/2}(n-1)\frac{s_n}{\sqrt{n}}\right] = 1 - \alpha \tag{1.7b}$$

Example 1.3 Determine the confidence interval that has a 99% probability of including the true value of a quantity when the average from 10 measurements is $\hat{x}_n$ and the sample variance is $s_n{}^2$. Compare the result with that when the population variance σ^2 is known.

For $10 - 1 = 9$ degrees of freedom, the t value for a $(1 - 0.99)/2 = 0.005$ tail area probability is $t_{0.995}(9) = 3.250$. The corresponding confidence interval is $\hat{x}_n \pm 3.25 s_n/\sqrt{10} = \hat{x}_n \pm 1.028 s_n$. Had we known σ, for a tail area of $(1 - 0.99)/2 = 0.005$ the normal distribution yields $k = 2.576$. Hence, the confidence interval would be $\hat{x}_n \pm 2.576\sigma/\sqrt{10} = \hat{x}_n \pm 0.815\sigma$, which is narrower than that when σ is unknown.

If s_n has been calculated from a sample of n results, perhaps from previous experiments, it is still possible to determine the confidence interval for a set of m data points by replacing $s_n/\sqrt{m}$ for $s_n/\sqrt{n}$ in (1.7b).

If there are systematic errors in addition to random errors, when calculating the mean of several readings, random errors cancel and only systematic errors remain. Because systematic errors are reproducible, they can be determined for some specified measurement conditions, and then the reading can be corrected when measuring under the same conditions. This calculation of the difference between the true value and the measured value is performed during the calibration process under some specified conditions. Furthermore, during that process the instrument is usually adjusted to eliminate that error. When making a single measurement, under the same conditions, only the random component of error remains.

In practice, however, during the calibration process, only systematic errors for some very specific conditions can be eliminated. Therefore, under different measurement conditions some systematic errors even greater than the random ones may be present. Product data sheets state these errors, usually through the range b_x having a given probability $1 - \alpha$ of enclosing the true value. The overall uncertainty can be then calculated by [5]

$$u_x = \pm t_{1-\alpha/2}\sqrt{\left(\frac{b_x}{2}\right)^2 + \left(\frac{s_n}{\sqrt{n}}\right)^2} \tag{1.8}$$

The usual confidence level in engineering is 95%, so that for $n > 31$, $t_{97.5} = 1.96$.

1.5 DYNAMIC CHARACTERISTICS

The sensor response to variable input signals differs from that exhibited when input signals are constant, which is described by static characteristics. The reason is the presence of energy-storing elements, such as inertial elements (mass, inductance, etc.) and capacitance (electric, thermal, fluid, etc.). The dynamic characteristics are the dynamic error and speed of response (time constant, delay). They describe the behavior of a sensor with applied variable input signals.

The *dynamic error* is the difference between the indicated value and the true value for the measured quantity, when the static error is zero. It describes the difference between a sensor's response to the same input magnitude, depending on whether the input is constant or variable with time.

The *speed of response* indicates how fast the measurement system reacts to changes in the input variable. A delay between the applied input and the corresponding output is irrelevant from the measurement point of view. But if the sensor is part of a control system, that delay may result in oscillations.

To determine the dynamic characteristics of a sensor, we must apply a variable quantity to its input. This input can take many different forms, but it is usual to study the response to transient inputs (impulse, step, ramp), periodic inputs (sinusoidal), or random inputs (white noise). In linear systems, where superposition holds, any one of these responses is enough to fully characterize the system. The selection of one input or another depends on the kind of sensor. For example, it is difficult to produce a temperature with sinusoidal variations, but it is easy to cause a sudden temperature change such as a step. On the other hand, it is easier to cause an impulse than to cause a step of acceleration.

To mathematically describe the behavior of a sensor, we assume that its input and output are related through a constant-coefficient linear differential equation, and that therefore we are dealing with a linear time-invariant system. Then the relation between sensor output and input can be expressed in a simple form, as a quotient, by taking the Laplace transform of each signal and the transfer function of the sensor [2]. Recall that the transfer function gives a general relation between output and input, but not between their instantaneous values. Sensor dynamic characteristics can then be studied for each applied input by classifying the sensors according to the order of their transfer function. It is generally not necessary to use models higher than second-order functions.

1.5.1 Zero-Order Measurement Systems

The output of a zero-order sensor is related to its input through an equation of the type

$$y(t) = k \cdot x(t) \tag{1.9}$$

Its behavior is characterized by its static sensitivity k and remains constant regardless of input frequency. Hence, its dynamic error and its delay are both zero.

An input–output relationship such as that in (1.9) requires that the sensor does not include any energy-storing element. This is, for example, the case of potentiometers applied to the measurement of linear and rotary displacements (Section 2.1). Using the notation of Figure 1.7, we have

$$y = V_r \frac{x}{x_m} \tag{1.10}$$

where $0 \leq x \leq x_m$ and V_r is a reference voltage. In this case $k = V_r/x_m$.

Models such as the previous one are always a mathematical abstraction because we cannot avoid the presence of imperfections that restrict the applicability of the model. For example, for the potentiometer, it is not possible to apply it to fast-varying movements because of the friction of the wiper.

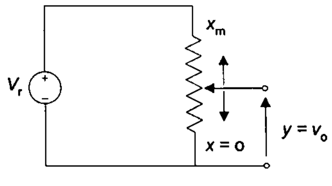

Figure 1.7 A linear potentiometer used as a position sensor is a zero-order sensor.

TABLE 1.3 Output Signal of a First-Order Measurement System for Different Common Test Inputs

Input	Output
Step $u(t)$	$k(1 - e^{-t/\tau})$
Ramp Rt	$Rkt - Rk\tau u(t) + Rk\tau e^{-t/\tau}$
Sinusoid A, ω	$\frac{kA\tau\omega e^{-t/\tau}}{1+\omega^2\tau^2} + \frac{kA}{\sqrt{1+\omega^2\tau^2}} \sin(\omega t + \phi)$ $\phi = \arctan(-\omega t)$

1.5.2 First-Order Measurement Systems

In a first-order sensor there is an element that stores energy and another one that dissipates it. The relationship between the input $x(t)$ and the output $y(t)$ is described by a differential equation with the form

$$a_1 \frac{dy(t)}{dt} + a_0 y(t) = x(t) \tag{1.11}$$

The corresponding transfer function is

$$\frac{Y(s)}{X(s)} = \frac{k}{\tau s + 1} \tag{1.12}$$

where $k = 1/a_0$ is the *static sensitivity* and $\tau = a_1/a_0$ is the system's *time constant*. The system's *corner (angular) frequency* is $\omega_c = 1/\tau$. Therefore, to characterize the system two parameters are necessary: k for the static response and ω_c or τ for the dynamic response.

Table 1.3 shows the expression of the output signal for each of the most common test inputs: step, ramp, and sinusoid. The derivation of the complete expressions can be found in most books on control theory or in reference 2. For the sinusoid the transient part of the output has been included. This is important when the reading is taken shortly after applying the input.

The dynamic error and delay of a first-order sensor depend on the input waveform. Table 1.4 shows the dynamic error and delay corresponding to the inputs considered in Table 1.3. The two values for the dynamic error for an input ramp correspond, respectively, to two different definitions:

$$e_d = y(t) - x(t) \tag{1.13}$$

$$e_d = y(t) - kx(t) \tag{1.14}$$

For step and sinusoidal inputs, only (1.14) has been used.

The availability of analytical expressions for the dynamic error may suggest

TABLE 1.4 Dynamic Error and Delay for a First-Order Measurement System for Different Common Test Inputs

Input	Dynamic Error	Delay
Step $u(t)$	0	τ
Ramp Rt	$R[t + k(\tau - t)]$ or $R\tau$	τ
Sinusoid A, ω	$1 - \dfrac{1}{\sqrt{1 + \omega^2\tau^2}}$	$\dfrac{\arctan \omega\tau}{\omega}$

that it can easily be corrected. In practice, however, the real input will seldom be as simple as the ones considered, and therefore it will not be possible to compensate for the dynamic error. Figure 1.8 shows the response to each of these input waveforms (see Problem 1.6).

An example of a first-order sensor is a thermometer based on a mass M with specific heat c (J/kg·K), heat transmission area A, and (convection) heat trans-

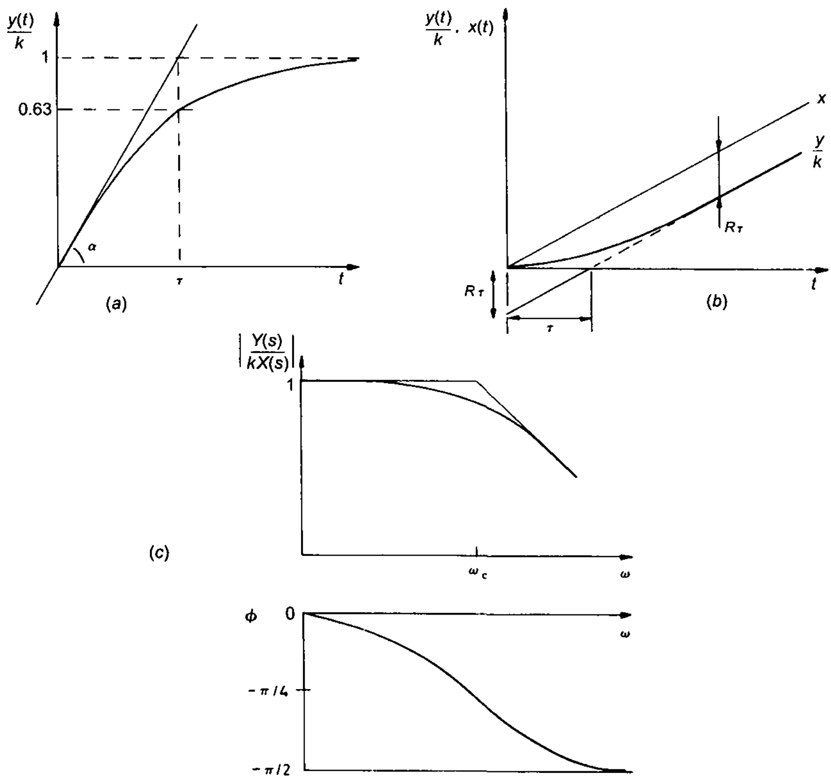

Figure 1.8 First-order system response to (a) a unit step input, (b) a ramp input, and (c) a sinusoidal input (amplitude modulus and phase).

fer coefficient h ($W/m^2{\cdot}K$). In steady state, energy balance yields

$$(\text{Heat in}) - (\text{Heat out}) = \text{Energy stored}$$

If we assume that the sensor does not lose any heat—for example, through its leads—and that its mass does not change (negligible expansion), if we call T_i its internal temperature when the external temperature is T_e, we have

$$hA(T_e - T_i)\,dt - 0 = Mc\,dT_i \tag{1.15}$$

$$\frac{dT_i}{dt} = \frac{hA}{Mc}(T_e - T_i) \tag{1.16}$$

By taking the Laplace transform and introducing $\tau = hA/Mc$, we obtain

$$\frac{T_i(s)}{T_e(s)} = \frac{1}{1 + \tau s} \tag{1.17}$$

Therefore, the resistance to heat transfer, along with the mass and thermal capacity, will determine the time constant and delay the sensor's temperature change. Nevertheless, once the sensor reaches a given temperature, its response is immediate. There is not any noticeable delay in sensing. The delay is in the sensor achieving the final temperature.

Example 1.4 The approximate time constant of a thermometer is determined by immersing it in a bath and noting the time it takes to reach 63% of the final reading. If the result is 28 s, determine the delay when measuring the temperature of a bath that is periodically changing 2 times per minute.

From the step response we have $\tau = 28$ s. From the last row in Table 1.4, the delay when measuring a cyclic variation will be

$$t_d = \frac{\arctan(\omega\tau)}{\omega}$$

The angular (radian) frequency of the temperature to measure is

$$\omega = 2\pi\frac{2 \text{ cycles}}{60 \text{ s}} = 0.209 \text{ rad/s}$$

The delay will be

$$t_d = \frac{\arctan\left(\dfrac{0.209 \text{ rad}}{1 \text{ s}} \times 28 \text{ s}\right)}{0.209 \text{ rad/s}} = 6.7 \text{ s}$$

1.5.3 Second-Order Measurement Systems

A second-order sensor contains two energy-storing elements and one energy-dissipating element. Its input $x(t)$ and output $y(t)$ are related by a second-order linear differential equation of the form

$$a_2 \frac{d^2 y(t)}{dt^2} + a_1 \frac{dy(t)}{dt} + a_0 y(t) = x(t) \tag{1.18}$$

The corresponding transfer function is

$$\frac{Y(s)}{X(s)} = \frac{k\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{1.19}$$

where k is the *static sensitivity,* ζ is the *damping ratio,* and ω_n is the *natural undamped angular frequency* for the sensor ($\omega_n = 2\pi f_n$). Two coefficients determine the dynamic behavior, while a single one determines the static behavior. Their expressions for the general second-order system modeled by (1.18) are

$$k = \frac{1}{a_0} \tag{1.20}$$

$$\omega_n^2 = \frac{a_0}{a_2} \tag{1.21}$$

$$\zeta = \frac{a_1}{2\sqrt{a_0 a_2}} \tag{1.22}$$

Notice that these three parameters are related and that a modification in one of them may change another one. Only a_0, a_1, and a_2 are independent.

Doebelin [2] details the procedure to obtain the output as a function of simple test input waveforms. Table 1.5 shows some results. Figure 1.9 shows their graphical characteristics. Note that the system behavior differs for $0 < \zeta < 1$ (underdamped case), $\zeta = 1$ (critically damped case), or $\zeta > 1$ (overdamped case). The initial transient has been omitted for the sinusoidal input.

Example 1.5 In a measurement system, a first-order sensor is replaced by a second-order sensor with the same natural (corner) frequency. Calculate the damping ratio to achieve the same -3 dB attenuation at that frequency.

A -3 dB attenuation means

$$3 = 20 \lg a$$

$$a = 10^{-3/20} = 0.707$$

From the last row in Table 1.5, the relative magnitude for a second-order response is

TABLE 1.5 Outputs of a Second-Order Measuring System for Different Common Test Inputs

Input	Output	
Unit step $u(t)$		
$0<\zeta<1$	$1-\dfrac{e^{-\delta t}}{\sqrt{1-\zeta^2}}\sin(\omega_d t+\phi)$	$\delta=\zeta\omega_n$ $\omega_d=\omega_n\sqrt{1-\zeta^2}$ $\phi=\arcsin\dfrac{\omega_d}{\omega_n}$
$\zeta=1$	$1-e^{-\delta t}(1+\omega_n t)$	
$\zeta>1$	$1+\dfrac{\omega_n}{2\sqrt{\zeta^2-1}}\left(\dfrac{e^{-at}}{a}-\dfrac{e^{-bt}}{b}\right)$	$a=\omega_n(\zeta+\sqrt{\zeta^2-1})$ $b=\omega_n(\zeta-\sqrt{\zeta^2-1})$
Ramp Rt		
$0<\zeta<1$	$R\left\{t-\dfrac{2\zeta}{\omega_n}\left[1-\dfrac{e^{-\zeta\omega_n t}}{2\zeta\sqrt{1-\zeta^2}}\sin(\sqrt{1-\zeta^2}\omega_n t+\phi)\right]\right\}$	$\phi=\arctan\left(\dfrac{2\zeta\sqrt{1-\zeta^2}}{2\zeta^2-1}\right)$
$\zeta=1$	$R\left\{t-\dfrac{2\zeta}{\omega_n}\left[1-\left(1+\dfrac{\omega_n\tau}{2}\right)e^{-\omega_n t}\right]\right\}$	
$\zeta>1$	$R\left\{t-\dfrac{2\zeta}{\omega_n}\left[1+\dfrac{2\zeta(-\zeta-\sqrt{\zeta^2-1}+1)}{4\zeta\sqrt{\zeta^2-1}}e^{-at}+\dfrac{2\zeta(-\zeta-\sqrt{\zeta^2-1}-1)}{4\zeta\sqrt{\zeta^2-1}}e^{-bt}\right]\right\}$	
Sinusoid A,ω	$\dfrac{kA}{\sqrt{\left(1-\dfrac{\omega^2}{\omega_n^2}\right)^2+\left(\dfrac{2\zeta\omega}{\omega_n}\right)^2}}\sin(\omega t-\phi)$	$\phi=\arctan\dfrac{2\zeta\omega/\omega_n}{1-\left(\dfrac{\omega}{\omega_n}\right)^2}$

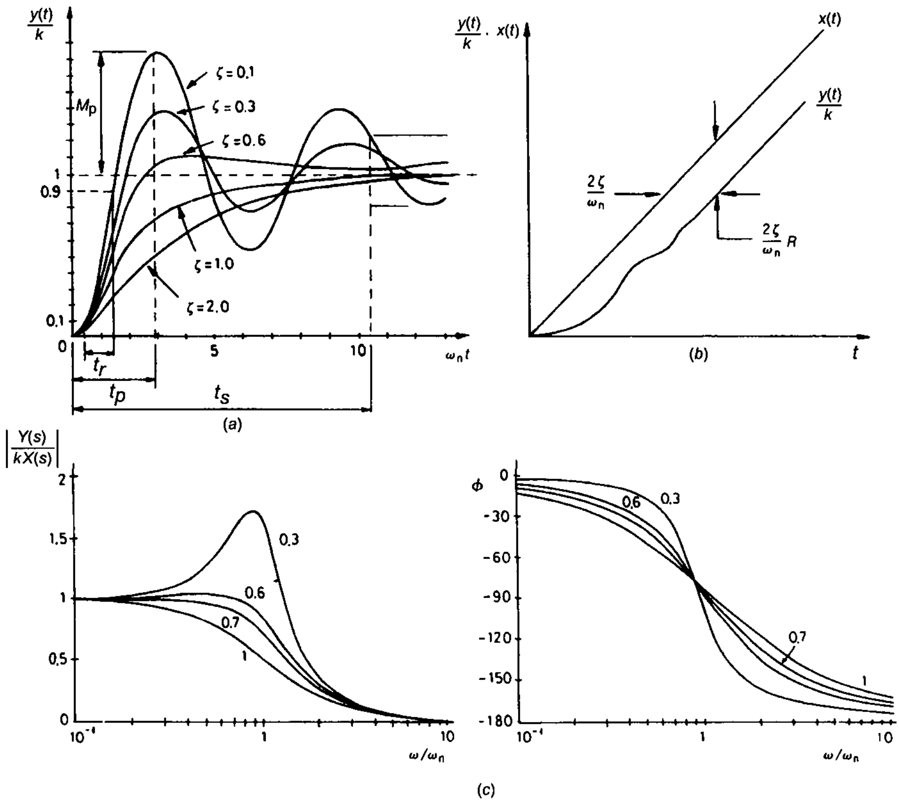

Figure 1.9 Second-order system response to (*a*) a unit step input, (*b*) a ramp input, and (*c*) a sinusoidal input (amplitude modulus and phase), for different damping ratios.

$$\frac{1}{\sqrt{\left(1-\dfrac{\omega^2}{\omega_n^2}\right)^2+\left(\dfrac{2\zeta\omega}{\omega_n}\right)^2}}$$

which reduces to $0.5/\zeta$ at ω_n. The condition to fulfill is $0.5/\zeta = 0.707$. This yields $\zeta = 0.707$.

The dynamic error and delay in a second-order system depend not only on the input waveform but also on ω_n and ζ. Their expressions are much more involved than in first-order systems, and to analyze them several factors related to ω_n and ζ are defined.

When the input is a unit step, if the system is overdamped ($\zeta > 1$) or is critically damped ($\zeta = 1$), there is neither overshoot nor steady-state dynamic error in the response.

In an underdamped system, $\zeta < 1$, the steady-state dynamic error is zero, but the speed and the overshoot in the transient response are related (Figure 1.9*a*). In general, the faster the speed, the larger the overshoot. The *rise time* t_r is the time spent to rise from 10% to 90% of the final output value, and it is given by

$$t_r = \frac{\arctan(-\omega_d/\delta)}{\omega_d} \tag{1.23}$$

where $\delta = \zeta\omega_n$ is the so-called *attenuation*, and $\omega_d = \omega_n\sqrt{1-\zeta^2}$ is the *natural damped angular frequency*.

The time elapsed to the first peak t_p is

$$t_p = \frac{\pi}{\omega_d} \tag{1.24}$$

and the maximum overshoot M_p is

$$M_p = e^{-(\delta/\omega_d)\pi} \tag{1.25}$$

The time for the output to settle within a defined band around the final value t_s, or *settling time*, depends on the width of that band. For $0 < \zeta < 0.9$, for a $\pm 2\%$ band, $t_s \cong 4/\delta$, and it is minimal when $\zeta = 0.76$; for a $\pm 5\%$ band, $t_s \cong 3/\delta$, and it is minimal when $\zeta = 0.68$. The speed of response is optimal for $0.5 < \zeta < 0.8$ [6].

Figure 1.9*a* may suggest that underdamped sensors are useless because of their large overshoot. But in practice the input will never be a perfect step, and the sensor behavior may be acceptable. That is the case with piezoelectric sensors (Section 6.2), for example. Nevertheless, a large overshoot can saturate an amplifier output (see Problem 1.8).

When the input is a ramp with slope R, the steady-state dynamic error is

$$e_d = \frac{2\zeta R}{\omega_n} \tag{1.26}$$

and the delay is $2\zeta/\omega_n$ (Figure 1.9*b*).

To describe the frequency response of a second-order system where $0 < \zeta < 0.707$, we note that the frequency of resonance is the same as the natural damped frequency

$$\omega_d = \omega_n\sqrt{1-2\zeta^2} \tag{1.27}$$

and the amplitude of that resonance at $\omega = \omega_d$ is M_r:

$$M_r = \frac{1}{2\zeta\sqrt{1-\zeta^2}} \tag{1.28}$$

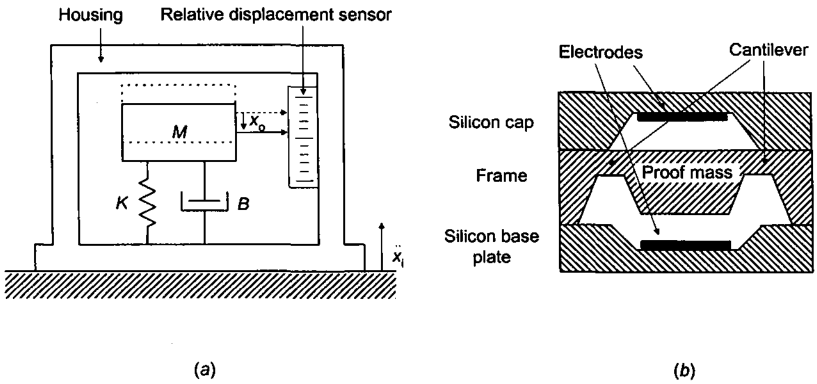

Figure 1.10 Second-order, underdamped sensors based on a mass–spring system. (*a*) The acceleration applied to the housing displaces the proof mass because of the force transmitted through its mechanical links. The spring constant K and viscous friction B do not necessarily belong to separate physical elements. (*b*) In this micromachined capacitive silicon accelerometer the applied acceleration flexes the cantilevers because of the force exerted by the proof mass, changing the capacitance between that mass surface and the fixed electrodes.

A simple example of a second-order sensor described by (1.19) is a thermometer covered for protection. In this case, we must add the heat capacity and thermal resistance of the covering to the heat capacity of the sensing element and heat conduction resistance from the medium where it is placed. The system has $\zeta > 1$. Liquid in glass manometers also have overdamped response (Problem 1.11).

Examples of underdamped systems are the mass–spring systems used to measure displacement, velocity, and acceleration in vibratory movements or in long-range missiles. They are also the heart of seismographs and micromachined accelerometers for airbag deployment in cars. Using the notation of Figure 1.10*a*, if we measure the displacement x_o of the mass M with respect to the armature fixed to the element undergoing an acceleration $\ddot{x}_i$, then the force on the mass (Newton's second law) is communicated through the spring deflection (Hooke's law) and the internal viscous friction. The force equation of the system is

$$M(\ddot{x}_i - \ddot{x}_o) = Kx_o + B\dot{x}_o \tag{1.29}$$

where K is the spring constant or stiffness and B is the viscous frictional coefficient. K and B represent different physical actions, but they are not necessarily separate elements. The Laplace transform of $\ddot{x}_i$ is $s^2X_i(s)$, from which we obtain

$$Ms^2X_i(s) = X_o(s)[K + Bs + Ms^2] \tag{1.30}$$

The transfer function is

$$\frac{X_o(s)}{\ddot{X}_i(s)} = \frac{X_o(s)}{s^2 X_i(s)} = \frac{M}{K} \frac{K/M}{s^2 + s(B/M) + K/M} \tag{1.31}$$

Therefore, $k = M/K$, $\zeta = B/(2\sqrt{KM})$, and $\omega_n = \sqrt{K/M}$. A large mass increases the sensitivity but reduces the natural frequency and the damping ratio. Stiffness increases the natural frequency but reduces the sensitivity and the damping ratio. Viscosity increases the damping ratio without affecting the sensitivity or the natural frequency. Micromachined accelerometers are very stiff and have small mass and friction. Therefore they have a large natural frequency but small sensitivity and damping ratio.

A potentiometer, a capacitive or inductive sensor, or a photodetector (with an ancillary light source and a shutter) can measure the displacement x_o of the proof mass. Alternatively, we can sense the stress of a flexing element holding the mass—for example, by using strain gages or a piezoelectric element. Figure 1.10*b* shows a capacitive micromachined silicon accelerometer based on a mass–spring system.

To consider also the acceleration of gravity when the axis of the accelerometer forms an angle θ with respect to the horizontal, the term $Mg \sin\theta$ has to be included in the right-hand member of (1.29). Then the output $y(t)$ would be defined as $x_o + (Mg \sin\theta)/K$, and its Laplace transform would be given by (1.31) with $Y(s)$ replacing $X_o(s)$.

If instead of the input acceleration we want to sense the displacement, we would multiply both sides of (1.31) by s^2 to obtain

$$\frac{X_o(s)}{X_i(s)} = \frac{M}{K} \frac{(K/M)s^2}{s^2 + s(B/M) + K/M} \tag{1.32}$$

From (1.31), the response for acceleration measurements is low-pass and ω_n must be higher than the maximal frequency variation of the acceleration to be measured. But for the measurement of vibration displacement—high-pass response (1.32)—ω_n must be lower than the frequency of the displacement and there is no dc response (see Problem 1.13).

1.6 OTHER SENSOR CHARACTERISTICS

Static and dynamic characteristics do not completely describe the behavior of a sensor. Table 1.6 lists other characteristics to consider in sensor selection, relative to the sensor and to the quantity to sense. In addition to those sensor characteristics, the measurement method must always be appropriate for the application. For example, there will be an error if, in measuring a flow, the insertion of the flowmeter significantly obstructs the conduit section.

TABLE 1.6 Characteristics to Consider in Sensor Selection

Quantity to Measure[a]	Output Characteristics	Supply Characteristics	Environmental Characteristics	Other Characteristics
Span	Sensitivity	Voltage	Ambient temperature	Reliability
Target accuracy	Noise floor	Current	Thermal shock	Operating life
Resolution	Signal: voltage, current, frequency	Available power	Temperature cycling	Overload protection
Stability	Signal type: single ended, differential, floating	Frequency (ac supply)	Humidity	Acquisition cost
Bandwidth	Impedance	Stability	Vibration	Weight, size
Response time	Code, if digital		Shock	Availability
Output impedance			Chemical agents	Cabling requirements
Extreme values			Explosion risks	Connector type
Interfering quantities			Dirt, dust	Mounting requirements
Modifying quantities			Immersion	Installation time
			Electromagnetic environment	State when failing
			Electrostatic discharges	Calibration and testing cost
			Ionizing radiation	Maintenance cost
				Replacement cost

[a] Sensor static and dynamic characteristics must be compatible with the requirements of the quantity to measure.

1.6.1 Input Characteristics: Impedance

The output impedance of the quantity to sense determines the input impedance needed for the sensor. Two examples illustrate this connection. To prevent the wiper in a potentiometer (Section 2.1) from losing contact with the resistive element, it is necessary for the wiper to exert a force on it. What would it happen if we desired to measure the movement of an element unable to overcome the friction between the wiper and the resistive element? This effect is not modeled by (1.9).

When we use a thermometer having a considerable mass to measure the temperature reached by a transistor, upon contact, wouldn't the thermometer cool the transistor and give a lower reading than the initial transistor temperature? Equation (1.17) would not describe that effect.

Neither the static nor the dynamic characteristics of sensors that we have defined allow us to describe the real behavior of the combined sensor-measured system. The description of a sensor or a measurement system through block diagrams ignores the fact that the sensor extracts some power from the measured system. When this power extraction modifies the value of the measured variable, we say that there is a *loading error*. Block diagrams are only appropriate when there is no energy interaction between blocks. The concept of input impedance allows us to determine when there will be a loading error.

When measuring a quantity x_1 there is always another quantity x_2 involved, such that the product x_1x_2 has the dimensions of power. For example, when measuring a force there is always a velocity; when measuring flow there is a drop in pressure; when measuring temperature there is a heat flow; when measuring an electric current there is a drop in voltage, and so on.

Nonmechanical variables are designed as *effort variables* if they are measured between two points or regions in the space (voltage, pressure, temperature), and they are designed as *flow variables* if they are measured at a point or region in the space (electric current, volume flow, heat flow). For mechanical variables the converse definitions are used, with effort variables measured at a point (force, torque) and flow variables measured between two points (linear velocity, angular velocity).

For an element that can be described through linear relations, the input impedance, $Z(s)$, is defined as the quotient between the Laplace transforms of an input effort variable and the associated flow variable [7]. The input admittance, $Y(s)$, is defined as the reciprocal of $Z(s)$. $Z(s)$ and $Y(s)$ usually change with frequency. When very low frequencies are considered, *stiffness* and *compliance* are used instead of impedance and admittance.

To have a minimal loading error, it is necessary for the input impedance to be very high when measuring an effort variable. If x_1 is an effort variable, then we obtain

$$Z(s) = \frac{X_1(s)}{X_2(s)} \tag{1.33}$$

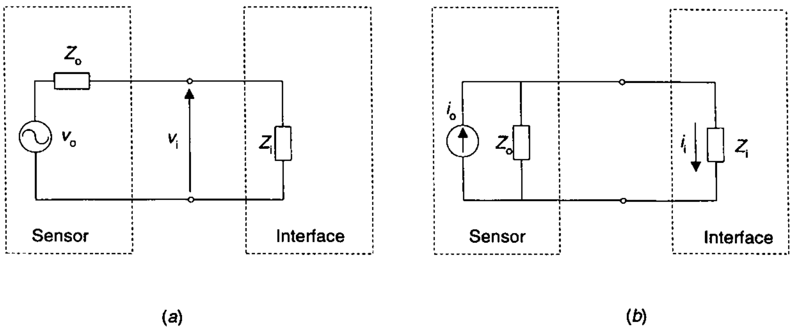

Figure 1.11 Interface circuits must have (*a*) high input impedance for sensors with voltage output and (*b*) low input impedance for sensors with current output.

The power drained from the measured system will be $P = x_1 x_2$; and if it is to be kept at minimum, x_2 must be as small as possible. Therefore the input impedance must be high.

To keep P very small when measuring a flow variable, it is necessary for x_1 to be very small, and that calls for a low input impedance (i.e., a high input admittance).

To obtain high-valued input impedances, it may be necessary to modify the value of components or to redesign the system and use active elements. For active elements, most of the power comes from an auxiliary power supply, and not from the measured system. Another option is to measure by using a balancing method because there is only a significant power drain when the input variable changes its value.

Sensor output impedance determines the input impedance needed for the interface circuit. A voltage output (Figure 1.11*a*) demands high input impedance in order for the sensed voltage

$$V_i = V_o \frac{Z_i}{Z_i + Z_o} \tag{1.34}$$

to be close to the sensor output voltage. Conversely, a current output (Figure 1.11*b*) demands low input impedance in order for the input current

$$I_i = I_o \frac{Z_o}{Z_i + Z_o} \tag{1.35}$$

to be close to the sensor output current.

1.6.2 Reliability

A sensor is reliable when it works without failure under specified conditions for a stated period. Reliability is described statistically: A high reliability means a probability close to 1 of performing as desired (i.e., units of that sensor seldom fail during the period considered). The *failure rate* λ is the number of failures of an item per unit measure of life (time, cycles), normalized to the number of surviving units. If in a time interval dt, $N_f(t)$ units from a batch of N fail and $N_s(t)$ survive, and life is measured in time units, the failure rate is

$$\lambda(t) = \frac{1}{N_s(t)}\frac{dN_f}{dt} \tag{1.36}$$

The reliability at any time t as a probability is

$$R(t) = \lim_{N\to\infty} \frac{N_s(t)}{N} \tag{1.37}$$

N will always be finite in practice. Hence, $R(t)$ can only be estimated. Since at any interval between $t = 0$ and any time later t, units either survive or fail,

$$N = N_s(t) + N_f(t) \tag{1.38}$$

Substituting into (1.37), differentiating, and applying (1.36) yields

$$\frac{dR(t)}{dt} = -\frac{1}{N}\frac{dN_f(t)}{dt} = -\frac{\lambda(t)N_s(t)}{N} = -\lambda(t)R(t) \tag{1.39}$$

Solving for $R(t)$, we obtain

$$R(t) = e^{-\int \lambda(t)\,dt} \tag{1.40}$$

Therefore, the reliability can be calculated from the failure rate, which is calculated from experiments that determine its reciprocal, the *mean time between failures* (MTBF):

$$\text{MTBF} = m = \frac{1}{\lambda} \tag{1.41}$$

Example 1.6 We test 50 units of a given accelerometer for 1000 h. If the failure rate is assumed constant and 2 units fail, determine the failure rate and MTBF.

From (1.36),

$$\lambda = \frac{1}{50}\frac{2}{1000\ \text{h}} = 40\frac{\text{failures}}{\text{milion hours}}$$

From (1.41),

$$\text{MTBF} = \frac{10^6\ \text{h}}{40} = 25000\ \text{h}$$

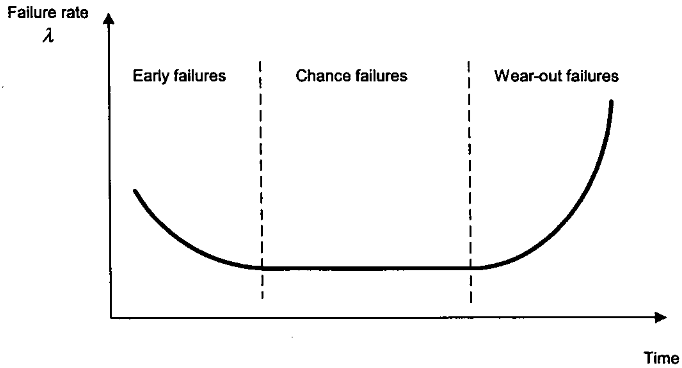

Figure 1.12 The failure rate of many devices follows a bathtub curve that determines three stages in a product life (infant mortality, useful life, and wear out stage) with different failure causes.

Experimental studies of many devices, including sensors, show that their failure rate is not constant but follows the trend in Figure 1.12 after obvious failures have been discarded. Some units of the initial population fail shortly after power up because of *early failures* or *break-in failures*, leading to the so-called infant mortality. Early failure result from microscopic defects in materials and from incorrect adjustments or positioning that went undetected during quality control. Electrical, mechanical, chemical, and thermal stresses during operation sometimes exceed those during product test, and they are withstood by normal units but not by inferior units. Early failures are excluded from MTBF calculations.

The flat segment in Figure 1.12 corresponds to the device useful life. λ is almost constant and is due to *chance failures* (intrinsic or stress-related failures) that result from randomly occurring stresses, the random distribution of material properties and random environmental conditions. Chance failures are present from the beginning, but early failures predominate at that stage.

Some time after placing different units of a device in service, they start to fail at an increasing rate. This is the wear-out stage, wherein parts fail because of the deterioration caused by thermal cycles, wear, fatigue, or any other condition that causes weakening under normal use. In this stage, wear-out failures predominate over chance failures.

Reliability is very important in sensors because they provide information for the control of control systems. Kumamoto and Henzel [8] analyze the reliability of systems that include sensors, alarms and feedback loops. Reference 9 analyzes reliability in depth.

1.7 PRIMARY SENSORS

Primary sensors convert measurands from physical quantities to other forms. We classify primary sensors here according to the measurand. Devices that

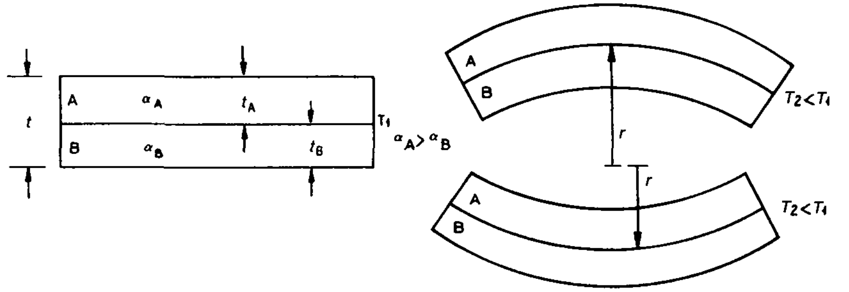

Figure 1.13 A bimetal consists of two metals with dissimilar thermal expansion coefficients, which deforms when temperature changes. Dimensions and curvature have been exaggerated to better illustrate the working principle. (From E. O. Doebelin, *Measurement Systems Application and Design*, 4th ed., copyright 1990. Reprinted by permission of McGraw-Hill, New York.)

have direct electric output are plain sensors and are discussed in Chapters 2, 4, and 6. Radiation-based measurement methods are described in Chapter 9. Khazan [11] and Fraden [12] describe additional primary sensors.

1.7.1 Temperature Sensors: Bimetals

A bimetal consists of two welded metal strips having different thermal expansion coefficients that are exposed to the same temperature. As temperature changes, the strip warps according to a uniform circular arc (Figure 1.13). If the metals have similar moduli of elasticity and thicknesses, the radius of curvature r, when changing from temperature T_1 to T_2, is [2]

$$r \cong \frac{2t}{3(\alpha_A - \alpha_B)(T_2 - T_1)} \tag{1.42}$$

where t is the total thickness of the piece and where α_A and α_B are the respective thermal expansion coefficients. Therefore the radius of curvature is inversely proportional to the temperature difference. A position or displacement sensor would yield a corresponding electric signal. Alternatively, the force exerted by a total or partially bonded or clamped element can be measured.

The thickness of common bimetal strips ranges from 10 μm to 3 mm. A metal having $\alpha_B < 0$ would yield a small r, hence high sensitivity. Because useful metals have $\alpha_B > 0$, bimetal strips combine a high-coefficient metal (proprietary iron–nickel–chrome alloys) with invar (steel and nickel alloy) that shows $\alpha = 1.7 \times 10^{-6}/°C$. Micromachined actuators (microvalves) use silicon and aluminum.

Bimetal strips are used in the range from −75 °C to +540 °C, and mostly from 0 °C to +300 °C. They are manufactured in the form of cantilever, spiral, helix, diaphragm, and so on, normally with a pointer fastened to one end of the strip, which indicates temperature on a dial. Bimetal strips are also used as actuators to directly open or close contacts (thermostats, on–off controls, starters for fluorescent lamps) and for overcurrent protection in electric circuits: The current along the bimetal heats it by Joule effect until reaching a temperature high enough to exert a mechanical force on a trigger device that opens the circuit and interrupts the current.

Other nonmeasurement applications of bimetal strips are the thermal compensation of temperature-sensitive devices and fire alarms. Their response is slow because of their large mass. Each October issue of *Measurements & Control* lists the manufacturers and types of bimetallic thermometers.

1.7.2 Pressure Sensors

Pressure measurement in liquids or gases is common, particularly in process control and in electronic engine control. Blood pressure measurement is very common for patient diagnosis and monitoring. Pressure is defined as the force per unit area. *Differential pressure* is the difference in pressure between two measurement points. *Gage pressure* is measured relative to ambient temperature. *Absolute pressure* is measured relative to a perfect vacuum. To measure a pressure, it is either compared with a known force or its effect on an elastic element is measured (deflection measurement). Table 1.7 shows some sensing

TABLE 1.7 Some Common Methods to Measure Fluid Pressure in Its Normal Range

1. Liquid column + level detection	
2. Elastic element	
2.1. Bourdon tube + displacement measurement:	Potentiometer
	LVDT
	Inductive sensor
	Digital encoder
2.2. Diaphragm + deformation measurement	
2.2.1. Central deformation[a]:	Potentiometer
	LVDT
	Inductive sensor
	Unbonded strain gages
	Cantilever and strain gages
	Vibrating wire
2.2.2. Global deformation:	Variable reluctance
	Capacitive sensor
	Optical sensor
	Piezoelectric sensor
2.2.3. Local deformation: strain gages:	Bonded foil
	Bonded semiconductor
	Deposited
	Sputtered (thin film)
	Diffused/implanted semiconductor

[a] Capsules and bellows yield larger displacements than diaphragms but suit only static pressures.

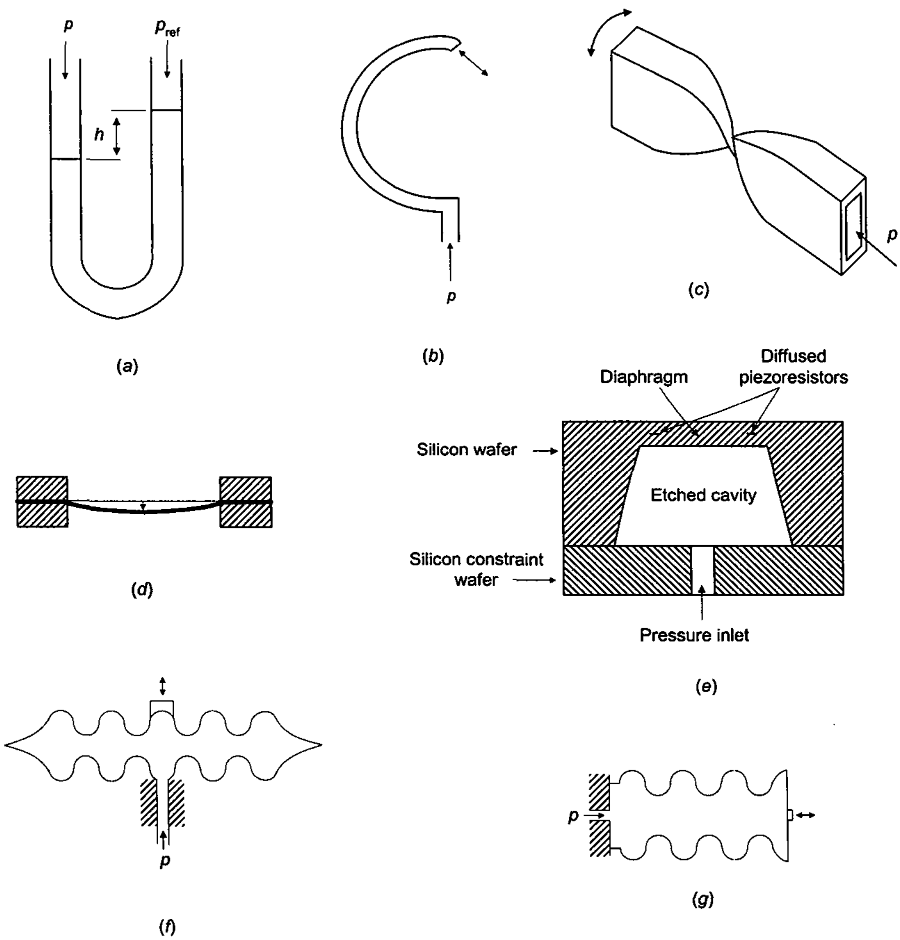

Figure 1.14 Primary pressure sensors. (*a*) Liquid-column U-tube manometer. The liquid must be compatible with the fluid for which pressure is to be measured, and the tube must withstand the mechanical stress. (*b*) C-shaped Bourdon tube. (*c*) Twisted Bourdon tube. (*d*) Membrane diaphragm. (*e*) Micromachined diaphragm. (*f*) Capsule. (*g*) Bellows. The area of the diaphragm in (*e*) is less than 1 mm^2. All other devices can measure up to several centimeters.

methods. Each issue of *Measurements & Control* lists the manufacturers of different pressure sensors: potentiometric (January); strain-gage and piezoresistive (April); capacitive (June); digital and reluctive (September); piezoelectric and liquid-column (October); and bellows, Bourdon tube, and diaphragm (December).

A liquid-column manometer such as the U-tube in Figure 1.14*a* compares the pressure to be measured with a reference pressure and yields a difference h of liquid level. When second order effects are disregarded, the result is

$$h = \frac{p - p_{\text{ref}}}{\rho g} \tag{1.43}$$

where ρ is the density of the liquid and g is the acceleration of gravity. A level sensor (photoelectric, float, etc.) yields an electric output signal.

Elastic elements deform under pressure until the internal stress balances the applied pressure. The material and its geometry determine the amplitude of the resulting displacement or deformation, hence the appropriate sensor (Table 1.7). Usual pressure sensors use the Bourdon tube, diaphragms, capsules, and bellows.

The Bourdon tube—patented by Eugene Bourdon in 1849—is a curved (Figure 1.14*b*) or twisted (Figure 1.14*c*), flattened metallic tube with one closed end. The tube is obtained by deforming a tube having a circular cross section. When pressure is applied through the open end, the tube tends to straighten. The displacement of the free end indicates the pressure applied. This displacement is not linear along its entire range, but is linear enough in short ranges. Displacement sensors yield an electric output signal. Tube configurations with greater displacements (spiral, helical) have large compliance and length that result in a small-frequency passband. The tube metal (brass, monel, steel) is selected to be compatible with the medium.

A diaphragm is a flexible circular plate consisting of a taut membrane or a clamped sheet that strains under the action of the pressure difference to be measured (Figure 1.14*d*). The sensor detects the deflection of the center of the diaphragm, its global deformation, or the local strain (by strain gages, Section 2.2). Some metals used are beryllium–copper, stainless steel, and nickel–copper alloys. A micromachined diaphragm is an etched silicon wafer with diffused or implanted gages that sense local strain (Figure 1.14*e*). Cars and hospitals use silicon pressure sensors by the millions. The diaphragm and elements bonded on it must be compatible with the medium and withstand the required temperature. Stainless steel diaphragms can protect sensing diaphragms from corrosive media, but in order to couple both diaphragms we need to interpose a fluid, which increases the sensor compliance and thermal sensitivity. Ceramic (96% Al_2O_3, 4% SiO_2) and sapphire (Al_2O_3) are highly immune to corrosive attack; but because they are very expensive, their use is restricted to the more demanding applications involving aggressive media, high temperature, or both.

For a thin plate with thickness t and radius R experiencing a pressure difference Δp across it, if the center deflection is $z < t/3$, we have [2]

$$z \cong \frac{3(1 - \nu^2)R^4}{16Et^3}\Delta p \tag{1.44}$$

where E is Young's modulus and ν the Poisson's ratio for the plate material. Large, flexible diaphragms undergo large deflection but have large compliance. Thin plates yield large deflections but are fragile. An alternative to sense the central deflection is to use a rod to transmit force to a cantilever beam with bonded strain gages, away from media temperature. Ceramic and some silicon pressure sensors rely on the capacitance change between an electrode applied on the diaphragm and one fixed electrode.

Piezoresistive sensors distributed on the diaphragm can sense radial and tangential strain. They are connected in a measurement bridge to add their signal and compensate temperature interference (Section 3.4.4).

Capsules and bellows yield larger displacements than diaphragms. A capsule (Figure 1.14*f*) consists of twin corrugated diaphragms joined by their external border and placed on opposite sides of the same chamber. A bellows (Figure 1.14*g*) is a flexible chamber with axial elongation that undergoes deflections larger than capsules, up to 10% of its length. Capsules and bellows are vibration- and acceleration-sensitive, do not withstand high overpressures, and have high compliance, hence poor dynamic response. Their displacement, however, can be sensed by an inexpensive potentiometer.

Pressure between contacting surfaces can be measured by a thin plastic film (Fuji Prescale Film) whose color increases for increasing pressure.

1.7.3 Flow Velocity and Flow-Rate Sensors

Flow is the movement of a fluid in a channel or in open or closed conduits. The flow rate is the quantity of matter, in volume or weight, that flows in a unit time. Flow rate is measured in all energy and mass transport processes to control or monitor those processes and for metering purposes—for example, water, gas, gasoline, diesel, and crude oil. Table 1.8 lists some measurement principles used in flowmeters. Chapters 28 and 29 in reference 13 discuss them. Each issue of *Measurements & Control* lists the manufacturers of different flowmeters: turbine (February); electromagnetic (April); anemometers and vortex (June); differential pressure, rotameters, and mass (September); positive displacement and ultrasonic (October); and open-channel, target, and flowmeters based on laminar flow elements (December).

A viscous or laminar flow is that of a fluid flowing along a straight smooth-walled and uniform transverse section conduit, where all particles have a trajectory parallel to the conduit walls and move in the same direction, each following a streamline. In turbulent flow, in contrast, some of the fluid particles have longitudinal and transverse velocity components—thus resulting in whirls—and only the average velocity is parallel to the axis of the conduit. In laminar flow, the fluid velocity profile across the conduit is parabolic. In turbulent flow, the fluid velocity profile is flatter.

The commonest flowmeters measure the drop in pressure across an obstruction inserted in the pressurized pipe in which we wish to measure the flow rate. Bernoulli's theorem relates fluid pressure, velocity, and height. It applies to an incompressible fluid experiencing only gravity as internal force (i.e., without friction) flowing in stationary movement and with no heat entering or leaving it. Any change in velocity produces an opposite change in pressure that equals the change of kinetic energy per unit of volume added to the change due to any

TABLE 1.8 Measurement Principles Used in Flowmeters

Input Quantity	Measurement Principle	Output Signal
Fluid velocity: local	Pitot probe	Differential pressure
	Thermal (hot wire anemometry)	Temperature
	Laser anemometry	Frequency shift
Fluid velocity: average	Electromagnetic	Voltage
	Ultrasound: transit time	Time
	Ultrasound: Doppler	Frequency
Volume flow rate[a]	Orifice plate	Differential pressure
	Venturi tube	Differential pressure
	Pitot probe	Differential pressure
	Flow nozzle and tube	Differential pressure
	Elbow	Differential pressure
	Laminar flow element	Differential pressure
	Impeller (paddlewheel)	Cycles, revolutions
	Positive displacement	Cycles, revolutions
	Target (drag force)	Force
	Turbine	Cycles, revolutions
	Variable area (rotameter)	Float displacement
	Variable area (weir, flume)	Level
	Vortex shedding	Frequency shift
Mass flow rate	Coriolis effect	Force
	Thermal transport	Temperature

[a] Volume flow rate can also be calculated by multiplying the average fluid velocity by the pipe cross section.

difference in level. That is, along a flow streamline we have

$$p + \rho g h + \frac{\rho v^2}{2} = \text{constant} \tag{1.45}$$

where p is the static pressure, ρ is the fluid density (incompressible), g is the acceleration of gravity, h is the height with respect to a reference level, and v is the fluid velocity at the point considered. When studying actual fluid flows, (1.45) is corrected by experimental coefficients.

The primary sensor in obstruction flowmeters is a restriction having constant cross section that obstructs the flow. For example, if we introduce in a pipe a plate having a hole, the fluid vein contracts, thereby changing from a cross section A_1 (that of the pipe) to a cross section A_2 (that of the hole) (Figure 1.15). Because of the principle of mass conservation, a cross-section change results in a corresponding change of velocity,

$$Q = A_1 v_1 = A_2 v_2 \tag{1.46}$$

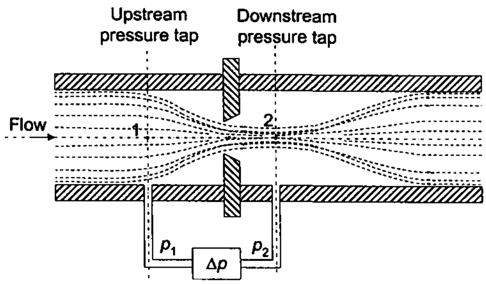

Figure 1.15 An orifice plate inserted in a pipe produces a drop in pressure related to the flow rate.

At the same time, from (1.45) we have

$$p_1 + \rho g h_1 + \frac{\rho v_1^2}{2} = p_2 + \rho g h_2 + \frac{\rho v_2^2}{2} \tag{1.47}$$

If $h_1 = h_2$, these two equations yield

$$v_2 = \sqrt{\frac{2(p_1 - p_2)}{\rho\left[1 - \left(\frac{A_1}{A_2}\right)^2\right]}} \tag{1.48}$$

Therefore, we can calculate the velocity from the drop in pressure across the plate, and we can determine the theoretical volumetric flow rate from $Q = A_2 v_2$. The real flow rate is somewhat lower and it is determined by experimentally calculating a correction coefficient, called a discharge factor, C_d, that depends on A_1, A_2, and other parameters. Then, $Q_r = C_d Q$. Tables in standards (ASME, ISO) give C_d for different pipe diameter and hole position and size, flow regimes, and pressure ports placement. For orifice plates we have $C_d \approx 0.6$.

Orifice meters produce a loss in pressure and cannot easily measure fluctuating flows unless the differential pressure sensor is fast enough, including the effects of the hydraulic connections. Flow nozzles and Venturi tubes (Figure 1.16) are based on the same principles but their internal shapes are not so blunt, thus reducing the loss of pressure (C_d can reach 0.97).

Variable-area flowmeters are primary sensors that apply Bernoulli's theorem and the principle of mass conservation in a way reciprocal to that described. They make the fluid pass section variable and keep the difference in pressure between both sides of the obstruction constant. The measured flow rate is then related to the area of the pass section.

The rotameter in Figure 1.17 applies this method. It consists of a uniform conic section tube and a grooved float inside it that is dragged by the fluid to a