CHAPTER 1

Basic Concepts

This chapter is devoted to topics that are common to all measurement devices.

Measurement devices can be characterized in several different ways. In regard to the measured value, some are continuous and some are discrete. In regard to time, some are continuous and some are sampled. In regard to their relationship to the process, some are in-line, some are on-line, and some are off-line.

The steady-state characteristics of a measurement device often determine its suitability for a given purpose. This includes its measurement range, its accuracy, its repeatability, the resolution of the measured value, and its turndown ratio. Measurement uncertainty is receiving increasingly more attention and will probably receive even more in the future.

Most measurement devices provide values for functions performed by other systems, including data acquisition and process control. The older interfaces consisted largely of current loops. Although most microprocessor-based transmitters also provide a current loop output, the trend is to use network communications with field devices, a technology generically referred to as *fieldbus*. However, serial communication has not entirely disappeared.

The sensor portion of a measurement device is generally exposed to process temperatures, process pressures, etc. Considerations such as ambient temperature and the hazardous area classification apply to the transmitter part of the measurement device. Proper enclosures are required for every measurement device.

The dynamic characteristics of a measurement device are especially important in applications such as process controls. Lags—first-order lags or transportation lags (dead times)—may be associated with the measurement device. Filtering and smoothing the measured value result in additional lags, so these technologies must be applied very carefully so as to not degrade the performance of the process controls.

Basic Process Measurements, by Cecil L. Smith

1.1. CONTINUOUS VS. DISCRETE MEASUREMENTS

Continuous and *discrete* refer to the type of value that is produced by the measurement device.

Continuous Measurements

The output of a continuous measurement device (often called a *transmitter*) indicates the current value of the variable being measured. The element *LT* (level transmitter) in Figure 1.1 represents a continuous measurement of the level in the evaporator. Provided the level is within the range covered by the measurement device, the output from the level transmitter indicates the level within the evaporator.

All continuous measurement devices are constrained by their measurement range, which in turn may be constrained by the technology, by the design parameters of the measurement device, by how the measurement device is connected to the process, and so on. The following terms pertain to the measurement range:

Lower-range value. Lower limit of the measurement range.

Upper-range value. Upper limit of the measurement range.

Span. Difference between the upper-range value and the lower-range value.

The output of a continuous measurement device is generally referred to as the *measured value* or *measured variable*. In process applications, the output of a continuous measurement device is often called the *process variable*.

Other examples of variables for which continuous measurements are available are temperature, pressure, flow, density, and composition. Parameters such

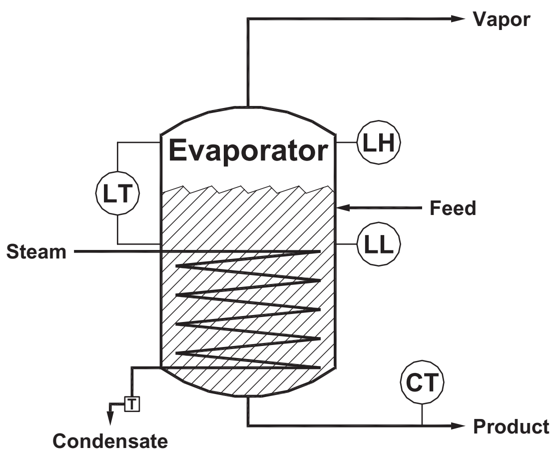

Figure 1.1. Measurements for an evaporator.

as accuracy, repeatability, turndown ratio, and resolution are associated with continuous measurements. These will be examined later in this chapter.

Discrete Measurements

The output of a discrete measurement device (often called a *switch*) is one of two states, depending on the value of the variable being measured. The elements *LH* (level high) and *LL* (level low) in Figure 1.1 represent discrete measurement devices in the form of level switches, one to detect that the liquid level is abnormally low and the other to detect that the liquid level is abnormally high.

The level switch basically indicates the presence or absence of liquid at a given point within the vessel, usually at the physical location of the switch. If the liquid level is above this location, the output of the level switch is one state. If the liquid level is below this location, the output of the level switch is the other state. In practice, there is always a small switching band (sometimes called the *deadband*) associated with a discrete measurement device.

The parameters associated with process switches are simpler than those associated with transmitters, and we will examine these parameters next.

Actuation and Reactuation The state of the process switch changes when the appropriate conditions are present within the process. Consider the level switches within the evaporator in Figure 1.1. Each level switch changes state when the level within the evaporator attains the location of that level switch.

These level switches are said to actuate on rising level. Similarly, a pressure switch actuates on rising pressure. The behavior of a level switch is as follows (most other switches behave in a similar manner):

Actuation point. This is the vessel level at which the switch changes state on rising level. Sometimes the actuation point is referred to as the *set point*. The actuation point of most level switches cannot be adjusted; the actuation point is determined by the physical location of the level switch. However, many pressure switches provide an adjustable actuation point. When the actuation point of a pressure switch is adjustable, the working pressure is the range of pressures over which the actuation point can be specified.

Reactuation point. This is the vessel level at which the switch changes state on falling level. This occurs at a level below the actuation point. The *deadband* is the difference between the actuation point and the reactuation point. In most switches, the deadband is fixed, but it is occasionally adjustable. A few pressure switches provide separate adjustments for the actuation point and the reactuation point (or the actuation point and the deadband).

This behavior is often represented as a diagram (Fig. 1.2).

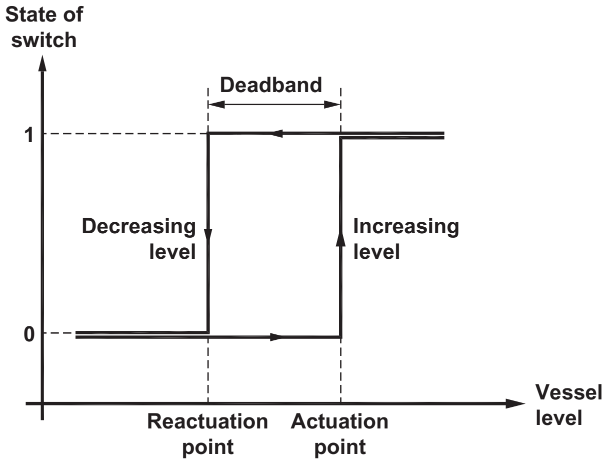

Figure 1.2. Switching logic for a level switch.

Normally Open and Normally Closed The terms *normally open* and *normally closed* are basically equipment terms, not process terms. The normal state of a switch in no way implies that the corresponding process conditions are normal.

The normal state of a switch is its state at ambient conditions. Some authors refer to this as its *shelf state*, and this is a good way to think of the normal state of a switch. The normal state is the state of the switch when removed from the process and placed in the warehouse. For a level switch, the normal state would not indicate the presence of liquid.

Within each switch, there is a contact whose state can be sensed. Figure 1.3 shows several possible configurations:

- A single-pole, single-throw (SPST), normally open (NO) switch provides only one contact whose shelf state is open. A level switch of this type would be referred to as a *normally open level switch*. On actuation, this contact closes.
- An SPST normally closed (NC) switch provides only one contact whose shelf state is closed. A level switch of this type would be referred to as a *normally closed level switch*. On actuation, this contact opens.
- A single-pole, double-throw (SPDT), switch provides a contact for both states of the switch. As illustrated in Figure 1.3, there are three wiring connections:

  Common. This is the return or ground for the electrical circuit.

  NO. In the normal, or shelf, state for the switch, this contact is open (or no continuity between the NO terminal and the common terminal). On actuation, this switch closes.

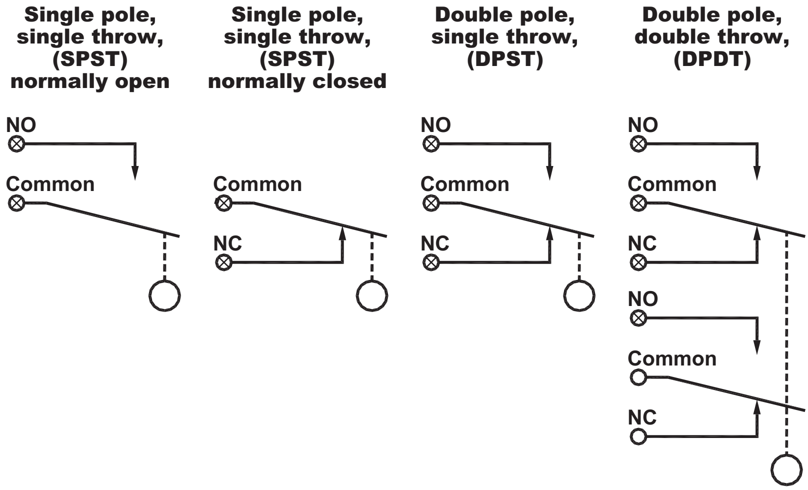

Figure 1.3. Wiring contacts for a level switch.

NC. In the normal, or shelf, state for the switch, this contact is closed (or continuity between the NC terminal and the common terminal). On actuation, this switch opens.

- A double pole, double throw (DPDT) switch is basically two double throw switches driven by the same mechanism. The switch provides six wiring connections.

The simple switches are single throw and must be ordered as either normally open or normally closed (the option is usually but not always available). Most switches designed for process applications are double pole and can be wired to either the normally open contact or the normally closed contact.

Wiring Diagram Symbols The symbols used in the wiring diagrams reflect the type of switch (level, pressure, etc.) and specify which contact (NO or NC) is used. Figure 1.4 provides symbols as commonly used in wiring diagrams. The symbols are provided as follows:

- Level switch that actuates on rising level.
- Pressure switch that actuates on rising pressure.
- Temperature switch that actuates on rising temperature.
- Flow switch that actuates on increasing flow.
- Physical contact, often referred to as a limit switch, that is used on two-position valves, on doors, etc.

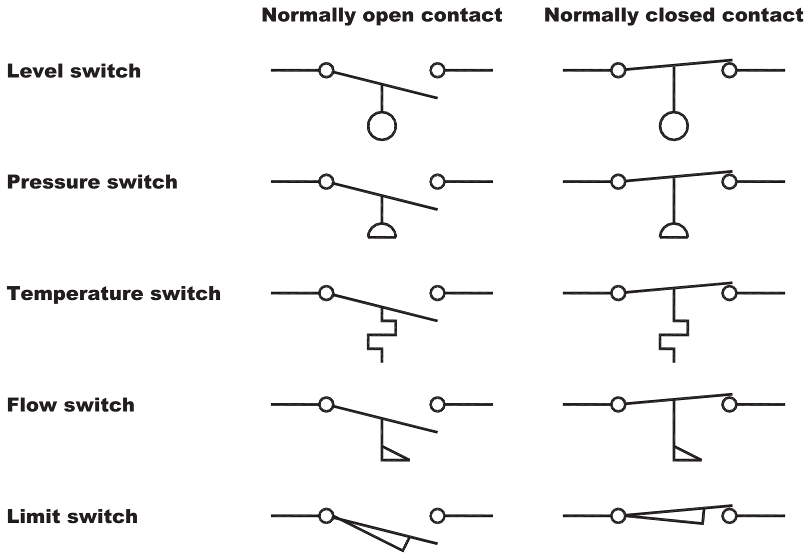

Figure 1.4. Symbols for various process switches in wiring diagrams.

Most companies have their preferred symbology for the various switches, but that given in Figure 1.4 is typical.

The required logic can be formulated using either the NC contact or the NO contact. The selection of which contact to use is determined by what conditions within the process are considered to be normal. The guiding principle is simple: When the process conditions are normal, all circuits that include a process switch must have continuity, which means that the wiring is to whichever contact (NO or NC) has continuity when the process conditions are normal. In this way, any failure in the circuit will indicate a problem. Otherwise, a defect in the circuit could easily go undetected and abnormal conditions would not be indicated.

Discrete Logic Process switches often provide the inputs to safety and shutdown systems. For the evaporator in Figure 1.1, the steam is to be blocked in either of the following conditions:

Low level. With an abnormally low level, the upper part of the tube bundle would not be submerged in liquid, which usually results in scaling or some other detrimental effect on the heat transfer surface.

High level. With an abnormally high level, liquid would be entrained into the vapor stream exiting the evaporator (many evaporators have a mist extractor in the top that must not be partially submerged in the liquid).

Another way of viewing this is to state the conditions or permissives that must be true for the steam block valve to open:

Evaporator LL switch indicates presence of liquid. Under this condition, the LL switch is actuated. The NO contact would be closed, thus providing continuity in an electrical circuit.

Evaporator LH switch does not indicate presence of liquid. Under this condition, the LH switch is not actuated. The NC contact would be closed, thus providing continuity in an electrical circuit.

Traditionally, such logic was implemented in hard-wired electrical circuits. Figure 1.5 presents the wiring diagram for a circuit that determines if the current process conditions permit the steam-block valve to be open. The *circle* represents a coil that indicates that it is okay for the steam-block valve to be open (contacts on this coil would be used in other circuits that open the steam-block valve). The coil is energized if power flows from the power rail to ground, passing through the coil. Power will flow if the normally open contact on the level low switch is closed and the normally closed contact on the level high switch is closed. Continuity is required for the steam-block valve to be open.

Today, such logic is more likely to be implemented in programmable electronic systems such as programmable logic controllers (PLCs). At least in the United States, the representation of this logic is usually by relay ladder diagrams that are similar to the wiring diagrams used for hard-wired implementations. However, this is discrete logic and can be represented and implemented in a number of ways, including Boolean expressions and sequential function charts.

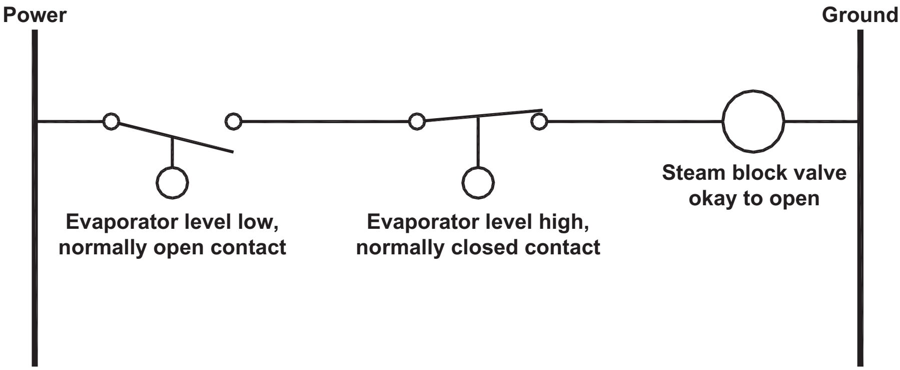

Figure 1.5. Permissive logic for a steam-block valve.

1.2. CONTINUOUS VS. SAMPLED MEASUREMENT

Continuous measurement and *sampled measurement* refer to the frequency with which the value from the measurement device is updated to reflect current conditions within the process.

Continuous Measurements

At every instant of time, a continuous measurement device provides a value that reflects the current value of the variable being measured. The level transmitter and both level switches in Figure 1.1 are continuous in this sense. Measurements for temperature, pressure, level, and flow are almost always continuous from the perspective of time.

When microprocessors are incorporated into the measurement device, the output is updated very frequently but technically not continuously. With update rates such as 10 times per second, the result is equivalent (from a process perspective) to a continuous measurement device, and such devices are normally included in the continuous category.

Sampled Measurements

The element CT (composition transmitter) on the product stream from the evaporator in Figure 1.1 is possibly a sampled measurement. A few composition analyzers are continuous, but many involve sampling. A sample is withdrawn from the process, and the analysis is performed on this sample. Sometimes a complete composition analysis (consisting of several values) is generated; sometimes the measurement is reduced to a single value, such as the ratio of two key components or the total amount of impurities.

The *sampling time* is the time between analyses. The value for the sampling time depends on the complexity of the analysis. It may be as short as a few seconds, or it could be several minutes. Sometimes analyzers are multiplexed between several process streams, which extends the sampling time even further for a given measured variable.

When a *sample-and-hold* capability is incorporated, a value for the output is available at every instant of time. However, the output reflects the results of the most recent analysis and will not change until a new analysis is performed. Consequently, such measured values are still considered to be sampled and not continuous.

1.3. IN-LINE, ON-LINE, AND OFF-LINE

In-line, *on-line*, and *off-line* pertain to the physical relationship between the measurement device and the process. This discussion will also refer to the two categories of properties:

Intensive. An intensive property does not depend on the amount of material present. Intensive properties include temperature, composition, and physical properties. The values for such properties can be obtained by withdrawing a sample of material from the process and then analyzing for the desired value. For example, one could withdraw a sample and then determine its temperature. Although rare for temperature measurement, composition measurements are routinely done in this manner.

Extensive. An extensive property depends on the amount of material present. Extensive properties include flow, weight, and level. Measurements of these cannot be performed on a sample. For example, it is not possible to determine the flow through a pipe by withdrawing a sample.

In-Line Measurements

An in-line measurement is connected in such a manner that the measurement device directly senses the conditions within the process. Most basic measurements (temperature, pressure, level, and flow) are in-line. However, very few composition measurements are in-line.

In-line is the preferred approach. Unfortunately, this option is not always available. Occasionally in-line measurements are available, but they have major concerns that must be addressed. For example, the composition of the product from a caustic evaporator can be inferred from the density. One approach to measuring density is to use a nuclear density gauge. Such gauges sense the density of the material flowing in the product pipe and are thus in-line measurements. But before such gauges are installed, the issues associated with having radioactive materials on-site must be addressed.

In-line measurements can be classified as *contact* and *noncontact*. Noncontact measurement devices perform their functions without any contact to the contents of the process. Examples of noncontact measurement devices include the following

- Radiation devices for measuring level or density.
- Pyrometers for measuring temperature.
- Clamp-on versions of ultrasonic flow meters (mounted externally to the pipe containing the fluid).
- Microwave radar level measurement, sometimes referred to as *noncontact radar*. Because the antenna is mounted above the surface to be detected, it is noncontact with respect to the liquid or solids; however, the antenna is exposed to the gases and vapors present above the surface, which over time can lead to deposits and buildups in some applications.

Noncontact does not always provide total immunity from process considerations. Consider the clamp-on ultrasonic flow meters. The transmitters/receivers are not exposed to the process fluids, but they must be bonded to the outside surface of the pipe. Therefore, the transmitters/receivers are exposed to temperatures only slightly different from that of the process fluid.

In-line measurements are also classified as *intrusive* or *nonintrusive*. A nonintrusive measurement device performs its functions without affecting the activities within the process in any way. Examples of nonintrusive measurement devices include the following:

- A magnetic flow meter must contact the fluid but does not provide any obstruction to fluid flow. The flow tube of such a meter is basically a straight length of pipe that would have the same effect on fluid flow as a spool section of the same length. There are no regions of low pressure that could lead to flashing or cavitation.
- Pressure transmitters require a process connection but one that usually does not intrude into the process. When flush connections are required to avoid dead spaces, capillary seal arrangements can be installed.
- The antenna of a microwave radar transmitter normally extends slightly into the process vessel. Technically, this is intrusive but only to a nominal degree. If necessary, a plastic seal can be installed to completely separate the antenna from the process. But whereas the exposed antenna can be inserted through a nozzle as small as 2 in., a 12 in. or larger opening is required for the seal arrangements.

On-Line Measurements

On-line measurements sense the conditions of materials that are withdrawn from the process. The most common example of this is associated with composition analyzers. Figure 1.6 shows an installation in which a sample is withdrawn from the process, transported to the physical location of the analyzer, and then conditioned or cleaned up in some manner to provide material for analysis. From the perspective of time, the analyzer may be continuous (such as infrared or ultraviolet) or may be sampling (chromatograph).

Known as a *sample system*, the equipment for withdrawing material from the process, transporting it to the analyzer location, and then removing unwanted constituents tends to be a source of problems. The design of such systems requires special skills in handling small streams in an industrial environment. Any lapses lead to maintenance problems. Unfortunately, for analyzers such as chromatographs, there seems to be no other alternative.

Because they are intended for sampling, on-line measurements can be used only for intensive properties.

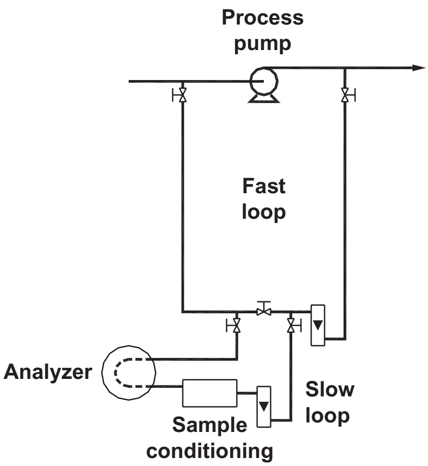

Figure 1.6. Sample loops for an on-line analyzer.

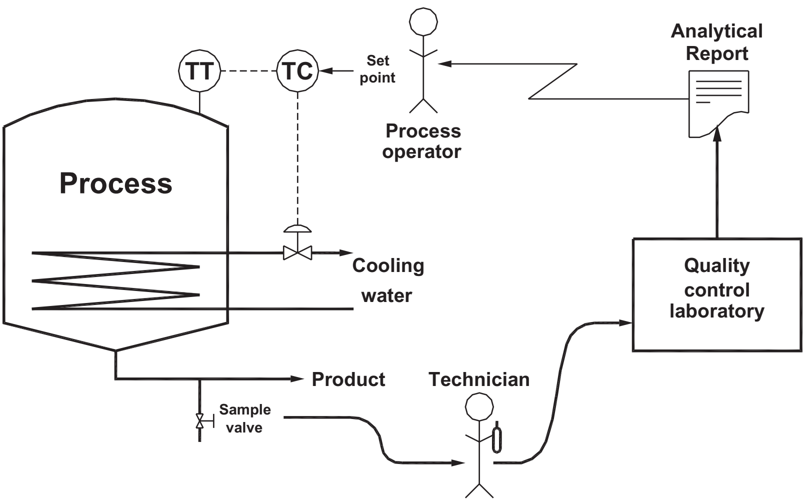

Figure 1.7. Off-line analysis.

Off-Line Measurements

Especially with the advent of robotics, almost any analysis could be implemented in an on-line fashion. But when the sampling interval is once every 4 or 8 hr, the cost justification becomes difficult. Furthermore, some very complex analyzers are much more suited to a laboratory environment (with lab technicians) than to a process environment (with process operators).

Figure 1.7 illustrates an off-line measurement. A sample is withdrawn from the process and transported to the laboratory, where the sample preparation is performed, the analysis is made, and the results are communicated back to the plant control room. From a control perspective, the major issue is how quickly the results are returned to the process operators. Technologies such as pneumatic conveying systems can speed the transport of the sample to the laboratory. Networking capabilities can make the results available in the control room as soon as they are obtained (and possibly verified) by the lab technicians.

Because they are intended for sampling, off-line measurements can be used only for intensive properties.

1.4. SIGNALS AND RESOLUTION

The output of a measurement device is generally referred to as a signal. In measurement applications, a signal is physical variable that in some way represents the process variable being measured. The signal is a mechanism for transfering information from one device (such as a measurement device) to another (such as a controller).

In process applications, the physical nature of the signal has evolved over the years and will probably continue to evolve. In the 1950s, most systems were pneumatic, using a 3- to 15-psi pneumatic signal. In the 1970s, electronic systems appeared, most using a 4- to 20-ma current loop to transmit information from one device to another. Both are analog transmission systems.

Eventually these will be replaced with digital communications using technology commonly referred to as fieldbus. These technologies permit several measured variables to be transmitted via a single physical conductor (a wire). The physical conductor can be eliminated by using technologies such as fiberoptics and wireless. However, all of these are digital communications systems.

We will later examine current loops and various forms of digital signal transmission systems, but not pneumatic systems.

Analog Values

In process applications, analog values are associated with signals such as the 3- to 15-psi pneumatic signal and the 4- to 20-ma current loop. Measurement devices such as a level transmitter can sense the level in the vessel, provided it is within the measurable range. The lower end of the measurable range corresponds to 3 psi or 4 ma; the upper end of the measurable range corresponds to 15 psi or 20 ma.

At least conceptually, an analog signal can represent any value of the vessel level, provided it is within the measurable range. In practice, there are limitations.

The resolution of the measurement device is the change in the process variable required to cause a change in the output signal. Basically, the resolution is the smallest change in the process variable that can be detected by the measurement device.

Digital Values

A digital value is an approximation to an analog value. The larger the number of digits in the digital value, the closer the approximation will be to the analog value. Computers represent data using the binary number system, and either a fixed-point or floating-point representation. But let's not get into the gory details of bits and bytes. We can illustrate our points using a fixed-point decimal representation.

Suppose the level in an evaporator can be measured over the range of 6 to 8 ft. Let's represent this as a digital value using a three-digit digital representation. The evaporator level is truly an analog value. Its conversion to the digital value is illustrated by the graph in Figure 1.8. The evaporator level is slowly increasing with a constant rate of change. However, the digital value increases in increments of 0.01 ft. Thus the minimum change that can be detected in this case is 0.01 ft. Consequently, the digital representation imposes a resolution of 0.01 ft.

In computer circles, the precision of a value is the number of digits following the decimal point. For example, the statement "Out.precision(2)" in C++ causes floating-point values written to stream Out to be formatted with two digits after the decimal point. When understood in this context, the precision of the digital value in Figure 1.8 would be 2. But within measurement circles, the term *precision* has been used (and misused) in so many ways that it is probably best to avoid it.

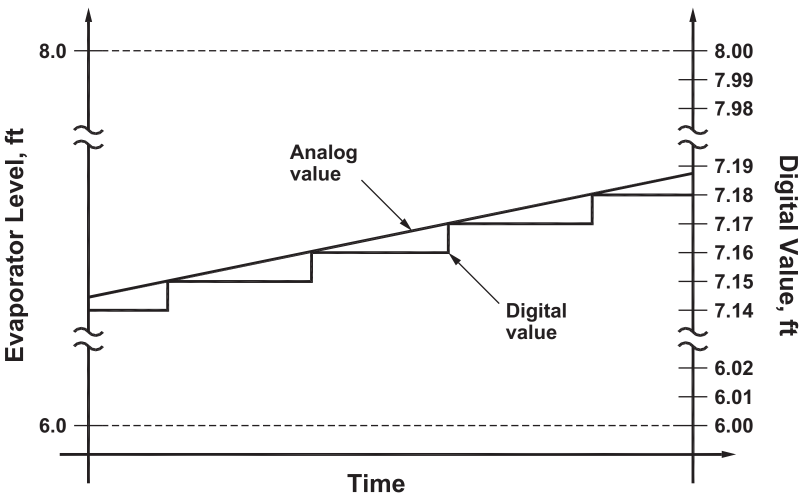

Figure 1.8. Digital representation of an analog value.

Resolution

When the measurement device provides a digital value, the best possible resolution is a change of 1 in the least significant digit of the digital value. However, it is not ensured that the measurement device can actually detect such a change. That is, the resolution can be no better than that imposed by the digital representation, but it can be worse.

The resolution can be expressed in engineering units (for example, 0.01 ft in Figure 1.8). Alternatively, it can be expressed relative to the measurable range (2.00 ft in Figure 1.8). On this basis, the resolution would be 0.01 ft in 2.00 ft, or 1 in 200 (usually read as *1 part in 200*). Most level measurement devices are capable of better resolution.

In determining the number of digits to be used in a digital value, the objective is for the resolution of the digital value to be sufficient to represent any meaningful change in the output of the measurement device. The specifications for most measurement devices state either the resolution or the repeatability (to be defined shortly). The resolution provided by the digital value should exceed both.

Voltage Inputs Consider a measurement device that outputs a 0- to 5-V DC signal that represents its measured value. An analog to digital (A/D) converter within the data acquisition or control system converts this voltage to a digital value.

Suppose a pressure measurement device generates 0 V when the pressure is 0 mm Hg absolute and 5 V when the pressure is 800 mm Hg absolute. An A/D converter with a resolution of 1 part in 4000 would generate the following digital values:

Quantity	Lower Range Value	Upper Range Value
Absolute pressure (mm Hg)	0	800
Input (V DC)	0	5
Raw value (count)	0	4000

A change of 1 in the digital value is often referred to as a *count* and cannot have a fractional part. Therefore, the resolution of the A/D converter is 1 count, which corresponds to 5 V/4000 counts = 0.00125 V/count = 1.25 mV/count. In terms of the pressure, the resolution is 800 mm Hg/4000 counts = 0.2 mm Hg/count, which is the smallest detectable change in pressure.

A 12-bit A/D converter provides 12 significant bits in the digital value and has a resolution of 1 part in 4096 = 2^{12}. The input voltage is converted to one of 2^{12} = 4096 states, which computer people always number 0, 1, 2, … , 4095. For industrial applications, the input voltage range may be either of the following:

- The input voltage range is 0 to 5.00 V. An input of 5.00 V corresponds to 4095, and the resolution is 1 part in 4096.
- Computer people just can't resist powers of 2! Even though often referred to as a 5-V A/D, the input voltage range is 0 to 5.12 V ($512 = 2^9$). Thus 5.12 V is converted to 4095 counts. An input of 5 V gives a count of 4000, so the resolution over 0 to 5 V is 1 part in 4000. The effective resolution is not quite 12 bits ($\log_2 4000 = 11.97$), but it is usually said to be 12 bits.

As in the previous example for the input from the pressure transmitter, we shall generally use a resolution of 1 part in 4000 for a 12-bit A/D. The reason is simple—the numbers are easier to work with. For the pressure transmitter, 800 mm Hg/4000 counts = 0.2 mm Hg/count, but 800 mm Hg/4096 counts = 0.1953125 mm Hg/count.

The A/D converters used in industrial systems have resolutions between 11 bits (1 part in 2000) and 15 bits (1 part in 32,000).

Current Inputs In industrial systems, most analog inputs are via a circuit known as a *current loop*. We will examine this in more detail later and explain the reasons for using DC current instead of DC voltage. The input range for the current signal is 4 to 20 ma. However, the A/D converter accepts volts, not current. By inserting a 250-Ω resistor (known as the *range resistor*) into the circuit, a current of 4 ma is converted to 1 V and a current of 20 ma is converted to 5 V.

Suppose a pressure measurement device generates 4 ma when the pressure is 0 mm Hg absolute and 20 ma when the pressure is 800 mm Hg absolute. An A/D converter with a resolution of 1 part in 4000 would generate the following digital values:

Quantity	Lower Range Value	Upper Range Value
Absolute pressure (mm Hg)	0	800
Current (ma)	4	20
Input (V DC)	1	5
Raw value (count)	800	4000

The effective input range is not 0 to 5 V; it is really 1 to 5 V. This decreases the effective resolution from 1 part in 4000 to 1 part in 3200. In terms of the pressure, the resolution is 800 mm Hg/3200 counts = 0.25 mm Hg/count, which is now the smallest detectable change in pressure.

Even though the A/D converter has a 12-bit resolution, the effective resolution of the input system is no longer 12 bits or 1 part in 4000. Instead, the effective resolution is 1 part in 3200. Sometimes this is stated as a number of bits, but with a fractional part. For this example, the resolution for the input value could be stated as $\log_2 3200 = 11.6$ bits.

At the expense of a small increase in complexity, some input systems convert the 4- to 20-ma signal into a 0- to 5-V signal so that the entire input range of the A/D can be used. Basically, one has to pay close attention to the details when examining the resolution of an input value.

Engineering Value Inputs With the incorporation of digital technology into the I/O equipment, providing a raw value in engineering units is becoming a common practice. One example is an resistance temperature detector (RTD) input card that performs the conversion to temperature units. Similar cards are available for thermocouples.

The *raw value* is the temperature in °C (or °F) with a resolution of 0.1 °C (or 0.1 °F). The raw value is represented as an integer number, so fractional parts are never present in the raw value. The location of the decimal point is at a fixed position, specifically, one decimal digit from the right. Therefore, a raw value of 1097 means a temperature of 109.7 °C (or 109.7 °F). Negative temperatures give negative raw values.

The limits on the measured value are not imposed by the I/O system; they are imposed by the RTD, specifically, –200° to 850 °C for the Class B RTDs normally installed in industrial processes. A resolution of 0.1 °C is a resolution of 1 part in 10,500 for the entire possible measurement range. But when measuring temperatures in water-based processes at atmospheric pressure (such as fermentation processes), the range of interest is 0° to 100 °C. A resolution of 0.1 °C becomes a resolution of 1 part in 1000 for the range of interest.

Display Resolution Vs. Internal Resolution When determining the resolution for a measured value, the tendency is to consider the display resolution from the perspective of what some human needs to see, such as, How many digits after the decimal point does the process operator need to see? or How many digits after the decimal point are required in the historical data records? This is the *precision* as the term is used in computer circles.

But the resolution is also important for whatever processing (such as control calculations) is being performed within the system. Suppose a process pressure is being measured over the range 0 to 10 psig. For values on the operator displays, a resolution of 0.1 psig is probably adequate. This is 1 part in 100, or 1% of the measurement range. Now suppose this pressure measurement is the input to a pressure controller with a gain of 5%/%. Each time the pressure changes by 0.1 psig (1% of the measurement range), the controller output will abruptly change by 1% × 5%/% = 5%. The bump to the process will be noticeable.

This issue is especially important for I/O systems that provide engineering value inputs. The resolution provided in the input value becomes the system's internal resolution. This resolution must be adequate for whatever internal processing, such as control calculations, is being performed by the system. If the internal resolution is more than required, the resolution can be reduced when the data are output to displays or historical files.

1.5. ZERO, SPAN, AND RANGE

As Figure 1.9 illustrates, the output of most modern transmitters varies linearly with the value of the measured variable or process variable. Such transmitters require only two adjustments:

Zero. Determines the value of the measured variable for which the transmitter output is equal to its lower-range value.

Span. Determines the change in the measured variable required to cause the transmitter output to change from is lower-range value to its upper-range value.

The zero and span adjustments determine the measurement range for the measured variable. The range can also be determined from the following two values, both in the engineering units of the measured variable:

Lower-range value. Equal to the zero adjustment.

Upper-range value. Equal to the zero adjustment plus the span adjustment.

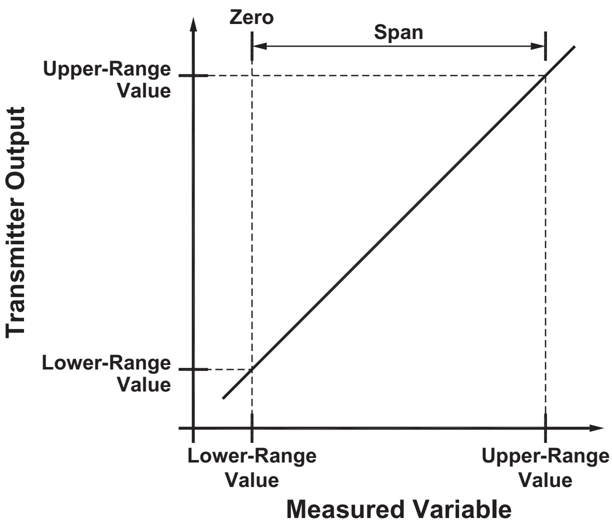

Figure 1.9. Operating line for a linear transmitter.

Specifying Zero and Span

Measurement devices based on the older technology contained physical adjustments for setting the zero and the span. The zero adjustment had to be set first, and then the span adjustment could be set.

But with microprocessors incorporated into measurement devices, the physical adjustments are replaced by software coefficients for the zero and span. Values for these coefficients can be obtained as follows:

- Specify the zero and span explicitly.
- Specify the zero and one point on the measurement device's operating line. The span is calculated from this point on the operating line.
- Specify one point on the measurement device's operating line and the span. The zero is calculated from this point on the operating line.
- Specify two points on the measurement device's operating line. The zero and span are calculated from these two points on the operating line.

Overrange

As the graph in Figure 1.9 suggests, many transmitters will function somewhat outside their measurement range. Such transmitters provide a small overrange. The amount of overrange varies, but typically is in the vicinity of 5% of the span. However, it is not recommended that the transmitter be routinely used in the measurement region provided by this overrange.

Some systems provide range alarms to provide an alert when the input from the measurement device is beyond the measurement range. These alarms usually include a deadband, a typical value being 0.5% of the span. If the lower-range value for the transmitter output is considered to be 0% and the upper-range value for the transmitter output is considered to be 100%, the typical logic for the range alarms is expressed as follows:

Occurrence. The range alarm is issued when the input from the measurement device is less than –0.5% or greater than 100.5%.

Return to normal. The range alarm returns to normal when the input from the measurement device is greater than or equal to 0% and less than or equal to 100%.

Zero Error

Figure 1.10 illustrates the behavior of a linear transmitter with an error in the specification for the zero. As shown, the transmitter output is above its lower-range value (typically 4 ma) when the variable being sensed is equal to its lower range value.

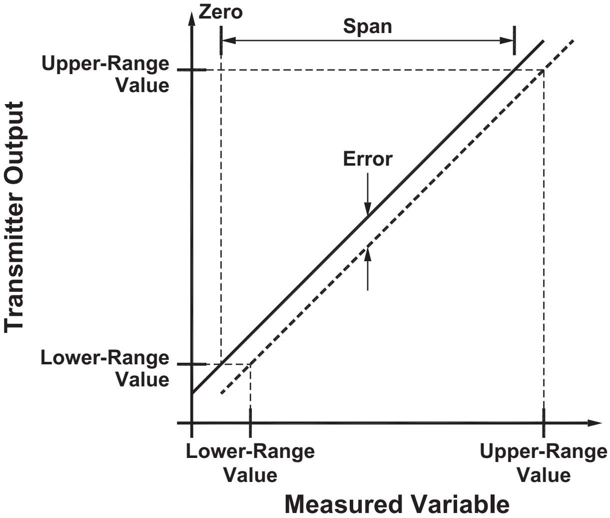

Figure 1.10. Zero error.

If the span specification is exact, any zero error results in a constant error over the entire measurement range. *Zero drift* is the change in the zero adjustment with time.

Zero drift normally causes the zero error to increase. One of the advantages of microprocessor-based (smart) measurement devices is that the zero error and zero drift are smaller than in conventional transmitters.

In some applications, a small zero error can have significant consequences. Consider a flow transmitter that provides the input to a flow controller. When a value of zero is specified for the set point, the flow controller is expected to completely close the valve. Whether this occurs or not depends on the nature of the zero error. There are two possibilities:

- The transmitter indicates zero flow when the flow is slightly positive. As the flow controller responds to the output of the flow transmitter, the flow controller will attempt to position the control valve to a very small opening to achieve the flow required for the transmitter output to be zero. This behavior is unacceptable.
- When the flow is zero, the transmitter output is a small value, suggesting a small flow. In attempting to drive the measured value of the flow to zero, the flow controller will decrease its output to the control valve as much as possible (to a value known as the lower output limit). This is the desired behavior.

As the latter is the preferred behavior, the zero for the flow measurement is often intentionally biased so that it will indicate zero for a slightly positive flow.

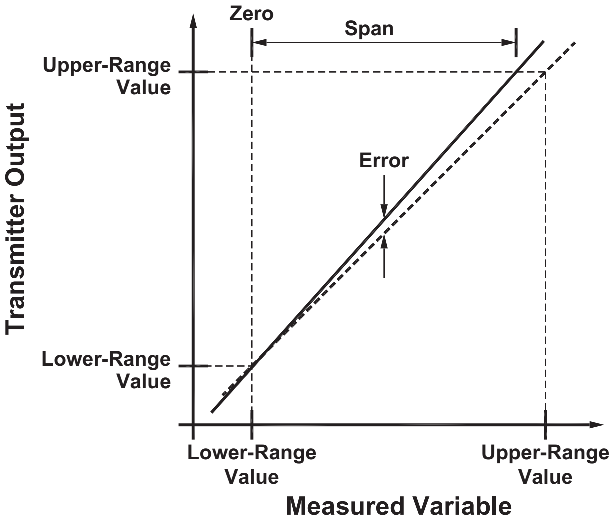

Figure 1.11. Span error.

Span Error

Figure 1.11 illustrates the behavior of a linear transmitter with an error in the specification for the span. In this example, a change in the measured variable from its lower-range value to its upper-range value gives a change in the transmitter output of less than its nominal span (16 ma for current loop installations).

If the zero specification is exact, the error is zero when the measured variable is equal to the lower-range value of the measurement range. The error increases as the process variable approaches the upper-range value of the measurement range.

Nonlinear Transmitter

Figure 1.12 illustrates the behavior of a transmitter whose output exhibits a modest degree of departure from linear behavior. If the zero and span specifications are exact, the error is zero when the measured variable is equal to either the lower-range value or the upper-range value of the measurement range.

The error in such a transmitter depends on the nature of the nonlinearity. This is often summarized as a statement for the maximum departure from linear behavior.

If the nature of the nonlinearity is known, digital systems permit a characterization function (or its equivalent) to be applied to its input value (which

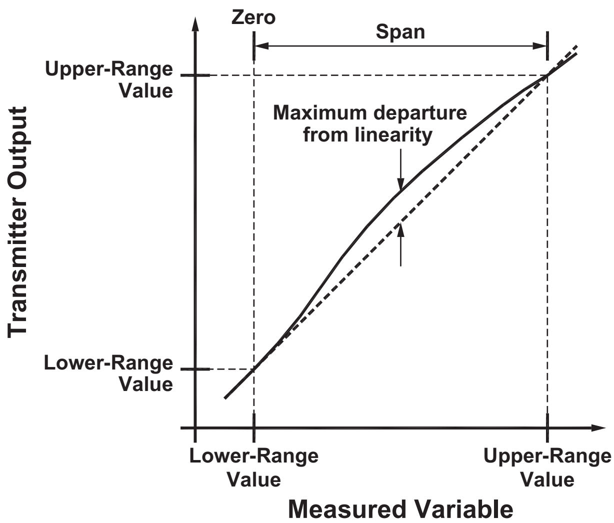

Figure 1.12. Nonlinear transmitter.

is the output of the transmitter) to compensate the measurement for its nonlinear behavior. For microprocessor-based measurement devices, the manufacturers normally incorporate such characterization functions into the measurement device itself, which usually reduces the maximum departure from linearity to an acceptable value.

Gain or Sensitivity

The gain or sensitivity of a device quantifies the degree to which a change in the input will affect the output of the device. For a measurement device, the input is considered to be the variable being measured, expressed in engineering units. The output is considered to be the output signal from the measurement device.

For a linear measurement device (including those with an insignificant departure from linearity), its gain or sensitivity (usually called the transmitter gain) is

$$\text{Transmitter gain} = \frac{\text{Span of the output signal}}{\text{Engineering span of the measured variable}}$$

The following calculations give the gain of a temperature transmitter with a measurement range of 50° to 250°F:

Signal Type	Signal Span	Gain or Sensitivity
Current loop	4–20 ma	$\frac{16\,\text{ma}}{200\,°\text{F}} = 0.08\,\text{ma}/°\text{F}$
Digital	0–100%	$\frac{100\%}{200\,°\text{F}} = 0.5\%/°\text{F}$

For electronic systems, the signal is the current loop with an output range of 4 to 20 ma. For digital systems, the output range is normally considered to be 0 to 100%. A gain or sensitivity of 0.5%/°F means that the transmitter output changes by 0.5% for each 1 °F change in temperature.

1.6. TURNDOWN RATIO AND RANGEABILITY

Although the turndown ratio can be applied to other situations, it is usually best understood for anything to which the term *throughput* is meaningful. Let's explain it in the context of a process such as a steam boiler.

Suppose the maximum capacity of a boiler is 50,000 lb/hr of steam. Does this mean that it is possible to operate the process at any steam rate below 50,000 lb/hr? Definitely not. In addition to the maximum steam rate, the designers of the boiler are given a specification for the minimum steam rate at which the boiler is to be operated. Alternatively, the designers can be given a specification for the boiler's turndown ratio, which is defined as follows:

$$\text{Turndown ratio} = \frac{\text{Maximum steam capacity for the boiler}}{\text{Minimum steam rate at which the boiler can be operated}}$$

If the boiler can be operated down to 20,000 lb/hr, its turndown ratio is 2.5 : 1.

Flow Measurements

In some applications, the flow is large at times but small at other times. How well a flow measurement device will perform in such applications depends on its turndown ratio, which is defined as follows:

$$\text{Turndown ratio} = \frac{\text{Maximum measurable flow}}{\text{Minimum nonzero measurable flow}}$$

The maximum measurable flow is usually taken to be the upper-range value for the measurement range. The lower-range value for the measurement range is usually zero. But we have to be careful with zero. Some flow meters, specifically, the vortex shedding meter, will indicate zero even if there is a small flow through the meter. The interest is the smallest nonzero flow that the meter can read in accordance with its specifications. The minimum measurable flow could be considered to be the smallest flow for which the flow measurement device can be calibrated.

One limitation of an orifice meter is that its turndown ratio is rather poor; typical values are in the range of 3:1 to perhaps 5:1. Other types of flow meters usually perform much better. For example, 50:1 is a typical value for the turndown ratio of a magnetic flow meter. If the capacity of such a meter is 100 gal/min, it can also accurately measure a flow of 2 gal/min.

Other Measurements

For a given measurement device, the maximum possible measurable value is the upper-range value. The minimum possible measurable value is the lower-range value. The maximum possible turndown ratio is as follows:

$$\text{Turndown ratio} = \frac{\text{Upper-range value}}{\text{Lower-range value}}$$

Obviously this equation has problems when the lower-range value is zero, which is unfortunately the case for most measurements for which the turndown ratio is of interest. We have already discussed flow measurements. Turndown ratio is often applied to pressure measurements, where the lower-range value is often zero (either zero absolute or zero gauge). Instead of using the lower-range value to compute the turndown ratio, the minimum nonzero measurable pressure must be used.

In industrial applications, turndown ratio is rarely of interest for temperature measurements, composition measurements, or measurements of other intensive properties.

Rangeability

The turndown ratio pertains to the performance of a measurement device with the existing settings for the zero and span. Rangeability basically assesses the ability of a given measurement device to be applied to a variety of applications. Rangeability is defined as follows:

$$\text{Rangeability} = \frac{\text{Maximum possible upper-range value}}{\text{Minimum possible upper-range value}}$$

Most measurement devices permit the upper-range value to be changed so that a given device can be applied to a range of applications. For flow measurements, some applications will require large flows; other applications will require much smaller flows. For example, if a flow measurement device can have a measurement range as low as 0 to 20 lpm but as high as 0 to 200 lpm, its rangeability is 10:1.

A high rangeability may not necessarily require a high turndown ratio. In one application, the flow may vary from 15 to 20 lpm, but in another application, the flow may vary from 150 to 200 lpm. In each application, a turndown ratio of 2:1 is adequate.

1.7. ACCURACY

Accuracy is generally thought of as the conformity of the measured value with the true value. However, how do we know the true value? We never know the true value.

To establish the accuracy of a measurement device, we compare the indicated value (the *measurand*) with that of a standard used for calibration or with a device whose accuracy is considered to be far better than that of our measurement device. An example is the use of a dead weight tester to determine the accuracy of a pressure measurement device.

Consequently, accuracy is the degree of conformity of the measured value with either a standard, reference, or other accepted value for the variable being measured. Accuracy is usually stated as the error in the measured value. Therefore, it can be argued that it is actually the inaccuracy of the measurement device that is being stated. However, the practice in the industry is to refer to these as stating the accuracy of the measurement device.

Bases for Stating Accuracy

The basis on which the accuracy of a measurement device is stated depends on the principles that are employed to design the measurement device. For a temperature transmitter with a measurement range of 50° to 250 °F, the four possibilities are of as follows if the current reading is 150 °F:

Basis	Example
Percent of span	±1% of 200 °F = ±2 °F
Percent of reading	±1% of 150 °F = ±1.5 °F
Percent of upper-range value	±1% of 250 °F = ±2.5 °F
Absolute accuracy	±1 °F

Process operations normally depend on the absolute accuracy stated as a value in the engineering units of the measured variable. To compare performance, all expressions of accuracy must be translated to this basis. However, if the configuration parameters of the measurement device are altered, the absolute accuracy may change.

Traditionally, accuracy was most frequently stated as percent of span. But with microprocessor-based measurement devices, percent of upper-range value has become very common. In some cases, these are equivalent. For example, the measurement range of most flow measurement devices is from zero to an upper-range value. In such cases, percent of span and percent of upper-range value are the same.

But consider a microprocessor-based temperature transmitter whose measurement range is 0° to 600°F. If the accuracy is stated as percent of upper-range value, changing the measurement range to 500° to 600°F has no effect on the absolute accuracy (in °F). But if the accuracy is stated as percent of span, changing the measurement range to 500° to 600°F improves the absolute accuracy by a factor of six.

Accuracy as Percent of Reading

For some applications, accuracy stated as percent of reading is preferred over accuracy expressed as either percent of span or percent of upper-range value. A flow meter whose accuracy is stated as percent of reading performs far better at low flows than does a flow transmitter whose accuracy is stated as percent of upper-range value. The error expressed as a percent of reading increases rapidly as the flow decreases. The relationship is as follows:

$$\begin{array}{c}\text{Accuracy as percent}\\ \text{of reading}\end{array} = \frac{(\text{Accuracy as percent of upper-range value}) \times F_{\text{MAX}}}{F}$$

where F_{MAX} = upper-range value for the flow measurement, F = measured value for flow.

For a measurement device with a stated accuracy of 1% of upper-range value, Figure 1.13 presents the accuracy as a percent of reading as a function of the measured value of the flow. On the basis for percent of reading, the accuracy is very poor at low flows. This behavior often limits the minimum flow for which the measured value is acceptable.

Limitations

Suppose a manufacturer states the accuracy of a viscosity measurement as "±2% for Newtonian fluids of low viscosity." In industrial practice, the viscosity of a Newtonian fluid of low viscosity seems rarely to be of much interest. But from the manufacturer's perspective, non-Newtonian behavior can be in a variety of forms. Basically, it becomes the responsibility of the user to translate this statement into something that applies to a specific application. The potential of a sale creates an incentive on the part of the manufacturer to assist the user in making this translation. However, the user will have to provide information on the behavior of the fluid whose viscosity is to be measured.

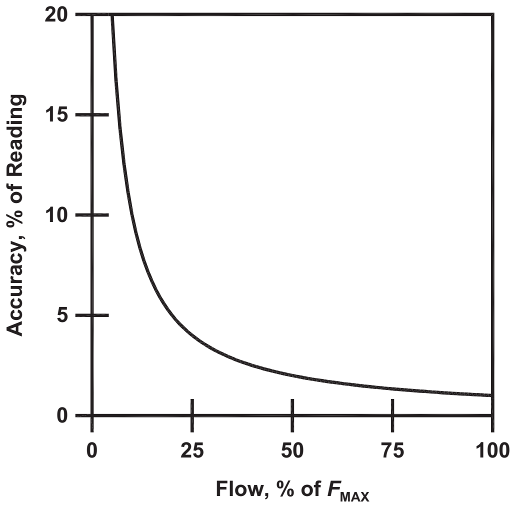

Figure 1.13. Absolute accuracy as a function of flow for a meter with a stated accuracy of 1% of the upper-range value.

Many statements of accuracy are accompanied by statements regarding process conditions. Furthermore, the accuracy is sometimes stated for measuring some property of air or water, often at or near ambient conditions. Conditions within the process can be far removed from such conditions.

Installation

A measurement device is not always directly exposed to the process variable whose value is of interest. For example, the temperature probe in Figure 1.14 is installed within a thermowell to measure the fluid temperature T_F. The accuracy statements from the supplier of the temperature probe will apply to the probe temperature T_P, not to the fluid temperature. Often we casually assume that the probe temperature is the same as the fluid temperature, but is this really correct?

Manufacturers often advise as to proper installation procedures, but it is ultimately the user's responsibility to install a measurement device properly. For temperature probes inserted into thermowells, spring-loaded assemblies apply pressure to ensure good metal-to-metal contact between the tip of the probe and the thermowell. Alternatively, the thermowell can be filled with a heat-conducting liquid.

Users are also advised to insert the thermowell at a location where the fluid motion is significant and to keep the mass of the probe as low as possible. But even if all such advice were followed, would the temperature of the probe be exactly the same as the fluid temperature? Probably not. At process operating conditions, how different can be difficult to ascertain.

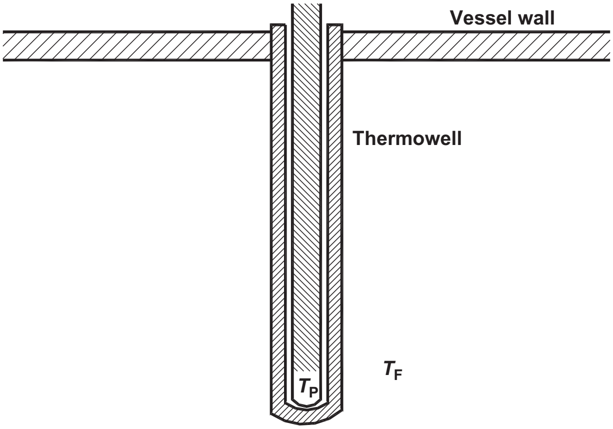

Figure 1.14. Temperature probe inserted into a thermowell.

1.8. REPEATABILITY

Although somewhat controversial, the contributions to the error in a measured value have traditionally been viewed as of two types:

Random. Errors of this type lead to the scatter observed when repeated measurements are made. This type of error can be quantified by computing the standard deviation or variance for a set of measurements.

Systematic. Errors of this type lead to a bias or offset between the measured value and the true value (which in practice is never known). Repeated measurements using the same device will not detect systematic errors. Even a set of measurements with a small standard deviation could be significantly different from the true value, whatever that is.

Quantifying the systematic error is far more difficult than quantifying the random error. Calibration attempts to address this issue by comparing the result to the value indicated by a more accurate measurement device. However, the possibility exists that there may be bias in the value from the more accurate device and even that this bias is the same as the bias in the device being calibrated. Comparing devices that rely on different measurement principles (for example, orifice meter vs. turbine meter) reduces the probability that the two biases will be the same, but not necessarily to zero.

Concept

Repeatability focuses only on the random errors and basically ignores systematic errors or biases. It applies to a given measurement device in a given application and can be viewed from two perspectives:

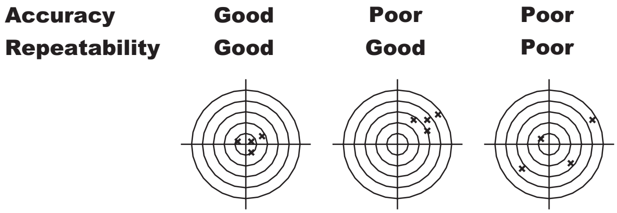

Figure 1.15. Accuracy and repeatability.

- The process conditions are exactly the same on two different occasions. The difference in the measured value on these two occasions reflects the repeatability (or actually the nonrepeatability) of the measurement device.
- The measured values are exactly the same on two different occasions. The difference in the process conditions on these two occasions reflects the repeatability (or actually the nonrepeatability) of the measurement device.

The scatter patterns for the targets in Figure 1.15 illustrate the concepts of accuracy and repeatability.

Repeatability is similar to accuracy in that it can be expressed on any of the following bases:

- Percent of span.
- Percent of reading.
- Percent of upper-range value.
- Absolute repeatability (in engineering units).

Accuracy Vs. Repeatability

For some measurement applications, accuracy is crucial, but for others, repeatability is crucial. Figure 1.7 illustrates an off-line analysis in which a sample is taken to the quality control (QC) lab for analysis, and the result's reported to the process operator. Suppose the operator adjusts the target for a temperature controller based on the results of the analysis. In this application:

QC lab analytical instrument. Good accuracy is required. Bias in this measured value must be minimized.

Process temperature measurement. Good repeatability is required. Bias in this measured value is not a problem. Often the operator is thinking in terms of changes; to make a certain change in the results of the analysis, the operator knows by experience what change is required in the process temperature. Modest errors in the measured value of the process temperature do not impair the operator's performance, provided the errors are the same from day to day.

The requirements for accuracy vs. repeatability are often summarized as follows:

- Regulatory control is focused on maintaining constant conditions within the process. This can be achieved provided the measurement device has good repeatability.
- Endeavors such as process optimization and constraint control depend on the accuracy of the measurement device.

Reproducibility

Repeatability applies to a given measurement device in a given application. *Reproducibility* applies to different measurement devices in the same application. There are two scenarios:

- A given measurement device fails and is replaced "in kind"—that is, maintenance removes the faulty device and installs a new one of the same model. The agreement (or actually the nonagreement) of the measured values from the two devices reflects the reproducibility for this specific model of the measurement device.
- A given measurement device is replaced by a measurement device based on a different measurement principle. For example, an orifice meter is replaced by a mass flow meter. How do the measured values from the mass flow meter agree with the measured values from the orifice meter? Rarely are the differences trivial. If you abruptly switch from basing process operations on the orifice meter to basing process operations on the mass flow meter, the consequences are likely to be noticeable. After installing the mass flow meter, you should continue to run the process based on the measured values from the orifice meter. Note the difference (the nonreproducibility) between the measured values from the two meters and establish targets for the mass flow meter that reflect the current manner of process operation. Using these targets will give a smooth transition from one meter to the other. Thereafter, the targets can be adjusted as warranted.

Precision

When reading specifications for measurement devices, you will encounter the term *precision*. But as used, this term unfortunately has a variety of definitions, in fact, so many that the National Institute for Standards and Technology (NIST) does not use it. But at least there is agreement on one aspect:

Precision is not another word for accuracy.

In some cases, precision is used as another term for repeatability. When a set of measured values is available, precision is sometimes considered to be the standard deviation of the measured values. Occasionally, precision is used as another term for resolution—that is, the number of significant figures in the measured value. Sometimes the manner in which precision is used will be clear, but too often it is not. Make no assumptions here. If you are relying on this term, have the supplier provide its definition and make sure you understand it.

Measurement Uncertainty

Suppose a temperature measurement device with a manufacturer-stated accuracy of ±0.5 °C indicates 152.1 °C. This is generally understood to mean that the true value is between 151.6 °C and 152.6 °C. How did the manufacturer conclude that the accuracy is ±0.5 °C? You will have to ask the manufacturer. There is no generally agreed on method for establishing assessments of the accuracy of measurement devices.

Does 152.1 °C with an accuracy of ±0.5 °C imply that the true value is more likely to be 152.1 °C than 151.6 °C or 152.6 °C? Except when there are specific reasons to believe otherwise, a normal or Gaussian distribution is usually assumed. But unless the manufacturer can provide additional information, no conclusions on the probability of values within the interval of 151.6 °C to 152.6 °C can be definitively made. Any probability distribution function, including those in Figure 1.16, is possible.

The current direction is to use the term *measurement uncertainty* in lieu of accuracy, repeatability, and the like to quantify the performance of a measurement device. Measurement uncertainty quantifies the probability associated with a measured value. NIST now views accuracy, repeatability, and so on as

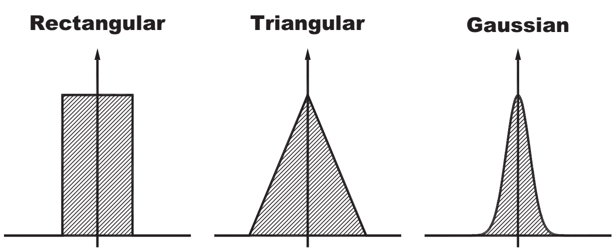

Figure 1.16. Possible distribution functions for a measured value.

qualitative terms. However, most manufacturers of measurement devices continue to state quantitative values for these.

1.9. MEASUREMENT UNCERTAINTY

To understand the need for developing an expression for the uncertainty associated with a measurement, the following statement must be recognized:

It is impossible to obtain the true value for any measurand.

Measurement is not an exact science. Therefore, it is imperative to agree on:

- How to express the uncertainty associated with a measured value.
- The approach used to obtain quantitative values for the uncertainty associated with a measured value.

The *International Vocabulary of Basic and General Terms in Metrology*[1] (VIM) defines *uncertainty* as a parameter associated with the result of a measurement that characterizes the dispersion of the values that could be reasonably attributed to the measurand.

The *Guide to the Expression of Uncertainty in Measurement*[2] (GUM) is the controlling document. However, it is generic to all types of measurement devices; documents specific to a given type of measurement device provide the details. For example, ISO 5167-2 *Measurement of Fluid Flow by Means of Pressure Differential Devices Inserted in Circular Cross-Section Conduits Running Full—Part 2: Orifice Plates*[3] provides a method for calculating the measurement uncertainty for an orifice meter.

Statistical Approach

When multiple measurements are available, statistical methods can be applied to obtain a value for the measurement uncertainty. Consider the following example of multiple measurements of a temperature:

Number of measurements	10
Measured values	152.2 °C, 151.7 °C, 152.1 °C, 152.2 °C, 152.0 °C, 152.1 °C, 152.0 °C, 152.3 °C, 152.2 °C, 152.2 °C
Mean value	152.1 °C
Standard deviation	0.17 °C
Degrees of freedom	9
Coverage factor	2.32 (95% confidence level)
Uncertainty	0.39 °C (95% confidence level)

The measurement uncertainty is computed as follows:

1. Compute the mean (152.1 °C) and standard deviation (σ = 0.17 °C).
2. With 10 measurements, there are 9 degrees of freedom. From the Student's *t*-table, the coverage factor *k* for a 95% level of confidence is 2.32 (actually, this is for 95.45%, which corresponds to 2σ for a normal distribution).
3. The measurement uncertainty is $k\sigma$ = 0.39 °C.

The measured value would be stated as 152.1 °C with a measurement uncertainty of 0.39 °C at a 95% level of confidence.

Continuous Measurements

There are a couple of problems with the statistical approach:

- For continuous measurement devices, multiple measurements are not available.
- The GUM does not view this approach favorably, mainly because no effort is expended to understand and analyze the measurement process.

Shortly we will discuss issues, such as false accepts and false rejects, associated with decisions made on measurements of *critical operating variables*. When a measured value is from multiple measurements made on one or more samples, the statistical approach can be used to establish the measurement uncertainty.

But what if one or more of the critical operating variables are continuous measurements? The most convenient approach would be to determine a measurement uncertainty for the continuous measurements that is equivalent to the measurement uncertainty established by the statistical approach. The measurement decision procedures would then be independent of whether the measurement is sampled or continuous.

Unfortunately, this proves to be a challenge. An approach will be described next, but the discussion also includes at least some of the criticisms of the approach.

The GUM Approach

The GUM provides a framework for evaluating measurement uncertainty based on a fundamental understanding of the measurement process. The following six steps summarize this approach.

1. Specify the measurand.
2. Model the measurement.

3. Quantify the contributors to uncertainty.
4. Determine the sensitivity coefficient for each contributor.
5. Combine the contributors.
6. Calculate the expanded uncertainty.

We will provide only an overview of this approach as it is applied to industrial measurement devices.

Critics have correctly observed that the resulting value for the measurement uncertainty is a calculated value that is not confirmed by subsequent testing and experimentation. One is advised to apply “reasonableness tests” to the final result and to be suspicious of very small and very large values for the uncertainty. One can even raise the issue of the uncertainty associated with the value obtained for the measurement uncertainty. Evaluating the measurement uncertainty is clearly not an exact process either.

The process should never be considered an exercise merely to get a number for the measurement uncertainty. The analysis quantifies the influence of each contributor to the overall measurement uncertainty. Knowing the influence of each contributor permits the technology pertaining to the measurement device to be enhanced. Despite its shortcomings, the GUM approach is superior to the alternatives and is clearly the direction of the future.

Measurement Model The analysis of a measurement device is based on a mathematical model that relates the measurement result y to the various inputs x_j that affect the measurement result:

$$y = f(x_1, x_2, \ldots, x_n)$$

where y = measurement result; x_j = input that affects measurement result; n = number of inputs that affect measurement result.

Clearly this model is specific to each type of measurement device. For example, ISO 5167-2[3] provides such an equation for the orifice meter.

Each input x_j is potentially a contributor to the measurement uncertainty. Any variability in an input will lead to variability in the measurement result. For an orifice plate, ISO 5167-2[3] specifies that the orifice must be circular to a certain tolerance, specifically, “no diameter shall differ by more than 0.05% from the value of the mean diameter.” This raises two issues:

- The uncertainty for the orifice diameter u_d must be quantified. The tolerances in the standard normally provide the basis for quantifying this uncertainty, but it is possible that a supplier will use tighter tolerances in the manufacturing process.
- The contribution of the uncertainty for the orifice diameter must be translated to uncertainty for the measured value of the flow. ISO 5167-2[3] provides an equation relating the flow to various quantities, one

of which is the orifice diameter. The equation permits the quantitative translation of uncertainty in the orifice diameter to uncertainty in the measured flow.

Sensitivity Coefficients The sensitivity coefficient c_j translates a small change in input x_j to a change in the measurement result y:

$$c_j = \frac{\partial y}{\partial x_j} \cong \frac{\Delta y}{\Delta x_j}$$

where c_j = sensitivity coefficient for input j.

A sensitivity coefficient can be thought of as the change in the measured value y that would result from a one-unit change in input x_j. The engineering units for the sensitivity coefficient are the engineering units for the measured value over the engineering units for the input.

The sensitivity coefficients can be evaluated either

Analytically. This involves taking the partial derivative of the measurement model. While preferred, this is not always possible, and even when possible, the resulting expression could be excessively complex.

Numerically. This is based on the finite difference approximation to the partial derivative. To perform the calculation, values must be provided for all inputs. Most measurement models are nonlinear equations, so the values provided for the inputs will affect the value calculated for the sensitivity coefficient.

Combined Standard Uncertainty For input x_j, the contribution to the uncertainty in the measured value is the product of the uncertainty u_j for that input and the sensitivity coefficient c_j for that input. The combined effect of all inputs gives the combined standard uncertainty u_c for the measured value. There are two possibilities:

Inputs are uncorrelated. The individual contributions are combined using the root mean square equation:

$$u_c = \left[\sum_{j=1}^{n} c_i^2 u_i^2 \right]^{1/2}$$

where u_c = combined standard uncertainty for measured value; u_j = uncertainty for input j.

Inputs are Correlated. Although we will not present it, an equation is available to compute the combined standard uncertainty for situations in which some degree of correlation exists between the inputs. The equation is more complex, but computationally presents no real problems. However, this equation contains another term, specifically the correlation coefficient $r(x_j, x_k)$, which characterizes the degree of correlation between input j and input k. Because we rarely have a good source for this term, it is usually estimated from the standard deviations of the two inputs.

It is tempting to make the assumption that the inputs are uncorrelated. But to the extent that this assumption is not correct, the value computed for the combined standard uncertainty will be low.

Expanded Uncertainty Measurement uncertainty is normally stated as $y \pm U$, where U is the expanded uncertainty. This interval encompasses the possible values for the measurand with a given level of confidence (typically 95%). To obtain the expanded uncertainty U, the combined standard uncertainty u_c is multiplied by the coverage factor k:

$$U = ku_c$$

where U = expanded undertainty; k = coverage factor.

The coverage factor k is obtained from the Student's t-table (extensive versions are provided in numerous reference books). To use this table requires values for the following:

Level of confidence. This is typically 95% (or 94.45% for 2σ), although 99.7% (for 3σ) is occasionally used.

Degrees of freedom ν_y *for the measurand.* This is computed from the degrees of freedom ν_j for each input using the Welch-Satterthwaite equation:

$$\nu_y = \frac{u_c^4}{\sum_{j=1}^{n} \frac{c_i^4 u_i^4}{\nu_j}}$$

For measurement devices, obtaining a value for the degrees of freedom ν_j for each input x_j is a major problem.

For a 95% confidence level, the minimum value of the coverage factor k is 2 (infinite degrees of freedom). As the number of degrees of freedom decreases, the coverage factor increases, but it is unlikely that k would exceed 3.

Degrees of Freedom When the uncertainty for an input is determined from a set of measurements, the degrees of freedom are $n - 1$, or one less than the number of measurements in the set. But for the orifice meter example used previously, the uncertainty for the orifice diameter is determined from specifications in the standard. How do we determine the degrees of freedom for this input?

One possibility is to assume that the degrees of freedom are infinite. If this is done for all inputs, the coverage factor k for a 95% confidence level would be 2, which is the lowest possible value. Making such an assumption would give the smallest possible value for the expanded uncertainty U.

The value for the uncertainty u_j for input j is not exact, which should be no surprise. Let Δu_j be the uncertainty for the uncertainty u_j. The relative uncertainty is $\Delta u_j/u_j$. The GUM provides the following equation for computing the degrees of freedom ν_j for input j from the relative uncertainty for input j:

$$\nu_j = \frac{0.5}{[\Delta u_j / u_j]^2}$$

But where do we get a value for the relative uncertainty? Unfortunately, no method other than professional judgment has been proposed for obtaining a value for the relative uncertainty $\Delta u_j/u_j$.

Current Status

I hope you now have a general idea of how to determine the measurement uncertainty for an industrial measurement device. Could you do it? Not based on only what has been presented in this rather superficial treatment of the subject. The evaluation of measurement uncertainty is a subject for an entire book.

In areas where a measurement test can be repeated to give multiple samples, a value for the measurement uncertainty now routinely accompanies test results. An informative source of information is the *A2LA Guide for the Estimation of Measurement Uncertainty in Testing*.[4]

For industrial measurement devices, progress has been much slower; most of the push is coming from the European community. The application of measurement uncertainty to industrial measurement devices is difficult, and the critics have adequately (and generally correctly) noted the problems. At this time, the manufacturers routinely state values for the traditional measures of accuracy, repeatability, etc., but not for measurement uncertainty. In the subsequent topics on specific types of measurement devices, this forces us to use the older measures in lieu of measurement uncertainty. This will change; the question is not *if*, but *when*. It is clear that measurement uncertainty is the direction of the future.

1.10. MEASUREMENT DECISION RISK

A process is operating properly if every critical operating variable is within a specified tolerance of a target. A product is acceptable if every product quality measure is within the tolerances stated in the specifications.

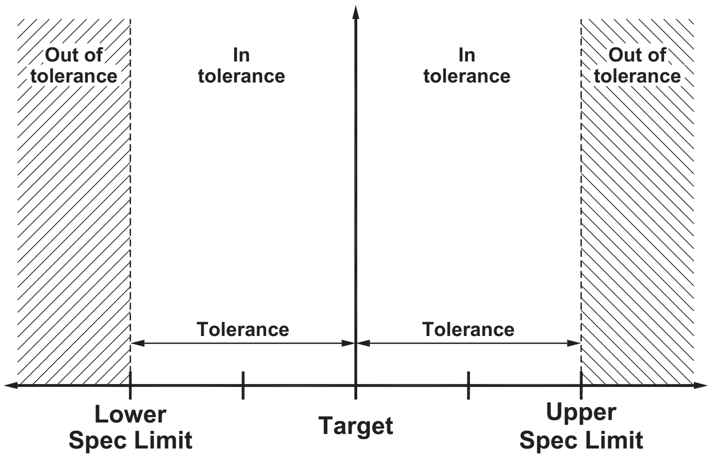

Figure 1.17. Symmetrical product tolerances.

The possibilities for process/product tolerances include the following:

Symmetrical. The same tolerance is applied to each side of the target. For example, a temperature should be 150 °C ± 10 °C. Temperatures above 160 °C or below 140 °C are unacceptable. A symmetrical tolerance is illustrated in Figure 1.17.

Nonsymmetrical. Different tolerances are applied to the two sides of the target. For example, a temperature should be 145 °C + 15 °C/–5 °C. Temperatures above 160 °C or below 140 °C are unacceptable.

One sided. A tolerance is applied to only one side of the target. For example, a temperature should be 145 °C + anything/–5 °C. Any temperature above 140 °C is acceptable.

All of these cases can be analyzed. However, we will consider only the symmetrical case.

Process/Product Distribution

Although the objective is to maintain the process or product exactly at the target, there will always be excursions from the target. The magnitude of the excursions depend on

- Magnitude of the upsets to the process.
- Performance of the process operators, control systems, etc. responsible for maintaining the process at the target.

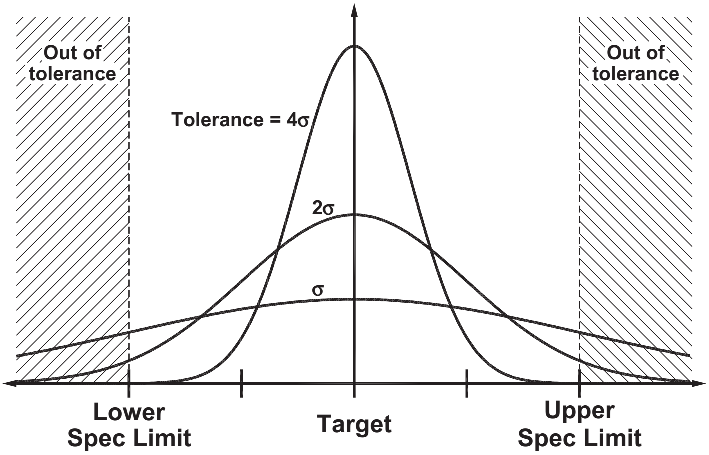

Figure 1.18. Distribution functions for tolerance equal to various multiples of the standard deviation.

The assumption is usually made that the variations are Gaussian in nature. If so, they can be characterized by their standard deviation σ. Figure 1.18 presents the distribution functions for three cases:

Tolerance equals twice the standard deviation σ. The out-of-tolerance is 4.6%. That is, the process is operating outside the acceptable limits approximately 4.6% of the time, or about 4.6% of the product is not within specifications.

Tolerance equals the standard deviation σ. The out-of-tolerance is 31.8%, which for most production operations is unacceptably large.

Tolerance equals four times the standard deviation σ. The out-of-tolerance is essentially zero.

Figure 1.18 applies to either of the following situations:

Constant standard deviation σ; variable tolerance. The process variations remain the same, and the product specifications are changed. The product specifications invariably become tighter, which means an increase in the out of tolerance for constant process variations (same process operations).

Constant tolerance; variable standard deviation σ. The product specifications remain the same, and the process variations are reduced. This case applies when automation technology is applied to provide more uniform process operations to reduce the out of tolerance.

The customary scenario is that the specifications become tighter, which provides the incentive to enhance the process and/or the controls to maintain the out of tolerance at an acceptable level.

Measurement Distribution

Figure 1.19 presents a Gaussian distribution function of the measurement results when the process is exactly at the target. This is commonly assumed unless there are specific reasons to believe otherwise.

If the process is indeed at the target, the measurement result is not necessarily equal to the target. But for the distribution function illustrated in Figure 1.19, the probability that the measurement result could be outside the tolerance limits for the process is extremely small. But when the process is close to one of the specification limits, there are four possibilities:

Process within limits; measurement result within limits. Process correctly considered to be in tolerance.

Process within limits; measurement result outside limits. Process incorrectly considered to be out of tolerance. For products, this is a *false reject*.

Process outside limits; measurement result within limits. Process incorrectly considered to be in tolerance. For products, this is a *false accept*.

Process outside limits; measurement result outside limits. Process correctly considered to be out of tolerance.

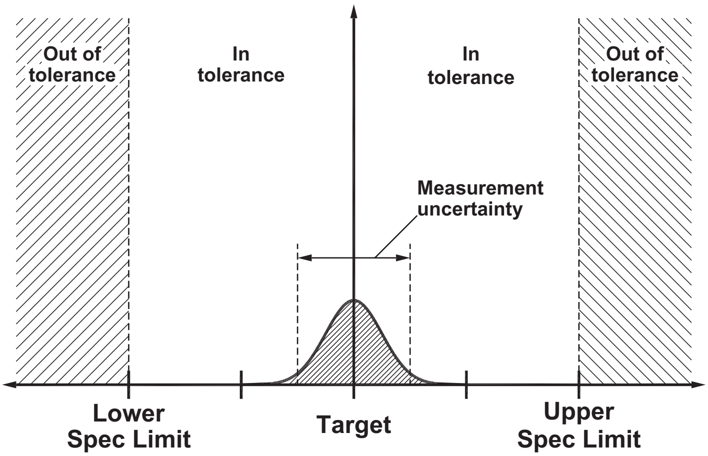

Figure 1.19. Distribution of measurement results, process operating at target.

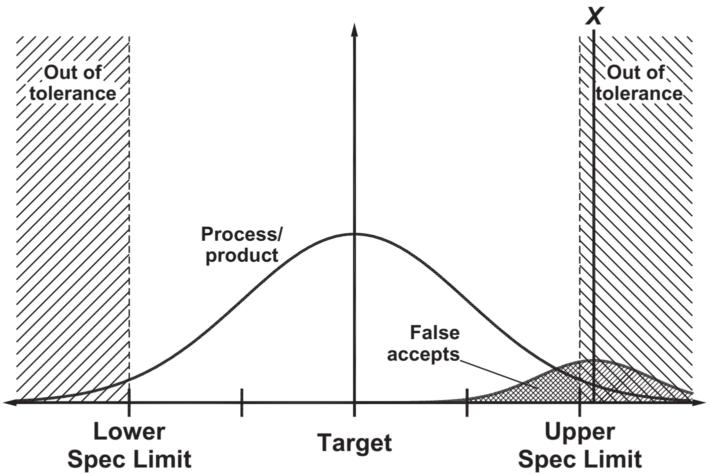

Figure 1.20. False accepts.

False Accepts Figure 1.20 depicts a situation in which the true value is slightly outside the upper specification limit. The process is actually out of tolerance. But because of the measurement uncertainty, the measurement result could be on the acceptable side of the upper specification limit. For products, this leads to what is referred to as false accepts. In industries like pharmaceuticals, false accepts are understandably a serious matter.

But before computing the level of false accepts, ask about the level of false accepts that is acceptable. The preferred answer is zero, but this is not achievable. Unfortunately, establishing a value for the acceptable level of false accepts involves many difficult issues. For a process that has been operating successfully for some time, one approach is to determine the level of false accepts with the procedures currently in use. Just knowing this number, however, can also have some side effects.

The level of false accepts depends on the following two ratios:

- The ratio of the process/product tolerance to the standard deviation for the process/product variations from target.
- The ratio of the process/product tolerance to the measurement uncertainty.

If the process/product tolerance is twice the process/product standard deviation and the process/product tolerance is four times the measurement uncertainty, the level of false accepts is approximately 0.8%. As either ratio increases, the level of false accepts decreases. However, there is a cost associated with increasing either ratio.

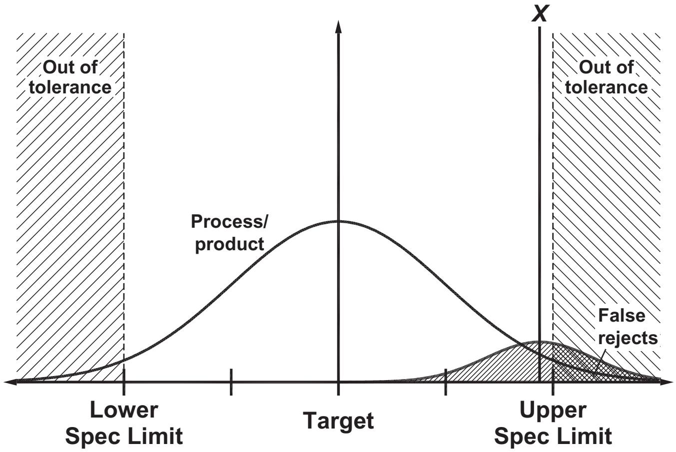

Figure 1.21. False rejects.

False Rejects Figure 1.21 depicts a situation in which the true value is slightly within the upper specification limit. The process is actually in tolerance. But because of the measurement uncertainty, the measurement result could be on the unacceptable side of the upper specification limit. For products, this leads to what is referred to as false rejects. An increase in the level of false rejects has the same consequences on process economics as a reduction in yield.

The level of false rejects also depends on the following two ratios:

- The ratio of the process/product tolerance to the standard deviation for the process/product variations from target.
- The ratio of the process/product tolerance to the measurement uncertainty.

If the process/product tolerance is twice the process/product standard deviation and the process/product tolerance is four times the measurement uncertainty, the level of false rejects is approximately 1.5%. As either ratio increases, the level of false rejects decreases. Again, there is a cost associated with increasing either ratio, but the economic benefits from reducing the level of false rejects just might offset these costs.

Accuracy Ratio

In the United States, a detailed analysis of measurement decision risk is usually avoided by relying on the accuracy ratio:

$$\text{Accuracy ratio} = \frac{\text{Process/product tolerance}}{\text{Measurement device accuracy or uncertainty}}$$

The objective of this approach is to ensure that measurement device errors are insignificant and may be ignored. The usual criterion for this is an accuracy ratio of 4:1, as in the case presented previously. For most applications, this accuracy ratio gives acceptable levels of false accepts (less than 1%) and false rejects (about 1.5%).

Figure 1.22 presents the distribution functions for process/product and measurement device for three cases:

Accuracy ratio of 8:1. The measurement device performance is better than necessary, which raises the potential for cost reductions.

Accuracy ratio of 4:1. This is the commonly desired accuracy ratio.

Accuracy ratio of 2:1. A detailed analysis of the measurement decision risk is recommended.

Guardbands

Figure 1.23 illustrates guardbanding, which is also known as *error budgeting*. The process/product specification limits are the same as before. But with guardbanding, the accept or reject decision is based on test limits instead of the specification limits. The tolerance for the test limits is smaller than the tolerance for the specification limits. The range of values between the test limit and the specification limit is the guardband. Guardbanding reduces the level of false accepts, but at the expense of increasing the level of false rejects.

Suppose a process temperature is to be 150 °C ± 10 °C. The temperature measurement device has a measurement uncertainty of 3 °C, giving an accuracy ratio of 10:3. If the level of false accepts must be 1% or less, we have three options:

- Increase the specification limits to 150 °C ± 12 °C. Such changes are more likely to be possible for process specifications than for product specifications, which are often set by the marketplace. But when the specification is on conditions within the process, this option should be put on the table. Such specifications are frequently set tighter than are really necessary.
- Install a temperature measurement device with a measurement uncertainty of 2.5 °C or less.
- Apply guardbanding. Using a guardband of 3 °C, we accept temperatures of only 150 °C ± 7 °C. Although the guardband is typically close to the measurement uncertainty, the guardband for a 1% level of false accepts can be calculated.

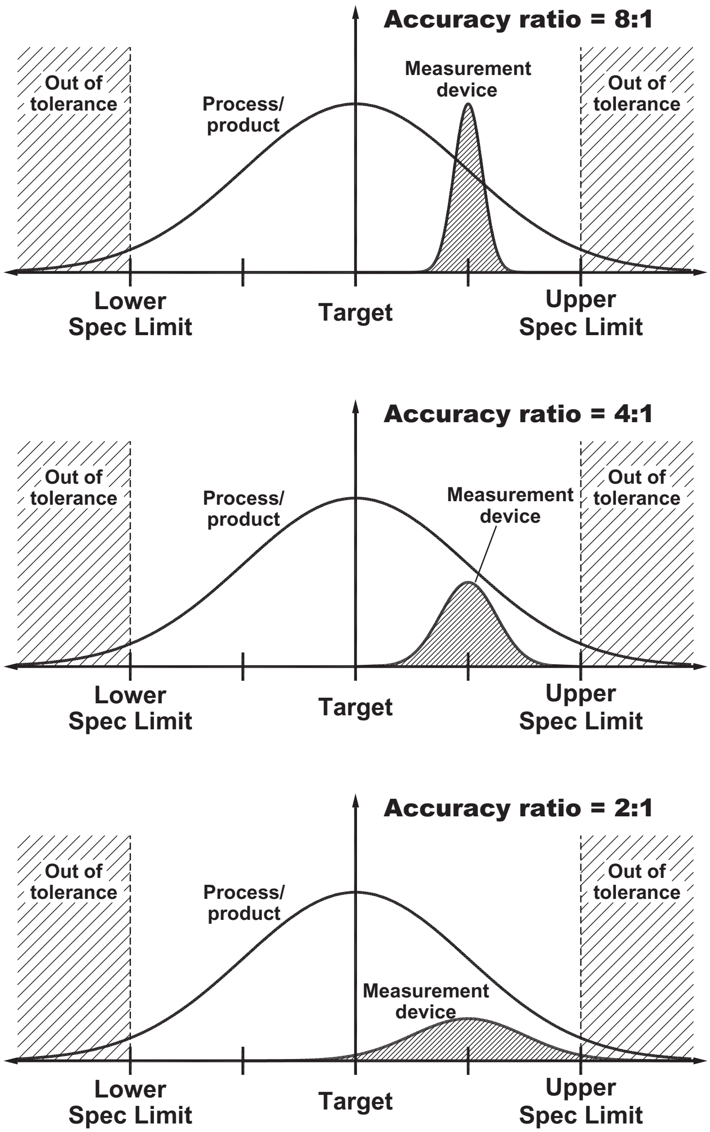

Figure 1.22. Accuracy ratio.

False Accepts with Guardbands Figure 1.24 depicts the use of a guardband when making an accept or reject decision based on a measurement with a significant uncertainty. For a true value slightly outside the acceptable range, the false accepts are also depicted. The level of false accepts is lower with the guardband (Fig. 1.24) than it is without the guardband (Fig. 1.20).

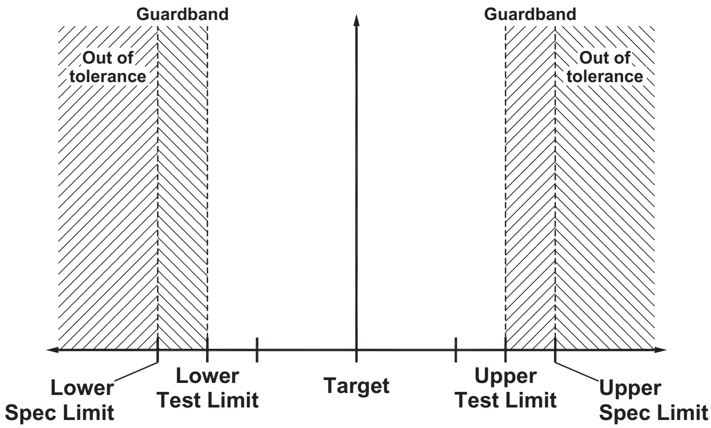

Figure 1.23. Guardbands.

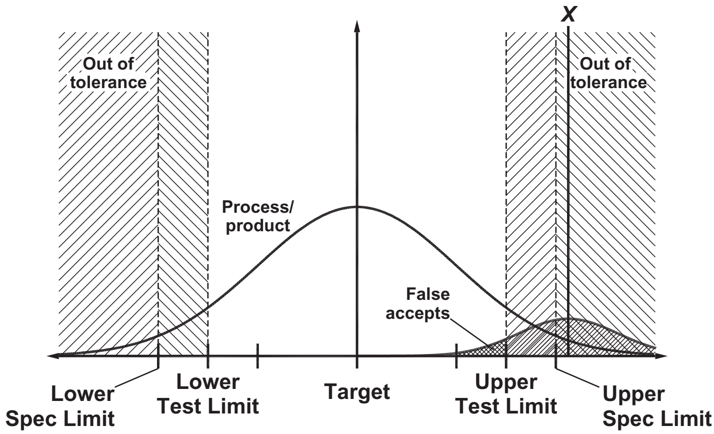

Figure 1.24. False accepts with guardbands.

Although the guardband is frequently set equal to the measurement uncertainty or accuracy, a more fundamental approach is as follows:

1. Establish an acceptable value for the level of false accepts.
2. Perform an analysis of the measurement decision risk to determine the value for the test limit that gives this level of false accepts.

As the measurement uncertainty increases, the magnitude of the guardband increases, and the test limit approaches the target. If the guardband is set equal to the measurement uncertainty, then for an accuracy ratio of 1:1 (measure-

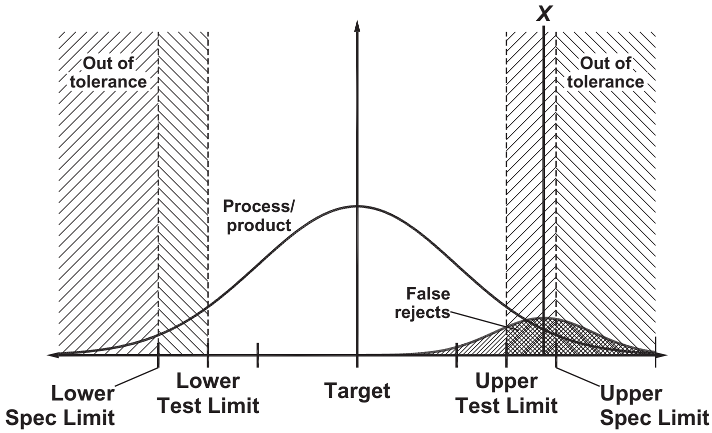

Figure 1.25. False rejects with guardbands.

ment uncertainty is equal to the process/product tolerance), the test limits are equal to the target. This situation is unacceptable.

False Rejects with Guardbands Nothing comes without a cost. The consequences of guardbands are the following:

- The level of false accepts decreases. This is the positive aspect of guardbands.
- The level of false rejects increases. This is the negative aspect of guardbands.

Increasing the level of false rejects has the same impact on production economics as a decrease in yield. This is the source of much of the controversy surrounding the use of guardbands.

Figure 1.25 depicts the use of a guardband when making an accept or reject decision based on a measurement with a significant uncertainty. For a true value that is barely within the acceptable range, the false rejects are also depicted. The level of false rejects is larger with the guardband (Fig. 1.25) than it is without the guardband (Fig. 1.21).

1.11. CALIBRATION

The responsibility of maintaining standards for various quantities (mass, length, time, etc.) rests with governmental agencies. In the United States, the agency is the NIST. Counterparts exist in Great Britain, France, Germany, and Russia.

Calibration must always assess the performance of the measurement device by comparing values from the measurement device to values indicated by the standards. The two common ways of doing this are

- Applying the measurement device to samples for which a standard value has been previously established.
- Comparing the values from the measurement device to values from another device that is known to be in close agreement with the standard values.

In some but not all cases, calibration encompasses making adjustments to the measurement device to improve the agreement between the measurement device and the standard values. Some measurement devices, such as glass stem thermometers, provide no such adjustments. In industrial measurement devices, a common practice has been to adjust the zero and/or span to provide the best agreement between the values from the measurement device and the standard values. But today, the more common alternative is simply to replace an inexpensive measurement device or to return an expensive measurement device to the supplier for recalibration and/or repair.

Calibration Lab

In years past, most calibrations were performed by instrument technicians or other local personnel. For calibrating pressure measurement devices, almost all instrument shops routinely used a dead weight pressure tester (a weight placed on a hydraulic cylinder produces a known pressure). The zero and span of the pressure transmitter were adjusted so that the values indicated by the transmitter were in agreement with the values from the dead weight test equipment.

Changes in technology are making such practices less common. As delivered, the microprocessor-based pressure transmitters provide sufficient accuracy for most process applications. When required, the manufacturer will provide a calibration certificate. The digital transmitters are also far more stable, especially in regard to zero drift. If an inexpensive device such as a pressure transmitter is suspect, replacing the device is both easier and less expensive than recalibrating it.

For those devices that do require calibration, it is usually easier to retain the services of an outside organization. Calibration laboratories provide such services for a range of measurement devices, either on a measurement device shipped to their site or by bringing their equipment to the plant site. Many suppliers of measurement devices also provide calibration services for their products. Today, calibration is not just doing the technical work; it also entails generating the necessary documentation for review by others.

Traceability

To obtain ISO 9000 certification, the measurement devices that affect product quality must be identified. These devices must be calibrated at appropriate intervals using equipment whose relationship to the national standards is known.

NIST provides calibration services, mainly to calibration labs and manufacturers of test equipment. NIST will calibrate a measurement device or will determine the measured value for an artifact, reporting both relative to their standards with appropriate documentation. This enables a test equipment supplier to provide a document that states the relationship between the measured values from one of its products to the device that it submitted to NIST for calibration.

The result is a chain of calibrations that starts with the national standard and ends with the calibration of the industrial measurement device. A typical calibration chain is as follows:

National Institute of Standards and Technology
⇑
Test meter manufacturer's calibration facility
⇑
Test meter at commercial calibration laboratory
⇑
Industrial measurement device

Traceability simply means that every link in the chain is known and documented. There can be any number of links in the chain, but obviously there is some incentive to minimize the number of links.

Uncertainty

Each link in the traceability chain adds some uncertainty to the assessment of the performance of a measurement device. The test accuracy ratio (TAR) is the ratio of the accuracy of the meter or artifact on which the calibration is based to the accuracy of the measurement device being calibrated. Traditional practice was to require a ratio of 4:1 or better.

The traditional approach is being replaced by approaches based on quantifying the uncertainty associated with each calibration in the traceability chain. IEC 17025, *General Requirements for the Competence of Testing and Calibration Laboratories*,[5] is currently the ultimate authority for calibration standards and procedures. IEC 17025 also specifies the documentation (including uncertainty data) that must accompany a calibration. Most calibration labs are now IEC 17025 accredited. Accreditation is making possible numerous mutual recognition agreements (MRAs), whereby a calibration performed by a lab in one country is recognized by other participating countries.

Check Standard

Some measurement applications are amenable to check standards; others are not. The following are two examples of check standards:

- Gas sample of known composition to be analyzed periodically by an analytical instrument.
- Sheet of known weight or thickness to be sensed by a weight or thickness gauge.

These examples are considered to be artifacts for which the expected results of the measurement are known with much less uncertainty that that associated with the measurement device. The use of a check standard permits data to be collected that can expose any deterioration of the measurement device with time.

Check standards are relatively easy to incorporate into the procedures for sampling analyzers such as gas chromatographs. To provide the check standard, a cylinder is prepared with a gas of known composition that is typical of the process samples (there are commercial organizations that specialize in supplying such cylinders). On a specified time interval, a sample from the cylinder is injected into the chromatograph in lieu of the process sample. Whether this is once an hour, once a day, or once a week has to be established based on prior experience on the specific analysis to be performed.

In sheet processing applications such as paper machines, the weight or thickness of the sheet is measured by a gauge that scans across the sheet. Between scans, the gauge can be positioned over a check standard so that the gauge can sense its weight or thickness.

1.12. MEASUREMENT DEVICE COMPONENTS

Figure 1.26 presents a schematic of the physical components of a measurement device. We want to examine the role of each.

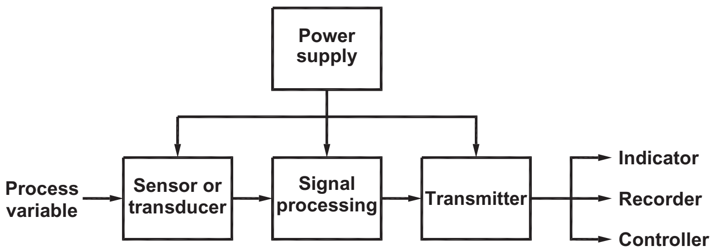

Figure 1.26. Components of a measurement device.

Sensor or Transducer

The process variable influences some characteristic of this component that can be used as the input to the signal processing components. Most rely on some basic physical principle. Examples are the following:

- Temperature affects the resistance of a noble metal such as platinum (RTD).
- The flow through an orifice affects the pressure drop across the orifice (orifice flow meter).
- The force on a displacer is the weight of the displacer less the weight of liquid being displaced (displacer level gauge).
- A conductor moving through a magnetic field generates a potential that is proportional to the velocity of the conductor and the strength of the magnetic field (magnetic flow meter).

Signal Processing

The objective of the signal processing components is to generate a signal that is linearly related to the process variable. The characteristic of the sensor that is affected by the process variable is usually not linearly related to the process variable. But because the nature of this relationship is known, the signal processing components can be designed so that the output from the signal processing components is linearly related to the process variable—that is, one function of the signal processing components is linearization.

For some measurement devices, other functions are required. One is to compensate for other variables that affect the measured value. For example, when using a thermocouple to measure process temperature, the output must be compensated for changes in the temperature of the reference junction.

Analog Transmitter

In electronic measurement devices, the output of the signal processing components is often a 0- to 1-V or a 0- to 5-V signal. If such a signal extends over any significant distance within an industrial facility, problems such as the following arise:

- The magnetic fields associated with power equipment affect the voltage levels within such signals.
- The voltage drop resulting from current flowing through the signal wires introduces errors into the signals.

Consequently, a medium that is not affected by such factors is required. For electronic signals, the medium normally chosen is current. An electronic transmitter converts the voltage signal into a current signal that is not affected by magnetic fields and resistances in the wires that carry the signal.

Digital Transmitter

The advantages of microprocessor technology for signal processing led to the development of the so-called smart transmitter. The capabilities of the microprocessor enable the signal from the basic sensor to be more accurately converted to a value in engineering units. More accurate linearization relationships are employed, and the signal from the sensor is compensated for the influence of temperature, pressure, etc.

To date, most smart transmitter products have retained the capability to transmit the measured value in analog form via a current loop. However, the use of current loops for signal transmission is gradually being replaced by digital transmission via a communications network. More on this shortly.

Power Supply

The signal processing and transmitter components of the measurement device always require a source of power. Some sensors require power, but many do not. For example, thermocouples generate a millivolt electrical potential that is a function of the difference in temperature between the two junctions. No external power is required.

The power supply is often not physically within an electronic measurement device itself. The measurement device must be located in the production area. In some facilities, this raises the issue of the presence of explosive vapors, which requires either of the following:

- Proper enclosures for the measurement device.
- Intrinsically safe designs of the electronic circuitry.

When the latter is pursued, advantages accrue from locating the power supply in a remote area where it will not be exposed to the explosive vapors.

1.13. CURRENT LOOP

At the turn of the millennium, the most common method for transmitting the measured value from the measurement device to a controller, computer, recorder, and so on, was via the 4- to 20-ma current loop illustrated in Figure 1.27. The correspondence between the current loop values and the measurement range is as follows:

$$4\,\text{ma}(0\%) \Leftrightarrow \text{Lower-range value}\,(\text{zero adjustment})$$

$$20\,\text{ma}(100\%) \Leftrightarrow \text{Upper-range value}\,(\text{zero adjustment} + \text{span adjustment})$$

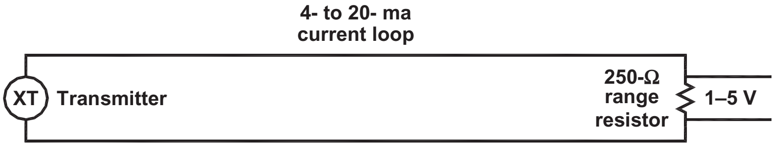

Figure 1.27. Current loop for signal transmission.

The current loop is said to provide a *live zero*—that is, an input of 0% of span is a current loop value of 4 ma. An input of 0 ma does not correspond to a valid measured value; in fact, this input means a failure of some type. It could possibly be due to loss of power or some other failure at the transmitter. It could also result from a disruption, or lack of continuity, in the circuit itself. Most input systems provide *open circuit detection* logic, which means that an input of zero is reported as an error.

Why Use Current?

In most industrial facilities, the measurement device is physically located at some distance from the controller, computer, recorder, etc. It may be less than 100 ft in small facilities, but could be over 1000 ft in large facilities such as refineries and paper mills. Two issues arise for voltage transmission:

- Some current will flow in the circuit, which results in a voltage loss due to the resistance of the wires.
- Most facilities contain power equipment, mainly electric motors, that generate magnetic fields of varying intensity. Such magnetic fields can lead to an electric potential—usually called an *induced voltage*—in electrical circuits. Good wiring practices, specifically shielding, can reduce the induced voltages to small values.

A major advantage of using a current loop for signal transmission is that the current flow is immune to these factors. Anything that influences the current flow around the circuit will affect the current flow at both the transmitter and the receiver. The transmitter must continuously monitor its current output and make the necessary adjustments to maintain the desired current flow. The performance of the current loop for signal transmission depends solely on how well the transmitter does this. The current flow at the receiver is exactly the same as the current flow at the transmitter.

Range Resistor

Current is used for signal transmission only. The internal workings of most electronic devices are based on voltages. Electronic analog (conventional) transmitters first represent the measured value as a voltage signal, which is subsequently converted to a current signal for transmission.

At the receiver end, the current signal can be converted to a voltage signal using a resistor, usually referred to as the *range resistor*. Figure 1.27 illustrates the simplest approach, which involves inserting a 250-Ω range resistor into the circuit to obtain the following conversion:

$$4\,\text{ma}(0\%) \Rightarrow 1\,\text{V}$$

$$20\,\text{ma}(100\%) \Rightarrow 5\,\text{V}$$

It is possible to have more than one range resistor in the circuit. This was common in conventional electronic systems, with one range resistor for the controller and a separate one for a display. Even some early digital systems inserted one range resistor for the input to the computer and a separate range resistor for the analog backup system. The maximum number of range resistors is determined by the maximum impedance in the current loop permitted by the transmitter. Today, most current loops contain only one range resistor—for the input to a digital system of some type.

More complex circuits can convert the current loop signal into a 0- to 5-V signal. For example, applying a –1.25-V bias to a 312.5-Ω range resistor converts 4 ma to 0 V and 20 ma to 5 V. This has the advantage of using the entire input range of the A/D converter. However, we will consider only the simple approach.

A/D Converter

As illustrated in Figure 1.28, the voltage across the range resistor is the input to an A/D converter, which converts the voltage input to a digital value. The resulting digital value is a short integer (16-bit) value that is generally referred to as the *raw value*. The input processing hardware and associated software direct the A/D converter to convert the input voltage to a digital value and store the result in the proper location in the *raw value table*.

The two key parameters for the A/D converter are

- Input voltage range.
- Resolution.

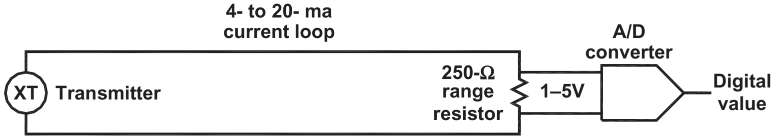

Figure 1.28. Current loop with A/D converter.

Other specifications apply, such as the time interval between conversions. But because most process applications are very slow, these are rarely an issue.

Input Voltage Range

For process applications, most A/D converters are 5-V converters. However, there are a couple of details that require explanation. Most current loop transmitters can output values somewhat over 20 ma. The exact overrange varies from one manufacturer to another. Most are on the order of 5% of the 16-ma span, giving a maximum value of 20.8 ma, or 5.2 V. Some 5-V A/D converters will process input values up to 5.12 V, which is a convenient value ($512 = 2^9$) for binary numbering systems. Although this may not accommodate the entire overrange available from the transmitter, it is sufficient.

A/D converters may be either unipolar (positive voltages only, or 0 to +5 V) or bipolar (positive and negative voltages, or –5 to +5 V). For current loop inputs, a unipolar A/D converter is sufficient. However, some systems will accept direct thermocouple inputs (after amplification). Because signals from thermocouples may be positive or negative, a bipolar A/D converter is required.

Resolution

Because most A/D converters generate binary numbers, the resolution is specified as the number of significant bits in the binary value. However, there is one subtle aspect that must be understood. For example, a 12-bit A/D converter provides a resolution of 1 part in $2^{12} = 4096$. This is applied to the input voltage range, which depends on whether the converter is unipolar or bipolar:

Unipolar, 12-bit converter. Input range is 0 to +5.12 V; resolution is 5.12/4096 = 1.25 mv.

Bipolar, 12-bit converter. Input range is –5.12 to +5.12 V; resolution is 10.24/4096 = 2.5 mv

In effect, a 12-bit bipolar A/D converter provides 11 data bits plus a sign bit, which is sometimes explicitly stated as *11 data bits plus sign*.

For process applications, most manufacturers choose A/D converters with 11 ($2^{11} = 2048$) or 12 ($2^{12} = 4096$) data bits. It is possible to obtain converters with 14 ($2^{14} = 16384$) or 15 ($2^{15} = 32768$) data bits. But unless the source of the input signal is extremely stable, the additional bits will be mostly noise.

Effective Resolution

If the input range is 0 to 5.12 V, an 11-bit unipolar A/D converter provides a resolution of 2.5 mv, which is 0.0488% of the input range. But when used for current loop inputs as in Figure 1.28, the resolution should really be expressed as a percent of the usable input range, which is 1 to 5V for the configuration in the figure.

For current loop inputs configured as in Figure 1.28, using an 11-bit A/D converter with an input range of 0 to 5.12V provides the following performance:

A/D converter input range	0 to 5.12V
A/D converter resolution	11 bits (2.5 mv or 1 part in 2048)
Resolution for input range	2.5 mv or 0.05% of a 0- to 5-V input range
A/D converter raw value range	0 to 2000 for a 0- to 5-volt input range
Effective input range	1 to 5V
Effective resolution	2.5 mv or 0.0625% of a 1- to 5-V input range
Effective raw value range	400 to 2000 for a 1- to 5-V input range
Effective resolution in bits	1 part in 1600 or 10.6 bits ($\log_2 1600 = 10.6$)

For a temperature transmitter with a measurement range of 50° to 250°F, the resolution in temperature units is 0.0625% of 200°F or 0.125°F. The repeatability of measurement devices such as coriolis flowmeters and RTDs can be 0.0625% or better, and thus the argument for A/D converters with 12 data bits. Especially in applications such as laboratories and pilot plants where the quality of the data is paramount, the resolution of the A/D converter deserves some attention.

Amplifiers and Multiplexers

Most input systems contain a few components in addition to those illustrated in Figure 1.28. To provide a stable input to the A/D converter, an amplifier is usually inserted between the voltage inputs from the range resistor and the A/D converter. For current loops, the amplifier does not alter the voltage level. But when used in applications such as direct thermocouple inputs, the amplifier increases the voltage level from the thermocouple's millivolt level to the 5-V range required by the A/D converter.

Most analog input cards provide multiple analog inputs, the number typically being in the range of 4 to 16 (computer people do like those powers of 2!). Some cards provide an A/D converter for each input; others use a multiplexer to switch the inputs onto a single A/D converter. This decision affects the input scan software provided by the manufacturer, but otherwise there should be no effect on users of the system.

1.14. POWER SUPPLY AND WIRING

There is one very important component missing from the current loop shown in Figure 1.28: a source of power. To date, this seems to require two strips of copper—that is, wires—between the source of power and the transmitter. Perhaps battery-powered transmitters will someday become the norm, but so far there has been little use of such transmitters.

Depending on the power requirements of the transmitter, the wiring configuration will be one of the following:

- Two wire.
- Three wire.
- Four wire.

Two-Wire Transmitters

As illustrated in Figure 1.29, in two-wire transmitters the data transmission and the power share the same two physical wires. The source of power may be integrated into the input-processing card, or may be provided by a separate power supply (as shown).

The power supply is typically a 24-V DC source of power. Suppose the current flow is at the lower range value of 4 ma. If the full 24-V were taken across the transmitter, the available power at the transmitter would be

$$\text{Power} = (4\,\text{ma})(24\,\text{V}) = 0.096\,\text{W}$$

The practical limit is somewhat less, but this is adequate for low-power measurement devices such as those for temperature and pressure.

When digital transmission is used, the digital transmission can also share the same two physical wires that are used to power the transmitter. The requirement for power reduces the attractiveness of technologies such as fiber optics and wireless. As long as you have to install physical wires to the transmitter for power, the possibility exists to use this same pair of wires for data transmission as well.

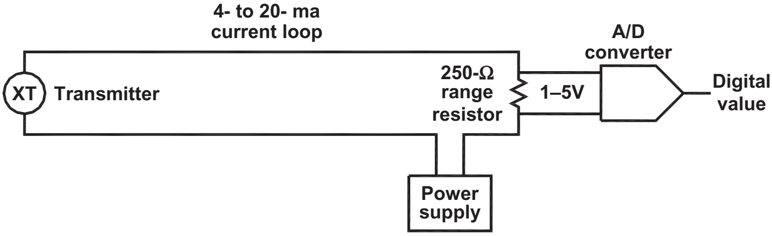

Figure 1.29. Two-wire transmitter.

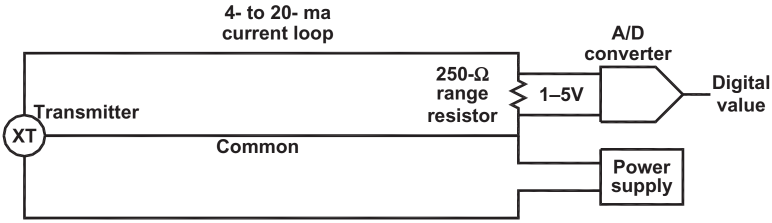

Figure 1.30. Three-wire transmitter.

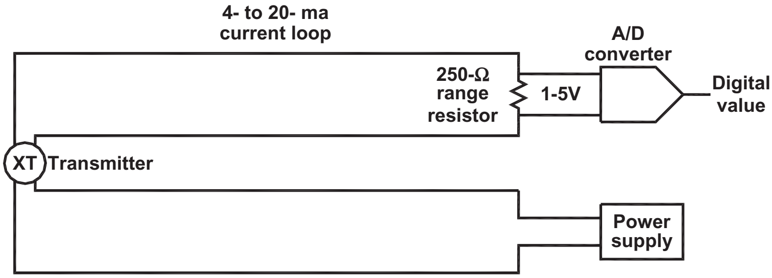

Figure 1.31. Four-wire transmitter.

Three-Wire Transmitters

In the three-wire transmitter shown in Figure 1.30, separate circuits are provided for the current loop and the power supply, but they share a common return, which is normally at ground potential. The power supply is always a DC supply that is separate from the input card. This configuration is typically used for transmitters that require a moderate amount of power.

Four-Wire Transmitters

Figure 1.31, illustrates a four-wire transmitter with separate circuits for the current loop and the power supply. There are no common elements between the two circuits. The power supply is usually AC, but could be DC (if DC, the three-wire configuration can usually be used). The power supply is always separate from the input card. There is no limit on the amount of power that can be supplied to the transmitter using this configuration. It is the choice for all transmitters with high power requirements and for all transmitters that require AC power.

1.15. SERIAL COMMUNICATIONS

Figure 1.32 illustrates using serial communications to transfer information from the transmitter to a digital system (such as a control system). The trans-

Figure 1.32. Signal transmission via serial communications.

mitter and the digital system are equipped with serial communications cards that provide one or more "ports" for serial interfaces. When the connection, or *link*, is from a port on the transmitter to a port on the digital system, the communications structure is said to be *point to point*.

The communications cards on a given link must be compatible. The common types are

RS-232. Most common, but limited to short distances (50 ft) and provides no electrical isolation between transmitter and receiver—that is, both must have same ground, which is a problem for industrial installations.

RS-485. Capable of longer distances and provides electrical isolation so that a common ground is not required. Consequently, RS-485 is more suitable for industrial installations.

If compatible cards are not available, modules that convert between RS-232 and RS-485 can be purchased.

Protocol

RS-232 and RS-485 are hardware standards only. That is, they specify the number of wires, the use of each wire, the voltage levels, etc. This permits binary information to be transmitted from the transmitter to the digital system but says nothing about the content. The protocol specifies the content of the transmission. For example, if the transmitter is to transmit a value to the digital system, the protocol specifies the exact sequence of characters to be transmitted.

A transmission could be entirely text information, such as the character string *127.8* followed by a carriage return and line feed (the *terminator*). Some transmitters can be configured for *continuous transmission*, by which they transmit the current value to the digital system on a fixed time interval. However, this is a very simple protocol, is one-way only (transmitter to digital system), can be used only in point-to-point configurations, and does not use the full potential of the serial communications interface. Therefore, more elaborate protocols are often used.

Software

A software module is required to convert the string of characters or bytes received from the transmitter to a numerical value for the measured variable.

Developing such software modules is complicated by the fact that there is no effective standard for the protocol. This has led some manufacturers to develop their own specifications for the protocol, resulting in a so-called proprietary protocol.

With some experience, such software modules can be developed quite quickly. However, experience has shown that these modules lead to support problems. Suppose one upgrades either the transmitter side of the link or digital system side of the link. Will the software module for the protocol still work? Usually it does, but occasionally it does not. Manufacturers have also been known to change the protocol during the life of a product. This would mean that the transmitters now being purchased use a different protocol from those previously installed. Basically, the software module for the protocol is a constant source of potential problems.

In industrial applications, the software module is commonly called a *driver*. However, this use of the term is not entirely consistent with its use in the context of computer operating systems.

Modbus

Developed in the 1970s by Modicon (a manufacturer of PLCs) for transferring information between PLCs or between a PLC and a host computer, Modbus is about the nearest thing to a standard for a protocol.[6] But even it is subject to extensions that, if used, will require customization of the software module. The original Modbus protocol is summarized as follows:

Byte	Contents
1	Start of transmission (STX) (02 hex)
2	Function code
3	Number bytes of user data (N)
4 to $N + 3$	User data (depends on function code)
$N + 4$	End of transmission (ETX) (03 hex)
$N + 5$ to $N + 6$	Cyclical redundancy check (CRC)

The second byte in the Modbus protocol is a function code. Modicon defined function codes for PLC applications, such as reading registers and writing registers. Some transmitters use these function codes. However, Modicon left a number of function codes undefined so that other users of the protocol could extend it. Some manufacturers of measurement devices have chosen to do so, often for good reason (measurement devices are quite different from PLCs). If the manufacturer has extended the Modbus protocol by defining function codes specific to its product, a standard Modbus software module would not support such function codes.

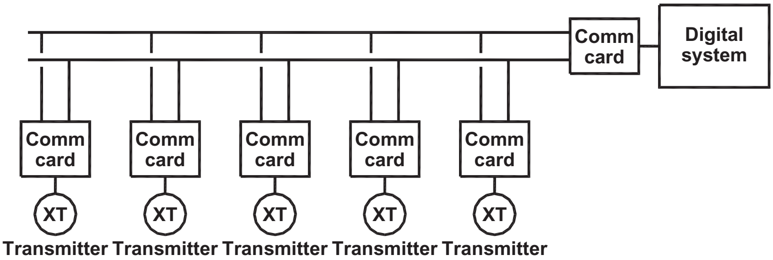

Figure 1.33. Multidropping.

Multidropping

Using point-to-point serial links would require the digital system to provide a serial port for each transmitter. This could lead to a large number of serial ports. For some computers, large numbers of serial ports are readily available, but others have severe limitations. The alternative is to use the multidropped arrangement illustrated in Figure 1.33.

To do so requires RS-485 communication cards for the hardware and a software module that can *poll* the individual transmitters. Each transmitter is assigned an address. The communications protocol must provide a field for the address of the transmitter (note that the standard Modbus protocol does not, although one could be easily added). To obtain the current value of the measured variable from a transmitter, the software in the digital system must issue a *read current value* request to the appropriate transmitter, which responds with the current value of its measured variable. This results in a slower scan rate for the measured variable, but 1-sec scans can normally be achieved with multidropped configurations containing eight or fewer transmitters.

Mass Flow Meters

Serial communications is not the interface of choice for most transmitters. However, suppose a measurement device produces more than one measured value. A good example of such a measurement device is a mass flow meter. The following measured values are available:

- Mass flow.
- Temperature.
- Density.
- Viscosity (if this option is purchased).

If a serial interface is used, all four values could be retrieved over the serial link. Using current loop technology, four current loops would be required. To make matters more complicated, one commercial supplier of mass flow meters provided only two current loop outputs in its products. The unit could be configured to transmit any of the listed values over a given current loop. However, the net result is that only two of the values could be retrieved using current loops.

Composition Analyzers

Composition analyzers such as chromatographs are sampling measurement devices. An analysis is generated at discrete intervals of time, depending on the characteristics of the separation column within the chromatograph. Furthermore, each analysis usually consists of multiple values, one for each component. In addition, the time that the sample was injected is also of interest.

Current loop interfaces do not handle such analyses very well. Occasionally the analysis can be reduced to the ratio of two key components, which can be transmitted using a current loop. But even in such situations, the time of the last analysis cannot be captured this way.

Weight Transmitters

Load cell technology is capable of resolutions far beyond what is practical with current loops. A 15-bit (data) A/D converter provides a resolution of 1 part in 32,768. Load cells are available with five-digit local displays, which provides a potential resolution of 1 part in 100,000. For example, the load cell that can display a weight from 0.0 to 6,000.0 kg with a five-digit display has a display resolution of 1 part in 60,000. Such a resolution cannot be achieved with current loop technology.

Suppose a 12-bit A/D converter with an input range of 0 to 5.12 V is used to process a current loop input from a load cell that displays weights up to 6,000 kg to a resolution of 0.1 kg (a five-digit display). Assuming a 250-Ω range resistor is inserted into the current loop circuit, a weight of 0.0 kg gives a raw value of 800. A weight of 6,000.0 gives a raw value of 4,000. A change of 1 count in the raw value corresponds to a change in weight of 6,000.0/3,200 = 1.9 kg. Whereas the load cell is capable of indicating a weight change of 0.1 kg, the smallest weight change that can be detected from the input signal is a change of 1.9 kg. Consequently, the value displayed by the digital system rarely agrees with the value in the local display of the weight transmitter.

1.16. SMART TRANSMITTERS

The marketing types have embraced the term *smart* to the extent that everything is now advertised as "smart"—even consumer products such as ovens and washing machines. The common denominator seems to be that a microprocessor has been somehow incorporated into the product. The predecessors to the smart transmitters were electronic analog transmitters. These are now usually referred to as *dumb* transmitters; obviously, nobody wants these anymore.

The microprocessor does indeed provide the opportunity for designers to incorporate some very useful features into the measurement device. The following list summarizes some of the possibilities:

- Checks on the internal electronics.
- Checks on environmental conditions, such as temperature, within the measurement device.
- Compensation of the measured value for conditions within the measurement device.
- Compensation of the measured value for other process conditions.
- Linearizing the output of the transmitter.
- Configuring the transmitter from a remote location.
- Automatic recalibration of the transmitter.

These features do indeed give the smart transmitters clear advantages over their predecessors.

Although not an explicit feature, a major advantage of smart transmitters is their versatility. Using the same model of a measurement device in multiple locations reduces spare parts requirements. If the requirements of the application change, a more versatile measurement device is likely to accommodate such changes. This is a major advantage in laboratory and pilot plant applications.

Remote Configuration

A smart transmitter can be configured remotely—that is, it is not necessary to go to the physical location of the transmitter, open the enclosure, and make adjustments. There are three options:

- A handheld battery-powered microprocessor-based configuration unit. Instrument technicians seem to like these and become very adept with them, but you may be surprised at their cost.
- Centralized device management software that maintains a database of the configurations. Usually this is integrated with the configuration software for the control system.
- PC-based device configuration software. Sometimes, but not always, the device management software part of the control system configuration software can be extracted and used in a stand-alone manner.

Remote configuration definitely facilitates maintenance—replace the measurement device and then load the configuration from either the handheld unit or the central database. This can be done rapidly and free of errors (provided the correct configuration data set is downloaded into the replacement).

One of the issues with remote configuration tools is that they should support products from multiple vendors. This is accomplished via device descriptions or profile descriptions, the most common being the following:

- Electronic device descriptions (EDDs).
- Profibus-process automation (profibus-PA) profile description.
- Highway Addressable Remote Transmitter (HART) device description.

The supplier of the device management software provides the descriptions available at the time the software was delivered. For later products, the manufacturer of the measurement device provides the description in the form of a software module that can be downloaded via the Internet.

HART can be used to communicate with devices connected via a 4- to 20-ma current loop. A high-frequency signal (1200 Hz represents binary 1; 2200 Hz represents binary 0) with an amplitude of 0.5 V or less is superimposed on the same two wires used for the current loop. A high-pass filter separates the communications signal; a low-pass filter separates the 4- to 20-ma signal.

Accuracy

One area that smart transmitters deliver better performance then do older models is in regard to accuracy. There are two aspects of this:

- The initial accuracy specifications are normally superior.
- The drift (change of parameters with time) is less.

Although complete automatic recalibration is rarely provided, many smart transmitters can automatically reset their zero adjustment. In conventional transmitters, zero drift contributed to a deterioration of the accuracy with time. When the zero can be reset automatically, this source of error is greatly reduced if not completely eliminated.

Interfaces

Being digital in nature, smart transmitters provide three options for interfacing with computers, control systems, etc.:

- Current loop.
- Network interface (usually referred to as *fieldbus*).
- Serial communications interface.

Almost all smart transmitters have current loop and fieldbus interfaces; the serial communications interface is less common. The current loop interface is

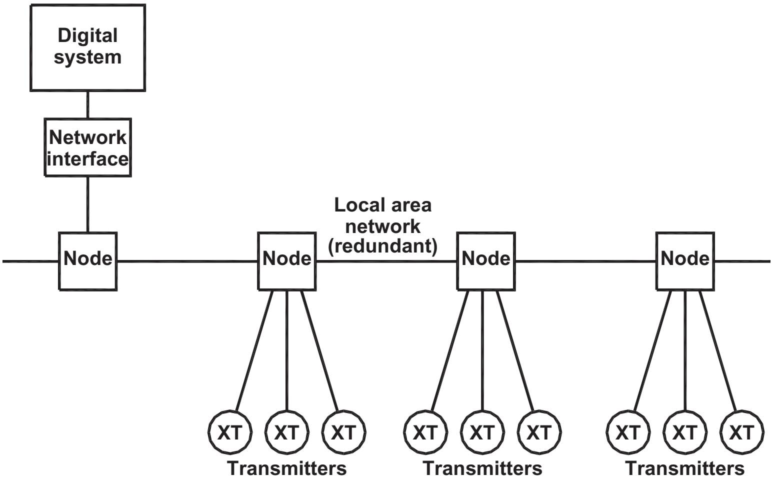

Figure 1.34. Fieldbus.

expected to gradually disappear as fieldbus networks become more widely installed.

Fieldbus

Fieldbus is basically a local area network (LAN) designed for communicating between field devices, including but not limited to smart transmitters, and digital systems such as controllers. Fieldbus uses LAN technology for communications and thus is flexible, two-way, capable of reading multiple values from a single transmitter, and so on.

Figure 1.34 presents a simplified structure for fieldbus. LAN technology is used to transfer information between the various nodes on the fieldbus network. The master node interfaces to the digital system, which provides data acquisition, process control, or other functions. The other nodes are slave nodes that communicate with the individual field devices, such as the transmitters.

Although most manufacturers seem to focus on reduced wiring costs as the major incentive for fieldbus, the interface's capabilities and flexibility prove far more advantageous. As compared to current loop installations, the initial commissioning of the system is much easier and can be accomplished in less time. Thereafter, addition of a new measurement device is greatly simplified.

In the process industries, fieldbus got off to a slow start. As is customary with digital technology, several designs for fieldbus appeared, all of which could meet the needs of process applications. The delays largely revolved around standards (or lack thereof) for the communications network:

- Adopting a standard for the protocol became embroiled in international politics. Although a standard is now available,[7] it is in reality *five* standards rolled into one document. That is, the standard provides five alternatives for fieldbus implementations. The marketplace will ultimately determine the winner.

- Integrating digital transmission into control systems required substantial systems efforts on the part of control system suppliers. Many suppliers were naturally reluctant to make such an investment until the standards issue was resolved.
- Without a standard, users were hesitant to install equipment that might be incompatible with the eventual standard.

The following process installations are encountered most frequently:

- Profibus, which appears to be prevailing in Europe.
- Foundation fieldbus, which appears to be prevailing in the United States.

Most manufacturers of measurement devices can supply a model for either of these (in addition to a model for current loop installations and possibly a model for serial communications).

1.17. ENVIRONMENTAL ISSUES

Environmental issues pertain to the electronics associated with the measurement device. These are not directly exposed to the process materials but instead are exposed to the ambient conditions that surround the location of the measurement device. The following must always be considered:

Temperature. This is the temperature at the location in which the electronics are physically located.

Atmosphere. The atmosphere surrounding the electronics can raise a number of issues.

Procedures. Let's illustrate with an example. In food-processing facilities, a crew regularly cleans up the area, which usually involves high-pressure water hoses or the like that could be directed at the measurement device.

Ambient Temperature

Ambient temperature applies to the signal processing and transmitter components of the measurement device. The manufacturer of a measurement device will state actual values, but typical limits are $-25°$ to $+80°C$. Sometimes the term *noncondensing* is appended; electronic equipment performs better when dry.

The physical location of the sensor or transducer is determined by the process variable that is the subject of the measurement. But for the components that provide the signal processing and the transmitter, the options are

Integrally mounted. The signal processing and the transmitter are within the same physical enclosure as the sensor or transducer.

Remotely mounted. The signal processing and the transmitter are in an enclosure that is separated by some distance from the physical location of the sensor or transducer. In most situations, this entails some distance of low-level (millivolt) signal wiring, which raises shielding, grounding, and other such issues.

The choice is often dictated by the process temperature at the location in which the measurement is required. Where the process temperature is elevated, active cooling can be considered. One approach is to provide a constant purge of instrument air through the enclosure containing the transmitter. A more extreme approach is to provide a water jacket around the enclosure. The downside of active cooling is that the air purge or water flow through the jacket must be maintained at all times. This always proves to be more difficult than one initially suspects; even a brief loss of air purge or water flow usually has adverse consequences.

Enclosures

Enclosures provide protection. The electronics within a measurement device are almost always protected from various elements of the environment by an appropriate enclosure. The requirements depend on the application, but the common requirements include the following:

Solids. Protect electronic equipment from foreign objects (or protect foreign objects such as hands and fingers from electronic equipment).

Liquids. Keep electronic equipment dry.

Corrosion. Protect electronic equipment from corrosive liquids or gases.

Explosive atmospheres. Either keep explosive atmospheres away from sources of ignition (purged enclosures) or contain the explosion in case of ignition (explosion-proof enclosures).

The manufacturer of a measurement device typically provides the enclosure, but it is the user's responsibility to make sure that the enclosure is appropriate to the environment within his or her facility.

The enclosure is accompanied by documentation that certifies its suitability for certain environments. These certifications are based on two standards:

NEMA Standard 250, a U.S. standard.

IEC 60529, a European standard.

Most other countries either use one of these directly or have a standard derived from one of these. Most manufacturers of enclosures can provide certifications appropriate for both standards.

NEMA Standard 250 The National Electrical Manufacturers Association (NEMA) is the U.S. manufacturer's organization that develops and maintains various standards. Also included are the test methods used by Underwriters Laboratories (UL) to verify that a product conforms to the standard.

NEMA Standard 250, *Enclosures for Electrical Equipment*,[8] defines 13 enclosure "types," some of which have letters appended to the type number to specify a variation of the basic type. Each type is briefly described as follows:

Type	Location	Brief Description
1	Indoor	General purpose
2	Indoor	Drip tight
3	Outdoor	Rain tight, sleet resistant
3R	Outdoor	Type 3 and dust tight
3S	Outdoor	Type 3R with external mechanism if ice laden
3X	Outdoor	Type 3 and corrosion resistant
3RX	Outdoor	Type 3R and corrosion resistant
3SX	Outdoor	Type 3S and corrosion resistant
4	Both	Watertight, dust tight
4X	Both	Type 4 and corrosion resistant
5	Indoor	Dust tight
6	Both	Submersible, watertight, and dust tight
6P	Both	Type 6 for prolonged submersions
7	Indoor	Explosion-proof (Class I, Groups A, B, C, D)
8	Both	Oil-immersed equipment (Class I, Groups A, B, C, D)
9	Indoor	Explosion-proof (Class I, Groups E, F, G)
10	Mines	Explosion-proof in methane or natural gas
11	Indoor	Corrosion resistant and dust tight, oil immersed
12	Indoor	Dust tight and drip tight
12K	Indoor	Type 12 with knockouts
13	Indoor	Oil tight and dust tight

This is intended only as a general guide. Consult the latest version of the standard for current and detailed information on NEMA types for enclosures.

For industrial measurement devices, enclosures generally have a NEMA 4X classification. Although this enclosure is corrosion resistant, the standard says nothing about what corrosive materials are resisted. NEMA 4X enclosures do not provide adequate protection in locations where explosive atmospheres may be present (see "Hazardous Area" in section 1.18).

IEC 60529 The International Electrotechnical Commission (IEC) develops and maintains international standards for electrical equipment and related technology. Through the European Commission on Electrotechnical Standardization (CENELEC), these standards are almost universally used in Europe.

IEC 60529, *Degrees of Protection Provided by Enclosures*,[9] defines ingress protection (IP) codes that are applied to enclosures. The format for the codes is IPxy, where x is a digit in the range 0 through 6 that pertains to protection from solid objects; y is a digit in the range 0 through 8 that pertains to protection against liquids. The following is a very brief description of the value of each digit:

Digit x	Protection from Solid Objects
0	No protection
1	Objects > 50 mm (hands)
2	Objects > 12 mm (fingers)
3	Objects > 2.5 mm (wires)
4	Objects > 1 mm (thin wires)
5	Dust (limited ingress)
6	Dust (totally protected)

Digit y	Protection from liquids
0	No protection
1	Vertically falling droplets
2	Sprays up to 15° from vertical
3	Sprays up to 60° from vertical
4	Sprays from all directions
5	Low-pressure jets of water
6	High-pressure jets of water
7	Low depth, brief immersion
8	Prolonged immersions

These lists are intended only as a general guide. Consult the latest version of the standard for current and detailed information on IP codes for enclosures.

For industrial measurement devices, enclosures generally have either an IP66 or an IP67 classification.

1.18. EXPLOSIVE ATMOSPHERES

For a flammable gas or vapor in air, the requirements for a fire or explosion are as follows:

- Concentration above the lower explosive limit (LEL).
- Concentration below the upper explosive limit (UEL).
- Source of ignition (spark, hot surface, etc).

The LEL and UEL depend on the specific component (or mixture of components) present in the air. The source of ignition also depends on the amount of energy needed to ignite the vapors.

Dust, fibers, and many other combustible solid materials in air can cause explosions that are more severe than flammable gases or vapors.

Hazardous Area

Any area in which flammable gases or vapors may be present continuously under normal operations, intermittently under normal operations, or only under abnormal operations is subject to a hazardous area classification. There are two options for hazardous area classifications:

- Class, division, and group as per NEC Article 500.
- Zone as per IEC 60079 or NEC Article 505.

Before any electrical equipment, including measurement devices, can be specified, the hazardous area classification must be established. The responsibility of the person specifying a measurement device for a given application is that the specifications be appropriate for the previously established classification of the hazardous area.

We will give only a brief introduction. The respective standards are the final word; but many of the suppliers also provide very informative and readable literature on the subject (for example, R. Stahl offers *Basics of Explosion Protection*[10] as a download from its website).

NEC Article 500 Developed by the National Fire Protection Agency (NFPA), the National Electric Code (NEC) is the governing standard on electrical safety within the United States. Article 500 pertains to the classification of hazardous locations,[11] and is summarized as follows:

Class. Pertains to the nature of the combustible material.

Class I. Flammable gases or vapors.

Class II. Combustible dust.

Class III. Easily ignitable fibers.

Division. Pertains to the frequency at which the combustible material is present.

Division 1. Present during normal operations.

Division 2. Not normally present.

Group. Pertains to the combustible material.

Group A. Acetylene.

Group B. Hydrogen, ethylene oxide, etc.

Group C. Ethyl ether, ethylene, etc.

Group D. Gasoline, hexane, etc.

Group E. Metal dust.

Group F. Carbon black, coal dust, etc.

Group G. Flour, starch, etc.

Division 2 is less demanding than Division 1. When functioning properly, most measurement devices (there are exceptions) would not provide a source of ignition—that is, they do not spark, they have no hot spots, etc. Even if exposed to an explosive atmosphere, there would be no detonation. For Division 2, the detonation would occur only if two failures occurred simultaneously:

- Explosive vapors are present (infrequent for the Division 2 classification).
- The measurement device fails (also infrequent).

For Division 1, the assumption is basically made that explosive vapors will be present at the time the measurement device fails. Therefore, either an appropriate enclosure must be provided for the measurement device or the measurement device must be designed so that no failure can lead to a spark or other source of ignition (intrinsically safe designs, to be discussed later).

IEC 60079 Developed by the IEC, standard 60079 is the governing standard on classification of hazardous areas used throughout Europe.[12] This standard introduced the concept of zones:

Zone 0. Flammable gases or vapors very frequently present or continuously present, even under normal operations.

Zone 1. Flammable gases or vapors occasionally present under normal operations.

Zone 2. Flammable gases or vapors present infrequently or for short periods (such as during abnormal operations).

Zones 20, 21, 22 are the corresponding zones for dusts.

Beginning in 1996, similar classifications were incorporated into the NEC, eventually evolving to NEC Article 505. While very similar, these two classifications are not identical. The general relationship between zones and divisions is typically considered to be as follows:

- Division 1 of NEC Standard 500 encompasses Zones 0 and 1 of IEC 60079/NEC Article 505.
- Division 2 of NEC Standard 500 is equivalent to Zone 2 of IEC 60079/NEC Article 505.

The splitting of Division 1 into two zones reflects the special issues that arise when explosive vapors are continuously present. When explosive vapors are present intermittently (Zone 1), the area can be monitored with a gas sniffer to make sure that no explosive conditions exist, which permits work to be performed on electrical equipment mounted within enclosures. But because explosive vapors are considered to be continuously present for Zone 0, this means that electrical equipment mounted within enclosures must be powered down to perform work. This usually means a plant shutdown.

Explosion-Proof Enclosures

Enclosures have been designed to contain any explosions. Flammable gases and vapors can enter such an enclosure, but should the mixture be explosive and be ignited by the electrical equipment inside, the enclosure is designed to contain the explosion in such a manner that it is not an ignition source for an external explosion. The surfaces of the enclosure must remain cool, and all escaping gases must be sufficiently cooled. Explosion-proof enclosures tend to be heavy.

The advantages and disadvantages of explosion-proof enclosures are summarized as follows:

- Sturdy, but expensive, enclosure.
- Well accepted in the United States, but not everywhere.
- Competent installers available, but installation cost high.
- When used in Zone 0, equipment within enclosure must be shut down for work to be performed.
- Only option for high-power electrical devices.
- Safe-area certification not required for electrical equipment.
- Cooling an issue for heat-generating electrical equipment.

From a measurement device perspective, cost often becomes an issue because the cost of the enclosure can easily exceed the cost of the measurement device itself.

Not only must the enclosures be technically appropriate but the supplier must provide the proper certifications regarding its approval by an organization such as Factory Mutual. It is then the purchaser's responsibility to make sure that the classification for the enclosure is acceptable for the hazard area classification of the location in which the measurement device and enclosure will be installed. All of this information must be retained; electrical inspectors like to look at these things.

Purged Enclosures

Flammable gases and vapors are kept away from the electrical equipment within the enclosure by pressurizing and purging. NFPA standard 496, *Purged and Pressurized Enclosures for Electrical Equipment*,[13] provides for three types:

X. Reduces classification from Division 1 to nonhazardous.
Y. Reduces classification from Division 1 to Division 2.
Z. Reduces classification from Division 2 to nonhazardous.

For type X, power is to be disconnected from the internal electrical components on loss of purge, as shown in Figure 1.35. The protective switch must be installed in the discharge line, and no valve or other restriction can be installed between the protective switch and the enclosure. Refer to the standard for additional requirements on purged enclosures.

In Europe, the requirements for pressurized enclosures are incorporated into IEC 60079 as *Part 2—Pressurized Enclosures*.[14] Actually, *pressurized* is a more appropriate term than *purged*—the instrumentation actually ensures that the pressure within the enclosure is above atmospheric, which means that any leaks are from inside to outside. The flow through the enclosure often depends on how well it is sealed.

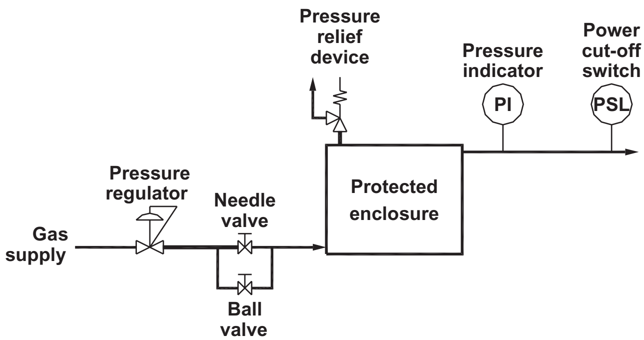

Figure 1.35. Pressurized enclosure.

Intrinsic Safety

For many applications, electrical circuits can be designed so as to be incapable of providing a spark or other source of ignition on circuit or component failures. Referred to as *intrinsically safe*, such designs eliminate the need for explosion-proof or purged enclosures.

Electrical devices are intrinsically safe only if certified as such. These certifications take two forms:

System. Elements of the circuit are approved as a system. The manufacturer submits all components of the circuit that contains the measurement device for approval. The measurement device is certified as intrinsically safe only when used in that circuit.

Entity. Individual components are approved for use in properly designed systems. This gives the user more flexibility but also imposes more responsibility on the user to get it right.

Intrinsically safe designs are possible only in low-power applications (which covers most but not all measurement devices). But low power alone does not mean intrinsically safe. The circuits must not be capable of storing sufficient energy, either via capacitance or inductance, that could lead to a spark.

To summarize, the advantages and disadvantages of intrinsically safe electronics are as follows:

- Expensive enclosures are not mandatory.
- Restricted to low-power applications.
- Maintenance can be performed without shutting off power.
- Imposes demands on engineering, especially when entity certifications are used.
- Must be installed and maintained by knowledgeable personnel.
- Excellent acceptance internationally.

Intrinsic Safety Barriers

Intrinsic safety barriers limit the current, voltage, and total energy delivered to a measurement device in a hazardous area. Two types are commercially available:

Galvanic isolators. These use transformers, optical isolators, or similar technology to meet the requirements for intrinsic safety. These require an external source of power (active devices).

Zener barriers. These use Zener diodes, resistors, and fuses to limit current and voltage. These do not require an external source of power (passive devices).

Most industrial installations use Zener barriers, and only these are described herein. There are several commercial suppliers of Zener intrinsic safety barriers.

The installation of the intrinsic safety barriers is always in the nonhazardous area. There are two possibilities:

Between the input/output (I/O) equipment and the measurement device. The measurement device must be intrinsically safe, but the I/O equipment does not (Fig. 1.36a). To date, most industrial installations have been in this fashion.

Between the communications/power supplies and the I/O equipment. The I/O equipment and all measurement devices must be intrinsically safe (Figure 1.36b). Being the leaders in the use of intrinsically safe installations, the European suppliers were the first to offer intrinsically safe I/O equipment that could be installed and serviced in the hazardous area.

Figure 1.37 illustrates the circuit for a simple Zener intrinsic safety barrier for positive polarity. The negative side of the power supply is grounded, so the

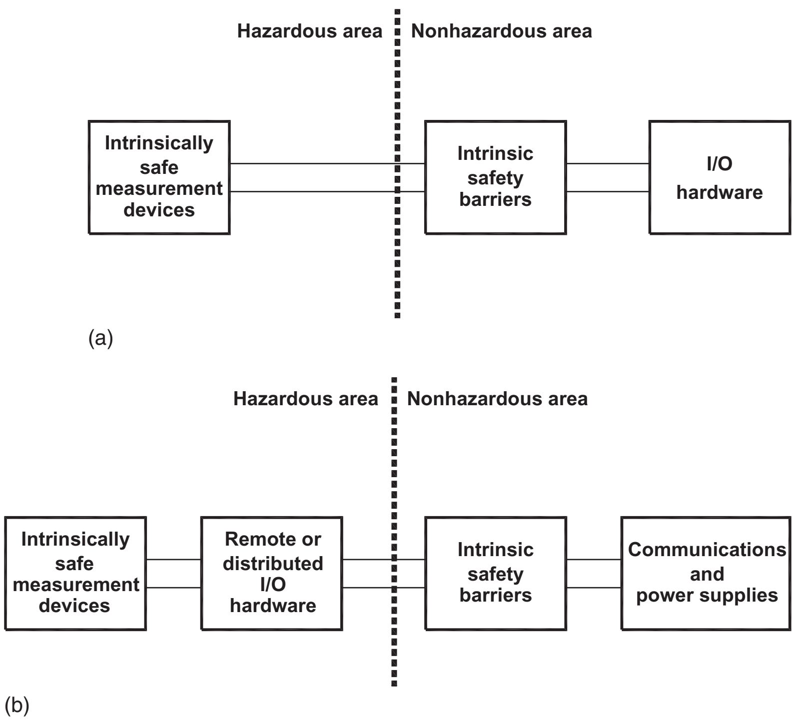

Figure 1.36. Intrinsic safety barriers. (a) Input/output (I/O) hardware in safe area. (b) I/O hardware in hazardous area.

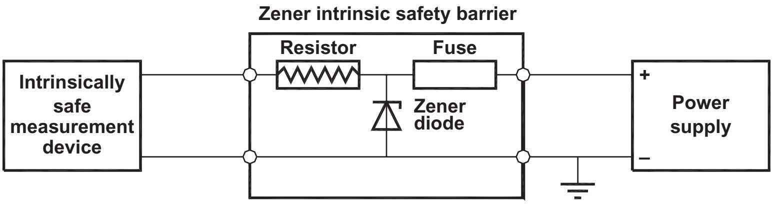

Figure 1.37. Single-channel intrinsic safety barrier, positive polarity.

safety barrier will experience only positive polarity. Circuits are available for negative polarity and alternating polarity. Single-channel means one side is grounded; dual-channel means that neither side is grounded. Consequently, the suppliers offer a large number of designs, with the selection also depending on

- Maximum voltage in safe are.
- Maximum current in safe area.
- Maximum power in safe area.
- Maximum permissible inductance in safe area.
- Maximum permissible capacitance in safe area.

Safety barriers are available for most signals, including

- Two-wire transmitters (4 to 20 ma).
- Thermocouples.
- RTDs (including three wire).
- Process switches (pressure, level, etc.).
- Outputs (to IP converters; 4 to 20 ma).
- Solenoid valves.
- Communication links.

Most suppliers provide a mounting rail or other mechanism that also provides an electrical ground to the barriers.

Attention must be paid to the details. The measurement device must be approved for use in the hazardous area in conjunction with an approved intrinsic safety barrier appropriate for that measurement device. The electrical inspectors start with the measurement device and then proceed to the safety barrier. They expect to see the proper documentation and only equipment with the appropriate certifications for the classification of the hazardous area.

1.19. MEASUREMENT DEVICE DYNAMICS

If the variable being sensed changes, how rapidly does the measurement device output respond to these changes? The answer to this question pertains to the dynamic characteristics of the measurement device. Measurement device dynamics arise more frequently in applications where the output of the measurement device is the input to a process control function.

Temperature Probe

In Figure 1.38, the fluid temperature T_F is measured using a bare temperature probe. Even for this simple configuration, the equations are relatively simple, provided certain assumptions are made. Finite element analysis are required for a more accurate analysis and can encompass the thermowell that is usually present.

Let's use the following notation:

A	Heat transfer area, m^2 or ft^2
c_P	Specific heat of probe, kcal/kg·°C or Btu/lb_m·°F
M	Mass of probe, kg or lb_m
T_F	Fluid temperature, °C or °F
T_P	Probe temperature, °C or °F
T_{ref}	Reference temperature, °C or °F
U	Heat transfer coefficient, kcal/min·m^2·°C or Btu/min·ft^2·°F

Let there be no steady-state error—that is, $T_P = T_F$ at equilibrium. Suppose the fluid temperature T_F increases abruptly by 5 °C. How does the probe tem-

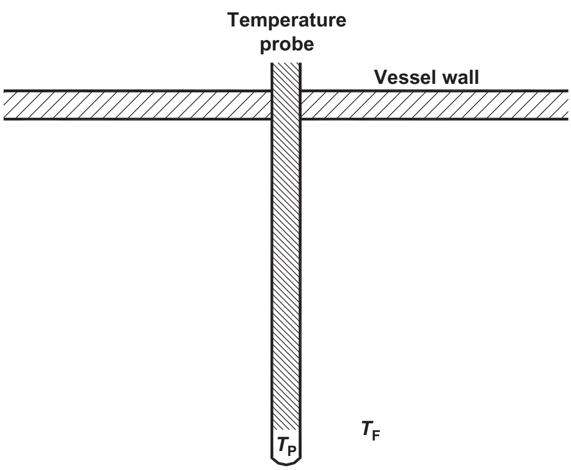

Figure 1.38. Bare probe for temperature measurement.

perature T_P respond? The probe temperature will not increase as rapidly as the fluid temperature. How rapidly the probe responds depends on two factors:

Heat transfer coefficient. The greater the heat transfer coefficient, the greater the heat transfer for a given temperature difference. Because this improves the response of the probe, the probe should be installed at a location where the fluid flows rapidly across the probe, which increases the heat transfer coefficient.

Mass of the probe. The greater the mass of the probe, the more heat that must be transferred to increase the temperature of the probe by 1 °C. For rapid response, the mass of the probe should be as small as possible.

Time Constant

An energy balance around the probe involves the following two terms:

Heat transfer rate	$UA(T_F - T_P)$
Energy content of the probe	$Mc_P(T_P - T_{ref})$

The rate of change of energy within the probe is equal to the heat transfer rate:

$$\frac{d}{dt}[Mc_P(T_P - T_{ref})] = Mc_P\frac{dT_P}{dt} = UA(T_F - T_P)$$

$$\frac{Mc_P}{UA}\frac{dT_P}{dt} + T_P = T_F$$

The coefficient Mc_P/UA has units of time:

Metric $$\frac{Mc_P}{UA} = \frac{(\text{kg})(\text{kcal/kg}\cdot{}^\circ\text{C})}{(\text{kcal/min}\cdot\text{m}^2\cdot{}^\circ\text{C})(\text{m}^2)} = \text{min}$$

English $$\frac{Mc_P}{UA} = \frac{(\text{Ib}_m)(\text{Btu/Ib}_m\cdot{}^\circ\text{F})}{(\text{Btu/min}\cdot\text{ft}^2\cdot{}^\circ\text{F})(\text{ft}^2)} = \text{min}$$

This coefficient is called the *time constant* of the probe. The value of the time constant determines how quickly the probe responds. A probe with a small mass M and a high heat transfer coefficient U responds quickly (value of Mc_P/UA is small) and is said to have a *short* time constant. Conversely, a probe with a large mass M and a low heat transfer coefficient U responds slowly (value of Mc_P/UA is large) and is said to have a *long* time constant.

First-Order Lag

Using the Greek letter τ for the time constant, the differential equation for the probe is written as follows:

$$\tau\frac{dT_P}{dt}+T_P=T_F$$

This is a first-order differential equation. Such processes are referred to as first-order lags.

Suppose the fluid temperature abruptly changes from 60° to 65 °C (a step change of +5 °C). The response of the probe temperature is presented in Figure 1.39 for τ = 1.0 min. When the step change is introduced, the probe temperature immediately begins to respond. But as the probe temperature approaches the fluid temperature, the rate of change becomes progressively slower in an exponentially decaying manner.

For the response illustrated in Figure 1.39, the time constant is 1.0 min. Note that the response attains 63% of the total change in 1.0 min. The 63% point is often used to characterize the speed of response of a process, a measurement device, a valve, etc.

Dead Time or Transportation Lag

In the illustration in Figure 1.40, material is being metered from a feed hopper through a rotary valve onto a belt conveyor. The velocity of the belt conveyor is V ft/min; the distance is L ft.

The time t required to transport material from one end of the conveyor to the other is the length L of the conveyor divided by the velocity V of the

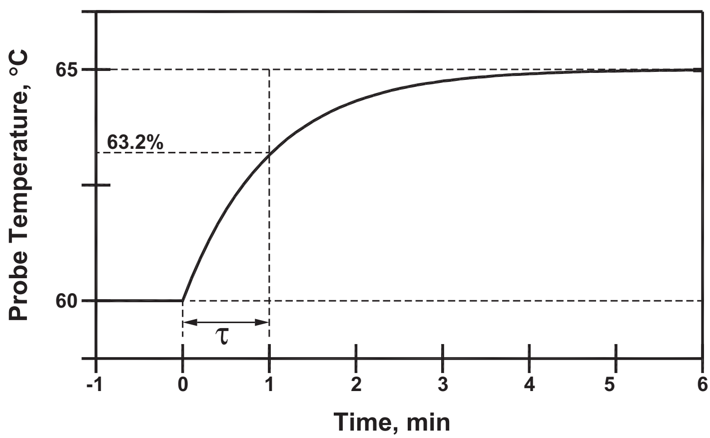

Figure 1.39. Response of a first-order lag to a step change in its input.

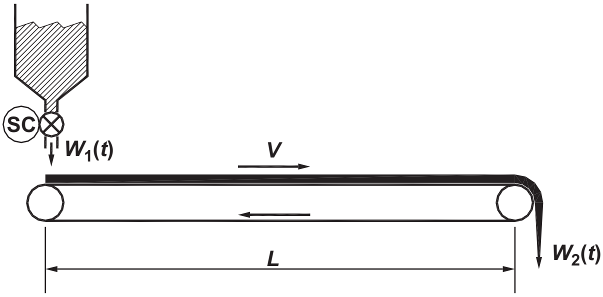

Figure 1.40. Belt conveyor.

conveyor. Although transportation time or transportation lag are more descriptive terms, this time is often referred to as the *dead time* and is designated by θ. For transportation processes,

$$\theta = \frac{L}{V}$$

Material is metered onto the belt conveyor at rate $W_1(t)$. Material falls off the other end of the conveyor at rate $W_2(t)$. The relationship between $W_1(t)$ and $W_2(t)$ is quite simple. The current value for $W_2(t)$ is the value of $W_1(t)$ at 1 dead time in the past. Mathematically, this is expressed as follows:

$$W_2(t) = W_1(t - \theta)$$

The conveyor in Figure 1.40 is a material transport system. Fluid flowing through pipes is another material transport system. Any material transport system exhibits dead time for parameters such as temperature and composition.

Measurement Device Location

When measuring temperatures, dead time can be introduced through the location of the measurement device. The process illustrated in Figure 1.41 mixes steam and cold water to produce hot water. The hot water temperature transmitter has been installed at a distance L from the hot water tank.

A belt conveyor is a material transport system; so is fluid flowing through pipes. The time required for the water to flow from the tank to the transmitter location is the distance L divided by the velocity V of the water flowing through the pipe. Locating the transmitter downstream of the tank introduces dead time:

Transmitter location introduces dead time for measurements of an intensive property (temperature, composition, physical properties, etc). But for measurements of an extensive property (flow, level, etc), sensor location does not

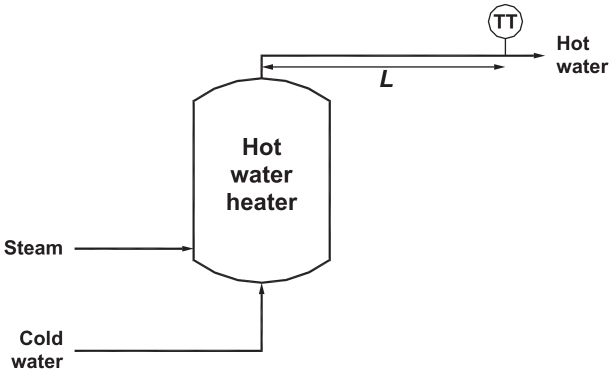

Figure 1.41. Introducing dead time through the location of the temperature transmitter.

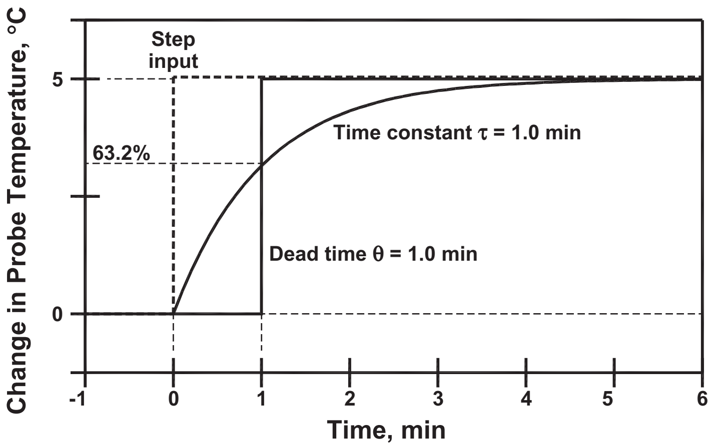

Figure 1.42. Responses to a step change in the input.

introduce dead time. That is, a flow transmitter located a distance L from the tank would indicate the same flow as one located at the tank.

Dead time is always associated with the transport of material from one location to another. This becomes a major concern for analyzer sample systems such as the one shown in Figure 1.6.

Step Response

A customary practice is to present responses as the change in the variable of interest from its starting or initial values. When presented in this manner, the responses always start at zero.

The responses to a step increase of 5 °C for two systems are shown in Figure 1.42:

- A system consisting of a single time constant, with $\tau = 1.0$ min.
- A system consisting of a pure dead time, with $\theta = 1.0$ min.

The time constant system reacts immediately, and in one time constant attains 63% of the total change. The dead time system exhibits no reaction at all until the dead time has elapsed, and then the change in the input immediately appears on the output.

For each, the 63% point corresponds to the lag in the system. However, it does not describe the nature of the lag. For the time constant system, it is a first-order lag; for the dead time system, it is a transportation lag.

When the measured variable is the input to a controller, dead time is far more detrimental to performance than a time constant.

Ramp Response

Figure 1.43 presents the responses to a ramp increase of 5 °C/min for two systems:

- A system consisting of a single time constant, with $\tau = 1.0$ min.
- A system consisting of a pure dead time, with $\theta = 1.0$ min.

The time constant system reacts immediately, but after some time (about 3 min, or 3τ) the response is a ramp that is 1 min later than the ramp input. The dead time system exhibits no reaction at all until the dead time has elapsed, and then the ramp immediately appears on the output. But after approximately 3 min, both outputs are ramps that are 1 min later than the ramp input. The time difference between the ramps is the lag in the system. For the

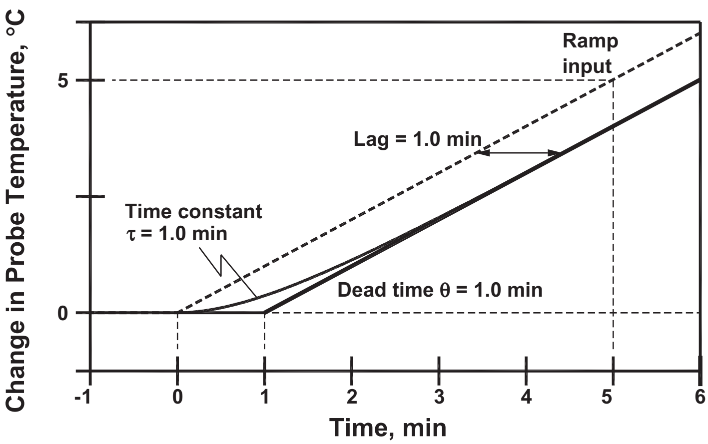

Figure 1.43. Responses to a ramp change in the input.

time constant system, this lag equals the value of the time constant. For the dead time system, this lag equals the value of the dead time.

Dynamic testing involves obtaining the response of the system to an input. We have exhibited this for step inputs and ramp inputs. Another form of the test is the pulse or "bump" test where the input is changed (increased or decreased) for some period of time and then returned to the original value.

Dynamic Performance Measures

When seeking data on the dynamic performance of a measurement device, expect to be disappointed. First, very little information is usually available on the dynamic performance of the measurement device. Second, whatever data are available are often based on applying the measurement device to air or water.

There is no uniform basis for expressing dynamic performance. Some state the time to attain 63% of the total response. This is the lag, but is it a first-order lag or a transportation lag? Most measurement devices behave more like a first-order lag, but a few (notably analyzers) exhibit dead time. Then you will encounter statements such as "90% of the response is attained in less than 20 sec." How do you interpret this?

The common denominator for comparing the dynamic performance of measurement devices has to be the lag. The relationship among the time t, the time constant τ, and the fraction response $c(t)$ at time t is as follows.

$$c(t) = 1 - e^{-t/\tau}$$

or

$$\tau = \frac{-t}{\ln[1 - c(t)]}$$

The values of $c(t)$ at integer multiples of the time constant τ are as follows:

Time	Response, %
τ	63.2
2τ	86.5
3τ	95.0
4τ	98.2
5τ	99.3
6τ	99.8

What is the lag for a measurement device that attains 90% of the change in 20 sec? The list does not contain an entry for 90%, so the lag must be computed from the formula:

$$\tau = \frac{-t}{\ln[1 - c(t)]} = \frac{-20}{\ln[1 - 0.9]} = 8.7\,\text{sec}$$

1.20. FILTERING AND SMOOTHING

The objective of any measurement is to provide a signal that indicates the current value of the process variable of interest. However, this is almost always complicated by the presence of some combination of the following:

Noise of process origin. There may be situations within the process that lead to variability in the variable of interest. Here are a couple of examples:

- Radar level transmitters can detect ripples on the surface caused by agitation within a vessel.
- Coriolis flowmeters respond rapidly enough to sense the pulsating flow from a positive displacement pump.

Measurement noise. This is especially prevalent in some measurements, such as weight gauges. However, it can also occur in flow, level, and pressure. When a thermowell is present, noise in temperature measurements is unusual and is likely the result of electrical problems. However, measurement noise is common on temperatures measured by pyrometers.

Stray electrical pickup. Every analog signal contains stray electrical pickup, the source being AC power at either 50 or 60 Hz. For current loop inputs, the analog input processing hardware in digital systems is specifically designed to remove this component.

Options for Smoothing

There are three possibilities for providing filtering or smoothing:

Between the process and the measurement device. This not recommended! However, it is done, often without official knowledge or sanction. For example, liquid level measurements are often installed in an external chamber that is connected to the process vessel at the bottom (liquid connection) and at the top (gas or vapor connection). Smoothing, usually to an unknown but excessive degree, is achieved by partially closing the isolation valve in the liquid (bottom) connection.

Within the measurement device. Most conventional transmitters provide a "damping" setting, usually via an uncalibrated screw adjustment. With no calibration, zero or no smoothing is the only setting for which the amount of smoothing is known. Smart transmitters also provide smoothing, but with the smoothing coefficient specified via a configuration parameter.

Within the data acquisition software. The input processing routines for digital systems provide the option for applying smoothing to the input. Before the advent of smart transmitters, this was the preferred approach.

The important aspect is to know exactly how much smoothing is being provided. For smart transmitters, providing smoothing within the transmitter is a viable alternative and is usually the preferred approach.

Frequency

A signal such as the output of a measurement device can be expressed as the sum of sinusoidal signals of various frequencies and amplitudes. The amplitude of the sinusoidal component at a given frequency can be computed by the Fourier integral:

$$G(j\omega) = \int_0^\infty x(t)e^{-j\omega t}dt$$

where $x(t)$ = signal; ω = frequency (radians/sec); t = time (sec).

The amplitude at frequency ω is $|G(j\omega)|$. A plot of $|G(j\omega)|$ as a function of frequency ω is a good way to characterize the nature of a signal. Unfortunately, such plots are rarely generated for the outputs from measurement devices. To obtain useful data from the Fourier integral, high scan rates (such as 100 samples per second) are normally required. The data acquisition equipment customarily installed in process applications cannot achieve such rates.

In process applications, the content of interest is usually at the lower frequencies. Noise of process origin is normally at a higher frequency, measurement noise is at an even higher frequency, and stray electrical pickup (at 50 or 60 Hz) is usually the highest component of significance.

The units for frequency in the Fourier integral are radians per second. The more common units are cycles per second, or Hertz (Hz). There are 2π radians per cycle, so multiply the frequency in Hertz by 2π to obtain the frequency in radians per second.

Filter

Smoothing is provided by a filter. In block diagrams, the filter is normally inserted between the measurement device and the digital system. Figure 1.44 illustrates a control loop that contains a filter. For analog equipment, the filter is a physical piece of hardware usually attached to the input terminals. For digital systems, the filter is provided in the software. With conventional (dumb) transmitters, the filter is normally incorporated into the data acquisition software. With smart transmitters, it is usually preferable to use the filter incorporated into the measurement device's software.

In process applications, the filter is a *low-pass filter* that attenuates the high-frequency components of a signal but not the low-frequency components. An

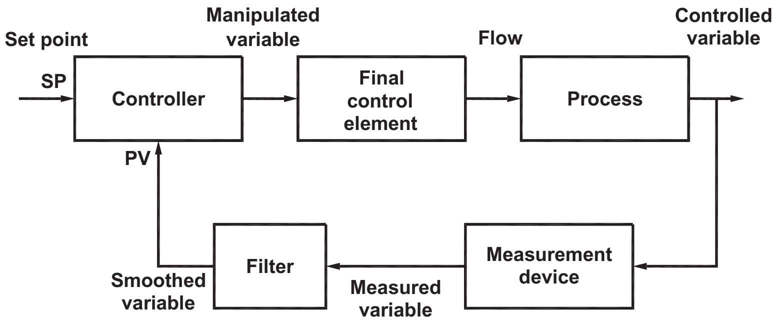

Figure 1.44. Control loop with a filter on the measured value.

ideal low-pass filter completely removes all components above a specified frequency (the cutoff frequency) and passes unaltered all components below this frequency. Unfortunately, such filters are unrealizable.

Practical low-pass filters do not provide a sharp cutoff as in the ideal filter. The cutoff frequency of a practical filter is normally defined as the frequency for which the attenuation factor is 0.891. The attenuation is greater for frequencies above the cutoff frequency (the higher the frequency, the greater the attenuation). Components at frequencies well below the cutoff frequency are passed basically unaltered, but some attenuation occurs for frequencies just below the cutoff frequency.

Exponential Filter

The exponential filter is described by the following differential equation.

$$\tau_F \frac{dy(t)}{dt} + y(t) = x(t)$$

where $x(t)$ = input to filter; $y(t)$ = output from filter; τ_F = filter time constant (sec); t = time (sec).

This equation is exactly the same as the equation for a first-order lag with time constant τ_F. The step response and the ramp response of the exponential filter also exhibit the same behavior as a first-order lag. Therefore, the following have the same effect:

- Applying an exponential filter to the output of a measurement device.
- Inserting a temperature measurement device into a thermowell.

In analog systems, the exponential filter is implemented in hardware, the most common being the resistor-capacitor (RC) network filter. The differential equation just given applies to the RC network filter, with τ_F = 1/RC. In digital systems, the exponential filter is implemented in software using a difference equation that is derived from the first-order lag equation as follows.

Differential equation for exponential filter:

$$\tau_F \frac{dy(t)}{dt} + y(t) = x(t)$$

Difference equation with backward difference to approximate the derivative:

$$\tau_F \frac{y_i - y_{i-1}}{\Delta t} + y_i = x_i$$

Solve difference equation for y_i:

$$y_i = \frac{\Delta t}{\tau_F + \Delta t} x_i + \frac{\tau_F}{\tau_F + \Delta t} y_{i-1}$$

Introduce smoothing coefficient k:

$$y_i = kx_i + (1-k)\,y_{i-1}$$

$$k = \frac{\Delta t}{\tau_F + \Delta t}$$

where Δt = time between execution (also known as the *sampling time*).

An alternate derivation gives $k = 1 - \exp(-\Delta t/\tau_F)$, but the expression for k derived above is equivalent for $\Delta t << \tau_F$. In some systems, the degree of smoothing is specified as the smoothing coefficient k (note that $0 < k < 1$) of the difference equation, but the trend is to specify the filter time constant τ_F.

Moving Average Filter

The moving average filter smoothes by computing the arithmetic average of some number N of consecutive input values.

$$y_i = \frac{1}{N} \sum_{j=0}^{N-1} x_{i-j}$$

The smoothing can be specified as either of the following:

- Averaging time T_A, usually in seconds. All input values received during this period of time are averaged.
- Number of input values N to be averaged.

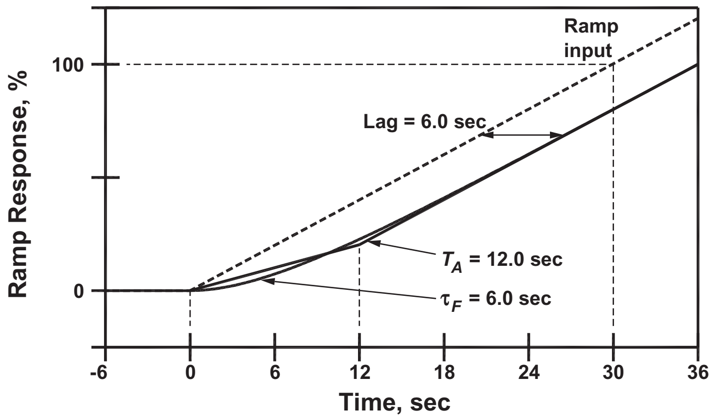

Figure 1.45. Effective lag for the exponential filter and the moving average filter.

Note that $N = T_A/\Delta t$, where Δt is the interval (in seconds) between values.

What is the origin of the term *moving*? The filter maintains a storage array for the required number N of input values. When a new input value is received, the new value replaces the oldest value in the storage array, and the average is recomputed. Unlike the exponential filter, there is no practical analog equivalent to the moving average filter.

Although not identical, the performance of the two filters is very similar when $T_A = 2\tau_F$. Figure 1.45 presents the ramp responses for an exponential filer with $\tau_F = 6$ sec and a moving average filter with $T_A = 12$ sec. The ramp response for the moving average filter clearly illustrates that the lag is $T_A/2$.

Use and Abuse of Filters

Specifying a value for the smoothing coefficient (τ_F for the exponential filter; T_A for the moving average filter) determines the cutoff frequency for the filter. To do so precisely, the first step is to characterize the input signal in terms of its components at various frequencies. In the process industries, this is rarely done, one reason being that it requires a scan rate higher than can be achieved by most industrial data acquisition systems.

In the process industries, the usual approach for setting the smoothing coefficient is to observe the variations in the measured variable and then increase the smoothing until the variations are insignificant. The typical result is excessive smoothing.

To make matters worse, filtering is often used to hide or obscure problems, either within the process, the measurement device, or the wiring. The presence of noise in the signal from a temperature measurement device inserted into a thermowell is most likely due to wiring problems. But instead of locating the wiring problem and correcting it, the temptation to hide the problem with smoothing is apparently irresistible. Coriolis flow meters respond rapidly enough to detect the pulsating flow from a positive displacement pump. A pulsation damper should be installed, properly pressurized, and otherwise in working order. Smoothing is not an alternative to a properly functioning pulsation damper.

LITERATURE CITED

1. International Organization for Standardization, *International vocabulary of metrology—Basic and general concepts and associated terms*, ISO/IEC Guide 99, ISO, Geneva, 2007.
2. International Organization for Standardization, *Guide to the Expression of Uncertainty in Measurement*, ISO/IEC Guide 98, ISO, Geneva, 1995.
3. International Organization for Standardization, *Measurement of Fluid Flow by Means of Pressure Differential Devices Inserted in Circular Cross-Section Conduits Running Full—Part 2: Orifice Plates*, ISO 5167-2, ISO, Geneva, 2003.
4. Adams, Thomas M., *A2LA Guide for the Estimation of Measurement Uncertainty in Testing*, G-104, American Association for Laboratory Accreditation, Frederick, MD, July 2002.
5. International Organization for Standardization, *General Requirements for the Competence of Testing and Calibration Laboratories*, ISO/IEC 17025, ISO, Geneva, 2005.
6. *Modicon Modbus Protocol Reference Guide*, PI–MBUS–300, Rev. J, Modicon, Int. Andover, MA, June 1996.
7. International Electrotechnical Commission, *Industrial Communications Networks—Fieldbus Specifications*, IEC 61158 (all parts), IEC, Geneva, 2007.
8. National Electrical Manufacturers Association, *Enclosures for Electrical Equipment (1000 Volts Maximum)*, NEMA Standards Publication 250-2003, NEMA, Rosslyn, VA, 2003.
9. International Electrotechnical Commission, *Degrees of Protection Provided by Enclosures (IP Code)*, IEC 60529, IEC, Geneva, 2001.
10. R. Stah/AG, *Basics of Explosion Protection*, available at www.rstahl.com.
11. National Fire Protection Association, *National Electrical Code*, NFPA 70, NFPA, Quincy, MA, 2008.
12. International Electrotechnical Commission, *Electrical Apparatus for Explosive Gas Atmospheres—Part 10: Classification of Hazardous Areas*, IEC 60079-10, IEC, Geneva, 2002.
13. National Fire Protection Association, *Standard for Purged and Pressurized Enclosures for Electrical Equipment*, NFPA 496, NFPA, Quincy, MA, 2008.
14. International Electrotechnical Commission, *Explosive Atmospheres—Part 2: Equipment Protection by Pressurized Enclosures*, IEC 60079-2, IEC, Geneva, 2007.