Peter H. Sydenham. "Static and Dynamic Characteristics of Instrumentation."

Copyright 2000 CRC Press LLC. <http://www.engnetbase.com>.

Static and Dynamic Characteristics of Instrumentation

Peter H. Sydenham
University of South Australia

Before we can begin to develop an understanding of the static and time changing characteristics of measurements, it is necessary to build a framework for understanding the process involved, setting down the main words used to describe concepts as we progress.

Measurement is the process by which relevant information about a system of interest is interpreted using the human thinking ability to define what is believed to be the new knowledge gained. This information may be obtained for purposes of controlling the behavior of the system (as in engineering applications) or for learning more about it (as in scientific investigations).

The basic entity needed to develop the knowledge is called *data*, and it is obtained with physical assemblies known as sensors that are used to observe or sense system variables. The terms *information* and *knowledge* tend to be used interchangeably to describe the entity resulting after data from one or more sensors have been processed to give more meaningful understanding. The individual variables being sensed are called *measurands.*

The most obvious way to make observations is to use the human senses of seeing, feeling, and hearing. This is often quite adequate or may be the only means possible. In many cases, however, sensors are used that have been devised by man to enhance or replace our natural sensors. The number and variety of sensors is very large indeed. Examples of man-made sensors are those used to measure temperature, pressure, or length. The process of sensing is often called *transduction*, being made with transducers. These man-made sensor assemblies, when coupled with the means to process the data into knowledge, are generally known as (measuring) instrumentation.

The degree of perfection of a measurement can only be determined if the goal of the measurement can be defined without error. Furthermore, instrumentation cannot be made to operate perfectly. Because of these two reasons alone, measuring instrumentation cannot give ideal sensing performance and it must be selected to suit the allowable error in a given situation.

© 1999 by CRC Press LLC

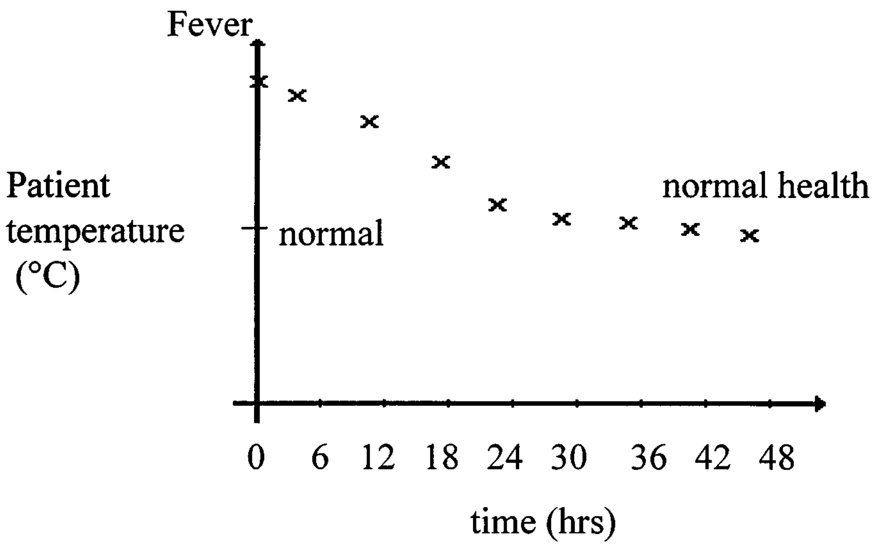

FIGURE 3.1 A patient's temperature chart shows changes taking place over time.

Measurement is a process of mapping actually occurring variables into equivalent values. Deviations from perfect measurement mappings are called *errors*: what we get as the result of measurement is not exactly what is being measured. A certain amount of error is allowable provided it is below the level of uncertainty we can accept in a given situation. As an example, consider two different needs to measure the measurand, time. The uncertainty to which we must measure it for daily purposes of attending a meeting is around a 1 min in 24 h. In orbiting satellite control, the time uncertainty needed must be as small as milliseconds in years. Instrumentation used for the former case costs a few dollars and is the watch we wear; the latter instrumentation costs thousands of dollars and is the size of a suitcase.

We often record measurand values as though they are constant entities, but they usually change in value as time passes. These "dynamic" variations will occur either as changes in the measurand itself or where the measuring instrumentation takes time to follow the changes in the measurand — in which case it may introduce unacceptable error.

For example, when a fever thermometer is used to measure a person's body temperature, we are looking to see if the person is at the normally expected value and, if it is not, to then look for changes over time as an indicator of his or her health. Figure 3.1 shows a chart of a patient's temperature. Obviously, if the thermometer gives errors in its use, wrong conclusions could be drawn. It could be in error due to incorrect calibration of the thermometer or because no allowance for the dynamic response of the thermometer itself was made.

Instrumentation, therefore, will only give adequately correct information if we understand the static and dynamic characteristics of both the measurand and the instrumentation. This, in turn, allows us to then decide if the error arising is small enough to accept.

As an example, consider the electronic signal amplifier in a sound system. It will be commonly quoted as having an amplification constant after feedback if applied to the basic amplifier of, say, 10. The actual amplification value is dependent on the frequency of the input signal, usually falling off as the frequency increases. The frequency response of the basic amplifier, before it is configured with feedback that markedly alters the response and lowers the amplification to get a stable operation, is shown as a graph of amplification gain versus input frequency. An example of the open loop gain of the basic amplifier is given in Figure 3.2. This lack of uniform gain over the frequency range results in error — the sound output is not a true enough representation of the input.

© 1999 by CRC Press LLC

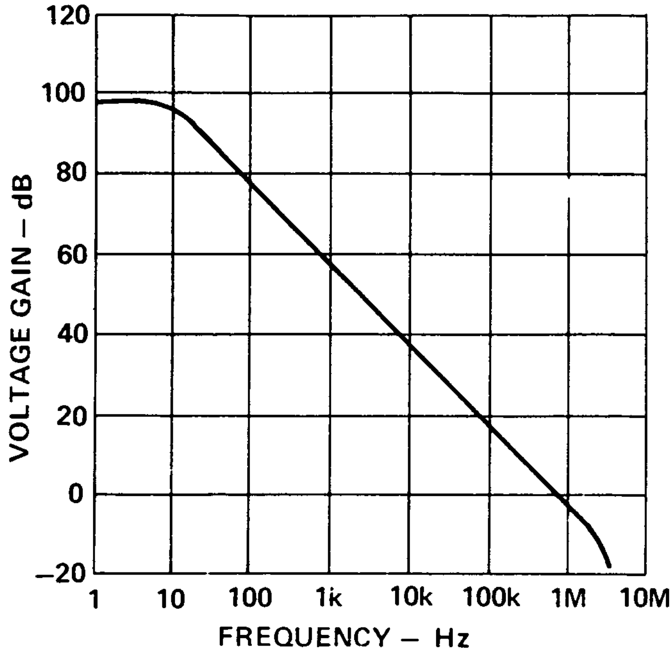

FIGURE 3.2 This graph shows how the amplification of an amplifier changes with input frequency.

Before we can delve more deeply into the static and dynamic characteristics of instrumentation, it is necessary to understand the difference in meaning between several basic terms used to describe the results of a measurement activity.

The correct terms to use are set down in documents called *standards*. Several standardized metrology terminologies exist but they are not consistent. It will be found that books on instrumentation and statements of instrument performance often use terms in different ways. Users of measurement information need to be constantly diligent in making sure that the statements made are interpreted correctly.

The three companion concepts about a measurement that need to be well understood are its *discrimination*, its *precision*, and its *accuracy*. These are too often used interchangeably — which is quite wrong to do because they cover quite different concepts, as will now be explained.

When making a measurement, the smallest increment that can be discerned is called the *discrimination*. (Although now officially declared as wrong to use, the term *resolution* still finds its way into books and reports as meaning discrimination.) The discrimination of a measurement is important to know because it tells if the sensing process is able to sense fine enough changes of the measurand.

Even if the discrimination is satisfactory, the value obtained from a repeated measurement will rarely give exactly the same value each time the same measurement is made under conditions of constant value of measurand. This is because errors arise in real systems. The spread of values obtained indicates the precision of the set of the measurements. The word *precision* is not a word describing a quality of the measurement and is incorrectly used as such. Two terms that should be used here are: *repeatability*, which describes the variation for a set of measurements made in a very short period; and the *reproducibility*, which is the same concept but now used for measurements made over a long period. As these terms describe the outcome of a set of values, there is need to be able to quote a single value to describe the overall result of the set. This is done using statistical methods that provide for calculation of the "mean value" of the set and the associated spread of values, called its *variance*.

The *accuracy* of a measurement is covered in more depth elsewhere so only an introduction to it is required here. Accuracy is the closeness of a measurement to the value defined to be the true value. This

© 1999 by CRC Press LLC

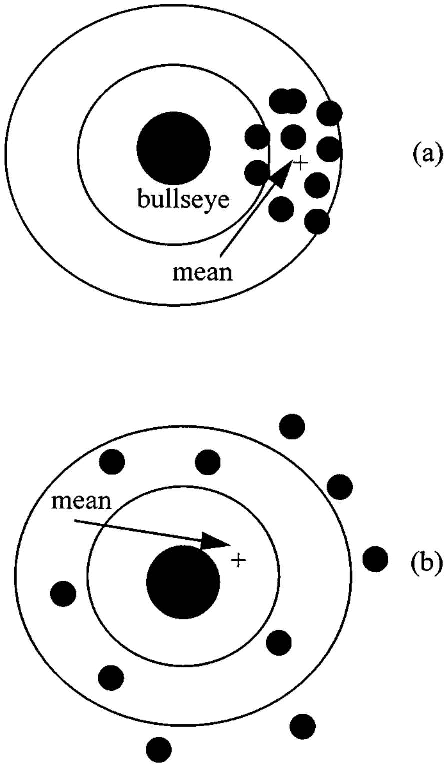

FIGURE 3.3 Two sets of arrow shots fired into a target allow understanding of the measurement concepts of discrimination, precision, and accuracy. (a) The target used for shooting arrows allows investigation of the terms used to describe the measurement result. (b) A different set of placements.

concept will become clearer when the following illustrative example is studied for it brings together the three terms into a single perspective of a typical measurement.

Consider then the situation of scoring an archer shooting arrows into a target as shown in Figure 3.3(*a*). The target has a central point — the bulls-eye. The objective for a perfect result is to get all arrows into the bulls-eye. The rings around the bulls-eye allow us to set up numeric measures of less-perfect shooting performance.

Discrimination is the distance at which we can just distinguish (i.e., discriminate) the placement of one arrow from another when they are very close. For an arrow, it is the thickness of the hole that decides the discrimination. Two close-by positions of the two arrows in Figure 3.3(*a*) cannot be separated easily. Use of thinner arrows would allow finer detail to be decided.

Repeatability is determined by measuring the spread of values of a set of arrows fired into the target over a short period. The smaller the spread, the more precise is the shooter. The shooter in Figure 3.3(*a*) is more precise than the shooter in Figure 3.3(*b*).

If the shooter returned to shoot each day over a long period, the results may not be the same each time for a shoot made over a short period. The mean and variance of the values are now called the *reproducibility* of the archer's performance.

Accuracy remains to be explained. This number describes how well the mean (the average) value of the shots sits with respect to the bulls-eye position. The set in Figure 3.3(*b*) is more accurate than the set in Figure 3.3(*a*) because the mean is nearer the bulls-eye (but less precise!).

At first sight, it might seem that the three concepts of discrimination, precision, and accuracy have a strict relationship in that a better measurement is always that with all three aspects made as high as is affordable. This is not so. They need to be set up to suit the needs of the application.

We are now in a position to explore the commonly met terms used to describe aspects of the static and the dynamic performance of measuring instrumentation.

© 1999 by CRC Press LLC

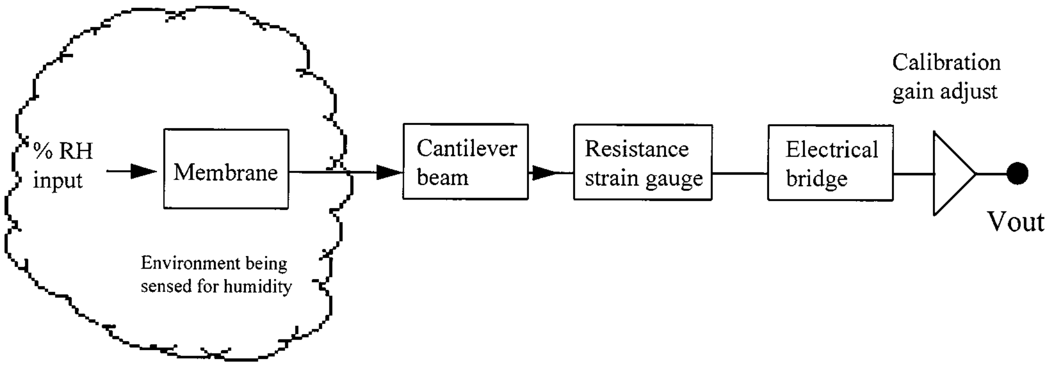

FIGURE 3.4 Instruments are formed from a connection of blocks. Each block can be represented by a conceptual and mathematical model. This example is of one type of humidity sensor.

3.1 Static Characteristics of Instrument Systems

Output/Input Relationship

Instrument systems are usually built up from a serial linkage of distinguishable building blocks. The actual physical assembly may not appear to be so but it can be broken down into a representative diagram of connected blocks. Figure 3.4 shows the block diagram representation of a humidity sensor. The sensor is activated by an input physical parameter and provides an output signal to the next block that processes the signal into a more appropriate state.

A key generic entity is, therefore, the relationship between the input and output of the block. As was pointed out earlier, all signals have a time characteristic, so we must consider the behavior of a block in terms of both the static and dynamic states.

The behavior of the static regime alone and the combined static and dynamic regime can be found through use of an appropriate mathematical model of each block. The mathematical description of system responses is easy to set up and use if the elements all act as linear systems and where addition of signals can be carried out in a linear additive manner. If nonlinearity exists in elements, then it becomes considerably more difficult — perhaps even quite impractical — to provide an easy to follow mathematical explanation. Fortunately, general description of instrument systems responses can be usually be adequately covered using the linear treatment.

The output/input ratio of the whole cascaded chain of blocks 1, 2, 3, etc. is given as:

$$[\text{output/input}]_{\text{total}} = [\text{output/input}]_1 \times [\text{output/input}]_2 \times [\text{output/input}]_3 \ldots$$

The output/input ratio of a block that includes both the static and dynamic characteristics is called the *transfer function* and is given the symbol G.

The equation for G can be written as two parts multiplied together. One expresses the static behavior of the block, that is, the value it has after all transient (time varying) effects have settled to their final state. The other part tells us how that value responds when the block is in its dynamic state. The static part is known as the *transfer characteristic* and is often all that is needed to be known for block description.

The static and dynamic response of the cascade of blocks is simply the multiplication of all individual blocks. As each block has its own part for the static and dynamic behavior, the cascade equations can be rearranged to separate the static from the dynamic parts and then by multiplying the static set and the dynamic set we get the overall response in the static and dynamic states. This is shown by the sequence of Equations 3.1 to 3.4.

© 1999 by CRC Press LLC

$$G_{total} = G_1 \times G_2 \times G_3 \ldots \tag{3.1}$$

$$= [\text{static} \times \text{dynamic}]_1 \times [\text{static} \times \text{dynamic}]_2 \times [\text{static} \times \text{dynamic}]_3 \ldots \tag{3.2}$$

$$= [\text{static}]_1 \times [\text{static}]_2 \times [\text{static}]_3 \ldots \times [\text{dynamic}]_1 \times [\text{dynamic}]_2 \times [\text{dynamic}]_3 \ldots \tag{3.3}$$

$$= [\text{static}]_{total} \times [\text{dynamic}]_{total} \tag{3.4}$$

An example will clarify this. A mercury-in-glass fever thermometer is placed in a patient's mouth. The indication slowly rises along the glass tube to reach the final value, the body temperature of the person. The slow rise seen in the indication is due to the time it takes for the mercury to heat up and expand up the tube. The static *sensitivity* will be expressed as so many scale divisions per degree and is all that is of interest in this application. The dynamic characteristic will be a time varying function that settles to unity after the transient effects have settled. This is merely an annoyance in this application but has to be allowed by waiting long enough before taking a reading. The wrong value will be viewed if taken before the transient has settled.

At this stage, we will now consider only the nature of the static characteristics of a chain; dynamic response is examined later.

If a sensor is the first stage of the chain, the static value of the gain for that stage is called the *sensitivity.* Where a sensor is not at the input, it is called the *amplification factor* or *gain.* It can take a value less than unity where it is then called the *attenuation.*

Sometimes, the instantaneous value of the signal is rapidly changing, yet the measurement aspect part is static. This arises when using ac signals in some forms of instrumentation where the amplitude of the waveform, not its frequency, is of interest. Here, the static value is referred to as its *steady state* transfer characteristic.

Sensitivity may be found from a plot of the input and output signals, wherein it is the slope of the graph. Such a graph, see Figure 3.5, tells much about the static behavior of the block.

The intercept value on the y-axis is the *offset* value being the output when the input is set to zero. Offset is not usually a desired situation and is seen as an error quantity. Where it is deliberately set up, it is called the *bias.*

The range on the x-axis, from zero to a safe maximum for use, is called the *range* or *span* and is often expressed as the zone between the 0% and 100% points. The ratio of the span that the output will cover

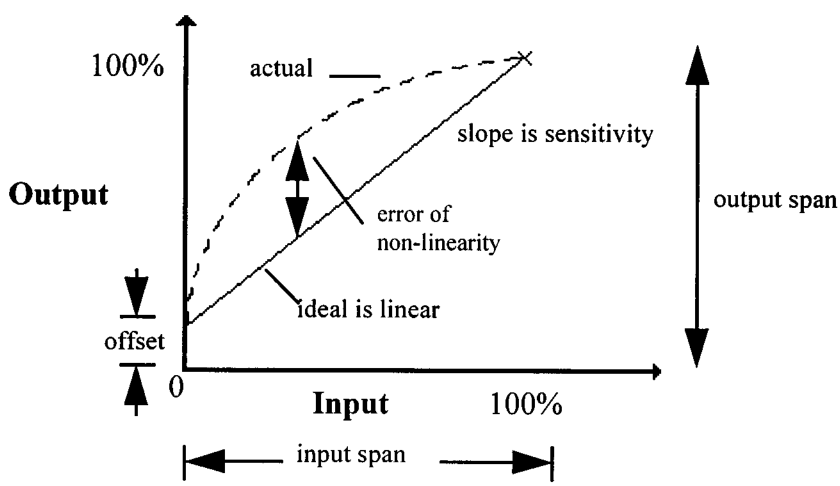

FIGURE 3.5 The graph relating input to output variables for an instrument block shows several distinctive static performance characteristics.

© 1999 by CRC Press LLC

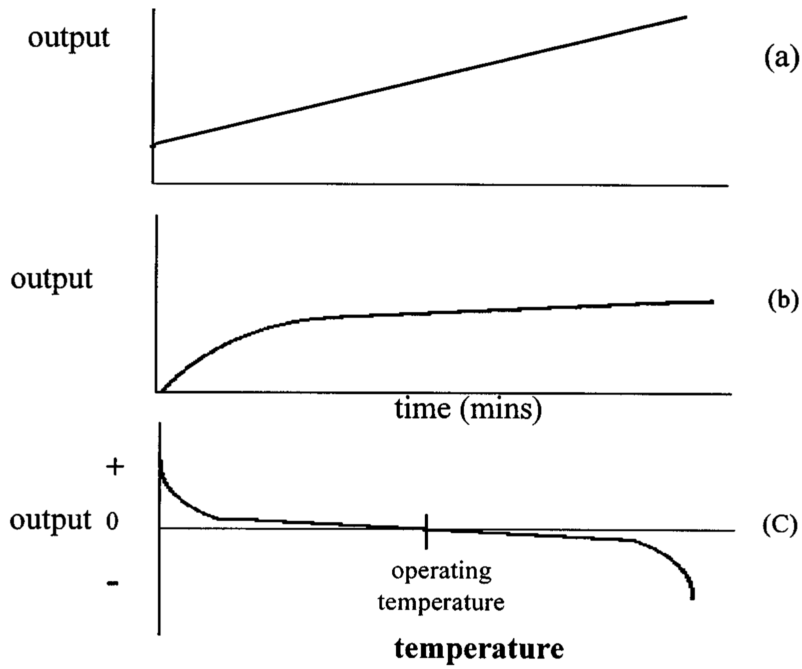

FIGURE 3.6 Drift in the performance of an instrument takes many forms: (*a*) drift over time for a spring balance; (*b*) how an electronic amplifier might settle over time to a final value after power is supplied; (*c*) drift, due to temperature, of an electronic amplifier varies with the actual temperature of operation.

for the related input range is known as the *dynamic range*. This can be a confusing term because it does not describe dynamic time behavior. It is particularly useful when describing the capability of such instruments as flow rate sensors — a simple orifice plate type may only be able to handle dynamic ranges of 3 to 4, whereas the laser Doppler method covers as much as 10^7 variation.

Drift

It is now necessary to consider a major problem of instrument performance called *instrument drift*. This is caused by variations taking place in the parts of the instrumentation over time. Prime sources occur as chemical structural changes and changing mechanical stresses. Drift is a complex phenomenon for which the observed effects are that the sensitivity and offset values vary. It also can alter the accuracy of the instrument differently at the various amplitudes of the signal present.

Detailed description of drift is not at all easy but it is possible to work satisfactorily with simplified values that give the average of a set of observations, this usually being quoted in a conservative manner. The first graph (*a*) in Figure 3.6 shows typical steady drift of a measuring spring component of a weighing balance. Figure 3.6(*b*) shows how an electronic amplifier might settle down after being turned on.

Drift is also caused by variations in environmental parameters such as temperature, pressure, and humidity that operate on the components. These are known as *influence parameters*. An example is the change of the resistance of an electrical resistor, this resistor forming the critical part of an electronic amplifier that sets its gain as its operating temperature changes.

Unfortunately, the observed effects of influence parameter induced drift often are the same as for time varying drift. Appropriate testing of blocks such as electronic amplifiers does allow the two to be separated to some extent. For example, altering only the temperature of the amplifier over a short period will quickly show its temperature dependence.

Drift due to influence parameters is graphed in much the same way as for time drift. Figure 3.6(*c*) shows the drift of an amplifier as temperature varies. Note that it depends significantly on the temperature

© 1999 by CRC Press LLC

of operation, implying that the best designs are built to operate at temperatures where the effect is minimum.

Careful consideration of the time and influence parameter causes of drift shows they are interrelated and often impossible to separate. Instrument designers are usually able to allow for these effects, but the cost of doing this rises sharply as the error level that can be tolerated is reduced.

Hysteresis and Backlash

Careful observation of the output/input relationship of a block will sometimes reveal different results as the signals vary in direction of the movement. Mechanical systems will often show a small difference in length as the direction of the applied force is reversed. The same effect arises as a magnetic field is reversed in a magnetic material. This characteristic is called *hysteresis*. Figure 3.7 is a generalized plot of the output/input relationship showing that a closed loop occurs. The effect usually gets smaller as the amplitude of successive excursions is reduced, this being one way to tolerate the effect. It is present in most materials. Special materials have been developed that exhibit low hysteresis for their application — transformer iron laminations and clock spring wire being examples.

Where this is caused by a mechanism that gives a sharp change, such as caused by the looseness of a joint in a mechanical joint, it is easy to detect and is known as *backlash*.

Saturation

So far, the discussion has been limited to signal levels that lie within acceptable ranges of amplitude. Real system blocks will sometimes have input signal levels that are larger than allowed. Here, the dominant errors that arise — *saturation* and *crossover distortion* — are investigated.

As mentioned above, the information bearing property of the signal can be carried as the instantaneous value of the signal or be carried as some characteristic of a rapidly varying ac signal. If the signal form is not amplified faithfully, the output will not have the same linearity and characteristics.

The gain of a block will usually fall off with increasing size of signal amplitude. A varying amplitude input signal, such as the steadily rising linear signal shown in Figure 3.8, will be amplified differently according to the gain/amplitude curve of the block. In uncompensated electronic amplifiers, the larger amplitudes are usually less amplified than at the median points.

At very low levels of input signal, two unwanted effects may arise. The first is that small signals are often amplified more than at the median levels. The second error characteristic arises in electronic amplifiers because the semiconductor elements possess a dead-zone in which no output occurs until a small threshold is exceeded. This effect causes crossover distortion in amplifiers.

If the signal is an ac waveform, see Figure 3.9, then the different levels of a cycle of the signal may not all be amplified equally. Figure 3.9(*a*) shows what occurs because the basic electronic amplifying elements are only able to amplify one polarity of signal. The signal is said to be *rectified*. Figure 3.9(*b*) shows the effect when the signal is too large and the top is not amplified. This is called *saturation* or *clipping*. (As with many physical effects, this effect is sometimes deliberately invoked in circuitry, an example being where it is used as a simple means to convert sine-waveform signals into a square waveform.) Crossover distortion is evident in Figure 3.9(*c*) as the signal passes from negative to positive polarity.

Where input signals are small, such as in sensitive sensor use, the form of analysis called *small signal* behavior is needed to reveal distortions. If the signals are comparatively large, as for digital signal considerations, a *large signal* analysis is used. Design difficulties arise when signals cover a wide dynamic range because it is not easy to allow for all of the various effects in a single design.

Bias

Sometimes, the electronic signal processing situation calls for the input signal to be processed at a higher average voltage or current than arises normally. Here a dc value is added to the input signal to raise the level to a higher state as shown in Figure 3.10. A need for this is met where only one polarity of signal

© 1999 by CRC Press LLC

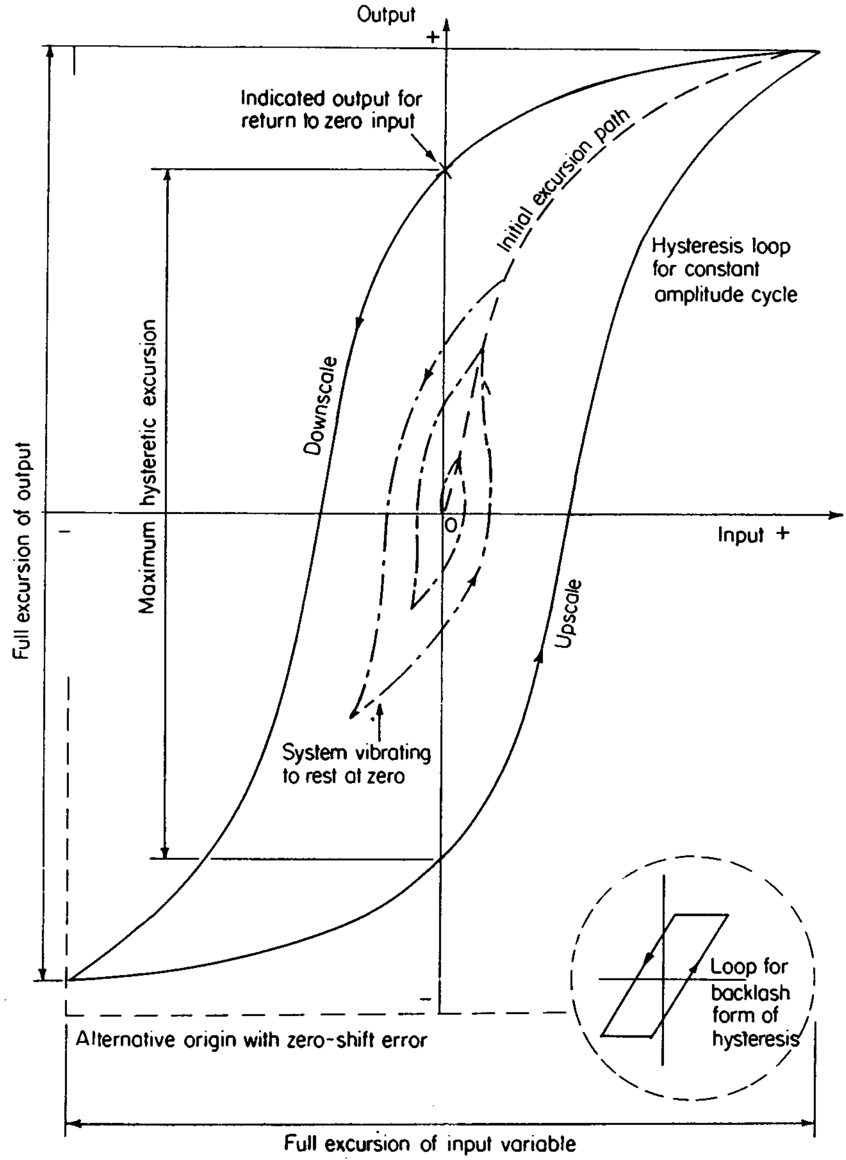

FIGURE 3.7 Generalized graph of output/input relationship where hysteresis is present. (From P. H. Sydenham, *Handbook of Measurement Science,* Vol. 2, Chichester, U.K., John Wiley & Sons, 1983. With permission.)

can be amplified by a single semiconductor element. Raising the level of all of the waveform equally takes all parts into the reasonably linear zone of an amplifier, allowing more faithful replication. If bias were not used here, then the lower half cycle would not be amplified, resulting in only the top half appearing in the output.

Error of Nonlinearity

Ideally, it is often desired that a strictly linear relationship exists between input and output signals in amplifiers. Practical units, however, will always have some degree of nonconformity, which is called the *nonlinearity.* If an instrument block has constant gain for all input signal levels, then the relationship graphing the input against the output will be a straight line; the relationship is then said to be linear.

© 1999 by CRC Press LLC

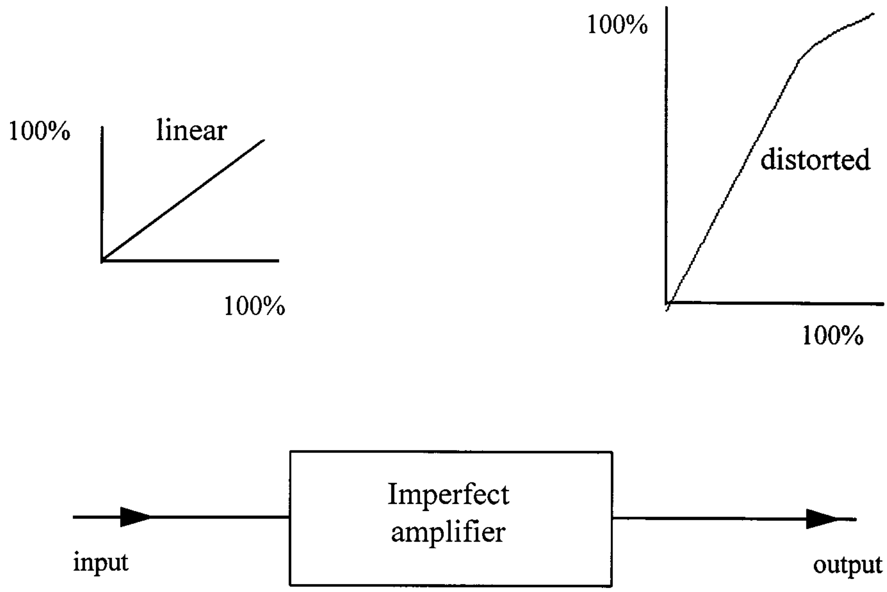

FIGURE 3.8 Nonlinear amplification can give rise to unwanted output distortion.

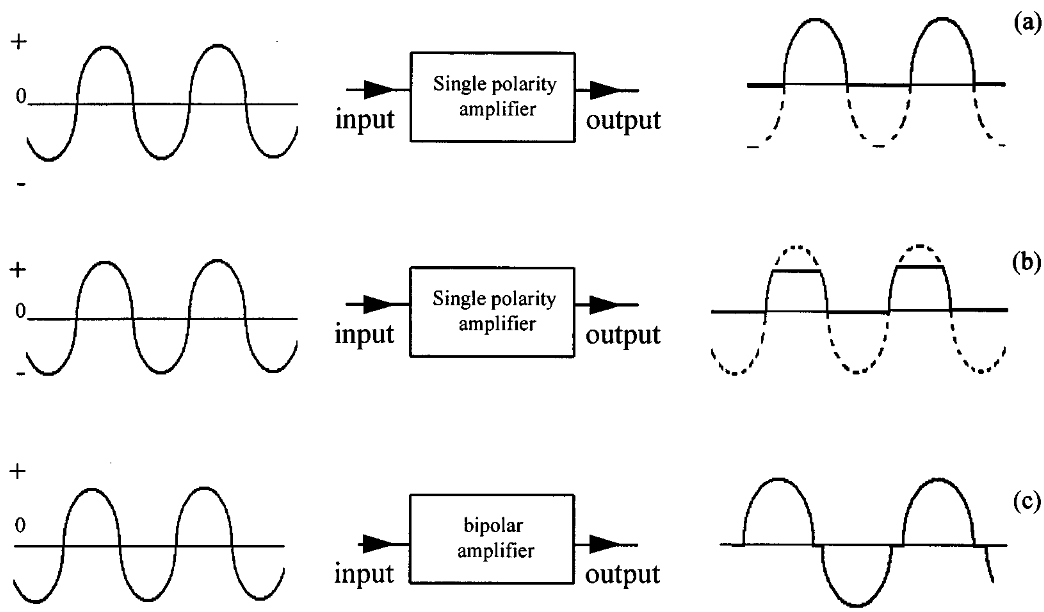

FIGURE 3.9 Blocks can incorrectly alter the shape of waveforms if saturation and crossover effects are not controlled: (*a*) rectification; (*b*) saturation; and (*c*) crossover distortion.

Linearity is the general term used to describe how close the actual response is compared with that ideal line. The correct way to describe the error here is as the *error of nonlinearity.* Note, however, that not all responses are required to be linear; another common one follows a logarithmic relationship.

Detailed description of this error is not easy for that would need a statement of the error values at all points of the plot. Practice has found that a shorthand statement can be made by quoting the maximum departure from the ideal as a ratio formed with the 100% value.

© 1999 by CRC Press LLC

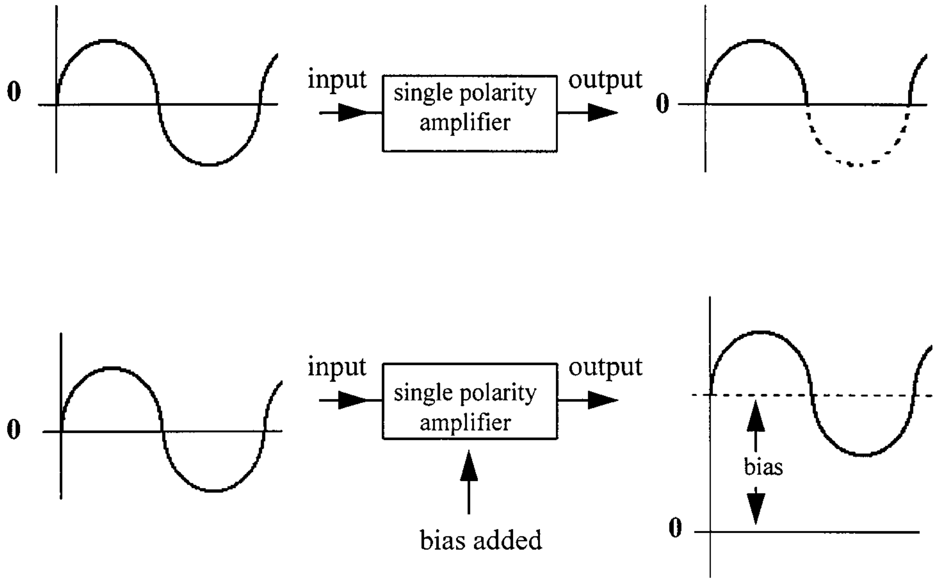

FIGURE 3.10 Bias is where a signal has all of its value raised by an equal amount. Shown here is an ac input waveform biased to be all of positive polarity.

Difficulties arise in expressing error of nonlinearity for there exist many ways to express this error. Figure 3.11 shows the four cases that usually arise. The difference arises in the way in which the ideal (called the "best fit") straight line can be set up. Figure 3.11(*a*) shows the line positioned by the usually calculated statistical averaging method of least squares fit; other forms of line fitting calculation are also used. This will yield the smallest magnitude of error calculation for the various kinds of line fitting but may not be appropriate for how the stage under assessment is used. Other, possibly more reasonable, options exist. Figure 3.11(*b*) constrains the best fit line to pass through the zero point. Figure 3.11(*c*) places the line between the expected 0% and the 100% points. There is still one more option, that where the theoretical line is not necessarily one of the above, yet is the ideal placement, Figure 3.11(*d*).

In practice then, instrument systems linearity can be expressed in several ways. Good certification practice requires that the method used to ascertain the error is stated along with the numerical result, but this is often not done. Note also that the error is the worst case and that part of the response may be much more linear.

The description of instrument performance is not a simple task. To accomplish this fully would require very detailed statements recording the performance at each and every point. That is often too cumbersome, so the instrument industry has developed many short-form statements that provide an adequate guide to the performance. This guide will be seen to be generally a conservative statement.

Many other descriptors exist for the static regime of an instrument. The reader is referred to the many standards documents that exist on instrument terminology; for example, see Reference [3].

3.2 Dynamic Characteristics of Instrument Systems

Dealing with Dynamic States

Measurement outcomes are rarely static over time. They will possess a dynamic component that must be understood for correct interpretation of the results. For example, a trace made on an ink pen chart recorder will be subject to the speed at which the pen can follow the input signal changes.

© 1999 by CRC Press LLC

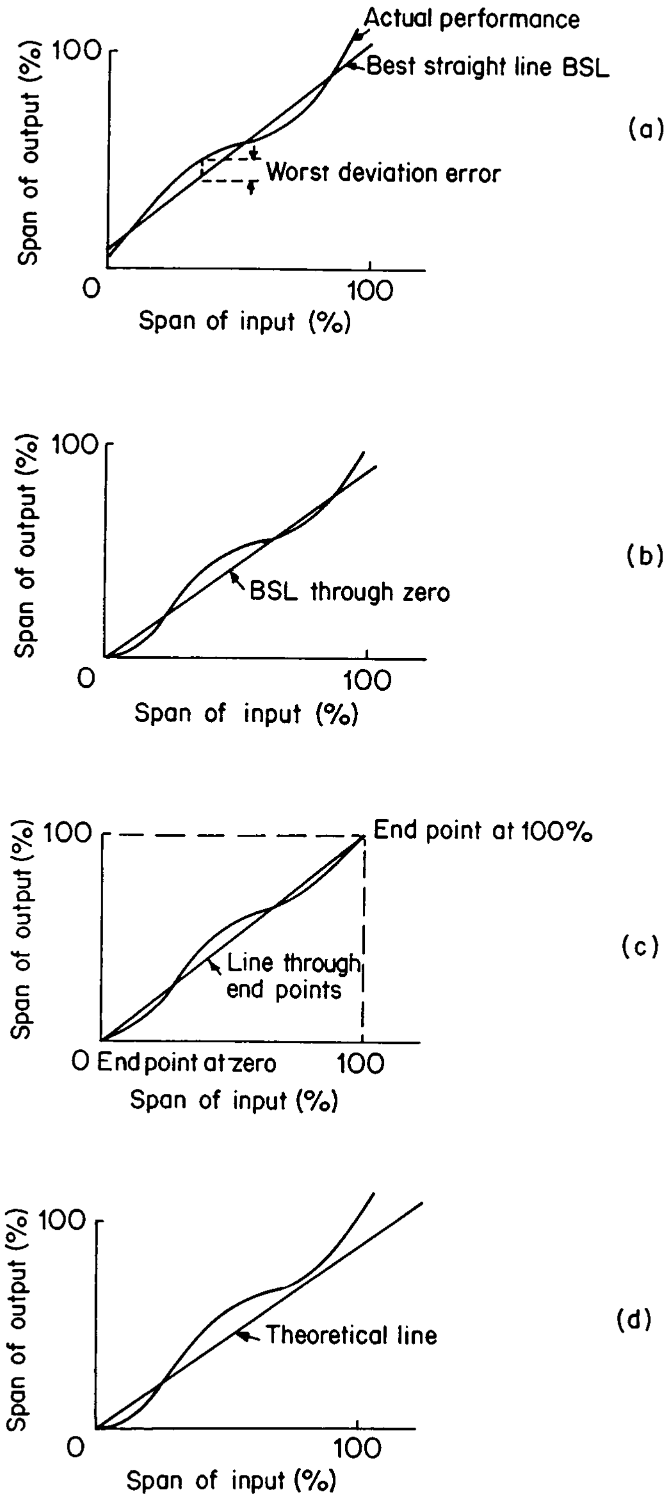

FIGURE 3.11 Error of nonlinearity can be expressed in four different ways: (*a*) best fit line (based on selected method used to decide this); (*b*) best fit line through zero; (*c*) line joining 0% and 100% points; and (*d*) theoretical line. (From P. H. Sydenham, *Handbook of Measurement Science,* Vol. 2, Chichester, U.K., John Wiley & Sons, 1983. With permission.)

© 1999 by CRC Press LLC

To properly appreciate instrumentation design and its use, it is now necessary to develop insight into the most commonly encountered types of dynamic response and to develop the mathematical modeling basis that allows us to make concise statements about responses.

If the transfer relationship for a block follows linear laws of performance, then a generic mathematical method of dynamic description can be used. Unfortunately, simple mathematical methods have not been found that can describe all types of instrument responses in a simplistic and uniform manner. If the behavior is nonlinear, then description with mathematical models becomes very difficult and might be impracticable. The behavior of nonlinear systems can, however, be studied as segments of linear behavior joined end to end. Here, digital computers are effectively used to model systems of any kind provided the user is prepared to spend time setting up an adequate model.

Now the mathematics used to describe linear dynamic systems can be introduced. This gives valuable insight into the expected behavior of instrumentation, and it is usually found that the response can be approximated as linear.

The modeled response at the output of a block G_{result} is obtained by multiplying the mathematical expression for the input signal G_{input} by the transfer function of the block under investigation G_{response}, as shown in Equation 3.5.

$$G_{\text{result}} = G_{\text{input}} \times G_{\text{response}} \tag{3.5}$$

To proceed, one needs to understand commonly encountered input functions and the various types of block characteristics. We begin with the former set: the so-called *forcing functions.*

Forcing Functions

Let us first develop an understanding of the various types of input signal used to perform tests. The most commonly used signals are shown in Figure 3.12. These each possess different valuable test features. For example, the sine-wave is the basis of analysis of all complex wave-shapes because they can be formed as a combination of various sine-waves, each having individual responses that add to give all other wave-shapes. The step function has intuitively obvious uses because input transients of this kind are commonly encountered. The ramp test function is used to present a more realistic input for those systems where it is not possible to obtain instantaneous step input changes, such as attempting to move a large mass by a limited size of force. Forcing functions are also chosen because they can be easily described by a simple mathematical expression, thus making mathematical analysis relatively straightforward.

Characteristic Equation Development

The behavior of a block that exhibits linear behavior is mathematically represented in the general form of expression given as Equation 3.6.

$$\ldots\ldots\ldots a_2 d^2 y / dt^2 + a_1 dy / dt + a_0 y = x(t) \tag{3.6}$$

Here, the coefficients a_2, a_1, and a_0 are constants dependent on the particular block of interest. The left-hand side of the equation is known as the *characteristic equation.* It is specific to the internal properties of the block and is not altered by the way the block is used.

The specific combination of forcing function input and block characteristic equation collectively decides the combined output response. Connections around the block, such as feedback from the output to the input, can alter the overall behavior significantly: such systems, however, are not dealt with in this section being in the domain of feedback control systems.

Solution of the combined behavior is obtained using Laplace transform methods to obtain the output responses in the time or the complex frequency domain. These mathematical methods might not be familiar to the reader, but this is not a serious difficulty for the cases most encountered in practice are

© 1999 by CRC Press LLC

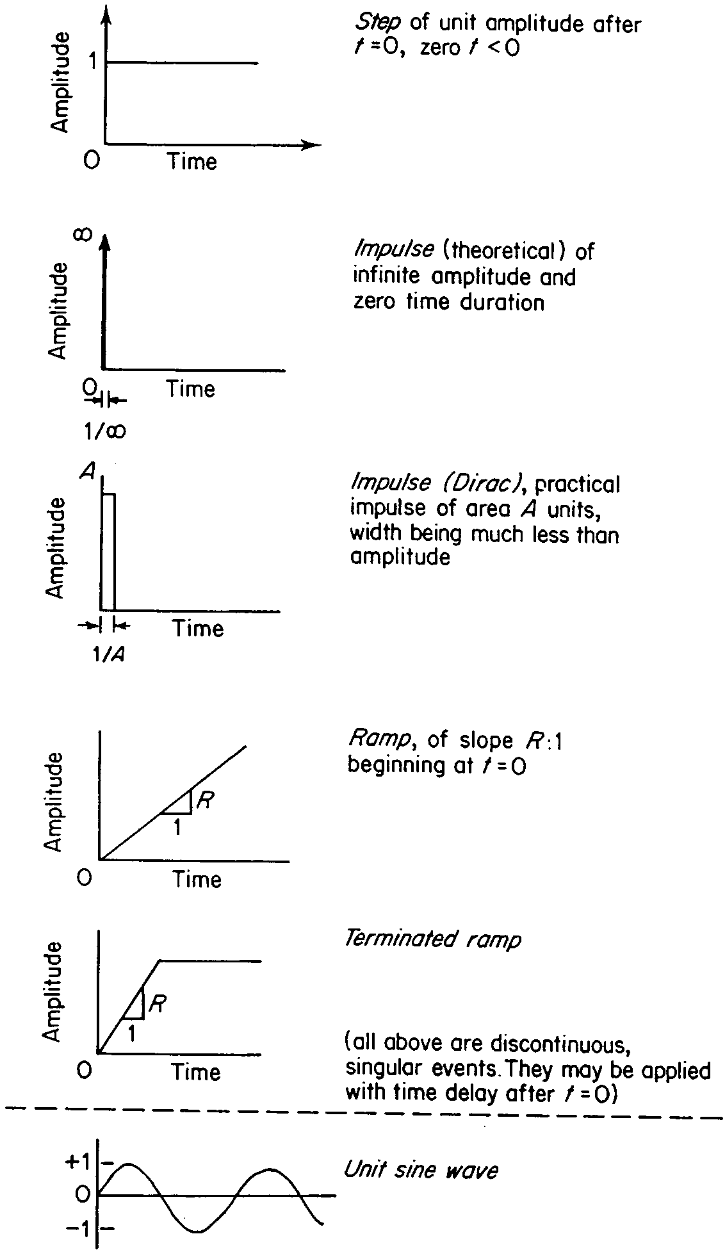

FIGURE 3.12 The dynamic response of a block can be investigated using a range of simple input forcing functions. (From P. H. Sydenham, *Handbook of Measurement Science,* Vol. 2, Chichester, U.K., John Wiley & Sons, 1983. With permission.)

well documented in terms that are easily comprehended, the mathematical process having been performed to yield results that can be used without the same level of mathematical ability. More depth of explanation can be obtained from [1] or any one of the many texts on energy systems analysis. Space here only allows an introduction; this account is linked to [1], Chapter 17, to allow the reader to access a fuller description where needed.

© 1999 by CRC Press LLC

The next step in understanding block behavior is to investigate the nature of Equation 3.6 as the number of derivative terms in the expression increases, Equations 3.7 to 3.10.

Zero order $$a_0 y = x(t) \tag{3.7}$$

First order $$a_1\, dy/dt + a_0 y = x(t) \tag{3.8}$$

Second order $$a_2\, d^2 y/dt^2 + a_1\, dy/dt + a_0 y = x(t) \tag{3.9}$$

*n*th order $$a_n\, d^n y/dt^n + a_{n-1} d^{n-1} y/dt^{n-1} + \ldots + a_0 y = x(t) \tag{3.10}$$

Note that specific names have been given to each order. The zero-order situation is not usually dealt with in texts because it has no time-dependent term and is thus seen to be trivial. It is an amplifier (or attenuator) of the forcing function with gain of a_0. It has infinite bandwidth without change in the amplification constant.

The highest order usually necessary to consider in first-cut instrument analysis is the second-order class. Higher-order systems do occur in practice and need analysis that is not easily summarized here. They also need deep expertise in their study. Computer-aided tools for systems analysis can be used to study the responses of systems.

Another step is now to rewrite the equations after Laplace transformation into the frequency domain. We then get the set of output/input Equations 3.11 to 3.14.

Zero order $$Y(s)/X(s) = 1 \tag{3.11}$$

First order $$Y(s)/X(s) = 1/(\tau s + 1) \tag{3.12}$$

Second order $$Y(s)/X(s) = 1/(\tau_1 s + 1)(\tau_2 s + 1) \tag{3.13}$$

*n*th order $$Y(s)/X(s) = 1/(\tau_1 s + 1)(\tau_2 s + 1)\ldots(\tau_n s + 1) \tag{3.14}$$

The terms $\tau_1, \ldots, \tau_n$ are called *time constants.* They are key system performance parameters.

Response of the Different Linear Systems Types

Space restrictions do not allow a detailed study of all of the various options. A selection is presented to show how they are analyzed and reported, that leading to how the instrumentation person can use certain standard charts in the study of the characteristics of blocks.

Zero-Order Blocks

To investigate the response of a block, multiply its frequency domain forms of equation for the characteristic equation with that of the chosen forcing function equation.

This is an interesting case because Equation 3.7 shows that the zero-order block has no frequency-dependent term (it has no time derivative term), so the output for all given inputs can only be of the same time form as the input. What can be changed is the amplitude given as the coefficient a_0. A shift in time (phase shift) of the output waveform with the input also does not occur as it can for the higher-order blocks.

© 1999 by CRC Press LLC

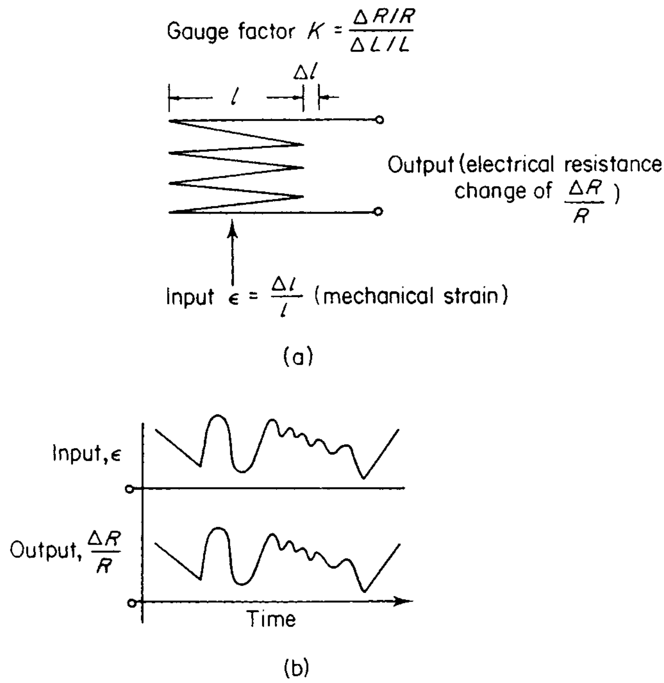

FIGURE 3.13 Input and output responses for a zero-order block: (*a*) strain gage physical and mathematical model; and (*b*) responses. (From P. H. Sydenham, *Handbook of Measurement Science,* Vol. 2, Chichester, U.K., John Wiley & Sons, 1983. With permission.)

This is the response often desired in instruments because it means that the block does not alter the time response. However, this is not always so because, in systems, design blocks are often chosen for their ability to change the time shape of signals in a known manner.

Although somewhat obvious, Figure 3.13, a resistive strain gage, is given to illustrate zero-order behavior.

First-Order Blocks

Here, Equation 3.8 is the relevant characteristic equation. There is a time-dependent term, so analysis is needed to see how this type of block behaves under dynamic conditions. The output response is different for each type of forcing function applied. Space limitations only allow the most commonly encountered cases — the step and the sine-wave input — to be introduced here. It is also only possible here to outline the method of analysis and to give the standardized charts that plot generalized behavior.

The step response of the first-order system is obtained by multiplying Equation 3.12 by the frequency domain equation for a step of amplitude A. The result is then transformed back into the time domain using Laplace transforms to yield the expression for the output, $y(t)$

$$y(t) = AK\left(1 - e^{t/\tau}\right) \tag{3.15}$$

where A is the amplitude of the step, K the static gain of the first-order block, t the time in consistent units, and τ the time constant associated with the block itself.

This is a tidy outcome because Equation 3.15 covers the step response for all first-order blocks, thus allowing it to be graphed in normalized manner, as given in Figure 3.14. The shape of the response is always of the same form. This means that the step response of a first-order system can be described as having "a step of AK amplitude with the time constant τ."

© 1999 by CRC Press LLC

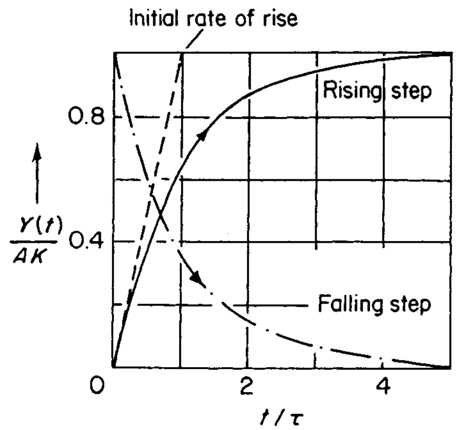

FIGURE 3.14 The step response for all first-order systems is covered by these two normalized graphs. (From P. H. Sydenham, *Handbook of Measurement Science,* Vol. 2, Chichester, U.K., John Wiley & Sons, 1983. With permission.)

If the input is a sine-wave, the output response is quite different; but again, it will be found that there is a general solution for all situations of this kind. As before, the input forcing equation is multiplied by the characteristic equation for the first-order block and Laplace transformation is used to get back to the time domain response. After rearrangement into two parts, this yields:

$$y(t) = \left[AK\tau\omega e^{-t/\tau} \Big/ \left(\tau^2\omega^2 + 1\right)\right] + \left[AK \Big/ \left(\tau^2\omega^2 + 1\right)^{1/2} \cdot \sin\left(\omega t + \phi\right)\right] \tag{3.16}$$

where ω is the signal frequency in angular radians, $\phi = \tan^{-1}(-\omega t)$, *A* the amplitude of the sine-wave input, *K* the gain of the first-order block, *t* the time in consistent units, and τ the time constant associated with the block.

The left side of the right-hand bracketed part is a short-lived, normally ignored, time transient that rapidly decays to zero, leaving a steady-state output that is the parameter of usual interest. Study of the steady-state part is best done by plotting it in a normalized way, as has been done in Figure 3.15.

These plots show that the amplitude of the output is always reduced as the frequency of the input signal rises and that there is always a phase lag action between the input and the output that can range from 0 to 90° but never be more than 90°. The extent of these effects depends on the particular coefficients of the block and input signal. These effects must be well understood when interpreting measurement results because substantial errors can arise with using first-order systems in an instrument chain.

Second-Order Blocks

If the second-order differential term is present, the response of a block is quite different, again responding in quite a spectacular manner with features that can either be wanted or unwanted.

As before, to obtain the output response, the block's characteristic function is multiplied by the chosen forcing function. However, to make the results more meaningful, we first carry out some simple substitution transformations.

The steps begin by transforming the second-order differential Equation 3.6 into its Laplace form to obtain:

$$X(s) = a_2 s^2 Y(s) + a_1 s Y(s) + a_0 Y(s) \tag{3.17}$$

This is then rearranged to yield:

$$G(s) = Y(s) \big/ X(s) = 1/a_0 \cdot 1 \Big/ \left\{\left(a_2/a_0\right)s^2 + \left(a_1/a_0\right)s + 1\right\} \tag{3.18}$$

© 1999 by CRC Press LLC

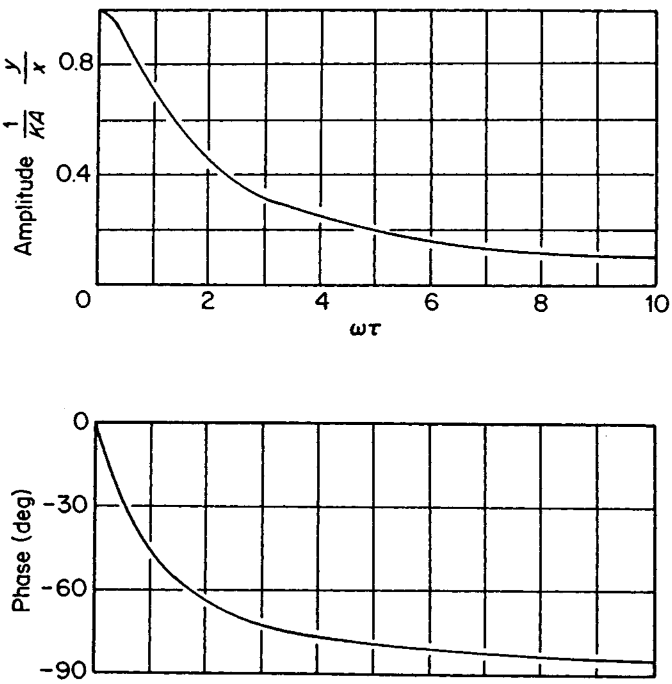

FIGURE 3.15 The amplitude and phase shift of the output of all first-order systems to a sine-wave input is shown by these two normalized curves; (*a*) amplitude and (*b*) phase. (From P. H. Sydenham, *Handbook of Measurement Science,* Vol. 2, Chichester, U.K., John Wiley & Sons, 1983. With permission.)

The coefficients can then be expressed in system performance terms as follows.

Angular natural frequency $$\omega_n = (a_0/a_2)^{1/2} \tag{3.19}$$

Damping ratio $$\zeta = a_1 \big/ 2(a_0 \cdot a_2)^{1/2} \tag{3.20}$$

Static gain $$K = 1/a_0 \tag{3.21}$$

These three variables have practical relevance, as will be seen when the various responses are plotted.

Using these transformed variables, the characteristic equation can be rewritten in two forms ready for investigation of output behavior to step and sine-wave inputs, as:

$$G(s) = K \Big/ \left\{ \left(1/\omega_n^2\right)s^2 + \left(2\zeta/\omega_n\right)s + 1 \right\} \tag{3.22}$$

and then as:

$$G(s) = K \Big/ \left(\tau^2 s^2 + 2\zeta\tau s + 1\right) \tag{3.23}$$

We are now ready to consider the behavior of the second-order system to the various forcing inputs.

First consider the step input. After forming the product of the forcing and characteristic functions, the time domain form can be plotted as shown in Figure 3.16.

© 1999 by CRC Press LLC

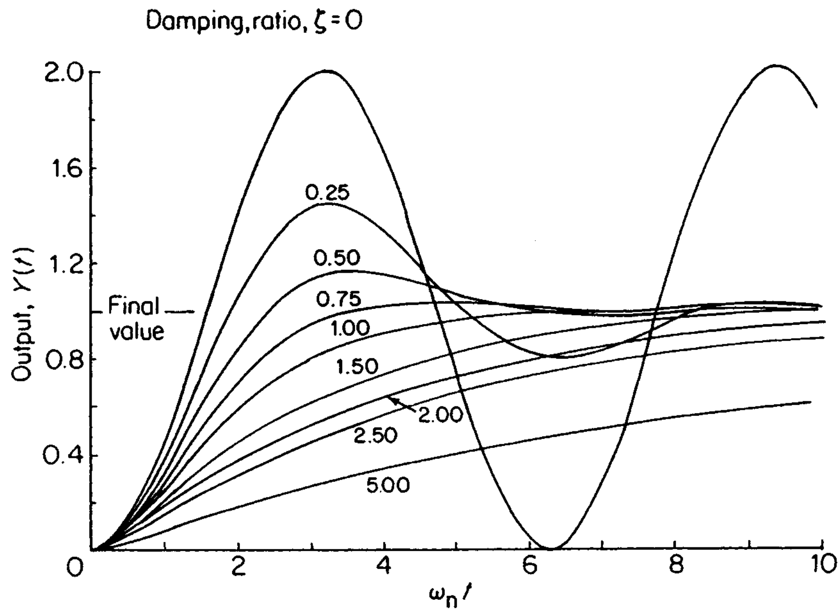

FIGURE 3.16 The response of second-order systems to a step input is seen from this normalized plot. (From P. H. Sydenham, *Handbook of Measurement Science,* Vol. 2, Chichester, U.K., John Wiley & Sons, 1983. With permission.)

This clearly shows that the response is strongly dependent on the damping ratio ζ value. If it is less than unity, it exhibits an oscillatory movement settling down to the final value. If the damping value is greater than unity, the response moves to the final value without oscillation. The often preferred state is to use a damping factor of unity, *critical damping.* The choice of response depends strongly on the applications, for all levels of damping ratio have use in practice, ranging from needing an oscillation that never ceases (zero damping) to the other extreme where a very gradual rate of change is desired.

A similar analysis is used to see how the second-order system responds to the sine-wave input. The two response plots obtained are shown in Figure 3.17: one for the amplitude response, and the other showing how the phase shifts as the frequency changes.

The most unexpected result is seen at the point where the gain rises to infinity for the zero damping state. This is called *resonance* and it occurs at the block's *natural frequency* for the zero damping state. Resonance can be a desirable feature, as in detecting a particular frequency in a radio frequency detection circuit, or it may be most undesirable, as when a mechanical system resonates, possibly to destruction. It can be seen that it is mostly controlled by the damping ratio. Note also that the phase shift for the second-order system ranges from 0 to 180°. This has important implications if the block is part of a feedback loop because as the frequency rises, the phase shift from the block will pass from stable negative feedback (less than 90°) to positive feedback (greater than 90°), causing unwanted oscillation.

More detail of the various other situations, including how to deal with higher orders, cascaded blocks of similar kind, and ramp inputs are covered elsewhere [1].

3.3 Calibration of Measurements

We have already introduced the concept of accuracy in making a measurement and how the uncertainty inherent in all measurements must be kept sufficiently small. The process and apparatus used to find out if a measurement is accurate enough is called *calibration.* It is achieved by comparing the result of a measurement with a method possessing a measurement performance that is generally agreed to have less uncertainty than that in the result obtained. The error arising within the calibration apparatus and process

© 1999 by CRC Press LLC

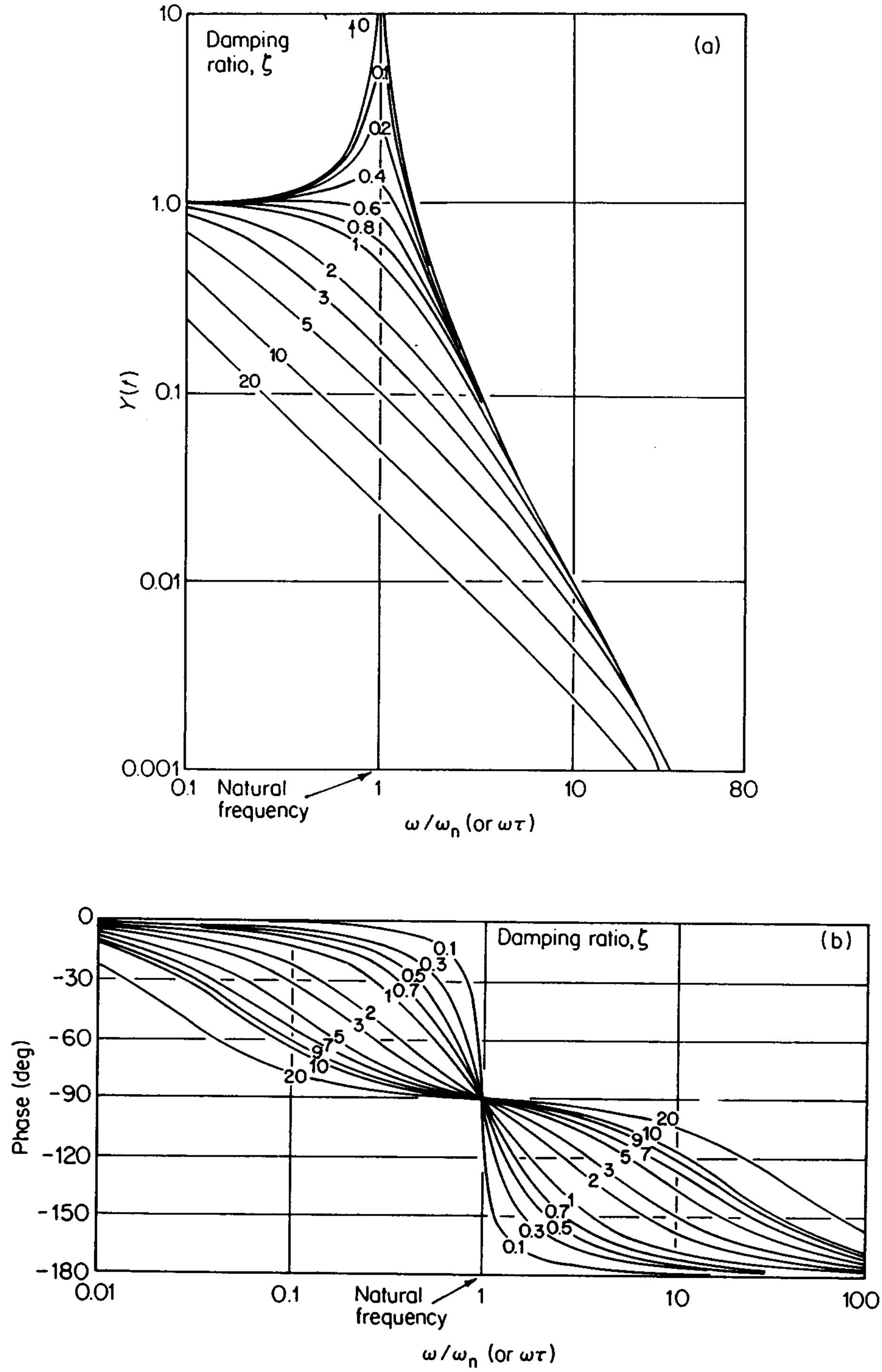

FIGURE 3.17 These two plots allow the behavior of second-order blocks with sine-wave inputs to be ascertained: (*a*) amplitude and (*b*) phase. (From P. H. Sydenham, *Handbook of Measurement Science,* Vol. 2, Chichester, U.K., John Wiley & Sons, 1983. With permission.)

© 1999 by CRC Press LLC

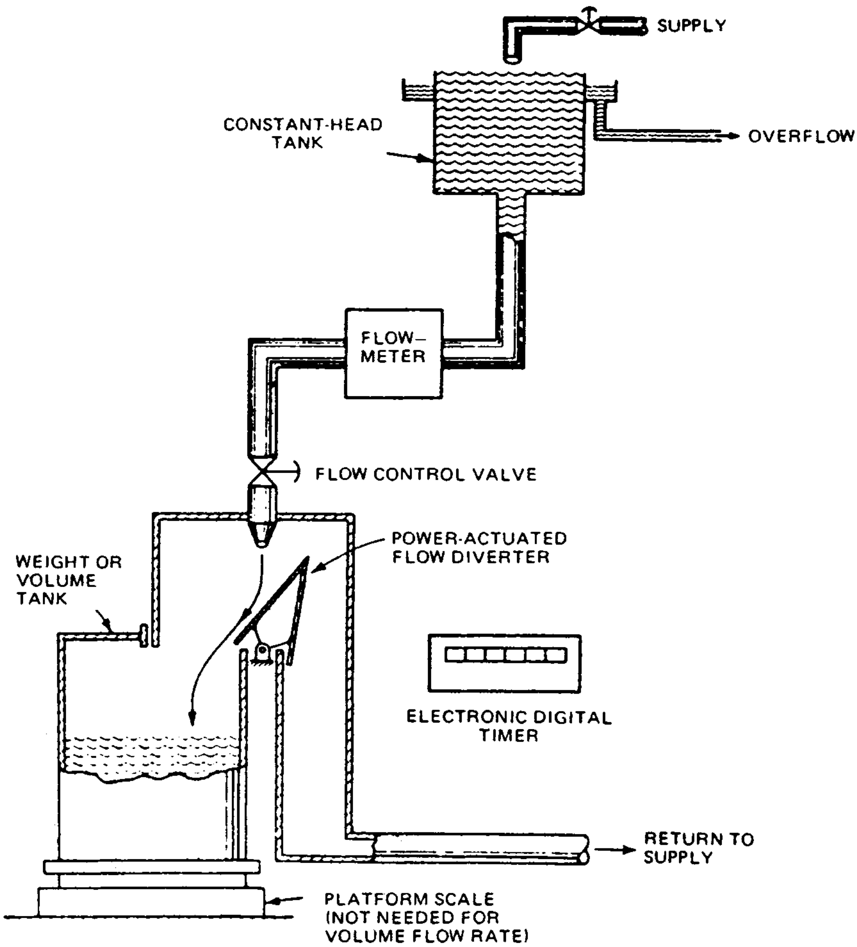

FIGURE 3.18 This practical example illustrates how flow meters are calibrated by passing a known quantity of fluid through the meter over a given time. (Originally published in P. H. Sydenham, *Transducers in Measurement and Control,* Adam Hilger, Bristol, IOP Publishing, Bristol, 1984. Copyright P. H. Sydenham.)

of comparison must necessarily be less than that required. This means that calibration is often an expensive process. Conducting a good calibration requires specialist expertise.

The method and apparatus for performing measurement instrumentation calibrations vary widely. An illustrative example of the comparison concept underlying them all is given in the calibration of flow meters, shown diagrammatically in Figure 3.18.

By the use of an overflowing vessel, the top tank provides a flow of water that remains constant because it comes from a constant height. The meter to be calibrated is placed in the downstream pipe.

The downstream is either deflected into the weigh tank or back to the supply. To make a measurement, the water is first set to flow to the supply. At the start of a test period, the water is rapidly and precisely deflected into the tank. After a given period, the water is again sent back to the supply. This then has filled the tank with a given amount of water for a given time period of flow. Calculations are then undertaken to work out the quantity of water flowing per unit time period, which is the *flow rate.* The meter was already registering a flow rate as a constant value. This is then compared with the weighed method to yield the error. Some thought will soon reveal many sources of error in the test apparatus, such as that the temperature of the water decides the volume that flows through and thus this must be allowed for in the calculations.

© 1999 by CRC Press LLC

It will also be clear that this calibration may not be carried out under the same conditions as the measurements are normally used. The art and science and difficulties inherent in carrying out quality calibration for temperature sensors are well exposed [2].

Calibration of instrumentation is a must for, without it, measurement results may be misleading and lead to costly aftermath situations. Conducting good calibration adds overhead cost to measurement but it is akin to taking out insurance. If that investment is made properly, it will assist in mitigating later penalties. For example, an incorrectly calibrated automatic cement batcher was used in making concrete for the structural frame of a multistory building. It took several days before concrete strength tests revealed the batcher had been out of calibration for a day with the result that the concrete already poured for three floors was not of adequate strength. By then, more stories had been poured on top. The defective floors had to be fully replaced at great cost. More resource put into the calibration process would have ensured that the batcher was working properly.

References

1. P. H. Sydenham, *Handbook of Measurement Science,* Vol. 2, Chichester, U.K.: John Wiley & Sons, 1983.
2. J. V. Nicholas and D. R. White, *Traceable Temperatures,* Chichester, U.K.: John Wiley & Sons, 1994.
3. British Standard Institution, *PD 6461: Vocabulary of Metrology,* London: BSI, 1995.

© 1999 by CRC Press LLC