

1

Measurement and Traceability

1.1 Introduction

Let's start with a thought experiment.

You've just bought this book, but you're concerned about the growing piles of unread books lying around the house. Impulsively you decide that it's time to build a bookshelf, and telephone the local timber merchants and ask them to deliver some suitable timber.

'How much?', they say.

'Eight metres ought to be enough', you reply.

'What's a metre?'

After some discussion and free exchange of ideas, you find that you and the timber merchants do not use the same measures of length. How can you communicate your needs?

Over the millennia, we have tended to use assorted body parts (feet, nails, inches, cubits, hands, palms, yards, etc.) to communicate measures of length. This might work well enough with the timber merchants but there are occasions when something more rigid and precise is required. The time also comes when we must interact with people other than the timber merchants, and eventually with the rest of the world. Then the only practical solution is for us all to share a single standard for each measure.

The concept of traceability has evolved to describe measurements that can be related to a common standard. To make traceability practical on a worldwide scale a huge number of people have to be involved defining the units, disseminating the units through calibration laboratories, manufacturing instruments, setting specifications for the manufacturers, and, finally, using the instruments and measurement results. Consequently, complex and extensive systems have evolved to provide for physical measurement standards, assessments of technical competence and the development of measurement protocols. For those of you whose business is measurement, achieving traceability and working with the measurement community will have a greater impact on your business than any technological advance in measurement technique.

This chapter is primarily background material answering some of the 'what' and 'why' questions of measurement, and describing how to interact with the rest of the measurement community. We begin by delving into measurement theory: what is a measurement, and why do we measure? This is followed by a short history of the evolution of temperature as a physical concept and the primary methods for measuring temperature. The last sections provide an overview of the systems that have evolved nationally and internationally to make traceability practical: the international system

of units, accreditation systems and documentary standards systems. We describe what they do and why they are there.

1.2 Measurement

1.2.1 What is a measurement?

Measurement is one of the most fundamental tasks carried out by humans, or any other species. All knowledge of the world is gained through the senses, and our ability to survive comes from our reactions to that knowledge. Our curiosity, intelligence and self-awareness are adaptations that allow us to prepare for the unexperienced, and our sensors and measuring instruments extend our knowledge of the universe beyond that acquirable from our senses alone.

For those of us working inside corporate organisations survival and risk have slightly different meanings from the biological sense, but whether we or our organisations make measurements for commerce, control or curiosity the ultimate motivation is the same: improving survival and minimising risk.

One way of identifying the reason for making a measurement is to determine how it affects our actions. For commercial measurements, the answer is simple because measurements are the basis of contracts. A buyer agrees to pay for a certain quantity of goods and his or her decision on whether to buy or not depends on the quantity and price. In commerce, measurements affect decisions about sale and purchase.

Measurements are often associated with control or regulatory mechanisms. In air-conditioning systems, temperature measurements determine whether heat flows are increased or decreased. Measurements of rust in cars control their roadworthiness, and measurements of toxicity control the quality of the environment. In each case the measurements precede decisions to increase or decrease, reject or accept, or to prosecute or not prosecute.

With curiosity-driven measurements the decisions are less obvious. In science, experiments are used as the basis for developing and testing theory or models of phenomena, and at each stage of the development a scientist makes a decision: does this work, does that work, what if we do this, etc.? Eighty years after Einstein published his general theory of relativity researchers are still making measurements to test the validity of the theory. Again measurements are made to aid in the making of decisions.

The accepted metrological definition of a measurement is in two parts:

Measurement:

The set of operations having the object of determining a value of a quantity.

Result of a measurement:

The value attributed to a measurand obtained by measurement.

While these two definitions are technically correct, they are so because of a tautology. The dictionary definition of a quantity is something that has a value, and a measurand is the thing being measured. Consequently, these are self-evident statements rather than definitions. More importantly, they are not especially helpful in distinguishing

a measurement from a meaningless assignment of numbers. We offer an alternative definition that applies to both the process and the result:

Measurement (alternative definition):

The symbolic representation of a state, event or attribute to aid in the making of a decision.

This definition highlights three important aspects of measurement not apparent from the first two definitions:

- (1) The results of measurement need not be numeric: grade A, red and sodium are all legitimate measurement results in the appropriate context. One of the most valuable aspects of symbolic representation is that we use the symbols in our models to make predictions. Mathematical models and numeric symbols particularly help to quantify predictions that might otherwise be qualitative or subjective.
- (2) Every measurement has a purpose. This is the distinction we seek between a meaningful measurement and meaningless assignment of numbers. In a great many measurements, especially outside the calibration laboratory, the purpose influences the design and outcome of the measurement. Consequently, measurement results may have meaning only within the context of that purpose. Results used for other purposes or gathered without purpose are potentially dangerous.
- (3) Decisions are associated with risks and rewards. This highlights the need to know the uncertainty in a measurement in order to assess the risk or optimise the measurement.

These three points are the main themes of Sections 1.2.2 to 1.2.4.

Exercise 1.1

Consider some of the measurements you make. What decisions do they influence?
What are some of the risks and rewards associated with those decisions?

1.2.2 Measurement scales

In this section, we look at some of the basic systems for the symbolic representation of measurement results: our measurement scales. Table 1.1 provides a coarse summary of the various types of measurement scale that we use. The classification given in the table is based on the algebraic properties underlying the measurement scales. While there are other systems of classification we have chosen this one in order to highlight the fundamental limitations associated with the different systems of representation.

Nominal scales

Measurements on nominal scales are essentially a naming, and often the results do not look like measurements at all. An example familiar to most is the number on a football player's shirt. In many sports, the number on a player's shirt indicates the position the

Table 1.1 A classification of measurement scales based on possible mathematical operations

Scale type	Description	Operations	Examples
Nominal	A renaming; can establish equivalence	=	Colours (red, blue) Team numbers (e.g. football) Stellar spectral types (O,B,A,F,G,...)
Ordinal	Can establish order	= < >	Moh hardness Rockwell hardness Beaufort wind scale Fahrenheit and Celsius scales
Interval	Can establish meaningful differences	= < > + -	Date, time of day Year Latitude and longitude Centigrade temperature scale
Metric or ratio	Can establish meaningful ratios	= < > + - $\times \div$	All SI scales, e.g. length Mass Frequency Thermodynamic temperature
Counting or natural	Counts of objects or events, an integer metric scale	= < > + - $\times \div$	Apples, buses Birthdays

sports person plays on the field. In Rugby Union, for example, the numbers 11 and 14 refer to the left and right wings. But the numbers carry no numeric meaning: it is meaningless to say that a number 14 player is any better than a number 11, or that the number 14 is twice as good as the number 7. However, it is possible to establish equivalence: a number 14 in one team plays in the same position and requires the same skills as the number 14 in another team. In this case, the measurement is performed when a coach determines the player's position.

Other examples of nominal scales include colours, the names of plants, and the classification of chemicals such as alcohols and acids. Indeed most of our language is based on nominal assignments. All of our verbs, nouns and adjectives are verbal and written symbols for actions, objects and attributes. One of the key aspects of nominal scales is that each classification or naming must have its own standard or definition. Thus, for example, each of us has to learn what red and blue or apples and pears are before we can make use of the symbols.

Ordinal scales

As its name implies, results on ordinal scales convey some sense of order. As with nominal scales, ordinal results need not be numerical. Perhaps one of the most famous ordinal scales, amongst English-speaking parents and children at least, is the one used by Goldilocks at the house of the three bears: too little, just right, and too much. This is the very same scale underlying many commercial transactions. A more typical example is the Moh hardness scale (Table 1.2), a scale once used by mineralogists to help identify minerals on the basis of hardness. The scale is constructed so that each mineral listed is harder than those below it. In this way the mineralogist can determine

Table 1.2 Moh's scale of hardness

Hardness	Mineral standard
10	Diamond
9	Sapphire
8	Topaz
7	Quartz
6	Feldspar
5	Apatite
4	Flourspar
3	Calcite
2	Gypsum
1	Talc

the hardness of an unknown mineral by determining which minerals scratch it. Strictly speaking, the scale measures resistance to scratching rather than hardness.

Other examples include the Beaufort wind strength scale, the Mercalli earthquake intensity scale, examination grades, credit ratings, library codes, and most of the early temperature scales. As can be seen from the examples, ordinal scales are characterised by several standards or definitions, one for each of the defined points on the scale.

In many cases ordinal scales also have a specified interpolating instrument that makes it possible to assign values in between the defined points. A very large number of chemical measurement scales are constructed this way, for example using standard solutions and a spectrometer to interpolate between the standards. The International Temperature Scale of 1990 (ITS-90) is also an ordinal scale, where the defined points are the various melting, freezing and triple points of pure substances, and the interpolating instruments include platinum resistance thermometers, vapour-pressure thermometers and radiation thermometers. With many ordinal measurements, the response of the interpolating instrument is known to be non-linear but it is still possible to tell when one sample has a greater concentration of a particular compound than another, or higher temperature than another.

Interval scales

Interval scales are those that are known to be linear in some fundamental sense, and are the simplest scale type to allow meaningful comparison of differences. Interval scales typically have an arbitrary zero. Familiar examples include the latitude and longitude scales, which are used to determine position on the surface of the earth. The longitude scale requires two standards to define it: the position of the zero, which is arbitrarily chosen to be Greenwich, and the number of degrees in a full revolution of the earth, which is arbitrarily chosen to be 360. It is possible to compare changes in longitude meaningfully, or to add and subtract intervals of longitude, but it is still not meaningful to talk about ratios. Statements such as 'a country at 40 degrees of longitude is twice the country at 20 degrees of longitude' are nonsense.

Other examples of interval scales include all of the time scales that we use to tell the time of day, date and year, and the 4 mA to 20 mA current loop representation used by many industrial instruments (a symbol need not be a squiggle on paper). One of the earliest thermodynamic temperature scales, the centigrade scale, was an interval scale based on the definition of the melting and boiling points of water at 0 °C and 100 °C

respectively. Because interval scales are the first that enable us to talk meaningfully about intervals, these are the first scales that allow us to do normal statistics, that is to calculate means and standard deviations. On more primitive scales, we have to adopt different statistical techniques that do not require the concept of meaningful intervals.

Metric scales

Metric scales are those that have a natural zero. On such scales, we can usefully talk about ratios and fractions. Metric scales include all of the familiar SI scales of length, mass, thermodynamic temperature, etc. On the mass scale, for example, we know exactly what zero mass means, we can add and subtract mass, and we can talk meaningfully about dividing a mass into fractions or doubling and tripling a mass. The key attribute of metric scales that distinguishes them from other scales is that only one standard is required to define each completely. The mass scale is defined in terms of the prototype kilogram stored in a safe in a basement of the Bureau International des Poids et Mesures (BIPM) in Paris. All other measurements reported on the mass scale are expressed as ratios with respect to the kilogram. The standard used to define the scale is known as the metric or the unit of the scale. Metric scales are also known as ratio scales, and the literal translation of the word metrology, from the Greek *metrologia*, is the study of ratios.

Perhaps the most important of our metric scales is also one of the earliest: the natural or counting scale. Early in our cultural development we learned to count apples, oranges, etc. However, it was apparently not until the time of the Greek mathematicians that it was recognised that numbers had properties independent of the objects. That is, the problem of ‘2 apples plus 3 apples’ is fundamentally the same as ‘2 oranges plus 3 oranges’, and can be generalised without reference to any object. However, we must remember that the measurement scales for counting oranges and apples are different because they have different metrics, one orange and one apple respectively, and one cannot take one apple from two oranges and obtain a meaningful result.

The log-ratio scales form a special class of interval scales that are actually based on metric quantities. Because of the very large range of values encountered, it is often convenient to transform metric measurements to a logarithmic scale. These scales are typically constructed as

$$\text{value on log scale} = \text{constant} \times \log(\text{value}/\text{reference value}).$$

There are two definitions required to define a log-ratio scale: the multiplying constant and the reference value. Examples of such scales include the various decibel scales, the visual magnitude of stars, and the Richter scale for the energy dissipated in earthquakes. On these scales equal intervals correspond to constant multiplying factors of the underlying metric quantity. An interval of 10 dB corresponds to a 10 times increase in power, five steps of visual magnitude correspond to 100 times decrease in the brightness of stars, and two steps on the Richter scale correspond to a 1000 times increase in the energy dissipated in an earthquake.

The progression of scales given above suggests that as the nature of quantities and measurements becomes well understood, the associated scales evolve towards metric scales. Science begins with classification — stamp collecting as Lord Rutherford called it. This evolutionary trend is common, but it is not universal. Some scales can never be

metric: colour will always be a three-dimensional scale based on two interval quantities and one metric quantity, and the Rockwell hardness scales will always be ordinal scales. Also, not all nominal scales are primitive. The models and mathematics underlying the current descriptions of fundamental particles, with demonstrably nominal attributes like beauty, charm, top and bottom, are amongst the most sophisticated that we have developed. Similarly the models and mathematics underlying chemical reactions, which are described by symbols like NaCl and H₂O, are also sophisticated.

The influence of scale type on traceability

In order to communicate results unambiguously it is necessary for each of us to share the same scale for a quantity and to have access to the standards that define the scale. For metric scales the traceability problem is relatively simple: all measurements have to be related to a single standard. For the other scale types, the traceability problem can be more complicated because more standards are required.

Many interval scales can be expressed in terms of metric quantities, so the traceability problem is not too difficult. The log-ratio scale, for example, requires a definition of the multiplying constant, which can be defined without error, and a reference value, which in most cases takes the place of the unit on metric scales. All of the time scales (time of day, year, etc.) rely on measurements of time interval (a metric quantity) and an arbitrarily defined zero. Angle scales, such as latitude and longitude, also rely on angle interval and an arbitrary zero.

Ordinal scales are the most problematic in respect of traceability. They require a minimum of two standards, and in many cases require an approved or specified interpolating instrument. To realise the Moh hardness scale the mineralogist has to carry samples of the 10 minerals that define the scale. A large number of measurements based on calibrated scales are in fact carried out on ordinal scales, and in particular many chemical measurements fall into this category.

Nominal scales typically have the greatest number of standards associated with them, usually one for each possible category on the scale. The standards may be descriptive or based on artefacts such as standard reference materials. At one time, for example, there was a descriptive definition for each of the elements in the periodic table, based on the distinguishing chemical and physical properties. Nowadays the definitions of the elements are based on the number of protons in the nucleus of an atom, which is derived from the natural scale.

The influence of scale type on treatment of uncertainty

According to the ISO *Guide for the expression of uncertainty in measurement* (ISO Guide), an uncertainty is the range of values that may reasonably be attributed to a measurand. Just how one goes about characterising a range depends on the scale type. The fewer mathematical operations that are permitted on a scale the fewer options there are available for statistical analysis. Table 1.3 summarises the various statistical operations available on the different measurement scales.

To clarify the meaning of uncertainty on nominal scales we have to remember that results are used to make decisions, and when we characterise uncertainties we are in fact concerned about the likelihood of making wrong decisions. The problem with

Table 1.3 The options available for characterising the dispersion of results (uncertainties) depend on the mathematical options available. Each scale has available the measures of average and dispersion as indicated and those of the scales above it

Scale type	Allowable operations	Statistical average	Statistical measure of dispersion
Nominal	=	Mode	Non-parametric
Ordinal	= < >	Median	Percentiles
Interval	= < > + -	Arithmetic mean	Standard deviation
Metric	= < > + - \times \div	Geometric or harmonic mean	Per cent deviation

nominal scales is that there are many ways of being wrong. Chemical tests for lead, for example, may confuse lead with other heavy metals: mercury, tin, cadmium, antimony and other neighbours of lead in the periodic table. The consequences of failing to identify the other metals correctly may all have different consequences, depending of course on the purpose of the measurements. With measurements on nominal scales it is usually necessary to consider every possible outcome of the measurements, so risk and uncertainty analyses can become complex.

With ordinal scales, the problem becomes much simpler because measurements can be wrong in only one of two ways, too big or too small (this is the Goldilocks scale), and sometimes only in one way (pass–fail). Risk is then evaluated strictly in terms of the distribution of probabilities for particular outcomes, which can usually be characterised simply in terms of percentiles.

With metric and interval scales, the distributions of possible results can often be expressed in terms of a model based on metric parameters. Thus all of the classical distributions, such as the normal, geometric, chi-square, Poisson, binomial, etc., can be used to characterise uncertainty. In this way, the description of the distributions can be rendered in terms of one or two numbers. This greatly simplifies risk and uncertainty analysis.

With metric scales, an additional possibility is available, namely geometric or harmonic analysis, which is based on distributions measured in terms of ratio rather than interval. An analysis of quantities measured on log-ratio scales using interval statistics is effectively a ratio analysis of the underlying metric quantity. Note that the ISO Guide strictly applies only to interval and metric scales, since the concepts of an arithmetic mean and standard deviation depend on meaningful measures of difference. However, many ordinal scales approximate metric scales or are sufficiently linear over small ranges to be treated as interval scales for the purposes of statistical analysis. Chapter 2 gives a detailed treatment of uncertainty in measurement as it is applied to metric and interval scales.

Exercise 1.2

Consider the following examples where numbers are assigned to objects or states. In what context could they be interpreted as measurements? Consider the numbers on a roulette wheel, street numbers, the ‘seed’ used to start a random number generator, a car numberplate, and a musical note indicated on a musical manuscript. To what type of measurement scale do they belong?

1.2.3 The problem of definition and the importance of purpose

Consider the statement

$$\text{water temperature} = 20^{\circ}\text{C} \pm 1^{\circ}\text{C}.$$

The description ‘water temperature’ is a shorthand statement for something like ‘the temperature of the chilled water on the outlet side of heat exchanger no. 2 on the Upside Downs site of the Moojoose Dairy Company’. The problem of clearly defining the measurand is called the *problem of definition*, and has two parts, one simple and one rather subtle and complex.

The first and simplest part of the problem of definition relates to the identification of the quantity measured. In principle, as given in the example above, all that is required is to provide sufficient information to allow the measurement to be repeated. Influences may also have to be specified, for example the operating conditions of the plant and the time of day that the measurement was made.

The second and most difficult part of the problem of definition relates to the technical definition of the attribute that is being measured. In this case, what do we mean by temperature? As we shall see in Chapter 4, the temperature of a system is strictly defined only in conditions of *thermal equilibrium*, that is no net flow of heat between any of the components of the system. The catch in the tail of this definition is that normally we are only interested in temperature because we want to understand something about the flow of heat from the system.

In the calibration laboratory our calibration baths and furnaces are designed to have a controlled volume that is in thermal equilibrium. A measure of the temperature has only one possible meaning and there can be no confusion. In contrast, in some industrial situations the system is so far from thermal equilibrium that different thermometers read different temperatures. A fluorescent tube is a good example. The electron temperature may be $30\,000^{\circ}\text{C}$, and the colour temperature of the radiation from the tube is about 5600°C . Yet the tube is cool enough to touch. In cases where there is no thermal equilibrium the concept of temperature is at the very least ambiguous, at worst meaningless. Yet any thermometer immersed into a non-equilibrium system will indicate a temperature. In order to make a meaningful measurement we have to understand the purpose of the measurement.

Heat is transported by conduction, convection and radiation. The thermal interaction of any system with a second system depends on the relative contributions of the different modes of heat flow. (This topic is covered in detail in Chapter 4.) If we are interested in a temperature that describes the thermal conditions experienced by the second system then the only meaningful way to define a temperature is to use a thermometer that duplicates the thermal conditions obtained when the second system is placed in thermal contact with the first. The definition of temperature is then determined as much by the thermometer as the system of interest. That is, the meaning of temperature is determined by the purpose of the measurements. With the fluorescent tube, the design of the thermometer and the results of the measurement depend on whether we are interested in the current density at the electrodes, the colour rendering properties of the lamp, or whether there is a burn risk to human fingers.

In the most precise temperature measurements, the insertion of a thermometer usually causes a measurable disturbance. Examples include the heat dissipated by resistance thermometers, the loss of radiation through apertures in blackbody cavities, and heat losses along the stems of all insertion thermometers. In most of these cases the system remains very close to equilibrium and the physics of the measurement process is well enough known to allow us to model the effects and apply corrections, or to enable the measurement to be designed so the effects are negligible. This topic is covered in detail in Chapter 4.

Even in large industrial plants modelling can be an effective solution to the problem. For example, sections of the wall of a large chemical reactor could be considered to have zero net heat flow; for example, a 100 kW flux into the reactor wall is balanced by a 100 kW flux out of the wall. So long as the thermometer is smaller than the volume over which the temperature can be said to be constant we can measure a meaningful temperature.

Where measurements take on high monetary value or are associated with high risks, it is important that the technique employed be acceptable to all parties. This is the rationale for many documentary standards, not just in temperature measurement but measurement in general. A particular example might be a safety standard where the temperature of a surface is measured using a thermocouple mounted in a ‘standard finger’, which duplicates the thermal properties of the human finger. In this way, the ‘temperature measurement’ has little utility as a measure of temperature but significant utility as an assessment of burn risk.

Measurements with definition problems are often the source of great argument. Unfortunately most scientists are passively taught to ignore the possibility of definition problems. Physicists in particular are taught that the aim of their science is to eliminate human subjectivity from their models of the universe. This is an admirable aim but there follow the inevitable problems of making conceptual definitions and models useful. To a theoretical physicist a length is the distance between two points. To a metrologist the most accurate measure of the length of a gauge block is the mean distance at 20 °C between the two end-planes, which are ideally parallel, with one plane defined optically and the other defined by a monolayer of oil and mechanical interference with a flat surface of the same mechanical finish. The first definition is conceptual, the second is practical and driven by a purpose: the need to disseminate practical standards of length at the least uncertainty. Thus for all of our measurements the quantities we measure are to a degree approximations to some conceptual ideal, with the non-ideal aspects of the measurement managed or controlled according to a purpose. Usually the further we get from the calibration laboratory the less control we have over the non-idealities or influence variables, and the more difficult our definition problems become.

The telltale sign of a definition problem is a measurement where the result seems to vary with the measurement technique. Such measurements tend to be subjective and a source of argument until the purpose has been properly identified. In thermometry, surfaces, gases and flames are particular problems.

Note that measurements made for one purpose may not be useful for another purpose. While documentary standards can be useful guides for factors affecting a particular measurement, blindly applying standards to measurements for which they

were not designed can lead to misleading or valueless results. Similarly, taking measurements of some handy attribute ‘just in case’ can also be risky if the measurement is later found to be ill suited to the required purpose.

Exercise 1.3

Think about alternative definitions of length and how they might depend on purpose. For example, the length of a pendulum, the length of an athlete’s jump in the long jump pit, the distance between Paris and Beijing by air or road. How do they compare to the theoretical physicist’s definition?

1.2.4 Decision, risk and uncertainty

All measurements eventually contribute to a decision, and associated with each decision are risk and reward. In this section we take a very brief look at the evaluation of risk, as this can affect the design and interpretation of measurements. The simplest cases occur when a single measurement leads to a pass or fail decision. In complex cases, multiple measurements contribute to a figure of merit that is the basis of decisions. In every case, the quality of the decision is affected by the uncertainty in the measurement. Figure 1.1 shows one example and Table 1.4 shows the possible outcomes of decisions based on the measurement.

Once the consequences of a decision have been identified, the decision process and measurement can be modified to maximise the rewards and minimise the risks. In order to avoid the high costs associated with the ‘incorrect-pass’ risks (Table 1.4) the usual strategy is to increase the pass–fail criterion, as in Figure 1.1. For this reason, industries supplying goods on the basis of net weight usually overfill their containers.

The required increase in the pass–fail criterion is determined by the uncertainty in the measurements. Thus by reducing the uncertainty one can also reduce risk or increase

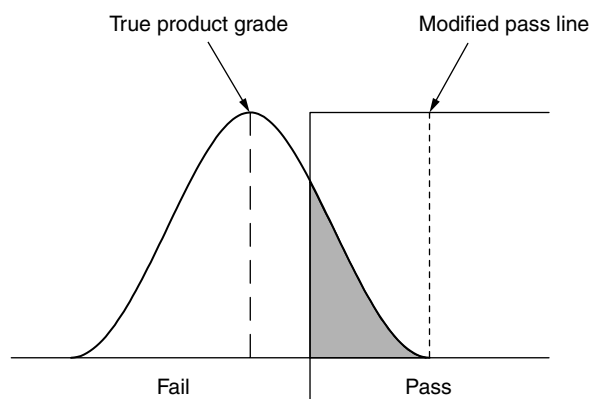


Figure 1.1 A simple decision made on the basis of a single measurement. If the true value is as indicated then, because of the uncertainty in the measurement, the result of the measurement may be anywhere under the curve. The shaded area indicates the probability of making a wrong decision

Table 1.4 The possible outcomes of decisions based on uncertain measurements

	Pass	Fail
Correct	Correct identification of satisfactory product or plant operation. This is where the company makes most of its profit	Correct identification of poor-quality product or sub-optimal operation. This eliminates risk of prosecution, plant failure or customer dissatisfaction
Incorrect	These decisions carry a high external risk or cost, possible plant failure, prosecution for substandard goods, or consumer dissatisfaction	These decisions carry an internal cost, often associated with unused plant capacity, wasted product, unnecessary reprocessing, or low prices for second-grade product

the rewards. In some cases the benefits of improved measurements can vastly outweigh the costs. Examples of this are found in the petroleum industry where measurements are made to 0.01 °C to enable thousands of tonnes of aviation fuel to be classified at the highest grade. This phenomenon is common where pass–fail criteria are defined or regulated, and large quantities of product are involved.

Commonly the most costly wrong decisions are associated with catastrophe: plant failure or huge costs from which there is no recovery. In these cases, the optimisation is straightforward: keep well away from the catastrophe. It is also common for industries to forget the internal costs associated with ‘incorrect–fail’, a reason for the emphasis on this quarter in quality management systems. From the measurement perspective the key factor that makes any optimisation possible is knowledge of the uncertainty in the measurement. In Chapter 2 we give a guide to expression of uncertainty in measurement. Readers interested in the optimisation of decision making should consult books on game theory, and may find risk analysis software available for spreadsheets useful.

Exercise 1.4

Do a risk benefit analysis on a temperature measurement system used for the temperature control of a baking oven. The decision table for this example is more complex than Table 1.4 since the measurement has three outcomes too high, just right and too low instead of the two-outcome pass–fail example of Table 1.4.

1.3 Temperature

In this section, we describe the evolution of temperature measurement from the perspectives of a classical historian, a physicist and a meteorologist. The main purpose is to provide a historical background for subsequent chapters and to illustrate the principles outlined in previous sections. The history of temperature measurement is a good example of the evolution of the measurement of a poorly understood ordinal quantity to a metric quantity founded on sound physical principles. The differences between the physicist’s and the meteorologist’s approaches also highlight differences between

scientific and applied measurements, and in particular the importance of purpose and accepted measurement protocol in the design of applied measurements.

1.3.1 The evolution of the temperature scale

The very first record of a temperature scale belongs to the Greek Galen (AD 130–200) who identified eight degrees of *temperamentum*, which he used to characterise the temperament of his patients and the effects of his medicines. He also defined a neutral temperature based on equal mixtures of boiling water and ice. However, he did not have a thermometer so the temperature was presumably assessed with the physician's hand. A more sophisticated version of Galen's scale of temperament surfaced again to help physicians in the sixteenth century.

In the late sixteenth century the first thermoscopes appeared (see Figure 1.2). These consisted of a glass bulb attached to a thin tube immersed in water. It is not certain who the inventor was, perhaps Galileo, but it seems likely that he was inspired by the pneumatic experiments of the Greeks Philo and Hero that were carried out in the first and second centuries BC. Thermoscopes were used for 50 years or so for both medical and meteorological experiments.

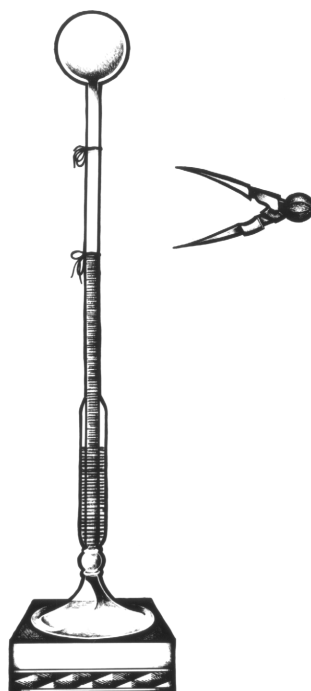


Figure 1.2 An early air thermometer or thermoscope. The thermoscope consists of a glass bulb with a long capillary attached immersed in water or oil. With an increase in the temperature, the air in the bulb expands lowering the fluid level. The scale was marked using two fixed points has indicated by the pieces of string, and interpolated using a pair of dividers to measure intermediate steps or degrees

Ferdinand II of Tuscany made the first sealed thermometer using wine spirit in 1641. This was the first device that we would recognise as a thermometer and was a major advance: a liquid sensor sealed against atmospheric pressure variations and evaporation, and degrees of temperature permanently marked on the stem. Subsequent developments of thermometers focused on improved methods of manufacture, especially on the choice of thermometric fluid and glass, and the method of fixing the scale.

By the early eighteenth century most liquid-in-glass thermometers were stable and had a reproducible scale. Typically the scales were marked using the temperatures of fixed points, such as melting snow, body temperature and boiling water (see Figure 1.3). The scale was then divided into a convenient number of steps or degrees. These scales are clearly ordinal scales with fixed points, and with the temperatures in between determined by interpolation using the expansion of the thermometric fluid, usually mercury. The familiar scales of Fahrenheit and Celsius are good examples of these types of scales. Elegant experiments involving the mixing of known volumes of hot and cold mercury, or hot and cold wine spirit, enabled experimenters to establish that mercury is a more linear thermometric fluid than spirit, but that both are non-linear to some extent.

In the late eighteenth century Gay-Lussac and Charles, building on the ingenious work of Amontons 100 years earlier, were both able to demonstrate that the thermal expansion coefficients of different gases were almost identical. Although Amontons had suggested that the linear expansion of gases with temperature implied that only one fixed point was required to calibrate a thermometer (i.e. to establish a metric scale), and a number of experiments determined values for absolute zero (-273.15°C), the suggestion was not adopted. Chappuis, working at the BIPM, refined gas thermometry further. He had been charged with the responsibility of calibrating a set of mercury-in-glass thermometers by gas thermometry. During a series of remarkable studies comparing temperatures determined using different gases he showed that the gas thermometer did in fact have a small gas species dependence, and that a scale based on hydrogen, although not ideal, was probably accurate to better than 0.01°C . In 1889 the Conférence Générale des Poids et Mesures (CGPM), at its first meeting, adopted the first official temperature scale, the ‘normal hydrogen’ scale. Because of the known dependence of the scale on the non-ideal properties of hydrogen the initial filling pressure of the thermometer was also specified. This scale was still not a metric scale; instead the scale was defined by fixing the interval between the ice point and steam point to be 100°C . The scale was an ordinal approximation to an interval scale based on two fixed points and an almost linear interpolating instrument.

The first proposals to use the variation of electrical resistance to measure temperature came from Davy in 1821 and later Siemens in 1861. However, it took the elaborate experiments and refinements in the construction of platinum resistance thermometers by Callendar to get the resistance thermometer accepted. By comparing the platinum resistance thermometer with the gas thermometer, Callendar and others were able to show that platinum has a parabolic or quadratic characteristic. In 1899 Callendar proposed a temperature scale that would be more practical than the ‘normal hydrogen’ scale, based on three fixed points: the ice point, the steam point and the melting point of sulphur. The scale was defined by fixing the interval between the ice point and steam point to be 100°C , with the sulphur point defined to be 444.5°C as determined by a gas thermometer calibrated at the other two points. Callendar’s proposal, again an ordinal

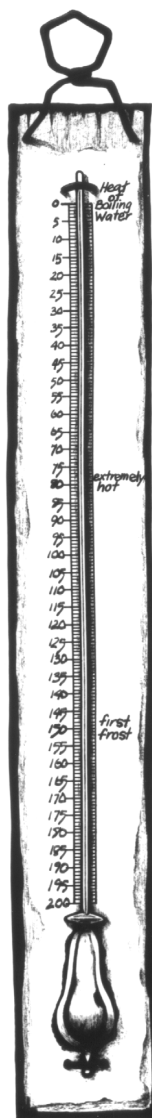


Figure 1.3 A thermometer from the mid-eighteenth century. It has a large bulb because of the large capillary. Note too the inverted scale, which was common in cooler countries where there was more interest in the number of degrees of cold than in the number of degrees of heat

approximation to an interval scale, was extended to a wider range of temperatures using more fixed points and was adopted in 1927.

The latter half of the nineteenth century saw the beginning of the golden age of physics and with it the development of the sciences of thermodynamics and statistical mechanics. Despite these theoretical developments giving meaning to temperature and suggesting a metric temperature scale, it took approximately 100 years before a metric scale was formally adopted. In 1960 the unit of temperature, the kelvin, was defined

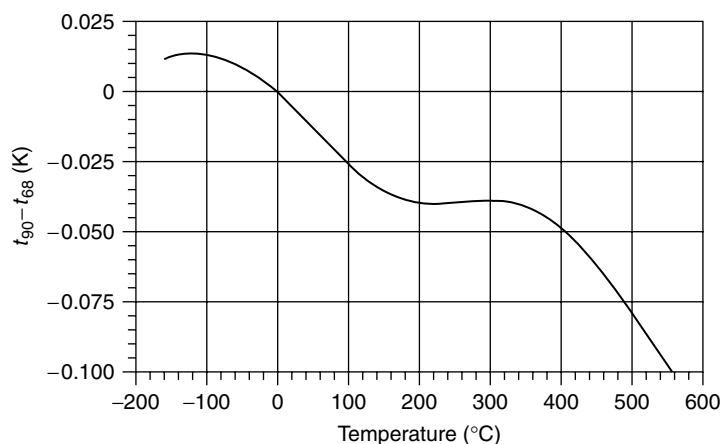


Figure 1.4 The differences between ITS-90 and the earlier scale IPTS-68

as $1/273.16$ of the temperature of the triple point of water. There was also a change in the way temperatures were reported. Prior to 1960, measurements on the temperature scale were reported as degrees Kelvin ($^{\circ}\text{K}$), meaning ‘steps’ on Kelvin’s scale. Since 1960 measurements have been reported as numbers of kelvin, kelvin being the unit for the metric scale.

The advantage of Callendar’s platinum resistance scale was that it was more highly reproducible, simpler and more practical to realise than a thermodynamic scale based purely on a gas thermometer. Even today, thermodynamic measurements prove to be extraordinarily difficult, and may cost as much as 20 person-years of effort to achieve accuracies an order of magnitude short of that necessary to support scientific research and commerce. The most practical solution is to adopt a ‘wire scale’, a scale based on highly reproducible thermometers, calibrated at fixed points, for which the temperatures have been determined by thermodynamic means.

As we have already noted the first wire scale was adopted in 1927 and covered the range from -190°C upwards. Since then there have been revisions in the international temperature scale occurring in 1948, 1968 and 1990. These revisions have provided improvements in respect of closer approximation of the thermodynamic temperature, improved interpolating equations, extensions to lower temperatures, and greater accessibility for users. Figure 1.4 shows the differences between ITS-90 and the International Practical Temperature Scale of 1968 (IPTS-68), and equations and a table of difference are given in Appendix A. Chapter 3 discusses ITS-90 in detail.

1.3.2 Thermodynamic temperature

For most materials temperature can be considered to be a measure of the density of heat in a body. While this interpretation appeals to intuition and is useful in many situations, it is not especially helpful when comparing different materials or materials near boiling points or other phase transitions. A better model is required. A thermodynamic analysis of Carnot engines (a particular form of ideal heat engine) shows that the efficiency of

reversible heat engines depends only on temperature. In particular the ratio of heat Q_1 taken in at a high temperature θ_1 to the heat Q_2 given out at a lower temperature θ_2 depends purely on the ratio of a function of the temperatures:

$$\frac{Q_1}{Q_2} = \frac{f(\theta_1)}{f(\theta_2)}, \quad (1.1)$$

where θ is any empirical measure of temperature. Kelvin's breakthrough was to recognise that the relationship could be used to define the temperature T :

$$\frac{Q_1}{Q_2} = \frac{T_1}{T_2}. \quad (1.2)$$

Kelvin was also able to show that this definition leads to an equation for ideal gases of the form

$$PV = \text{constant} \times T, \quad (1.3)$$

so that Kelvin's definition of temperature is equivalent to the gas scale originally proposed by Amontons, and implemented by Chappuis in 1889. While Kelvin's approach provides a definition of temperature, it does not provide much insight into the nature of the temperature. A diagram of a gas thermometer designed by Kelvin is shown in Figure 1.5.

It took the combined work of Maxwell, Boltzmann and Gibbs, using what is now known as statistical mechanics, to solve the problem. By considering the movement

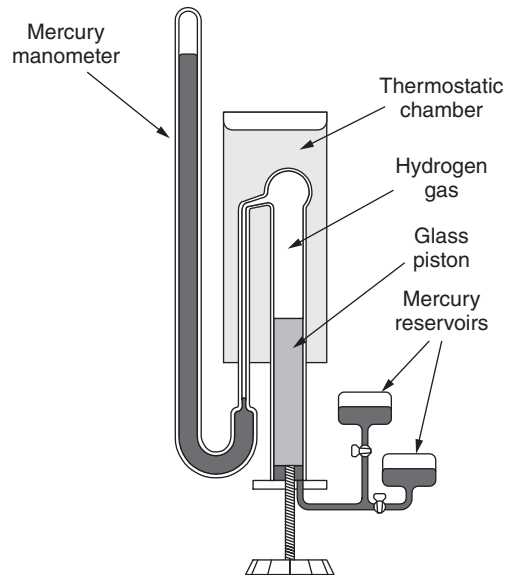


Figure 1.5 A constant-pressure hydrogen-gas thermometer designed by Kelvin. The glass piston is adjusted so that the pressure from the mercury manometer is constant. Under these conditions the volume of gas in the thermostatic chamber is proportional to temperature. The volume of the chamber is indicated by the displacement of the piston. The other mercury containers provide seals for the piston

Table 1.5 Some of the equations of state that have been used as the basis for thermometers to measure the thermodynamic temperature

Thermometer	Equation of state
Gas thermometer: pressure, P , and volume, V , of gas versus number of molecules, n , and temperature	$PV = nkT = NRT$
Total radiation thermometer: total radiance, L , versus temperature	$L = \frac{2\pi^5 k^4}{15c^2 h^3} T^4$
Spectral band radiation thermometer: spectral radiance, L_λ , versus wavelength, λ , and temperature	$L_\lambda = \frac{2hc^2}{\lambda^5} \left[\exp\left(\frac{hc}{\lambda kT}\right) - 1 \right]^{-1}$
Acoustic thermometer: speed of sound, c_s , versus specific heat ratio, γ , molecular mass, m , and temperature	$c_s^2 = \frac{\gamma kT}{m} = \frac{\gamma RT}{M}$
Noise thermometer: mean square noise voltage V_T^2 versus real part of impedance, Z , bandwidth, Δf , and temperature	$V_T^2 = 4kT \operatorname{Re}(Z) \Delta f$
molar gas constant, $R = 8.314\,447\,2 \text{ J mol}^{-1} \text{ K}^{-1}$	speed of light, $c = 299\,792\,458 \text{ m s}^{-1}$
Planck's constant, $h = 6.626\,068\,76 \times 10^{-34} \text{ J s}$	Boltzmann's constant, $k = 1.380\,650\,3 \times 10^{-23} \text{ J K}^{-1}$

and collisions of individual atoms in a closed box they were able to show that thermal equilibrium requires the mean kinetic energy of all the atoms to be the same. When applied to an ideal gas they obtained the result

$$PV = \text{constant} \times \langle mv^2/2 \rangle, \quad (1.4)$$

where $\langle mv^2/2 \rangle$ is the average kinetic energy of each atom in the gas. Comparison of this equation with Equation (1.3) shows that the *temperature* is proportional to the average kinetic energy of each atom, which finally gives a meaning to temperature that we can comprehend relatively easily. Note that the total kinetic energy of molecular gases is higher than that for monatomic gases because they can rotate and vibrate; in that case the temperature is proportional to the mean translational kinetic energy.

When the principles of thermodynamics and statistical mechanics are applied to other idealised systems we obtain equations of state that may be exploited in thermodynamic thermometers. A few of the equations and the corresponding thermometers are described in Table 1.5. All of these thermometers have been, and continue to be, used to measure temperature on the thermodynamic scale. Unfortunately the accuracy of thermodynamic thermometers falls well short of the repeatability and ease of use required for both research and commerce. Figure 1.6 summarises the latest measurements of the differences between the thermodynamic scale and ITS-90. The distributions of results presented are indicative of the accuracy of the thermodynamic thermometers. By comparison the ITS-90 scale achieves a repeatability of about 1 mK over the same temperature range.

1.3.3 Meteorological temperatures

In contrast with thermodynamic measurements, which represent the best efforts to measure temperature in accordance with the theoretical concepts, meteorological

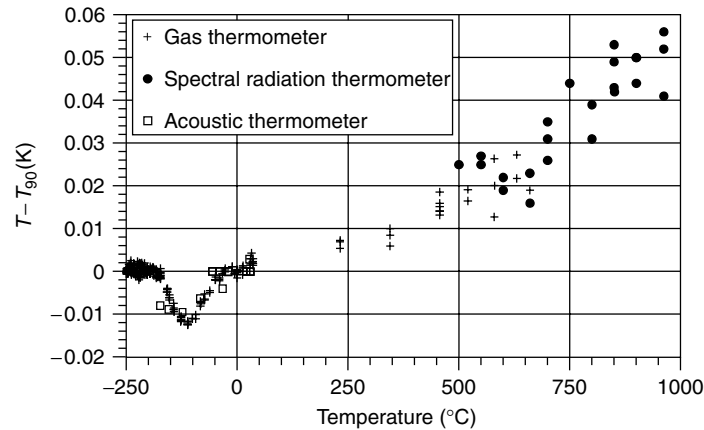


Figure 1.6 The latest measurements of the differences between ITS-90 and the thermodynamic temperature scale

temperature measurements are amongst the most empirical. In this section we discuss meteorological temperature in more detail to highlight some of the difficulties of giving meaning to temperature measurements when the system of interest is not in thermal equilibrium.

Some of the earliest thermometry observations recorded the change in temperature during the day. It was hoped that the measurements would correspond to how hot or cold a person felt, but this was not always the case. Lakes felt warm in winter and cool in summer, but the thermometers indicated otherwise. Problems also arose when people compared observations. An observer at one site, where the thermometer was kept in a living room heated by a fire, would find that the temperature variation was less than that of an observer who kept the thermometer in a spare room. Another observer, who thought that thermometers should be located outside the window on the sunny side of the house, found an even wider temperature variation.

It took some time before everybody was convinced that meteorological readings should be taken outdoors, even though temperature variations could be greater than for measurements made indoors. Debate continued as measurement techniques became more refined: should the thermometer be near a building, shaded from the sun, protected from the wind, and how large should the thermometer be?

Finally, after about 150 years, meteorologists settled on the Stevenson screen, as shown in Figure 1.7, to protect and mount the thermometer. The screen has double-louvred walls with the louvres sloping in opposite directions to allow the air to circulate yet block all radiation from direct access to the chamber. The screen must be mounted at a fixed height above the ground, a minimum distance from buildings and trees, and the paint used on the screens is specified. The thermometer is also specified and must be calibrated to 0.1°C . Despite the tight specification based on 250 years of development the screens are only expected to be reproducible to about 1.5°C !

So what then is the correct meteorological temperature? Ideally, we want to measure the air temperature, but this is not possible with a contact or immersion-type thermometer. In an ideal situation, such as a stirred liquid, the thermometer is in very good thermal contact with the liquid and very poor thermal contact with everything

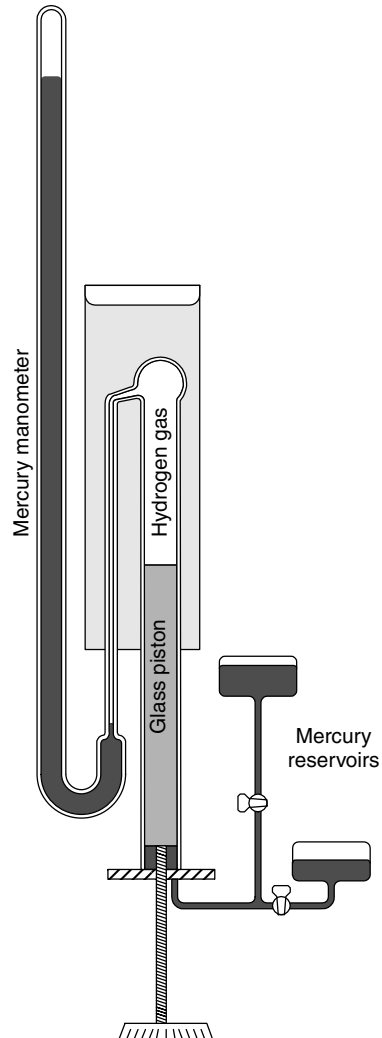


Figure 1.7 An early constant-pressure hydrogen gas thermometer designed by Kelvin. The volume of gas in the thermostatted chamber can be altered by adjusting the glass piston so that the pressure from the mercury manometer is constant. The other mercury containers provide sealing for the piston

else around it. Therefore, it reads a temperature very close to that of the liquid. Air on the other hand has a very poor thermal conductivity, a very low thermal mass, is transparent to most infrared radiation and is extremely viscous (its kinematic viscosity is similar to that of treacle!). In air, a thermometer is in poor contact with everything. The still air that could be said to be in immediate contact with the thermometer may have a mass much less than a gram, far less than the mass of the thermometer. If the conditions become windy, effectively increasing the mass of air in contact with the thermometer, then thermal contact with the air can improve by more than 100 times. At room temperature, everything radiates infrared radiation totalling about 500 W m^{-2} ,

so the thermometer is in radiative contact with literally everything around it. This radiative contact is not trivial and is very difficult to control. The relative effect of conduction and radiation also depends on the size of the thermometer.

In the final analysis the thermometer measures a very complex weighted average of the temperature of the air and all of the other objects around it, and it proves to be practically impossible to eliminate the effects of the other objects. However, with the use of screens and the other constraints it is practical to establish a measurement protocol that controls the most significant influence effects and delivers a result close to the physical definition. This ensures that results in different parts of the world are comparable. Thus with meteorological measurements the temperature has an uncertain physical meaning but does provide an internationally accepted basis for comparing meteorological conditions related to human comfort.

The use of the Stevenson screen highlights a traceability issue common to all measurements, which is essentially a problem of giving meaning to the measured quantity. Whereas a physicist (or any other theoretical scientist) may have a concise conceptual definition of the quantity measured, in almost every practical measurement there are problems leading to ambiguities or uncertainty. Therefore, if the measurement is to have meaning or be comparable with similar measurements made elsewhere we must follow accepted measurement protocols. In this example, the ‘meteorological temperature’ is defined by international standards for the screens and thermometers, and anyone who wants to claim to have measured the meteorological temperature must conform to those standards.

1.4 Traceability

1.4.1 Defining traceability

The ISO definition of traceability is:

Traceability:

The property of a result of a measurement whereby it can be related to appropriate standards, generally national or international standards, through an unbroken chain of comparisons each with a stated uncertainty.

There is considerable scope for variation in the interpretation of this definition. In order to clarify the meaning let us consider a few of the possible interpretations and investigate their limitations. While we discuss these options, we should keep in mind that the purpose of traceability is to enable different users, potentially on opposite sides of the world, to compare measurement results meaningfully.

Case 1: The dictionary interpretation

The word traceable has a wide range of colloquial meanings, the most appropriate being ‘able to be followed to the source’. This adds nothing to the ISO definition, but does highlight two important points. The ISO definition tells us where the chain begins and ends. Specifically, it begins with the measurement result, not with the instrument. Secondly the uncertainty provides a measure of the proximity to the source of traceability.

Case 2: The measurement scale interpretation

Following our exposition in Section 1.2.2 on measurement scales we could argue that traceability simply requires all measuring instruments to use the same measurement scale. However, this is not a very discerning interpretation. If we look around us we find that almost all of our measuring instruments indicate results in terms of the SI scales, so we must conclude that almost every measurement is already traceable. It has also happened on occasions that very good thermometers, such as standard platinum resistance thermometers, have been ‘calibrated’ using very poor thermometers such as thermocouples. In such cases the scale on the calibrated thermometer is not as good as it could or should be. Thus it is useful to have a measure of the quality of the scale and, as we noted in Section 1.2.4, it is necessary to know the uncertainty to make sensible decisions. Both arguments provide the rationale for the requirement in the ISO definition for stating the uncertainties.

Case 3: The filing cabinet interpretation

Let us suppose that for each measuring instrument used to make traceable measurements there is a calibration certificate kept in a filing cabinet, which allows the location of another filing cabinet to be traced, and so on, until a filing cabinet containing a certificate for the primary standard is found in a national measurement institute. Let us further suppose that each of these certificates appears to be a good and useful certificate; that is, each reports the uncertainties in measurements obtained with the instrument, under the appropriate specified conditions. The problem here is that possession of a certificate is no assurance that the final measurement, or indeed any of the measurements, have been carried out competently.

Case 4: The laboratory accreditation interpretation

In this case each of the laboratories involved in the calibration of the instruments and the final measurements is required to have the filing cabinets and certificates, but is additionally subject to an independent and expert audit of the entire measurement process. In this scenario, there is no real opportunity for measurement results to become corrupted in any sense (unless we question the competence of the accrediting body, and there are accreditation processes for them too). This requirement for an assessment of technical competence is not stated in the ISO definition of traceability, but accreditation to ISO 17025 *General requirements for the competence of testing and calibration laboratories* has become the practical working definition of traceability.

To summarise, we can interpret the ISO definition to mean:

Traceability (alternative definition):

The ability to demonstrate the accuracy of a measurement result in terms of appropriate national or international standards.

For thermometry, the appropriate standard is the SI kelvin.

1.4.2 Achieving traceability

So how is traceability achieved in practice? Clearly a substantial community effort is required. In general there are three main requirements for an international measurement system, as follows.

A source of primary physical standards

Primary physical standards are required to provide a unique definition of the measurement scales. Easily the most important source is the SI, *Système International d'unités*, which defines a system of seven base measurement scales, and primary physical standards for the unit for each of those scales. The scales for a large number of other metric quantities are also derived from the seven base scales. The SI is managed and maintained under a diplomatic treaty originally signed in 1875.

A large number of measurements are also made on non-SI scales, especially nominal, ordinal and interval scales, and these scales also require standards. These include for example: proprietary colour scales owned by the manufacturers of printing inks; scales based on tightly specified testing machines such as for Rockwell hardness, engine octane rating for fuel, and viscosity; and chemical scales for complex mixtures that are based on certified reference materials.

A source of documentary standards

With difficult measurements, as we found with meteorological temperatures, it is sometimes necessary to agree on a measurement protocol in order to be able to make comparable measurements. This is true also for measurements of viscosity, thermal resistance and electrical resistance; indeed almost all temperature-related product tests. Test methods for these quantities are usually standardised and published as documentary standards.

Documentary standards are also used to define other protocols, some of which have a direct effect on measurements. These include standardised responses for platinum resistance thermometers and thermocouples, mechanical specifications for electrical instruments and parts, software specifications and interfaces, and quality assurance systems. While such standards may not have a direct impact on traceability they do benefit the measurement community by ensuring that sensors, instrumentation and systems made by different manufacturers are equivalent or compatible.

The documentary standards system is a rather mixed group of national and international organisations, and some centred on professional societies or particular industries.

A source of independent third-party assessors

Measurements are quite unlike many of the products that we buy. When we buy a television set or a banana we can see something of the quality of the product that we are buying. With a measurement there is generally no way to tell from the result we receive whether the measurement is of good or bad quality. Measurements are also different in another way. Whereas a television manufacturer might like to change the

model each year or provide additional marketing gimmicks, a measurement sold on the basis of being made in terms of the ITS-90 temperature scale should mean no more and no less. In a sense the terms ‘kelvin’, the other SI units and other primary standards have accepted meanings, and the measurement community cannot afford to allow those meanings to become corrupted. If changes in meaning were allowed a supplier of measurements would be able to tender on the basis of the largest kelvin or the shortest metre. Chaos would soon result.

Given that a calibration or testing laboratory has followed documentary standards and calibrated its equipment, it must demonstrate that it has conformed to the community expectation in respect of measurement standards and technical procedures. The most important accreditation bodies are those accrediting to ISO 17025 *General requirements for the competence of testing and calibration laboratories*. Most countries now have organisations that offer accreditation and are recognised through mutual recognition agreements and the International Laboratory Accreditation Co-operation (ILAC). This ensures that the results from laboratories endorsed by these organisations are accepted in other countries. Because of the importance of measurement in the development and utilisation of technology, and the technological standing of countries as perceived by their trading partners, accrediting organisations are usually established under government regulations.

There are also other third-party assessment schemes. These are often associated with specific industries such as the military or aviation industry. Increasingly, however, these schemes are evolving towards and merging with the ISO 17025 system.

In the following sections we investigate these three components in more detail.

1.5 The SI

1.5.1 The metre convention

In a museum in Utrecht lies a thermometer with 18 different scales marked on a wide board behind the thermometer tube. The early 1700s were a time when each thermometer manufacturer had its own proprietary scale and natural philosophers were only beginning to appreciate the value of meteorological observations that could be compared with those made at other places. This was a lesson that traders have known since at least the time of the pharaohs. Nowadays temperature is one of the most measured quantities and is associated with commerce and technology reaching across every continent. Were we still to have 18 scales, disasters like the recent NASA Mars Climate Orbiter fiasco, which was caused by the confusion of SI and Imperial units for the small translational forces imparted by the jets that rotated the satellite, would be commonplace. The only way to be sure of the clear communication of measurement results is for all users of the measurements to share a common system of measurement.

The seeds of the SI system were planted by King Louis XVI at the time of the French Revolution; he wanted a decimal system of measurement. The metre was defined to be 1/10 000 000 of the distance between the Equator and the North Pole as measured along the quadrant that passes through Paris. (That the earth has a circumference that is almost exactly 40 000 km is no coincidence.) After the definition of the metre, the kilogram was defined to be the weight of 1 cubic decimetre of water. Platinum

artefact standards representing these two quantities were deposited in the Archives de la République in Paris. All measures of length and weight were derived from these standards, and hence these scales were amongst the first to be implemented as truly metric scales.

By the middle of the nineteenth century international trade was on the increase and a number of eminent scientists and industrialists of the time foresaw the value of a universal system of measurement and lobbied their governments intensively. The metric system was the obvious choice since it was well established in several European countries. Seventy-five years after King Louis XVI founded the metric system, on 20 May 1875, 17 nations signed a diplomatic treaty, the *Convention du Mètre*. This established an international organisation, and a laboratory, the Bureau International des Poids et Mesures (BIPM), on land set aside by the French government, with the responsibility of maintaining standards of measurement for the metre, the kilogram and the second. Over the years the responsibility has been extended to cover the ampere, the kelvin, the candela and the mole. Approximately 50 countries are now signatories to the Metre Convention and almost all of the world's 190+ independent states use the SI. However, the importance of the treaty is not so much that it establishes a metric system of units but rather that nations agree on the meaning of the units.

Figure 1.8 shows the various organs of the *Convention du Mètre*. Delegates from member countries meet at the Conférence Générale des Poids et Mesures (CGPM) at four-yearly intervals to approve the latest recommendations on improvements to the SI and the operation of the organisation. Nearly all of the recommendations come from the other parts of the organisational structure to which member nations have the opportunity to contribute. The consultative committees in particular are made up from experts in each measurement discipline, usually from the world's national measurement standards laboratories, and have the responsibility for overseeing and encouraging appropriate research into the various units. This ensures that the units have sufficient accuracy to meet the ever-increasing needs of commerce and technology, and can be made readily available to those who need them.

When the Metre Convention was signed it was envisaged that the BIPM would maintain the primary standards for all of the various scales. However, scales based on a single artefact are troublesome. As the number of measurements traceable to that artefact increases so too does the value of the artefact. The need to protect it from damage becomes paramount and it becomes very difficult to disseminate many measurements at the highest accuracy. The alternative, which has been pursued vigorously, is to base the standards on fundamental physical constants, in the same way that temperature is defined in terms of the triple point of water. This ensures that the standards can be rebuilt easily if damaged and in principle copied by every national standards laboratory. Nowadays only one artefact standard is maintained at the BIPM, namely the kilogram. It is hoped that in the near future, the kilogram will join the other units and be defined in terms of fundamental physical constants.

1.5.2 The SI units and conventions

The SI units are divided into two classes: base units and derived units. In principle, some of the base units are unnecessary since they can be related to each other through

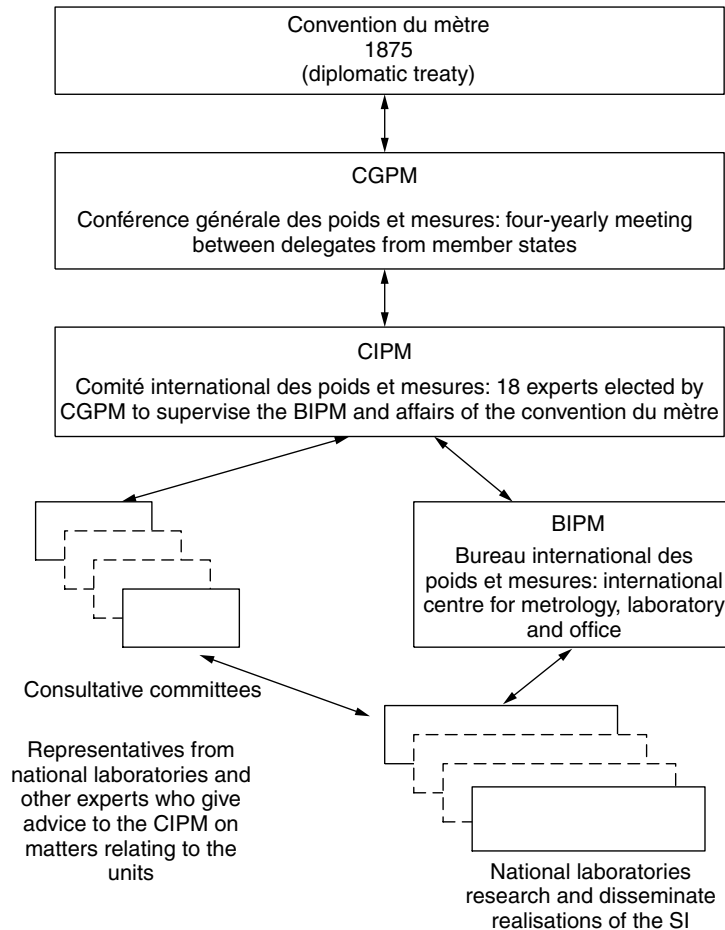


Figure 1.8 The structure and responsibilities of the various organs of the Metre Convention

measurements of the fundamental physical constants; however, they are necessary for the most accurate measurements and are regarded as dimensionally independent. The current SI definitions for the base units are as follows:

The second (s), the unit of time interval: The second is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the caesium atom.

The metre (m), the unit of length: The metre is the length of the path travelled by light in a vacuum during a time interval of $1/299\,792\,458$ of a second.

The kilogram (kg), the unit of mass: The kilogram is the unit of mass, equal to the mass of the international prototype kilogram.

The ampere (A), the unit of electric current: The ampere is that constant current which, if maintained in two straight parallel conductors of infinite length, of circular

cross-section and placed 1 metre apart in a vacuum, would produce between these conductors a force equal to $2\pi \times 10^{-7}$ newton, per metre of length.

The kelvin (K), the unit of thermodynamic temperature: The kelvin is the fraction $1/273.16$ of the thermodynamic temperature of the triple point of water.

The candela (cd), the unit of luminous intensity: The luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency 540×10^{12} hertz and has a radiant intensity in that direction of $1/683$ of a watt per steradian.

The mole (mol), the unit of amount of substance: The amount of substance of a system that contains as many elementary entities as there are atoms in 0.012 kilograms of carbon-12. When the mole is used, the elementary entities must be specified and may be atoms, molecules, ions, electrons, other particles, or specified groups of such particles.

The derived units are formed from products and powers of the base units according to the algebraic formulae linking the quantities involved. Thus, for example, velocity is measured as metres per second, the ratio of two of the base units. A few of the derived units prove to be so useful that they are given special names and symbols; these are given in Table 1.6.

In order to avoid confusion in the presentation of results the SI conventions should be adhered to.

Table 1.6 The SI derived units with special names and symbols

Derived quantity	SI derived unit	
	Special name	Symbol
Plane angle	radian	$\text{rad} = \text{m} \cdot \text{m}^{-1}$
Solid angle	steradian	$\text{sr} = \text{m}^2 \cdot \text{m}^{-2}$
Frequency	hertz	$\text{Hz} = \text{s}^{-1}$
Force	newton	$\text{N} = \text{kg} \cdot \text{m} \cdot \text{s}^{-2}$
Pressure	pascal	$\text{Pa} = \text{kg} \cdot \text{m}^{-1} \cdot \text{s}^{-2}$
Energy	joule	$\text{J} = \text{kg} \cdot \text{m}^2 \cdot \text{s}^{-2}$
Power, radiant flux	watt	$\text{W} = \text{kg} \cdot \text{m}^2 \cdot \text{s}^{-3}$
Electric charge	coulomb	$\text{C} = \text{A} \cdot \text{s}$
Electric potential difference	volt	$\text{V} = \text{kg} \cdot \text{m}^2 \cdot \text{s}^{-3} \cdot \text{A}^{-1}$
Capacitance	farad	$\text{F} = \text{kg}^{-1} \cdot \text{m}^{-2} \cdot \text{s}^4 \cdot \text{A}^2$
Electric resistance	ohm	$\Omega = \text{kg} \cdot \text{m}^2 \cdot \text{s}^{-3} \cdot \text{A}^{-2}$
Electric conductance	siemens	$\text{S} = \Omega^{-1}$
Magnetic flux	weber	$\text{Wb} = \text{kg} \cdot \text{m}^2 \cdot \text{s}^{-2} \cdot \text{A}^{-1}$
Magnetic flux density	tesla	$\text{T} = \text{kg} \cdot \text{s}^{-2} \cdot \text{A}^{-1}$
Inductance	henry	$\text{H} = \text{kg} \cdot \text{m}^2 \cdot \text{s}^{-2} \cdot \text{A}^{-2}$
Celsius temperature	degree Celsius	$^{\circ}\text{C} = \text{K}$
Luminous flux	lumen	$\text{lm} = \text{cd} \cdot \text{sr}$
Illuminance	lux	$\text{lx} = \text{cd} \cdot \text{m}^{-2}$
Activity	becquerel	$\text{Bq} = \text{s}^{-1}$
Absorbed dose	gray	$\text{Gy} = \text{m}^2 \cdot \text{s}^{-2}$
Dose equivalent	sievert	$\text{SV} = \text{m}^2 \cdot \text{s}^{-2}$

Use of names for units

- When written in full, the names of all SI units start with a lower-case letter, except at the beginning of a sentence; for example, kelvin not Kelvin or degrees kelvin, degrees Celsius not Degrees Celsius.
- The symbols are lower case except when named after a person. Hence K is the symbol for kelvin. When written in full, the names of the units may be made plural according to the rules of English grammar; for example, ‘temperature difference in kelvins’.

Use of the symbols for units

- Symbols should be used to denote the units when reporting numerical results, and the full name when referring to units in written text. The symbol should be separated from the last digit by a single space, e.g. 273.15 K not 273.15 K.
- When reporting quantities with compound units formed by the product of two or more units, the unit symbols should be separated by a half-high dot, dot or a space; for example, for metre-kelvin: m.K or m.K or m K, but not mK, which implies millikelvin.
- When reporting quantities with compound units formed by ratios of two or more units, exponentiation or a single solidus may be used. Parentheses should be used to prevent ambiguities: for example, W/m² or W.m⁻²; J/(kg.°C) or J.kg⁻¹°C⁻¹ not J/kg/°C.

Decimal points and commas

- Numbers less than one should have a single zero before the decimal point. A comma should be used as the decimal point. In English-speaking countries a dot on the line is more commonly used, e.g. 0.1 °C or 0,1 °C, but not .1 °C.

Table 1.7 The most commonly used SI prefixes

Factor	Prefix	Symbol
10 ¹²	tera	T
10 ⁹	giga	G
10 ⁶	mega	M
10 ³	kilo	k
10 ²	hecto	h
10	deca	da
10 ⁻¹	deci	d
10 ⁻²	centi	c
10 ⁻³	milli	m
10 ⁻⁶	micro	μ
10 ⁻⁹	nano	n
10 ⁻¹²	pico	p
10 ⁻¹⁵	femto	f

- To facilitate the reading of numbers with many digits, the digits may be separated into groups of three counting from the decimal point. The groups should be separated by a space, never a comma, which may be confused for a decimal point.

Use of prefixes for symbols

- The most commonly used prefixes are given in Table 1.7.
- When joining a prefix and SI unit symbol, there is no space between the prefix symbol and the unit symbol, e.g. 10 mK or 10 m°C, not 10 m K.

1.6 Documentary Standards

From a traceability perspective the most important contribution of documentary standards to thermometry is in the area of test methods. There are very few properties of materials that do not change with temperature, and consequently a very high percentage of test methods involve the measurement of temperature.

Documentary standards also cover a number of aspects of interest to thermometrists, including:

- Specifications for the response of platinum resistance thermometers and thermocouples, and the dimensions and scales for liquid-in-glass thermometers.
- Colour codes for thermocouple lead wires, and colours and dimensions of thermocouple plugs and sockets.
- The materials and dimensions for sheathing materials and thermowells (the protective pockets used for mounting thermocouples in industrial plants).
- Dimensional and electrical specifications for industrial instrumentation such as temperature controllers.
- Specifications for furnaces and ovens, especially those used for heat treatment and sterilisation.
- Electrical and communication standards for instrument interfaces.
- Quality assurance and laboratory accreditation systems.

A short summary of major standards organisations relevant to thermometry is given in Table 1.8.

It is notable that only a few of the standards organisations are truly international. Consequently, there may be differences between standards from different organisations, apparently for the same device or protocol. Particular examples include the standards for platinum resistance thermometers and thermocouples, where there are small differences. As the standards have been revised following the change to the ITS-90 temperature scale, many of the standards have become harmonised.

It is beyond the scope of this book to catalogue all of the temperature-related standards available from these organisations. Your local standards organisations will have catalogues available and may be able to advise which standards are relevant. In addition, most of the organisations now have Internet sites, with good search engines and on-line shops. Many of the organisations also have application guides and manuals

Table 1.8 Some of the larger standards organisations that produce thermometry-related documentary standards

Acronym	Title	Internet address
ANSI	American National Standards Institute	www.ansi.org
ASTM	American Society for Testing and Materials	www.astm.org
BSI	British Standards Institution	www.bsi-global.com
DIN	Deutsches Institut für Normung	www.din.de
IEC	International Electrotechnical Commission	www.iec.ch
IP	Institute of Petroleum	www.petroleum.co.uk
ISO	International Organisation for Standardisation	www.iso.ch
JIS	Japanese Industrial Standards	www.tokyo.jsa.or.jp
OIML	International Organisation for Legal Metrology	www.oiml.org

for particular measurement disciplines, and booklets giving detailed information on the SI.

Exercise 1.5

Spend an hour or so visiting some of the Internet sites given in Table 1.8 and search for standards and documents relating to temperature. If you have responsibilities for product testing you should try searching on a few relevant keywords.

1.7 Laboratory Accreditation to ISO/IEC 17025

The standard, ISO/IEC 17025:1999 *General requirements for the competence of testing and calibration laboratories*, has evolved from the ISO Guide 25 of the same name. The standard applies to all calibrations and testing laboratories whether using standard, non-standard or laboratory-developed methods. The standard has two groups of requirements: managerial and technical.

The managerial requirements are equivalent to those required under the ISO 9001 and ISO 9002 quality systems, and include the following:

- The laboratory's management must be committed to a quality scheme by ensuring that policies and objectives are communicated to, and understood and implemented by, all laboratory personnel.
- Independence and financial stability of the laboratory are desirable. Where the laboratory is part of a larger organisation it is particularly important for the laboratory to act independently. There should be no conflicts of interest.
- Quality systems must be properly documented. Procedures must be written to cover the responsibilities of the staff. Documentation control is needed to ensure that the staff use the latest procedures.
- New work must be reviewed to understand its requirements and determine whether the laboratory can carry it out.

- Records should be made and kept for all work. Regulatory or contractual requirements may determine the length of time to hold records.
- Complaints and corrective action procedures are essential, especially for the control of sub-standard testing work.
- Control over procurement of equipment, consumables and services requires procedures to see that they are of appropriate quality; for example, calibration certificates supplied with test equipment.

The technical requirements include the following:

- Staff should be properly qualified and a regular training programme should be in place.
- Accommodation and environment provisions must be adequate for staff, test equipment and test samples.
- There must be adequate management and control of test equipment. This is an essential feature for traceability, which translates into the whole life of a piece of test equipment being properly documented.
- Calibration of test equipment must be carried out in a timely and proficient manner.
- Test methods, whether in-house or standard methods, must be validated.
- If sampling is involved in any of the work then the laboratory should have a sampling plan and procedures.
- There must be procedures for the identification of items for test, and for safe handling and storage, to ensure the integrity of the item.
- Test reports and certificates must be well specified in terms of content and format.
- The laboratory should participate in proficiency testing programmes or employ other statistical techniques, where appropriate, to enhance the confidence in procedures.

Because calibration laboratories provide an important link in the traceability chain to the SI, and may affect many clients downstream from the laboratory, the accreditation process for calibration laboratories tends to be more stringent than that for testing laboratories, especially in respect of the care of instruments and in the assessment of uncertainty. This is recognised in the ISO 17025 standard. In some countries, separate accreditation bodies deal with testing and calibration.

If you or your laboratory is considering accreditation you should remember that the accreditation authority is not a customer but a supplier of a service. You are paying the authority to represent your clients' interests, and you should be treated as their client. It is not a regulatory authority, although it may request conformance to documentary standards if that is what your clients expect, and will judge whether you are conforming to any documentary standards that you nominate. Remember too that when it assesses your laboratory it acts on behalf of its other clients who expect it to help maintain the integrity of the various parts of the measurement system. The service it provides is the acceptance of your test and measurement results by a wide range of international customers.

Users of measurement results should be careful to distinguish between accreditation and certification. In the quality industry, accreditation is a jargon term applied strictly

to companies assessed for competence according to the ISO 17025 standard. When applied to organisations, certification simply means that the company has implemented an ISO 9000 management system, which has no explicit requirement for technical competence.

1.8 National Measurement System

Having knowledge of what is necessary for good measurements (i.e. to achieve traceability) is of no use unless the theory can be put into practice. Much of what is required is beyond the direct influence of any individual. Fortunately most governments have taken an interest in their country's ability to measure because it affects the country's wealth and welfare, and standing with trading partners. Planning of the national measurement system (NMS) by government also ensures that the necessary services are available to enable regulatory and contractual requirements to be met by industry. The total of the measurement services inside a country can be considered together as the NMS, a concept that has grown in importance over the last few decades.

Good measurement practice requires several services to be readily available:

- calibration of instruments;
- training of staff in measurement techniques;
- regulation of trade measurement;
- endorsement of results by accreditation;
- specifications and procedures for measurements;
- supply of measuring instruments; and
- repair and servicing of instruments.

In this section we consider the first three of these topics. Accreditation and documentary standards have been covered already in sufficient detail, and repair and supply of servicing is normally provided by non-government organisations.

Figure 1.9 outlines the formal components of a national measurement system that are external to its users, that is calibration, specification and accreditation. The figure gives an outline only; in a well-organised NMS there are many components, including second- and third-tier calibration laboratories and standards committees, and many more interactions.

For measurements to have legal standing countries must have laws providing for measurement units, and for the laws to be effective a national measurement institute (NMI) is required to hold primary measurement standards. Such laws provide a basis for national and international trade, consumer protection, and environmental and health management. Where the best accuracies are required the NMI will realise the physical definitions of the units in accordance with the SI definitions and BIPM guidelines, and maintain these as the primary standards. International obligations can also be met by using reference standards calibrated by the BIPM or other NMIs.

The infrastructure requirement of maintaining uniform measures is the most important of the NMI's responsibilities and is met most simply by providing calibration services. While this satisfies the prime function in support of legislation and trade,

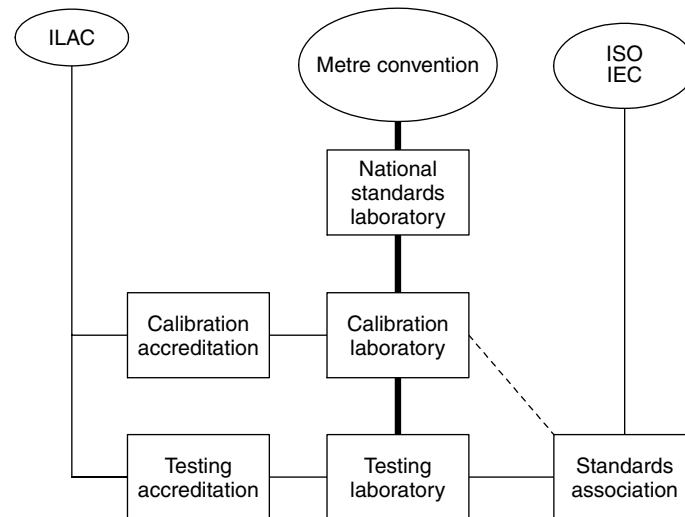


Figure 1.9 Traceability links in a national measurement system. The main traceability path is shown by the heavy line. Documentary input that affects the path is shown by the other lines. Accreditation bodies for testing and calibration may be separate from each other

the end-users of the NMI's 'product' are generally distributed wherever the country's products are exported. In order to satisfy all clients the NMI must not only ensure that the country's measures are uniform but also establish the credibility of its standards on the international stage. To this end the BIPM has extended its responsibilities to the maintenance of a mutual recognition arrangement (MRA) that enables one country to recognise another's measurement standards. To participate in the MRA the NMI must be an associate signatory to the Metre Convention, participate in international comparisons of measurement standards, have a quality system equivalent to ISO 17025, and publish relevant contributions to metrological science. The MRA, which laid the 'ground rules' for recognition, was signed in October 1999, and will gradually take effect over the next five years or so.

The NMIs also provide a pool of expertise that makes a valuable contribution to a country. The links with other NMIs and familiarity with new measurement techniques often mean that the NMIs provide a way of introducing new measurement technology to a country, and of keeping up with international trends. To this end most NMIs offer training courses and participate in accreditation assessments, as well as carrying out consultancy for specific clients.

Generally the NMI has no powers of enforcement, and experience shows that within a country some enforcement is required to protect consumers and the general public. To this end most countries have a legal metrology organisation. Its responsibility is to ensure that traders' weights and measures are correct and that the public is not being defrauded in any transactions. Again there is an international connection through the Organisation Internationale de Metrologie Légale (OIML). It provides experience and resources to aid countries in the establishment and enforcement of appropriate law.

Training is essential throughout the whole NMS to ensure that those involved are technically competent to make measurements. Unlike the other three components, there

is no recognised organisational structure to achieve this, and it is rare to find tertiary education institutes providing formal measurement training. Usually measurement expertise is absorbed by osmosis as a part of other technical training courses, or passed on as lore from other staff. Indeed much of the subject matter lacks a formal basis for an educational curriculum. In many countries, most of the government-operated bodies in the NMS offer short training courses in their respective disciplines.

Exercise 1.6

Draw a diagram similar to Figure 1.9 for your temperature measurements. Then note how the procedures differ for any other measurements you make. If possible obtain the names and addresses of the organisations involved in these procedures and the name of a contact person. Include any linkages to organisations outside your NMS.

Indicate in the diagram if you are subject to more than one accreditation body or standards association and show their international linkages. Scientific users can give the relevant scientific references instead of the organisations. Are your measurements traceable according to the definitions given in this chapter?

Further Reading

Measurement and measurement scales

- C W Churchman and P Ratooch (1959) *Measurement: Definitions and theories*, John Wiley, New York.
- B Ellis (1968) *Basic Concepts of Measurement*, Cambridge University Press, Cambridge.
- Handbook of Measurement Science*, John Wiley, Chichester.
- Vol 1 *Theoretical Fundamentals* Ed P H Sydenham (1982).
- Vol 2 *Practical Fundamentals* Ed P H Sydenham (1983).
- Vol 3 *Elements of Change*, Ed P H Sydenham and of R Thorn (1992).
- P H Sydenham, N H Hancock and R Thorn (1989) *Introduction to Measurement Science and Engineering*, John Wiley, Chichester.

Historical developments

- A Klein (1974) *The Science of Measurement: A Historical Survey*, Dover, New York.
- W E Knowles-Middleton (1966) *A History of the Thermometer and its Use in Meteorology*, Johns Hopkins University Press, Baltimore, MD.
- T J Quinn (1990) *Temperature*, 2nd Edition, Academic Press, London.

Thermodynamic measurements

- T J Quinn (1990) *Temperature*, 2nd Edition, Academic Press, London.
- R L Rusby, R P Hudson, M Durieux, K Grohmann, H J Jung, P P M Steur and J V Nicholas (1996) The status of thermodynamic thermometry, *Metrologia* **33**, 409–414.
- J F Schooley (1986) *Thermometry*, CRC Press, Boca Raton.

The SI

ISO 31-0:1992 *Quantities and Units –General Principles*, International Organisation for Standardisation, Geneva.

The International System of Units (1998) BIPM, Paris.

Quality assurance and accreditation

G W Roberts (1983) *Quality Assurance in Research and Development*, Marcel Dekker, New York.

General reading on temperature measurement

R P Benedict (1984) *Fundamentals of Temperature, Pressure, and Flow Measurements*, 3rd Edition, John Wiley, New York.

R E Bentley (1998) *Handbook of Temperature Measurement, Vols 1, 2 and 3*, Springer-Verlag, Singapore.

T W Kerlin and R L Shepard (1982) *Industrial Temperature Measurement*, Instrument Society of America, Research Triangle, NC.

T D McGee (1988) *Principles and Methods of Temperature Measurement*, John Wiley, New York.

L Michalski, K Eckersdorf and J McGhee (1991) *Temperature Measurement*, John Wiley, Chichester.