AN INTRODUCTION TO INSTRUMENTED INDENTATION TESTING

No. 92

Nigel Jennett
ABSTRACT
Instrumented Indentation is the future of hardness testing. It is able to measure both hardness and elastic modulus from the same indentation experiment. It is able to do anything conventional hardness testing can do but is also one of the only methods to be able to measure both elastic and plastic properties of small volumes of materials. For example, it can measure the elastic modulus of a sub micron thick coating with an uncertainty comparable to state-of-the-art uniaxial tensile testing of a steel bar. This guide aims to provide a rapid introduction to the technique and is structured to cater both for new users and those operating at higher levels of complexity and/or accuracy. Sections marked for “advanced users” contain the fine detail that may not be relevant for every user, being of most interest to those operating at high resolution and accuracy.
Acknowledgements

The Author would like to thank and acknowledge the many collaborators who have contributed to the development of the measurement good practice identified in this document. In particular the partners of the INDICOAT EC project contract No. SMT4-CT98-2249: Determination of Hardness and Modulus of Thin Films and Coatings by Nanoindentation (INDICOAT). The author also acknowledges the financial support of the National Measurement System Directorate of the UK Department of Trade and Industry.
1 Introduction ................................................................. 1
  1.1 Introduction .................................................................. 2

2 The basis of instrumented indentation .................................. 3
  2.1 Introducing the Force – Displacement Curve .................. 4
  2.2 Martens hardness scale, HM ........................................ 6
  2.3 Indentation hardness, $H_{IT}$ ........................................... 7
  2.4 Indentation modulus ..................................................... 7
  2.5 Measurement of Coating Properties .............................. 8

3 Standard testing procedure .............................................. 11
  3.1 ISO 14577 .................................................................... 12

4 Indenter geometry .......................................................... 13
  4.1 Choosing the right tool for the job ................................. 14
  4.2 Indenter geometry considerations for advanced users ...... 16
    4.2.1 “Self-similar” indenters .............................................. 16
    4.2.2 Vickers ................................................................. 16
    4.2.3 Berkovich indenter .................................................. 16
    4.2.4 Cube Corner .......................................................... 17
    4.2.5 Spherical indenters .................................................. 17
  4.3 Flowcharts .................................................................... 19

5 How to set-up a test ....................................................... 23
  5.1 Sample mounting ........................................................ 24
  5.2 Sample cleaning ........................................................ 25
  5.3 Cleaning the indenter .................................................... 25
  5.4 Choosing a test environment ......................................... 26
  5.5 Environmental temperature control (active & passive) ... 26
  5.6 Vibration isolation (active & passive) ............................ 28
  5.7 Typical Indentation cycle .............................................. 29

6 How to calculate results ................................................. 31
  6.1 Zero Point determination .............................................. 32
  6.2 Thermal drift correction .............................................. 33
  6.3 Frame compliance correction ....................................... 34
  6.4 Fitting to the force removal curve ................................. 34

7 Measurement artefacts to consider ................................... 37
  7.1 Discarding bad data .................................................... 39
    7.1.1 What to look for ..................................................... 39
    7.1.2 Criteria to discard ................................................ 39

8 Equipment calibrations .................................................. 41

9 Measurement uncertainty ............................................... 47

10 References ................................................................. 49
Introduction
1.1 Introduction

Hardness testing has been used for over a century as a quick and simple test for distinguishing materials and as a quality assurance test. Hardness has conventionally been defined as the resistance of a material to permanent deformation by another harder material with measurement being made after the test force has been removed, such that elastic deformation is ignored e.g. Tabor[1]. Instrumented indentation testing, however, provides the ability to measure the indenter penetration under the applied force throughout the testing cycle and is therefore able to measure both the plastic and elastic deformation of the material under test [2,3,4,5,6,7]. Conventional hardness testing uses optical microscopy to measure the lateral size of the residual indentation. Instrumented indentation is, however, able to infer the size of an indentation from the measured penetration depth [7,8,9,10]. In both cases, knowledge of the indenter geometry is required. Instrumented indentation renders a separate measurement of the indent size unnecessary and allows indentation testing at indent sizes much smaller than previously possible [5]. Instrumented (Nano)-indentation is one of the very few techniques that can measure both the elastic and plastic properties of very small volumes of materials (∼ 10⁻²¹ m³) and can measure coatings ∼100 nm thick with an uncertainty equal to or lower than that obtained by uniaxial tensile tests when testing macroscopic test-pieces [11]. Since the advent of its commercial availability in the 1980’s, instrumented indentation has produced a tremendous increase in capability allowing for the measurement of other properties such as modulus, creep and visco-elasticity as well as hardness [7,9,12,13,14]. When the ability of instrumented indentation to measure modulus is combined with acoustical measurements, the combined data provides non-destructive measurement of Poisson’s ratio and the thickness or density of coatings with low uncertainty [15,16].

It is an indicator of the continued rapid development of instrumentation in this field that typical indentation depths have followed a Moore’s Law like reduction. For example, in the 1990’s, typical instrumented indentation depths were of the order of 1 μm. Now indentations are regularly 100nm or less and depths of the order of 10 nm are not uncommon [17,18,19]. Interestingly, higher force instruments are also now available, capable of applying forces up to 1500 N at least. Indeed at the National Physical Laboratory, many of the UK Hardness scales are defined by an instrumented indenter that is programmed to simulate conventional hardness testing cycles.
The basis of instrumented indentation
2.1 Introducing the Force – Displacement Curve

The key feature of an instrumented indentation instrument is its ability to measure indenter penetration, $h$, under the applied force, $F$, throughout the testing cycle (i.e. as a function of time). As the force is increased, the instrument measures both the plastic and elastic deformation of the material under test. As the force is removed the indentation recovers elastically and the stiffness of the contact may be related to the elastic modulus of the material under test using knowledge of the area of contact.

The principle of a typical system is shown in Figure 1. The indenter is mounted on an indenter shaft, which is suspended by flexure elements and is pressed against the test sample by a force actuator (commonly either an electromagnetic, electrostatic or piezoelectric system). Most commonly, a parallel plate capacitor is used to measure the displacement of the shaft (and therefore the indenter) into the sample. By instrumenting the indentation displacement, indents no longer have to be inspected to determine their size and can therefore be much smaller.

Figure 1: Schematic of instrumented indentation test system

For larger indents (indentation depths greater than 6 µm) the indentation area of contact can be calculated from measurements of indentation depth assuming the indenter has the “ideal” or “perfect geometry” (the error in area is less than 1% for a Vickers or Berkovich indenter that complies with ISO 14577-2 [20]). For smaller indent depths, more detailed knowledge of the indenter shape as a function of depth is required and the area as a function of depth (the “area function”) is calibrated for each indenter. Currently the best method to use for this is an Atomic Force Microscope that has in-built metrology (independent linear displacement transducers that have a calibration traceable to National Standards) [21, 22], see Figure 2.

Figure 2: 3D Image and Indenter area function obtained by metrological Atomic Force Microscopy of a modified Berkovich indenter
In a typical quasi-static indentation cycle, force is progressively applied and the indenter is steadily pressed into the sample. As a result, the sample deforms elastically and plastically under the indenter until a maximum force is reached.

The force may be held for a while before being progressively removed, the sample relaxing elastically as the indenter is removed from the surface. The key feature is that the applied force and measured displacement are recorded either continuously or at frequent time points throughout this cycle. The immediate output from instrumented indentation is therefore the force-displacement curve, Figure 3. This in itself can be extremely instructive, for example, the ratio of the areas under the force increasing and decreasing curves is an immediate indicator of the balance between elastic and plastic deformation of the material under test.

![Figure 3: Typical force displacement curve showing measured (F, F_max, h, h_max, h_f) and derived parameters (S, h_c, h_c)](image)

Figure 4 shows the schematic representation of the indent at maximum applied force and after the force has been removed. The gradient of the force-displacement curve as the force is removed is a measurement of the contact compliance $C = 1/S$ from which the contact depth, $h_c$, can be calculated by taking into account the bowing of the surface in response to the exact shape of the indenter. This surface bowing is calculated using contact mechanics equations solutions provided by Sneddon [23,24]. In reality the actual contact depth is also a function of the test material response to indentation; 'sinking-in' and 'piling-up' of the material around the indenter affect the value, reducing and increasing the actual contact depth respectively. Pile up and sink in are a function of material processing history and, as in conventional hardness testing, are not included in the standard analysis.

![Figure 4: Schematic of indentation showing the displacements observed during an indentation experiment (after Oliver and Pharr [9]). The overlay shows the elastic response as a combination of the sample and the indenter considered as two springs in series. If the sample is coated, then the spring of the sample is itself a combination of the coating and substrate ‘springs’ in series.](image)
The measured values of contact compliance, maximum depth and maximum force may be used to calculate the contact depth, the indentation modulus $E_{IT}$, the indentation hardness $H_{IT}$ (equivalent to Meyer Hardness and similar to $HV$ for a Vickers indenter) and the Martens hardness $HM$ (previously known as Universal hardness, $HU$) of a material [20,25,26]. In addition, time-dependent material properties may be investigated by measuring the continued displacement under a constant applied force. Toughness properties may also be investigated e.g. by calculating the ratio of elastic to plastic work done during an indentation [27,28].

For convenience, ISO 14577 [20] defines three ranges of instrumented indentation: Macro, Micro and Nano. From a user’s perspective, the Micro range is divided into two, since the standard allows the assumption of ideal indenter geometry in the analysis of indentations greater than 6 µm deep. These ranges are shown in Table 1 and broadly reflect the three main types of instruments commercially available at the time the standard was written.

<table>
<thead>
<tr>
<th>SCALE</th>
<th>FORCE RANGE</th>
<th>DEPTH RANGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macro range</td>
<td>$2 \text{ N} &lt; F &lt; 30\text{ 000 N}$</td>
<td></td>
</tr>
<tr>
<td>Micro range (upper)</td>
<td>$F &lt; 2 \text{ N}$</td>
<td>$h &gt; 6 \text{ µm}$</td>
</tr>
<tr>
<td>Micro range (lower)</td>
<td>$F &lt; 2 \text{ N}$</td>
<td>$0.2 \text{ µm} &lt; h &lt; 6 \text{ µm}$</td>
</tr>
<tr>
<td>Nano</td>
<td></td>
<td>$h \leq 0.2 \text{ µm}$</td>
</tr>
</tbody>
</table>

Table 1: Ranges of instrumented indentation

The standard allows a wide range of materials properties to be calculated from instrumented indentation data, many defined for a wide range of indenter geometries: cube corner, Knoop, ball indenters (including "hardmetal" ball indenters with diameters ranging from 0.5 mm to 10 mm), and diamond sphero-conical indenters with a range of flank angles, e.g. 30º, 45º, and 60º, and radius of curvature varying from 0.500 mm $> R > 0.050 \text{ mm}$. However, Martens Hardness, $HM$, is defined only for Vickers and Berkovich indenters.

### 2.2 Martens hardness scale, $HM$

The Martens hardness value, $HM$, is calculated by dividing the test force $F$ by the surface area of the indenter penetrating beyond the original surface of the test piece $A_s(h)$.

\[
\text{a) Vickers indenter: } \quad HM = \frac{F}{A_s(h)} = \frac{F}{26.43 \cdot h^2} \\
\text{b) Berkovich indenter: } \quad HM = \frac{F}{A_s(h)} = \frac{F}{26.44 \cdot h^2}
\]

The benefit of this method is that a Martens hardness value is easy to obtain for the upper Micro and Macro ranges, as it only requires the measurement of maximum force and maximum displacement and an indenter geometry that conforms to ISO 14577-2. However, this ease of use reduces at lower ranges because the calculation of $HM$ for low Micro and Nano ranges is complicated by the requirement to use the actual surface area of the indenter at each penetration depth. This indenter surface area function is not as easy to obtain as a measurement of projected (Cross-sectional) area. An important consideration for the user is that, for any range, the $HM$ value is a combination of elastic and plastic indentation responses and so has no direct correlation either to a physical property such as elastic modulus, or to traditional hardness scales.
2.3 Indentation hardness, $H_{IT}$

The indentation hardness $H_{IT}$ is closely related to Meyer Hardness in conventional testing and is the average indentation pressure required to cause permanent deformation in a material. Thus $H_{IT}$ is calculated as the ratio of the maximum applied test force, $F$, divided by the projected area of the indenter in contact with the test piece at maximum applied force:

$$H_{IT} = \frac{F}{A(h_c)}$$

The projected contact area $A(h_c)$ is calculated from knowledge of the geometry of the indenter and the measured stiffness of the contact. In very plastic materials, the area of contact under moderate forces is very similar to that of the remaining indent after the test force has been fully removed. For these materials, the indentation hardness is closely proportional to the Vickers hardness value, the difference being that $HV$ divides force by surface area of contact and instrumented indentation divides force by projected area. Both hardness measurements assume a known geometry for the indenter: the Vickers scale uses an optical measure of the projected contact area and assumes a perfectly pyramidal indenter geometry to calculate the surface area of contact; instrumented indentation uses the measured displacement data and a measured shape of the indenter, which makes allowance for tip rounding and other common deviations, to calculate the projected contact area. In practice, tip rounding means that the two scales diverge as the indentation contact becomes more elastic until, at indentation pressures below the yield point, the indentation becomes fully elastic, $HV$ becomes infinite and the calculated $H_{IT}$ value is no longer a measure of plasticity, although it remains a measurement of the average indentation pressure.

2.4 Indentation modulus

The indentation modulus $E_{IT}$ is calculated from the slope of the unloading curve through the formula:

$$E_{IT} = (1 - \nu_{IT}^2) \left\{ \frac{2}{\sqrt{\pi}} \frac{\sqrt{A(h_c)}}{S} - \frac{(1 - \nu_{indenter}^2)}{E_{indenter}} \right\}^{-1}$$

where:
- $\nu_{IT}$ = the Poisson's ratio of the test piece
- $\nu_{indenter}$ = the Poisson's ratio of the indenter
- $S$ = the slope of the tangent of the force/indentation curve during the unloading cycle (Figure 1)
- $h_c$ = the contact depth value, which is dependent on the shape of the indenter

The full procedure for this is described in ISO 14577 part 1 [20]. For homogeneous and isotropic materials, where the Poisson’s ratio is known, $E_{IT}$ approaches the Young's modulus of the material. For an anisotropic material, the value is a 3D average of the crystallographic moduli (e.g. Voigt-Reuss-Hill [29,30,31]).
2.5 Measurement of Coating Properties

When measuring the hardness and modulus of coatings by instrumented indentation testing, the fact that each indentation is a composite response of both the coating and the substrate must be taken into account [32,33,34]. For modulus measurement there is, in principle, no ‘safe depth’ at which a coating-only value may be obtained, since the coating and substrate are effectively responding as two springs in series. However, procedures are now developed [35,36,37,38] that enable the ‘coating-only’ properties to be extracted effectively and reliably from a composite response and these procedures are contained in a fourth part for ISO141577 that has recently been published as a draft international standard. These procedures include a normalisation procedure that allows measurements from different thickness coatings made by different indenter geometries to be directly compared and plotted on the same graph. Indenter geometry differences are handled by using the real indenter area functions to calculate a/tc, a dimensionless relative indentation size, where a is the effective indentation radius estimated as \((\frac{Ac}{\pi})^{\frac{1}{2}}\) and tc is the nominal coating thickness.

![Figure 5: Variation of E\text{IT}\* with a/tc for gold coating on nickel (three different indenters), where a is the radius of contact estimated as (Ac/\pi)^{\frac{1}{2}} and tc is the nominal coating thickness.](image)

\[ E = E_{IT}^* (1 - v^2) \]

Figure 5 shows results for the plane strain indentation modulus of gold on nickel for a number of different experiments. It can be seen that, as the ratio of indentation radius to coating thickness is reduced, the composite modulus response reduces. The intercept value of modulus represents the plane strain modulus for the coating. Figure 6 shows a similar plot of hardness as a function of the ratio of indentation depth to coating thickness for three different thickness DLC coatings. For two of the coatings, there is a common plateau in the composite hardness response, which represents the hardness of the DLC coating. In the case of the thinnest coating, there is a maximum in the hardness response, but this is much lower than that obtained for the two thicker coatings. This can be explained by considering the depth at which the maximum occurs in the von Mises shear stress under the indenter. This depth is related only to the indentation size and the indenter tip radius. For equally sized indentations, this stress maximum moves deeper into the substrate if the coating thickness is reduced. Thus, for the thinnest coating, the substrate has begun to yield before the coating has reached its yield point. This new method is clearly an improvement on the old rules of
thumb for coating-only hardness measurement (e.g. Bückle’s <10% rule [39, 40]) as it can take account of the fact that, even for normal radii of indenter tips, indentation at shallow depths is essentially elastic and it is only when the mean indentation pressure reaches that required to create a fully developed plastic zone in the coating, that a plateau hardness value is reached. The position and extent of such plateaux will depend on the ratio of coating to substrate yield strengths and the new procedure is able to adapt directly to that. In this example of DLC on hardened tool steel, the plateau is achieved between 8% and 15% of the coating thickness and indentations at much lower relative depths would in fact yield an incorrect coating hardness result.

![Diagram](image)

**Figure 6**: Variation of $H_{IT}$ with ratio of permanent indentation depth to coating thickness for DLC (Diamond-Like Carbon) on steel. The solid horizontal line marks the hardness of the steel substrate.
Standard testing procedure
3.1 ISO 14577

ISO 14577 is the international standard for instrumented indentation of bulk materials and coatings [20]. The structure and content of the standard may be briefly described as follows:

**Part 1** defines standard methods for obtaining a valid data set of force, displacement and time points from which materials properties may be calculated. It also defines standard ways to calculate Martens Hardness ($H_M$), Indentation Hardness ($H_{IT}$), Indentation modulus ($E_{IT}$), Indentation Creep ($C_{IT}$), Indentation relaxation ($R_{IT}$) and Indentation work ($W_{total}$, $W_{elast}$, $W_{plast}$).

**Part 2** defines the geometries of acceptable indenters, the specification of acceptable force, displacement (and time) sensors and the requirements for calibration and validation of the test machine for force, displacement, time (indentation cycle), indenter shape (area function) and instrument frame compliance.

**Part 3** defines the specification for an acceptable calibration machine and gives a standard method for calibration of reference blocks.

**Part 4** (published in 2007, precursor versions published in [37,38]) provides a standard method for measuring hardness and plain strain modulus of coatings. It also provides refined methods for verification of indenter area function and instrument frame compliance.
Indenter geometry
4.1 Choosing the right tool for the job

ISO 14577 does not place any specific restriction on the indenter geometries that may be used to determine materials properties by indentation. Most instrumented indentation systems have the ability to change the indenter, some more readily than others. Being able to choose the most appropriate indenter geometry for a particular task is, without doubt, an advantage. Any particular indenter geometry has both merits and drawbacks. The most important attribute of an indenter, however, is the validity of its area function. There is little use in choosing a theoretically more appropriate geometry if that means using an indenter with an unknown or invalid area function. Similarly, it is only worth changing the indenter for one with a different geometry if it will act to decrease the total measurement uncertainty.

In general, hardness testing is best done with sharp pointed, high aspect ratio indenters that generate plastic deformation at low indentation depths and forces.

![Diagram showing von Mises Shear stress for indenter on coating](image)

Self-similar (constant aspect ratio) indenters are typically used, as they return a depth independent hardness value in a homogeneous material. In particular, the measurement of the hardness of coatings is only possible if plastic deformation is induced in the coating and not in the substrate. The shear stresses generated in an indented material vary with depth and reach a maximum value at approximately $0.47(A_c/\pi)^{1/2}$ below the indenter tip, where $A_c$ is the projected area of contact. Unless the yield stress of the substrate is vastly higher than that of the coating, it is best to keep the maximum stress to be inside the coating, see Figure 7 to Figure 9. If the contact area is not small enough, the maximum of the generated stress will always be in the substrate and may exceed the substrate yield stress before the coating yield stress is reached. This is best achieved by using high aspect ratio indenters with a small radius of curvature at the tip, so that high indentation pressures are reached at low depths and a small area of contact.
Modulus testing is, in many ways, less demanding of indenter geometry, since all that is required is an accurate knowledge of the projected area of contact as a function of contact depth. In practice, a number of measurement problems, such as fracture and creep, can be avoided if the indentation contact remains purely elastic. Larger radius of curvature tips remain in the elastic regime for a larger range of forces, although the indentation depths are reduced and care has to be taken that an increase in the depth related uncertainties does not swamp the gains made by reduction in the other uncertainties.

Figure 8: AFM image of indentation into amorphous-alumina coated aluminium. The extensive yielding of the substrate dwarfs the indent in the coating.

Figure 9: Schematic of an indentation where the substrate has yielded before the coating.
4.2 Indenter geometry considerations for advanced users

The following sections are aimed at advanced users and attempt to summarise some of the issues that might contribute to particular indenter geometry choices. Standard users should skip these sections and go directly to the flow charts in Section 4.3.

4.2.1 “Self-similar” indenters

In this context, a “self-similar” indenter is one that has the same aspect ratio at all indentation depths. All sharp pointed cones and pyramids fall into this category. ISO 14577 parts 1 to 3 [20] do not place any specific restriction on the indenter geometries that may be used to determine materials properties by indentation. This is even the case for Indentation Hardness, where, in all cases, the average indentation pressure is a measurable parameter. However, the measured indentation pressure (and so indentation hardness) is only depth independent for self-similar indenters; for any other geometry the value returned will vary with indentation depth. Self-similar indenters are however essential to implement the ISO14577 part 4 method for measuring the hardness of a coating [37,38]. This method depends on the assumption of depth independent hardness to determine the point at which the substrate is starting to affect the measured property. Self-similar indenter geometries have the additional advantage that they tend to a sharp point. This makes them ideal for measuring the hardness in thin coatings or shallow surface regions. The maximum shear stress is usually generated directly below the contact point at a depth proportional to the indenter radius and so a low tip radius is necessary if the hardness of material near to the surface is to be measured. Other than cones, most self-similar geometries are faceted (Vickers, Berkovich, Cube corner). The edges where two facets meet are relatively sharp and can cause cracks which radiate out from the ‘corners’ of the indentation. This can be desirable if adhesion or fracture toughness is being investigated, but in general is to be avoided. The tendency to fracture reduces as the opening angle of the indenter and/or the tip radius increases.

4.2.2 Vickers

This geometry is identical to that used in the conventional Vickers hardness scale and so comparison of data is in principle easiest. However, the four sides of the pyramid frequently do not meet at a point. In conventional HV scales this causes a direct error, which is managed by limiting the acceptable size of the ‘line of conjunction’ to 500 nm. For the smaller indentations typical in instrumented indentation this is a large deviation in geometry and for this reason a three-sided equivalent, the Berkovich, was invented.

4.2.3 Berkovich indenter

Invented as an improvement on the Vickers geometry, this is possibly the most popular indenter geometry in instrumented indentation. Since three facets cannot form a line of conjunction and must meet at a point, this geometry can be used down to very small depths. Ultimately a real indenter is not infinitely sharp and has a rounded end. Thus a critical parameter is the radius of this tip, since this determines the depth below which the indenter has effectively a spherical geometry. It is perhaps of interest to the historically inclined that the facet angle of a Berkovich was originally specified to be 65°3’ which corresponds to a three-sided pyramid with the same surface area to depth ratio as a Vickers four-sided pyramid (facet angle 68°) [41]. The most commonly used version of the Berkovich is now the “modified Berkovich” with a facet angle of 65.3°. It is not clear if this was due
to propagation of a typographical error or to the fortuitous fact that this angle ensures that the cross sectional area to depth ratio of a modified Berkovich is the same as that of a Vickers indenter.

4.2.4 Cube Corner

This is also a three-sided indenter, but has a much higher aspect ratio (lower opening angle) than the Berkovich. As such it is favoured for indenting very thin coatings or for generating cracks more easily. The literature suggests that this geometry is more difficult to calibrate and may exhibit an indentation response that deviates from the standard contact mechanics employed in ISO 14577:2002. At very shallow depths its main advantage is an ability to support a tip with a very small radius of curvature. It is consequently more fragile than other geometries and has a tendency to wear more rapidly.

4.2.5 Spherical indenters

A spherical geometry has the advantage that it is the most amenable to analysis by standard Hertzian contact mechanics. The high initial area of contact also tends to increase the force range over which an indent remains fully elastic, which can be a significant advantage if the elastic response of a thin film or multilayer structure is to be studied or data is to be compared to models. Elasticity can be modelled analytically, whereas plastic deformation usually requires numerical methods, such as Finite Element Analysis, which can be more expensive and more computer intensive. Despite the well-established conventional hardness scales of Rockwell and Brinell, the sphere has traditionally been avoided for the determination of indentation hardness, principally because the indentation hardness number obtained is no longer depth independent. It does, however, represent the mean indentation pressure. New research is beginning to relate the mean pressure of indentation and relative indentation size to properties such as yield stress and empirical algorithms now exist to obtain tensile properties such as stress-strain curves by indentation, see Figure 10 [42,43,44,45,46].

Unfortunately, it is extremely difficult to manufacture a spherical diamond. Most diamond ‘spheres’ are polyhedral and/or vary in radius with indentation depth. An example is given in Figure 11. This shows a nominal 10 µm radius indenter which was observed to leave polyhedral indents and was measured to have a radius varying between 35 µm at the tip to approximately 9 µm at larger depths. The wear rate of the different crystal directions in diamond varies by approximately two orders of magnitude and so it takes considerable quality control to obtain a spherical indenter that is sufficiently constant in geometry to apply Hertzian mechanics with a single value for radius. More typically a radius function has to be used e.g. Jennett and Bushby [47]. Diamond is the preferred indenter material because of its extremely high modulus and hardness. If a compromise on these properties is acceptable, spherical indenters can be made from other materials, such as tungsten carbide ‘hardmetal’, ruby or sapphire, which are routinely manufactured into highly spherical ball bearings.
Figure 10: FEA vs. experiment for oxygen-free copper, from Lim & Chaudhri [42] demonstrating an increase in yield stress with decrease in indenter radius. Plots of mean pressure (Pm) vs. ratio of indent radius to tip radius (a/r) are analogous to true stress vs. strain.

Figure 11: Contour plot of Metrological AFM data taken from a spherical indenter. Note the severe polyhedral nature and the rapid change in tip radius
4.3 Flowcharts

The following flowcharts describe how to select indenter geometry and indentation parameters to measure coating properties. Figure 12 refers to the indentation modulus of a coating and Figure 13 to the indentation hardness of a coating. To enable the analysis to deal with coatings of all thicknesses and indenters of all shapes, the indentation sizes are normalised as \((a/t_c)\): the ratio of the contact radius \(a = (A_c/\pi)^{\frac{1}{2}}\) to the coating thickness, \(t_c\). This normalisation may be used for both hardness and modulus, but, if self-similar geometry indenters are used, the more intuitively obvious normalisation of \(h_c/t_c\) may be used for hardness measurement.

Many of the choices are equally valid for the indentation of bulk materials, although there is no need to indent over a range of depths (indentation sizes) unless a variation of properties with depth is being investigated. Where a surface has been modified, however, the ‘case’ depth can be considered to be the coating thickness.
To measure Indentation Modulus of Coating

Choose suitable indenter

Perform test indentations (two depths)

Coating fractures?

Yes

Change indenter? Increase indenter radius or opening angle

Are forces below fracture limit Possible?

Yes

Purely elastic response?

Yes

Extrapolate Linear Fit to find coating EIT* at a/te = 0

No

Hold at F_max until creep rate reduced

Coating creeps?

No

Increase radius?

No

Perform indentations over the range zero to a/te < 2 (hard or brittle) a/te < 1.5 (ductile)

Input or estimate Poisson ratio to calculate coating EIT

Quality Assurance Check
Inspect indentations for cracking and pileup using Optics/SEM/AFM

Figure 12: Flow chart for selection of indenter geometry and indentation parameters to measure coating properties: Indentation modulus of coating. \(a\) is the radius of contact estimated as \((Ac/\pi)^{\frac{1}{2}}\) and \(t_c\) is the nominal coating thickness.)
Figure 13: Flow chart for selection of indenter geometry and indentation parameters to measure coating properties: Indentation hardness of coating ($h_c$ is the indentation contact depth and $t_c$ is the nominal coating thickness).
How to set-up a test
### 5.1 Sample mounting

The key issues for sample mounting are that the region to be indented should be perpendicular to the indentation direction and the whole sample mounted into the instrument in a way that does not add any significant compliance (i.e. the sample does not move or flex when indented). A useful tip is to apply some pressure when bonding samples to an instrument mounting-stub. It is surprising (but true) that resting a sample onto a drop of cyanoacrylate adhesive, without pressing the material surfaces together, can significantly increase the measured contact compliance. When using hot-melt wax, it is best to move the sample around in the molten wax to ensure the removal of air bubbles, minimise the wax thickness whilst maximising the surface area of the bond. This is important because the wax or glue will tend to have a low elastic modulus (~1 GPa) compared to most inorganic materials and so its contribution to the absolute compliance of the contact is reduced as the area of mounting contact increases and the glue/wax layer thickness decreases.

Any additional movement that is not directly related to the indenter penetrating the sample will distort the results. This includes drift and vibration, which are dealt with later. Other than that, the area to be indented should be of low surface roughness if possible. The effect of surface roughness is to make detection and definition of the surface position (i.e. the zero point of the indentation) and area of contact with the indenter more uncertain, see Figure 14.

A good guide is that the Ra of a surface should be less than 5% of the indentation contact depth $h_c$. However, re-preparation of a surface can alter its properties and so it is often best not to attempt to reduce surface roughness unless it is the dominant cause of uncertainty in the measurement.

![Figure 14: Indentation into two surfaces with the same Ra but different lateral frequency of roughness resulting in (a) a stiffer contact, (b) a compliant contact.](image)
5.2 Sample cleaning

It is important that foreign bodies such as dirt, fluids or dust do not get into the indentation contact and distort the result. The surface should therefore be as clean as possible. If the surface requires cleaning, then the method chosen must not alter the surface properties. For this reason, it usually best to keep to gentle solvents such as alcohols, taking care that water or other residues are not deposited onto the surface in the process. For example, a jet of subliming carbon dioxide may be used if the surface temperature is maintained above the dew point. The use of ultrasound is known to increase surface damage and this can be a problem when cleaning some coated surfaces.

5.3 Cleaning the indenter

It is often forgotten that cleanliness of the indenter is as important as that of the sample. Experience has shown that a diamond surface is very attractive to dirt and dust particles and a “tide mark” of loose debris can build up around the indenter, see Figure 15. Of more concern is the possibility that sample material sticks temporarily or permanently to the indenter. This can change both the nature of the contact response and the effective shape and size of the indenter. For this reason, it is good practice to regularly indent into a reference material, at least between each sample tested, so that temporary contamination can be detected and bad data identified.
If an indenter does become contaminated, it is important to exercise extreme care when cleaning it. Any damage will change the area function and invalidate any future experiments. In particular, sideways forces applied to the indenter can damage the tip very easily. A full cleaning procedure is given in Annex D of ISO 14577 part 1. This recommends inspection in an optical microscope with a magnification of x400 or greater to determine if an indenter is contaminated. The cleaning procedure consists of a series of increasingly aggressive options. These start by using the surface tension forces in a drop (or cotton wool ball) of solvent; move on to indenting into freshly exposed surfaces of expanded polystyrene (the plasticizer is a good cleaning agent); then on to indenting into wood or aluminium in an attempt to get the contaminant to stick to the cleaning sample rather than the indenter. The principle of the procedure is to increase the severity until the contamination is removed and then finish with a solvent step to remove any residue from previous cleaning steps.

5.4 Choosing a test environment

The ideal environment is completely stable in temperature and relative humidity and is free from any vibration or electrical or magnetic disturbance. Real environments are rarely like this and so it is often necessary to protect the instrument from disturbances. If possible, it is wise to conduct experiments on Certified Reference Materials (CRM) to determine which environmental disturbances are causing a significant effect. The homogeneity and reproducibility of CRMs make them ideal for characterising instrument uncertainty. In this way, optimising effort can be targeted to best effect.

5.5 Environmental temperature control (active & passive)

Even very small and slow changes in temperature, can result in a large displacement drift rate due to differential thermal expansion. The level of effort required to minimise drift will be determined by the speed and depth of the indentations being made, but there are a few simple precautions that can be taken. Allow time for the samples and the instrument to reach thermal equilibrium. This can be accelerated by minimising the time spent handling the instrument and samples, using a shallow water bath to cool down samples mounted by hot wax methods, or simply mounting samples in advance of testing and keeping them in or near the instrument enclosure so that they reach thermal equilibrium while waiting to be tested.
It is wise to shield the instrument from draughts and create a passive space that can come to thermal equilibrium. A thermally insulating shield will tend to gear down any external changes in temperature and delay their effect. The external environment can then change temperature about a constant average value without disturbing the instrument temperature, which will track the long-term average and not “see” the short-term fluctuations, see Figure 16. This is particularly important if the environment is air-conditioned or is temperature controlled with coarse PID feedback heaters, which tend to cause a saw-tooth temperature variation with a short time period. Residual thermal drift can be measured using a constant force hold period. If the drift rate is constant over the time taken by an indentation cycle, the drift can be calculated from the measured drift rate and the time stamp on each displacement point and subtracted from the measured displacement. This is why, even in a poorly controlled environment, it is important to arrange for the instrument temperature to vary slowly e.g. by using shielding to damp out oscillations. Rapid changes in temperature would result in a change in sign of the drift rate and thus any attempt at correction may make things worse if the drift rate were to change or to reverse during a cycle.

It is important, wherever possible, to keep any intermittent sources of heat outside the enclosure. In some older instruments, it may be necessary to ensure that the indenter shaft is made of invar between the indenter and the displacement transducer so that the heat generated in the force transducer coil does not cause thermal expansion of the indenter shaft. It is also important to ensure that stage stepper motors are configured to be normally off, so that they do not cause heating of the instrument enclosure. In other cases, the speed or voltage of a stage motor may need to be limited to reduce the thermal disturbance it causes to the test environment.

Figure 16: Use of a thermal insulating enclosure both smoothes out temperature variations and gears down the effect of a temperature rise.
5.6 Vibration isolation (active & passive)

A thermally insulating enclosure will also provide some protection from airborne vibrations. Lining the enclosure with acoustically damping material will help further. High frequencies are easier to absorb than low frequencies, but generally it is the low frequencies that are the biggest problem. Passive absorption of low frequencies requires layers of dense material. Footfall and building vibrations can, ultimately, be reduced by excluding people from the instrument environment. Alternatively, it is common to use active or passive anti-vibration tables. The simplest of these consists of a heavy slab on top of soft springs. For this reason air springs, football bladders and rubber mats are popular. Elastomeric shock cords also work well, are cheap and have the advantage that the system’s resonant frequency can be easily tuned simply by altering the pendulum length of the suspension cords. It is useful if the springs also have some form of damping as this reduces the build up of oscillations at the resonant frequency. Loose sand beds can provide good damping for some floor mounted systems, at least until the sand becomes compacted. More sophisticated anti-vibration damping systems use carefully set up systems of opposing springs that generate an effective negative compliance or use piezo or electromagnetic actuators with feedback and feed-forward control.

[Advanced users: Usually the most sensitive system available to the user for determining the performance of an anti-vibration solution is the instrumented indenter itself. An easy way to evaluate vibrational noise is to get the instrument to perform an indentation cycle consisting of a long, constant force, hold period in contact with a hard surface. The average drift rate is obtained from a linear fit to the data. The standard deviation of the displacement signal, after subtraction of the linear drift, is a sensitive measure of the vibration of the instrument. It is instructive to compare the amplitude of the vibration in contact with that experienced when the tip is out of contact. The amplitude out of contact multiplied by the spring stiffness is a useful estimate of the vibration-induced force uncertainty.

Diagnosing and solving a vibrational problem can be assisted by determination of the resonant frequency causing the problem. Scanned probe microscopes (SPM) have fast enough data rates to be able to generate an FFT of the displacement vs. time data to identify problem resonances. It is increasingly common for instrumented indentation systems to incorporate SPM systems and they, themselves, are increasing in data rate. Active anti-vibration platforms are also sometimes equipped with a diagnostic capability. Thus, a high level of sophistication in environmental characterisation is increasingly possible. Transient episodes of vibration (e.g. due to footfalls, people shouting or doors closing) will show up as a burst of higher amplitude noise. A good quality control practice is always to perform a few test indentations on a reference material before and/or after each set of experiments. Inspection of the hold periods in the reference material indentations is a reliable way of identifying increased uncertainty due to vibration.

Figure 17 and Figure 18 show data from QC indentations into a CRM, which identified a vibration problem. Plots of the thermal drift data (after subtraction of the average drift rate) in Figure 17 show vibration induced displacement uncertainty of nearly 4 nm compared to the normal uncertainty of < 0.5 nm. Figure 18a shows how the vibration affected the force-displacement data. The cause was a connecting cable being stretched taut and ‘bridging’ the anti-vibration table. When this cable was slackened and proper strain relief put in place, the result was as shown in Figure 18b.]
5.7 Typical Indentation cycle

ISO 14577 requires that the indentation cycle be described in the test report. This is because the handling of creep etc. has a significant effect on the hardness value. For instance, the hardness will change depending on the absolute amount of creep that is allowed to occur before the force is removed. Thus, although a valid strategy to compensate for a high creep rate and a poor thermal stability in the environment is to increase the force application and removal rates, this will only yield the same modulus value; the hardness value will depend on the total amount of creep incurred, which is proportional to the total time under force (including during force application) [48]. ISO 14577-1 suggests that a typical indentation cycle is as follows:

- Approach < 20 nm/s
- Optional Thermal drift Hold at constant applied force typically 30-60 s
- Constant rate of force increase 30 s
- Hold at constant applied force 10 s (or until creep rate is acceptable)
- Constant rate force decrease to 10% of maximum force 27 s
- Optional Thermal drift Hold at constant applied force typically 30-60 s
- Reduce applied force to zero 3 s

However, if a data rate ~100 Hz is possible, it is not unreasonable to reduce each part of the cycle to 1-2 s and the total to less than 10 s if the creep rate is acceptable. This makes hardness values difficult to compare between fast and slow instruments, but does bring the cycle back into line with conventional hardness testing.
Figure 18: Quality control data obtained before and after diagnosis and remedy of a vibration problem. Each Force-displacement plot is a single ISO14577-1 typical indentation cycle performed on a Certified Reference Material.
How to calculate results
The raw data generated by an instrumented indentation instrument requires correction before it can be analysed to obtain materials property values. The main corrections necessary are to assign or refine the assignment of the zero point in force and displacement, correct for displacement drift and remove the effects of instrument compliance.

6.1 Zero Point determination

All of the contact mechanics equations used to calculate the indentation depth and area of contact assume that the indentation force displacement curve starts when the indenter first touches the surface of the material being tested. In other words, the applied force is zero when the indentation depth is also zero. An error in the zero point causes directly proportional errors in hardness and modulus. For self-similar indenters, a 1% error or uncertainty in the zero point displacement results in an uncertainty or error of 1% in modulus and 2% in hardness.

In practice, all machines approach the surface and start the force-increasing curve from a small but finite indentation depth and applied force. In some instruments, the approach data is recorded for later re-inclusion into the force-displacement curve once the zero point has been determined. In other instruments, the data is not recorded. If the data is recorded, it can be inspected after the indentation and the data point where the stiffness increases can be taken to be the first touch on the sample and therefore the zero point. In dynamic indentation there is a large phase change when the indenter touches the surface, which can be easily detected. These surface detection methods can be a sensitive measure of contact with a low zero point uncertainty of the order of 0.1nm. This method’s only disadvantage is that it can be too sensitive to asperity contact when the surface (or indenter) has significant roughness.

In systems that do not record the approach data, the only option is to either accept the force and displacement offset as a fixed error, or fit the force-increasing curve and extrapolate back to zero. In principle, most indenter tips are effectively spherical at first touch and so fitting using the Hertz equation not unreasonable. In the case of asperity contact, however, a straight line fit may be more appropriate. The uncertainty in the back extrapolation method will be affected by:

- Inhomogeneity of the sample (including the cracking of native oxide layers),
- surface roughness (of indenter and sample surface),
- the indenter geometry,
- the noise on the data (e.g. due to vibration),
- how well the mathematical function fits the trend in the data,
- the length of extrapolation (i.e. the size of the initial force).

When indenting surfaces of homogeneous materials with an average roughness (at the indentation site) of less than 1nm, where the data has a displacement noise of less than 1nm and is extrapolated back from an initial force of between 2 µN and 50 µN using a mathematical function that closely describes the form of the data, it is possible to achieve zero point uncertainties in the range 0.1 nm to 2 nm. However, this estimate should not be relied upon in any specific instance as the actual uncertainty can be strongly affected by any or all of the factors identified above.
6.2 Thermal drift correction

Instrumented indenters will always exhibit some displacement drift. Sometimes this may be insignificant when compared with the indentation depths of the indent and therefore may be safely ignored. In other cases, particularly for shallow indentations in the Nano and low micro ranges, this may be a significant proportion of the total indentation depth. Measuring the displacement drift rate at a constant contact force and using the time stamp on each data point to calculate and apply a displacement correction can correct for drift. A decision tree from the INDICOAT report is given in Figure 19.

![Decision tree for optimising the strategy for correcting data for residual thermal drift](image)

[Advanced Users: The underlying source of drift is usually attributed to differential thermal expansion. However, it is useful for the advanced user to be aware of other possible sources of drift. Immediately after contact, or a removal of force, capillary water or other condensed liquid layers on the surface can be squeezed out or drawn back into a contact, causing an apparent displacement change [49]. Similarly, the stage may take a few seconds to settle after a move while lubrication films are squeezed out. Some stage motors can generate significant bursts of heat even for short moves. If the force of contact is high enough to exceed the elastic limit of the material, plastic deformation and indentation creep may occur. In the case of power law creep materials, it may take a while for the creep rate to decay away. Indeed, in some instances, the creep may not be at the indenter. Some machines measure differential displacement of the indenter with respect to the instrument, which sits on a hard sapphire reference ring that is in contact with the sample itself. In some instances, this ring can indent the sample surface and exhibit indentation creep, which appears as displacement drift. Most of these effects are transitory and it is best to allow 10 – 20 s for them to decay away before measuring the underlying drift rate. This can be done by holding for 60 s but fitting only the last 40 s of data.]

---

Figure 19: Decision tree for optimising the strategy for correcting data for residual thermal drift (after INDICOAT final report [38]).
6.3 Frame compliance correction

Frame compliance is the amount of additional displacement recorded by the instrument due to its own load frame flexing when a force is applied, see Figure 20. Typical instrumented indentation instruments range in frame compliance from 0.1 nm/N to 5 nm/mN and this can cause a significant error in both the indentation depth and in the measured contact compliance. Instruments that measure indentation displacement referenced to the top of the sample can appear to be around an order of magnitude stiffer than this, because most of the frame flexure is “outside the measurement loop”. Since frame compliance is very reproducible, it is normal to calibrate an instrument, either directly or indirectly using a CRM, to provide a value that is used to correct all subsequent data [22,38,50]. It is wise to check the frame compliance periodically, e.g. when the indenter has been changed, to ensure that nothing has changed. Experience has shown that slack mounting of indenters or a failure to bolt stages down tightly can easily increase the frame compliance by a factor of two or more.

6.4 Fitting to the force removal curve

The contact compliance is determined by differentiating a mathematical fit to the data taken during force removal. For highly plastic materials, the Doerner and Nix method makes use of the fact that the force removal curve has very little curvature and an easy approximation to the necessary tangent is to use a linear fit to the top portion of the force removal curve (typically 30%) instead. A more accurate model for the force removal curve is a power law, where the exponent, \( m \), depends on the geometry of the indenter. Some analysis software forces the exponent to be the theoretical value (i.e. 2 for self-similar indenters and 1.5 for spherical contacts) but others allow the exponent to vary and this does tend to result in a better (lower sum of residuals) fit. The tangent is obtained by analytical differentiation and is usually evaluated at the maximum indentation force. Commonly the top 80% of the force removal curve is fitted, however, if the curve is not continuous over that range, for example a fracture occurs, then a shorter portion may be fitted.

A more serious problem occurs if the data in the force removal curve is distorted by residual displacement creep or uncorrected drift, at \( F_{\text{max}} \) see Figure 20. Drift can be corrected for if the drift rate is known, but creep is not so easy to deal with. Many materials exhibit creep, i.e. a continued increase in indentation depth at constant force. For most metals, this is genuine power law creep behaviour and some glasses exhibit a similar effect. It is therefore wise to use a hold at \( F_{\text{max}} \) to allow the creep rate to reduce. The creep rate reduces with time for two main reasons. There is an exponential decrease in creep rate for power law creep materials held at a constant stress. Secondly, indenters usually have a geometry such that their projected area increases as indentation depth increases. Thus, under a constant force, the indentation mean pressure reduces as the material creeps. Power law creep is very sensitive to a reduction in stress and so the creep rate tends to die away quite rapidly if it is given a little time. Severe creep rate at the point of force removal results in a “nose” on the force removal curve, where the positive going creep rate beats the elastic recovery of the material. Blind fitting of such curves can generate negative slopes and thus negative elastic moduli, which are clearly unphysical (i.e. impossible) results. The error (uncertainty) in the tangent to the force removal curve is directly proportional to the ratio of the creep rate at \( F_{\text{max}} \) to the force.

Modern instruments have increasingly higher data acquisition rates and it is tempting to speed up the indentation to “beat” the drift rate. This is a valid strategy, but care should be taken that there are not transiently high and changing drift rates captured in the indentation cycle that violate the assumption of a constant, linear drift rate throughout the experiment. [22,38,50].
removal rate. Many instruments’ own software allows the user to export indentation data into a spreadsheet compatible format. A simple linear fit to the last part of the $F_{\text{max}}$ hold data is a rapid indicator of the scale of the problem. If creep is a problem, it is best to use an asymmetric indentation cycle, i.e. a slow force application, a longer hold than usual, followed by a maximum rate force removal. Note that it is necessary to capture enough points in the force removal curve to obtain a valid and accurate fit. The maximum rate of force removal is therefore limited by both the speed of the instrument force control circuit and the maximum data capture rate.

![Figure 20: Frame compliance causes a force dependent displacement (green dotted line) that is reversed upon force removal (blue line). Uncorrected thermal drift and any creep allowed to occur cause a permanent increase in displacement (red lines).](image)

**Advanced Users:**
If a material is visco-elastic or visco-plastic, then it is likely that all hold periods will exhibit creep due to viscosity effects. Furthermore, the creep rate at $F_{\text{max}}$ will be positive, whilst the creep rate at 90% force removal will be negative and the indentation data will show a very large negative going step, roughly matching the positive going creep in the $F_{\text{max}}$ hold. Estimating the thermal drift from such a hold period is impossible. The changing viscous response during force removal will also affect the shape of the force removal curve, invalidating estimates of contact stiffness obtained from the force removal slope. An answer to this problem is to switch from quasi-static indentation to dynamic indentation (also called continuous stiffness, continuous compliance and nano-DMA). Dynamic indentation uses a superimposed oscillation (usually force in a force controlled instrument) and a phase lock amplifier to detect the amplitude and phase of the resulting oscillatory displacement. As this is almost equivalent to an instantaneous measurement of the slope of the force removal curve, the original patent coined the term “continuous stiffness” measurement (CSM) [51]. The stiffness and phase angle can be used in a kinematic model of the system to obtain both loss and storage modulus of a visco-elastic material, assuming it is operating in the region of frequency and temperature that is above its glass transition temperature (Tg) and below its visco-elastic time constant. CSM does not, however, solve the thermal drift measurement problem. The best way is to separate the thermal drift from the visco-elasticity is by holding contact with a hard reference surface immediately before and after an indentation to determine the drift rate. (Indeed it is possible here to turn the situation to advantage by creating a time weighted drift rate if the two measurements indicate a difference in drift rate over time).
If dynamic indentation modes are not available, choosing a fixed cycle and accepting that the data will provide only a repeatable comparison between materials is one approach. Other approaches are to make the indentation cycle as fast as possible, in an attempt to “beat” the time constant of the material; or to cool the sample below its Tg.

Some materials, although not viscous, still exhibit some elastic hysteresis, which distorts the functional form of the curve away from a pure power law. This can be easily detected if the fit is subtracted from the data and the residuals plotted as a function of depth. A good fit will have residuals that are evenly distributed about zero. If the data is of a different functional form to the function being fitted, there will tend to be a systematic bias to the residuals. It is not unusual for a fit to be restricted to the top 40% of the force removal curve in these cases. It may be of interest to those operating at the highest resolution, that even much loved and trusted materials, such as fused silica, do not exactly obey a power function over their force removal curves. Recent research is indicating that a refinement to the Oliver and Pharr contact mechanics used in ISO14577 will solve this problem and it is hoped that this will be incorporated in the 2007 revision of the standard. An early description of this approach is contained in the INDICOAT project final report [38].

Although the best point in the indentation cycle for a thermal drift hold is arguably at the very start of the indentation, when the chances are greatest that the contact is fully elastic, it is often useful, for diagnostic purposes, to introduce an additional hold near the end of the force removal curve. If creep is a problem and the material is soft enough that it yields under the contact force, the drift rate measured at contact may be affected by creep. The drift rate at 90% force removal, however, is much less likely to be affected by power law creep and may therefore be more reliable assuming other effects are not present.

A further problem can be noise in the force removal curve at or near $F_{max}$. This can be due to force control issues in some high force instruments or simply due to poor vibration isolation. It is common for power law fits not to be weighted and so noise at the start of the force removal curve can have a disproportionate effect on the gradient. There is no easy answer to this, but some users have reported success in fitting as far up the curve as possible, e.g. to 98% of $F_{max}$ and then extrapolating the curve to $F_{max}$ to obtain the correct estimate of $h_{max}$. This is not a standardised procedure. Extrapolation is always dangerous and this should only be attempted if the user is willing to inspect individually each curve and each fit to ensure that the strategy is not introducing more error than the problem it is intended to solve.]
Measurement artefacts to consider
Anyone who has performed a hardness test will know that some materials pile up around the indentation and some appear to sink in. What is less well appreciated is that this is not a fixed material property and may depend on the work-hardened state of a material and upon the level of constraint upon the development of the plastic zone beneath the indentation. Soft or annealed material tends to sink in. Work-hardened material tends to pile up. Interestingly, Figure 21 shows from AFM line traces that the aspect ratio of the pile up increases with indentation depth, which tends to indicate that the area in contact with the piled up material may not be able to support as much pressure as the contact area with material below the average surface plane. Figure 22 shows how shallow indents into an aluminium coating sink in and yet deeper indentations pile up as the plastic zone is constrained by the hard (unyielding) substrate.

![Figure 21: Line traces through spherical indentations at different applied forces for a t = 0.31 µm Al film on BK7. The units of the X and Y axes are nanometres. (from Spary, Jennett & Bushby [43])]()

![Figure 22: Berkovich indents at increasing forces (depths) into 1050 nm Al on Bk7; indentation forces 3 mN, 10 mN and 50 mN, left to right respectively (taken from INDICOAT final report [38])]()
7.1 Discarding bad data

7.1.1 What to look for

There is a trend towards automating the analysis of indentation data. This is useful in the sense that it allows higher instrument throughput and better statistical sampling but there can be problems with outliers.

As with any data set, the temptation to discard data simply because it does not fit the expected answer should be resisted strenuously. Conversely, data that is clearly incorrect should be discarded to prevent it artificially increasing the error and uncertainty of the results. Proper validation of the indentation data set requires visual inspection of each indentation and each load-displacement curve (including attendant fits, tangents and drift calculations).

7.1.2 Criteria to discard

The most rigorous form of data quality control is to inspect everything: The physical indentation to check for pile up, sink in, fracture etc.; The data itself, i.e. the corrected force-displacement curve, the fits to both the thermal drift and the force removal data, the tangent and the residuals to the force removal curve fit. There are a number of checks that can be made which may indicate that data should be further corrected or validly discarded:

1. Observed discontinuities in the force removal curve, in particular, any inside the fitted region. These are usually caused by fracture or pressure-induced phase transitions. Unless these are specifically being studied, they are best avoided.
2. There are grounds to discard the data if there is large displacement noise or if the thermal drift data is not coincident with the fit towards the end of the hold period, or if the drift rate is variable and therefore unlikely to be represented by the drift rate calculated. If two hold periods have been used, a very rigorous check is that the thermal drift correction from one results in the correction of the drift in the second. Care is needed, however, as not all drift is thermally induced and so it is not unusual for there to be some residual discontinuity in the data.
3. If the data does not begin at the force and displacement zero point, the correct zero point needs determination and any corrections applied to the whole data set to ensure that each point is correctly referenced to the force and displacement zero point.
4. If there are large and systematic residuals to the fit to the force removal curve, there may be a problem with the fit range.

[Advanced User: Traditional hardness does not take the effect of pile up into account. In Vickers or Knoop hardness scales, where the measured parameter is the indentation diagonal, it is known that the positions of the corners of such indents are little affected by pile up, which occurs mainly on the indenter facets. Thus the measured indent size will reduce when pile up occurs making the material appear harder. Rockwell indentation is similarly afflicted. The Rockwell number is based on the difference in indentation depth at a small pre-load before and after applying the maximum test force. Now the indentation depth is limited by the maximum indentation mean pressure that the material can support, i.e. a fixed contact area will be supported by the material. This, however, will include the pile up and so the instrument will record a smaller depth difference, making the material appear harder, when pile-up occurs. Instrumented indentation is therefore not unique in being unable to take account of any pile up when estimating the contact area by a contact mechanics equation.]
5. If the tangent to the force removal curve fit is not tangential but passes inside the data, there is a problem.
6. If the tangent has a negative slope or there is a visible “nose” on the force removal curve, e.g. due to creep, the data is invalid and the ratio of creep rate to force removal should be calculated to indicate the scale of the problem.
7. If the force removal curve ends at a significantly non-zero force, this may indicate a significant error in the data due to an incorrect spring stiffness calibration or additional surface forces acting on the indenter.
8. If the hold at maximum force indicates a large positive creep rate and the hold at 90% force removal indicates a large negative creep (recovery), the data is likely dominated by viscosity effects.
Equipment calibrations
Calibration is the means by which it is established that the milli-Newton being applied and the nanometre being measured in one instrument are the same as in another instrument in another place and at another time. Although it is possible to calibrate to any arbitrary unit, reproducibility between instruments is only guaranteed if they are both calibrated to the same unit and that this unit is maintained to be constant. This is achieved by ensuring that the calibrations done are traceable to the absolute definition of the (SI) unit required. In practice this means being able to trace an unbroken series of calibrations back to the “realisation” of the absolute SI unit that is maintained in a National Measurement Institute (NMI). The Mutual Recognition Agreement of 1999 assures that each NMI SI unit is exactly equivalent to another and that you are therefore equivalently calibrated to any other traceably calibrated instrument in the world. Since it is the business of NMIs to maintain the stability of their calibrations, traceable calibration is likely to be the lowest uncertainty route for reproducibility over time. Since traceable calibrations always have a statement of uncertainty attached, this also makes calculation of the test uncertainty budget easier.

The key calibrations necessary for accurate instrumented indentation are:

- Force
- Displacement
- Time
- Indenter Area function
- Frame Compliance

Force calibration can be achieved by hanging weights on the indenter shaft, but it is often easier to simply indent into an electronic top-pan balance or a calibrated load cell, as this allows many more points to be obtained for the same effort.

Displacement calibration is most directly obtained by using an interferometer, particularly if it is a differential path design so that vibration and thermal drift are less of a problem. Newtons rings methods have been used, but these tend to be unacceptably fiddly to set up and time consuming to perform. A more user-friendly approach is to use a transfer calibration device such as a calibrated closed-loop piezo driven stage. This has excellent reproducibility and more than adequate resolution. Various “quick and dirty” approaches are also possible, where a greater uncertainty can be tolerated. These include using the instrument sample stage as a calibrator.

Time is less critical unless very accurate comparisons are being made between instruments as this is used mainly for drift rate corrections that require only a repeatable internal measure. If exact cycle times and creep rates are to be measured, however, then time must also be calibrated accurately.

Indenter Area function may be calibrated directly e.g. by AFM [21,22], or indirectly by indentation into a certified reference material (CRM) and reversing the usual equations by inserting the certified reference value and obtaining the corresponding indentation area for that particular indentation depth [22,47,52]. There is greater uncertainty involved in the indirect method as the value obtained includes all the uncertainties of the instrument in addition to any uncertainty in the CRM itself. It is tempting to use the indirect method as the only requirements are machine time and a CRM or two, however, it can be a lot of work to obtain a high quality area function by this method as many indentations are required to obtain ‘good statistics’. Some instruments are able to use a multiple partial unloading method, whereby the load is increased and then decreased a fraction to allow the contact stiffness and so the modulus or hardness to be evaluated at intermediate points in the force range. This reduces the time spent moving the stage, but does reduce the statistical sampling of the
CRM as fewer indentation positions are used. Even using multiple unloading, the resolution of the indirect method can never match that of AFM.

Figure 23 compares indirect indentation area function data (separated individual symbols) with that obtained from a traceably calibrated Thermo-Microscopes M5 AFM in metrological (‘zdet’) mode. The indentation data has sufficient uncertainty that a continuous polynomial fit through it would miss the real kink in the data that is due to the Berkovich tip being damaged. The AFM data has less uncertainty and has over 4000 points over the same depth range and easily picks out the damage in the indenter. An added advantage of the AFM, is that the damage is imaged too, see Figure 24.

Most indenters have an easily described area function far from the tip. Cones and pyramids are a simple square law of indentation depth. The problem is fitting a mathematical function that is able to describe the shape of the indenter at the tip where it is more spherical. High order polynomials or reciprocal power series are a popular choice but the high density of data points in the AFM makes it possible to use a much more accurate mathematical function to fit the data than could be the case with the restricted number of points in the indentation area function. The statistical expression of this is the number of degrees of freedom the data set has, which is one less than the number of points in the data set. If a mathematical fit uses more parameters or coefficients than there are degrees of freedom then the fit will pass through every point, fitting the noise in the data rather than averaging out the noise to obtain the underlying function. With 4000 degrees of freedom available a 2nd order B-spline function with 23 knots (or more) is easily possible [11]. This is a much better fit to a spherical tip transition as can be seen from a plot of the residuals in Figure 25.

Figure 23: From Aldrich-Smith, Jennett and Hangen [53]. Area function for a damaged Berkovich indenter. The individual symbols represent a number of indirect area functions obtained by indentation into a series of different reference materials. The apparently continuous line with a kink at the damage point is in fact 4000 AFM data points obtained over the same depth range.
Figure 24: From Aldrich-Smith, Jennett and Hangen [53]. Metrological AFM images of the same damaged Berkovich indenter as in Figure 23. Left image is 5 µm x 5 µm and right image is 2 µm x 2 µm. Both images are 512 x 512 pixels.

Figure 25: From Jennett and Meneve [11]. Residuals to a 23 knot spline fit are evenly distributed about zero demonstrating an excellent fit. The increase in residuals near zero is due to the proportionate increase in the uncertainties due to finite pixel size.
Calibration of frame compliance is usually done by indentation into a CRM. Direct measurement of frame compliance would be useful, as it would make the calibration independent of other calibrations, in particular the indenter area function. The problem with this approach is that it usually requires making measurements only after replacing the indenter with a much stiffer connection between sample stage and indenter shaft. Direct compliance measurement therefore requires performance of a full indirect validation, once the indenter has been replaced, to ensure poor mounting of the indenter or a defect in the indenter itself have not added additional frame compliance. Indirect methods have suffered (until recently) from the lack of certified reference materials. To avoid the use of a reference value for the materials used, a series of indentations at different forces (depths) are performed and a plot of compliance vs. $1/\sqrt{Ac}$ is used to obtain the intercept, which is the frame compliance. This has the problem of high sensitivity of the extrapolation to the standard error in the data. If, however, a CRM is used together with a direct measurement of the indenter area function, then the contact stiffness and therefore frame compliance can be calculated for each indentation. This method can significantly reduce the uncertainty in the value of frame compliance as the standard error of a mean is a factor of $\sqrt{N}$ less than the standard deviation of the N data points.
Measurement uncertainty
Every measurement has an associated uncertainty, which is an expression of how accurate the measurement is. Whether it is how finely divided a measurement scale is, or how reproducible a measurement is, there is always a finite uncertainty associated. Most measurements have a total or combined uncertainty that is made up of a number of uncertainties associated with the individual measurements and calibrations that, together, made up the measurement concerned.

In a high precision or high accuracy measurement such as instrumented indentation, an in-depth uncertainty analysis can be complicated. It is, however, to the user’s great advantage to do this. Most uncertainty “budgets” have one or two dominant uncertainties. These are the individual components in the total combined uncertainty that are so large that they render the total uncertainty somewhat insensitive to increases or variations in the smaller uncertainties. The benefit is therefore in identifying these dominant uncertainties so that they can be reduced (the lowest cost to benefit activity for improving measurement quality) or so that unnecessary effort spent controlling the insignificant uncertainties can be abandoned (same quality for less effort). Since uncertainty budgets are in fact an algorithm for calculating the uncertainty for any measurement, this also means that ‘what-if’ scenarios can be run. This is useful if an environmental change occurs or a smaller uncertainty is required for a special measurement, as the easiest route to the necessary uncertainty can be identified. Conversely, should the quality control of a lower-grade batch of components become necessary, the measurement uncertainty limits can be relaxed appropriately.

The exact algorithm for the calculation of uncertainty in instrumented indentation is not universal. It depends on instrument type, the calibration methods used, the environmental controls in place, the indentation cycle used, the material property being measured and the type of material being tested. The simplest uncertainty budget must however consider the following key uncertainties:

- Force calibration and noise level
- Displacement calibration and noise level
- Time calibration (cycle repeatability)
- Zero point calculation
- Area function
- Frame compliance
- Thermal drift correction
- Repeatability of the measurement
- Uncertainty in any fit to the data and associated slope
- Uncertainty of any extrapolation

A number of these uncertainties can be taken directly from the certificates issued after traceable calibration. Others are best estimated from prior measurements using CRMs to obtain a generically applicable value that minimises any double counting of material induced variability in the estimate. The combined uncertainty of the measurement must also include the repeatability of the measurement, e.g. derived from the standard deviation of repeated indentation results. This is an important contribution as a significant component of the total uncertainty directly results from the state of the actual material being tested (e.g. homogeneity and surface roughness).

There are a number of sources for assistance in the calculation of uncertainties. Accreditation services, such as the UK Accreditation Service (UKAS), often run training courses in the mechanics of combining uncertainties. There are also guides to the calculation of uncertainties such as the ISO “GUM” (Guide to uncertainty in measurement) [54] and the UKAS guide to the “GUM”, M 3003 [55]. Suppliers of CRMs should also provide assistance in uncertainty measurement and calculation.
References
References


42. Y Y Lim and M M Chaudhri. Phil. Mag., **A79** (1999) pp2979


