We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

3,900
Open access books available

116,000
International authors and editors

120M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Uncertainties and Errors in Nanoindentation

Jaroslav Menčík

1. Introduction

Nanoindentation provides information on mechanical properties by calculation from indenter load and displacement. The basic formulae for these computations are based on certain assumptions on the material response, which are valid only under some conditions. If a factor is neglected that is important at these conditions, the results can be wrong. The user of nanoindentation techniques thus should be aware of possible influences and errors, in order to prepare the tests and process the data so as to avoid or reduce them.

Various works deal with errors in nanoindentation measurements, for example (Menčík & Swain, 1995), Chapter 4 in (Fischer-Cripps, 2004), or some of the papers quoted in the following text. This chapter provides an overview of sources of uncertainties and errors in instrumented testing. The individual cases will be discussed and formulae given, showing the relationships between the errors in input quantities and in the calculated ones. This can help in the assessment whether a certain influence should be considered, and in the formulation of demands on the tests and the accuracy of input data.

This chapter is divided according to the following sources of uncertainties and errors:
- Properties of the tested material
- Models used for data evaluation
- Properties of the indenter
- Device properties
- Specimen properties
- Initial depth of penetration
- Temperature changes, drift
- Contact profile, pile-up
- Indentation size effect
- Surface forces and adhesion
- Scatter of measured values
2. Properties of the tested material

Errors can arise if the measurement is not done under appropriate conditions. The tests can be arranged in a more effective way if more is known about the tested material in advance: whether it is elastic-plastic or brittle, and if the deformations depend on time. It is also important to know something about its microstructure, basic components, their size and the influence of possible anisotropy. Until now, most of the papers on nanoindentation were devoted to isotropic materials, and so also is this chapter. The study of anisotropic properties is beyond its scope.

The first information on mechanical properties is obtained from a simple load-unload test, which provides a general idea and basic characteristics such as hardness and elastic modulus and also informs about irreversible processes (Fig. 1). All this can serve as the starting point for more detailed study. A pointed indenter is often used, but spherical indenter may be advantageous, as it enables continuous change of the mean contact pressure and characteristic strain. It can be used for the study in elastic regime, with reversible deformations, but also in the elastic-plastic regime, and enables the construction of stress-strain curves and determination of the yield strength and strain-hardening parameters of a ductile material, important also for the assessment of pile-up effects (see Section 9). In brittle materials, spherical indenter enables the determination of the limit stresses for crack formation or other irreversible phenomena, such as stress-induced phase transformations, densification, or generation of dislocations.

![Figure 1. Typical load-depth curves of indentation into elastic-plastic materials – a schematic (after Menčík, 1996). a – highly elastic material, b – hard elastoplastic material, c – soft elastoplastic material, d – formation of cracks during loading, e – phase transformation during unloading, f – delamination of the indented coating from substrate](image-url)

For components with coatings, also the coating strength and adhesion to the substrate are important. If delamination or another kind of coating damage occurs, the measured values of hardness or elastic modulus will be wrong. Nevertheless, nanoindentation can be used also for characterization of these phenomena.

Many materials exhibit time-dependent response, for example polymers or biological tissues, but also metals under certain conditions. Examples are shown in Figs. 2 and 3: the dotted curve in Fig. 2 illustrates distortion of unloading curve due to delayed deforming. Fig. 3 shows gradual indenter penetration into a polymer under constant
load. Instrumented indentation enables the determination of parameters for characterization of viscoelastic response. A very important question is whether the deformations will be only delayed-reversible, or if also irreversible creep will be present. A role can be played by the magnitude of acting stresses. Low stress sometimes causes only reversible delayed elastic deforming, but if it exceeds some limit value, also irreversible viscous flow can occur, sometimes together with plastic (time-independent) deformations. One should be aware that pointed indenters always cause high stresses at the tip. Thus, if a low-stress response is to be studied, a spherical indenter must be used to ensure a sufficiently low stress level.

After the first information on the time-dependent response has been obtained, one can better specify the conditions of further exploration: the time course of the tests: velocity of the initial load increase, the dwell under constant load and the velocity of unloading, or the frequency if harmonic loading is used.

Properties of some materials depend on the environment (temperature, humidity). The measurements should be done in an environment corresponding to the assumed use of the material. This is especially important for biological materials, such as tissues or bones, where the properties of a dry specimen can significantly differ from those existing in human body. The tests, therefore, are often done on samples immersed in a suitable solution. Some materials also undergo slow changes during long time. For example, the properties of a child bone differ significantly from those of an old person. Gradual changes occur also in wood, concrete, plastics and some other materials.

It should be known in advance what structural details of the material will be studied, as some phenomena are evident only at certain loads and imprint size. Information on the material microstructure and the size of its building units is very important. Most materials are inhomogeneous on microscale. If indentation is done under so high load that many structural units are contained in the indentation-influenced volume, the indenter response reflects the average property of this volume. If properties of the individual phases are to be studied, appropriately lower load must be used. For more, see Section 12.

Under low loads (mN), with imprints smaller than several µm, hardness of some materials increases with decreasing load. This, so-called indentation size effect is treated in Section 10. Under even lower loads, tens of µN and less, another material-related problem can appear. Surface forces and adhesion between the indenter and sample are no more negligible, and can dramatically change the load-depth curves, especially for soft materials such as gels or biological tissues. This must be taken into account in the preparation of measurement and their evaluation, otherwise significant errors can appear; cf. Section 11.

The test conditions must also respect if the measured property depends on temperature. In any case, the temperatures of the specimen, device and environment should be constant during the test, otherwise additional errors appear due to thermal dilatations; see Section 8.
3. Models used for data evaluation

The quality and accuracy of results depend substantially on the quality and adequacy of the models used for data evaluation. A well known example is the increase of accuracy in the determination of hardness and elastic modulus, which was attained by calculating the contact stiffness via a power-law approximation of unloading curve (Oliver & Pharr, 1992) instead of the originally suggested linear approximation. The models are often only empirical and a matter of our choice. Examples are the function for approximation of unloading curve, of a creep curve for a viscoelastic material, or of the response during indentation into a coated sample. Sometimes, the same response can be described reasonably well by several models. When planning a test or evaluating the data, one should be aware that the models for determination of material properties are valid only under some conditions and within certain range of parameters (load magnitude, test duration, etc.). The purpose of the test should therefore always be known in advance, as it is important for the choice of a model with suitable accuracy.

In this chapter, the following sources of model-related errors will be discussed: 1. Calculation of contact depth, 2. Time-dependent response, 3. Indentation into coated specimens, 4. Others.

3.1. Models for contact depth

The most often determined mechanical properties are hardness $H$ and elastic modulus $E$, calculated in depth-sensing indentation tests as:

$$H = \frac{F}{A} \times \frac{E_r}{E} \left( 2 \frac{A^{1/2}}{h} \right).$$

$F$ is the indenter load, $E_r$ is the reduced (= sample + indenter) modulus, defined by Eq. (18), $S$ is the contact stiffness, determined as the slope of the unloading curve at the beginning of unloading ($S = dF/dh$), and $A$ is the contact area, calculated from the contact depth $h_c$ (Fig. 2), which is related to the total indenter displacement $h$ as

$$h_c = \beta h - \epsilon F / S;$$

$\beta$ and $\epsilon$ are constants. Usually, $\epsilon = 0.75$ is used for a spherical or pointed indenter, and $\beta = 1$ for spherical and 1.05 for Berkovich indenter. These values are suitable for highly elastic materials, such as hard metals, glasses and ceramics. In substances with easy plastic flow, the material sometimes piles-up around the indenter, so that the above values of $\beta$ and $\epsilon$ can lead to underestimated values of $h_c$. For such cases, Bec et al. (1996) and Fujisawa & Swain (2006) recommend $\beta = 1.2$ and $\epsilon = 1$. (For pile-up, see Section 9.) $S$ is mostly calculated from a power-law approximation of the unloading curve proposed by Oliver & Pharr (1992), $F = A(h - h_f)^m$, with empirical constants $m$ and $h_f$.

Equations (1) and (2) are valid for elastic as well as elastic-plastic indentation and any indenter shape. For elastic contact with a spherical indenter (of radius $R$) also Hertz’ formulae (Johnson, 1985) may be used:
Uncertainties and Errors in Nanoindentation

\[ h = \frac{9F^2}{16RE_c^2}, \quad h_c = h/2. \]  \hspace{1cm} (3)

\textbf{Figure 2.} Load-depth curve of an indentation test. \( F \) – indenter load, \( h \) – displacement (penetration), \( h_c \) – contact depth. Dotted curve – distortion due to delayed viscoelastic response during unloading

\section*{3.2. Models for time dependent response}

The load response of viscoelastic materials depends on the load and its duration. The indenter often continues penetrating into the specimen even under constant load (Fig. 3). As a consequence, the apparent elastic modulus and hardness decrease with time. This should be considered in the characterization of mechanical properties. Nevertheless, they are often determined also from common load-unload tests. Here, one more error can occur. Due to delayed response, the unloading part of the \( F-h \) curve is often distorted – more convex than for elastic materials (dotted curve in Fig. 2). As a consequence, the apparent contact stiffness \( S \), determined from the slope of the upper part of unloading curve, is higher than the actual value. This could lead to errors in the determination of contact depth and all derived quantities, such as \( H \) and \( E \).

The influence of viscoelastic after-effects can be reduced in various ways. Often, a dwell is inserted between the loading and unloading. According to Chudoba & Richter (2001), the effect of delayed deforming on the unloading curve may be neglected if the creep velocity has decreased so that the depth of penetration grows less than 1\% per minute. The influence of creep at the end of dwell can also be mitigated by using the effective contact stiffness \( S \), as proposed by Feng & Ngan (2002):

\[ \frac{1}{S} = \frac{1}{S_{\text{app}}} + \frac{\dot{h}_d}{|F_u|}, \]  \hspace{1cm} (4)

\( S_{\text{app}} \) is the apparent stiffness, obtained from the unloading curve by the standard procedure; \( \dot{h}_d \) is the indenter velocity at the end of dwell, and \( F_u \) is the load decrease rate at the beginning of unloading. A disadvantage remains that the indenter depth at the beginning of
unloading after the dwell is larger than at the end of loading. This results in larger contact area and lower apparent hardness. Therefore, it is recommended to use relatively fast loading followed immediately by fast unloading (Cheng & Cheng, 2005). Nevertheless, it is generally insufficient to characterize materials, which flow under load, only by a single value of hardness or elastic modulus. Better are universal rheological models, created from springs and dashpots, which may also be used in the finite element analysis.

The parameters in these models can be obtained by fitting the time course of indenter penetration (Fig. 3) by a suitable creep function, generally in the form

\[ h^m(t) = K \psi(F, J, t) ; \]  

\( m \) and \( K \) are constants related to the indenter shape. For pointed indenters, \( m = 2 \) and \( K = \pi/(2 \tan \alpha) \); \( \alpha \) is the semiangle of the equivalent cone (70.3° for Berkovich and Vickers). For a spherical indenter, \( m = 3/2 \) and \( K = 3/(4\sqrt{R}) \); \( R \) is the tip radius. \( \psi(F, J, t) \) is a response function depending on the load, material and time (Menčík et al., 2009, 2010). \( J \) is the creep compliance function, which expresses the material response to the unit step load. For monotonic load and instant deformation accompanied by reversible delayed deforming, 

\[ J(t) = \sum_{j=1}^{n} C_j \left[ 1 - \exp\left(-t/\tau_j\right) \right] \]  

with constants \( C_0, C_j \) and \( \tau \). The instantaneous compliance \( C_0 \) corresponds to a spring, and the terms with \( C_j \) correspond to Kelvin-Voigt bodies (a spring in parallel with a dashpot). For harmonic loading, trigonometrical series can be used (Menčík et al., 2004, 2011a).

The parameters in a model for viscoelastic response are usually determined from the time course of indenter penetration under constant load (Oyen, 2006, Menčík et al., 2009, 2010). However, the displacement was also influenced by the initial period of load-increase. This can be taken into account by correction factors. If the rate of load increase is constant (\( dF/dt = \text{const} \)), the creep compliance function (6) can be modified as follows (Oyen, 2006):

\[ J(t) = C_0 + \sum_{j=1}^{n} C_j \left[ 1 - \rho \exp\left(-t/\tau_j\right) \right] , \]
with ramp correction factors

\[ r_j = \frac{t_j}{t_R} \left( \exp \left( \frac{t_R}{\tau_j} \right) - 1 \right); \]  

(8)

\( \tau_j \) is so-called retardation time, \( t_R \) is the duration of load increase to the nominal value \( F \). This correction is more reliable for shorter times \( t_R \), where \( \rho_j \) are not very different from 1.

The practical procedure is as follows. The indenter is loaded to the nominal load \( F \), which is then held constant. The time course of its penetration under constant load is fitted by a function of type (Menčík, 2011c, Menčík et al., 2009, 2010)

\[ h^m(t) = F K_0 - \Sigma C_j \exp(-t/\tau_j), \]  

(9)

following from Eqs. (5) and (7). The constants \( B_0, D, \) and \( \tau \) are obtained using the least squares method. Then, the correction factors \( \rho_j \) are calculated via Eq. (8). The constants \( C_j \) in the creep function (7) are obtained as \( C_j = D_j/\rho_j \). Finally, \( C_0 \) is determined as \( C_0 = B_0 - \Sigma C_j \). If no permanent deformations occurred, \( C_0 \) is related to the elastic modulus as \( C_0 = 1/E_r \).

When choosing a model, one should know whether the time-dependent deformations are reversible or irreversible. Kelvin-Voigt elements are suitable for reversible deformations. If irreversible deformations appear, a plastic element in series with other elements must be added for time-independent deformations, or a dashpot for viscous flow; for details see Menčík et al. (2009, 2010) or Menčík (2011c). The information on the reversibility can be obtained from the unloading response – whether the deformations disappear or not. The indentation experiments should be arranged with respect to the assumed use of the material. If the stresses in operation will be low and the deformations reversible, also the test stress should be low, and a spherical indenter must be used.

In the choice of a model for time-dependent response, the following experience can be used (Menčík, 2010, 2011b, 2011c). Delayed deforming is modeled well by Kelvin-Voigt bodies. One K-V body is effective only within a limited time interval, roughly for \( 0.03 \tau < t < 3 \tau \), where \( \tau \) is the retardation time – a material parameter of this body. For example, a K-V body with \( \tau = 5 \) s is effective from about 1.5 s to 15 s. If the response during a long time should be modeled, more K-V bodies must be arranged in the model, with retardation times \( \tau \) scaled roughly by a factor of ten. It is thus useful to know approximately the duration of the process to be modeled, and the demanded accuracy. Models should not be complicated more than necessary. For long lasting processes, higher accuracy is usually needed for longer times, and lower for the initial period of load increase. The response can often be approximated similarly well by several models with different number of elements. Also the characteristic times \( \tau \) in more complex models may vary in some range, with only small influence on the quality of the fit. Only sometimes it is possible to relate a particular K-V body (or the pertinent retardation time \( \tau \)) with some process in the material. Experience with continuous stiffness measurement can be found in Pharr et al. (2009). Errors caused by the difference between harmonic and quasistatic stiffnesses in the evaluation of monotonic load-unload tests (in CSM mode) were discussed in (Menčík et al., 2005).
 Generally, it is not difficult to find a good approximation for the response of a viscoelastic material within certain time interval. However, caution is necessary if the model should be used for predictions of behavior in longer times, especially if the indenter movement has not ended during the test. Example of differences in extrapolation for various models are shown in Fig. 3.

3.3. Models for indentation response of coated components

The response is influenced by the properties of the coating and substrate. The influence of substrate is negligible only if hardness of a relatively soft coating is to be determined and the indentation depth is smaller than 10% of the coating thickness. Otherwise it must be considered, especially if elastic modulus is measured, because the substrate influences the elastic response beginning from the minimum load. The determination of the coating property $X_c$ (Young modulus, hardness, etc.) consists of measuring this property for several depths $h$, fitting the obtained $X(h)$ values by a suitable function (Fig. 4), and extrapolating to zero depth; $X_c = X(0)$ is found as one constant in the regression function. If CSM or DMA test mode (small harmonic loading superimposed on monotonously growing load) is used, all $X(h)$ data can be obtained in one test. General form of the fitting function is

$$X(\xi) = X_c \Phi(\xi) + X_s [1 - \Phi(\xi)] ;$$  

(10)

![Figure 4. Apparent elastic modulus for various indenter penetration into a coating (a schematic).](image)

$E_s$ – substrate modulus, $a$ – contact radius, $t$ – coating thickness

$X_c$ and $X_s$ is the pertinent property of the coating and the substrate, and $\Phi(\xi)$ is a non-dimensional weight function of parameter $\xi$, defined as the ratio of the penetration depth to the coating thickness, $\xi = h/t$, or the ratio of contact radius to the coating thickness, $\xi = a/t$. Function $\Phi(\xi)$ decreases from 1 for $\xi = 0$ (or $h = 0$) to zero for $h >> t$. The advantage of the dimensionless form of Eq. (10) is its universality: the data obtained on specimens of various thicknesses can be combined together, so that the fitting is based on more measured values.

Various functions for obtaining the coating elastic modulus are compared in Menčík et al. (1997). As most universal appeared the function proposed by Gao et al. (1992). Later, several other models were proposed. Gao’s function is based on the analytical solution of penetration of a punch into a semiinfinite elastic body with a layer on its surface. Some
models, e.g. by Doerner & Nix (1986), Bhattacharya & Nix (1988), Rar, Song & Pharr (2001), Saha & Nix (2002) or Hay & Crawford (2011) correspond to the series arrangement of springs, representing the coating and substrate. All these models have theoretical substantiation and can lead to relatively reliable results. But even simple empirical models, based on exponential function $\Phi = \exp (-k\xi)$, can yield good results if the measurements are done for very small depths of penetration compared to the coating thickness. Extrapolation of the measured $X(\xi)$ values to zero depth ($\xi = 0$) should be as small as possible.

The indentation response varies between that for the coating and the substrate. The error in the determination of coating modulus will be smaller if it is close to that of the substrate. Sometimes it is possible to find a suitable substrate. It may also be useful to make the measurements for the coating deposited on two substrates: one stiffer than the coating, and the other more compliant (Menčík et al., 1997). The coating property will lie between the values obtained by extrapolating the $X(\xi)$ data for both samples to $\xi = 0$.

When fitting the measured $X(\xi)$ data by the $\Phi(\xi)$ function, one usually inserts the substrate property, $X_s$, as a known constant into Eq. (10). However, also this value is associated with some uncertainty and can contribute to the error in the determination of coating modulus. The pertinent substrate property thus should be known as accurately as possible. Some role is also played by the Poisson’s ratio of the coating and substrate.

The principle of determination of coating hardness is similar. This topic is treated in literature, e.g. Jönsson & Hogmark (1984), Burnett & Rickerby, 1987, Korsunsky et al., 1998). Four combinations of properties can appear: hard coating on a hard substrate, hard coating on a compliant (and ductile) substrate, soft coating on a compliant substrate, and a soft coating on a hard substrate. Approaches and functions $\Phi$ for various cases are also described in Menčík et al. (1997) and Menčík (1996). As it follows from Eq. (10), the accuracy of the determination of coating hardness depends also on the accuracy of determination the substrate hardness and the coating thickness $t$, appearing in the terms $h/t$ or $a/t$. Sometimes, errors arise if the original (nominal) coating thickness is used in the evaluation of the $X(\xi)$ data, especially for soft coatings. High pressure under indenter can make the coating thinner. A correction for the thickness reduction can bring significant increase in accuracy of hardness as well as elastic modulus (Menčík et al., 1997).

Other sources of errors exist, as well. If a relatively soft coating was deposited on a hard and stiff substrate, the coating material can pile-up around the indenter. The contact area and stiffness are thus larger and do not correspond to the above models (cf. Section 9). Another problem occurs if the coating debonds from the substrate or if cracks arise in the coating or substrate during the indentation. In such cases the contact compliance becomes higher. These events are indicated by changes in the load-unload $F - h$ curves (Fig. 1d,f), whose shape is thus an important source of information. Also, indentation tests leading to coating delamination or nucleation of cracks can be used for the determination of material fracture toughness, interface strength or some adhesion parameters. More about delamination by indentation can be found in Rother & Dietrich (1984), Marshall & Evans (1984), Matthewson (1986) or Li & Siegmund (2004). Determination of fracture toughness
by indentation is treated in Niihara (1983), Niihara et al. (1982), Lawn et al. (1980) and Zhang et al. (2010).

3.4. Other cases

In the study of elastic-plastic behavior, the model for strain-hardening can be important; see, e.g., Field & Swain (1993), Herbert et al. (2001) and Cao & Lu (2004). If a material, consisting of several phases, is studied, the results depend on the indenter load and the number of phases considered in the model for statistical analysis; cf. Section 12. At very low loads, also surface (adhesive) forces can play a role, as explained in Section 11.

4. Indenter properties

Models for data evaluation assume certain shape and mechanical properties of the indenter. However, the actual parameters can deviate from the nominal values, used in the formulae.

4.1. Shape

The formulae for determination of mechanical properties usually work with contact area $A$. The indenters are mostly pointed or spherical. For an ideal pointed (conical) indenter,

$$A_{id} = kh_c^2; \quad (11)$$

$k = \frac{\pi}{\tan(\alpha)^2}; \quad \alpha$ is the semiapical angle. For pyramidal indenters, an equivalent cone is worked with. The angle of the equivalent cone for Berkovich indenter is $\alpha \approx 24,5^\circ$.

The shape of a real indenter often deviates from the ideal one. The tip of a pointed indenter is never ideally sharp, but somewhat blunt, with the radius several tens of nm or more (Fig. 5). This deviation is important especially for very small depths of penetration. The simplest correction uses the addition of a small constant $\delta$ to the measured depth, $h_{corr} = h + \delta$. The constant $\delta$ can be obtained by fitting the series of $F(h)$ points by the expression $F = c (h + \delta)^2$. However, also the actual angle $\alpha$ can slightly differ from the nominal value, and even vary with depth. More reliable is indenter calibration. It provides so-called indenter area function, which expresses the projected contact area as a function of contact depth, $A = A(h_c)$. Especially for very small depths of penetration (less than several tens of nm), the difference between $A_{true}$ and $A_{id}$, and the corresponding errors in the determination of $A$ or $E$ can amount several tens of percent.

![Figure 5. Shape of a real pointed indenter (a schematic). Contact profile without pileup (solid line) and with pileup (dashed line). $h_{tr}$ – transition between the spherical tip and conical part](image-url)
One method of indenter calibration determines the contact area via imaging the imprints by
scanning- or atomic force microscope. More often, the indenter area function is calculated
from the contact stiffness, obtained by indenting a sample of known elastic modulus:

\[ A(h_c) = \pi \left[ S(h_c) \right]^2 / \left( 4E^2 \right). \]  \hspace{1cm} (12)

The measurements must be done for a sufficient number of depths. The stiffness \( S \) can be
obtained either in quasistatic manner from unloading curves, or in dynamic way, using
small oscillations superimposed on the basic load (CSM or DMA mode). In this case, the
area function \( A(h_c) \) can be obtained in one test. The measured \( A(h_c) \) values are then
approximated by a suitable analytical expression, e.g. a polynomial (Oliver & Pharr, 1992):

\[ A(h_c) = C_0 h_c^2 + C_1 h_c^{1/2} + C_2 h_c^{1/4} + C_3 h_c^{1/8} + \ldots \]  \hspace{1cm} (13)

Indenter calibration is often done on fused silica, as its surface around the indenter deforms
similarly to theoretical elastic models, without pile-up.

The contact area under a spherical indenter of radius \( R \) is

\[ A_{id} = \pi a^2 = \pi \left( 2Rh_c - h_c^2 \right) = 2\pi Rh_c. \]  \hspace{1cm} (14)

If the deformations are elastic, Hertz’s formulae may be used as well. However, spherical
indenters of small radius are never perfectly spherical, a role being sometimes played by
anisotropy of properties (Menčík, 2011d). The indenters must thus be calibrated. Again, the
indenter area function \( A(h_c) \) can be determined via the unloading contact stiffness and
formula (12). Sometimes, Hertzian formulae are used, which work with indenter radius \( R \).
In such case, the effective indenter radius, corresponding to certain contact depth, \( R_{eff}(h_c) \),
should be used. \( R_{eff} \) can be calculated from the actual contact area as

\[ R_{eff}(h_c) = \frac{A(h_c)}{2\pi h_c}. \]  \hspace{1cm} (15)

or from the elastic displacement of the indenter:

\[ R_{eff}(h) = \left( 9 / 16 \right) F^2 E^2 \pi^{-2} h^{-3}. \]  \hspace{1cm} (16)

This \( R_{eff} \) represents the average radius for the indenter penetration \( h \). If the indenter shape is
not exactly spherical, various \( R_{eff} \) values are obtained for various depths of penetration.
In this way it is possible to construct the calibration curve \( R_{eff}(h) \). Figure 6 shows calibration
curve for a spherical indenter of nominal radius 200\( \mu \)m. The curve was obtained from the
indenter penetration into fused silica, whose reduced modulus was assumed \( E_r = 70 \) GPa.
The maximum load was \( F = 50 \) mN and the corresponding depth \( h = 130 \) nm. One can see that the
effective radius for this depth \( (R_{eff} = 134 \mu m) \) differs significantly from the nominal value 200
\( \mu m \). It is obvious that this indenter cannot be used for the evaluation of data for penetration
less than 10 nm, and the results for depths up to 20 nm are not very reliable. Other examples
of indenter calibration can be found in Oliver & Pharr (2004) or Nohava & Menčík (2012).
With pointed or spherical indenters, one should be aware that both shapes are, in principle, sphericonical. The tip is approximately spherical, but later it turns gradually into a cone or a pyramid. Thus, if one wants to use the formulae for spherical indenters, the contact depth should not be larger than the transition point between the sphere and cone ($h_{tr}$ in Fig. 5). On the other hand, the formulae for a conical or pyramidal indenter are sufficiently accurate only if the round part at the tip is negligible compared to the total depth of penetration. In other cases, indenter area function or some correction must be used. The distance of the transition point from the tip is 

$$h_{tr} = R(1 - \cos \alpha). \quad (17)$$

Sometimes, a cylindrical indenter („rigid punch“) is used, for example for very compliant materials. Its advantage is constant contact area, independent of the penetration. However, the edge between the planar face and cylindrical part is not ideally sharp, and at extremely small depths of penetration the contact area slightly changes with depth. Moreover, high stress concentration at the edges can cause irreversible changes in the specimen, and thus some nonlinearity in the response.

![Figure 6. Effective radius $R_{eff}$ of the spherical indenter as a function of penetration $h$. Nominal radius $R = 200$ µm. (After Menčík & Nohava, 2011)](image)

The indenter roughness influences the measurement similarly as the specimen roughness (Section 6). Despite of high hardness, indenters gradually get worn, especially if very hard materials are repeatedly tested. Therefore, it is necessary to check their actual shape from time to time, and recalibrate it, if necessary.

### 4.2. Mechanical properties

Nanoindentation is very suitable for the determination of elastic modulus in small volumes. However, it is so-called reduced modulus $E_r$, which is obtained from Eq. (1). This modulus is influenced by the Young modulus of the specimen ($E_s$) and the indenter ($E_i$):

$$\frac{1}{E_r} = \frac{1}{E_s} + \frac{1}{E_i}; \quad (18)$$

$E^\prime$ means $E/(1 - \nu^2)$; $\nu$ is Poisson’s ratio. The specimen modulus is calculated from $E_r$ as
Its value thus depends on the composite modulus $E_r$ but also on the modulus of the indenter and Poisson’s ratio of the indenter and sample. These values, however, are often only assumed, for example $E_i = 1000$ GPa and $\nu = 0.07$ for diamond, and the actual values can be different. (According to literature, Young modulus of diamond can vary between 800 and 1100 GPa, and indenters are made also from other materials, SiC, tungsten carbide, steel, etc.). Thus, if incorrect value of $E_i$ is inserted into Eq. (19), the calculated sample modulus $E_s$ is also wrong. The pertinent error will be non-negligible especially if relatively stiff materials are tested. The relative error in the sample modulus, $E_s'/E_{s,a}'$, expressed as the function of the relative error of indenter modulus $E_{i,a}/E_i$ and the ratio of the specimen’s modulus and indenter modulus, $E_{s,a}/E_i$, is:

$$
\frac{E_s'}{E_{s,a}'} = 1 \left[ 1 + \frac{E_{s,a}}{E_{i,a}} \left( \frac{1 - \nu_s^2}{1 - \nu_i^2} \right) \right],
$$

(20)

The subscript $a$ denotes apparent value, i.e. the value obtained from the test ($E_{s,a}'$), or assumed ($E_{i,a}'$). Actually, we do not get the true specimen value from the test; we only try to obtain it from the apparent values. The situation is depicted in Fig. 7. One can see that the error in the determination of sample modulus, caused by the use of incorrect value $E_i$, is negligible for $E_s/E_i \leq 0.01$, but can amount several percent for $E_s/E_i > 0.1$, more for larger error in $E_i$ and higher modulus of the tested material.

**Figure 7.** Relative error of sample modulus ($E_s/E_{s,a}$) as a function of relative error of indenter modulus $E_{i,a}/E_i$ and the ratio of specimen and indenter modulus $E_{s,a}/E_i$. Subscripts: $s$ – sample, $i$ – indenter, $a$ – apparent

The specimen modulus depends also on Poisson’s ratio. As it follows from Eq. (19), the relative error is

$$
\frac{E_{s,a}}{E_s} = \left( 1 - \nu_s^2 \right) \left( \frac{1 - \nu_i^2}{E_i} \right).
$$

(21)
For example, if the Poisson’s ratio $\nu_s,a = 0.2$ were used for the specimen instead of the correct value 0.3, the error in the determination of Young modulus would be about 5%.

5. Device properties

The results of measurement are influenced to some extent by the compliance of the device. The force acting on the indenter causes not only its penetration into the specimen, but also the deformation of the indenter shaft, the specimen holder and other parts of the device, all denoted together as the instrument frame. The total indenter displacement, measured by a sensor, $h_m$, is thus larger than the depth of penetration, $h$, by the frame deformation $h = C_f F$, where $F$ is the load and $C_f$ is the frame compliance. The actual depth of penetration is

$$h = h_m - C_f F.$$  \hfill (22)

$C_f$ is one of the input quantities in the software for nanoindentation data processing. If an incorrect apparent value $C_{f,a}$ is used instead of the actual value $C_f$, it will cause an error in the depth of penetration $h_a$, contact stiffness $S_a$ and all quantities calculated from them. The contact stiffness $S$ is the reciprocal of the contact compliance, defined as $C = 1/S = dh/dF$. With $h$ given by Eq. (22), the apparent stiffness is

$$S_a = \left[ \frac{dh_m}{dF - C_{f,a}} \right]^{-1}.$$  \hfill (23)

If the value $C_{f,a}$ is (for example) smaller than the actual frame compliance $C_f$, the apparent depth $h_a (= h_m - C_{f,a}F)$ will be larger than the actual depth $h$, the apparent contact compliance $C_a$ will be higher and the apparent contact stiffness $S_a$ smaller. The apparent contact depth,

$$h_{c,a} = h_a - \varepsilon C_a F = h_m - \varepsilon F d h_m / dF - C_{f,a} F (1 - \varepsilon),$$  \hfill (24)

will be larger than the actual depth. As a consequence, the apparent contact area will be larger and hardness lower than their actual values. Also the apparent elastic modulus, Eq. (1), will be lower than its actual value. For example, the results for a silicate glass (Berkovich indenter, $F = 50$ mN) were: $h = 694$ nm, $h_c = 527.9$ nm, $S = 223785$ N/m, $H = 7.32$ GPa and $E = 72.94$ GPa. If, instead of the correct frame compliance $C_f$, contact stiffness $S$ and all quantities calculated from them. The contact stiffness $S$ is the reciprocal of the contact compliance, defined as $C = 1/S = dh/dF$. With $h$ given by Eq. (22), the apparent stiffness is

$$S_a = \left[ \frac{dh_m}{dF - C_{f,a}} \right]^{-1}.$$  \hfill (23)

If the value $C_{f,a}$ is (for example) smaller than the actual frame compliance $C_f$, the apparent depth $h_a (= h_m - C_{f,a}F)$ will be larger than the actual depth $h$, the apparent contact compliance $C_a$ will be higher and the apparent contact stiffness $S_a$ smaller. The apparent contact depth,

$$h_{c,a} = h_a - \varepsilon C_a F = h_m - \varepsilon F d h_m / dF - C_{f,a} F (1 - \varepsilon),$$  \hfill (24)

will be larger than the actual depth. As a consequence, the apparent contact area will be larger and hardness lower than their actual values. Also the apparent elastic modulus, Eq. (1), will be lower than its actual value. For example, the results for a silicate glass (Berkovich indenter, $F = 50$ mN) were: $h = 694$ nm, $h_c = 527.9$ nm, $S = 223785$ N/m, $H = 7.32$ GPa and $E = 72.94$ GPa. If, instead of the correct frame compliance $C_f$, contact stiffness $S$ and all quantities calculated from them. The contact stiffness $S$ is the reciprocal of the contact compliance, defined as $C = 1/S = dh/dF$. With $h$ given by Eq. (22), the apparent stiffness is

$$S_a = \left[ \frac{dh_m}{dF - C_{f,a}} \right]^{-1}.$$  \hfill (23)

If the value $C_{f,a}$ is (for example) smaller than the actual frame compliance $C_f$, the apparent depth $h_a (= h_m - C_{f,a}F)$ will be larger than the actual depth $h$, the apparent contact compliance $C_a$ will be higher and the apparent contact stiffness $S_a$ smaller. The apparent contact depth,

$$h_{c,a} = h_a - \varepsilon C_a F = h_m - \varepsilon F d h_m / dF - C_{f,a} F (1 - \varepsilon),$$  \hfill (24)

will be larger than the actual depth. As a consequence, the apparent contact area will be larger and hardness lower than their actual values. Also the apparent elastic modulus, Eq. (1), will be lower than its actual value. For example, the results for a silicate glass (Berkovich indenter, $F = 50$ mN) were: $h = 694$ nm, $h_c = 527.9$ nm, $S = 223785$ N/m, $H = 7.32$ GPa and $E = 72.94$ GPa. If, instead of the correct frame compliance $C_f$, contact stiffness $S$ and all quantities calculated from them. The contact stiffness $S$ is the reciprocal of the contact compliance, defined as $C = 1/S = dh/dF$. With $h$ given by Eq. (22), the apparent stiffness is

$$S_a = \left[ \frac{dh_m}{dF - C_{f,a}} \right]^{-1}.$$  \hfill (23)

If the value $C_{f,a}$ is (for example) smaller than the actual frame compliance $C_f$, the apparent depth $h_a (= h_m - C_{f,a}F)$ will be larger than the actual depth $h$, the apparent contact compliance $C_a$ will be higher and the apparent contact stiffness $S_a$ smaller. The apparent contact depth,

$$h_{c,a} = h_a - \varepsilon C_a F = h_m - \varepsilon F d h_m / dF - C_{f,a} F (1 - \varepsilon),$$  \hfill (24)

will be larger than the actual depth. As a consequence, the apparent contact area will be larger and hardness lower than their actual values. Also the apparent elastic modulus, Eq. (1), will be lower than its actual value. For example, the results for a silicate glass (Berkovich indenter, $F = 50$ mN) were: $h = 694$ nm, $h_c = 527.9$ nm, $S = 223785$ N/m, $H = 7.32$ GPa and $E = 72.94$ GPa. If, instead of the correct frame compliance $C_f$, contact stiffness $S$ and all quantities calculated from them. The contact stiffness $S$ is the reciprocal of the contact compliance, defined as $C = 1/S = dh/dF$. With $h$ given by Eq. (22), the apparent stiffness is

$$S_a = \left[ \frac{dh_m}{dF - C_{f,a}} \right]^{-1}.$$  \hfill (23)

If the value $C_{f,a}$ is (for example) smaller than the actual frame compliance $C_f$, the apparent depth $h_a (= h_m - C_{f,a}F)$ will be larger than the actual depth $h$, the apparent contact compliance $C_a$ will be higher and the apparent contact stiffness $S_a$ smaller. The apparent contact depth,

$$h_{c,a} = h_a - \varepsilon C_a F = h_m - \varepsilon F d h_m / dF - C_{f,a} F (1 - \varepsilon),$$  \hfill (24)

will be larger than the actual depth. As a consequence, the apparent contact area will be larger and hardness lower than their actual values. Also the apparent elastic modulus, Eq. (1), will be lower than its actual value. For example, the results for a silicate glass (Berkovich indenter, $F = 50$ mN) were: $h = 694$ nm, $h_c = 527.9$ nm, $S = 223785$ N/m, $H = 7.32$ GPa and $E = 72.94$ GPa. If, instead of the correct frame compliance $C_f$, contact stiffness $S$ and all quantities calculated from them. The contact stiffness $S$ is the reciprocal of the contact compliance, defined as $C = 1/S = dh/dF$. With $h$ given by Eq. (22), the apparent stiffness is

$$S_a = \left[ \frac{dh_m}{dF - C_{f,a}} \right]^{-1}.$$  \hfill (23)

If the value $C_{f,a}$ is (for example) smaller than the actual frame compliance $C_f$, the apparent depth $h_a (= h_m - C_{f,a}F)$ will be larger than the actual depth $h$, the apparent contact compliance $C_a$ will be higher and the apparent contact stiffness $S_a$ smaller. The apparent contact depth,

$$h_{c,a} = h_a - \varepsilon C_a F = h_m - \varepsilon F d h_m / dF - C_{f,a} F (1 - \varepsilon),$$  \hfill (24)

will be larger than the actual depth. As a consequence, the apparent contact area will be larger and hardness lower than their actual values. Also the apparent elastic modulus, Eq. (1), will be lower than its actual value. For example, the results for a silicate glass (Berkovich indenter, $F = 50$ mN) were: $h = 694$ nm, $h_c = 527.9$ nm, $S = 223785$ N/m, $H = 7.32$ GPa and $E = 72.94$ GPa. If, instead of the correct frame compliance $C_f$, contact stiffness $S$ and all quantities calculated from them. The contact stiffness $S$ is the reciprocal of the contact compliance, defined as $C = 1/S = dh/dF$. With $h$ given by Eq. (22), the apparent stiffness is

$$S_a = \left[ \frac{dh_m}{dF - C_{f,a}} \right]^{-1}.$$  \hfill (23)

If the value $C_{f,a}$ is (for example) smaller than the actual frame compliance $C_f$, the apparent depth $h_a (= h_m - C_{f,a}F)$ will be larger than the actual depth $h$, the apparent contact compliance $C_a$ will be higher and the apparent contact stiffness $S_a$ smaller. The apparent contact depth,
The compliance $C_t$ can be found by calibration. Several methods exist. A simple procedure is based on the general relationship between the load and depth. For a spherical indenter:

$$h_m = C_s F^{2/3} + C_f F;$$

(25)

$c_s$ characterizes the compliance of contact with a spherical indenter. (For purely elastic deformations, $C_s = [(9/16)R/E]^3$. ) Fitting the series of $h_m(F)$ data by Eq. (25) will yield the specimen-related compliance $C_t$ and the frame compliance $C_f$. As Eq. (25) is nonlinear, a suitable fitting tool must be used (e.g. the solver in Excel is possible), with the optimization criterion “minimum of the sum of squared differences between the measured $h_m(F)$ and calculated $h_{calc}(F)$ values”. As the instrument compliance is usually very low, it is recommended to make the calibration measurements up to relatively high loads. On the other hand, the deformations should be only elastic, so that the calibration block must be stiff and hard, and the indenter radius $R$ must be at least several hundreds of μm. As Eq. (25) is based on Hertz formula for a spherical contact, the tip radius $R$ should be constant within the used range of penetration depths. This is better achievable with larger radius.

Another method uses the fact that the derivative of Eq. (25) with respect to $F$,

$$\frac{dh_m}{dF} = (2/3)C_s F^{-1/3} + C_{t(r)}$$

(26)

represents a straight line in coordinates $dh_m/dF$ vs. $F^{-1/3}$. The frame compliance $C_f$ can be obtained as its intercept for $F^{-1/3} \to 0$. A disadvantage is the necessity of calculating the derivatives $dh_m/dF$. Equation (26) can also be expressed as a function of depth. More about calibration can be found in Fischer-Cripps (2004) or in manuals for nanoindenters.

The results of measurement can also be influenced by the internal noise in the measuring system, by vibrations, and by the way the indenter movement is controlled. In tests done under harmonic load, also the dynamic properties of the device, its resonant frequency and damping ability can play a role. For details, the reader is referred to (e.g.) TestWorks (2002).

6. Specimen properties

The results of nanoindentation measurements are also influenced by the surface roughness and by the changes in the surface layer, caused by the specimen preparation. Reliable determination of the properties of a coating needs also a good knowledge of the substrate properties and coating thickness.

6.1. Surface roughness

No surface is ideally smooth. Even the best polished specimens have surface undulations of height from several nanometers to several tens of nanometers, and with the distances between individual hills or crests tens of nanometers and more. As a consequence, the indenter comes first into contact with the highest asperities (Fig. 8). During the load increase, the tips of the highest asperities become compressed and the indenter gradually comes into contact with lower asperities, etc. The specimen surface conforms more and more with the indenter and the
response gradually approaches to that for the perfect contact. The overall response resembles that for a specimen with a thin layer of more compliant material on its surface.

A role is played not only by the shape of the asperities and by the distribution of their heights, but also by their density, by the radius and shape of the indenter tip and also by the elastic-plastic properties of the specimen. The situation resembles two nonlinear springs in series, one corresponding to the contact macrogeometry and the other to the microscopic asperities. At a first approximation, elastic contact and Hertz theory may be assumed. The influence of surface roughness can be assessed by the nondimensional roughness parameter, defined by Johnson (1976, 1985) as

$$\rho = \sigma_s R / a^2;$$  (27)

$\sigma_s$ is the standard deviation of the height of surface asperities, $R$ is the indenter radius, and $a$ is the contact radius. As $a^2/R \approx 2 h_c$, the parameter $\rho$ is closely related with the ratio of the standard deviation of asperities height to the depth of penetration $h_c$. According to Johnson (1976), the influence of surface roughness on the elastic contact stiffness is small for $\rho < 0.05$. With ductile materials, also plastic flow must be considered.

6.2. Change of surface properties

The measured values depend also on the changes, undergone by the specimen surface during its preparation. In specimens of ductile materials, the properties of surface layer can be changed by polishing or another method of processing. Due to plastic flow, residual stresses can arise in this layer, and some materials also strain-harden. If brittle materials are prepared for nanoindentation, a very thin surface layer can be damaged by grinding and polishing. Hard abrasive grains cause irreversible deformations on the surface and below it and can nucleate submicroscopic cracks here, which reduce the contact stiffness.

Sometimes, the initial part of the load-depth curve looks like that in Fig. 9. The reason is a very thin hard oxidic layer, formed on the specimen surface during its preparation. It has properties different from the genuine specimen properties, and complicates also the identification of the first contact between the indenter and specimen.

7. Initial depth of penetration

One source of uncertainty in nanoindentation tests is the indenter penetration into the sample at the minimum load. The depth measurement can start only after the first contact
has been made. This is at the minimum measurable load \( F_i \), but the corresponding depth \( h_i \) (Fig. 9) is unknown yet. Nevertheless, it should be added to the measured depth \( h_m \), so that the actual depth is

\[ h = h_m + h_i. \]  

(28)

The approximate magnitude of the initial penetration, \( h_i \), can be obtained by fitting the initial part of the \( F-h \) curve and extrapolating it to zero load. Every indenter has a less or more rounded tip, and a reasonable approximation is

\[ h = k F^n, \]  

(29)

where \( k \) and \( n \) are constants. A suitable shape of the fitting curve is

\[ h_t = k F^n - h_i \quad (= h_{m,\text{calc}}). \]  

(30)

Only the lower part (about 10 – 20 \%) of the \( h-F \) curve should be used for the fitting.

---

**Figure 9.** Determination of initial depth of penetration (a schematic). The indenter response at low loads was influenced by a thin hard oxidic layer on the surface. \( F_i \) – initial load, \( h_i \) – initial depth of penetration, \( h_i' \) – initial depth obtained by extrapolation of measured depths (crosses) without considering the surface layer.

Three notes must be made here: 1) Sometimes, the shape of the load-depth curve at very low loads is influenced by a very thin oxidic layer on the specimen surface, formed here during its preparation (Fig. 9). The decision about what part of the \( h-F \) curve should be used for the determination of \( h_i \) in such case needs some experience. 2) The question of the initial depth of penetration is less important for advanced indentation devices with very high load and depth sensitivity, enabling the use of very low loads. 3) The above approach for the determination of the initial depth of penetration is not suitable for very compliant materials such as gels, where also adhesive forces play a role; see Section 11.

8. Temperature changes, drift

Errors can be caused by a gradual change of specimen or instrument dimensions due to temperature changes. If the specimen temperature at the test beginning is higher, it decreases during the measurement, and its dimensions become smaller. The sample surface
moves in the same direction as the indenter tip, so that the apparent depth of indenter penetration, \( h_a \), will be larger (Fig. 10), and so also the apparent contact depth, \( h_{c,a} \). Thus, the apparent hardness, \( H_a \), will be lower. The situation with elastic modulus is more complex. If the specimen temperature decreases only during the loading period, the apparent modulus (I) will be lower because of larger apparent contact depth and area. If, however, the temperature decreases during unloading, the indenter withdrawal from the contracting specimen lasts a shorter time and the seeming contact stiffness \( S_a \) will be higher and the apparent modulus \( E_a \) can be higher (cf. the higher slope of the unloading part of the \( F - h \) diagram in Fig. 10); an opposite role is played by the larger depth. The effect of temperature decrease during the unloading is thus more complex than during loading. The opposite situation occurs if the specimen temperature during the test increases. Similar effects to the sample temperature changes can be caused by the device temperature changes.

![Figure 10. Load – penetration diagram without drift (solid lines) and with drift (dashed lines)](image)

Several approaches are used to mitigate the thermal drift effects. The simplest way is to make the measurements only if all temperatures (specimen, nanoindenter) are constant during the test. Sufficient time before measurements is necessary for temperature equalization. The device is usually placed in a cabinet in order to minimize the random temperature variations. Another approach corrects the measured \( h_a \) values by subtracting the displacement caused by thermal drift. In this case, a dwell several tens of seconds is inserted into the unloading part of the test cycle (usually when the indenter load has dropped to 10% of the maximum value) to measure the velocity of the indenter movement due to the drift. This velocity is then used in a model for drift correction.

The influence of temperature is also reduced by making the critical parts of the device from materials with extremely low coefficient of thermal expansion. One device of this kind (UNHT made by CSM Instruments) uses also a reference shaft with a spherical tip, parallel with the indenter and close to it (several mm). This reference shaft, with its own actuator and sensor, is in contact with the specimen surface, so that the indenter tip displacement is measured not from the nanoindenter body, but from the specimen surface. With all these sophisticated tools, the drift in very advanced devices is lower than 1 nm per minute.

The influence of drift is also smaller if the indenter is loaded by a small harmonic signal. If the loading cycle does not last more than about \( 0.01 - 0.1 \) s, the creep or drift effects are
negligible. On the other hand, if viscoelastic materials are tested, their response depends on the frequency of loading, and the harmonic contact stiffness $S_h$ (and thus also elastic modulus) usually increases with the excitation frequency. However, if small harmonic oscillations are superimposed on the slowly growing load (CSM mode), the total indenter displacement is the same (and influenced by the drift in similar way) as in the common quasistatic load-unload cycle (Menčík et al., 2005). With harmonic loading also a danger exists that the dissipated energy can change the temperature and viscosity in the loaded volume, and thus also its mechanical properties.

9. Contact profile, pile-up

In depth-sensing tests, contact depth $h_c$ and area $A$ are not measured, but calculated from the total indenter displacement $h$ via Eqs (2) and (11, 12, 13). In Equation (2) one usually assumes $\varepsilon = 0.75$, which follows from the analysis of elastic contact with a pointed or spherical indenter. According to this model, the surface around the indenter sinks down. If plastic deformations appear, ductile material is sometimes squeezed out upwards around the indenter, Fig. 5. If this pile-up occurs, the actual contact area is larger than the apparent values, and the material seems to be stiffer. If no correction for pileup is done, the calculated (apparent) hardness and elastic modulus are higher than in reality.

Important for pileup is the ratio of the elastic modulus and yield strength of the material, its strain-hardening behavior, and also the indenter shape. Pile-up around a pointed indenter is typical of materials, which have high ratio of elastic modulus to yield stress and do not strain-harden. Examples are relatively soft metals, which were cold-worked. High $E/Y$ ratio (low hardness) means relatively large plastic zone, and thus a large volume of material around the indenter, which could be displaced by plastic flow upwards. In materials with strain-hardening, the yield stress increases during plastic deforming. The intensively deformed surface layer around the indenter becomes harder and constrains the upward flow of the material. For hard and highly elastic materials, such as glasses and ceramics, the $E/H$ ratio is lower, the size of the plastic zone is small and the volume of the material displaced by the indenter can be accommodated by the elastic deformations of the material around. In this case, pile-up is not observed.

The error in the calculated contact area and hardness due to incorrect use of elastic formulae for materials with significant pile-up can be as high as 60% (Oliver & Pharr, 1992). For elastic modulus it is lower, but it can also amount up to several tens of percent, as it follows from the role of contact area $A$ in Eqs. (1). The situation with pointed indenters was studied by comparison of the measured $h_c(h)$ and $A(h)$ values with those obtained by direct imaging of the imprints (McElhaney et al., 1998) and by the finite element modeling (Bolshakov & Pharr, 1998, Hay et al., 1998). Several conclusions can be made from these studies. Pileup can be expected for $E/H = 10^2$ and more. The stress field around a pointed indenter is self-similar and the degree of pileup does not depend on the depth of penetration. A simple quantity to assess the material propensity to pile-up is the ratio of the final depth of the imprint after unloading to the total depth under load, $h_f/h$. Pileup is large if $h_f/h$ is close to 1.0 and the work-hardening
is small. If \( h/h < 0.7 \), none or only very small pileup is observed regardless whether the material work-hardens or not (Oliver & Pharr, 2004, Bolshakov & Pharr, 1998).

The situation with spherical indenters is more complex. The mean contact pressure and the stresses and strains increase with the depth of penetration. The representative strain is usually expressed as the ratio of the contact radius \( a \) and indenter radius \( R \). At relatively low loads, the deformations are only elastic, and the material around the indenter sinks. Plastic flow starts at the mean contact pressure \( p_m = 1.1 \sigma_Y \), but the Hertz formulae for the displacement \( h(F) \) are approximately valid till \( p_m = 1.6 \sigma_Y \), when the volume of the plastically deformed zone beneath the indenter tip becomes so large that this material is pushed upwards around the indenter (Mesarovic & Fleck, 1999). The tendency to pile-up increases with increasing ratio \( E/Y \) and the relative depth \( h/R \) of indenter penetration. From a finite element study (Mesarovic & Fleck, 1999) it follows that for a non-strain-hardening material with \( E/Y = 200 \), the pile-up is negligible for \( h/R < 0.001 \), and should be considered for \( h/R > 0.01 \). For materials exhibiting strain-hardening, the tendency to pile-up is smaller.

Pile-up can also occur if a relatively soft coating on a hard substrate is indented: the coating material is displaced upwards from the contact zone around the indenter more than in a homogeneous specimen. The opposite case are materials, which densify under high compressive stresses. Here, sinking-in of the surface is more pronounced than in common elastic materials.

More about pile-up can be found in (Fischer-Cripps, 2004) and (Oliver & Pharr, 1992); some methods for its corrections were proposed in (Cheng & Cheng, 1998). In all cases with a suspicion of errors due to pile-up or sink-in, direct observation of the imprint (for example via AFM) is recommended for obtaining more reliable values of the contact area.

10. Indentation size effect

It is well known that hardness of metals, measured by a pointed indenter at small depths of penetration, often increases with decreasing depth of penetration (Fig. 11). This so-called indentation size effect (ISE) becomes observable for imprint depths smaller than several µm. ISE can appear also with a spherical indenter, though here the apparent hardness does not depend on depth, but increases with decreasing radius. Even some ceramics exhibit indentation size effect to a small extent.

There are several reasons for depth dependence of hardness in metals. Sometimes, a thin hard oxidic layer has been formed on the specimen surface during its preparation (Fig. 9). Surface layer can also become harder due to strain hardening or residual stresses caused by plastic flow during polishing. Another reason is the bluntness of the tip of every real indenter. For example, the tip of a Berkovich pyramid is roughly spherical with the radius 50 – 100 nm. As a consequence, the actual contact area for very small distances from the tip is much larger than at the same distance for an ideal indenter (Fig. 5). For a given load, the penetration of a real indenter is thus smaller, and the hardness appears to be higher. With increasing depth, the influence of the tip bluntness becomes smaller and the measured
hardness values approach gradually to those for the ideal shape. However, this is no actual indentation size effect, but an error, which can be avoided by a proper indenter calibration.

Nevertheless, even if these errors are removed, there still may be some dependence of hardness on indentation depth. According to the classical continuum mechanics, hardness of a homogeneous material, measured by a self-similar indenter, should be constant. However, plastic deformation of real crystalline materials is facilitated by the presence of dislocations, more with their higher density. If the indentation depth is large, the influenced volume contains many dislocations, and the plastic flow is insensitive to this depth. If the size of the indented volume becomes comparable with the distance between dislocations, less dislocations are present, and the average contact stress for the initiation of plastic flow must be higher. The role of dislocations also explains why the size effect is more significant in annealed samples than in cold-worked ones of the same basic material, as the latter contain more dislocations created during the mechanical treatment.

![Figure 11. Depth dependence of hardness (ISE) of a single crystal Cu (after Nix & Gao, 1998)](image)

A semiempirical formula has been proposed for the depth dependence of hardness (Nix & Gao, 1998):

\[
H = H_0 \left[1 + \left( \frac{h^*}{h} \right)^{3/2} \right],
\]

(31)

\(H\) is the hardness for a given depth \(h\) of indentation, \(H_0\) is the macro-hardness (corresponding to large indents), and \(h^*\) is a characteristic length that depends on the shear modulus, on \(H_0\) and on indenter shape. The constant \(h^*\) (and \(H_0\)) for the tested material can be obtained by fitting a series of \(H(h)\) values. It appears that a correct description of material response in micro or nano scale needs – in addition to „macro“ parameters such as elastic modulus or yield strength – also a parameter, describing characteristic dimension of the microstructure. This material length scale may be defined, e.g., as (Nix & Gao, 1998):

\[
l^* \approx \frac{L^2}{b} = c / (b \rho),
\]

(32)

where \(c\) is a non-dimensional constant, \(L\) (m) is the mean spacing between dislocations, \(b\) (m) is the Burgers vector, and \(\rho\) (m\(^{-2}\)) is the average density of dislocations, calculated as the
number of dislocation lines crossing a unit area (or as the total length of dislocation lines in a certain volume, divided by this volume). Material length scale \( l' \) can be related to the characteristic length \( h' \). Also other formulations for \( l' \) exist.

Equation (31) indicates that hardness is less dependent on depth for smaller values of \( h' \) (and \( l' \)), i.e. for higher density of dislocations. Nix & Gao (1998) have shown that hardness does not depend much on the depth of penetration if the material is intrinsically hard. An example is fused quartz, whose hardness is essentially independent of depth of indentation.

Indentation size effect also occurs if hardness is measured by a spherical indenter. In this case, however, it depends on the indenter radius rather than on the indentation depth. The contact-influenced volume, corresponding to a certain depth of penetration, is smaller for smaller indenter radius \( R \), so that it will contain less pre-existing dislocations. Beginning from certain small \( R \), hardness will increase with decreasing tip radius. Formula (31) may also be used for a spherical indenter, if the depth \( h \) is replaced by the curvature radius \( R_p \) of the residual surface impression and \( h' \) is replaced by \( R' = c'/h_p \) (Swadener et al., 2002).

The relationship between the imprint size and the measured hardness is influenced by several factors and can be more complex. If the size of the influenced volume is smaller than the distance between the dislocations, the plastic flow starts only when the local stress reaches a value high enough to generate new dislocations. A role is also played by the fact that the strain field under the indenter is strongly inhomogeneous. This is considered in the strain gradient plasticity theory (Nix & Gao, 1998). Nevertheless, Gerberich et al. (2002) presume that gradient plasticity models do not apply at very shallow depths. Moreover, for very small depths of indentation by a spherical tip, the surface-related phenomena become important, such as friction or consumption of energy for the creation of larger surface as the indenter penetrates into the originally flat specimen.

In brittle materials, contact compliance is influenced by the presence of cracks in the contact region. Cracks can be present here from the manufacture or can be created during the test. Glass objects contain many tiny surface cracks, nucleated in mutual sliding with other bodies. Ceramics contains weak places from manufacture (broken crystalline grains, cracks from machining or polishing, etc.). Cracks are also formed during indentation (Zhu et al., 2008), for example at the imprint edges. For larger contact area or volume a higher probability exists that a larger pre-existing defect will be present in the stressed region, making the contact compliance higher. If the contact area is smaller, less cracks are present (or even none), and higher contact pressure is necessary for their nucleation or growth.

More about indentation size effect can be found in the above cited references and also in (Sargent, 1989, Huang et al., 2006, Spary et al., 2006, Rashid & Al-Rub, 2007, Sangwal, 2009). The most important conclusion is that for small imprints, the information on the indent size must always be given together with the hardness value and other details of the test.
11. Surface forces and adhesion

The models for instrumented indentation usually consider only the action of indenter load. However, this is correct only if this load is higher than several tens of μN. If it is lower, also adhesive forces should be considered, otherwise the results can be erroneous. Atoms of solids attract mutually, but the range of action of these attractive forces is very short, about 1 nm. Atoms in a very thin surface layer act also onto any other matter close to the surface, and this can result in adhesive joining. If the distance between two surfaces drops below some very small value, the bodies sometimes cling spontaneously and remain in close contact even without action of external force, and separate suddenly only after they are pulled apart by a certain tensile force (Fig. 12). In contact of relatively stiff solids, adhesion is usually not observed. Due to surface roughness, both bodies come into contact only at a limited number of matching asperities. The area of close contacts is minute, and the resultant attractive force is insufficient to compress the “hills”. If, however, both surfaces are very smooth (roughness of the order of nm) or very compliant, the surfaces adjust one another better and the adhesion is stronger.

![Figure 12. Indentation into a soft material influenced by adhesion and viscoelasticity (a schematic).](image)


Adhesion can also influence nanoindentation measurements, especially for compliant materials and very low loads, from μN to mN. Here, surface forces change the character of load-displacement curves (Gupta et al, 2007) and must be considered in the preparation and evaluation of the tests.

Several theories of adhesion exist; for an overview see Israelachvili (1992). Derjaguin, Muller and Toporov (1975) have developed theory of contact of stiff bodies (DMT), which is used, for example, in AFM indentation. For common compliant materials, the “Johnson–Kendall–Roberts” (JKR) model, proposed by Johnson et al. (1971), is used most often. It assumes spherical shape of contacting bodies and is based on Hertz’ theory of elastic contact, but considers also adhesive forces. Due to them, the contact deformations differ from those by Hertz. The actual contact radius $a$ is larger and during unloading it remains finite even to zero load ($F = 0$), for which:

$$a_0 = \left[9 \pi R^2 W / (2E_s) \right]^{1/3};$$  \hspace{1cm} (33)
$R$ and $E_r$ are the equivalent radius and elastic modulus of contacting bodies, and $W$ is the specific surface energy of the contact of both bodies. The solids keep together up to a small tensile force, and suddenly separate at

$$F_{ad} = \left(\frac{3}{2}\right) \pi R W,$$

with the corresponding contact radius $a_{ad} = a_0/4^{1/3}$. For example, adhesion force $F_{ad}$ of a glass sphere of $R = 100 \, \mu\text{m}$ on rubber ($W = 62 \, \text{mJ/m}^2$) is $29 \, \mu\text{N}$.

The resultant adhesive force depends also on the roughness of the surfaces in contact, on the viscosity of one or both bodies and also on the presence of a liquid at the contact. The influence of surface roughness was studied by Johnson (1976) and Tabor (1977) on asperities of spherical shape and randomly distributed heights. Each asperity adheres until its distance from the counter-surface exceeds some critical value, $\delta_c$. The ratio of the standard deviation $\sigma$ of asperity heights and the distance $\delta_c$ at separation, called elastic adhesion index $\alpha = \sigma/\delta_c$, can be related with the ratio of the force $F(\sigma)$, required to compress an asperity by $\sigma$, and the force of adhesion for that asperity, $F_{ad}$. Figure 13 shows how the apparent adhesion force decreases with increasing roughness. For $\alpha = 0$, this force is maximum and corresponds to the JKR model. For $\alpha = 1$ it drops to 50%, for $\alpha = 2$ it drops to 10%, and becomes negligible for $\alpha > 3$.

Adhesion is stronger if one or both bodies are very compliant so that the surfaces can easily adjust one another. Examples are gels, soft elastomers, or biological tissues. Their response, however, is visco-elastic, which also influences adhesion. Generally, the force at which two viscoelastic bodies get adhered during mutual approaching, is lower than the force needed for their separation on unloading (Fig. 12). A role is played by the decreasing apparent modulus $E(t)$ and growing contact area with time; the force for separation thus corresponds to a larger area. Moreover, this force also depends on the rate of separation – it is higher for faster separation. Greenwood & Johnson (1981) came to the conclusion that during the mutual catching or separation, the surface energy $W$ in the JKR model should be modified by the ratio of instantaneous and asymptotic modulus, $E_0/E_\infty$. A fracture mechanics approach to adhesion was presented by Maugis & Barquins (1978). Recently, an alternative model was proposed by Kesari et al. (2010).

Figure 13. Relative decrease of adhesive force $(F/F_{ad})$ with growing relative surface roughness $\alpha = \sigma/\delta_c$; $\sigma$ – standard deviation of asperity heights, $\delta_c$ – critical distance at the loss of adhesion. (A schematic, after Johnson, 1976)
If a thin layer of liquid is present between contacting bodies, it contributes to their mutual attraction by its surface tension. This is also the case in nanoindentation (Fig. 14):

1. Tensile surface stress $\gamma$ acts along the circumference of the liquid ring around the indenter-sample contact and attracts the indenter to the specimen by the force $F_\gamma = 2\pi a \gamma$, where $a$ is the contact radius and $\gamma$ is the surface tension between the liquid and the air or another surrounding environment.

2. If the liquid wets the indenter and specimen, its shape is torroidal and the surface tension causes underpressure in it, $p = \gamma/r$, which attracts the indenter to the specimen by the force

$$F_\text{up} = -\pi a^2 p = \pi a^2 \gamma / r;$$

(35)

$r$ is the meniscus radius (Israelachvili, 1992). This capillary force is much bigger than $F_\gamma$ and can also be expressed by the contact angle $\phi$ between the liquid and indenter or specimen. With some simplifications, the attractive force is (Israelachvili, 1992)

$$F_\text{up} \approx 4\pi \gamma R \cos\phi.$$  

(36)

For example, water ($\gamma = 73$ mN/m) between the sample and indenter with $R = 100$ µm (and $\phi = 20^\circ$) can increase the adhesion force by 86 µN.

One interesting phenomenon must be mentioned. If the liquid wets the surface well (the contact angle $\phi$ is small), even its vapours can condense spontaneously in a narrow space between two surfaces, and the corresponding attractive force can appear even without obvious presence of a liquid, just under sufficient concentration of its vapours. This also can happen between a spherical indenter and a specimen. For more, see Israelachvili (1992).

The presence of adhesive forces influences the load-depth curves and must be considered in the evaluation of low-load indentation tests. Elastic modulus is usually determined from a load-unload test by the Oliver & Pharr (1992) procedure. With a spherical indenter and elastic deformations, $E$ can also be calculated from the expression (Gupta et al., 2007):

$$E_{r, H} = \left[ \frac{3}{R F} \right]^{1/2},$$

(37)

which follows from the Hertz formula, relating the indenter load $F$ and displacement $h$. $R$ is the indenter radius and $S$ is the contact stiffness at the nominal load. The advantage of Eq. (37)
is that one does not need to know the contact area $A$ nor the instant of the first contact, necessary for the determination of indenter penetration $h$. If the surface forces are not negligible, the relationship between load and displacement is not proportional to $h^{3/2}$, but nearly linear for very low loads (Gupta et al., 2007). If the JKR model for adhesion is assumed, it is possible to determine the reduced elastic modulus using a spherical indenter and formula (37), but modified as follows (Carrillo et al, 2005, 2006, Gupta et al., 2007, Menčík, 2012):

$$F_{r, JKR} = F_{r, H} \times \Phi;$$

$$\Phi = \left[ \frac{2}{3} \right] \left[ 1 + \frac{1}{\psi} \right]^{3/2} + 1 \right]^{1/2} / \left[ 1 + 2 \psi^{-1} \left[ 1 + \left( 1 + \psi \right)^{1/2} \right]^{1/2} \right];$$

$\Phi$ is the correction factor (Fig. 15), and $\psi$ is the ratio of indenter load and adhesion force, $F/F_{ad}$. The determination of $E$ needs (in addition to $F$ and $R$) the knowledge of adhesion force $F_{ad}$ and contact stiffness $S$. They both can be ascertained in one load-unload test.

![Figure 15. Ratio $\Phi$ of E-modulus calculated with the consideration of adhesive forces ($E_{JKR}$) and without ($E_{Hertz}$), plotted as a function of the ratio of the load $F$ and adhesive force $F_{ad}$](image)

12. Scatter of measured values

In principle, there are three kinds of errors: random, systematic, and gross. The gross errors occur due to extreme conditions in a particular test, such as random mechanical shock or sudden fluctuation of temperature. It usually does not appear if the test is repeated. Systematic errors are caused by inadequate conditions of measurement or data processing.

They can be revealed by the use of another method. Random errors are observable in repeated measurements, and can be reduced by processing the data by statistical methods.

Random scatter of individual values is caused by the noise in the device, by random variations of test conditions, and, especially, by the variability of material properties and inhomogeneities. At low loads, the results are influenced more by the random surface roughness of the indented place. Sometimes, porosity plays a role. Common also is the variability of coating properties along the surface; the range of values measured over a large area can be much wider than for the area of size less than 1x1 mm². Statistical data
processing is necessary, and the number of tests should be adjusted to the scatter of the data.

If a certain material parameter $X$ is to be determined, for example elastic modulus, one usually wants to ascertain it as accurately as possible. The uncertainty due to random scatter can be reduced by repeating the measurements. The confidence interval, which contains the true mean value $\mu_X$ with probability $\alpha$, is

$$
\mu_X = \bar{x} \pm t_{\alpha,n-1} \frac{s}{\sqrt{n}},
$$

where $\bar{x}$ is the average value of $x$ (calculated as $\Sigma x_i/n$), $s_x$ is the standard deviation, $t_{\alpha,n-1}$ is the $\alpha$-critical value of the Student’s distribution for $n-1$ degrees of freedom, and $n$ is the number of measurements. For $n > 30$, the width of the confidence interval is roughly indirectly proportional to $\sqrt{n}$. In this way, the parameter $X$ could, in principle, be determined with any accuracy, just by increasing $n$. In reality, the number of tests is a compromise between the demanded accuracy and the costs of measurements.

The scatter of measured values often reflects the inherent variability of the pertinent property. For example, strength of a brittle material, such as ceramics, is given by the size of internal defects, which is random and depends on the raw materials and manufacturing conditions. Large scatter is also typical of biological samples. In such cases, there is no single value of the parameter $X$, and the quantity should be described by its probability distribution. The knowledge of this distribution allows the determination of confidence interval, containing certain percentage of all possible values or probable extreme value, for example the minimum strength that will not be attained only with some allowable probability, e.g. 1%. Analytical distributions, common in materials sciences (normal, log-normal, Weibull, etc.) are described by their parameters; the procedures for their determination can be found in textbooks on statistics. The number of measurements must be so high as to allow reliable determination of the distribution parameters. In some cases, the histogram of measured values cannot be approximated by a simple analytical function, and the use of order statistics is more useful.

Nanoindentation is also used for the investigation of heterogeneous materials consisting of several phases. Examples are concrete, ceramics, polycrystalline metals, plastics with hard particles dispersed in soft matrix, etc. Here, the measured values depend on the imprint size and thus also on the load. Large imprints, caused by high load, encompass many particles and the matrix between them, and give the average properties of the compound. If, on the contrary, the imprint is small compared to the size of the individual particles, and is near the centre of a particle (or matrix) and sufficiently far from the interface with another phase, the property of the phase alone is obtained. However, reliable indenter positioning is possible only if the individual particles or phases are visible in the nanoindenter’s microscope. Sometimes, the phases cannot be distinguished optically. In such case, a grid of several hundred indents is created, e.g. 20x20, distributed uniformly over an area. The histogram of the measured property (Fig. 16) often exhibits several, less or more pronounced, peaks. This indicates that the measured values are a mixture of two or more probability distributions, each for one phase (Constantinides et al., 2006). In such
case, the decomposition of the resultant distribution can yield the distribution parameters of the individual phases. For a composite of N-phases, with Gaussian distribution each, the probability density of the measured quantity \( X \) is given as the sum of densities of the individual quantities,

\[
f(X) = \sum_{j=1}^{N} \frac{p_j}{\sqrt{2\pi}\sigma_j} \exp \left( -\frac{(X - \mu_j)^2}{2\sigma_j^2} \right), \quad \sum_{j=1}^{N} p_j = 1,
\]

where \( \mu_j \) and \( \sigma_j \) are the mean and standard deviation of the \( j \)-th phase \((j = 1, 2, \ldots, N)\), and \( p_j \) is the relative proportion of this phase. The procedure for obtaining all parameters is described in (Constantinides et al., 2006, Sorelli et al., 2008) or the chapter by J. Němeček in this book; examples can be found in Němeček & Lukeš (2010), Nohava & Haušild (2010) and Nohava et al. (2012).

![Figure 16. Distribution of property X in a two-phase material (a schematic). f – relative frequency (\( n_j/n \); \( n \) is the total number of indents), A, B – distribution of property A or B alone, C – distribution of all X-values, obtained by grid indentation.](image)

Until now, this method has been used mostly for systems consisting of two phases \((N = 2)\). Unfortunately, some indents occur at or near the interfaces of both phases, and this influences the response (similarly like the response to indentation into a coating is influenced by the substrate). The interaction depends on the indent size and on the difference of properties of the phases; see, e.g., Buršík (2011) or Němeček & Lukeš (2010). Sometimes, therefore, trimodal Gaussian distribution is assumed, with the third fictitious „phase” pertaining to the influenced results. In this case, \( N = 3 \) and the genuine properties of the phases 1 and 2 can be determined better. Examples of application are described in (Nohava et al., 2012). In principle, this method can be used for materials consisting of more phases. However, the number of indents in such analysis must be very high, especially if the probability distributions of individual quantities overlap and the differences between the properties of individual phases are low. It can also be recommended to make several grids over a larger area.
13. Summary and conclusion

Accuracy of determination of mechanical properties by nanoindentation can be impaired for various reasons, such as inadequate model for data processing or omission of some specific properties or factors, such as pileup, surface forces or temperature changes. In this chapter, individual sources of errors were reviewed and explained and formulae or advice given for their avoiding or reducing.

Author details
Jaroslav Menšík
University of Pardubice, Czech Republic

Acknowledgement
The author highly appreciates the effort of InTech Publishing House to facilitate access to knowledge to everyone. He is also grateful for its help in the preparation of this chapter.

14. References


