Chapter 1
Surfaces in Contact

Friction and wear depend on the characteristics of the mating surfaces. The difficulty to explain and to predict with high accuracy such phenomena reflects the complex nature of the surfaces, which is determined by the material properties (such as the microstructure), the geometrical irregularities, the presence of oxides due to the interaction with the surrounding atmosphere, and the presence of organic molecules, water vapour or other impurities adsorbed from the environment. Therefore, when two bodies are brought into close contact, the relevant features of their surfaces determine the nature of the interaction, which has a mechanical character, with the formation of a stress and strain field in the contact region, and a physical-chemical nature, with the establishment of physical or chemical bonds.

To quantitatively evaluate the contact stresses, it is convenient to introduce the concept of *smooth surface*, i.e., of a surface free from geometrical irregularities. This is obviously an ideal vision since it is impossible to produce smooth surfaces at a molecular level. Using the contact mechanics and, in particular, the theoretical analysis developed by Hertz for linear elastic bodies under this assumption, useful relations for the contact stresses and deformations can be obtained. They can be profitably employed when the bodies are in elastic and frictionless contact, with the assumption that the radius of the contacting bodies is large compared to the contact zone size.

This chapter firstly introduces the main concepts of contact mechanics and the types of material response to the contact stresses. Subsequently, the micro-geometrical characteristics of real surfaces are illustrated, showing how they affect the contact at the microscopic level. The chapter is completed with the analysis of the phenomenon of adhesion, which depends on the physical-chemical interactions in the contact region.

G. Straffelini, *Friction and Wear*, Springer Tracts in Mechanical Engineering,
DOI 10.1007/978-3-319-05894-8_1

1.1 Contact Between Ideal Surfaces

1.1.1 Elastic Contact

From a geometrical point of view, the contact between two bodies may be conformal or non-conformal (Fig. 1.1). A conformal contact occurs when the mating surfaces fit nearly together. Such a contact occurs, for example, in sliding bearings (between bearing and shaft) or in drawing processes (between wire and tool). If the contacting profiles are rather different, the contact is non-conformal, and it theoretically occurs at a point or along a line. A point contact is present, for example, in rolling bearings (between ball and seat), whereas a line contact occurs in gears (between tooth and tooth).

In the case of conformal contact, the nominal area of contact (A_n) has a finite extension and its determination is straightforward. Also in the case of non-conformal contact, the nominal area of contact is finite because of the local deformations. Figure 1.2a schematically shows the contact between two spheres of radius R_1 and

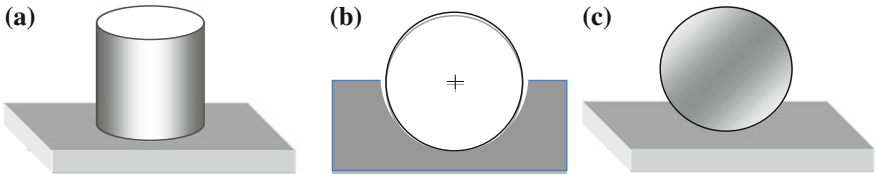

Fig. 1.1 Contacts between ideal surfaces. Conformal contact between the base of a cylinder and a plane (**a**), and in a sliding bearing (**b**). Non-conformal contact between a sphere and a plane (**c**)

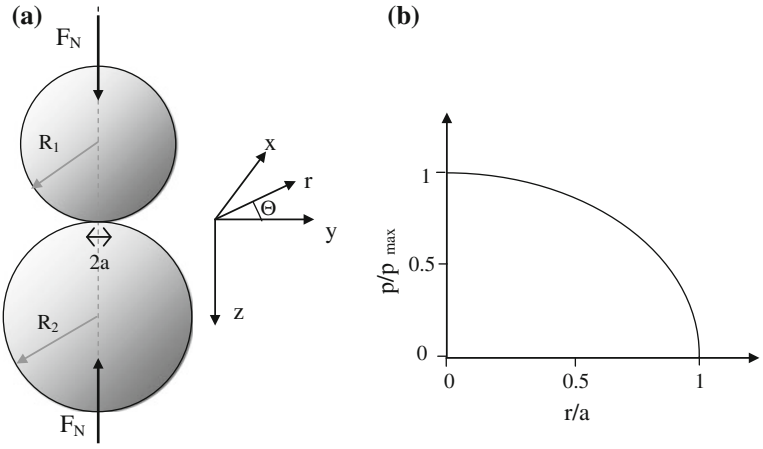

Fig. 1.2 **a** Point contact between two spheres and definition of the coordinates (the x-y plane is the contact plane, and the z axis lies along the load line directed positively into the lower sphere); **b** variation of contact pressure as a function of distance r from the centre of the contact area (for z = 0)

R_2 (the subscripts 1 and 2 refer to the two bodies). In this case, the contact is theoretically in a point. Indicating with E_1 and E_2 the elastic moduli of the materials of the two spheres, the nominal (or apparent) area of contact can be evaluated from the following relationship:

$$A_n = \pi \cdot a^2 \tag{1.1}$$

where a is the radius of the circular contact region, given by:

$$a = \left(\frac{3 \cdot F_N \cdot R'}{2E'}\right)^{1/3} \tag{1.2}$$

where F_N is the applied normal force, R' and E' are the reduced radius of curvature and the effective modulus of elasticity, respectively. They are defined as such:

$$\frac{1}{R'} = \frac{1}{R_1} + \frac{1}{R_2} \tag{1.3}$$

$$\frac{1}{E'} = \frac{1}{2} \cdot \left(\frac{1 - \nu_1^2}{E_1} + \frac{1 - \nu_2^2}{E_2}\right) \tag{1.4}$$

where ν_1 and ν_2 are the Poisson's ratios of the materials of the two spheres.

Figure 1.2b shows the contact pressure distribution (for z = 0). According to the Hertz theory, it is semi-elliptical:

$$p = -\sigma_z(z = 0) = p_{\max} \cdot \sqrt{1 - \left(\frac{r}{a}\right)^2} \tag{1.5}$$

and the maximum value, which occurs at the centre of the contact, i.e., at r = 0, and is known as the *Hertzian pressure*, is given by:

$$p_{\max} = \frac{3F_N}{2\pi a^2}. \tag{1.6}$$

The evolution of the corresponding surface stresses, in polar coordinates, is illustrated in Fig. 1.3a for $\nu_1 = \nu_2 = 0.3$. Note that a tensile radial stress develops at the edge of the nominal area of contact. Its maximum value (for r = a) is given by the following:

$$\sigma_r = p_{\max} \frac{1 - 2\nu}{3} \tag{1.7}$$

The stresses along the load line (z axis) and for r = 0 are shown in Fig. 1.3b. For symmetry reasons, they are also principal stresses. At the surface: $\sigma_z = -p_{\max}$ and

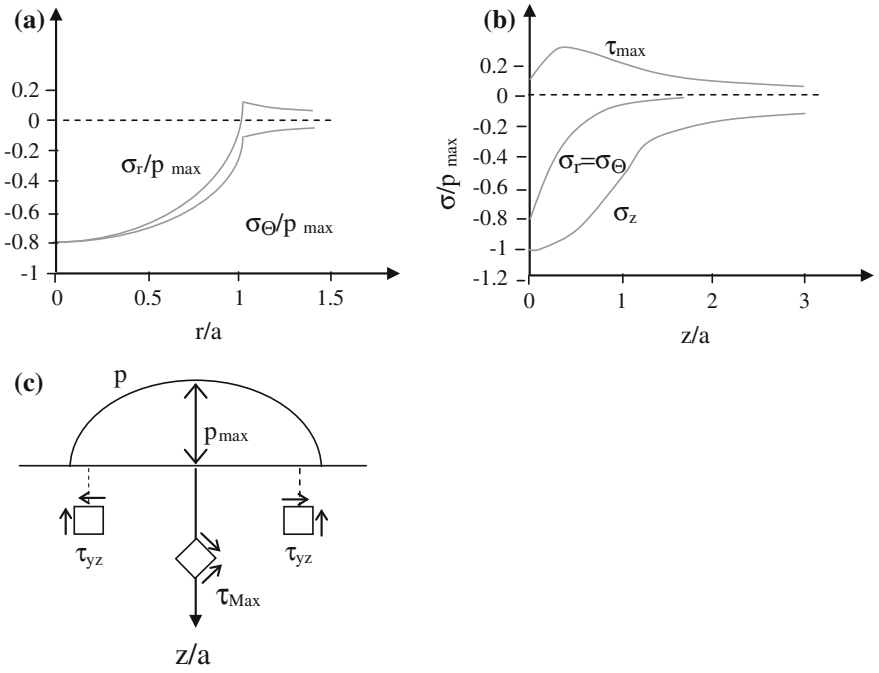

Fig. 1.3 Elastic contact of spheres (ν = 0.3): **a** distribution of normalized Hertzian stresses σ_r and σ_Θ at the surface (z = 0); **b** distribution of normalized stresses σ_r, σ_Θ, σ_z and τ_{max} along the z axis, i.e., moving inside one of the two spheres; **c** schematization showing the maximum values of τ_{max} and τ_{yz}

$\sigma_r = \sigma_\Theta = (1/2 - (1 + \nu))\, p_{max} = -0.8\, p_{max}$ for ν = 0.3. In Fig. 1.3b the *maximum shear stress* (τ_{max}) distribution along the z-axis is also shown. τ_{max} is defined by:

$$\tau_{max} = \frac{1}{2}|\sigma_z - \sigma_r| \tag{1.8}$$

It is oriented at 45° with respect to the contact surface and reaches its maximum value, equal to $\tau_{Max} = 0.31 p_{max}$ at a distance $z_m = 0.48a$ from the surface (these values are obtained with ν = 0.3).

The stress field far from the z-axis is characterized by the presence of stresses whose modulus is lower than at the load line. Of particular importance is the occurrence of a shear stress, τ_{yz} (i.e., normal to the z and y axes), which is due to the lateral displacement of material beneath the flattened contact area. It is parallel to the contact surface and is maximum at a depth of 0.5a and at a distance of ± 0.87a from the z-axis. Its maximum value is $0.25 p_{max}$. See the schematization in Fig. 1.3c.

The parameters a, p_{max}, τ_{Max}, z_m, and the nominal pressure p_o (defined by F_N/A_n), for the configurations sphere/sphere (point contact) and cylinder/cylinder (line contact) are listed in Table 1.1. In the table, the relations for the mutual

Table 1.1 Equations for the calculation of the contact parameters for elastic solids

	Sphere/sphere (point contact)	Cylinder/cylinder (line contact)
a	$\left(\frac{3F_N R'}{2E'}\right)^{1/3}$	$\sqrt{\frac{8F_N R'}{\pi E' L}}$
δ	a^2/R'	$\frac{2F_N}{\pi L}\left\{\frac{1-\nu_1^2}{E_1}\left[\ln\left(\frac{4R_1}{a}\right)-\frac{1}{2}\right]+\frac{1-\nu_2^2}{E_2}\left[\ln\left(\frac{4R_2}{a}\right)-\frac{1}{2}\right]\right\}$
p_{max}	$3F_N/2\pi a^2$	$2F_N/\pi aL$
p_o	$0.67p_{max}$	$0.78p_{max}$
τ_{Max}	$0.31p_{max}$	$0.3p_{max}$
z_m	0.48a	0.786a

The parameters τ_{Max} and z_m in case of the sphere/sphere contact are determined considering $\nu = 0.3$. The cylinder/plane and sphere/plane contacts are special cases of the cylinder/cylinder and sphere/sphere contacts occurring when R_1 (or R_2) tends to infinity

displacements, δ, along the load line are also reported. The parameters a and δ are representative of the local elastic strains (for further details on the contact stresses and displacements see, for example, Refs. [1–3]).

For a conformal contact, the nominal contact pressure is simply given by the ratio between the applied normal force and the nominal area of contact (A_n). The stress distribution in the contact area, however, is strongly affected by the stress concentration exerted by the edges. If the radius of curvature of the edge tends to zero, the pressure at the edge tends to infinity, as schematically shown in Fig. 1.4a, for a contact between a punch and a plane. Such stress intensification can be alleviated by rounding the edges. Figure 1.4b schematically shows the contact between a punch with rounded corners and a plane. A method to calculate the contact pressures in this case is reported in Ref. [4].

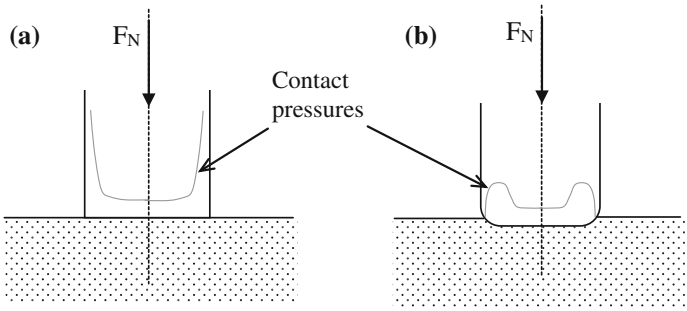

Fig. 1.4 Pressure distribution for a contact between a punch and a plane. In **a** the radius of curvature of the edges tends to 0, while in **b** the effect of a rounding of the edges is shown

1.1.2 Viscoelastic Contact

Some materials, like polymers, may display a particular deformation behaviour that is affected by elastic, viscoelastic and plastic processes. Following the application of a stress σ, the total deformation ε_t, is thus given by the sum of three terms: the instantaneous elastic deformation, ε_1, the viscoelastic deformation, ε_2, and the plastic deformation, ε_3:

$$\varepsilon_t = \varepsilon_1 + \varepsilon_2 + \varepsilon_3 = \frac{\sigma}{E} + \frac{\sigma}{E_r} \cdot \left(1 - e^{\frac{E_r \cdot t}{\eta_r}}\right) + \frac{\sigma \cdot t}{\eta_0} \tag{1.9}$$

where E is the elastic modulus, E_r is the viscoelastic modulus, η_r is a damper parameter, η_o is the viscosity parameter and t is time. As a consequence, the nominal area of contact is greater than that predicted by the Hertzian theory, and it increases with time. As an example, Fig. 1.5 shows the experimental dependence of the contact deformation displacement, δ, as a function of time in the case of a polypropylene (PP) sphere pressed against an optically transparent plane [5]. In polymers the viscoelastic contribution is particularly marked if the material contains an amorphous phase and is at a temperature above its glass transition temperature, T_g. Polypropylene is a semi-crystalline polymer (with about 60 % of amorphous phase) and has a T_g of −14 °C (see Sect. 6.8 for further details on the properties of polymers for tribological applications).

Similarly to elastic deformations, even viscoelastic deformations are recoverable, although not instantaneously but over a period of time after unloading. In addition, energy losses are associated to the viscoelastic loading and unloading cycles. Such energy dissipation may produce a noticeable material heating (especially in the case of cycling loading) because of the very low thermal conductivity of polymers. On the contrary, plastic deformations are permanent. The viscoelastic and plastic processes strongly depend on temperature and their intensity increases as temperature is increased, especially above the T_g-temperature of the polymer.

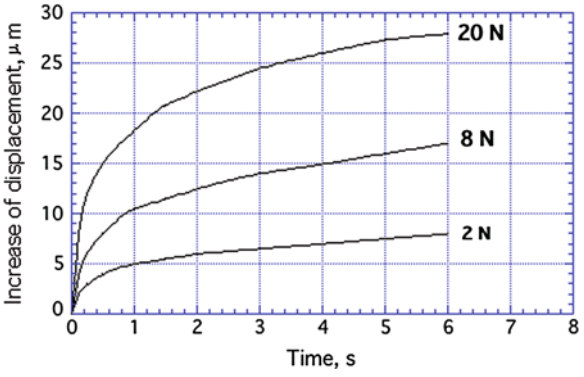

Fig. 1.5 Variation of load line displacement as a function of time for a polypropylene sphere pressed against a plane (modified from [5])

1.1.3 Elastic-Plastic and Fully Plastic Contacts

If a material behaves in a ductile manner, the applied contact force may induce a localized plastic deformation if the equivalent stress at the most critical point reaches the *uniaxial yield stress* of the material, indicated with σ_Y. In such a case, the contact is no longer elastic but *elastic-plastic* [6]. In the case of conformal contact, yield starts at the surface first, possibly at the edges. In case of non-conformal contact, yield starts first at the depth z_m, when τ_{Max} reaches the *shear yield stress* τ_Y given by $\sigma_Y/2$ (following the Tresca yield criterion). This means that subsurface localized yield starts when the Hertzian pressure, p_{max}, becomes equal to $1.61\sigma_Y$ for a point contact, and to $1.67\sigma_Y$ for a line contact. The schematic of Fig. 1.6a illustrates the condition of elastic-plastic contact in the case of a sphere/plane contact (where the plane has a lower hardness than the sphere). By further increasing the applied load, the size of the plastic zone also increases. If the applied load is removed when the contact pressure is below a specific limit, the additional loads, of the same magnitude, which are possibly applied, give rise to elastic deformations only. This phenomenon is also defined by the term *elastic shakedown* and will be further considered in the next chapter.

If the applied load is high to the point that the plastic deformation reaches the surface, the contact becomes *fully plastic* (Fig. 1.6b). The increase in the plastic zone size from first yield is made difficult by the local stress triaxiality. Fully plastic contact is therefore achieved when the nominal pressure (p_0) reaches a critical value, called the *yield pressure* (p_Y) that is greater than the uniaxial yield stress. Typically, $p_Y = b\sigma_Y$, where b is a constant greater than unity, which takes into account the difficulty of the spreading of plasticity (it depends on the geometry of the contact, the applied load, and the materials properties). As an example, in the case of a sphere in contact with a plane, it was obtained that b = 2.8 using the slip-line field theory.

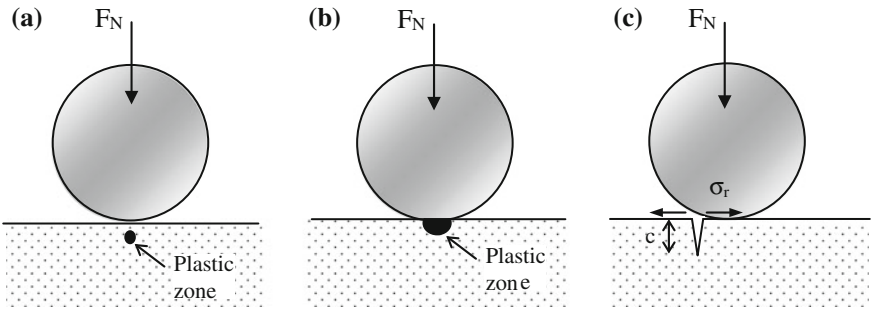

Fig. 1.6 Elasto-plastic (**a**) fully plastic (**b**) and brittle contact (**c**) between a sphere and a plane

1.1.4 Brittle Contact

If the yield strength of a material is high and its fracture toughness is low, the increase in the applied force may lead to brittle fracture at the contact surface. The contact, in this case, is *brittle* [7]. A brittle contact may take place when a micro-crack is present on the surface of one of the two mating bodies and such micro-crack is subject to a critical opening tensile stress. Consider, for example, Fig. 1.6c that displays a sphere in contact with a plane that contains a surface micro-crack, right in correspondence with the outer edge of the nominal area of contact. Brittle contact occurs if the local radial tensile stress (σ_r), given by the relation 1.7, is greater than the critical value, σ_F, given by the following relation:

$$\sigma_F = \frac{K_{IC}}{1.12\sqrt{\pi c}} \tag{1.10}$$

where c is the length of the micro-crack and K_{IC} is the fracture toughness of the material of the plane. Surface cracks formed in this way are often called *C-cracks*, since their shape is conical. A brittle contact may take place even in the absence of a surface micro-crack if the load is applied in a very concentrated point, as is the case of the angular contact of a ceramic particle. Figure 1.7 shows the mechanism proposed by Lawn and Swain [8]. The applied load may promote the formation of a

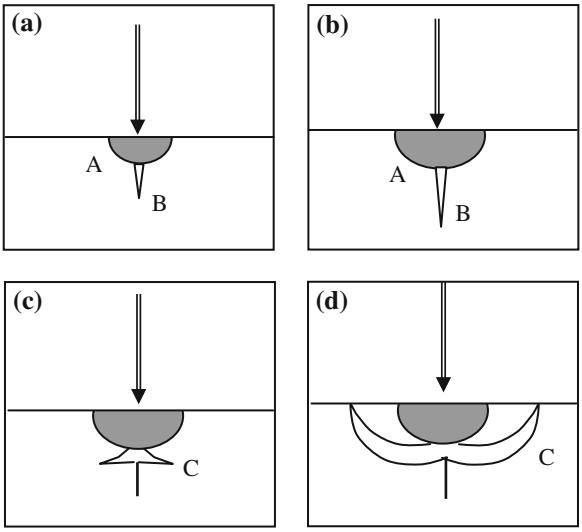

Fig. 1.7 Formation of radial and lateral cracks by the Lawn and Swain mechanism (modified from [8]). **a** Plastic deformation at the angular contact (*A*) with the formation and opening of a crack perpendicular to the surface (*B*); **b** propagation of the crack with increasing the applied load; **c** load removal followed by closing of the perpendicular crack and formation of new radial cracks (*C*) by the action of local residual stresses; **d** spalling when the radial cracks reach the surface

very concentrated plastic deformation due to the large and compressive local stresses. If it is increased over a critical value, a crack is formed, and it propagates towards the interior of the plane, perpendicularly to the surface. If the load is then removed, the local residual stresses may induce the formation of radial cracks that are parallel to the surface. Such cracks are formed if the initially applied load exceeds a critical value, which is proportional to the so-called *brittleness index*, given by the ratio H/K_{IC}, where H is the material hardness.

1.1.5 Materials Response to the Contact Stresses

The materials constituting the bodies in contact can differently deform to the applied stresses. The kind of response depends on the applied force, the properties of the materials' involved, and the geometry of the bodies in contact. The most important materials' properties are the elastic properties, such as the longitudinal modulus of elasticity and Poisson's ratio, the yield strength, hardness and fracture toughness. Table 1.2 shows such properties for some engineering metals, ceramics, and polymers.

Table 1.2 Selected mechanical properties of metals, ceramics and polymers

Materials	E (GPa)	σ_Y (MPa)	H (kg/mm^2)	K_{IC}, (MPa $m^{1/2}$)	E/σ_Y	E/H	E/K_{Ic} ($m^{-1/2}$)	H/K_I ($m^{-1/2}$)
Metals								
Ferritic-pearlitic steel	207	400	200	140	517.5	103.5	1478	14.3
Heat treated steel	207	1200	430	80	172.5	49	2600	54
Phosphorus bronze	110	350	120	70	314.3	91.7	1571	17.1
Aluminum alloys 6061 T6	70	275	100	25	254.5	71.3	2800	39.2
Ceramics								
Glass	72	3600	500	0.7	20	14.4	102,800	7140
Alumina	380	5230	1400	4	72.6	27.1	95,000	3500
Si nitride	310	4250	1800	4	72.9	17.2	77,500	4500
Si carbide	410	10,000	3000	4	41	13.7	102,500	7500
Polymers								
Nylon	3	60	12	3	50	25	1000	40
HD PE	1	30		2	33		500	
PMMA	3	70	30	1	95.5	28.6	3000	300

For ceramic materials σ_Y was obtained from hardness measurements using the relationship: $\sigma_Y = H/3$ [9]. Hardness, H, is defined by the resistance to surface penetration under a given force. In the text, it is always expressed in kg/mm^2, as obtained from Brinell or Vickers tests

The onset of the elastic-plastic contact is relatively easy in materials with high values of the ratio E/σ_Y (or E/H), which is called the *plasticity index*. A brittle contact, if surface micro-cracks are present, is rather easy in materials characterized by high values of the ratio E/K_{Ic}, while the brittle contact due to action of a concentrated point load is easy in materials with high values of the ratio σ_Y/K_{Ic} (both ratios are *brittleness indices*). In Table 1.2, the calculated indices of the selected materials are also reported. The data show that metallic materials would easily provide elastic-plastic (and fully plastic) contact, while a brittle contact is more common for ceramics. Polymeric materials do not provide easily both plastic and brittle contacts, although they are characterized by a low yield strength and a low toughness. Polymers may give a viscoelastic contact if the temperature is greater than T_g.

1.2 Surface Roughness

The concepts and relationships reported in the previous paragraph are strictly valid for ideally smooth surfaces. However, a close look at the real surface of a solid body (Fig. 1.8, first zoom) shows that it is not really smooth but consists of asperities and valleys of variable height, typically between 0.1 and a few micrometres. Furthermore, a surface is usually characterized by the presence of defects such as scratches, holes, cracks, and inclusions, having dimensions up to 10 μm or more. Thus each solid surface is characterized by a certain *roughness*.

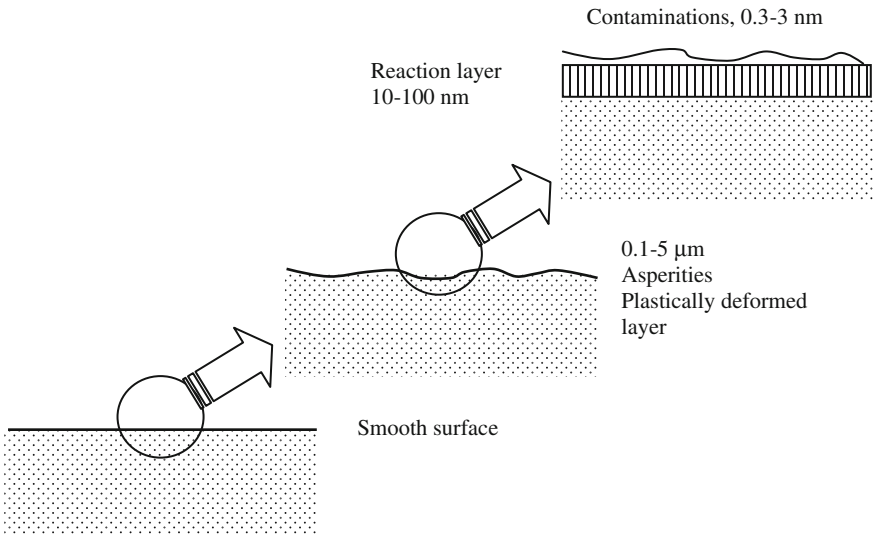

Fig. 1.8 Schematic of the microstructural characteristics of a material surface

The surface topography is quantified by different geometric parameters. One parameter, which is widely used in engineering applications, is the *average roughness* (also known as centre line average—CLA), which is indicated with R_a and is defined by the following relation:

$$R_a = \frac{1}{n}\sum_{i=1}^{n} |y_i| \tag{1.11}$$

where y_i are the distances from the mean line of n of points of the roughness profile, ideally obtained by a surface normal section (Fig. 1.9). Another important engineering parameter is the *root-mean-square roughness*. It is indicated with R_q (or with RMS) and is defined by the following relation:

$$R_q = \sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i^2}. \tag{1.12}$$

For a Gaussian distribution of the distances (or, heights) from the mean line, the ratio $R_q/R_a \approx 1.25$. The parameters R_a and R_q are typically expressed in micrometres and are often evaluated from data obtained using a *stylus profilometer*. With this method, a diamond stylus is moved along the surface and the vertical movement is measured. Several non-contact methods have been also developed, such as optical and capacitive methods, which are better suited for soft materials. As an example, Fig. 1.10 shows the roughness profiles of an AISI D2 tool steel surface polished to obtain two different finishing levels. As shown in Table 1.3, the surfaces with higher roughness are those obtained by foundry processes and hot working processes. After cold working (often in the presence of lubrication) and machining, surfaces with lower surface roughness are obtained (see, for example, Ref. [10]). With super finishing processes R_a-values down to 0.025 μm (or even less) may be achieved [11].

The height readings obtained using the stylus profilometer or other methods can be utilized to calculate several roughness parameters according to international standards. This is often done with the measuring instruments built in software. In addition to *amplitude parameters*, such as R_a and R_q (and relevant standard deviations), *spacing parameters*, such as the mean spacing of adjacent asperities, or *hybrid parameters* can be calculated [12]. The hybrid parameters are a combination

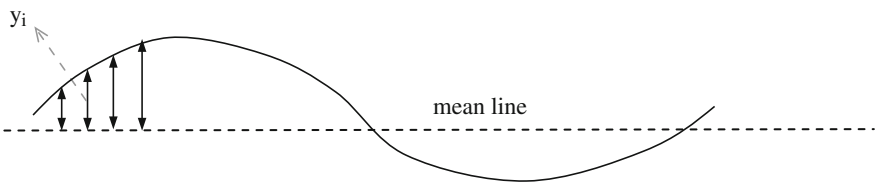

Fig. 1.9 Schematic of roughness profile for the evaluation of parameters R_a and R_q

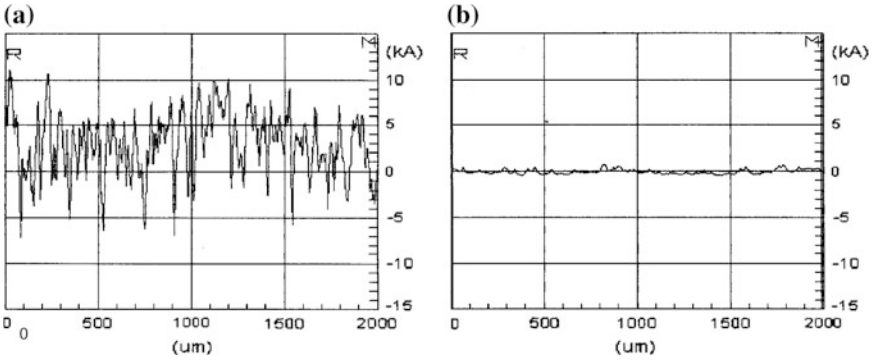

Fig. 1.10 Roughness profiles of an AISI D2 tool steel surface metallographically polished to obtain two different finishing levels: **a** R_a = 0.22 μm; **b** R_a = 0.02 μm (mirror finish). Please note that the x- and y-scales are very different, and the graphs thus give a distorted view of roughness. The asperities and valleys are not so sharp as they appear in the roughness profiles!

Table 1.3 Typical roughness values, R_a in micrometres, of surfaces obtained with different production methods

Foundry operations		Cold working processes	
Sand casting	8–25	Stamping	0.6–5
Shell casting	1.5–4	Rolling	0.16–2
Die casting	0.8–1.6	Drawing	0.5–3
Hot working processes		*Machining*	
Forging	4–15	Turning	0.4–3
Rolling	10–25	Grinding	0.1–1.2
Extrusion	0.8–4	Lapping	0.05–0.4

of amplitude and spacing properties. An example is the *mean slope of the profile* that varies on average from 1° and 5° and is generally less than 10°. It may be evaluated from surface topography measurements data by calculating all slopes between each two successive points of the profile and then making their average. If obtained in this way, it is also indicated with Δ_a. An experimental relationship between Δ_a and R_a is [13]:

$$\Delta_a = 0.108 \cdot R_a^{1.165} \tag{1.13}$$

where Δ_a is in degrees and R_a in micrometres. It has been obtained by stylus measurements on lapped and ground stainless steel plates. It has to be considered, however, that Eq. 1.13 is valid only for the given conditions. In fact, surfaces with different Δ_a-values can display similar amplitude parameters and, in general, one parameter is not able to characterize alone the topography of surfaces [14].

If the surface of an engineering material is observed even closer (Fig. 1.8, second zoom), the existence of different surface layers can be recognized:

(1) A deformed layer;
(2) A reacted layer;
(3) A contaminated layer;

The characteristics of the plastically deformed layer depend on the material and the manufacturing processes. Typically, in metals a work-hardened layer is formed with possible local microstructural modifications (such as the formation of white layers that will be considered in Sect. 6.1.1). The reacted layer is formed spontaneously due to exposure to the surrounding environment. In metals, an oxide layer is formed after exposure in air. In ferrous alloys, a mixture of Fe_3O_4 and Fe_2O_3 is observed at the top surface layer. In addition, FeO may be present in an intermediate layer above the bulk. A very thin and compact layer of chromium oxide, Cr_2O_3, covers stainless steel. On the surface of aluminium alloys a thin layer of amorphous Al_2O_3 oxide is present, possibly covered by a thicker and porous layer of hydrate oxide. On the surface of copper two oxide layers may be present: an innermost Cu_2O covered by CuO. The surface oxides have a typical thickness of 10 nm. They are well bonded to the metal if the ratio between their specific volume and that of the underlying metal is greater than 1. This happens in most metals such as iron, aluminium, copper, nickel, cobalt, zinc, chromium, molybdenum (and their alloys). In titanium, however, this ratio is less than 1. Surface oxides may be also present on non-oxide ceramics such as carbides or nitrides. In polymers, no oxides are present on their surface.

The outermost contaminated layer is made up of water vapour, hydrocarbons, and gases and has a typical thickness of 2 nm. The contaminants are mainly physically adsorbed on the surface through rather weak van der Waals forces. The adsorption of hydrocarbons is particularly important if the component is operating close to lubricated machinery, since oils tend to vaporize.

1.3 Real Area of Contact

When two surfaces are brought into contact only few asperities actually touch each other, as shown schematically in Fig. 1.11. The *real area of contact* (A_r) is therefore given by the sum of the individual areas (A_i) that form at each contact spot:

$$A_r = \sum_{i=1}^{N} A_i \tag{1.14}$$

where N is the number of contacting asperities. A_r is therefore smaller than the nominal area of contact, A_n (it is typically 10^{-2}–10^{-6} A_n).

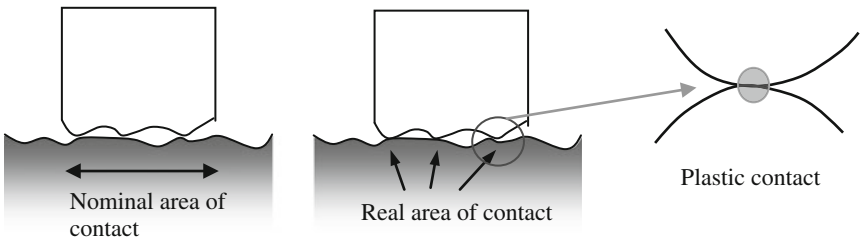

Fig. 1.11 Definition of nominal and real area of contact and lay out of the junction between two plastic asperity contact

The contact at the asperities may be elastic, plastic, or mixed. Greenwood and Williamson proposed a simplified model assuming the contact between spherically shaped asperities with the same radius of curvature and that follow a Gaussian height distribution [15]. They proposed the following index:

$$\Psi = \frac{E^*}{H}\sqrt{\frac{\sigma_s}{R_s}} \tag{1.15}$$

where $E^* = E'/2$, H is the hardness of the softest material in contact, σ_s is the composite standard deviation of the asperity height distribution, and R_s is the composite asperity radius ($1/R_s = 1/R_1 + 1/R_2$). Both σ_s and R_s can be obtained from profilometry data. Greenwood and Williamson obtained that when $\Psi < 0.61$ the elastic contact at the asperity tips dominates, whereas when $\Psi > 1$ the plastic contact dominates. Therefore, the kind of contact depends on the plasticity index and the roughness characteristics of the surfaces. With reference to the data listed in Table 1.2, it can be then argued that in metals the contact at the asperities is almost always plastic since the plasticity index is quite high. On the contrary, in ceramics or polymers the asperity contacts may be prevailing elastic if their surface roughness is sufficiently low. In practice, when engineering polymers and ceramics are involved, the contact at the asperities may be assumed to be mixed.

A simple relation for A_r can be obtained in the case of plastic contacts at the asperities (Fig. 1.11). At equilibrium:

$$F_N = \sum_{i=1}^{N} p_Y A_i = p_Y A_r \tag{1.16}$$

where N is the number of asperity junctions. Then

$$A_r = \frac{F_N}{p_Y} \tag{1.17}$$

The yield pressure (p_Y) is given by $b\sigma_Y$ as described in Sect. 1.1.3, and b depends on the geometry of the asperities. To a first approximation $p_Y = H$, where

H is the hardness of the material or, better, its microhardness. In the case of elastic contacts a relation for estimating A_r can be obtained by the Greenwood and Williamson model:

$$A_r \cong \frac{3.2F_N}{E^* \sqrt{\frac{\sigma_s}{R_s}}} \tag{1.18}$$

In either plastic or elastic contacts, A_r is thus independent from A_n and it is proportional to the normal load, F_N. In plastic contacts, A_r decreases as the hardness of the softest mating material is increased, whereas in elastic contacts, A_r decreases as E* is increased and the composite roughness is decreased (in particular, as the radius if the asperity tips is increased and the scatter in the height distribution is decreased). It is argued that as normal load is increased, the number N of asperity contacts is increased as well, whereas the average contact size is almost independent from load. In fact, as load is increased the size of pre-existing contact spots increases, but new contacts of smaller size also form. N can be estimated by assuming that each junction is circular in shape, with a mean radius r:

$$N = \frac{A_r}{\pi r^2} = \frac{F_N}{p_Y \pi r^2} \tag{1.19}$$

This relation is valid if the nominal pressure is low (roughly less than $\sigma_Y/2$). In fact, if the applied load becomes so high that A_r tends to A_n, the average size of the contact spots increases and N decreases (at the limit, when $A_r = A_n$ only one macro junction is formed).

The experimental evaluation of r is quite difficult. In the case of metals, typical values reported in the literature are of the order of 10^{-5}–10^{-6} m, and it is common that r is inversely proportional to p_Y [16].

1.4 Adhesion Between Surfaces in Contact

If two bodies are brought into contact with an applied normal force, a force may be required to pull the surfaces apart after removal of the normal force (Fig. 1.12a). This force is called *adhesion force*, since the phenomenon of adhesion at the asperity contacts is responsible for its appearance. As an example, Fig. 1.12b shows the measured force between a bioskin probe and different surfaces. The maximum negative force can be taken as the adhesion force.

The adhesion between two surfaces may be due to mechanical, chemical, or physical interactions at the areas where the materials are in intimate contact, possibly favoured by local intense plastic deformations (Fig. 1.11). In the contact of engineering surfaces, the adhesion at these *junctions* is mainly due to the formation

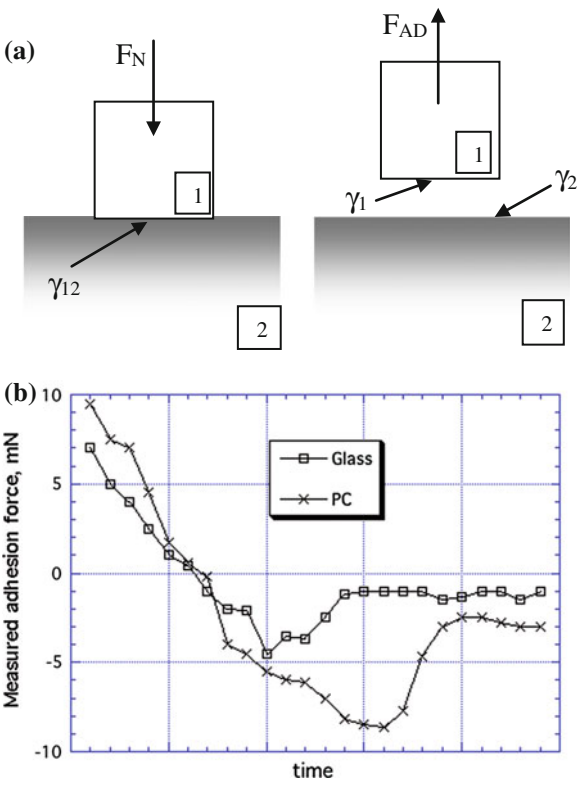

Fig. 1.12 **a** Adhesion between two bodies in contact and definition of adhesion force, F_{AD}. **b** Adhesion test between a bioskin probe and two different surfaces. The probe was firstly pushed against the surfaces with a force of 1 N. Then it was retracted and the force between the probe and the surface was recorded. More negative force shows that the probe was more stuck to the surface due to the adhesion (courtesy of Duvefelt and Olofsson)

of rather weak *van der Waals bonds* and also *hydrogen bonds* in polymers. They involve surface atoms that have free unsaturated bonds, or dipole-dipole interactions between polar molecules. Strong interactions, characterized by chemical bonding or interdiffusion phenomena, can occur during the deposition of thin coatings on various substrates, as discussed in Sect. 7.5.2.

A theoretical evaluation of the adhesion forces is quite difficult. Because of this, the thermodynamic concept of the *work of adhesion* per unit area has been introduced. It is usually indicated with W_{12} (where the subscripts refer to the two materials in contact) and represents the energy that must be theoretically supplied to separate two surfaces in contact. It is defined as the following:

$$W_{12} = \gamma_1 + \gamma_2 - \gamma_{12} \tag{1.20}$$

where γ_1 and γ_2 are the *surface energies* of the two bodies (more precisely, the surface energies of the interfaces with the surrounding environment) and γ_{12} is the surface energy of the interface that the two bodies form when they are in contact. Surface energies vary between 1 and 3 J/m^2 for clean metals, between 0.1 and 0.5 for ceramics and are lower than 0.1 J/m^2 for polymers [17].

Table 1.4 Values of surface energy for different materials, the data relate to clean surfaces and are taken from Refs. [17, 18]

Metal	γ (J/m^2)	Ceramic	γ (J/m^2)	Polymer	γ (J/m^2)
Fe	1.5	Al_2O_3	0.8	HDPE	0.035
Cu	1.1	ZrO_2	0.53	PMMA	0.045
Al	0.9	TiC	0.9	PA 6	0.05
Ni	1.7	ZrC	0.6	PVC	0.045
Ag	0.9			PTFE	0.018
Pb	0.45				
Cr	1				

In Table 1.4, typical values of the surface energy for some materials are listed. In the case of metals, the reported values refer to clean surfaces but with the presence of an unavoidable (albeit thin) native oxide layer. The possible presence of organic contaminants on the surfaces strongly affects their surface energy values and, in general, tends to reduce them.

Since the determination of γ_{12} is quite difficult, the work of adhesion is better estimated by the following relationship:

$$W_{12} = c(\gamma_1 + \gamma_2) \tag{1.21}$$

where c is a constant that is 1 for the contact between identical materials, and decreases as the *tribological compatibility* between the materials is decreased. The definition of tribological compatibility is not simple. Following Rabinowicz [18], two metals can be considered *compatible* when their phase diagram shows that they have high mutual solubility (>1 %) and are capable of forming intermetallic compounds. Two metals are *partially compatible* if they show a limited mutual solid solubility, between 0.1 and 1 %. They are *partially incompatible* if their mutual solid solubility is less than 0.1 %. Finally, two metals are *incompatible* if their mutual solubility is negligible. Following this approach, Rabinowicz has determined the compatibility chart shown in Fig. 1.13. With reference to experimental data, the *compatibility parameter*, c, is then set to 1 for identical metals, 0.5 for compatible metals, 0.32 for partially compatible metals, 0.2 for partially incompatible metals and 0.12 for incompatible metals. In the contact between ceramics, c can be set to 0.6 for compatible ceramics (such as two oxides or two nitrides), and 0.36 for incompatible ceramics. In the contact between polymers, c typically ranges from 0.8 to 0.95 [19]. The contacts between metals and ceramics, metals and polymers and ceramics and polymers can be assumed to be tribologically incompatible, and c can be set to 0.12 in every case.

In real contacts the adhesion force, F_{AD}, is expected to be proportional to the product of the work of adhesion and the real area of contact, A_r, since the van der Waals bonds form only at the asperity contacts. We may thus write: $F_{AD} \approx W_{12} A_r$. Considering Eq. 1.17 we can thus obtain an expression for the so-called *adhesion*

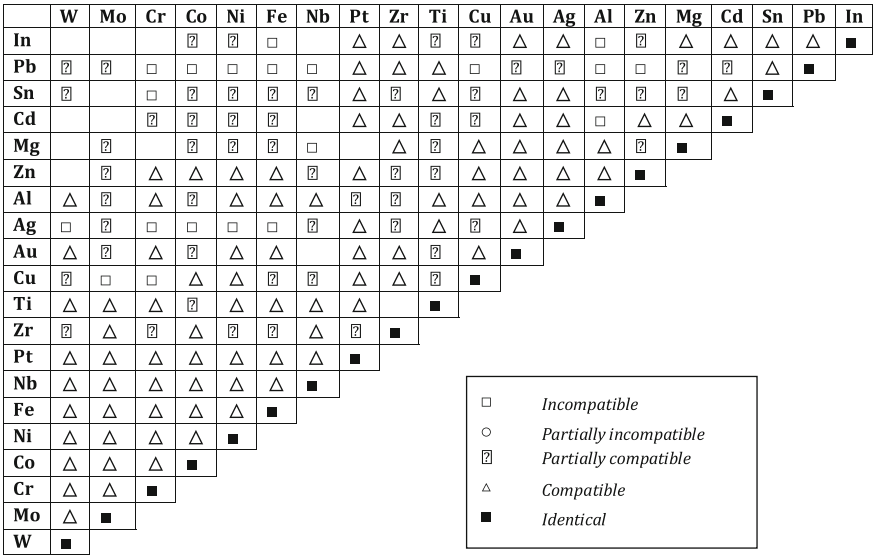

Fig. 1.13 Compatibility chart for metals (modified from [18])

coefficient, given by the ratio F_{AD}/F_N, where F_N is the load applied to establish the contact:

$$\frac{F_{AD}}{F_N} \propto \frac{W_{12} \cdot A_r}{p_Y \cdot A_r} \cong \frac{W_{12}}{H} \tag{1.22}$$

where H is the hardness of the softest material in contact (if the two materials are different). In Table 1.5, the ratio W_{12}/H is calculated for some material pairs. It can be noted that iron and silver are incompatible and the relative W_{12}/H-value is lower than that corresponding to the Fe–Fe pair. However, iron and lead are also incompatible, but their W_{12}/H-ratio is quite high. This is because of the low hardness of lead that gives rise to a large real area of contact. The Fe-polymer

Table 1.5 Calculation of the W_{12}/H-values for some metal pairs. H is the hardness of the softer metal in contact

	c	W_{12} (J/m^2)	H (kg/mm^2)	W_{12}/H (10^{-7} m)
Fe-Fe	1	3	80	0.038
Fe-Ag	0.12	0.29	50	0.006
Fe-Pb	0.12	0.23	4	0.058
Cu-Cu	1	2.2	80	0.028
Fe-Cu	0.32	0.83	80	0.01
Fe-Polymer	0.12	<0.2	10	<0.002

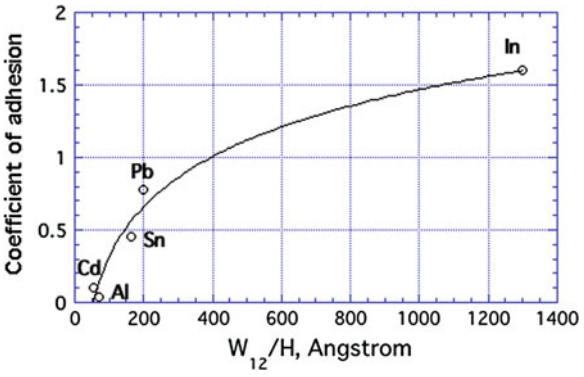

Fig. 1.14 Coefficient of adhesion versus W_{12}/H (H is the hardness of the soft metal) obtained by McFerlane and Tabor (modified from [18, 22])

contact is characterized by a low W_{12}/H-value because of the very low work of adhesion given by the low γ-values of polymers.

The experimental verification of relation 1.22 by tensile pulling on the interface is quite difficult for different reasons. First of all, it is common experience that adhesion is usually very poor in ordinary conditions. This is because surfaces are usually very contaminated and W_{12} is thus quite low. In order to carry out *pull-off force* measurements with available instruments, the surfaces have to be very clean. But despite this, two more and interrelated effects may render the experimental determination of F_{AD} very difficult. They are connected to the elastic deformations at the contact asperities. After unloading, the elastic part of the local deformation is released and the spring-back favours the detachment of the junctions. Such an effect is particularly pronounced when an elastic contact at the asperities is prevailing and when the dispersion in asperity heights (given by σ_s) is quite large [20]. The highest adhesion and the highest F_{AD}-values are thus achieved when the contacting materials are soft and the local elastic deformations are low, and when all asperities are of the same height and thus the junctions broke simultaneously during unloading [21].

In Fig. 1.14, the classic results of McFerlane and Tabor are shown [18, 22]. The graph shows the experimental coefficient of adhesion of a clean steel ball pressed for 1000 s against different clean planes made of soft metals. The experimental data highlight an increasing trend of the coefficient of adhesion with the ratio W_{12}/H, in good agreement with Eq. 1.22.

References

1. K.L. Johnson, *Contact Mechanics* (Cambridge University Press, Cambridge, 1985)
2. J.E. Shigley, C.R. Mischke, R.G. Budynas, *Mechanical Engineering Design*, 7th edn. (McGraw-Hill, New York, 2004)
3. G.W. Stachowiak, A.W. Batchelor, *Engineering Tribology* (Elsevier, New York, 1993)

4. M. Ciavarella, D.A. Hills, G. Monno, Influence of rounded edges on indentation by a flat punch. Proc. Inst. Mech. Eng. Part C: J. Mech. Eng. Sci. **212**, 319–328 (1998)
5. H. Czichos, in *Polymer Wear and Its Control*, ed by L.H. Lee (American Chemical Society, Washington, 1985), pp. 3–26
6. F.P. Bowden, D. Tabor, *Friction and Lubrication* (Methuen and Co. Ltd., London, 1956)
7. S.J. Sharp, M.F. Ashby, N.A. Fleck, Material response under static and sliding indentation loads. Acta Metall. Mater. **41**, 685–692 (1993)
8. B.R. Lawn, M.V. Swain, J. Mat. Sci. **10**, 113–122 (1975)
9. M.F. Ashby, R.H. Jones, *Engineering Materials* (Pergamon Press, Oxford, 1980)
10. S. Kalpakjian, *Manufacturing Processes for Engineering Materials* (Addison-Wesley, Reading, 1984)
11. G. Sroka, L. Winkelman, Superfinishing gears—the state of the art. Gear Technol. **20**, 6 (2003)
12. E.S. Gadelmawla, M.M. Koura, T.M.A. Maksoud, I.M. Elewa, H.H. Soliman, Roughness parameters. J. Mater. Process. Technol. **123**, 133–145 (2002)
13. D. Dawson et al., in *Polymer Wear and Its Control*, ed. by L.H. Lee (American Chemical Society, Washington, 1985), pp. 171–187
14. P.J. Blau, *Friction Science and Technology* (Marcel Dekker Inc., New York, 1996)
15. J.A. Greenwood, J.B.P. Williamson, Contact of nominally flat surfaces. Proc. R. Soc. (Lond) A, **295**, 295–300 (1966)
16. S.C. Lim, M.F. Ashby, Wear-mechanism maps. Acta Metall. Mater. **35**, 1–24 (1987)
17. D.S. Ramai, L.P. DeMeio, K.L. Mittal, *Fundamentals of Adhesion and Interfaces*, VSP (1995)
18. E. Rabinowicz, *Friction and Wear of Materials*, 2nd edn. (Wiley, New York, 1995)
19. K.H. Czichos, K.H. Habig, *Tribologie Handbuch*, *Reibung und Verlschleiss* (Vieweg, Braunschweig, 1992)
20. K.L. Johnson, Mechanics of adhesion. Tribol. Int. **31**, 413–418 (1998)
21. M. Benedetti, D. Bortoluzzi, M. Da Lio, V. Fontanari The influence of adhesion and sub-newton pull-off forces on the release of objects in outer space. Trans. ASME **128**, 828–840 (2006)
22. J.S. McFarlane, D. Tabor, Adhesion of solids on the influence of surface films. Proc. R. Soc. A **202**, 224–243 (1950)