

3

Contact Between Solid Surfaces

- 3.1 [Introduction](#)
- 3.2 [Hertzian Contacts](#)
 - Contact of Spheres • Contact of Two Cylinders with Axes Parallel • Contact of Two Cylinders with Inclined Axes • Contact of Bodies of Arbitrary Shape • Surface and Subsurface Stresses • Surface Traction and Sliding Contact • Cylinder Sliding Perpendicular to Its Axis and Sliding Spheres
- 3.3 [Non-Hertzian Contacts](#)
 - Flat Rigid Planar Punch • Flat Rigid Axisymmetric Punch • Indentation by an Angular Wedge • Indentation by a Cone
- 3.4 [Numerical Methods for Contact Mechanics](#)
 - Matrix Inversion Method • Variational Method • Finite Element Method
- 3.5 [Experimental Methods for Contact Mechanics](#)
 - Real and Nominal Area of Contact Measurement • Experimental Contact Stress Analysis
- 3.6 [Further Aspects](#)
 - Contact of Rough Surfaces • Loading Beyond the Elastic Limit • Repeated Contacts — The Role of Residual Stresses and Shakedown • Contacts on Layered Solids

John A. Williams
Cambridge University

Rob S. Dwyer-Joyce
The University of Sheffield

3.1 Introduction

When two solid surfaces are loaded together there will always be some distortion of each of them. Deformations may be purely elastic or may involve some additional plastic, and so permanent, changes in shape. Such deflections and modifications in the surface profiles of the components can be viewed at two different scales. Consider, for example, the contact between a heavily loaded roller and the inner and outer races in a rolling element bearing. The degree of flattening of the rollers can be expressed as a proportion of their radii, i.e., at a relatively *macroscopic* scale. On the other hand, since on the *microscale* no real surface, such as those of either the roller or the race, can be truly smooth, it follows that when these two solid bodies are pushed into contact they will touch initially at a discrete number of points or asperities. Some deformation of the material occurs on a very small scale at, or very close to, these areas of true contact. It is within these regions that the stresses are generated whose total effect is just to balance the applied load. Classical contact mechanics assumes the deforming materials to be isotropic and homogeneous; in principle, its results can be applied both to global contacts and to those between interacting asperities.

3.2 Hertzian Contacts

When any two curved bodies of different radii of curvature are brought into contact they will initially touch at either a point or along a line. With the application of the smallest load, elastic deformation enlarges these into contact areas across which the loads are distributed as pressures. The first analysis of this situation was presented by Heinrich Hertz in 1881 and is based on the following assumptions:

- i. The surfaces are continuous, smooth, nonconforming and frictionless,
- ii. The size of the contact area is small compared to the size of the bodies, i.e., the strains associated with the deformations are small,
- iii. Each solid can be considered to behave as an elastic half-space in the vicinity of the contact zone, and
- iv. The gap h between the undeformed surfaces can be approximated by an expression of the form

$$h = Ax^2 + By^2 \quad (3.1)$$

where x and y are orthogonal coordinates lying in the common tangent plane to the two surfaces.

Although strictly, idealization (iv) requires parabolic surface profiles, by implication Hertzian analysis is relevant to the contact of spheres, cylinders, and ellipsoids. Progress over the last century in contact mechanics can be seen as the gradual relaxation of the restrictions imposed by the original Hertzian treatment; for an exhaustive review, see Johnson (1985).

3.2.1 Contact of Spheres

If two elastic spheres 1 and 2 of radii R_1 and R_2 are pressed into contact with force P , as in [Figure 3.1\(a\)](#), then the resultant circular contact area has radius a such that

$$a = \{3PR/4E\}^{1/3} \quad (3.2)$$

where E^* is the contact modulus defined by

$$\frac{1}{E^*} = \frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2} \quad (3.3)$$

and R , the reduced radius of curvature, is related to those of the individual components by the relation

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} \quad (3.4)$$

Convex surfaces are taken to have positive radii of curvature; those of concave surfaces are negative.

The resulting pressure distribution $p(r)$ is semielliptical, i.e., of the form

$$p(r) = p_0 \left\{1 - r^2/a^2\right\}^{1/2}, \quad \text{where } r^2 = x^2 + y^2 \quad (3.5)$$

Such a distribution is shown in [Figure 3.1b](#) and is characteristic of Hertzian contacts. The maximum pressure p_0 which occurs on the axis of symmetry and the mean pressure p_m are given by

$$p_0 = \frac{3}{2} p_m = 3P/2\pi a^2 \quad (3.6)$$

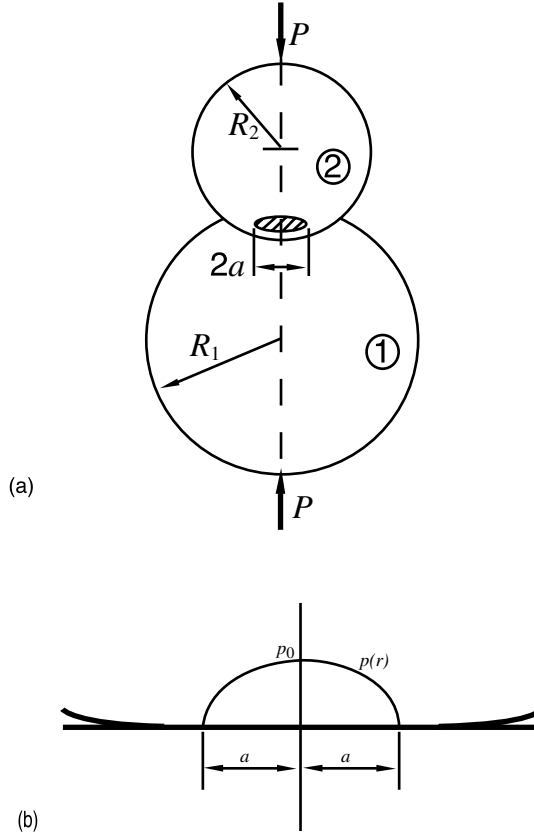


FIGURE 3.1 (a) Spheres in elastic contact; (b) the resulting semi-elliptical pressure distribution.

p_0 is sometimes known as the Hertz stress. Under this loading, the centers of the two spheres move together by the small displacement Δ where

$$\Delta = a^2/R = a\pi p_0/2E^* = \left\{9P^2/16RE^{*2}\right\}^{1/3} \quad (3.7)$$

If one of the loaded solids is, in fact, a plane surface then its effective radius becomes infinite so that the reduced radius of the contact is numerically equal to that of the opposing sphere.

3.2.2 Contact of Two Cylinders with Axes Parallel

If two circular cylinders with radii R_1 and R_2 are pressed together by a force per unit length of magnitude P with their axes parallel, as shown in Figure 3.2, then the contact patch will be of half-width b such that

$$b = \left\{2PR/\pi E^*\right\}^{1/2} \quad (3.8)$$

where R and E^* are the reduced radius of contact and the contact modulus defined by Equations 3.3 and 3.4, respectively. Contact pressure is again semielliptical such that

$$p(x) = p_0 \left\{1 - r^2/b^2\right\}^{1/2} \quad (3.9)$$

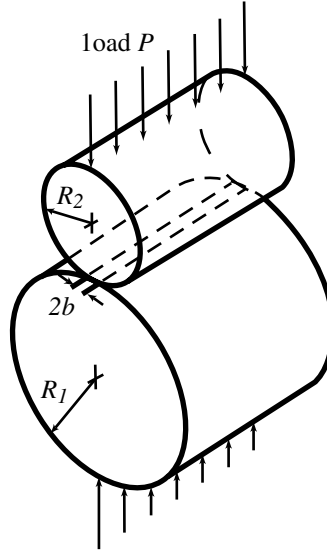


FIGURE 3.2 Parallel cylinders in elastic contact; the intensity of the loading is P per unit length.

where peak pressure

$$p_0 = \{PE^*/\pi R\}^{1/2} \quad (3.10)$$

and coordinate x is measured in a direction perpendicular to that of the cylinder axes.

The mean pressure p_m over the contact strip is equal to $P/2b$ and is given by

$$p_m = \pi p_0 / 4$$

The axes of the cylinders move together by a small distance Δ where

$$\Delta = (1 - \nu_1^2) \left[\ln(4R_1/b) - 1/2 \right] / E_1 + (1 - \nu_2^2) \left[\ln(4R_2/b) - 1/2 \right] / E_2 \quad (3.11)$$

Once again, if one of the surfaces is planar, then the value of R will be equal to the radius of curvature of the other.

3.2.3 Contact of Two Cylinders with Inclined Axes

Suppose the lower cylinder is of radius R_1 and the upper of R_2 , as shown in Figure 3.3. The cylinders touch at the point O but their axes are inclined at an angle θ . Ox_1y_1 is a set of Cartesian axes with Oy_1 along a generator of the lower cylinder, and Ox_2 a second similar set with Oy_2 along a generator of the upper cylinder.

Close to the origin we can approximate the circular section of each cylinder by a parabolic profile and so write that the separation h of the two solid surfaces at the point P, which has coordinates (x_1, y_1) in the first coordinate system or (x_2, y_2) in the second, is given by

$$h \approx \frac{x_1^2}{2R_1} + \frac{x_2^2}{2R_2} \quad (3.12)$$

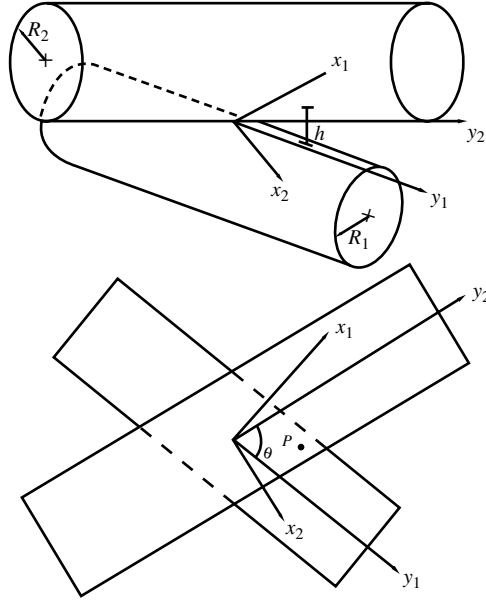


FIGURE 3.3 The geometry of two circular cylinders in skewed contact.

It is possible to choose a common set of axes (Oxy) , in which Ox is inclined to the Ox_1 -direction by the angle α , such that Equation 3.1 is recovered, provided that A and B satisfy the conditions

$$B - A = \frac{1}{2} \left\{ 1/R_1^2 + 1/R_2^2 + (2/R_1 R_2) \cos 2\theta \right\}^{1/2} \quad (3.13)$$

and

$$B + A = \frac{1}{2} \left\{ 1/R_1 + 1/R_2 \right\} \quad (3.14)$$

and α is given by the solution of

$$\frac{R_2}{R_1} \sin 2\alpha = \sin 2(\theta - \alpha) \quad (3.15)$$

Equation 3.1 can then be written as

$$h = \frac{x^2}{2R'} + \frac{y^2}{2R''} \quad (3.16)$$

by defining

$$R' = 1/2A \quad \text{and} \quad R'' = 1/2B \quad (3.17)$$

R' and R'' are known as the principal radii of *relative curvature*.

It is evident from Equation 3.16 that contours of constant gap h between the undeformed surfaces will be *ellipses*, the lengths of whose axes are in the ratio $(R'/R'')^{1/2}$.

When a normal load P is applied, the point of contact spreads into an elliptical area with semi-axes a and b , such that the eccentricity, i.e., the ratio b/a , is *independent* of the load and depends only on the ratio of R'/R'' . For mildly elliptical contacts, i.e., $A/B < 5$, then the ratio b/a is given by

$$b/a \approx (A/B)^{2/3} \quad (3.18)$$

An “equivalent” radius R_e can be defined as

$$R_e = (R'' \times R')^{1/2} = \frac{1}{2}(AB)^{-1/2} \quad (3.19)$$

and this used to estimate the contact area or Hertz stress by using the circular contact equations with R replaced by R_e . The approach of the bodies can be calculated using Equation 3.7 but with R replaced by $(AB)^{-1/2}$ rather than R_e .

If the two cylinders make contact with their axes *parallel* so that $\theta = 0$, it follows from Equations 3.13, 3.14, and 3.17 that

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

i.e. Equation 3.3 is recovered.

On the other hand, if the axes of the two cylinders are *perpendicular* to one another $\theta = 90^\circ$; then $R' = R_1$ and $R'' = R_2$. It follows that

$$h = \frac{x^2}{2R_1} + \frac{y^2}{2R_2} \quad (3.20)$$

Consequently, for the particular case of a pair of *equal* cylinders crossing at an angle of 90° , contours of constant surface separation close at the contact point will be circles.

3.2.4 Contact of Bodies of Arbitrary Shape

Suppose that the profile of each body closer to the origin can be expressed as

$$\left. \begin{aligned} z_1 &= \frac{x_1^2}{2R_1'} + \frac{y_1^2}{2R_1''} \\ z_2 &= -\left\{ \frac{x_2^2}{2R_2'} + \frac{y_2^2}{2R_2''} \right\} \end{aligned} \right\} \quad (3.21)$$

where the directions of the axes of each body are chosen to coincide with the principal curvatures of that body. In general the two sets of axes may be inclined to each other at an arbitrary angle θ , as shown in [Figure 3.4a](#).

The coordinates can be transformed to a common set of axes (Oxy) inclined at α to x_1 and β to x_2 , as shown. The gap h between the surfaces can then be written as

$$h = z_1 - z_2 = Ax^2 + By^2 + Cxy \quad (3.22)$$

where $C = 0.5\{(1/R_2') - (1/R_2'')\}\sin 2\beta - 0.5\{(1/R_1') - (1/R_1'')\}\sin 2\alpha$.

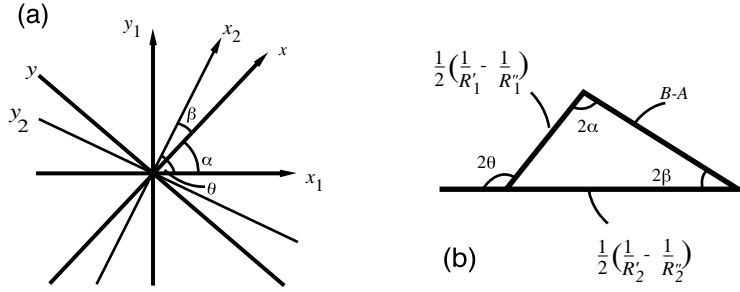


FIGURE 3.4 (a) Arbitrary curved bodies in contact; (b) geometric relations.

The condition that C should vanish, so that Equation 3.1 is recovered, is satisfied by the triangle shown in Figure 3.4b with the result that

$$\begin{aligned}
 B - A &= 0.5 \left\{ 1/R_1' - 1/R_1'' \right\} \sin 2\alpha + 0.5 \left\{ 1/R_2' - 1/R_2'' \right\} \sin 2\beta \\
 &= 0.5 \left\{ \left[1/R_1' - 1/R_1'' \right]^2 + \left[1/R_2' - 1/R_2'' \right]^2 + \left[1/R_1' - 1/R_1'' \right] \left[1/R_2' - 1/R_2'' \right] \cos 2\theta \right\}^{1/2}
 \end{aligned}$$

Finally,

$$A + B = 0.5 \left\{ 1/R_1' + 1/R_1'' + 1/R_2' + R_2'' \right\} \quad (3.23)$$

from which the values of A (= 1/2R') and B (= 1/2R'') can be found.

3.2.5 Surface and Subsurface Stresses

Stresses that develop in conditions of nonconformal contacts can be among the most severe in any mechanical device; the situation of a Hertzian contact is of particular interest because the most heavily loaded elements of material, and thus the site of initial plastic yielding, lies not at the surface but a small distance beneath it. (This is in contrast to the situation when the curvature of the surfaces is not continuous, for example in circumstances in which one of the bodies has the profile of a wedge or cone, when the site of maximum stress lies adjacent to its apex [see Section 3.3]). Closed form analytical solutions for subsurface stresses are known for only a limited number of special cases.

Nominal Point Contacts

In the case of two spheres pressed into contact (or the equivalent geometry of two circular cylinders with their axes inclined perpendicularly to each other), the pressure within the contact patch (i.e., when $r \leq a$) is given by Equation 3.5 and the surface stresses (i.e., at $z = 0$) are given in polar coordinates by the expressions

$$\left. \begin{aligned}
 \sigma_{rr}/p_0 &= \frac{1-2\nu}{3} \left(a^2/r^2 \right) \left\{ 1 - \left(1 - r^2/a^2 \right)^{3/2} \right\} - \left(1 - r^2/a^2 \right)^{1/2} \\
 \sigma_{\theta\theta}/p_0 &= -\frac{1-2\nu}{3} \left(a^2/r^2 \right) \left\{ 1 - \left(1 - r^2/a^2 \right)^{3/2} \right\} - 2\nu \left(1 - r^2/a^2 \right)^{1/2} \\
 \sigma_{zz}/p_0 &= -p(r)/p_0 = -\left(1 - r^2/a^2 \right)^{1/2}
 \end{aligned} \right\} \quad (3.24)$$

Outside the contact patch ($r \geq a$)

$$\sigma_n/p_0 = -\sigma_{\theta\theta}/p_0 = (1-2\nu)a^2/3r^2 \quad (3.25)$$

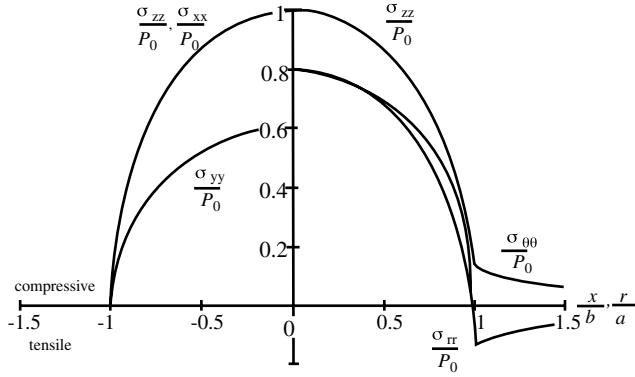


FIGURE 3.5 Stress distributions at the surface of a line contact (left-hand side) and point contact (right-hand side); Poisson's ratio $\nu = 0.3$.

The form of these surface stress distributions is illustrated in Figure 3.5 for the case $\nu = 0.3$. The radial stress is tensile outside the loaded circle and has its maximum value at the edge of the loaded zone when $r = a$. This is the site of the largest tensile stress occurring anywhere in the material and is held responsible for the ring cracks often observed in the contact of brittle materials. At the center of the contact the radial stress is compressive and of magnitude $(1 - 2\nu)p_0/2$. Thus for an incompressible material, with $\nu = 0.5$, the stress at the origin is hydrostatic.

Within the material stresses along the axis of symmetry ($r = 0$) are given by

$$\left. \begin{aligned} \sigma_{rr}/p_0 = \sigma_{\theta\theta}/p_0 &= -(1+\nu) \left\{ 1 - (z/a) \arctan(a/z) \right\} - \frac{1}{2} \left(1 + z^2/a^2 \right)^{-1} \\ \sigma_{zz}/p_0 &= - \left(1 + z^2/a^2 \right)^{-1} \end{aligned} \right\} \quad (3.26)$$

This stress distribution for a half space with $\nu = 0.3$ is shown in Figure 3.6a. Along the z -axis, σ_{rr} , $\sigma_{\theta\theta}$, and σ_{zz} are principal stresses and so the principal shear stress τ_1 is of magnitude $\frac{1}{2} |\sigma_{zz} - \sigma_{rr}|$ and this is also shown in the figure: its maximum value lies at a depth of $0.48a$ below the surface and here $\tau_{\max}/p_0 = 0.31$.

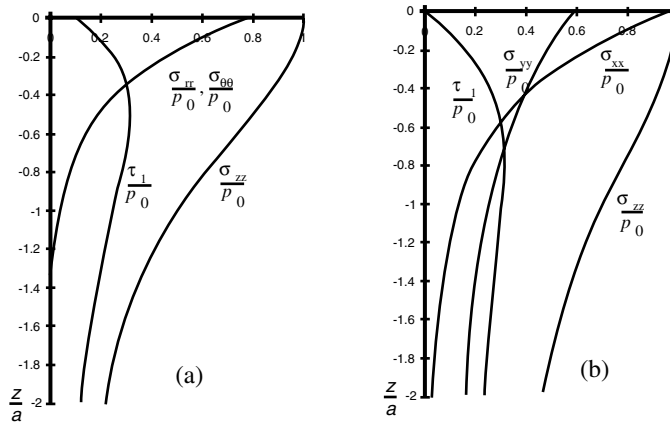


FIGURE 3.6 (a) Elastic contact of spheres; subsurface stresses along the axis of symmetry; (b) elastic contact of cylinders; subsurface stresses along the axis of symmetry.

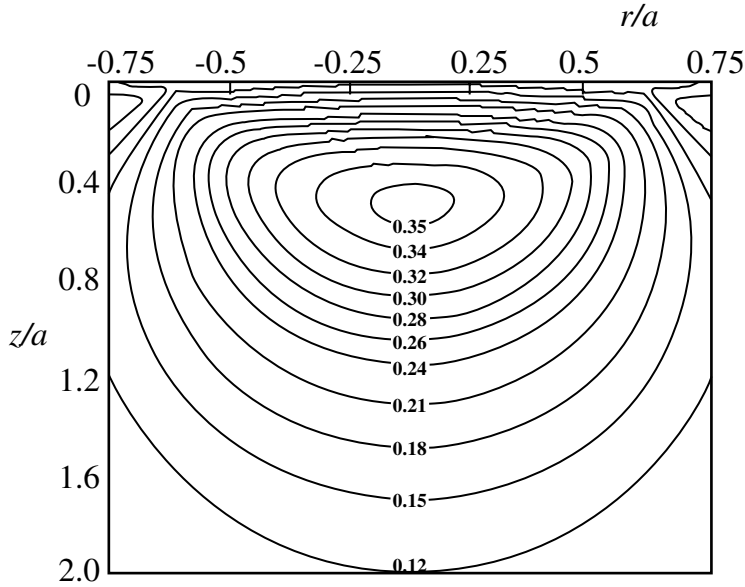


FIGURE 3.7 Contours of maximum shear stress normalized by Hertz stress p_0 , beneath nominal circular point contact of radius a in material with $\nu = 0.3$.

TABLE 3.1 Normalized Maximum Shear Stress τ_{\max}/p_0 and Its Location for Materials with Various Poisson's Ratios

Poisson's Ratio	Material	τ_{\max}/p_0	z/a
0.2	Glass	0.335	0.45
0.30	Steel	0.310	0.48
0.33	Aluminium	0.303	0.50
0.5	Rubber	0.267	0.55

Since this is the greatest shear stress in the material, its position is the likely site of the initiation of plasticity. The value of τ_{\max}/p_0 and its location is somewhat influenced by the value of Poisson's ratio (Table 3.1).

Away from the axis of symmetry the stresses can be computed in a number of ways, such as those described by Hamilton and Goodman (1966), Hamilton (1983), and Sackfield and Hills (1983); these are discussed in Hills, Nowell, and Sackfield (1993). Contours of maximum shear stress, plotted as τ_{\max}/p_0 , for such a nominal circular point contact are shown in Figure 3.7.

Nominal Line Contacts

If the axes of the cylinders are taken to lie parallel with the y -axis, then surface stresses are given by the equations

$$\begin{aligned}
 p(x) &= p_0 \left\{ 1 - x^2/b^2 \right\}^{1/2} \quad \text{where peak pressure } p_0 = \left\{ PE^*/\pi R \right\}^{1/2} \\
 &\quad \left. \begin{aligned} &\text{together with } \sigma_{yy}/p_0 = -2\nu \left(1 - x^2/b^2 \right)^{1/2} \\ &\text{and } \sigma_{zz}/p_0 = -p(x)/p_0 = -\left(1 - x^2/b^2 \right)^{1/2} \end{aligned} \right\} \quad (3.27)
 \end{aligned}$$

Along the line of loading symmetry, (i.e., the z -axis)

$$\left. \begin{aligned} \sigma_{xx}/p_0 &= -\left(1+2z^2/b^2\right)\left(1+z^2/b^2\right)^{-1/2} - 2z/b \\ \sigma_{zz}/p_0 &= -\left(1+z^2/b^2\right)^{-1/2} \\ \text{and } \tau_1/p_0 &= z/b - \left(z^2/b^2\right)\left(1-z^2/b^2\right)^{-1/2} \end{aligned} \right\} \quad (3.28)$$

The value of τ_{\max}/p_0 is now 0.30 and occurs now at a depth of $z = 0.78b$ as illustrated in [Figure 3.6b](#). The stresses σ_{xx} , σ_{zz} , and τ_{\max} are all independent of the value of ν although not so the third, out-of-plane, stress σ_{yy} ; for plane strain conditions

$$\left. \begin{aligned} \sigma_{yy} &= \nu(\sigma_{xx} + \sigma_{zz}) \\ \text{so that } \sigma_{yy}/p_0 &= -2\nu \left\{ \left(1+z^2/b^2\right)^{1/2} + z/b \right\} \end{aligned} \right\} \quad (3.29)$$

Away from the axis of symmetry the stress components can be evaluated using the method of complex potentials (Muskhelishvili, 1949), although solutions to the problem in a real form have been published by several authors (Beeching and Nichols, 1948; Poritsky, 1950; Smith and Liu, 1953; Sackfield and Hills 1983). McEwen (1948) expresses the stress at a general point (x,z) in terms of m and n defined by

$$\left. \begin{aligned} m^2 &= 0.5 \left\{ \left[\left(b^2 - x^2 + z^2\right)^2 + 4x^2 z^2 \right]^{1/2} + \left(b^2 - x^2 + z^2\right) \right\} \\ n^2 &= 0.5 \left\{ \left[\left(b^2 - x^2 + z^2\right)^2 + 4x^2 z^2 \right]^{1/2} - \left(b^2 - x^2 + z^2\right) \right\} \end{aligned} \right\} \quad (3.30)$$

where the signs of m and n are the same as the signs of x and z , respectively; then

$$\left. \begin{aligned} \sigma_{xx}/p_0 &= m \left\{ 1 + \left(z^2 + n^2\right) / \left(m^2 + n^2\right) \right\} / b + 2z/b \\ \sigma_{zz}/p_0 &= -m \left\{ 1 - \left(z^2 - n^2\right) / \left(m^2 + n^2\right) \right\} / b \\ \tau_{xz}/p_0 &= n \left\{ \left(m^2 - z^2\right) / \left(m^2 + n^2\right) \right\} / b \end{aligned} \right\} \quad (3.31)$$

Contours of maximum shear stress under such a line contact in a material with Poisson's ratio 0.3 are shown in [Figure 3.8](#).

Elliptical Contacts

Where the contact geometry leads to an elliptical contact patch, it is more difficult to find explicit solutions for the internal stress field. Thomas and Hoersch (1930) examined the problem on the axis of symmetry and later Fessler and Ollerton (1957) and Ollerton and Morley (1963) determined in closed form values of σ_{zz} and τ_1 . A more general solution is due to Sackfield and Hills (1983), and [Figure 3.9](#), taken from this, illustrates the change in location of the point of severest stress along the axis $x = y = 0$ in terms of the eccentricity parameter k equal to the ratio of b/a , the minor to major axes of the contact patch.

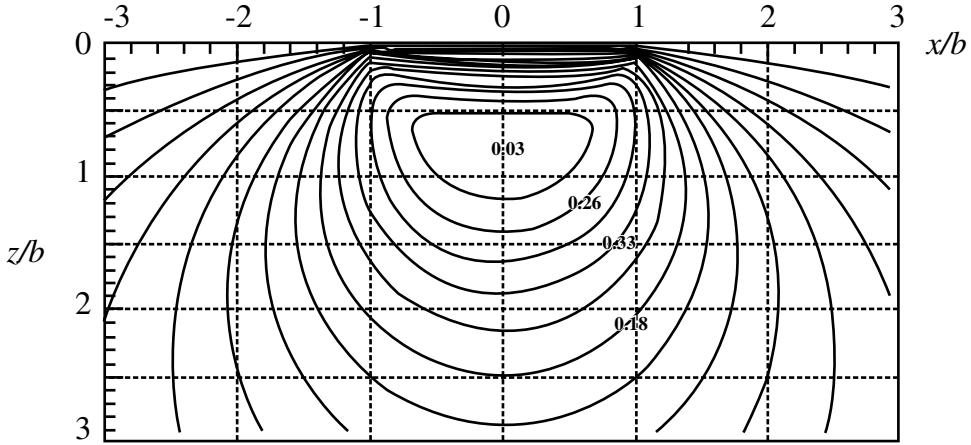


FIGURE 3.8 Contours of maximum shear stress normalized by Hertz stress p_0 , beneath nominal line contact in material with $\nu = 0.3$.

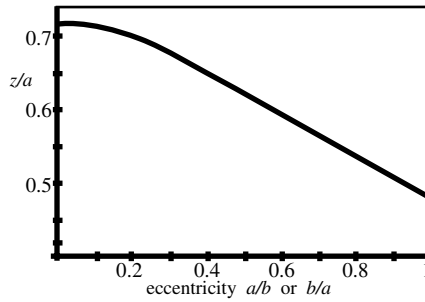


FIGURE 3.9 Effect of ellipticity of contact patch measured as ratio a/b or b/a on location of point of severest stress.

3.2.6 Surface Traction and Sliding Contact

Relative Motion of Surfaces: Rolling, Sliding, and Spin

Two solid bodies in contact are said to roll together if there is a difference in the components of their angular velocities measured along an axis parallel to their common tangent plane. Consider two cylinders 1 and 2 touching along a common generator, as shown in Figure 3.10a. If the angular velocity ω_1 is not equal to ω_2 , both in magnitude and sense (i.e., clockwise or counterclockwise), the two cylinders are in rolling contact. The *angular velocity of roll* has a magnitude which is equal to the difference between ω_1 and ω_2 . In this figure, the Oxy plane is the common tangent plane; if there is any difference in the components of the linear velocities of the two points in contact at O within this plane, then, as well as rolling on one another, the surfaces also have a relative *sliding velocity*. In Figure 3.10a this sliding velocity is equal in magnitude to $U_1 - U_2$ and thus equal to $|R_1\omega_1 - R_2\omega_2|$.

The term *rolling velocity* \bar{U} must be carefully defined, especially if the point of contact between the solid surfaces is not at rest, in other words if one or both of the centers of the cylinders has a linear motion. If, however, bodies 1 and 2 are rotating about *fixed* centers then the rolling velocity $\bar{U} = \frac{1}{2} |U_1 + U_2|$.

In the case of a more general three-dimensional contact, e.g., that of two spheres in contact Figure 3.10b, or of a sphere on a plane, there can also be a difference in the components of the angular velocities of the two bodies in the direction of the common normal, i.e., at right angles to the common tangent plane. This represents a *spinning* motion, and in Figure 3.10b is of magnitude $|\omega_{1z} - \omega_{2z}|$.

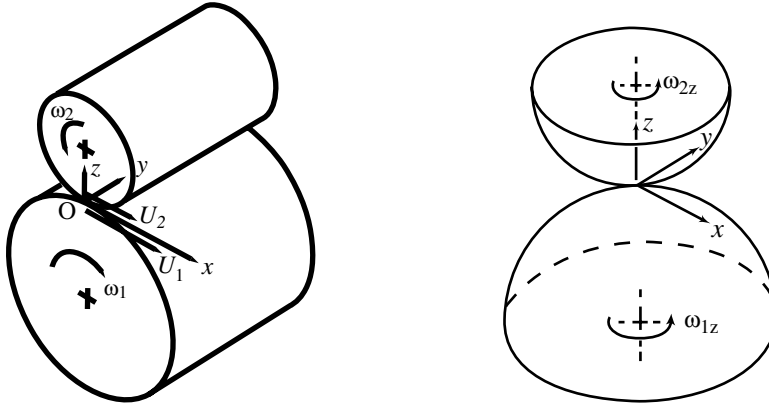


FIGURE 3.10 (a) Two cylinders 1 and 2 in contact with their axes parallel: there can be rolling and sliding at the point of contact; (b) in the three-dimensional case, Body 2 can also spin relative to Body 1.

In the effectively two-dimensional case of Figure 3.10a, when the cylinders are pressed together by a normal load, the contact will spread over a finite area which can be determined for this nonconformal contact by Hertz theory (Equations 3.8 through 3.11). Once account is taken of the fact that the contact spreads into a strip which has a finite width, rather than occurring along a line, the specification of “sliding” is not quite so straightforward. This is because, in the presence of friction (or *traction*) between the surfaces, some pairs of points in contact at the interface may “slip” relative to one another while others “stick.” A difference between the tangential surface strains in the two bodies in the region of “stick” can lead to a small degree of overall relative movement, which is known as *creep*.

3.2.7 Cylinder Sliding Perpendicular to Its Axis and Sliding Spheres

Gross Sliding

The simplest case is one in which there is gross sliding of the interface. If the contact is frictionless, then the contact stresses, both surface and subsurface, are unaffected by the sliding motion. However, if there is an imposed sliding force, this affects both the distribution of stresses and the contact geometry. If the two solids have the same elastic constants, then any tangential force transmitted between them gives rise to equal and opposite normal displacements of any point on the interface. Thus any “warping” of one surface conforms exactly to that of the other and does not disturb the distribution of normal pressure. The size and shape of the contact are then fixed by the profiles of the two surfaces, and the normal load is independent of the traction. With solids of differing elastic properties this is no longer the case, and tangential tractions do interact with the normal pressure so that the contact area and pressure distribution is no longer symmetrical (Buffler, 1959). However, these effects are generally small, especially if the coefficient of friction is small so that in practical cases it can usually be assumed that the stresses and deformations due to normal and tangential loadings are independent of one another and can be superposed to find resultant values.

Assuming that Amontons’ law holds at every point at the interface, i.e., that the local surface tangential stress $q(x)$ is given by

$$|q(x)| = \mu |p(x)|$$

where $p(x)$ is given by Equation 3.27.

Hence

$$|q(x)| = \frac{2\pi P}{\pi b^2} \left\{ 1 - x^2/b^2 \right\}^{1/2} \quad (3.32)$$

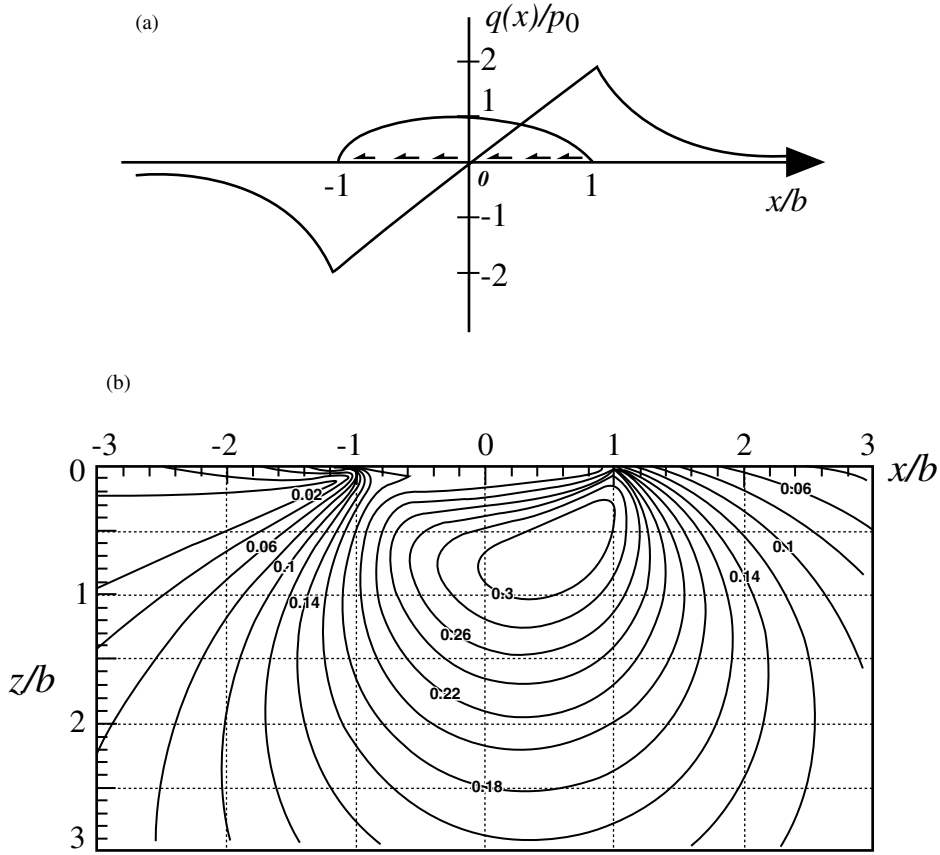


FIGURE 3.11 (a) Surface stresses due to friction traction at a nominal line contact; (b) contours of principal shear stress beneath a sliding elastic line contact with $\mu = 0.2$.

The stress components generated by the application of this form of loading have been evaluated by several workers (McEwen, 1949; Poritsky, 1950; Smith and Liu, 1953; Sackfield and Hills, 1983b).

At the surface ($z = 0$) the stress component parallel to the surface reduces to the expressions

$$\left. \begin{aligned} \sigma_{xx}/\mu p_0 &= -2x/b \quad \text{for } |x| \leq b \\ \text{and } \sigma_{xx}/\mu p_0 &= -2 \left\{ x/b \pm \left(x^2/b^2 - 1 \right)^{1/2} \right\} \quad \text{for } |x| > b \end{aligned} \right\} \quad (3.33)$$

These are illustrated in Figure 3.11a. Whatever the coefficient of friction, the maximum resultant surface tensile stress in sliding occurs at the trailing edge of the contact and is of magnitude $2\mu p_0$. The application of a surface traction has two effects: first, it disrupts the symmetry of the subsurface stress distribution of Figure 3.8, and, second, it increases the level of stress at any given depth and gradually moves the location of the most highly stressed element of material toward the loaded surface, as illustrated in Figure 3.11b.

In the case of sphere sliding on a flat, again neglecting any interference between normal pressure and tangential loading arising from a difference in elastic constants of the two solids, the distribution of traction is given by

$$|q(r)| = \frac{3\mu P}{2\pi a^3} \left(1 - r^2/a^2 \right)^{1/2} \quad (3.34)$$

where a is the contact patch radius. Explicit equations for subsurface stresses are available in Goodman and Hamilton (1966), Hamilton (1983), and Sackfield and Hills (1983). Once again, as the coefficient of friction increases, so the region of maximum shear stress both moves from a subsurface location toward the surface and becomes more intense. The site of potential yield occurs at the surface when μ exceeds about 0.3. The normal contact of elastic spheres introduces a radial tension at $r = a$ of magnitude $(1 - 2\nu)/3 \approx 0.3p_0$. The effect of the tangential traction is to add to the tension at one side of the contact and to subtract from it at the other. The maximum tension, which occurs at the point $(-a, 0)$ rises to $0.5p_0$ and $1.0p_0$ for $\mu = 0.25$ and $\mu = 0.5$, respectively. This result is comparable to that from the two-dimensional case. The analysis has been extended to elliptical contacts by Bryant and Keer (1982) and Sackfield and Hills (1983b), who show that the contact pressure for first yield is almost independent of the shape of the contact ellipse.

Incipient Sliding — Microslip

The application of a tangential force which is less than the limiting force of friction will not give rise to sliding motion but will induce interfacial frictional tractions and elements of relative shearing or microslip at the interface. In general, over the center of the contact region, normal stresses may be sufficient to prevent relative motion at the interface so that the surfaces are effectively within a “stick” region. However, at the periphery of the contact, there may be relative surface displacements as a result of interfacial “slip.” Difficulty arises in the solution of such problems because the division of the contact into “stick” and “slip” regions is not known *ab initio*: a useful first step is to assume that slip occurs nowhere in the contact zone. Slip is then most likely in those regions where the local tangential stress, so found, exceeds its possible limiting values. In the case of two cylinders in line contact, slip occurs first at the edges of the contact: over the central region, the greater values of normal pressure inhibit any tangential motion. In these circumstances the interfacial shear stress $q(x)$ is given by

$$q(x) = \mu p_0 \left\{ \left(1 - x^2/b^2\right)^{1/2} - \left(1 - x^2/c^2\right)^{1/2} \right\} \quad (3.35)$$

where b is the half width of the whole contact, and c the half-width of the central sticking region. This distribution is illustrated in [Figure 3.12a](#). The width of the central zone, i.e., the value of dimension c , is dependent on the magnitude of the applied tangential force Q ,

$$c/b = \left\{ 1 - Q/\mu P \right\}^{1/2} \quad (3.36)$$

The physical behavior of the contact is summarized by this equation. If the normal load P is kept constant while the tangential force on the junction Q is varied, as soon as Q increases from zero, microslip begins at the edges of the contact. As Q grows, so this slipping region spreads inward according to Equation 3.36. When Q approaches the limiting value of μP the sticking region has degenerated to a line at $x = 0$. When Q reaches magnitude μP gross sliding at the interface is initiated. The change in the relative size of the sticking region with the increase in the applied tangential load is illustrated in [Figure 3.12b](#).

A similar argument in the case of a sphere sliding on a flat leads to a central circular “sticking” region of radius c , whose value can be found from the applied tangential force Q by the equation

$$c/a = \left\{ 1 - Q/\mu P \right\}^{1/3} \quad (3.37)$$

Contact analyses of this sort (Johnson, 1985) are particularly relevant when the applied forces vary or oscillate with time. Microslip at the interface between two surfaces which are subjected to vibration, often in combination with corrosion, produces the characteristic surface damage known as “fretting.” In components which also carry a high steady stress, fretting can lead to premature failure by fatigue.

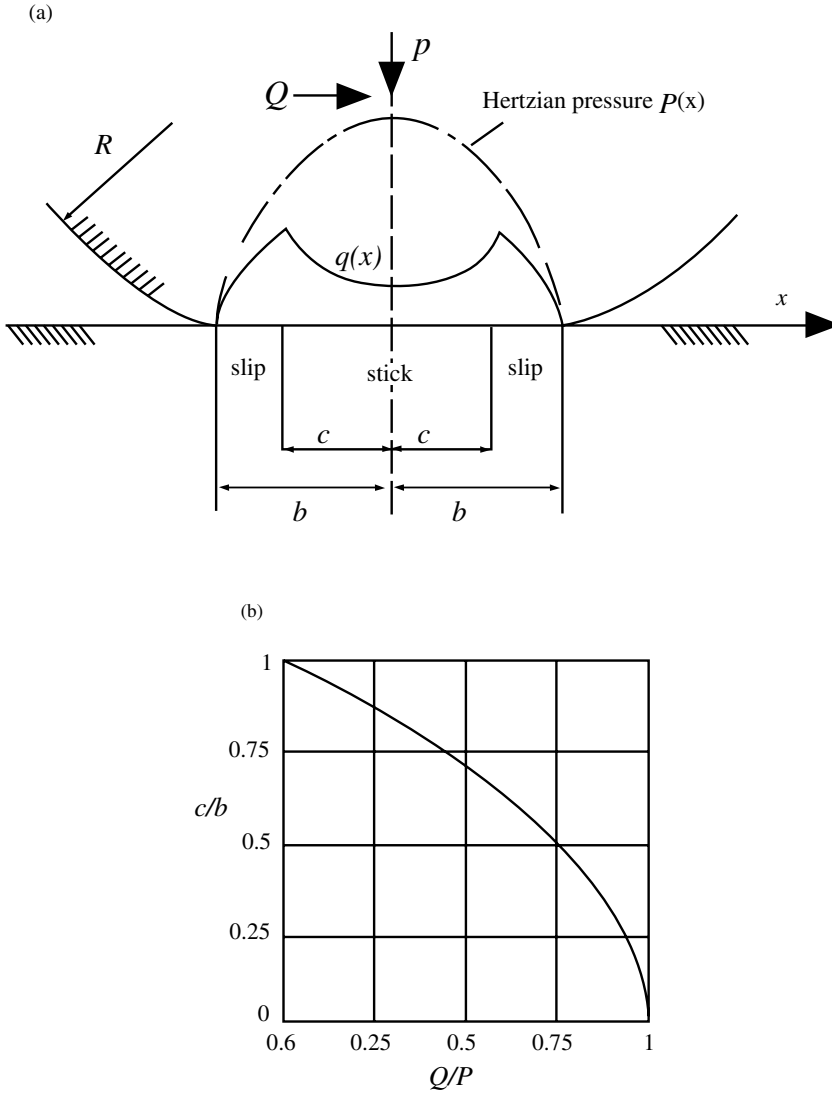


FIGURE 3.12 (a) Contact between cylinders with parallel axes: surface traction due to the application of traction load of intensity Q ; (b) reduction of central sticking region with applied tangential load for two cylinders in nominal line contact.

Free and Tractive Rolling

The term *free rolling* is usually used for conditions in which there is no net traction, while *tractive rolling* is reserved for those situations in which such tangential forces exist. Tractive rolling is typified by the driving or braking wheels of a vehicle, where the interfacial tangential or frictional force acts to change the velocity of the mass in motion.

If there is no transmission of torque between two cylinders in line contact, then the distribution of shear stress at the interface is given by the equation

$$\frac{q(x)}{p_0} = \frac{\beta}{\pi} \left\{ 1 - x^2/b^2 \right\}^{1/2} \log \left| (b+x)/(b-x) \right| \quad (3.38)$$

TABLE 3.2 Values of the Parameter b for Various Materials; Body 2 is Steel

Body 1	β
Rubber	~ 0
Perspex/Plexiglas	0.19
Glass	0.21
Duralumin	0.12
Cast iron	-0.24

in which β is a factor which depends on the elastic constants of the two materials, where

$$\beta = \frac{1}{2} \left\{ \frac{(1-2\nu_1)(1+\nu_1)/E_1 - (1-2\nu_2)(1+\nu_2)/E_2}{(1-\nu_1^2)/E_1 + (1-\nu_2^2)/E_2} \right\} \quad (3.39)$$

If the two solids are different, then $|\beta| > 0$. The values of β for a number of fairly common materials chosen as Body 1 are given in Table 3.2; in each case Body 2 is steel.

Tractive Rolling

When either a driving or braking torque is transmitted through a rolling contact, the contact zone between the solids will, in general, contain both sticking and slipping zones, but, unlike the situation in sliding contact, the sticking zone will not be symmetrically disposed about the line joining the geometric centers of the cylinders. Consider a case in which two elastically similar cylinders, loaded together by a normal force of intensity of P per unit length, are rotating so that a torque is transmitted across the interface. Referring to Figure 3.13, we suppose that cylinder 1 is the *driving* roller and cylinder 2 the *driven*.

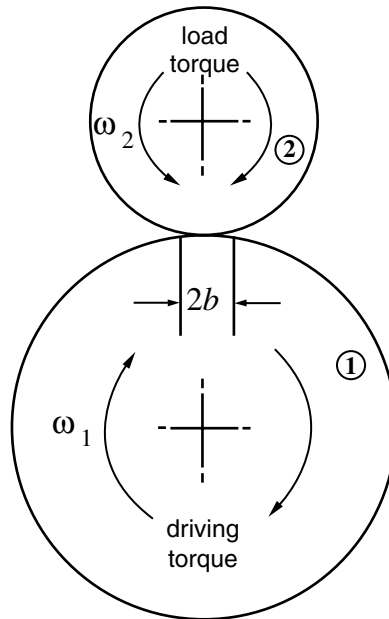


FIGURE 3.13 Rolling contacts; the lower cylinder 1 drives the upper cylinder 2 through frictional effects at the interface which is of width $2b$: in free rolling there is no load torque applied to the driven cylinder.

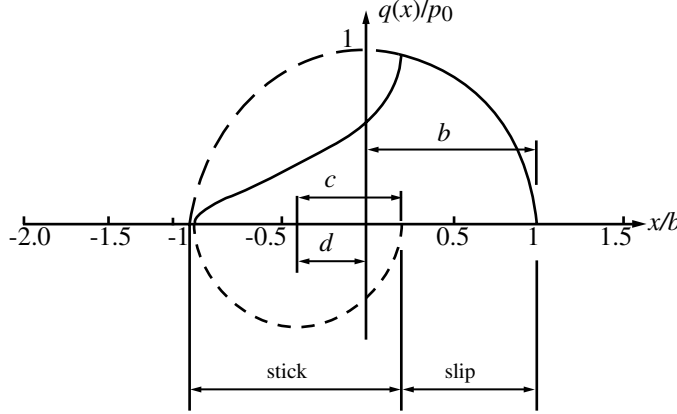


FIGURE 3.14 Distribution of surface traction under tractive rolling of similar cylinders plotted for the case $c/b = 0.6$ and $Q/\mu P = 0.64$.

By analogy with the stationary situation, the contact area will be divided into regions of “stick” and “slip”. In the static case there was a central sticking region, extending from $-c < x < +c$, with side slip zones disposed symmetrically. Under tractive rolling the pattern is displaced so that there is no slipping at the leading edge, as illustrated in Figure 3.14. The distribution of surface tractive stresses is then given by

$$q(x) = \mu p_0 \left\{ \left(1 - x^2/b^2 \right) \right\}^{1/2} - \mu p_0 (c/d) \left\{ 1 - (x+d)^2/c^2 \right\}^{1/2} \quad \text{for } -b < x < c-d$$

and

$$q(x) = \mu p_0 \left\{ 1 - x^2/b^2 \right\}^{1/2} \quad \text{for } c-d < x < b \quad (3.40)$$

Dimensions c and d are related to the applied normal load P and tangential load Q by the relations

$$c/d = 1 - d/b = \left\{ 1 - Q/\mu P \right\}^{1/2}$$

3.3. Non-Hertzian Contacts

The preceding section has described analysis for the contact of curved bodies where the gap between the undeformed surfaces can be expressed as a quadratic (i.e., by Equation 3.1). In this section, elastic contacts between some bodies of other shapes are considered.

3.3.1. Flat Rigid Planar Punch

Consider a flat-ended punch, of width $2b$ and of infinite length in the y -direction, pressed onto an elastic half-space with a force per unit length P , as shown in Figure 3.15. The surface of the punch is assumed frictionless so there is no restriction to lateral displacement of material under the punch.

Contact occurs across the width of the punch, $2b$. A theoretically infinite value of the contact stress is reached at the shoulders of the punch. The pressure distribution is given by:

$$p(x) = \frac{P}{\pi} \left\{ b^2 - x^2 \right\}^{-1/2} \quad (3.41)$$

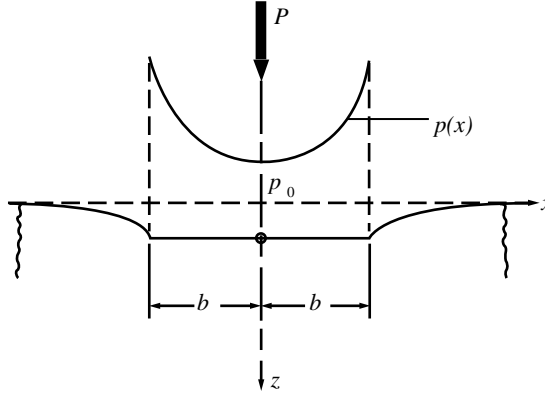


FIGURE 3.15 Indentation by a flat-ended rigid planar punch.

The mean pressure over the contact strip is given by:

$$p_m = P/2b = \pi p_0/2 \quad (3.42)$$

As in all line contact problems, the deflection can only be presented relative to some datum. The normal deflection u_z of the surface, outside the contact region, is given by:

$$u_z = \delta - \frac{2(1-\nu^2)P}{\pi E} \ln \left[\left(x/b \right) + \left\{ x^2/b^2 - 1 \right\}^{1/2} \right] \quad (3.43)$$

where δ is the normal deflection at an arbitrary datum point.

3.3.2 Flat Rigid Axisymmetric Punch

In this case the punch has a circular section of radius a . It is pressed onto an elastic half-space with a force P . Again the surface of the punch is assumed frictionless.

Contact occurs across a circle of radius a and the resulting pressure distribution (Figure 3.16) is (Timoshenko and Goodier, 1951)

$$p(r) = \frac{P}{2\pi a} \left\{ a^2 - r^2 \right\}^{-1/2} \quad (3.44)$$

The mean pressure p_m over the contact circle is given by

$$p_m = P/\pi a^2 = \frac{p_0}{2} \quad (3.45)$$

The penetration Δ of the punch is given by

$$\Delta = P(1-\nu^2)/2Ea \quad (3.46)$$

3.3.3 Indentation by an Angular Wedge

If a two-dimensional wedge of semi-angle α is pressed onto a frictionless elastic half-space with a force per unit length, P then contact is made over a rectangular region of semi-width b , such that

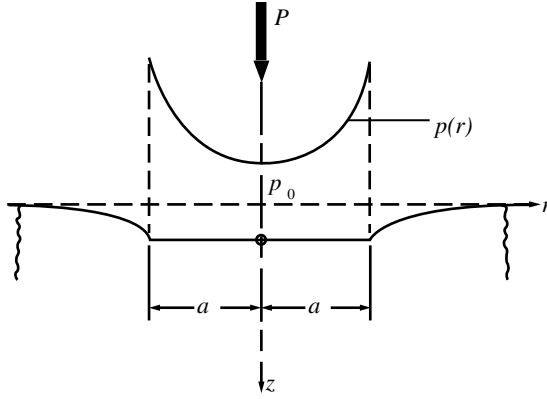


FIGURE 3.16 Indentation by a flat-ended rigid axisymmetric punch.

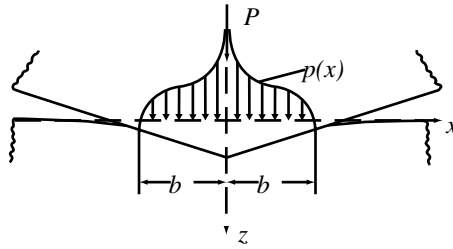


FIGURE 3.17 Indentation by a two-dimensional blunt wedge.

$$b = P / (E^* \cot \alpha) \quad (3.47)$$

This time the pressure distribution (Figure 3.17) has a singularity at the wedge apex and is given by

$$p(x) = \frac{E^* \cot \alpha}{\pi} \cosh^{-1}(b/x) \quad (3.48)$$

Equations 3.47 and 3.48 are appropriate for blunt wedges only (α is close to 90°) such that deformations are small and linear elastic theory is applicable.

3.3.4 Indentation by a Cone

If a cone of semi-angle α is pressed onto a frictionless elastic half-space with a force per unit length equal to P , then contact is made over a circular region of radius a ,

$$a = \left\{ \frac{2P}{\pi E^* \cot \alpha} \right\}^{1/2} \quad (3.49)$$

The pressure distribution again has a singularity at the cone apex and is given by (Love 1939)

$$p(r) = \frac{E^* \cot \alpha}{2} \cosh^{-1}(a/r) \quad (3.50)$$

The extent of the penetration Δ of the cone is given by

$$\Delta = P/4E^* a \quad (3.51)$$

Equations 3.49 to 3.51 are appropriate for large angle cones only (α is close to 90°) such that deformations are small and linear elastic theory is applicable.

3.4 Numerical Methods for Contact Mechanics

Many contact problems will not fall into the preceding categories (Hertzian elastic or frictionless cones/punches) and require modeling by some numerical means. Common problems of this sort are those that involve friction and partial slip, complex geometry, nonlinearity of elastic properties, or plasticity.

3.4.1 Matrix Inversion Method

For contact problems where the half-space assumption holds (i.e., the surfaces deform under contact loading as if they were elastic half-spaces), then a surface discretization method may be employed. The basic principle involves dividing the surface into a number of discrete elements. Some form of pressure distribution is then assumed to act on each element (Bentall and Johnson, 1967). Figure 3.18 shows three possible unit pressure distributions for a two-dimensional (i.e., line contact) problem; point loading, uniform pressure, or an overlapping triangular pressure distribution.

The normal displacement at a node point i due to the loading at node point j is then expressed from the appropriate elasticity equation for the chosen unit loading. Hills et al. (1993) provide a useful summary of these traction elements. For example, for the uniform pressure elements (Figure 3.18b) the normal displacement at element i by a pressure applied at node j is given by

$$(u_z)_i = -\frac{(1-\nu^2)}{\pi E} p_i \left\{ (x+a) \ln \left[(x+a)/a \right]^2 - (x-a) \ln \left[(x-a)/a \right]^2 \right\} + \text{constant} \quad (3.52)$$

where $2a$ is the spacing between nodal points and x is the distance between nodes i and j . The constant is usually removed by considering the displacements relative to some arbitrary datum.

The displacement $(u_z)_i$ of the point i due to the summation of the loadings at each of the points j is then expressed by a matrix equation of the form

$$(u_z)_i = \sum_{j=1}^n C_{ij} p_j \quad (3.53)$$

The influence coefficient matrix C_{ij} is obtained from the appropriate expressions for the deflection caused by the chosen elemental pressure distribution. For the uniform pressure element example, this gives

$$C_{ij}(k) = \frac{1}{2\pi} \left\{ (k+1) \ln(k+1)^2 - (k-1) \ln(k-1)^2 \right\} + \text{constant} \quad (3.54)$$

where k is the number of elements between nodes i and j (i.e., $k = i - j$).

In many cases the region of contact will not initially be known and an interactive procedure is adopted. A region of contact is guessed and a set of displacements u determined from the interpenetration of the

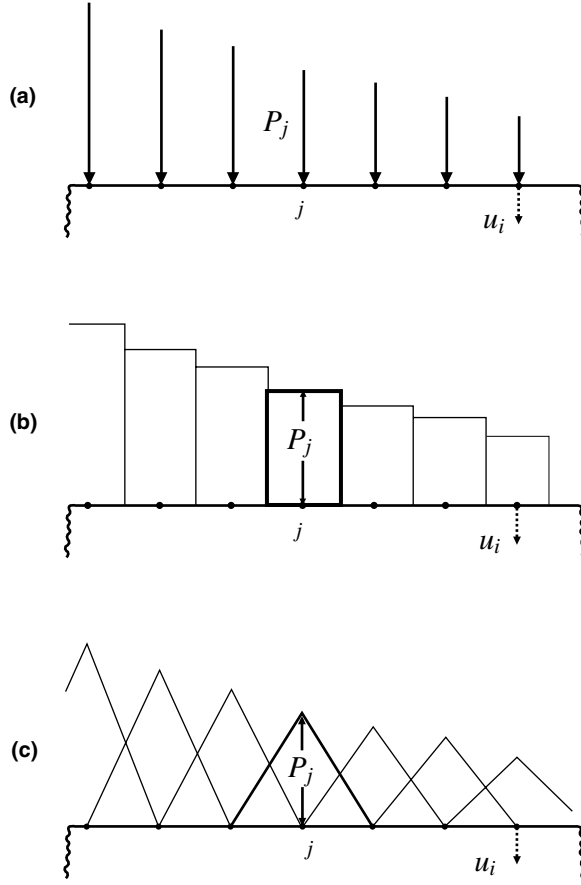


FIGURE 3.18 Discretization of a contact surface and the application of element loading; (a) point loading at each node; (b) uniform pressure acting on each element; (c) overlapping triangular pressure distributions.

two bodies. Some numerical matrix inversion procedure then gives the corresponding pressures p . The set of pressures will contain some negative values; these are removed from the assumed contact area for the next iteration. A solution is found when the pressure is positive at all points within the region of contact and zero elsewhere. The total load is then found from the sum of the elemental pressures

$$P = A \sum_{j=1}^n p_j \quad (3.55)$$

where A is a constant depending on the pressure element chosen (for the uniform pressure element example A is simply the width of the element, and the load is expressed per unit contact length). This method has been used extensively for determining the contact of rough surfaces (Webster and Sayles, 1986; Snidle and Evans, 1994). Real digitized surface roughness data are recorded, and this method used to determine real areas of contact when the profile is loaded against some counterface. Figure 3.19 shows the results from one such model; a rough surface has been loaded against a smooth cylindrical body. The resulting displaced surfaces and pressure distributions are shown.

Once a surface distribution is known it is straightforward to determine the subsurface stress field (again assuming linear elastic behavior). The stress at any location is the summation of the stresses caused by each of the surface pressure elements. Again for the line contact uniform pressure element example,

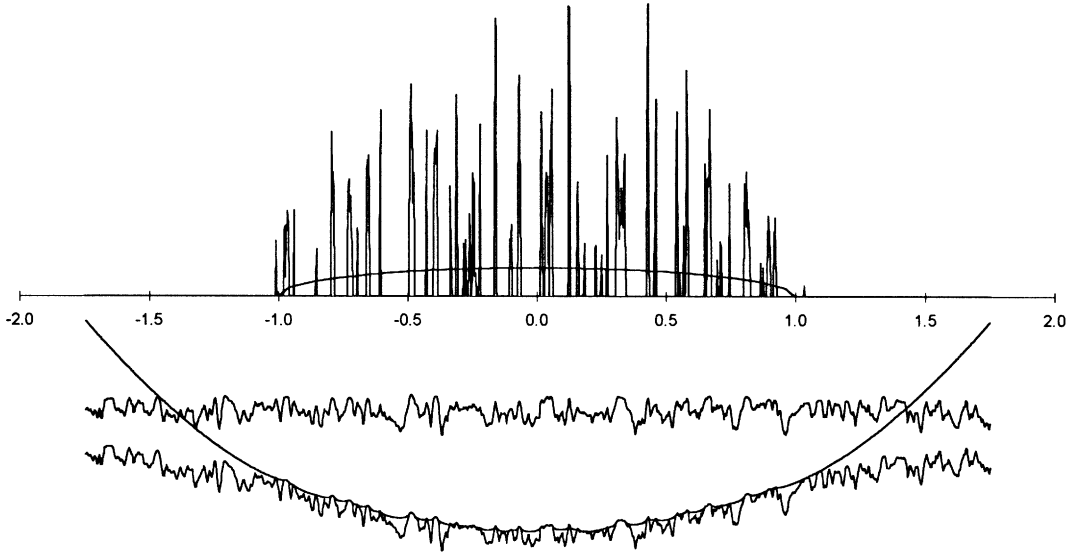


FIGURE 3.19 Numerical model of the contact between a planar rough surface and a smooth cylinder; (a) the predicted pressure distribution is shown compared to the Hertz smooth surface solution; (b) original and deformed surfaces.

the stresses at point (x, z) beneath a single pressure element, p , located at the surface ($z = 0$) and $-a < x < +a$ are (Johnson 1985)

$$\begin{aligned}\sigma_{xx} &= -\frac{p}{2\pi} \left\{ 2(\theta_1 - \theta_2) + (\sin 2\theta_1 - \sin 2\theta_2) \right\} \\ \sigma_{zz} &= -\frac{p}{2\pi} \left\{ 2(\theta_1 - \theta_2) - (\sin 2\theta_1 - \sin 2\theta_2) \right\} \\ \tau_{xz} &= \frac{p}{2\pi} \left\{ \cos 2\theta_1 - \cos 2\theta_2 \right\} \\ \tan \theta_1 &= z/(x-a) \quad \text{and} \quad \tan \theta_2 = z/(x+a)\end{aligned}\tag{3.56}$$

Figure 3.20 shows a contour plot of the von Mises stress beneath the contact between a rough sphere and a smooth half-space (Lee and Ren 1994) determined using this numerical approach. The stress peaks associated with the roughness tend to be restricted to the near-surface region: the deeper stresses are similar to those for a smooth contact.

3.4.2 Variational Method

It is possible to use the approach described above in cases where friction is present at the interface. In these cases both shear and normal tractions result in surface displacements. However, this further complicates the solution method since the regions of stick and slip within the contact must also be determined. The matrix inversion method is then very intensive in computational time.

An alternative approach which is more useful in these cases is the variational method (Kalker, 1990). The variational principle states that the true stress field within a solid, loaded at its boundary, exists when the complementary potential energy is at a minimum value. The complementary potential energy V^*

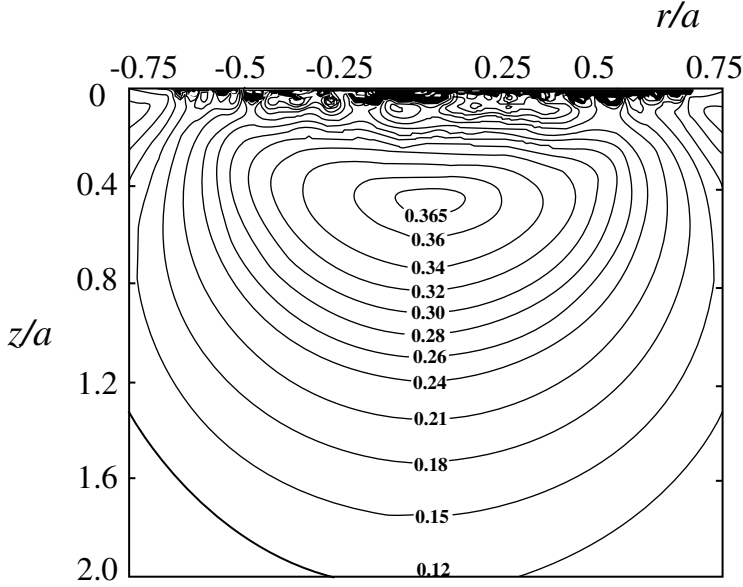


FIGURE 3.20 Numerically determined contour plot of von Mises stress beneath the contact between a rough sphere and a smooth half-space. (From Lee, S.C. and Ren, N. (1994), The sub-surface stress field created by three-dimensional rough bodies in contact with traction, *Tribol. Trans.*, 37, 615-621. With permission.)

can be expressed in terms of the stored strain energy and the forces and displacements prescribed at the surface

$$V^* = \frac{1}{2} \int_{i=1}^n p(u_{z1} + u_{z2}) dS + \int_S p(h_i - \delta) dS \quad (3.57)$$

where δ is the overlap of the two surfaces and $h(x,y)$ is their initial separation. S is the surface over which the pressure p acts, causing normal displacements u_z . Again this can be written in a discretized form as

$$V^* = \frac{1}{2} p_i \sum_{j=1}^n p_j \left(\sum_{j=1}^n C_{ij} p_j \right) + A \sum_{i=1}^n p_i (h_i - \delta) \quad (3.58)$$

where A is a constant depending on the pressure element used. The solution proceeds by determining the values of p_i which minimize V^* subject to $p_i > 0$. This has been done using a simplex method (Kalker and van Randen, 1972) or by using a direct quadratic programming technique (Tian and Bushan, 1996); the latter being more computationally efficient. Figure 3.21 shows the deformed surface predicted when this method is applied to the three-dimensional contact of a rough surface an a rigid sphere.

3.4.3 Finite Element Method

Where the contact problem involves nonlinear material behavior such as plasticity or when the half-space assumption is no longer valid (i.e., when the dimensions of the body are not large compared with the regions of contact), then it may be necessary to employ a finite element method. A number of commercial finite element codes are available which are suitable for the solution of frictional contact problems (e.g., ABAQUS, DYNA3D, MARC, NASTRAN, RADIOSS, ANSYS).

The contacting bodies are divided into an array of either two-dimensional axisymmetric or planar elements or three-dimensional brick elements. A set of gap or contact elements is defined between the surface nodes which are likely to come into contact during loading. A friction coefficient may be defined for the contact. Appropriate materials properties, boundary conditions, and nodal loadings are also specified.

Modern programs have graphical interfaces to assist in the generation of meshes. The “preprocessing” program then generates an input file to use with the main analysis program. Also available are “postprocessing” packages, which generate graphical presentations of results such as stress or strain contours and deformed mesh shapes.

Typically, contact problems involve loads distributed over small areas of contact. This results in rapid changes in stress in the elements surrounding the contact. Therefore, the mesh in this region must be refined adequately to describe these high stress gradients. Thus even a simple Hertz contact analysis needs a fine mesh for accurate results (Figure 3.22). The nonlinearities associated with contact, stick-slip behavior, and plasticity mean that the solution is loading-path dependent. This requires that the applied load is divided into a number of small steps and a solution found at each iteration. These two features can often lead to long computation times.

The availability and ease of use of commercial FEA codes has led to their widespread adoption in both industrial and academic environments. Many fundamental tribological phenomena have been investigated by this method (e.g., elastic plastic contact, layered bodies, repeated contact cycles, indentation and hardness testing, impacts, metal forming, rough surface contact, and crack propagation). Likewise, countless industrial products and processes with contact aspects are modeled with these techniques.

3.5 Experimental Methods for Contact Mechanics

In the experimental analysis of contact problems for machine element design, it is likely that the measurement of contact (both real and apparent) and contact stress are of interest. There are several experimental techniques for determining these parameters.

3.5.1 Real and Nominal Area of Contact Measurement

Where the surfaces of the bodies are rough, contact is made only at the asperity summits. The true area of contact can be significantly lower than the geometrical (or nominal) area of contact. The techniques of determining this real area of contact can be categorized into electrical, optical, acoustic, and surface-coating methods.

Electrical and Thermal Resistance

Measurement of the electrical resistance between contacting surfaces can give information about the true area of contact (Holm, 1967; Bowden and Tabor, 1939). This must be achieved by means of a current-potential method because the resistance of the contact is small (of the order 10^{-3} to 10^{-6} ohms) compared with that of the leads. If the region of contact between two bodies is made up of n discrete contact points of radius a_i , the total contact resistance is

$$R = \rho \div \sum_{i=1}^n a_i \quad (3.59)$$

where ρ is the resistivity of the material (ohm/m). The resistance R depends on the contact spot radius (and *not* on the spot area). It is therefore not possible to determine the real area of contact directly, using this method, unless some assumption is made about either the size or number of individual contact spots. Equation 3.59 only holds if the contact points are far apart. In addition, the presence of surface

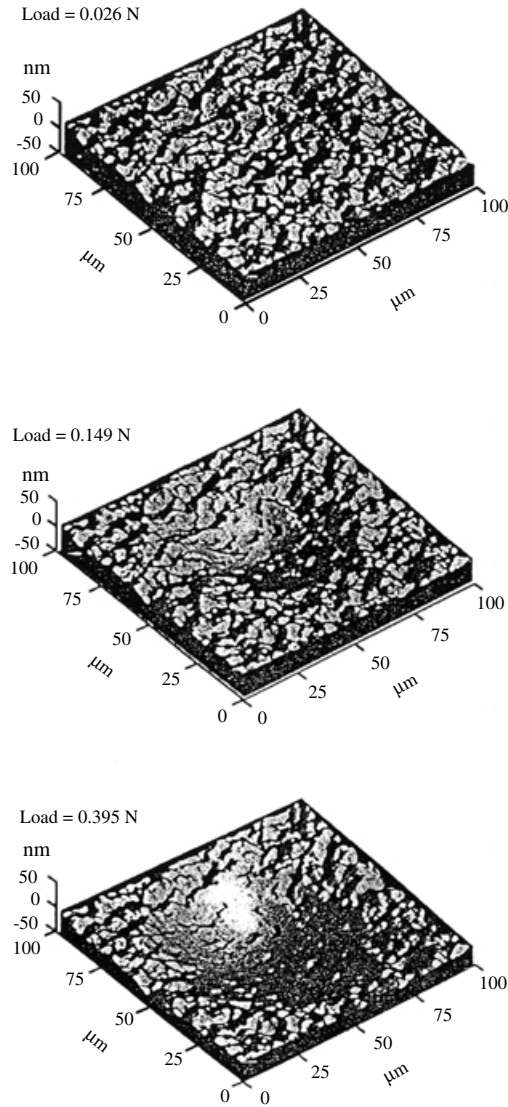


FIGURE 3.21 Numerical model of the contact between a three-dimensional rough surface and a smooth rigid sphere. (From Tian, X. and Bhushan, B. (1996), A numerical three-dimensional model for the contact of rough surfaces by variational principle, *Trans. ASME J. Trib.*, 118, 33-42. With permission.)

oxide films can have a significant effect on the contact resistance, and it is for these reasons the electrical resistance method is restricted to qualitative measurement.

A similar method is the measurement of heat flux across the interface (Newcomb, 1957). This method has the advantage of being less susceptible to surface films. However, heat flow is not limited to junctions, as some may occur across air gaps.

Optical Methods

If one of the contacting bodies is transparent to light, then the regions of contact can be observed directly. There are a number of ways in which this principle has been realized, including direct methods, frustrated internal reflection and optical interference.

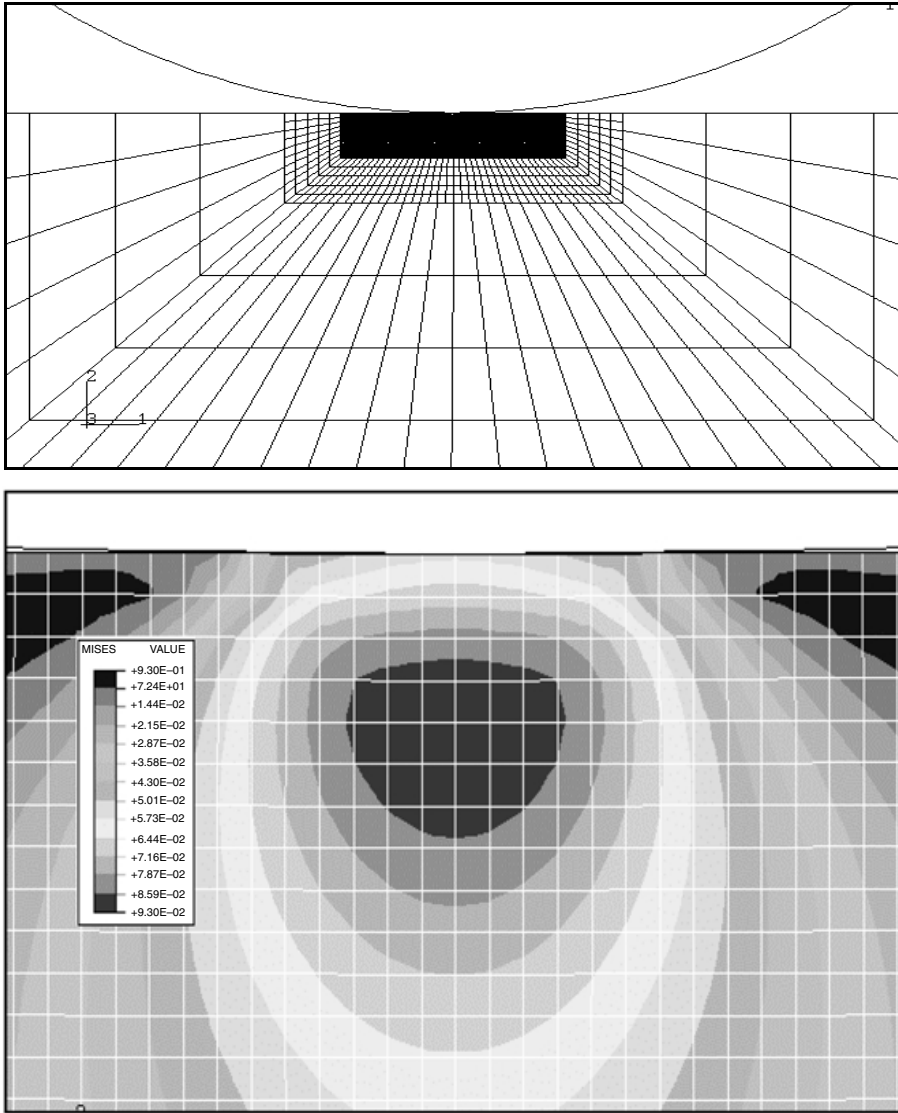


FIGURE 3.22 (a) A typical finite element mesh for use with a planar frictional contact problem; (b) contour plots of von Mises stress obtained by finite element method.

A soft metal may be pressed against a glass surface and the number and size of the regions of contact measured directly (Dyson and Hirst, 1954). On the other hand, if both bodies are transparent, then a light beam may be directed through the interface. The beam will pass through regions of contact without deflection but will be scattered at air gaps. Contact areas then show up as bright spots against a grey background (Kragelsky, 1965).

Alternatively, the specimen surface may be loaded against a prism and a parallel beam of light directed at the interface at grazing incidence (Kragelsky and Demkin, 1960). At regions of noncontact the beam is internally reflected, while at the contact spots this is frustrated. Contact spots are then viewed in the reflected light.

A third method relies on the optical interference between two light beams, one of which is reflected from the top surface and one from the bottom (Bailey and Courtney-Pratt, 1955). The two surfaces,

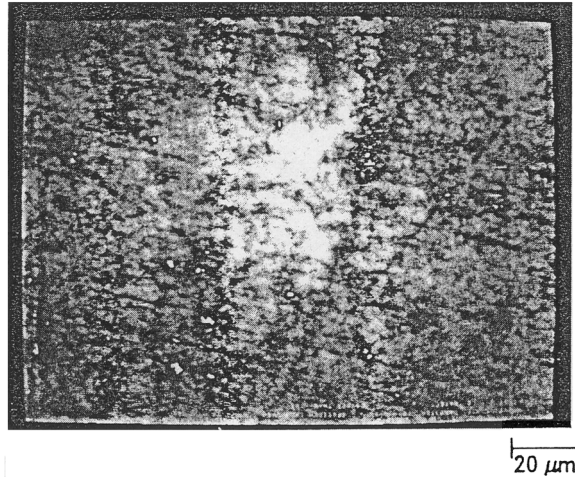


FIGURE 3.23 Optical interferogram of the contact between a flat disk and a glass lens. The real area of contact is related to the proportion of light spots in the image. (From Bhushan B. and Dugger M. T. (1990) Real contact area measurements on magnetic rigid discs, *Wear*, 37, 41-50. With permission.)

which must be transparent, are coated with thin metallic films so that they are partially reflective. A beam of light is directed through the transparent specimen. Interference occurs between that part of the beam which reflects at the first surface and that which reflects at the second. Constructive interference occurs when the gap between the two specimens is equal to a whole number of wavelengths. Figure 3.23 shows an optical interferogram of the contact between a flat disk and a glass lens (Bhushan and Dugger, 1990).

Ultrasonic Reflection

A wave of ultrasound incident at an interface between two materials will transmit through regions of contact and be reflected back at air gaps. This phenomenon can be used to investigate the true area of contact at an interface (Kendall and Tabor, 1971; Drinkwater and Dwyer-Joyce, 1996). An ultrasonic transducer is mounted on one of the bodies and is excited to emit a broad band (typically 5 to 20 MHz) longitudinal pulse. The reflected pulse is received by the same transducer, amplified, and its signal stored on a computer. The amplitude of the reflected pulse can be divided by that of the incident pulse to give a reflection coefficient $|R|$. Provided that the wavelength of the signal is large compared with the size of the asperity contact, this is dependent on K the stiffness of the interface

$$|R| = \left\{ 1 + \left(2K / \omega z \right)^2 \right\}^{-1/2} \quad (3.60)$$

where ω is the angular frequency of the wave, and z is the acoustic impedance of the material (the product of the wave speed and density). The stiffness of the interface is defined as the contact pressure required to cause unit approach of the surfaces; this depends on both the number and size of individual contact spots as well as their proximity. The stiffness varies from zero to infinity as the ratio of the real to nominal area of the contact varies from zero to 100%.

This method may also be used to determine the extent of a contact region. An ultrasonic transducer is scanned across the interface. Where contact is made the wave will be transmitted; reflection occurs at regions of noncontact. The resolution of this technique is relatively coarse and is currently of most use in applications where the contact region is relatively large. Figure 3.24 shows a scan of the interface between a rough rubber sphere and a smooth steel flat. The resolution of the method is such that

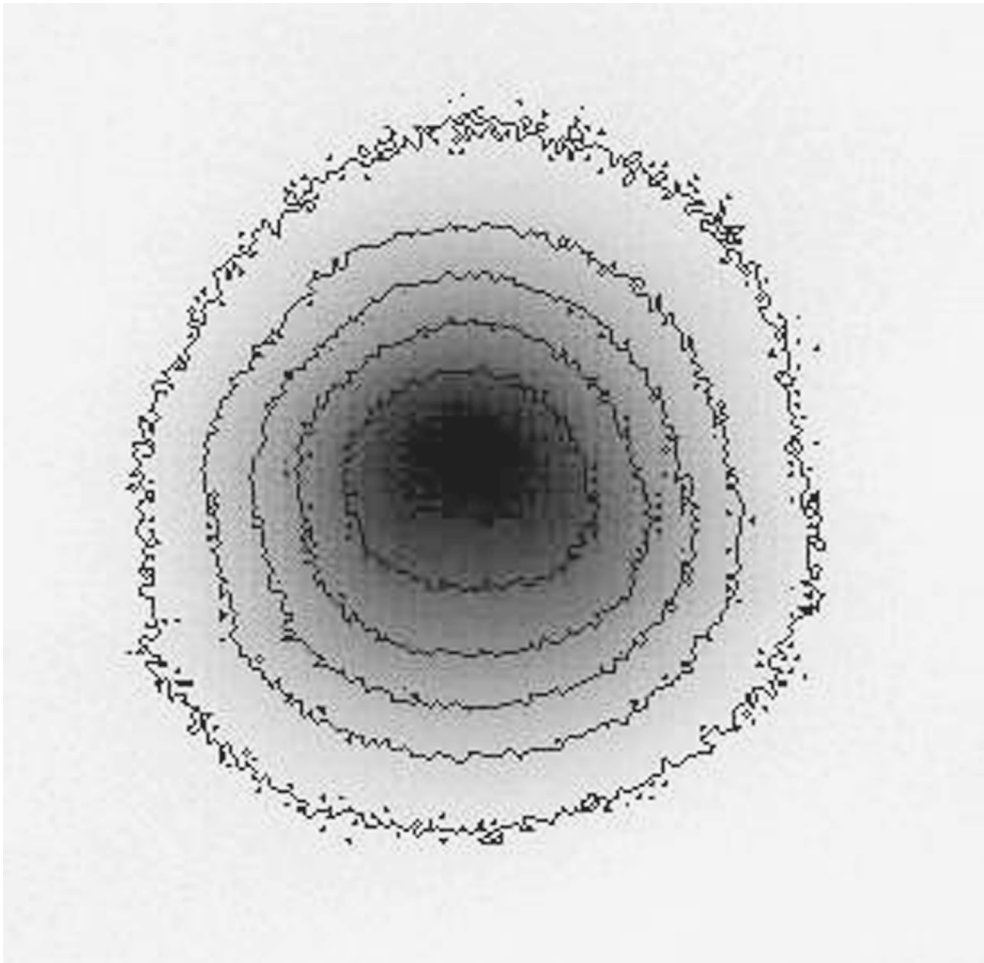


FIGURE 3.24 Ultrasonic scan of the contact between a steel ball and a steel flat (diameter 1.3 mm). The levels of shading indicate the reflection coefficient (the proportion of an incident wave reflected at the interface). Thus regions of higher pressure will result in increased conformity of the roughness and thus reduced reflection.

individual contact regions are not visible; rather regions of conformity (high percentage contact) are observed.

Thin Surface Coatings

A thin, soft metal (e.g., copper, silver, or gold) coating is applied to one of the surfaces by chemical or physical deposition and the bodies are then loaded together as required. The thin layer within the region of contact is deformed. The extent of the contact region can be observed by a change in the appearance of the deposited film.

Alternatively, a radioactive or thin fluorescent paint coating may be applied to one surface. This is then loaded against the counterface. The extent of transfer can be determined quantitatively by, for example, use of some form of radioactivity counter.

3.5.2 Experimental Contact Stress Analysis

Photoelasticity and Caustics

The contacting bodies are modeled in photoelastic material, such as polycarbonate or epoxy resin. For two-dimensional applications a planar model is fabricated and loaded in a polariscope (a light source

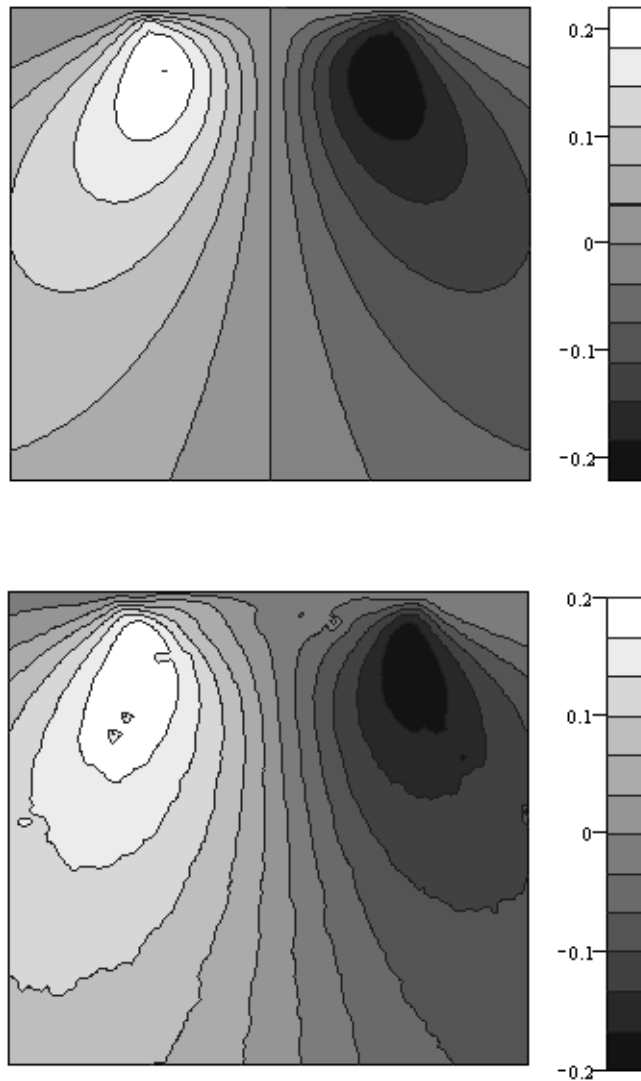


FIGURE 3.25 Photoelastic fringe pattern of the contact between a cylinder and half-space. Images have been obtained by a stress freezing method and are digitized for an automated analysis. (From Burguete, R.L. and Patterson, E.A., (1997), A photoelastic study of contact between a cylinder and a half-space, *Exp. Mech.*, 37, 314-323. With permission.)

with two polarizing filters). The isochromatic fringe pattern gives contours of principal stress difference ($\sigma_1 - \sigma_2$). For three-dimensional models a stress freezing technique must be employed. The most common method is to apply the loading to the model at elevated temperature and allow to cool under load. The deformation remains frozen into the structure, which may then be cut into appropriate slices for analysis in a polariscope. This method has been used by Ollerton and Haines (1963) to study elliptical contact subjected to normal and tangential traction. Full field imaging and automated stress separation of contact problems has been studied by Burguete and Patterson (1997). Figure 3.25 shows a photoelastic fringe pattern from the contact between a cylinder and a flat surface compared with the theoretical solution.

A related application is the method of caustics. The application of stresses causes deflection and a change in the refractive index of the material. If a planar specimen is illuminated with parallel incident light, rays which pass through the stressed regions are deflected. The light distribution on an image plane behind the specimen is no longer uniform. The boundary between the light and dark regions is known

as the caustic curve. This analysis is generally carried out by finding the boundary loading which gives a theoretical caustic curve closest to that observed (Theocaris and Stassinakis, 1978).

Contact Pressure Pads and Films

For larger-scale contacts, commercially available thin film pressure transducers can have an application. The transducer consists of a matrix of conductive elements whose resistance varies with applied load. The elements are mounted on a thin flexible sheet which is placed between the contacting bodies. Typically, the smallest transducers are of the order of 1 mm^2 and are sensitive to pressures up to 150 MPa. They are used routinely in the motor industry to study human interactions with the vehicle.

Die impregnated films can also be used as pressure-sensitive devices. The film is placed between the contacting bodies, which are then loaded. The film releases a dye at controlled pressure levels. After unloading, the film is then removed and examined; the color depth is proportional to applied pressure.

Resistive Microtransducers

Small elements of manganin or titanium are deposited onto a surface. When this microtransducer is located inside a contact, the pressure causes the resistance to increase. Typically the transducers are deposited through a shaped mask by radio frequency sputtering (for metal surfaces a thin layer of insulating silica is deposited first). They have been used to measure pressures in rolling and sliding lubricated contacts (Hamilton and Moore 1971, Kannel et al. 1965). The transducers are also sensitive to temperature rises which must be taken into consideration during any measurement.

3.6 Further Aspects

3.6.1 Contact of Rough Surfaces

Nominally Flat Surfaces

Having established the modes of behavior of a single contact spot, it is now possible to consider what happens when two real surfaces, that is, those carrying a large number of asperities with a range of summit heights, are brought into loaded contact. The elastic stresses depend on the *relative* profiles of the surfaces, described by the reduced radius of curvature R and the contact modulus E^* . This means that, for the purposes of analysis, all the deformable surface imperfections can be considered to be concentrated on one surface, while the second surface is both rigid and plane. Suppose the rough surface to consist of N hills or asperities whose heights z above the mean level vary in some statistical fashion. This distribution can be described by the probability density function $\phi(z)$, which must be such that

$$\int_{-\infty}^{+\infty} \phi(z) dz = 1 \quad (3.61)$$

so that all the summits are included. In their often-quoted treatment, Greenwood and Williamson (1966) assumed that the tips or summits of the asperities were spherical and all had the same characteristic radius of curvature R_s . When a normal load is applied, the rigid surface moves toward the mean level of the rough surface so that, when the separation is d , it will have made contact with all the asperities for which $z_s > d$. The number n of such contacts will be given by

$$n = N \int_d^{\infty} \phi(z) dz \quad (3.62)$$

In the case of a rough surface that exhibits perfect plasticity, deformation of each contact spot will occur at the same normal pressure p_m . It follows immediately that as long as p_m remains constant the

total real area of contact between the two solids will be in direct proportion to the applied load W . The actual number of individual contact points will depend very much on the nature of the function $\phi(z)$.

However, when the contact is *purely* elastic, each asperity contact spot can be treated as a separate Hertzian contact between a sphere and a flat. This might occur either because the normal load is insufficiently large to cause significant plastic flow, or perhaps more realistically, because after repeated loading, the shakedown mechanism has given rise to elastic conditions (see Section 3.6.3).

The total load W is given by summing up all n such contacts, so that

$$W = \frac{4}{3} NE^* R_s^{1/2} \int_d^\infty \phi(z) (z-d)^{3/2} dz \quad (3.63)$$

and the total real area of contact A is the sum of all the n contacts, so that

$$A = N\pi R_s \int_d^\infty \phi(z) (z-d) dz \quad (3.64)$$

The numerical values of n , W , and A given by Equations 3.62, 3.63, and 3.64, clearly depend on the form of the function of $\phi(z)$. In their original paper Greenwood and Williamson considered a random or Gaussian distribution of summit heights and evaluated these integrals numerically; however, they also suggested that an analytical treatment is possible if we make use of the observation that, whatever the precise form of the bell shaped distribution of $\phi(z)$, it is only the most prominent asperities, i.e., those of large z , which will take part in the surface interactions. Within this region we can reasonably model $\phi(z)$ by an exponential expression of the form

$$\phi(z) = \lambda \exp(-\lambda z)$$

where λ is a suitable constant. In such a case, the number of points of contact n , and the real contact area A , are both *strictly proportional* to the applied load W . As the separation of the surfaces decreases, although the area of any particular contact spot grows, more asperities make contact so that the *average* size of the patches of real contact, i.e., A/n , remains constant. Provided the surface contains a *range* of surface roughnesses the direct proportionality between applied load and area of contact extends throughout the *elastic* range of material response as well as the plastic regime.

A consequence of the area of real contact being proportional to the load is that the actual mean contact pressure is constant. Its value p_m is equal to the load W divided by the area of contact A

$$p_m = \frac{W}{A} = \frac{4}{3\pi} \frac{E^*}{R_s^{1/2}} C_1 \quad (3.65)$$

and is therefore independent of the separation of the surfaces. When the numerical value of the coefficient C_1 is substituted, Equation 3.65 becomes

$$p_m \approx 0.39E^* \sqrt{\sigma_s/R_s} \quad (3.66)$$

where σ_s is the standard deviation of the surface roughness. Since each asperity contact is being modeled as being equivalent to that between a sphere and a plane, the results of the Hertz analysis can be used to predict those conditions which will just lead to the initiation of plastic deformation. In such a Hertzian contact the maximum shear stress occurs beneath the surface and is of magnitude $0.47p_m$. Setting this

shear stress equal to the flow stress of the material k , we can see that the critical value of p_m to cause yielding is equal to $2.15k$. Since the flow stress is simply related to the hardness of the material H ($\approx 5.6k$) there is likely to be some plastic yielding if

$$p_m \geq \frac{2.15 \times H}{5.66} \quad (3.67)$$

Combining this with the expression for p_m from Equation 3.65 the condition for the onset of plasticity becomes that

$$\frac{E^*}{H} \sqrt{\sigma_s / R_s} \geq C \quad (3.68)$$

where C is a constant which, although depending to some extent on the form of the height distribution, is of the order of unity. The nondimensional group on the left-hand side of Equation 3.68 is sometimes known as the *plasticity index* Ψ , and neatly describes the likely deformation characteristics of a rough surface. If its value is very much *less* than unity, then asperity deformation is likely to be entirely elastic; on the other hand, if Ψ is significantly *greater* than 1 the response is predominantly plastic. The factor $\sqrt{\sigma_s / R_s}$ is associated with the mean slope of the surface and can be obtained, approximately at least, from the profilometry data.

Elastic Contact of Rough Curved Surfaces

Most tribological contacts are between nonconforming or curved surfaces. The question then naturally arises, under what circumstances can the surface roughnesses on such loaded curved surfaces be safely neglected in calculating the contact stresses (using, for example, the Hertzian analysis) and under what circumstances can they not. For the situation to be amenable to quantitative treatment it is necessary for the two ranges of scales involved to be very different. These are typified on the one hand by the radius of curvature of the mean level of the surfaces, say R , and on the other by height and spatial distribution of the asperities. The nominal contact area will contain many deforming asperities, and these then act as a compliant layer lying over the surface of the body. Contact extends over a larger area than would be the case if the solids were perfectly smooth.

The details of the quantitative solution are beyond the scope of this discussion. They can be found in Johnson (1985) and examples in Bhushan (1996, 1998). However, the conclusions can be conveniently expressed if a nondimensional parameter χ is defined to describe the influence of the surface roughness by the relation

$$\chi = \sigma_s R / a^2 = \sigma_s \left\{ 16 R E^{*2} / 9 W^2 \right\}^{1/3} \quad (3.69)$$

here a is the contact radius for smooth surfaces according to the Hertz analysis. The smoother the surfaces, then the *smaller* the value of χ (since it depends on σ_s). The influence of χ on the effective pressure distribution for point contacts can be seen in [Figure 3.26](#), which illustrates the extent of the contact zone for two values of χ which correspond to smooth ($\chi = 0.05$) and rough ($\chi = 2$) engineering surfaces. In this plot the pressure is normalized by the maximum Hertz pressure for smooth surfaces, the radial distance by the Hertzian contact radius a , and the dotted curve represents the Hertzian semi-ellipse. With a rough surface the peak pressure is much below the normal Hertzian value and falls more or less asymptotically to zero over an area very much greater than the Hertzian area. However, if χ is less than about 0.1 both the pressure curve and the contact radius are really quite close to Hertzian values. Thus the Hertz theory for perfectly smooth surfaces can be used to calculate the contact stresses on real engineering surfaces without introducing numerical errors of more than a few percent, provided the nondimensional group $(\sigma_s R / a^2)$ is less than about 0.1.

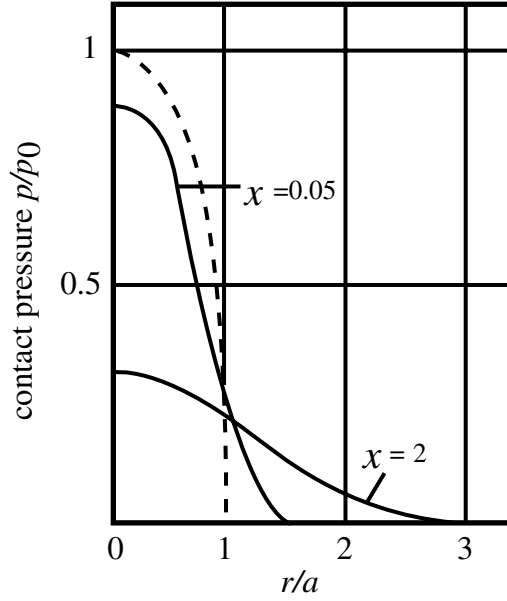


FIGURE 3.26 The effect of the contact parameter χ on the contact between rough, curved surfaces.

3.6.2 Loading Beyond the Elastic Limit

Limit of Elasticity

There must be some value of load for which the material of the weaker surface of the contact pair starts to deform *plastically* rather than elastically: to estimate this critical value an appropriate yield criterion must be applied. The yield of ductile metals is usually taken to be governed by either the Tresca maximum shear stress criterion or the von Mises strain energy criterion.

In the case of the axial contact of cylinders the condition of plane strain deformation ensures that the stress component σ_{yy} is the intermediate principal stress. The Tresca criterion involves equating the maximum principal shear stress (which from Section 3.2.5) is equal to $0.30p_0$ to the shear yield stress k (which is numerically equal to $Y/2$). Hence, the critical value of the peak pressure p_0^Y is given by

$$p_0^Y = 3.3 k = 1.67 Y \quad (3.70)$$

The corresponding value of the mean pressure p_m^Y is, from Equation 3.10,

$$p_0^Y = \frac{\pi}{4} p_m^Y \approx 2.59 k = 1.3Y \quad (3.71)$$

The load per unit length of the contact P^Y required to bring material to the point of yield can be found from Equation 3.10,

$$P^Y = \frac{\pi R}{E^*} p_0^Y = 8.76 \frac{RY^2}{E^*} \quad (3.72)$$

Even when some yielding has taken place, the scale of the changes of shape must be small. This is because initial yield has occurred *beneath* the surface so that the plastic zone is still totally surrounded by a region in which the stresses and strains are still elastic: this limits the extent of plastic deformation

since the plastic strains must be of the same order as the adjacent elastic strains. The material displaced by the flattening of the contact is accommodated by an *elastic* expansion of the surrounding hinterland.

If the normal load on the contact is increased still further, beyond P^Y , then the plastic zone grows until eventually it breaks out at the free surface. Once this occurs the constraint on the scale of the plastic deformation offered by the elastic hinterland is much reduced; the plastic strains can become very much larger with significant changes in the surface profile. In an analysis of this situation, it is then reasonable to consider the elastic strain components to be negligibly small in comparison to the plastic. Provided the material does not strain harden to any appreciable extent, it may be idealized as a *rigid perfectly plastic* solid which flows under a constant shear *flow* stress k .

The situation can now be analyzed by one of the standard methods for dealing with large strain plastic forming processes; for example the *slip line field* technique in two dimensions, or the approximate, but more readily applied, *upper bound* method in both two- and three-dimensional geometries (see, for example, Johnson and Mellor, 1964, or Hosford and Caddell, 1983). The upper bound method depends on postulating some mechanism of plastic flow which is compatible with both the external constraints on the body and with volume constancy. Equating the rate at which the external load does work to the rate at which energy is absorbed within the material provides an upper bound on the true maximum load the contact can sustain.

In the case of a single, loaded line contact the upper bound theorem leads to estimates of the form

$$p_m = 5.66k = 2.83Y. \quad (3.73)$$

When the same methods are applied to other indentation geometries, it is found that although the numerical value of the ratio p_m/Y depends to some extent both on details of the geometry and the frictional conditions on the surface it is always close to the value 3. Thus we can see that after subsurface yielding is initiated (at an interfacial mean pressure of about Y) there is a transitional range, involving values of p_m between Y and $3Y$, when the plastic flow is constrained by the surrounding elastic material and within which the elastic strains cannot be neglected in comparison with the plastic. These three ranges of loading — elastic, elastic/plastic, and fully plastic — are characteristic of nearly all engineering structures subjected to loadings of increasing severity. The pressure p_m can be thought of as a measure of the indentation hardness H of the material, that is, the smallest value of normal pressure required to bring about significant plastic deformation under a rigid indenter.

Elastic/Plastic Contact

While problems in the intermediate elastic–plastic region are, in principle, susceptible to analysis, solutions are difficult to obtain in all but the simplest of geometries. Difficulties arise in finding solutions which satisfy both the equations of equilibrium and compatibility on both sides of the boundary between elastically and plastically stressed material because the size and shape of the elastic/plastic boundary is not known *a priori*. In the case of the contact between a flat surface and a comparatively blunt indenter, i.e., one with an included angle of more than about 140° , the experimental observation that the subsurface displacements are approximately radial from the point of first contact, with roughly cylindrical or spherical contours of equal strain, has led to a simplified approximate model (Johnson, 1970).

The pressure beneath the indenter is a function of the single nondimensional group $E \tan \theta / Y$ where θ is the angle between the face of the indenter of its local tangent and the surface. This group can be interpreted as the ratio of the strain *imposed* by the indenter (which is related to its “sharpness” or the value of $\tan \theta$) to the strain *capacity* of the material as measured by the elastic strain at yield (i.e., Y/E); the group $E \tan \theta / Y$ is thus a measure of the severity of the loading. The three regions of behavior, elastic, elastic–plastic, and fully plastic, can be displayed on a graph of p_m/Y vs. $E \tan \theta / Y$. The elasticity of the indenter can be taken into account by replacing E by E^* (defined as in Equation 3.3). Such a “map” of material behavior is shown in [Figure 3.27](#). First yield for a spherical indenter occurs at about $p_m/Y = 1.1$ and for a cone at about $p_m/Y = 0.5$. The fully plastic state is reached when the group $E^* \tan \theta / Y$ has a numerical value in the range of 30 to 40.

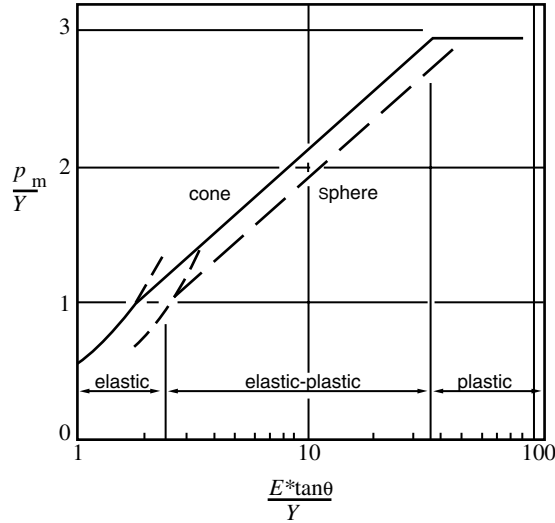


FIGURE 3.27 Indentation of a half-space by spheres and cones. p_m/Y is the ratio of the mean contact pressure to the material yield stress, and $E^*\tan\theta/Y$ is a measure of the severity of loading.

The same gradual transition from purely elastic contact to fully plastic behavior occurs when the contact is subject to a tangential load applied simultaneously with the normal pressure. The presence of a surface traction moves the point of initial yield from the Oz axis of symmetry and brings it closer to the surface as shown by comparing Figures 3.8 and 3.11. As the frictional traction at the surface gets larger, a stage is reached when the position of maximum shear occurs actually *at* the surface rather than beneath it. This effect is illustrated in Figure 3.28 curve A in which the ratio of the maximum Hertz pressure to the shear flow stress, i.e., to p_0/k , which is just sufficient to cause initial plastic flow, is plotted against the traction or friction or traction coefficient Q/P . Figure 3.28 is thus effectively a second map

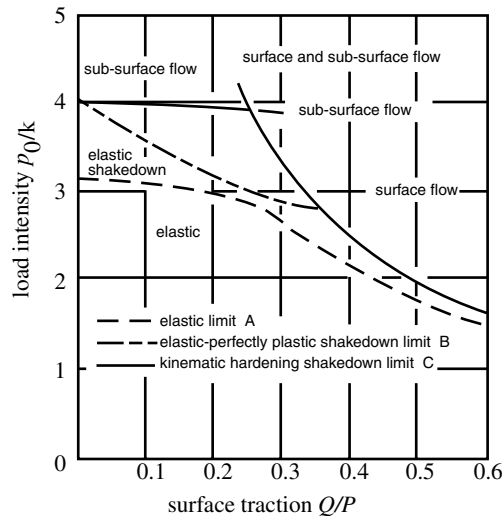


FIGURE 3.28 Effect of surface traction Q/P on the maximum Hertz pressure p_0 for either yielding or shakedown of material with shear flow stress k : curve A elastic contact, first yield, Tresca; curve B elastic-perfectly plastic behavior; curve C, after shakedown, for a kinematically hardening material. (From Johnson, K.L. (1985), *Contact Mechanics*, Cambridge University Press, Cambridge. With permission.)

which can be used to determine both the likelihood of plastic deformation and whether this is at or below the surface. It applies to a single contact at which the severity of the loading, described by both the ratio of the peak Hertzian pressure to the uniaxial yield strength and the traction coefficient Q/P , is known. In the case of a line contact there is a slight complicating factor in that it is possible for deformation to occur by spread of material in the out-of-plane, or Oy -direction; however, this mode of lateral flow is favored only over a fairly narrow range of traction coefficients.

Figure 3.28 illustrates an important general point: if the friction or traction coefficient is less than about 0.3 then the plastic zone beneath a sliding contact is contained within the bulk and does not break through to the free surface. The plastic strains must therefore be comparatively small and of the order of the elastic strains. But if the traction coefficient exceeds 0.3 then the flow field extends to the surface for both line and point contacts. Having lost the constraining effect of the elastic boundary it follows that the plastic strains can now become much larger.

3.6.3 Repeated Contacts — The Role of Residual Stresses and Shakedown

In many practical contacts the loaded surfaces have to withstand many repeated passes of the load. If, in the first pass, the elastic limit is exceeded (so that the operating point of the contact is above curve A in Figure 3.28) an element of plastic deformation will take place, and thus when the load is removed some system of residual stresses will remain in the material. In the second pass of the load the material is subjected to the combined action of the applied contact stresses *and* the system of residual stress left behind from the first pass. Generally speaking, such residual stresses are protective, in the sense that they will make yielding less likely in the second pass than was the case in the first. It is possible that after a few passes the residual stress system builds up to such an extent that the applied load can be carried entirely elastically. This process is known as *elastic shakedown*. Shakedown can be investigated either by repeated application of numerical stress analysis (usually FEM, see Section 3.4.3 and, for example Ham et al., 1988, or Kulkarni et al., 1991) or by appealing to an important theorem in stress analysis; this states that if *any* distribution of residual stress can be found which, together with the elastic stresses due to the load, constitute a system of stress which is everywhere within the elastic limit of the material then the system *will* shake down. The strength of the theorem lies in the fact that the system of residual stress that is used for the calculation does not have to be the real one — as long as it is a possible one. The important practical implication is that, providing a suitable set of residual stresses is possible, the material will find them.

When this argument is applied to the case of a repeated Hertzian contact between cylinders, the condition for elastic shakedown to occur turns out to be that the peak Hertzian pressure p_0 must be no greater than $4.00k$. Using the Tresca criterion the critical value of p_0 for yield on the first application of the load is $3.3k$. This means that if the repeated load is such that $3.3k < p_0 < 4.0k$ then, although there will be some yield in the first cycle of operation, shakedown *will* occur. As the total load P on the contact is proportional to $(p_0)^2$ it follows that the ratio of the shakedown limit P^S to the elastic limit P^Y is given by

$$\frac{P^S}{P^Y} = \left\{ \frac{4.0}{3.3} \right\}^2 = 1.47 \quad (3.74)$$

In other words the load on such a contact can be nearly 50% greater than that needed to bring about yield on the first cycle before conditions are so severe that continuous plastic deformation will be produced on *every* subsequent cycle.

A similar analysis can be made of a Hertzian line contact which also carries tractive load Q . As the traction on the surface is increased, so the limiting value of the maximum Hertz pressure for which shakedown is possible reduces. This effect is illustrated in Figure 3.28 curve B, from which it can be seen that as the traction grows, so the allowable interval between the load for first yield and that for shakedown becomes narrower.

The effect of strain hardening can also lead to an apparent shakedown phenomenon irrespective of the contact geometry. With repeated deformation the effective value of k , the material shear strength, may rise in the surface layers, so that a load which initially exceeds the shakedown limit subsequently lies below it. Further refinements of the shakedown argument are possible for materials which strain harden in particular ways. For example, Figure 3.28 curve C illustrates the shakedown curve for a material that exhibits kinematic hardening. In a material which exhibits some form of nonlinear strain-hardening (which is typical of many engineering alloys) the situation of incremental plastic collapse or *plastic ratchetting* can arise; here gross deformation of the surface and close-to-surface material may accumulate after millions of repeated loading cycles, each of which contributes an element of plastic strain. Shakedown can also play a part in rolling contacts, such as a ball on a flat, in which with repeated traversals an originally plane surface becomes grooved, so increasing the contact area and reducing the contact stresses (see Williams, Dyson, and Kapoor, 1999).

3.6.4 Contacts on Layered Solids

Stress and deformation analyses for the case of a rigid cylindrical indenter making contact with a half-space carrying an elastic surface layer with differing mechanical properties from the bulk have been carried out by a number of authors (Burminster, 1945; Hannah, 1951; Gladwell, 1976; Chen, 1971; Chen and Engel, 1972; Jaffar and Savage, 1988). The usual method of solution is to first calculate the local surface deformations from the degree of nominal overlap of the two surfaces and then to solve, either analytically or numerically, the stress-deformation integral equation, so arriving at a consistent distribution of normal pressure at the interface: this can then be integrated to give the normal load. Gupta and Walowit (1974) and Kannel and Dow (1986) obtained generalized solutions for elastic cylindrical indenters against layered elastic solids. Walowit and Pinkus (1982) and King and O'Sullivan (1987) analyzed the stress field associated with an elastic cylindrical indenter sliding with traction over an elastic half-space with a single surface layer (see Figure 3.29a), and this analysis was extended to multiple layers by Elsharkaway and Hamrock (1993). O'Sullivan and King (1988) examined the case of a spherical indenter loaded sliding over a single layered surface. Results from analyses of this sort can be validated and extended by finite element analyses such as those of Ihara (1986), Komvopoulos (1988), and Arnell (1993, 1994). Figure 3.29, taken from O'Sullivan and King, illustrates the effect of both harder and softer elastic layers on the surface pressure distribution and local deformation. In these results Poisson's ratio is taken as 0.3 throughout and the indenter is assumed rigid and of radius R . In Figure 3.29b the pressure is normalized by p_0 which is the Hertz pressure under the center of the indenter for the case $E_1 = E_2$ when the radius of contact a is equal to layer thickness. Figure 3.29c illustrates the variation with indentation depth δ with the normal load P ; this is normalized by load P_0 , the load for the situation when $E_1 = E_2$ and $a = h$. The more compliant layer gives, as expected, enhanced penetration depths and significant changes to the pressure profile. Subsurface stress components, σ_{xx} , σ_{zz} , and τ_{xz} , are also much modified by the presence of both the surface layer and any surface friction or traction. A stiffer surface layer behaves rather like a thin elastic plate bonded to the substrate, large bending stresses can develop giving a tensile σ_{xx} at the interface, which can be significant for the growth of cracks normal to the interface. The shear stress τ_{xz} is also enhanced by the stiffness of the surface layer, although its magnitude decays rapidly with depth z . High interface shear stress can adversely effect the bonding between layer and substrate. Layers which are of lower modulus than the bulk generally lead to reduced stresses but, since they are also likely to be of lower hardness, provide lower resistance to wear or surface damage. To fully exploit the superior tribological properties of harder (and generally stiffer) surface coatings, softer intermediate layers may be appropriate. Similar effects have been found in the case of nominal point rather than line contacts (Kral et al., 1997; Lovell, 1998).

Relatively little work has been done on the analysis of the elastic-plastic contact of layered surfaces despite the fact that plastic deformation is a major concern in the design and performance of engineering components. Failure of a hard coating-soft substrate system under some tribological situations may be caused not by conventional wear but rather by debonding of the coating from the substrate (adhesive

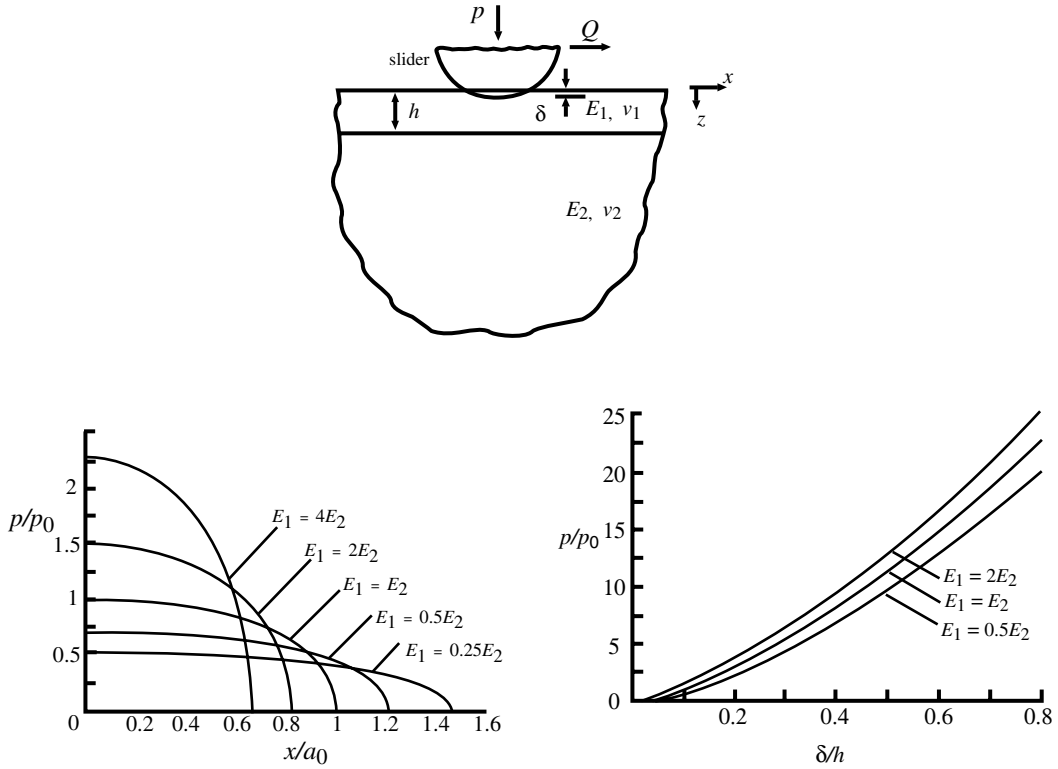


FIGURE 3.29 (a) Spherical or cylindrical indenter in contact with a layered half-space; (b) normal contact pressure profiles beneath a frictionless rigid spherical indenter for an elastic layered half-space with different values of E_1/E_2 , a_0 is the Hertzian semi-dimension when $E_1 = E_2$; (c) indentation depth vs. applied load, P_0 is the normal load for the case $E_1 = E_2$ and $a_0 = h$. (Adapted from O'Sullivan, T.C. and King, R.B. (1988), Sliding contact stress field due to a spherical indenter on a layered elastic half-space, *Trans. ASME J. Trib.*, 110, 2345-240. With permission.)

failure), fracture of the coating (cohesive failure), or even subsurface fracture (substrate failure). Accordingly, it is valuable to have a picture of the spatial distribution of plastic stresses and of the position of the initiation and subsequent development of the plastic zone. Based on an FEM analysis of a hard coating (TiN) on a soft substrate (Ti) indented by a rigid cylinder, Komvopoulos (1989) found that coating thickness has a significant effect on the initiation and development of the plastic zone. Deformation maps can be constructed which define the deformation mode (elastic or plastic) in terms of such parameters as maximum pressure and coating thickness.

More recent work (see Mao et al., 1994) has used a rigorous elastic-plastic FEM analysis to study the initiation and development of the plastic zone in various TiN coating-substrate systems, including TiN-Al, TiN-Ti, and TiN-HSS (high-speed steel) indented by a rigid ball of 100 μm diameter. The results confirm that both the coating thickness and substrate strength have a significant influence on both the plastic deformation behavior and the load-bearing capacity. In most coating-substrate systems, plastic deformation is initiated first in the substrate at the coating-substrate interface; plastic deformation does not initiate in the harder coating until a large plastic zone has been developed in the substrate. Increasing the coating thickness and/or substrate strength increases the resistance of the composite to plastic deformation. When the substrate is relatively soft and the coating is relatively thin, plastic deformation is concentrated in the substrate. With increasing coating thickness and substrate strength, plastic deformation in the coating becomes more evident, such that when a very thick TiN coating on (relatively strong) HSS is indented by the rigid ball, plastic deformation is initiated in the coating rather than the

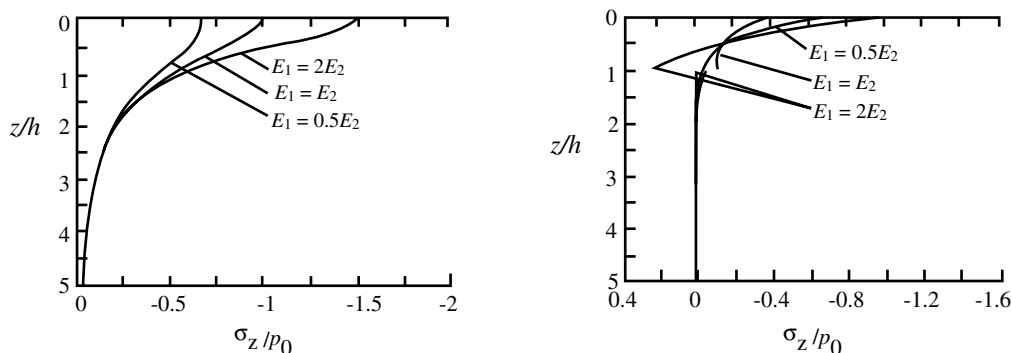


FIGURE 3.30 Normal and shear stress components under the center of a rigid spherical frictionless indenter sliding over a layered elastic half-space with different values of E_1/E_2 ; the radius of contact a_0 is equal to layer thickness h when $E_1 = E_2$.

substrate. Only when the coating is very thin does its presence degrade the load-bearing capacity of HSS. This indicates the importance of depositing a sufficiently thick coating in order to optimize the tribological performance of coated systems.

Tian and Saka (1991) also investigated the interaction between a rigid cylinder and a two-layer elastic-plastic half-space. As the normal load increases, plastic deformation initiates in the substrate for a thin interlayer (specifically when the interlayer thickness was less than 0.7% of cylinder radius), whereas, for a thicker interlayer, plastic deformation could be initiated at the interface between the top layer and the interlayer. In the elastic-plastic regime, the contact pressure distribution became flatter and the peak pressure moved slightly away from the contact center as the plastic zone breaks out to the contact interface. Sliding contact was also considered and it was found that surface deformation, location of initial yielding, as well as the values of the stresses and strains along the interfaces between layers strongly depended on the coefficient of friction. Yielding initiated on the surface when the friction coefficient was high and in the subsurface when the friction coefficient was low. Surface strains, especially shear strains, were large owing to the element of unconstrained deformation. The completion of further analyses of the contact of such multilayered solids with representative surface topographies, especially under conditions of repeated loading, should enable their optimization to be more readily achieved than by empirical developments, which are currently most commonly used.

References

- Arnell, R.D. (1994), Stress distribution in layered surfaces subject to tribological forces, *Phys. Stat. Sol. (a)*, 145, 247-254.
- Arnell, R.D. and Djabella, H. (1993), Two-dimensional finite element analysis of elastic stresses in double-layer systems under combined surface normal and tangential loads, *Thin Solid Films*, 226, 65-73.
- Beeching, R. and Nichols, W. (1948), Theoretical discussion of pitting failure in gears, *Proc. Instn Mech. Engrs.*, 158, 317-324.
- Bentall, R.H. and Johnson, K.L. (1967), Slip in the rolling contact of two dissimilar elastic materials, *Int. J. Mech. Sci.*, 9, 389-397.
- Bhushan B. and Dugger M. T. (1990) Real contact area measurements on magnetic rigid discs, *Wear*, 37, 41-50.
- Bhushan, B. (1996), Contact mechanics of rough surfaces in tribology: single asperity contact, *Appl. Mech. Rev.*, 49(5), 275-298.
- Bhushan, B. (1998), Contact mechanics of rough surfaces in tribology: multiple asperity contact, *Trib. Lett.*, 4, 1-35.

- Bowden, F.P. and Tabor, D. (1939), The area of contact between stationary and moving surfaces, *Proc. R. Soc. A*, 169, 391-402.
- Bryant, M. D. and Keer, L. M. (1982), Rough contact between elastically and geometrically similar curved bodies, *ASME J. Appl. Mech.*, 49, 435-441.
- Bufler, H. (1959), Zur theorie der rollenden reibung, *Ing. Arch.*, 27, 137-142.
- Burguete, R.L. and Patterson, E.A., (1997), A photoelastic study of contact between a cylinder and a half-space, *Exp. Mech.*, 37, 314-323.
- Burminster, D. M. (1945), General theory of stresses and displacements in layered systems, *J. Appl. Phys.*, 16, 89-94.
- Chen W.T. (1971), Computation of stresses and displacements in a layered elastic medium, *Int. J. Engng. Sci.*, 9, 775-800.
- Chen W.T. and Engel, P.A. (1972), Impact and contact stress analysis in multi-layered media, *Int. J. Solids Struc.*, 8, 1257-1281.
- Dyson, J. and Hirst, W., (1954), The true area of contact between solids, *Proc. Phys. Soc. (London)*, 67B, 309-312.
- Elsharkaway, A.A. and Hamrock, B.J. (1993), Numerical solution for dry sliding line contact of multi-layered elastic bodies, *ASME J. Tribology*, 115, 237-245.
- Fessler, H. and Ollerton, E. (1957), Contact stresses in toroids under radial loads, *Brit. J. Appl. Phys.*, 8, 387-393.
- Gladwell, G.M.L. (1976), Some unbounded problems in plane elasticity theory, *ASME J. Appl. Mech.*, 43, 263-267.
- Gladwell, G.M.L. (1980) *Contact Problems in the Classical Theory of Elasticity*, Sijhoff and Noordhoff, Amsterdam.
- Greenwood, J.A. and Williamson, J.P.B. (1966), Contact of nominally flat surfaces, *Proc. R. Soc. A*, 295, 300-330.
- Gupta P.K. and Walowit, J.A. (1974) Contact stresses between an elastic cylinder and a layered elastic solid, *ASME J. Lub. Tech.*, 96, 259-257.
- Ham, G., Rubin, C.A., Hahn, G.T., and Bhargava, V. (1988), Elasto-plastic finite element analysis of repeated two-dimensional rolling sliding contacts, *ASME J. Trib.*, 110, 44-49.
- Hamilton, G.M. (1983), Explicit equations for the stresses beneath a sliding spherical contact, *Proc. Instn. Mech. Engrs.*, 197C, 53-61.
- Hamilton, G.M. and Goodman, L.E. (1966), The stress field created by a circular sliding contact, *ASME J. Appl. Mech.*, 33, 371-376.
- Hamilton, G.M. and Moore, S.L., (1971), Deformation and pressure in an elasto-hydrodynamic contact, *Proc. R. Soc.*, 332A, 313-330.
- Hannah, M. (1951) Contact stress and deformation in a thin elastic layer, *Quart. J. Appl. Maths.*, 4, 94-105.
- Hills, D. A., Nowell, D., and Sackfield, A. (1993), *Mechanics of Elastic Contacts*, Butterworth-Heinemann, Oxford.
- Holm, R. (1967), *Electric Contacts*, Springer-Verlag, New York.
- Hosford, W.F. and Caddell, R.M. (1983), *Metal-forming. Mechanics and Metallurgy*, Prentice-Hall, New York.
- Ihara, T., Shaw, M.C., and Bhushan, B. (1986), Finite element analysis of contact stress and strain in an elastic film on a rigid substrate, *ASME J. Trib.*, 108, 527-533 and 534-539.
- Jaffar, M.J. and Savage, M.D. (1988), On the numerical solution of line contact problems involving bonded and unbonded strips, *J. Strain Anal.*, 23, 67-77.
- Johnson, K.L. (1970), The correlation of indentation experiments, *J. Mech. Phys. Solids*, 18, 115-123.
- Johnson, K.L. (1985), *Contact Mechanics*, Cambridge University Press, Cambridge.
- Johnson, W. and Mellor, P.B. (1964), *Engineering Plasticity*, van Nostrand, New York.
- Kalker, J.J. and van Randen, Y. (1972), A minimum principle for frictionless elastic contact with application to non-Hertzian half-space contact problems, *J. Engng. Math.*, 6 193-206.

- Kalker, J.J. (1990), *Three Dimensional Elastic Bodies in Rolling Contact*, Kluwer Academic Publishers, Dordrecht.
- Kannel, J.W. and Dow, T.A. (1986), Analysis of traction forces in a precision traction drive, *ASME J. Trib.*, 108, 403-410.
- Kannel, J.W., Walowit, J.A., Bell, J.C., and Allen, C.M. (1965), The determination of stresses in rolling contact elements, *Trans. ASME J. Lub. Tech.*, 89, 453-465.
- Kendall, K. and Tabor, D. (1971), An ultrasonic study of the area of contact between stationary and sliding surfaces, *Proc. R. Soc. A*, 323, 321-340.
- King, R.B. and O'Sullivan, T.C. (1987), Sliding contact stresses in a two-dimensional layered elastic half-space, *Int J. Sol. Structs*, 23, 581-697.
- Komvopoulos, K. (1988), Finite element analysis of a layered elastic solid in normal contact with a rigid surface, *Trans. ASME J. Trib.*, 110, 477-485.
- Komvopoulos, K. (1989), Elastic-plastic finite element analysis of indented layered media, *Trans ASME J. Trib.*, 111, 430.
- Kragelsky, I.V. (1965), *Friction and Wear*, Butterworths, London.
- Kragelsky, I.V. and Demkin, N.B. (1960), Determination of the true area of contact of elastic solids, *Trans. ASME, Friction and Wear in Machinery*, New York, 14, 30-53.
- Kral, E.R. and Komvopoulos, K. (1997), Three dimensional finite element analysis of subsurface stress and strain fields due to sliding contact on an elastic-plastic layered medium, *Trans. ASME J. Trib.*, 111, 332-341.
- Kulkarni, S., Hahn, G.T., Rubin, C.A., and Bhargave, S. (1991), Elasto-plastic finite element analysis of three dimensional pure rolling contact above the shakedown limit, *ASME J. Appl. Mech.*, 58, 347-353.
- Lee, S.C. and Ren, N. (1994), The sub-surface stress field created by three-dimensional rough bodies in contact with traction, *Tribol. Trans.*, 37, 615-621.
- Love, A.E.H. (1939), Bousinesq's problem for a rigid cone, *Quart. J. Math.*, 10, 161-170.
- Lovell, M.R. (1998), Analysis of contact between transversely isotropic coated surfaces, *Wear*, 214, 165-174.
- Mao, K., Sun, Y., and Bell, T. (1994), Contact mechanics of engineering surfaces: state of the art, *Surface Engng.*, 10(4), 297-306.
- McEwen, E. (1949), Stresses in elastic cylinders in contact along a generatrix, *Philos. Mag.*, 40, 454-463.
- Muskhelishvili, N.I. (1949), *Some Basic Problems of the Mathematical Theory of Elasticity*, 3rd ed., Moscow (English translation, by J.R.M. Radok, Noordhoff, 1953).
- Newcomb, T.P. (1957), Communication, *Proc. Conf. Lubrication and Wear*, Inst. Mech. Eng., London, 837-838.
- Ollerton, E. and Haines, D.J. (1963), Contact stress distributions on elliptical contact surfaces subjected to radial and tangential forces, *Proc. Instn. Mech. Engrs.*, 177, 95-101.
- Ollerton, E. and Morley, J.W.W. (1963), Fatigue strength of rail steel in rolling contact, in *Proc. Symposium on Fatigue in Rolling Contact*, IMech E, London.
- O'Sullivan, T.C. and King, R.B. (1988), Sliding contact stress field due to a spherical indenter on a layered elastic half-space, *Trans. ASME J. Trib.*, 110, 235-240.
- Poritsky, H. (1950), Stresses and deflections of cylindrical bodies in contact, *Trans. ASME J. Appl. Mech.*, 17, 191-201.
- Sackfield, A. and Hills, D. A. (1983a), Some useful results in the classical Hertz contact problem, *J. Strain Anal.*, 18, 101-108.
- Sackfield, A. and Hills, D.A. (1983b), A note on the Hertz contact problem: correlation of standard formulae, *J. Strain Anal.*, 18, 195-201.
- Smith, J.O. and Liu, C.K. (1953) Stresses due to tangential and normal loads on an elastic solid, *Trans. ASME J. Appl. Mechs.*, 20, 157-166.
- Snidle, R.W. and Evans. H.P. (1994), A simple method of elastic contact simulation, *Proc. Insns. Mech. Engrs. Part J*, 208, 291-293.

- Theocaris, P.S. and Stassinakis, C.A., (1978), The elastic contact of two disks by the method of caustics, *Exp. Mech.*, 18, 409-415.
- Thomas, H.R. and Hoersch, V.A. (1930), Stresses due to the pressure of one elastic solid on another, *Univ. of Illinois Engng. Exptal. Station Bulletin No 212*.
- Tian, H. and Saka, N. (1991), Finite element analysis of an elastic-plastic two-layer half-space; normal contact, *Wear*, 148, 47-68.
- Tian, H. and Saka, N. (1991), Finite element analysis of an elastic-plastic two-layer half-space; sliding contact, *Wear*, 148, 261-272.
- Tian, X. and Bhushan, B. (1996), A numerical three-dimensional model for the contact of rough surfaces by variational principle, *Trans. ASME J. Trib.*, 118, 33-42.
- Timoshenko, S.P. and Goodier, J.N. (1951), *Theory of Elasticity*, McGraw-Hill, New York.
- Walowit, J.A. and Pinkus, O. (1982), Effect of friction between cylinders and rubber staves of finite thickness, *Trans. ASME J. Lub. Tech.*, 104, 255-261.
- Webster, M.N. and Sayles, R.S. (1986), A numerical model for the elastic frictionless contact of real rough surfaces, *Trans. ASME J. Trib.*, 108, 314-320.
- Williams, J.A., Dyson, I.N., and Kapoor, A. (1999), Repeated loading, residual stresses, shakedown and tribology, *J. Mater. Res.*, 14(4) 1548-1559.