Chapter 3 Contact Resistance Model with Adhesion between Contact Surfaces

In this chapter, I develop a contact resistance model that includes adhesion between contact surfaces. This chapter is organized as follows. First, the physical origin of adhesion is introduced, with reference to van der Waal forces and metallic bonds. The modeling of adhesion in elastic contact between a sphere and a flat is then discussed in detail, and the well-known JKR and DMT models are described. A model of adhesion in the elasto-plastic and plastic deformation regimes is described. This is followed by a discussion of the unloading and separation of contacts – it is seen that following elasto-plastic or plastic deformation, contacts might separate adhesively at the contact interface, or cohesively, within a contact body, accompanied by material transfer. All of the above work leads to a single asperity, elastic-plastic, load-unload model. I then describe how this can be applied to a multi-asperity model. Specific features of measured contact resistance characteristics are then analyzed with the help of the model. These include the shape of the contact resistance characteristic during loading and unloading, the adherence force, and the evolution of these characteristics as the contact is cycled.

3.1 Physical origin of adhesion

When two surfaces are in contact, there are surface forces at the interface tending to hold the surfaces together (a more detailed treatment of the discussion in this section is available in [Maugis 1999]). In some cases, these forces are mainly due to ionic, covalent or metallic bonds between atoms at the surfaces. However, independent of the nature of
the surface, there are certain weaker, longer-range forces which exist between any pair of surfaces. They are commonly referred to as van der Waals forces and result from electrostatic inter-atomic or inter-molecular interactions due to the permanent, induced or instantaneous dipole moments of the atoms or molecules. The van der Waals interaction potential (or free energy) \( U(r) \) between a pair of atoms or molecules a distance \( r \) apart is of the form

\[
U(r) = -C/r^6. 
\]  

(3.1)

The interaction force between a pair of molecules is therefore given by

\[
F(r) = -dU/dr = -6C/r^7, 
\]

(3.2)

the negative sign indicating that the force is attractive.

The van der Waals interaction between two macroscopic bodies may be found by summing the interaction energies of all the atoms in one body with all the atoms in the other. The van der Waals interaction potential (per unit area) between two flat surfaces a distance \( z \) apart, derived from equation (3.1), is:

\[
U(z) = -A/12\pi z^2, 
\]

(3.3)

where \( A \) is the Hamaker constant, and is associated with the material properties of the two bodies. \( A \) is found to be \( \sim 10^{-19} \) J for most condensed phases (liquids or solids) in air. When the two surfaces are in contact (\( z \sim 2 \) Å), the interaction energy (or adhesion energy, \( w \)) is therefore typically -50 mJ/m\(^2\). Surfaces are normally characterized by the surface energy \( \gamma \), defined as the energy required to create unit area of the surface, and hence equal to half the adhesion energy in magnitude for two identical surfaces.
3.1.1 The Lennard-Jones potential

At separations smaller than the molecular radius, electron clouds of interacting molecules overlap, resulting in a repulsive force. The short-range repulsion and longer-range van der Waals attraction between a pair of atoms or molecules are commonly represented by a semi-empirical formula known as the Lennard-Jones potential:

\[ U(r) = D/r^{12} - C/r^6. \]  \hspace{1cm} (3.4)

Based on this equation, the potential between two parallel surfaces a distance \( z \) apart can be calculated as:

\[ U(z) = -\frac{\pi}{12\pi z_0^2} \left( \frac{Z_0}{z} \right)^2 - \frac{1}{4} \left( \frac{Z_0}{z} \right)^8 \]. \hspace{1cm} (3.5)

where \( A \) is the Hamaker constant, and \( Z_0 \) is the equilibrium distance between the surfaces.

The minimum potential is the adhesion energy, with the magnitude:

\[ U_0 = w = \frac{A}{16\pi z_0^2}. \] \hspace{1cm} (3.6)

The force per unit area between the two plates is given as the derivative of the potential,

\[ \sigma(z) = \frac{A}{6\pi z_0^3} \left[ \left( \frac{Z_0}{z} \right)^3 - \left( \frac{Z_0}{z} \right)^9 \right]. \] \hspace{1cm} (3.7)
Equations 3.6 and 3.7 are plotted in Figures 3.1 and 3.2 respectively, in non-dimensional units. By definition, the force per unit area is negative for separations smaller than \( Z_0 \), and represents an elastic compressive stress. The force per unit area is positive for separations larger than \( Z_0 \), and reaches a maximum value at \( z = Z_m = \sqrt[3]{3} Z_0 = 1.2 Z_0 \), before decreasing at larger separations. Between \( Z_0 \) and \( Z_m \), the force per unit area is regarded as an elastic traction stress, which is noticeably non-linear with respect to \( z \). At separations greater than \( Z_m \), the force is regarded as a non-contact adhesive force. However, the difference is semantic – both forces arise from van der Waals interaction.

### 3.1.2 Adhesion between metals

When two metal surfaces are in contact, the surface forces holding them together are metallic bonds, which are much stronger than van der Waals forces. Metals have surface

![Figure 3.1 Lennard Jones potential (normalized by the equilibrium potential, or the adhesion energy) as a function of separation (normalized by the equilibrium separation \( Z_0 \)).](image)
energies in the range 1-3 J/m\(^2\), while the surface energy of molecular solids, determined by van der Waals forces, is \(\sim 25\) mJ/m\(^2\), as discussed above. However, metallic bonds are short-range forces, effective only at inter-atomic distances, while van der Waals forces are effective over much larger separations, up to several nm. Metal surfaces in atmosphere are usually covered by physisorbed or chemisorbed species which screen the short-range forces. If the adsorbate itself does not form ionic or covalent bonds, the surface energy is smaller, and determined mainly by van der Waals forces. For example, Israelachvilli and Tabor (1972) have shown that clean surfaces of mica (with an intrinsic monolayer of water) have a surface energy of 300 mJ/m\(^2\), while mica covered with a monolayer of stearic acid has a surface energy of 30 mJ/m\(^2\).
Researchers have computed the interaction potential between two surfaces in the range of metallic bonds. Rose et al. (1981) have shown that it is possible to represent the interaction potential of all metals by a single curve by plotting the normalized potential \( U^* = U/U_0 \) as a function of a scaled separation, \( z^* = (z-Z_0)/\lambda \), where \( U_0 \) is the minimum potential or surface energy, \( \lambda \) is the Thomas-Fermi screening length (the characteristic length of metallic bonds, generally of the order of 0.5 Å), and \( Z_0 \) is the equilibrium distance between the surfaces. \( \lambda \) in atomic units is given by:

\[
\lambda = \frac{1}{3} \left( \frac{243 \pi}{64} \right)^{1/6} n^{-1/6}(Z_0),
\]

(3.8)

where \( n(Z_0) \) is the electron density at the equilibrium separation. Ferrante et al. (1982) plotted the normalized potential as a function of \( z^* \) for a number of metal surfaces (reproduced in Figure 3.3). The plot can be fitted to the curve

\[
U(z^*) = -U_0 (1 + \beta z^*) e^{-\beta z^*},
\]

(3.9)

with \( \beta = 0.9 \). By differentiation of the interaction potential, the force per unit area is obtained as:

\[
\sigma(z^*) = U_0 \beta^2 z^* e^{-\beta z^*},
\]

(3.10)
Figure 3.3 Interaction potential between metallic surfaces, obtained empirically by Rose, et al. (1981).

Figure 3.4 Interaction force (per unit area) between metallic surfaces, calculated from the interaction potential given by Rose, et al. (1981).
Using the definition of the elastic modulus as earlier, the theoretical stress can be shown to be \( \sigma_{th} \sim 0.5 \sqrt{\frac{Ew}{Z_0}} \). Equation 3.10 is plotted in Figure 3.4, with the force per unit area normalized with respect to the theoretical stress. Clearly, at large separations (say \( z^* > 5 \)), the interaction potential and force per unit area decay nearly exponentially with separation. Therefore, metallic bonds exert short-range forces relative to van der Waals forces (which only decay as \( \sim z^{-3} \)).

In chapter 1, experimental measurements of metallic surface energies were discussed. As stated there, there is a wide variation in measured results, primarily due to the difficulties associated with getting a perfectly clean surface, and a truly single-asperity contact. However, many authors assume an upper bound of 2-2.5 J/m\(^2\) on the adhesion energy \( w \) for a perfectly clean metallic contact. A value of 2.5 J/m\(^2\) is used here for model calculations. The atomic radius of gold is 1.8 Å – accordingly, the equilibrium separation \( Z_0 \) between two surfaces in contact is assumed to be 3.6 Å. As in the previous chapter, the values of the elastic modulus \( E \) and the hardness \( H \) are assumed to be 91 GPa and 2.2 GPa respectively. Note that, based on these values, the theoretical stress \( \sigma_{th} \) can be calculated to be approximately 13 GPa.

3.2 Adhesive elastic contact and fracture mechanics

The elastic loading, unloading, and eventual separation of a sphere and a flat has been shown by Maugis (1992, 1999) to be analogous to the retreat and advance of an “external circular crack”, which can be more easily visualized as a deeply notched bar. Maugis has
used the fracture mechanics approach to reconcile the two well-known models of adhesive elastic contact, the JKR and DMT models, which will be introduced later. In this chapter, I will follow the same approach. To do this, it is necessary to first provide a very brief introduction to the fundamental concept of crack propagation.

Consider a double cantilever of width $b$ whose arms are kept open by a wedge, such that the ends of the arms are a fixed distance $\delta$ apart (Figure 3.5). It is possible for a crack to form at the junction of the arms; the crack advances by the progressive breaking of bonds at the tip of the crack, like unzipping a zipper. The length of the crack is determined by the availability of the work required to break the bonds. An elastic energy $U_E$ is stored in the bent arms of the cantilever. If the contact area between the arms could decrease by an area $dA$ (that is, the crack could advance a distance $dA/b$), the elastic energy stored in the bent arms would decrease by an amount $(\partial U_E / \partial A)_\delta dA$. Here, the suffix $\delta$ indicates that

Figure 3.5 Double cantilever with arms separated by a wedge.
the perturbation occurs at a fixed position of the wedge, that is, without any displacement of the force exerted by the wedge. Therefore, the potential energy of the system, $U_p$, remains constant. At the same time, the work required to break surface bonds over an area $dA$ is $wdA$, where $w$ is the adhesion energy. The crack can only advance if the decrease in the elastic energy can provide the work required to advance the crack, that is, if $-\left(\frac{\partial U_E}{\partial A}\right) > w$. Similarly, the crack will recede if $-\left(\frac{\partial U_E}{\partial A}\right) < w$. Therefore, the equilibrium position of the crack is reached when $G = -\left(\frac{\partial U_E}{\partial A}\right) = w$. This is a central concept in fracture mechanics. The equilibrium $G=w$ may be stable or unstable, depending on whether $(\partial G/\partial A)$ is positive or negative. If the equilibrium is unstable, crack propagation is initiated at $G=w$.

The derivative of the elastic energy with respect to crack area is called the strain energy release rate. The quantity was introduced by Irwin (1952), and is denoted by $G$ after Griffith, who first brought the concept of surface energy into solid mechanics (Griffith 1920). An equilibrium condition can also be obtained by considering perturbations at a constant load, rather than constant displacement. In this case, $G$ is equal to $-\left(\frac{\partial U_E}{\partial A} + \frac{\partial U_p}{\partial A}\right)$, where the subscript $F$ denotes that the energies change at the constant load $F$. 
3.3 Elastic contact of a rigid cylindrical punch with an elastic flat

The stresses and displacements in an elastic contact between a rigid cylindrical punch and an elastic flat were partially solved by Boussinesq, and solved completely by Sneddon (1946). When the flat end of a rigid cylindrical punch of radius $a$ is pressed with a force $F$ into the surface of a half-space of elasticity $E$, the penetration of the punch is given by:

$$\alpha = \frac{1 - \nu^2}{2E} \frac{F}{a}. \quad (3.11)$$

Within the area of contact, the normal stress is given by:

$$\sigma_z(\rho, 0) = -\frac{F}{2\pi a^2} \frac{1}{\sqrt{1 - \rho^2}}, \quad (3.12)$$

where $\rho = r/a$ is the normalized radial distance from the center of the cylinder. Evidently, the normal stress is compressive within the contact area, and infinite at the contact boundary. The result is physically unrealistic at the contact boundary because of the assumption that there is a sharply defined contact boundary, and that linear elasticity holds everywhere within this boundary, and there are no interaction forces outside the boundary. In reality, there is an annular boundary “zone”. In this zone, the separation between the surfaces increases gradually, the stress-strain relationship becomes progressively non-linear, and the stress changes from compressive to tensile (as in Figures 3.2 and 3.4).

The adherence force (or separation force), for the flat punch was first calculated by Kendall (1971). We consider the situation in which the load $F$ is a traction force, trying to separate the punch from the flat. This can be considered analogous to the fracture of a
deeply notched bar (or equivalently, the propagation of an external circular crack). Since the cylindrical punch is rigid and its contact surface is flat, the contact area is independent of the adhesion force. Equation 3.11, relating the load to the penetration, can therefore be used along with the crack equilibrium condition \( G = \omega \) to determine the adherence force. We assume that the contacting bodies are pulled apart with a constant load \( F \), and the contact area is variable (constrained by the maximum area \( \pi a^2 \)). The potential energy is given by:

\[
U_p = -F \alpha = -\frac{F^2(1 - \nu^2)}{2aE},
\]

and the elastic energy is given by:

\[
U_E = \int F(\alpha) d\alpha = \frac{F^2(1 - \nu^2)}{4aE}.
\]

The strain energy release rate is

\[
G = \left( \frac{\partial U_E}{\partial A} + \frac{\partial U_p}{\partial A} \right)_F = \frac{F^2(1 - \nu^2)}{8\pi a^3 E}
\]

The equilibrium \( G = \omega \) is reached at a traction force \( F_{adh} \), given by:

\[
F_{adh} = \sqrt{8\pi a^3 \omega/(1 - \nu^2)}
\]

The derivative of \( G \) with respect to \( A \) is

\[
\left( \frac{\partial G}{\partial A} \right)_F = -\frac{3F^2(1 - \nu^2)}{16\pi^2 a^5 E} < 0.
\]

Therefore, the equilibrium \( G = \omega \) is always unstable. Now consider that equilibrium is reached at \( F = F_{adh} \), and the contact radius is \( a_1, a > a_1 > 0 \). If there is a small perturbation momentarily increasing \( a_1 \), \( G \) decreases, making \( G < \omega \), and the crack recedes, further
increasing $a_1$, but only until $a_1 = a$. On the other hand, if the perturbation decreases $a_1$, making $G > w$, the crack advances, further reducing $a_1$, until the contact area becomes zero and the contacting bodies separate. Therefore, $F_{adh}$ is the adherence force.

The adherence force $F_{adh}$ can be generalized for the case of contact between an elastic flat punch of elasticity $E_1$ and a half-space of elasticity $E_2$:

$$F_{adh} = \sqrt{6\pi a^3 Kn},$$

(3.18)

where $\frac{1}{K} = \frac{3}{4} \left( 1 - \frac{v_1^2}{E_1} + \frac{1 - v_2^2}{E_2} \right)$ [Maugis 1999].

### 3.4 Elastic adhesion between a sphere and a flat surface

First consider the Hertz model of elastic contact between a sphere and a flat. This model assumes that stresses inside the contact region are compressive, and therefore ignores adhesive forces. The normal stress inside the contact region has a hemispherical distribution (Figure 3.6), with a maximum value at the center, and becoming zero at the boundary of the contact spot:

$$\sigma_z(\rho,0) = -\frac{3}{2}\frac{F}{\pi a^2} \sqrt{1-\rho^2},$$

(3.19)

where $\rho$ is the normalized radial distance from the center of the contact, that is, $\rho = r / a$.

In a real contact, tensile stresses exist inside and outside the contact area, which make the contact area larger than predicted by the Hertz model. I will now describe two well-known models of adhesive elastic contact. In the Johnson-Kendall-Roberts (JKR) model
[Johnson 1971], adhesive forces are assumed to be confined to the contact area, and a new stress distribution and new relationships between the contact force, area and vertical deformation are calculated. In the Derjaguin-Muller-Topolov (DMT) model [Derjaguin 1975], adhesive forces are assumed to exist only in an annular region around the contact spot, and add to the applied contact force, resulting in a larger “effective” load. The Hertz stress distribution is assumed to hold inside the contact area, and the Hertz deformation equations still apply with the applied load replaced by the effective load.

### 3.5 The JKR Model

Let us assume that contact between a sphere of radius $R$ and a flat surface under a contact force $F$, in the presence of adhesive forces acting only within the area of contact, results in a contact radius $a$. Now let us first consider adhesionless contact under an equivalent load $F_1 (F_1 > F)$, such that the contact radius is $a$. That is,

$$a = \left( \frac{F_1 R}{K} \right)^{1/3}. \quad (3.20)$$

The vertical penetration, $\alpha_1$, is therefore,

$$\alpha_1 = \frac{a^2}{R} = \left( \frac{F_1^2}{K^2 R} \right)^{1/3}. \quad (3.21)$$

The stored elastic energy, $U_{E1}$, is given by:

$$U_{E1} = \int F_1 d\alpha_1 = \frac{2}{5} \frac{F_1 a^2}{R} = \frac{2}{5} \frac{a^2 K}{R^2}. \quad (3.22)$$

Starting from load $F_1$, the load is gradually reduced to $F$, while keeping the contact radius constant at $a$, and at the same time, increasing the energy of adhesion from 0 to $w$. Since
the area of contact remains constant, this is equivalent to applying a traction force \( (F_f - F) \) on a flat punch of radius \( a \). Consequently, the vertical penetration decreases by an amount (from equation 3.11):

\[
\Delta \alpha = \frac{2(F_f - F)}{3aK}.
\]  

(3.23)

The stored elastic energy changes by an amount (from equation 3.14):

\[
\Delta U_E = \frac{F_f^2 - F^2}{3aK}.
\]  

(3.24)

Therefore, the final penetration, \( \alpha \), is given by:

\[
\alpha = \alpha_f - \Delta \alpha = \frac{a^2}{3R} + \frac{2F}{3aK},
\]  

(3.25)

and the final stored elastic energy is given by:

\[
U_E = U_E^f - \Delta U_E = \frac{F^2}{3aK} + \frac{a^2K}{15R^2}.
\]  

(3.26)

Using the definition of the strain energy release rate, \( G = (\partial U_E / \partial A)_\delta \), it can be shown that:

\[
G = \frac{(a^3K / R - F)^2}{6\pi a^3K}.
\]  

(3.27)

At equilibrium, \( G = w \). After writing \( a \) in terms of the Hertzian load \( F_f \), equation (3.27) and the equilibrium condition can be used to obtain a quadratic equation in terms of \( F_f \), whose solution is given by:

\[
F_f = F + 3\pi wR \pm \sqrt{6\pi wRF + (3\pi wR)^2}
\]  

(3.28)
An unique solution can be obtained for $F_i$ (and hence for the contact radius $a = \left(\frac{F_i R}{K}\right)^{1/3}$), by imposing the condition for the stability of the $G = w$ equilibrium, 

$$\left(\frac{\partial G}{\partial A}\right)_F > 0.$$  

By differentiating equation 3.27, we obtain:

$$\left(\frac{\partial G}{\partial A}\right)_F = \frac{F_i^2 - F^2}{4\pi^2 a^5 K}. \quad (3.29)$$  

Replacing $F$ in the above equation with the value of $F$ obtained from equation (3.27), and also using the expression $a = \left(\frac{F_i R}{K}\right)^{1/3}$, we obtain:

$$\left(\frac{\partial G}{\partial A}\right)_F = \frac{F_i - F - 3\pi wR}{2\pi^2 a^3 R}. \quad (3.30)$$  

A stable equilibrium is obtained only if $F_i - F - 3\pi wR > 0$. Therefore, from equation 3.28, the contact radius is given by:

$$F_i = \frac{a^3 K}{R} = F + 3\pi w R + \sqrt{6\pi w R F + (3\pi w R)^2} \quad (3.31)$$

The vertical penetration $\alpha$ can be obtained by superposing the penetration for the Hertz model under a compressive force $F$, and the flat punch under a tensile force $F_i - F$. Therefore, from equations 3.21 and 3.11,

$$\alpha = \frac{a^2}{3R} + \frac{2F}{3aK} = \frac{a^2}{R} - \sqrt{\frac{8\pi w a}{3K}} \quad (3.32)$$
The normal stress in the contact area can be obtained by adding the Hertz stress distribution with the normal stress distribution for a flat punch under a tensile force $F_1 - F$. Therefore, from equations 3.19 and 3.12,

$$
\sigma_z(\rho,0) = \frac{F_1 - F}{2\pi a^2} \frac{1}{\sqrt{1 - \rho^2}} - \frac{3}{2} \frac{F}{\pi a^2} \sqrt{1 - \rho^2}.
$$

(3.33)

The above stress distribution is shown in Figure 3.6, along with the stress distribution given by the Hertz model. Evidently, the contact area has a central compressive zone, surrounded by an annular tensile zone (in which adhesive forces act), with the stress becoming infinite at the edge of the contact area. Again, the unphysical result at the contact boundary occurs because of the assumption of a sharp contact boundary with linear elasticity assumed to exist over the entire contact area, and interaction forces assumed to vanish abruptly outside the contact area. As mentioned earlier, in reality there is an annular boundary region in which the separation increases gradually, and the stress-strain relation is non-linear.

The shapes of the deformed sphere and flat are shown schematically in Figure 3.7 (a) (a calculation of the deformed profiles can be found in (Kendall 1971)). For comparison, the profiles obtained in the Hertz model are shown in Figure 3.7 (b). In the Hertz model, the contact bodies meet tangentially at the edge of the contact region. In the JKR model, because of the additional flat punch-like displacement of the contact bodies, the surfaces of the contact bodies are normal to each other at the edge of the contact region, with a “necking” of the sphere, and a corresponding “lip” in the deformed flat.
Equation 3.31 permits a negative contact force $F$. That is, contact persists even under a tensile load, due to the adhesion forces represented by a non-zero $w$. At a critical tensile load $F_{adh}$, at which the contact radius is $a_{\text{min}}$, the right hand side of equation 3.31 becomes zero:

$$
\left( \frac{\partial G}{\partial A} \right)_F = \frac{a_{\text{min}}^3 K / R - F_c - 3\pi w R}{2\pi^2 a_{\text{min}}^2 R} = 0.
$$

(3.34)

The equilibrium becomes unstable, and the contacts separate. By setting $F = F_{adh}$ and $F_1 = \frac{a_{\text{min}}^3 K}{R}$ in equation 3.31, we obtain:
Figure 3.7 Schematic representation of the deformed profiles of the contact bodies in the JKR model (a) and in the Hertz model (b)

\[ F_{adh} = -\frac{3}{2} \pi w R. \]  

(3.35)

And from equation 3.27,

\[ a_{\text{min}} = \left( \frac{3}{2} \frac{\pi w R^2}{K} \right)^{1/3}. \]  

(3.36)

\( F_{adh} \) is numerically equal to the adherence force; clearly, the model predicts that the contact bodies separate abruptly at a non-zero contact radius \( a_{\text{min}} \). Note that for a single asperity of radius \( R=0.1 \) µm, using \( w=2.5 \) J/m\(^2\), the adherence force is \( F_{adh} = 1.2 \) µN, and the contact radius at separation is 11.3 nm.

From equations 3.32 and 3.36, the vertical penetration at contact separation is given by:
\[ a_{\text{min}} = \frac{2}{3R} \left( \frac{\pi^2 w^2 R}{12K^2} \right)^{1/3}. \]  

This ends the discussion of the JKR model. The JKR model gives the contact radius as a function of contact force in the presence of adhesion (equation 3.31), and also gives the relationship between the vertical penetration and the contact radius (equation 3.32). The model also predicts an abrupt separation of the contacts at a critical tensile load \( F_{\text{adh}} \), at a finite contact radius \( a_{\text{min}} \) (equations 3.35 and 3.36).

### 3.6 The DMT Model

The fundamental assumption of the DMT model is that the Hertz stress distribution in the contacting bodies is not altered by the presence of attractive surface forces. Unlike the JKR model, the authors do not neglect forces in the adhesive zone surrounding the contact area. However, they calculate these forces while assuming that the Hertz deformation model still holds.

The authors use a thermodynamic approach to calculate the adherence force (Derjaguin 1975). The net work done in bringing a sphere in contact with a flat is represented as:

\[ U = U_E - U_S, \]  

Here \( U_E \) denotes the stored elastic energy, and \( U_S \) is the work done by adhesive forces in the adhesive boundary zone.
Since the Hertz stress distribution is assumed to hold, the derivative of $U_E$ with respect to the vertical penetration is the Hertz load ($F_E = \frac{\partial U_E}{\partial \alpha} = \frac{a^3 K}{R}$), and the total contact force is therefore related to the contact radius as:

$$F = \frac{\partial U_E}{\partial \alpha} - \frac{\partial U_S}{\partial \alpha} = \frac{a^3 K}{R} - F_s.$$  \hfill (3.39)

The adhesive force, $F_s$, is obtained as follows. The adhesive energy is given by:

$$U_S = \int_0^\infty \phi(z_g(r)))2\pi rdr,$$  \hfill (3.40)

where $r$ is the radial distance from the centre of the contact spot, $z_g$ is the gap between the surfaces of the contacting bodies, and $\phi$ is the surface potential. Within the contact area ($r \leq a$), $z_g$ is assumed to be constant (roughly equal to the inter-atomic spacing, denoted by $Z_0$). Outside the contact area, $z_g$ is given by the surface profiles determined from the Hertz model. Derjaguin (1975) shows this to be:

$$z_g(r) = \frac{1}{\pi R} \left( a \sqrt{r^2 - a^2} + r^2 \tan^{-1} \frac{\sqrt{r^2 - a^2}}{a} \right) + Z_0.$$

$\phi$ is assumed to have a form such as the Lennard Jones potential (equation 3.5). Denoting the two components of adhesive energy (within and outside the contact spot of radius $a$) by $U'_s$ and $U''_s$ respectively, the total adhesive force can be written as:
\[ F_s = \frac{dU^I_s}{d\alpha} + \frac{dU^\|_s}{d\alpha} = F_s^I + F_s^\|. \]  

(3.42)

where \( \alpha \) is the vertical penetration, \( \frac{a^2}{R} \). Since

\[ U^I_s = \pi a^2 \varphi(Z_0) = \pi a^2 w, \]  

(3.43)

\( F_s^I \) is given by:

\[ F_s^I = \pi Rw. \]  

(3.44)

That is, \( F_s^I \) is independent of the vertical penetration and therefore independent of the contact force. Irrespective of the specific form of the potential function \( \varphi \), it can be shown that when contact separation is approached, that is, \( \alpha \to 0 \), \( F_s^\| \to \pi Rw \) [Derjaguin 1975]. If \( \varphi \) is assumed to be the Lennard Jones potential, \( F_s^\| \) is seen to approach zero as \( \alpha \) increases. Therefore, the adherence force is \( 2\pi Rw \), and at large contact forces, the adhesive force tends to \( F_s = \pi Rw \). Unlike the JKR model, there is no hysteresis in the loading/unloading characteristics. As the contacts are unloaded, \( a \) and \( \alpha \) progressively decrease until, at a tensile load of \( 2\pi Rw \), a point contact is reached.

While the predicted adherence force has been experimentally verified for contact geometries in the DMT regime, the predicted adhesion force at higher contact loads is not reliable. This is because of the ad hoc assumption of the Hertz profiles for the deformed bodies, which is clearly not consistent with the presence of adhesive forces not taken into account by the Hertz model.
3.7 Validity of JKR and DMT models

As discussed earlier, the JKR model neglects adhesive forces outside the contact area, while the DMT model ignores the tensile stress within the contact area. The stress distributions obtained in the two models are shown schematically in Figures 3.8 (a) and (b). Accordingly, the JKR model is accurate when the non-linear boundary zone is small compared to the contact area; the DMT model becomes more accurate when the boundary zone becomes larger. The transition was first quantified by Tabor (1976), for a lightly loaded contact \( F \ll F_1 \), the apparent Hertz load, as follows. Consider a simplified representation of the deformation profiles, in which the flat is assumed to be rigid, and the sphere has the equivalent elastic modulus \( K \). The more pronounced the “neck” region of the sphere, the larger is the separation between the two surfaces immediately outside the contact region, and the smaller the extent of the annular non-contact adhesive region. In the JKR model, the height of the neck region can be approximated as the flat punch displacement introduced to account for adhesion within the contact region. Denoting the height of the contact as \( h \), equation 3.23 can be re-written as

\[
h = \frac{2(F_i - F)}{3aK}.
\]

(3.45)

Since \( F \ll F_i \), we assume \( F = 0 \), and \( a \approx a_{\text{min}} \). Substituting \( F_i = \frac{a_{\text{min}}^3K}{R} \) (the Hertz model), we obtain

\[
h \approx 2 \left( \frac{Rw^2}{K^2} \right)^{1/3}.
\]

(3.46)
If $h$ is large compared to the characteristic range over which adhesion forces are active (in the case of metallic bonds, the Thomas-Fermi screening length $\lambda$, here assumed to be equal to 0.5 Å), the boundary zone can be neglected, and the JKR model is applicable. Conversely, when $h << \lambda$, the neck disappears, and the DMT model becomes applicable. Note that equation 3.46 indicates that the JKR model is likely to be applicable in the case of relatively elastic solids, large asperity radius, and a large energy of adhesion. In our case, using the values $R=0.1$ μm, $w=2.5$ J/m$^2$, $E=91$ GPa and $\nu=0.5$ (therefore $K=81$ GPa), we obtain $h \approx 18\lambda$, which indicates that we are in the JKR regime. The JKR model is used in the elastic deformation regime (load and unload) in subsequent model calculations.

More general solutions (applicable in both the JKR and DMT regimes) have been given. Muller et. al. (1992) have given a numerical solution. This solution uses the Lennard Jones function to represent the surface potential, and calculates the deformed profiles and stresses in a self-consistent manner.

Maugis (1992) provided a solution based on an analogy with the Dugdale model of crack propagation [Dugdale 1960]. There are long range adhesive forces on the exposed surfaces near the tip of a crack, and localized plastic yielding at the tip of the crack, similar to the boundary zone in the flat punch and sphere-flat contacts. These effects are neglected in classical fracture mechanics, causing a singularity in the modeled stress at the crack tip. In the Dugdale model of crack propagation, these effects are accounted for by a zone of finite width at the crack tip with a constant tensile stress $\sigma_0$. This is applied
by Maugis to the flat punch component of the normal stress distribution in the JKR model, with $\sigma_0$ assumed to be equal to the theoretical stress. The stress distribution is shown schematically in Figure 3.8(c). The width of the non-contact adhesive zone is set so as to cancel the singular term in the stress distribution of equation 3.33. The resulting model agrees with the JKR and DMT models in the respective regimes, and predicts a transition between the two models.

3.8 Elasto-plastic and plastic adhesion between sphere and flat surface

We use the approach of Maugis (1984), which is based on deformation models given by Studman (1976) and Roy Chowdhury and Pollock (1981). In the elastic regime, the deformation is assumed to be governed by the JKR model (equations 3.31 and 3.32). The criterion for the elastic-to-elasto-plastic transition is assumed to be the same as in the adhesionless case. That is, the transition is assumed to occur at a contact radius $a_e$ at which the “Hertzian” load ($= \frac{a^3K}{R}$) becomes equal to the load given by the modified Studman elasto-plastic model (equation 2.6). Therefore, using the values $R=0.1$ $\mu$m, $H=3Y=2.2$ GPa, and $E=91$ GPa, we have $a_e = 3.9 \frac{RY}{E} = 3.1$ nm in our case. Note that this is smaller than the minimum stable contact radius under the elastic JKR model, $a_{min}=11.3$ nm (from equation 3.36), which implies that when 0.1 $\mu$m radius asperities are being loaded, contact is always either elasto-plastic or purely plastic.
In the elasto-plastic and plastic regimes, the stress distribution and contact geometry are assumed to be those for a purely plastic contact between a sphere and flat, with no adhesion. We define a quantity $p_m$, as the contact pressure corresponding to the compressive component of the stress distribution. The contact pressure $p_m$ is assumed to vary with load according to the modified version of the Studman model given by Maugis.

Figure 3.8 Normal stress profile in the JKR model (a), the DMT model (b), and Maugis’ model based on the Dugdale transition (c).
(1984), defined in the adhesionless model of the previous chapter. Re-writing equation 2.6,
\[ p_m = Y(1.1 + 0.7 \ln \frac{Ea}{3.9YR}). \]  
(3.47)

We denote the work corresponding to the compressive contact pressure \( p_m \) by \( U_{PL} \). The work done by adhesive forces is assumed to be \( U_S = \pi a^2w \); that is, adhesive forces outside the contact area are neglected. Therefore, the net work required to create a contact area of radius \( a \) can be written as:
\[ U = U_{PL} - U_S = \int_0^a \pi a^2 p_m d\alpha - \pi a^2w. \]  
(3.48)

Since the contact geometry is assumed to be that for a purely plastic contact without adhesion, \( \alpha = \frac{a^2}{2R} \), and the contact force is related to the contact area as:
\[ F = \frac{\partial U_{PL}}{\partial \alpha} - \frac{\partial U_S}{\partial \alpha} = \pi a^2 p_m - 2\pi Rw. \]  
(3.49)

Eventually, the contact pressure becomes equal to the hardness \( H (=3Y) \) – from equation 3.47, this occurs at a critical contact radius \( a_p \approx 60YR/E \) (with \( R=0.1 \) \( \mu \)m, \( H=3Y=2.2 \) GPa, and \( E=91 \) GPa, we have \( a_p=49 \) nm). Thereafter, the deformation is purely plastic, and the contact pressure remains constant. In the purely plastic regime,
\[ F = \frac{\partial U_{PL}}{\partial \alpha} - \frac{\partial U_S}{\partial \alpha} = \pi a^2 H - 2\pi Rw. \]  
(3.50)
In our case, we cannot assume the geometrical relationship $\alpha = \frac{a^2}{2R}$ since, as in the adhesionless contact model, the relationship between $\alpha$ and $a$ in the plastic regime should have continuity with the $\alpha$-$a$ relationship in the elastic regime (that is, equation 3.32 of the JKR model). As the load increases, the effect of the adhesion force becomes progressively smaller. Therefore, as $\alpha$ increases, the $\alpha$-$a$ relationship should tend towards $\alpha = \frac{a^2}{2R}$. Based on these considerations, we assume the following relationship between $\alpha$ and $a$ in the elasto-plastic and plastic regimes:

$$\alpha = \frac{a^2}{2R} + \frac{\alpha_e}{2} - \frac{1}{2} \sqrt{\frac{8\pi a_e w}{3K}},$$

(3.51)

where $\alpha_e$ is the critical value of the vertical penetration at which the elastic-to-elasto-plastic transition occurs, and $a_e$ is the corresponding critical contact radius. Since the additional terms introduced in the above equation are independent of $\alpha$ and $a$, equations 3.46 and 3.48 still apply in relating the contact radius to the contact force.

Since the effect of the adhesive forces on the stress distribution decreases progressively in the plastic regime, the recovered radius of curvature of the asperity, $R_{eff}$, is assumed to be the same as it would be in the absence of adhesion (Johnson (1977)). Therefore, following equation 2.8,

$$R_{eff} = \frac{a_j^3 K}{\pi a_j^2 p_m} = \frac{a_j K}{\pi p_m}$$

(3.52)

in the elasto-plastic regime, and
\[ R_{\text{eff}} = \frac{a_f K}{\pi H} \quad (3.53) \]

in the plastic regime.

Figure 3.9 shows the modeled variation of the contact spot radius with load, using the JKR model and the elasto-plastic / plastic deformation model of this section, for a single asperity of radius 0.1 \( \mu \text{m} \). The contact radius variation in the absence of adhesion is shown for comparison. Figure 3.10 shows the modeled variation of the deformed asperity radius with load. Clearly, at loads less than 0.1 \( \mu \text{N} \), the contact radius and deformed asperity radius are determined mainly by the surface energy, and are relatively independent of the load. At higher loads, adhesion becomes less important, and both quantities approach the values predicted by the adhesionless model.
Figure 3.9 Modeled variation of contact spot radius with load when a single 0.1 micron asperity is pressed into a flat surface, in the presence and absence of adhesion. The first three legends correspond to different segments of the lower (adhesionless) characteristic, the last two legends correspond to segments of the upper (adhesive) characteristic. Note that adhesive contact is elasto-plastic even at zero load. At high loads, the effect of adhesion is small – therefore, the purely plastic segments of the two characteristics nearly coincide.

3.9 Unloading and separation of contact bodies

If the contact only deforms elastically during loading, the loading is reversible, and the contact radius and vertical penetration follow the JKR model (equations 3.31 and 3.32) during unloading, until the adherence force $F_{adh}$ is reached, at a contact radius $a_{min}$, and vertical penetration $\alpha_{min}$, and the contacts separate abruptly.
Figure 3.10 Modeled variation of effective asperity radius $R_{eff}$ as a function of the maximum load on the first cycle. Initial asperity radius is 0.1 micron.

If the loading is elasto-plastic or plastic, three different unloading modes are possible [Maugis 1984]. When the contact is gradually unloaded starting from a maximum contact force $F_f$, the initial unloading is equivalent to the elastic unloading of a flat punch. Therefore, in this regime, the contact radius remains constant ($a=a_f$), and (following equation 3.11) the vertical deformation varies as:

$$\alpha - \alpha_f = \frac{3(F - F_f)}{2K a_f}.$$  \hspace{1cm} (3.54)

This initial regime persists until one of the following three conditions is met:

a) The average (tensile) axial stress becomes equal to $H$ ($F = F_d = -\pi a_f^2 H$) – the contacts separate at this force, and the separation is “ductile”, similar to the rupture of a deeply notched bar, and may occur within one of the contact bodies rather than at the interface.
b) The contact force reaches a value \( F_m \) at which the JKR equilibrium condition is met

\[
F = F_m = \frac{a_f^3 K}{R_{eff}} - \sqrt{6\pi a_f^3 K w},
\]

and the contact radius \( a_f \) is smaller than the minimum stable contact radius under the JKR regime,

\[ a_{\text{min}} = \left( \frac{3}{2} \frac{\pi w R_{eff}}{2K} \right)^{1/3}. \]

Since this equilibrium is unstable, the contacts separate. The separation mode is similar to that described in the flat punch and JKR models, that is, it occurs at the interface, and is similar to the propagation of an external circular crack. This is commonly referred to as “brittle” separation.

c) The contact force reaches the JKR equilibrium value \( F_m \), and \( a_f > a_{\text{min}} \), therefore the equilibrium is stable. Subsequent unloading of the contact follows the JKR model, and the contact radius decreases progressively with contact force until the JKR equilibrium becomes unstable (\( F = F_b = -\frac{3}{2} \pi w R_{eff} \), \( a = a_{\text{min}} \)), and the contacts separate. Again, the separation is brittle.

We refer to the above three modes by the symbols representing their respective adherence forces, \( F_d, F_m \) and \( F_b \). Evidently, the separation mode that actually occurs is the one with the smallest (absolute) adherence force, provided the geometric conditions for that mode are satisfied. Therefore, the range of final contact radius \( a_f \) for each mode to occur can be shown to be as follows:
**Ductile (F\textsubscript{d}) mode:** following elasto-plastic loading - \( a_f < \frac{6Kw}{\pi(H + p_m)^2} \); following plastic loading - \( a_f < \frac{3Kw}{2\pi H^2} \).

**Brittle (F\textsubscript{m}) mode:** following elasto-plastic loading - \( \frac{6Kw}{\pi(H + p_m)^2} < a_f < \frac{3Kw}{2\pi p_m^2} \); not possible after plastic loading.

**Brittle (F\textsubscript{b}) mode:** following elasto-plastic loading - \( a_f > \frac{3Kw}{2\pi H^2} \); following plastic loading - \( a_f > \frac{3Kw}{2\pi H^2} \).

In general, ductile separation is more likely to occur at small loads, and in solids with high elastic modulus, low hardness and high surface energy. Figure 3.11 shows the variation of adherence force and separation mode as a function of the final load \( F_f \), for a single asperity with original radius of curvature 0.1 \( \mu \)m. As the load increases, the separation mode changes from \( F_{d} \) to \( F_{m} \), and then to \( F_{b} \).

Figure 3.12 shows the variation of contact radius corresponding to a single asperity as the asperity is unloaded. The three unloading curves correspond to three different values of the final load \( F_f \). Unloading from \( F_f=0.7 \) \( \mu \)N results in ductile separation, unloading from \( F_f=1.9 \) \( \mu \)N results in \( F_{m} \) separation, and unloading from \( F_f=4 \) \( \mu \)N results in \( F_{b} \) separation. The figure shows that in the first two modes, the contact radius remains constant until contact separation, while in the third mode, the contact radius first remains constant and then decreases by a relatively small amount before contact separation.
3.10 Subsequent load-unload cycles

In the deformation model without adhesion, we used the commonly-made assumption that plastic deformation only occurs on the first load cycle, and subsequent loading to the same maximum load results in the same contact radius by elastic (Hertzian) deformation. The deformed asperity radius was calculated based on this assumption. In the elasto-plastic/plastic deformation models with adhesion, the assumption is less straightforward – the deformed asperity radius is such that if the stress distribution under maximum load on a subsequent load cycle is approximated as Hertzian, the mean pressure is equal to the compressive pressure under elasto-plastic/plastic deformation on the first load cycle. This model does not result in purely elastic deformation on subsequent load cycles. Instead, applying the same load on a second cycle results in more elasto-plastic or plastic deformation, and a larger contact and asperity radius than on the previous load cycle. Over multiple load cycles, the model would predict sufficient flattening of the asperity that eventually subsequent loading becomes elastic and reversible. However, it is not clear that this is physically realistic, particularly since the model is not intended to give an accurate representation of the deformed asperity after plastic deformation. In general, there does not appear to be published work on analytical modeling of repeated loading-unloading of asperities in the presence of adhesion.
Figure 3.11 Modeled variation of adherence force and separation mode with ultimate contact force $F_f$ on a single asperity.

Figure 3.12 Modeled variation of contact radius with contact force while loading and unloading a single asperity ($R=0.1 \mu m$). The curved line represents the loading curve, while the three nearly flat lines represent unloading after loading to three different contact load values.
There are other concerns in predicting the deformed asperity radius after a load-unload cycle. Clearly, if the separation of the contacts is ductile, and therefore does not necessarily occur at the contact interface, the asperity shape defined by the maximum load may not be preserved after the contacts separate. Also, if the interface is perfectly clean and consists of metallic bonds (as assumed in model calculations, with $w=2.5 \text{ J/m}^2$), it essentially “disappears”, or becomes indistinguishable from the bulk. Therefore, even if the contact separation is predicted to be brittle, the contacts may not separate precisely along the interface.

Accordingly, I do not attempt to model a subsequent load cycle, based on plastic deformation on the previous cycle, as was done in the model without adhesion. However, at the end of this chapter, I present a somewhat qualitative discussion of changes in measured characteristics with cycling of the contacts. This includes a discussion of changes predicted by the model with changes in two model parameters, namely the asperity radius, and the roughness parameter $\sigma$.

3.11 Multiple asperity model with adhesion

The adhesive deformation model developed in this chapter can now be applied to the multiple-asperity system of the previous chapter. Consider $N$ asperities, each with an end radius of curvature $R$, and heights $z_1 > z_2 > ... > z_N$, and let the separation between the reference planes be $d$ for a given contact force $F$, such that $z_n > d > z_{n+1}$. Asperities $1,2,...,n$ come into contact, and the vertical penetration of asperity $i$ is given by
\[ \alpha_i = z_i - d. \]  

(3.55)

For a given \( \alpha_i \), the contact radius \( a_i \) is given by the appropriate model: elastic (JKR), elasto-plastic, or plastic. However, as mentioned earlier, any finite contact with 0.1 \( \mu m \) asperities results in elasto-plastic or plastic deformation, so that equation 3.51 applies. The force required to produce this deformation, \( F_i \), is given by equations 3.49 and 3.50 respectively in the elasto-plastic and plastic regimes. The total contact force is the sum of the loads on individual asperities, \( F = \sum_{i=1}^{n} F_i \). For a given separation between the reference planes, the force on each asperity, and the radius of each of the corresponding contact spots can be obtained using the previously stated equations. As in Chapter 2, model calculations are performed for two different surfaces, each covered with 100 0.1 micron asperities; the smoother surface has an asperity height distribution with standard deviation (\( \sigma \)) = 0.01 \( \mu m \), and the rougher surface has an asperity height distribution with standard deviation (\( \sigma \)) = 0.1 \( \mu m \). Figure 3.13 shows the number of contact asperities, and the radii of a few of the asperities, plotted as a function of increasing contact force. For the smoother surface, most asperities are in elasto-plastic contact between 10-300 \( \mu N \); for the rougher surface, both elasto-plastic and plastic contact occurs in this force range.
Figure 3.13 Modeled variation of number of contacting spots, and contact spot radii with contact force for a surface with roughness $\sigma=0.01$ micron ((a) & (b)), and $\sigma=0.10$ micron ((c) & (d)). The initial asperity radius is 0.1 micron.

After the contact force has reached a desired maximum, $F = F_f$, the unloading and separation of the asperities have to be modeled, while increasing the separation between the reference planes, $d$, in stages. If an asperity has only undergone elastic deformation, its contact radius and load are governed by the JKR equations 3.31 and 3.32 during unloading. If an asperity has been plastically deformed, it has a new effective radius of curvature, $R = R_{eff,j}$, given by equation 3.52 or 3.53. As discussed in section 3.10, the
separation mode of the asperity can be determined based on the final contact radius. In the case of ductile and $F_m$ modes, the force on the asperity is given by equation 3.54, while the contact radius remains constant. When the appropriate adherence force ($F_d$ or $F_m$) is reached, contact is lost at the asperity. In the case of the $F_b$ mode, unloading of the asperity is initially similar to the $F_m$ mode until an unloading force $F=F_m$ is reached; unloading beyond this point follows the JKR model, with the unloading force and contact radius given by equations 3.31 and 3.32, until the unloading force reaches the JKR adherence force $F = F_b = -\frac{3}{2}\pi w R_{eff}$. 

3.12 Model calculations and discussion

Figure 3.14 shows the model prediction for the number of asperities in contact, and their modes of separation while unloading from two different maximum loads – 70 µN, and 170 µN. For the rougher contact surface ($\sigma=0.1$ µm), all the asperities separate in $F_b$ mode. For the smoother surface ($\sigma=0.01$ µm), a small number of asperities (the shortest ones, bearing the smallest loads) separate first either in ductile or $F_m$ mode. In all cases, a number of asperities sometimes lose contact simultaneously at the same unloading force – this is particularly noticeable at the final contact separation in the $\sigma=0.01$ µm case. This is the result of a domino effect: when a single asperity loses contact, the unloading force that was acting on it is re-distributed among the remaining asperities, which may cause another asperity to lose contact, and so on.
Figure 3.14 Modeled variation of number of contacting spots while unloading from maximum loads of 70 µN and 170 µN respectively with surface roughness $\sigma=0.01 \mu m$ ((a) & (b)), and $\sigma=0.1 \mu m$ ((c) & (d)). For the rougher surface, all the contacting asperities separate in brittle ($F_b$) mode.

The model can now be used to calculate the contact resistance characteristic of a contact interface with multiple asperities. In Figure 3.15, modeled characteristics are for contacts with $\sigma=0.01 \mu m$, and $R=0.1 \mu m$, being loaded up to 1 mN. Comparison with the model
Figure 3.15 Modeled contact resistance as a function of load, in the presence of adhesion ($w = 2.5 \text{ J/m}^2$), and without adhesion. The rough surface has 100 asperities with $R = 0.1 \mu\text{m}$ and $\sigma = 0.01 \mu\text{m}$.

Without adhesion shows, as expected, that adhesion results in a significantly lower contact resistance at small contact loads, say less than 10 $\mu\text{N}$.

Figure 3.16 shows the contact resistance during a load-unload cycle, with the same model.
parameters, and with maximum loads of 70 µN and 170 µN. The contact resistance remains nearly constant while unloading, within a few mΩ. This is because contacting asperities all undergo elasto-plastic or plastic deformation, so that the radii of the contact spots remain constant during the flat punch-like unloading (in the $F_d$ or $F_m$ mode), or nearly constant (in the $F_b$ mode).

We now look more closely at the effects of the important model parameters – $\sigma$ and asperity radius $R$ – on specific aspects of the contact resistance characteristics – the number of asperities in contact, the contact resistance bounds at a given load, the adherence force, and the sensitivity of contact resistance to contact force while loading and unloading.

Figure 3.17 Predicted number of asperities in contact, at a load of 70 µN. The same calculations are plotted as a function of the roughness parameter $\sigma$ in (a), and as a function of the asperity radius $R$. The contact surface has 100 asperities in each case.
3.12.1 Number of contact spots

The number of asperities in contact at a load of 70 µN is plotted over a range of values of $\sigma$ and $R$ in Figure 3.17. Expectedly, the number of contacting asperities decreases with increase in roughness. This number also decreases as $R$ increases, because the vertical penetration or flattening of individual asperities at a given load decreases as $R$ increases. Based on SEM images, the smallest $\sigma$ used in these calculations is 0.003 µm (3 nm). SEM images also show that at least 5-10 asperities are in contact at 70 µN. Therefore, the calculations indicate that $\sigma$ is physically unlikely to be much larger than 0.1 µm. Also, $R$ is probably no greater than 1 µm, since only 7 asperities are predicted to be in contact when $R=1$ µm and $\sigma=0.003$ µm.

![Figure 3.18 Predicted adherence force at a load of 70 µN. The same calculations are plotted as a function of the roughness parameter $\sigma$ in (a), and as a function of the asperity radius $R$. The contact surface has 100 asperities in each case.](image-url)

3.12.2 Contact adherence

The effects of $\sigma$ and $R$ on the adherence force are shown in Figure 3.18, once again assuming an ultimate contact force $F_f=100$ $\mu$N. $\sigma$ is seen to have a strong effect. There are two reasons for the higher adherence of smoother surfaces. For a single asperity, the dependence of the adherence force on the ultimate contact load is much weaker than a linear dependence (very approximately, $F_{adh} \sim F_f^{1/3}$, if we ignore surface forces while loading, and assume brittle separation). Therefore, a larger number of contacting asperities at a given maximum load results in a larger adherence force. Secondly, during unloading of contacts with rougher surfaces, when the shortest asperities are separating, the tallest asperities (which suffer the most deformation) are still under compressive load, and “help” the short asperities to separate by pushing the contact surfaces apart. This effect becomes more pronounced with higher values of maximum load. If the contact surfaces are smooth, nearly all the asperities are under tensile load during separation.

The effect of $R$ on the adherence force is evidently weak, and the observed variation with $R$ appears to be random. This is due to the fact that different numbers of asperities are in contact at different values of $R$, and therefore the surface roughness is effectively somewhat different at each value of $R$, even though the roughness parameter $\sigma$ is constant.
3.12.3 Contact resistance

The dependence of the contact resistance lower and upper bounds on $R$ and $\sigma$ is shown in Figure 3.19. The upper bound depends only on the total contact area, while the lower bound also depends on the number of asperities in contact – for a given total contact area, the larger the number of contacting asperities, the smaller is the resistance. As surface roughness decreases, the total contact area at a given load increases. This is due to the same reason discussed earlier – with a smoother surface and a large number of contacting lightly loaded asperities, surface forces have a relatively large effect on the contact area at each asperity. Therefore, the upper bound on the contact resistance decreases as surface roughness decreases. Since the number of contacting asperities also increases as surface roughness decreases, the lower bound decreases more sharply than does the upper bound.

Figure 3.19 Predicted contact resistance bounds at a load of 70 $\mu$N. The same calculations are plotted as a function of the roughness parameter $\sigma$ in (a), and as a function of the asperity radius $R$. The contact surface has 100 asperities in each case.
Figure 3.19 (b) shows that the upper bound decreases with increase in $R$. This is because most asperities are in the elasto-plastic regime, in which the contact spot size for a given load on an asperity increases with $R$; consequently, the total contact area increases with $R$. However, the lower bound has the opposite trend with respect to $R$ – in this case, the number of contact spots decreases with increase in $R$, and this is a stronger effect than that of the increase in the total contact area. In all cases, the bounds approach each other as the number of asperities in contact decreases.

3.12.4 Sensitivity of resistance to force while loading

Finally, we look at the effects of $R$ and $\sigma$ on the “shape” of the contact resistance characteristic while loading. Figures 3.20 and 3.21 show that at large values of $\sigma$ and $R$, the contact resistance is less sensitive to force while loading – that is, the resistance is relatively small at small loads, and then changes relatively little with load. First consider the effect of roughness (Figure 3.20). At small loads, two surfaces with different levels of roughness have similar (small) numbers of asperities in contact (the tallest asperities), and therefore have similar contact resistance. As the load increases, the smoother surface has more asperities in contact, and also a larger total contact area, as discussed earlier; therefore, its contact resistance decreases more sharply. Now consider the effect of the asperity radius (Figure 3.21). At small loads, surfaces with different asperity radii each have a few asperities in light contact. Since the asperities are lightly loaded, surface forces have an important effect on the contact area; surface forces are higher at blunt asperities, and therefore the surface with blunt asperities has lower contact resistance. As
Figure 3.20 Modeled contact resistance characteristics with two different values of $\sigma$, $\sigma=0.1$ $\mu$m (a), and $\sigma=0.003$ $\mu$m (b). In each case, the contact surface has 100 asperities with $R=0.1$ $\mu$m.

As the contact force increases, the effect of surface forces decreases, and the two surfaces have similar contact resistance.

3.13 Measurements and discussion
At the end of chapter 2, hysteresis in the measured load-unload contact resistance characteristic of a microswitch was used to motivate the study of adhesion. The measured characteristic of Figure 2.19 is reproduced here in Figure 3.22 (a). During unloading, actuation voltages smaller than the threshold voltage correspond to a tensile unloading force, and the voltage at which the switch opens corresponds to the contact adherence force. The contact force model of Chapter 2 (Figure 2.16) can also be used to determine the contact force while unloading (including the adherence force, at which the contact separate); it results in a negative contact force for actuation voltages smaller than the threshold voltage. The modeled variation of contact force with gate voltage of Figure 2.17 is re-plotted in Figure 3.23, with the negative portion of the force characteristic included. Using this model, the measured characteristic of Figure 3.22 (a) is plotted in terms of the contact force in Figure 3.23 (b). This figure shows that the contact was

![Figure 3.21 Modeled contact resistance characteristics with two different values of asperity radius, R=1 μm (a), and R=0.1 μm (b) (re-plotted from Figure 3.16). In each case, the contact surface has 100 asperities with σ=0.01 μm.](image-url)

95
Figure 3.22 Contact resistance of a microswitch, measured as a function of gate voltage while loading and unloading (a), and plotted as a function of contact force (b). 

loaded up to 70 μN, and when unloaded, opened at a (tensile) force of -22 μN. The corresponding adherence force predicted by the models are –31 μN with σ=0.01 μm and R=0.1 μm, and –15 μN with σ=0.1 μm and R=0.1 μm. After the switch turns on, the measured resistance decreases slowly with increasing load, from 0.5 Ω to 0.25 Ω. The resistance while unloading is less sensitive to the load than while loading, but not as insensitive as predicted by the models. There is reasonable agreement with the σ=0.01 μm, R=0.1 μm model (Figure 3.16 (a)) in terms of the loading segment of the characteristic, and the adherence force.
However, comparison of the models with measured contact resistance characteristics is complicated by large variations in the measured characteristics over the lifetime of the switch, which become evident if the measurement is repeated at intervals while cycling the contacts. Repeating in part the discussion of measurements at the end of Chapter 2, the load-unload characteristic changes as follows as a switch is cycled (see Figure 3.24): the switch turns on more abruptly, with a sharper knee in the loading segment; the contact resistance becomes smaller and less sensitive to the contact force; and the adherence force increases – commonly, switches with a mechanical restoring force of 70 $\mu$N per contact stick permanently closed when cycled $10^4$-$10^6$ cycles with 50 $\mu$N or greater contact force. Another feature visible in Figure 3.24 is that the contact resistance continues to decrease somewhat in the initial portion of the unloading curve, before starting to increase. This is the result of a time-dependent effect: when the switch is held

![Graph](image)

*Figure 3.23 Modeled variation of contact force with gate-to-source actuation voltage for the microswitch geometry of Figure 2.14. Actuation voltages below the threshold voltage of 63 V correspond to a negative, or adhesive, contact force.*
Figure 3.24 Measured contact resistance while loading and unloading a single switch: the measurements were made while actuating the switch for the first time (a), and after the switch was cycled 10, 1000 and 10000 times ((b), (c) and (d)). The switch was cycled at an actuation voltage of 75 V (contact force of 30 µN) and current of 4 mA.

closed with a constant contact force, the contact resistance decreases with time, typically by 10-20% over tens of seconds, or a few minutes. The mechanism responsible for this is unknown. However, electromigration is a strong possibility – this is briefly discussed in Chapter 4.
The evolution of the measured characteristics was studied quantitatively, specifically to
determine the effects of the contact force and the number of cycles on the evolution of
the contact resistance and the adherence force. This study was performed as follows. A
switch was subjected to 3 consecutive load-unload cycles of the type shown in Figure
3.22, each time being loaded to a different maximum force. The switch was then cycled
(turned on and off without making resistance measurements); at logarithmic intervals, the
cycling was halted, and the set of 3 load-unload cycles was repeated – the order of the
three cycles was randomized in each set. Data was collected from a group of 5 switches
from a single wafer (Table 3.1). The maximum loads on the three load-unload cycles
were approximately 10 µN, 70 µN, and 170 µN.

A linear regression analysis of the minimum contact resistance on all load-unload cycles,
as a function of the logarithm of elapsed switch cycles, maximum load during load-
unload cycle, and switch number (which represents unaccounted factors), shows that only
the number of cycles is statistically significant (Figures 3.25 (a) and (b), the effect of
switch number is not shown). Linear regression analysis of the contact adherence as a
function of logarithm of elapsed cycles, maximum load during load-unload cycle, and
switch number shows a statistically significant dependence on all 3 factors, with the
effect of the elapsed cycles stronger than the other two factors (Figure 3.25 (c) and (d)).
There is also a reasonably strong correlation between the minimum contact resistance and
the adherence force over all data points, with a correlation coefficient of 0.58.
Table 3.1 Load unload data for a group of 5 switches from a single wafer

Clearly, contact resistance (and its sensitivity to the contact force) and contact adherence are affected much more strongly by cycling the contacts up to $10^5$ cycles than by varying the contact force from 10 $\mu$N to 170 $\mu$N. Broadly speaking, cycling can cause two kinds of changes to the contacts. First, the topology of the surface may evolve with cycling, that is, the roughness of the surface, and the radii of the asperities may change. Second, the surface properties (specifically, the surface energy) may change with cycling.
Figure 3.25 Linear regression analysis of the data of Table 1. The solid lines in Figures 3.25(a) and (b) are the leverage plots of the minimum contact resistance (in Ω) as a function of the log of number of cycles, and the maximum contact force (μN), respectively. Solid lines in Figures 3.25 (c) and (d) are the leverage plots of the contact adherence (μN) as a function of the log of number of cycles, and the maximum contact force (μN), respectively. The dashed lines on either side of the solid lines represent 95% probability limits. The horizontal dashed line in each graph is the sample mean. Generally, an effect is considered to be statistically significant only if both 95% probability limits intersect the sample mean (the effect in (b) is not statistically significant).
One possibility is progressive material transfer between the contact bodies (due to either ductile or brittle rupture, if the contact interface is perfectly metallic, as discussed earlier). This would remove the contact film and cause more clean metallic surface to be exposed. It could also cause a gradual smoothening of the surface and/or blunting of asperities; since the highest and sharpest asperities bear the highest loads, they would be the likeliest sites of material transfer, due to the cleaner contact interface. Some other mechanisms that could cause some of these changes are briefly discussed in Chapter 4.

I will end this chapter with two pieces of discussion. First, I present a somewhat simplistic model of a contact surface partially covered by an electrically insulating film, whose extent decreases with cycling. Next, we consider the extent to which the evolution of measured characteristics could be caused by changes in the surface topology, based on the model results shown in section 3.12.
3.13.1 Changes in insulating film

The initial changes in measured contact resistance characteristics (for example, Figures 3.24 (a) and (b)) are almost certainly caused by an insulating film at the contact interface being gradually removed by repeated contact. It is possible that this phenomenon may at least partially explain the evolution during the remainder of the life-cycle as well (corresponding to Figures 3.24 (a) and (b)). A way to model this is to assume that at any particular stage of the contact evolution, only asperities bearing a load greater than a certain threshold are in intimate contact. At the asperities with individual loads smaller than the threshold, contact is interrupted by an insulating film. At these asperities, the surface energy is assumed to be zero – that is, the asperities deform according to the adhesionless model of Chapter 2, and do not contribute to the adherence force while

![Contact Resistance vs Load and Unload](chart.png)

**Figure 3.26** Contact resistance as a function of loading and unloading force, as given by a contact model with a “surface film effect”, with an asperity threshold force of 5 µN (a). The characteristic given by the original model is re-plotted in (b). The contact surface is identical in both models, 100 asperities with σ=0.01 µm, and R=0.1 µm.
unloading. Also, the corresponding contact spots are assumed to be electrically non-conducting. As the contacts are cycled, the threshold force gradually decreases, bringing more lightly loaded asperities into electrical contact. The load-unload characteristic given by this model is shown in Figure 3.26(a), with a maximum contact force of 70 μN, and with a threshold force of 5 μN. The original model (with all contacting asperities assumed to be in clean, metallic contact) is re-plotted for comparison in Figure 3.21(b). Compared to the original model, the introduction of the threshold force results in the switch turning on only when the contact force is about 9 μN (when the first contacting asperity is loaded to 5 μN), and results in a somewhat more gradual decrease in the contact resistance at low contact forces. The ultimate contact resistance bounds are higher, and the adherence force is smaller (22 μN, as compared to 31 μN in the original model). All these changes are qualitatively consistent with the differences between measured characteristics at a small number of switch cycles (which would correspond to the model with a threshold force) and measured characteristics after a large number of switch cycles (corresponding to the model without a threshold force).

3.13.2 Changes in topology

The predicted effect of $R$ on contact resistance characteristics is weak. On the other hand, $\sigma$ clearly has a strong effect on the contact adherence force, and also the contact resistance. For example, if the contact surface was initially equivalent to a surface with $\sigma=0.01 \, \mu m$, and $R=0.1 \, \mu m$, and evolved with cycling to a surface with $\sigma=0.003 \, \mu m$, and $R=0.1 \, \mu m$, the adherence force corresponding to a maximum load of 70 μN would increase from 31 μN to 62 μN, and the lower bound on the contact resistance at 70 μN
would decrease from 0.11 Ω to 0.06 Ω. Qualitatively, this could explain the kind of transition in adherence force and contact resistance we observe between 1000 and 10000 switch cycles (Figure 3.27). However, a progressive change in $\sigma$ clearly does not account for the progressive force-insensitivity of the contact resistance while loading; consequently, with increase in switch cycles, the model becomes increasingly inaccurate in predicting the contact resistance at small loads while loading.

One way to explain a very flat loading characteristic is to postulate that the asperities tend towards having flat tops, and that the flat surface tends towards having “mating” concavities that fit perfectly with the asperities, so that there is eventually no significant plastic deformation, and a large and force-independent contact area is obtained immediately upon contact. In the limiting case, this is equivalent to representing the rough surface by a set of flat punches, all of nearly the same height (as we have seen, the unloading uncharacteristic is flat because the asperities unload like flat punches). It is possible to perform a rough calculation based on this idea. Let the rough surface have $n$ flat punch asperities, all with the same height and with the same radius $a$, such that the total contact area is of $3\times10^{-14}$ m$^2$, which corresponds to a contact force of 70 µN, assuming plastic deformation and neglecting the effect of adhesive forces. Defining an effective contact radius $a_{\text{eff}}=0.1$ µm corresponding to this contact area, the total adherence force is either $n^{1/4} \sqrt{6\pi a_{\text{eff}}^3 K w}$ (see equation 3.18) or $\pi a_{\text{eff}}^2 H$, depending on whether the separation is brittle or ductile. Assuming $n$ is in the range 10-50, separation will be ductile, and the adherence force is 86 µN. The contact resistance will be completely independent of force while loading and unloading; for $n=10$, the contact
Figure 3.27 Comparison of modeled contact resistance characteristics at two different values of $\sigma$ (0.01 $\mu$m in (a), and 0.003 $\mu$m in (b)), with measured characteristics after 1000 cycles (a) and 10000 cycles (b) respectively. In either case, the rough surface is modeled by 100 asperities with $R=0.1$ $\mu$m.

Contact resistance bounds are 0.2 $\Omega$ - 0.06 $\Omega$; for $n=50$, the bounds are 0.2 $\Omega$ - 0.03 $\Omega$. These are all physically realistic values at a large number of cycles.

References


