B

Babbitt Bearing

► Hydrodynamic Journal Bearings

Ball and Roller Bearing Heat Transfer

► Rolling Bearing Heat Transfer and Temperature

Ball and Roller Bearings

► Rolling Element Bearings, History

Ball Bearings for Radial-Thrust Loading

► Angular Contact Bearings

Band Clutch or Brake

► Wet Clutch Friction Material: The Surfaced Groove Effect

Base Fluid

► Mineral Oil Base Fluids

Base Oil

► Mineral Oil Base Fluids

Base Stock

► Mineral Oil Base Fluids

Basic Concepts in Adhesion Science

Robert H. Lacombe
Materials Science and Technology, CONFERENCES, Hopewell Junction, NY, USA

Synonyms

Dipole-dipole interactions; Dispersion forces; Electromagnetic interactions and adhesion; Surface energy and adhesion; Surface tension and adhesion; Van der Waals interactions

Definition

Adhesion science is the study of the mutual interaction of material bodies at their impinging surfaces or shared interfaces, and also the exploration of the consequences of said interactions on the thermal-mechanical stability and physical behavior of coatings, laminate structures, and other adhering bodies.

Scientific Fundamentals

The problem of the adhesion of materials in nature has the curious property of being both a sticky subject and at the same time a slippery one, too. The sticky aspects are readily apparent in the many common phenomena we observe in everyday life such as the behavior of glues, tapes, and the sliding friction behavior of nearly every macroscopic object. The slippery aspect of adhesion

Q.J. Wang, Y.-W. Chung (eds.), *Encyclopedia of Tribology*, DOI 10.1007/978-0-387-92897-5,

behavior arises when trying to understand the fundamental mechanisms behind the common adhesion phenomena that are routinely observed in the natural world. The basic difficulty arises from the fact that all adhesion behavior arises from interactions at the atomic and molecular level, which gives rise to an extremely complex, many-body problem involving the mutual interaction of a nearly uncountable number of interacting bodies, each of which is acted upon by all the others. To begin to understand this problem requires the use of approximating assumptions, each of which works well enough within a certain regime, but none of which holds over the entire range of behavior.

The approach of this short entry will be to work from the bottom up. Starting with an elementary treatment of two simple atoms and working up by degrees to the adhesion behavior of relatively small macroscopic entities.

At the Bottom all Adhesion-Related Interactions are Electromagnetic

By the time the twentieth century was half over it was well established that the properties of all common matter were fundamentally governed by electromagnetic interactions. Classical electromagnetism was well explained by Maxwell's equations and quantum theory explained how the electromagnetic interactions between the fundamental particles such as the proton and the electron served to bind atomic matter together. What was happening within the atomic nucleus was still problematic since an entirely different set of forces were operating, but these forces held the typical nucleus together so strongly and within such a small volume that, for the purposes of understanding the chemical nature of matter, the nucleus could be assumed to be a point mass and a source of positive charge that attracted and bound the electrons. To get some idea of the sizes and distances involved, consider that a typical atomic nucleus is on the order of 3.4×10^{-15} m, the electromagnetic radius of the electron comes to 2.8×10^{-15} m, and within a typical atom electrons and protons are separated on average by roughly a Bohr radius, which is close to 5.3×10^{-11} m. From these numbers one readily perceives that atomic matter amounts to relatively small point masses separated by huge distances. To put the matter in better perspective simply scale the numbers up by a factor of 10^{13}. The nucleus then would have a radius of 3.4 cm, which is roughly a billard ball, and the electron would come to 2.8 cm, or a large grape, and they would be separated on average by a distance of 530 m, or roughly half a kilometer! Thus one gets the impression that the world at the atomic level is mostly empty space. Or is it? What is it that fills all that space and gives the impression that atomic matter is something sturdy and substantive? The answer in a word is the electromagnetic field. Technically, according to Maxwell's equations, when charges are present and moving around then both electric and magnetic fields have to be present. However, the electric field tends to be substantially stronger than the magnetic, so for purposes of illustration we can focus on the electric field only. Going back to lessons from Physics 101, the fundamental relation governing the force of attraction or repulsion between two charged particles is given by

$$F = \frac{1}{4\pi\varepsilon_0}\,\frac{q_1 q_2}{R^2} \tag{1}$$

where:

F = Attractive/repulsive force (N)
q_1,q_2 = Particle charges (coulomb)
R = Radial separation between particles (m)
ε_0 = Permitivity of free space ($coul^2/N\ m^2$)

To get some idea of the implications of (1) it helps to put in a few numbers and see what comes out. The prefactor $1/(4\pi\varepsilon_0)$ is roughly 9×10^9. Thus, assuming two particles each with opposite charges of 1 coul and separated by a distance of 1 m, the force attracting them would be nine billion newtons! One had better step back rather quickly because the resulting collision would be rather spectacular, amounting to a small nuclear explosion. This explains why in the natural world most matter is approximately neutral since the forces due to charge separation mount up rather rapidly. For another example consider a simple capacitor consisting of two conducting plates each 1 cm^2 in area and separated by a gap of 1 mm. If opposite charges of roughly 9×10^{-9} C (9 nanocoulombs) are put on each plate then the voltage between the plates would rise to 10,000 V or more than the breakdown voltage of air.

The force in (1) is moderated by the electric field between the charges. Each charge is the source of an electric field given by

$$E = \frac{1}{4\pi\varepsilon_0}\,\frac{q}{R^2} \tag{2}$$

This field is indeed a quite palpable quantity and is what fills all that apparently empty space in atomic matter. In fact the energy density associated with a general electric field E is

$$U = \frac{1}{2}\varepsilon_0\, \boldsymbol{E}\cdot\boldsymbol{E} \tag{3}$$

U is an energy density with units of Joules per cubic meter (J/m^3). Now the solidness of a typical slab of matter is generally gaged by its elastic modulus Y, which normally

Basic Concepts in Adhesion Science, Table 1 Electromagnetic energy density in the neighborhood of an electron at representative atomic distances estimated from (4)

Distance from electron (angstrom)	Energy density/ (electromagnetic modulus GPa)	Comments
4.0	2.8	Roughly the modulus of a thermoplastic (i.e., polystyrene)
1.37	208.0	Modulus of metal such as steel
1.0	732.0	Modulus of refractory material like silicon carbide or diamond

has the units Pascals (Pa – N/m^2) but also has the units of an energy density (J/m^3 since N/m^2 is equivalent to $N\text{-}m/m^3$, which is J/m^3). The basic unit of charge floating around in any solid is the charge on the electron, which is 1.6×10^{-19} coul. We can get some idea about the electromagnetic energy density close to a typical electron by combining (2) and (3) and using the electron charge for q. Condensed down to essentials the following formula emerges:

$$U = \frac{2300}{\pi}\frac{1}{r^4} \tag{4}$$

where r is in angstrom units and U in units of gigapascals (GPa) where a gigapascal is the same as 10^9 J/m^3 and is the more common label for an elastic modulus. Table 1 summarizes the electromagnetic energy density at a few representative atomic distances.

Table 1 clearly shows that the electromagnetic energy density associated with the electron's electric field can account, at least in a heuristic fashion, for the elastic properties of nearly all forms of common matter. Thus, knowing the electron charge density in a solid should also allow one to estimate the attractive fields responsible for the adhesion force. The problem is how to go about it.

From the force estimates based on (1) it is clear why matter is basically neutral with few or no free charges floating around. Thus the direct use of (1) is of no help in explaining the adhesion force. However, even though the typical solid has no free electrons (i.e., no electrons not simultaneously coupled to an opposite proton), there is still an enormous amount of electronic charge present and, at any point in time, it is not necessarily uniformly distributed. That is, a certain amount of charge separation can be present, giving rise to nonuniform charge distributions that can exert their own local electromagnetic influence. One way to sneak up on this problem is through what is known as the multipole expansion formula for the potential field due to a general charge distribution (This is a standard formula from classical electromagnetic field theory worked out in a number of texts. A classic example is Jackson (1962)):

$$\Phi(r) = \frac{1}{4\pi\varepsilon_0}\left[\frac{q}{r} + \frac{\boldsymbol{p}\cdot\boldsymbol{r}}{r^3} + \frac{1}{2}\sum_{ij} Q_{ij}\frac{r_i\, r_j}{r^5} + \ldots\right] \tag{5}$$

where:

Φ (r) = Electrostatic field potential (volt)
q = Free charge in coulombs (coul)
p = Dipole moment of charge distribution (coul m)
r = Radius vector from origin with magnitude $r = (\mathbf{r}\cdot\mathbf{r})^{1/2}$ (m)
Q_{ij} = Quadrupole moment of charge distribution (coul m^2)
r_i, r_j = Cartesian components of radius vector r (m)

Equation (5) neatly dissects any general charge distribution into components of ever-decreasing strength. If an amount of free charge q is present, this is accounted for by the first term, which is the potential responsible for the classic coulomb force and electric field given by (1) and (2). If no free charge is present the next term dominates, which is due to a simple dipole charge separation. Higher-order terms are also present such as the quadrupole contribution, but these terms are substantially weaker than the first two and are generally neglected unless circumstances dictate otherwise. Nuclear physicists in particular have to grapple with these terms in order to sort out the charge distribution of atomic nuclei. For present purposes, since the assumption is that essentially no free charge is present, attention focuses on the second dipole term. It is well known that many molecules have dipole moments due to asymmetric electronegativities of their component atoms. A carbon atom bonded to any of the halogens is a typical example. Thus there is good reason to believe that dipole fields will be present in otherwise neutral matter and these fields will interact with the general charge distribution present. In particular, and again from classical electromagnetic theory, it is known that an electric field can give rise to a local polarization charge density given by

$$\rho_{pol} = -\nabla\cdot\boldsymbol{P} \tag{6}$$

where:

ρ_{pol} = Polarization charge density ($coul/m^3$)
$\mathbf{P} = N\alpha\varepsilon_0 E$

N = Number of charges/unit volume (m^{-3})
α = Polarizability (m^3)
E = Electric field (volt/m)

The components of the electric field E in (6) arising from the dipole moment p are given by (Note that spherical coordinates are assumed here where the basic variables are the radius distance r, the declination from the *z* axis θ, and the azimuthal angle φ, which measures the angular displacement of the projection of the radius vector r on the *x,y* plane from the *x* axis.)

$$\begin{aligned} E_r &= \frac{2p\cos\theta}{r^3} \quad (radial\ component) \\ E_\theta &= \frac{p\sin\theta}{r^3} \quad (theta\ component) \\ E_\phi &= 0 \qquad (azimuthal\ component) \end{aligned} \tag{7}$$

Equations (6) and (7) come from the standard theory of electrostatics and define what is called the polarization field due to a dipole charge distribution. Thus a dipole moment in an electrically neutral molecule gives rise to an electric field, which in turn can induce another polarization charge density in the local neighborhood. Now the local energy density associated with the polarization charge interacting with the induced field is given by another standard formula:

$$U = \Phi(\boldsymbol{r})\ \rho_{pol}(\boldsymbol{r}) \tag{8}$$

In (8) U is the energy density (J/m^3) generated by the potential interacting with the charge density ρ_{pol}. The potential and polarization charge ρ_{pol} can be substituted from (6) and (7) with the help of (5) to give the following formula for the dipole-induced dipole energy density:

$$U = -\frac{N\alpha p^2\cos^2\theta}{16\pi\varepsilon_0 r^6} \tag{9}$$

Equation (9) represents the dipole-induced dipole contribution to what is commonly called the Debye interaction (in honor of Peter Debye, who did pioneering work on dipole interactions). The main feature of (9) is the $1/r^6$ spacial dependence of the interaction, which makes it a very short-range effect.

Though (9) gets somewhat closer to understanding the nature of electromagnetic interactions in charge neutral materials, the story of the interactions between elements of neutral matter does not end here. It is well known that the attractive force exists even when no permanent dipoles are anywhere present. What now enters the picture are the so-called London dispersion forces, where one imagines that due to fluctuations in one region a net dipole moment is momentarily created, which then induces further charge separation and thus another dipole in the local neighborhood, which gives rise to yet another net interaction. At this stage classical electromagnetic theory alone is of no further use in trying to sort out what is going on. The need now is to understand in more detail the dynamics of charges under the influence of electromagnetic interactions, which means going into the realm of quantum mechanics.

Quantum Mechanics of Adhesion Interactions

In general solving the problem of electromagnetic interactions in solid materials requires a quantum mechanical formulation. Unfortunately, this approach typically leads to insuperable mathematical problems. However, some insight can be gained by looking at simple systems and making approximating assumptions. The simplest nontrivial system imaginable is the hydrogen atom, where one electron interacts with one proton. This problem has been heavily studied mainly because it is one of a very small number of problems that can be treated exactly in a mathematical sense. Many treatments have been given, but a clear and concise treatment is that of Feynman from his classic Lectures on Physics (See, in particular Feynman et al. (1963)). In this case the hamiltonian operator is:

$$H = -\frac{h^2}{8\pi^2 m}\nabla^2 - \frac{e^2}{4\pi\varepsilon_0 r} \tag{10}$$

where:

e = Electron charge (coul)
m = Electron mass (kg)
ε_0 = Permitivity of free space ($coul^2/N\ m^2$)
h = Planck's constant = 6.63×10^{-34} (J s)
r = Radial distance between electron and proton at any given time
∇^2 = Laplacian operator most conveniently expressed in spherical coordinates

The eigenvalues of H are the allowed energy levels that the atom can occupy and are given by:

$$E_n = -\frac{me^4}{8h^2{\varepsilon_0}^2}\left(\frac{1}{n^2}\right) \quad n = 1, 2, 3\cdots\infty \tag{11}$$

According to (11) the electron can occupy any one of an infinite number of energy levels, but the one of most interest is the ground state, which is obtained when n = 1. This is the state of lowest energy and therefore the one most likely to be occupied under normal circumstances.

The corresponding wave function ψ_1 yields a charge distribution $\psi_1^*\psi_1$, which is spherically symmetric and contained within a sphere of radius roughly 1/2 Å or 5×10^{-11} m. This is given by what is called the Bohr radius $r_B = \varepsilon_0 h^2/(\pi m e^2)$.

Technically, since the charge density falls off exponentially with increasing radial distance, the electron can spend some time outside the Bohr radius but not much. The exponential tail of the charge distribution can be important, however, in explaining another strange quantum phenomenon known as electron tunneling, whereby an electron manages to tunnel its way through an energy barrier that would be strictly forbidden from the point of view of classical mechanics.

As seen from a distance, the hydrogen atom is an electrically neutral spherically symmetric object with no net dipole or quadrupole moment whatever. It should therefore be electromagnetically invisible to any other hydrogen atom that is suitably far away. This would certainly be the case for the hydrogen atoms held in a container at room temperature and pressure where the mean spacing between the atoms is something like 10^{-7} m or more than 10,000 Bohr radii distant. The case of interest here is when the atoms get much closer together, as they can at very high pressure or in the liquid but not too close. Thus consider two hydrogen atoms separated by a distance R much larger than the Bohr radius but much smaller that the spacing in the ideal gas state as shown in Fig. 1.

For the configuration shown in Fig. 1 the electrons are close enough to perceive that there is definite structure present but not so close that problems arise due to the fact that the electron has an internal spin degree of freedom that gives rise to a magnetic moment and further makes it a Fermi particle that must obey the Pauli exclusion principle. In this simpler case the interaction energy of the system is given by:

$$\Phi = \frac{e^2}{4\pi\varepsilon_0}\left[\frac{1}{\boldsymbol{R}} + \frac{1}{\boldsymbol{R} + \boldsymbol{r}_2 - \boldsymbol{r}_1}\right] + \frac{e^2}{4\pi\varepsilon_0}\left[\frac{1}{\boldsymbol{R} + \boldsymbol{r}_2} + \frac{1}{\boldsymbol{R} - \boldsymbol{r}_1}\right] \tag{12}$$

Basic Concepts in Adhesion Science, Fig. 1 Schematic diagram of two hydrogen atoms separated by a distance large compared to the Bohr radius but close compared to the mean spacing in an ideal gas

Even though (12) looks reasonably simple, putting it into the hamiltonian operator would still present a rather difficult problem. Therefore, at this point use is made of the fact that R is much larger than r_1 and r_2 allowing (12) to be expanded in powers of 1/R. We find that the first nonvanishing term occurs at second order giving rise to the following simplified version of the interaction energy:

$$\Phi = -\frac{e^2}{4\pi\varepsilon_0}\left[\frac{\boldsymbol{r}_1 \cdot \boldsymbol{r}_2}{R^3} - \frac{3(r_1 \cdot \boldsymbol{R})(r_2 \cdot \boldsymbol{R})}{R^5}\right] \tag{13}$$

Equation (13) may not look much simpler than (12) but it has the definite advantage of being small since it is on the order of $1/R^3$. Also note that (13) is precisely the interaction energy we would expect for two interacting dipoles. Thus the two hydrogen atoms in Fig. 1 appear to be interacting through a weak dipole-dipole mechanism. What this will do is cause the energy levels of each atom to shift slightly, and since the interaction is relatively weak it is not necessary to solve the whole problem again from scratch, but use can be made of the well-known results for the isolated atom to calculate the energy shift using perturbation theory (What is presented here is intended to give a brief overview of the methodology of perturbation theory. To go into details would lead to very deep water very fast and would not be all that illuminating to the topic at hand. Perturbation theory is covered in a number of standard texts such as *Introduction to Quantum Theory* (Park 1964). At a more advanced level one can consult Gottfried (1966)). Essentially what the perturbation approach comes down to is this. Starting with some hamiltonian H for which we know all the amplitude functions ψ_n^0 and energy levels E_n^0 it is desired to solve the problem where H is perturbed by some small interaction ΔH. Since ΔH is small it is expected that the changes in ψ_n^0 and E_n^0 will also be small, so it is possible to write the wave function and energy levels for the perturbed system as

$$\begin{aligned} \psi_n &= \psi_n^0 + \lambda\psi_n^1 + \lambda^2\psi_n^1 + \cdots \\ E_n &= E_n^0 + \lambda E_n^1 + \lambda^2 E_n^0 + \cdots \end{aligned} \tag{14}$$

In (14) ψ_n^0 and E_n^0 are the known wave functions and energy levels of the unperturbed system. ψ_n and E_n are the wave functions and energies of the new system that are to be found. The ψ_n^1, ψ_n^2, ... and E_n^1, E_n^2, ... are the first-, second-, and higher-order corrections that must be determined. The parameter λ is a number less than 1 used to serve as an expansion marker to keep track of the order of the terms in the expansion. One usually hopes that the first-order corrections will be good enough and if that doesn't work out then the second-order corrections must be calculated. It gets very cumbersome if one has to go

higher than second order, which fortunately does not occur very often. It turns out that for the interaction energy given by (13) the first-order corrections turn out to be zero due to the symmetry of the hydrogen ground state. Going to second order fortunately gives a finite result, which turns out to be

$$\Delta E = \frac{e^2}{4\pi\varepsilon_0 r_B}\left(\frac{r_B}{R}\right)^6 \zeta \qquad (15)$$

where:

$r_B = \varepsilon_0 h^2/(\pi m e^2)$ the Bohr radius
ς = Dimensionless constant ~ 6.5

Again, the ubiquitous $(1/R^6)$ spacial dependence of the interaction energy appears just as was found in the dipole-induced dipole case, though in this case it arises under circumstances where there are no permanent dipoles.

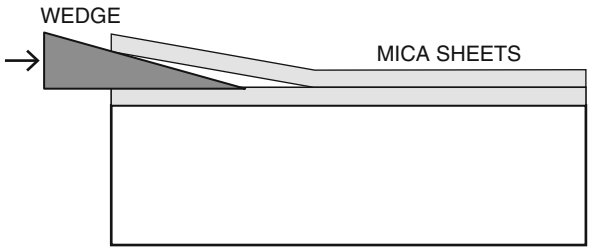
Equation (15) is the simplest example of the interaction energy arising from what are commonly referred to as dispersion or van der Waals interactions. Although this result has been derived fairly rigorously from first principles, it was arrived at using the simplest nontrivial system imaginable. However, the physical arguments used to derive (15) hold up for even more complex atomic and molecular systems and the van der Waals interaction has been used successfully to understand the behavior of dense gases, liquids, and simple molecular solids under the assumption of purely electrically neutral interactions with no free charges or dipole interactions. As a note of caution, one has to be aware that as the distance between the two atoms in Fig. 1 gets too small, then the assumptions leading to (13) break down and all the machinery of perturbation theory is now invalid. Worse yet, the spin degrees of freedom of the electrons enter the problem and account has to be taken of the fact that electrons are Fermi particles that require special treatment within quantum theory. Pushing the two atoms too close together results in the formation of a covalent bond the treatment, which requires full industrial-strength quantum theory that is far beyond what can be treated here.

Adhesion Interaction of Macroscopic Bodies: The Next Level – Clean Surfaces in Close Contact

Up to this point it is established that there can be at least three different types of electromagnetic interactions that can lead to adhesion forces between atoms and molecules. The most obvious candidate is the coulomb force given by (1), which is active when free charges are present. However, it is clear that for the most part macroscopic bodies are electrically neutral due to the enormous forces that can build up when free charges are present. Though there are many common instances where free charges can cause macroscopic objects to cling to one another, the coulomb force is not the dominant mechanism for macroscopic adhesion between material bodies. The next candidate would be the dipole-dipole interactions that are quite common in many materials. However, dipole interactions are strongly directional and can be either attractive or repulsive depending on the orientation of the interacting molecules. Also, in many materials the dipoles are oriented haphazardly, giving rise to a strong tendency for the net interaction to cancel out. Finally there is the van der Waals dispersion force, which is always attractive and always present in neutral matter.

Early Work of Obreimoff

Under ordinary circumstances, one does not get the immediate impression that solid bodies can interact simply by bringing the two together. The reasons for this are readily understandable since for any solid large enough to hold in your fingers, or even with tweezers, the adhesion forces involved are very small. In addition, most surfaces tend to be quite rough and contaminated with all manner of foreign substances. All of these conditions tend to suppress strong adhesion interactions. However, there are exceptions to be found and one of them is thin mica sheets. Obreimoff (1930), who was working in Russia early in the twentieth century where an exceptional type of muscovite mica was available, noticed that relatively large sheets of perfectly clean, smooth, and flat mica could attract each other quite strongly. Curious to investigate this phenomenon more closely he set up a simple experiment whereby a wedge could be used to separate two sheets in a controlled manner, as shown schematically in Fig. 2. By gently taping a wedge between the two mica sheets, they could be separated in a controlled manner. The deformation of the top sheet could also be monitored by interferometric methods, essentially Newton ring type phenomena that one readily sees by bringing a curved lens into contact with a flat reflecting surface. By this method Obreimoff demonstrated that the sheet being lifted of was behaving like an elastic beam just as one would expect if it were a piece of sheet metal being lifted off the floor by one edge. What is of interest though is that upon removing the wedge the sheets were observed to rejoin quite readily and that the energy recovered was not always equal to the energy spent in splitting the sheets. Further it made a difference whether the experiment was performed in air or under vacuum conditions. These earliest experiments on separating mica sheets gave the world one of



Basic Concepts in Adhesion Science, Fig. 2 Schematic of Obreimoff experiment. A wedge is gently tapped between two mica sheets. The deformation of the top sheet can be monitored by interferometric methods

the earliest glimpses into the nature of the atomic and molecular forces that bind all matter.

Improved Observations of Tabor and Winterton

The work of Obreimoff was a good start but there were a number of problems with the method. First, using a wedge to split the mica introduced drag forces associated with driving the wedge. In addition, during the splitting process a significant amount of charge separation between the sheets could occur giving rise to electrostatic forces that made themselves known through electrical discharges. However, the use of nice, flat, clean mica sheets was definitely a good idea and Tabor and Winterton (1969) took up on it while making a number of substantial improvements.

The first improvement was to avoid the use of a wedge for separating the sheets. To get around this problem they came up with the idea of bending the sheets into the form of cylindrical sections by gluing them to cylindrically shaped substrates. By bringing the two cylindrical sections together at a right angle to each other they could get a carefully controlled circular contact area. The second improvement involved bringing the two cylindrical sections together in a precisely controlled way by using a piezoelectric actuator and measuring the force acting between the two surfaces by mounting one of the cylinders on a carefully fabricated cantilever beam with precisely known bending behavior.

Finally, by applying a thin reflecting metal coating to the back of each sheet they could observe the interference behavior of multiple reflections between the two surfaces instead of the standard Newton's rings as was done in the Obreimoff experiment. With the Newton ring method one has a measurement sensitivity down to a few tenths of a micron, whereas the multiple reflection method gets one down to a few nanometers.

The Surface Force Apparatus

Building upon the work of Obreimoff and Winterton, Israelachvili and Tabor (1972) further refined the experiment, leading to what is now called the *surface force apparatus* shown schematically in Fig. 3. In this experiment sheets of muscovite mica are applied to cylindrical glass surfaces and arranged at a right angle to each other in order to give a circular contact region when brought together. The gap between the cylinders is precisely controlled by a piezoelectric actuator and the gap distance is accurately measured by multiple reflection interferometry down to a resolution of 3 nm. The force between the cylinders is measured by a carefully calibrated cantilevered spring. Upon bringing the two cylinders increasingly closer together an attractive force between the cylinders is noted. As the cylinders approach still closer, a point is reached where the cylinders jump into contact as the cantilever spring is no longer strong enough to balance the attractive force. Figure 4 illustrates typical data for two crossed mica cylinders.

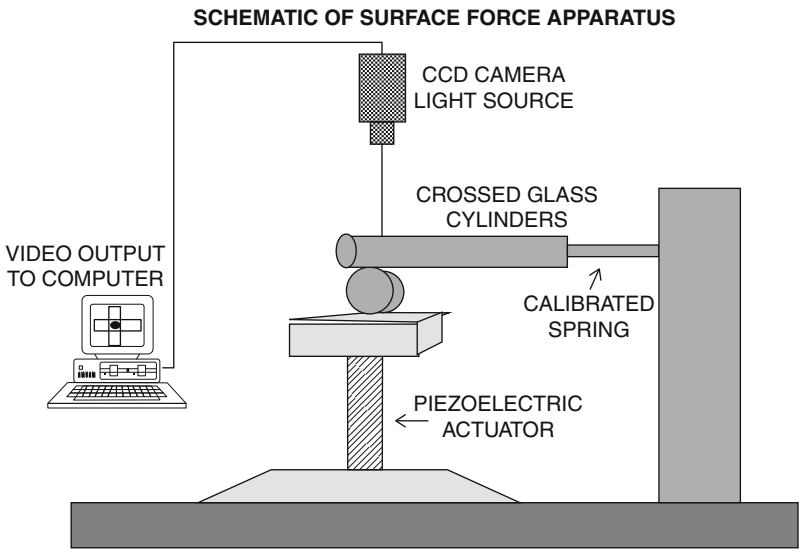
It was apparently Hamaker (1937) who concluded that if two neutral bodies are brought close enough together then van der Waals London dispersion forces would kick in and give a net attraction. His approach to computing what the force should be was to divide each of the approaching solids up into small segments and assume a pair wise interaction of the segments according to the $(1/R^6)$ behavior and add them all up to get the net attraction. Though there are some problems with these assumptions, they were good enough for the time and allowed Hamaker to come up with the result given in (16), which fits the data in Fig. 4 remarkably well. Details of Hamaker-type calculations for various solid shapes have been worked out in the text by Rimai and Quesnel (2001). For the two crossed cylinder case we get the following formula for the interaction energy per unit area:

$$F/A = -\frac{D}{12z^2} \quad \textit{Two crossed cylinders} \qquad (16)$$

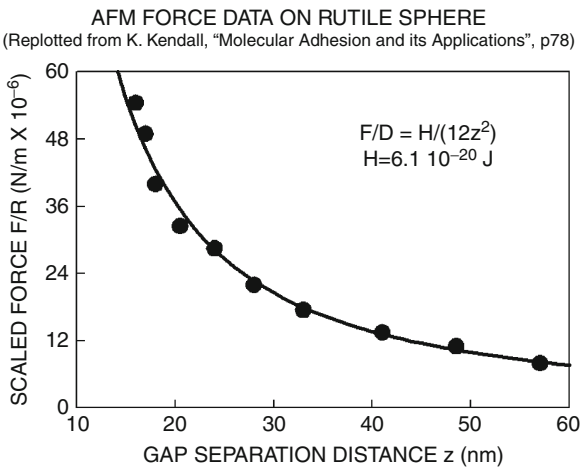
where:

F = Attractive force (N)
D = Cylinder diameter (m)
z = Separation distance between sheets (m)
A = Hamaker constant (roughly 10^{-19} J)

These results show that the attractive forces between solid surfaces can be explained by electromagnetic interactions arising either from the presence of dipoles or from the quantum mechanical effect of fluctuating charge distributions giving rise to transient dipole moments, which can interact on average to give the same $1/R^6$ behavior that



Basic Concepts in Adhesion Science, Fig. 3 Schematic of surface force apparatus resulting from work of Obreimoff, Winterton, Tabor and Israelachvil



Basic Concepts in Adhesion Science, Fig. 4 Sample surface force data for crossed mica cylinders using surface force apparatus

arises when permanent dipoles are present. These so-called dispersion forces can then be used to derive the Hamaker interaction in (16). This was all very nice but was only the beginning of the story. There are conceptual problems with the Hamaker approach to adding up all the pair wise interactions since, if there is an intervening medium between the two bodies, it will materially complicate matters by introducing various types of many body effects. To address these problems Lifshitz (The original work of Lifshitz and collaborators can be found in the following: Lifshitz (1956), Dzyaloshinskii et al. (1960), Dzyaloshinskii et al. (1961)) and co-workers came up with another approach to working out the problem of how the short-range molecular and atomic interactions give rise to the macroscopic attraction between bulk material objects. At this point the story starts to become not only more interesting but stranger as well. From Lifshitz's point of view the real actor was the electromagnetic field that moderates the interactions between atoms and molecules and, as demonstrated for the case of simple hydrogen atoms above, could also be shown to give rise to the London van der Waals dispersion forces.

To really understand what the electromagnetic field is doing one has to go beyond classical electrodynamics and Schrödinger equation–style quantum theory to the then-emerging concepts of quantum electrodynamics (QED). What one has to realize is that even after the development of classical electrodynamics and the quantum theory of the early twentieth century there was still a big hole in the understanding of electromagnetism, which essentially arose from the neglect of relativistic effects. Classical electrodynamics as worked out by Maxwell was relativistically correct from the outset. In fact, the best approach to understanding Einstein's famous theory is through electrodynamics. However, the early quantum theory was not relativistically correct. In particular, the Schrödinger-style hamiltonian in (10) does not account for relativistic effects. Though this was not a major problem early on,

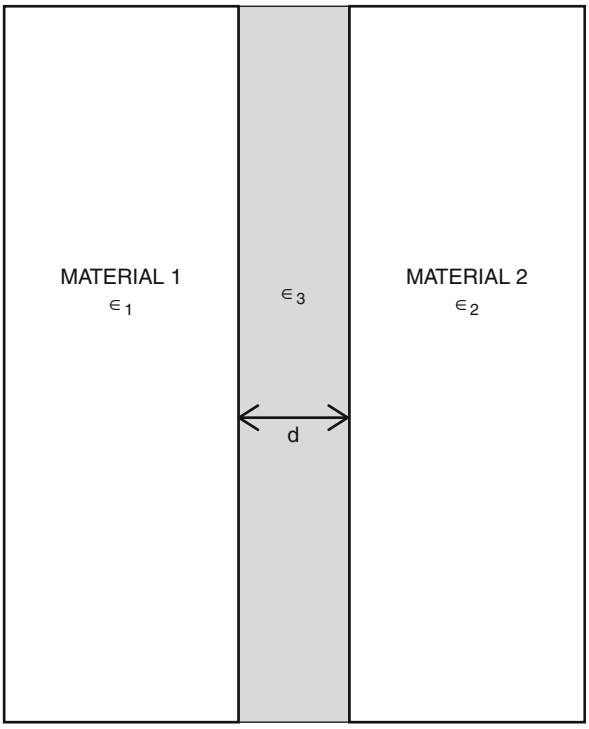
it was a conceptual difficulty that had to be worked out. This was all eventually taken care of by roughly mid-century. P. A. M. Dirac formulated his now famous Dirac equation Dirac (1928), which gave the relativistically correct hamiltonian for the electron and, as a clever bonus, also predicted the existence of the positron ahead of the experimentalists. Dyson, Feynman, Schwinger, and Tomonaga later followed up by formulating the complete and relativistically correct theory of electromagnetic radiation (The interested reader can follow up on the details of the development of quantum electrodynamics (QED) in the following review article: Schweber (1994)). In QED the source of the electromagnetic field is the photon that can be considered as a sort of quasi-particle. The typical photon can be in one of two possible polarization states and always travels at the speed of light from its creation to its eventual destruction.

Although single photons are rather feeble, they have the capacity of adding up under the superposition principle thereby creating an electromagnetic field of nearly any magnitude. Now Lifshitz took the approach that all the interactions were through the exchange of photons (i.e., electromagnetic radiation) between the bodies in question. Thus, using the principles of QED he was able to work out the formula for the Hamaker force by an entirely different approach than the original work of Hamaker described above. The result for two material planes separated by a third medium, as shown in Fig. 5, came to the following (A fairly concise and informative discussion of the Lifshitz theory can be found in Moy and Neumann (1996). A far more detailed review has been given in French (2000)):

$$
\begin{aligned}
A_{132} &= \frac{3kT}{2}\sum_{n=0}^{\infty/}\sum_{m=1}^{\infty}\frac{\Delta_{13}(n)\Delta_{23}(n)]^m}{m^3}\\
\Delta_{13}(n) &= \frac{\varepsilon_1(i\zeta_n)-\varepsilon_3(i\zeta_n)}{\varepsilon_1(i\zeta_n)+\varepsilon_3(i\zeta_n)}\\
\Delta_{23}(n) &= \frac{\varepsilon_2(i\zeta_n)-\varepsilon_3(i\zeta_n)}{\varepsilon_2(i\zeta_n)+\varepsilon_3(i\zeta_n)}\\
\zeta_n &= \frac{4\pi^2 kT}{h}n
\end{aligned}
\qquad (17)
$$

where:

A_{123} = Hamaker constant for the interaction of materials 1 and 2 separated by material 3
k = Boltzman constant (1.38×10^{-23} J/molecule °K)
h = Planck's constant 6.63×10^{-34}(J s)
T = Temperature (°K)
$\varepsilon_1, \varepsilon_2, \varepsilon_3$ = Complex dielectric constants for materials 1, 2, and 3 respectively



Basic Concepts in Adhesion Science, Fig. 5 Material configuration corresponding to the Hamaker constant calculation in (18)

$i = (-1)^{1/2}$ imaginary unit from complex variable theory
ς_n = nth response frequency

What is interesting about (17) is that all material properties are now represented by the complex dielectric constant, which is a measure of the material's response to an applied electromagnetic field. The dielectric constant is a function of frequency and the frequency range of interest is that in which a typical electron lives.

The energy of a photon is given by $E = h\nu$ (h = Planck's constant and ν is the frequency) and it will interact with electrons of roughly the same energy. Going back to our hydrogen atom the energy of the ground state is given by (11) with n = 1, which works out to 2.17×10^{-18} J (or −13.6 eV (eV), which are the more common units). Equating electron and photon energies we see that the hydrogen electron will interact with radiation near the frequency:

$$\nu = \left(\left|\left(2.17 \times 10^{-18}\mathrm{J}\right)/\left(\mathrm{h} = 6.63 \times 10^{-34}\mathrm{Js}\right)\right|\right) = 3.27 \times 10^{15}\mathrm{Hz}.$$

This puts the frequency in the range of ultraviolet radiation. In general most materials have energy levels that interact with radiation ranging from the infrared

through the visible to the ultraviolet, with the visible to ultraviolet range being the most common. The implication is that measuring the absorption and refractive index behavior of materials in the visible to ultraviolet range will allow for predictions about the short-range attractions. Very interesting indeed. Not only that but the story is about to get more interesting yet, as demonstrated by one further topic that bears on this discussion and is by far the most curious aspect of surface interactions in the quantum world – the Casimir effect.

Exotic Phenomena: The Casimir Effect

In 1948 the Dutch physicist Hendrik B. G. Casimir, while working on problems of colloid suspensions for Philips Research Labs, predicted that two uncharged parallel metal plates will be subject to a force pressing them together given by the following expression: (A short review of the literature on the Casimir effect can be found in the following articles: Lambrecht (2002), Lamoreaux (1997))

$$\frac{F_c}{S} = \frac{\pi hc}{480 d^4} \tag{18}$$

where:

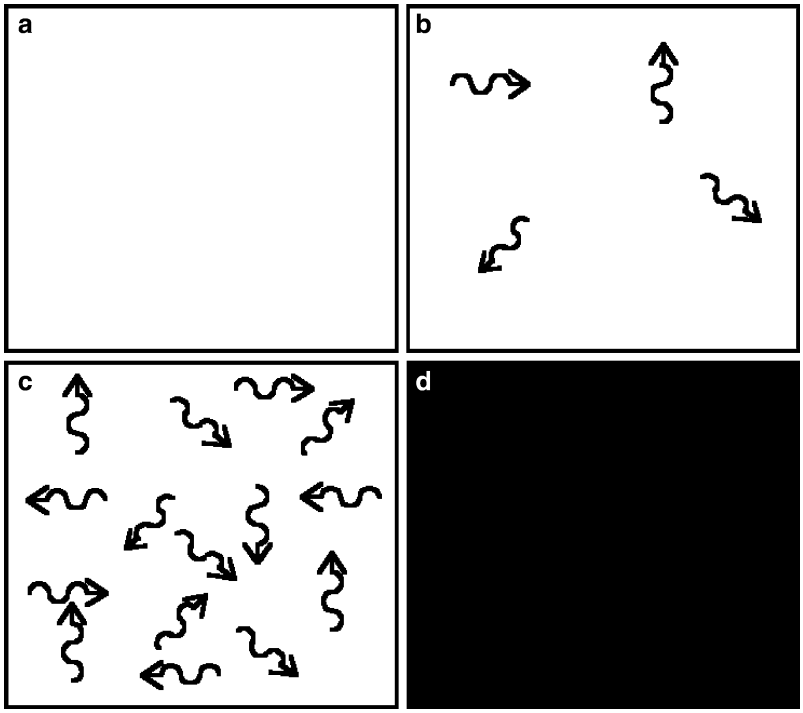
F_c = Attracting force
h = Planck's constant
c = Speed of light
d = Distance between the two plates
S = Surface area of plates

Although (18) bears a curious resemblance to Hamaker's equation the appearance is deceptive since in the Hamaker case the interaction results from the presence of atomic and molecular matter. With the Casimir effect the interaction arises out of completely nothing being wholly independent of any material property, as (18) clearly shows. How can this be?

In the realm of common sense, nothing is precisely that, a total lack of anything of interest. However, once one enters the realm of quantum electrodynamics nothing takes on an entirely different significance. A look at Fig. 6 gives a crude sort of analogy. In Fig. 6a we have a representation of the vacuum (i.e., a perfectly empty box). In Fig. 6b a small amount of radiation is added and Fig. 6c adds still more radiation. Finally, Fig. 6d goes to the ultimate limit and adds all possible radiation and thus is a representation of absolutely everything. Or is it? If we change the color of the photons from black to white and the background from white to black (i.e., look at the negative photograph) then the picture is precisely reversed and Fig. 6d becomes the vacuum and 6A everything! This is precisely the picture of the vacuum offered by quantum electrodynamics. The vacuum is nothing more than a seething sea of radiation of all magnitudes, phases, and directions such that on average everything tends to cancel out but not quite perfectly. There are fluctuations such that at any given moment a virtual photon is created for a brief interval and then immediately sucked back into the milieu. Since these virtual photons are exceedingly feeble they go completely unnoticed under ordinary circumstances. However, as we start making very small micro-electro-mechanical systems (MEMS), such as tiny sensors and actuators, that are fabricated on silicon substrates unexpected phenomena start to emerge. The Casimir effect can start to have a significant impact on this technology and thus is no longer simply an academic curiosity.

But just what is this mysterious effect? The easiest way to look at it is through a simple analogy. Say we have two small, uncharged metal plates sitting in a vacuum at a macroscopic separation of a few centimeters. To within any degree of measurable accuracy there is no interaction between them. The virtual photons of the vacuum are popping up and vanishing like so many bubbles in a boiling kettle of water, but they are so feeble that their existence is completely unnoticed. Now start to bring the two plates together in a very careful way, all the time monitoring for the presence of any net attraction between them. As the separation distance starts to get into the range of 10 nm and less an interesting effect starts to take hold. The virtual photons that are being continually created and destroyed in the gap between the plates have distinct wavelengths, and as the plates get closer and closer they start to act as a sort of microwave cavity that only allows certain wavelengths to exist. In effect then the creation of virtual photons in a certain range of wavelengths is being suppressed between the plates but not on the outside. Since the photons are electromagnetic waves they exert a pressure and since fewer can exist in the gap between the plates than outside there is a net pressure difference between the outside and inside that creates the Casimir attraction. Its not unlike the situation that would exist if the plates were in a chamber filled with some ideal gas and by some mechanism the space between the plates was being evacuated. The gas outside would generate a net pressure forcing the plates together.

A better analogy comes from the era of sailing ships. Two ships sitting on the ocean with very little breeze blowing but very heavy swells present can experience something close to the same phenomenon. When the ships are far apart the waves buffet them more or less randomly from all sides and on average they tend to hold their positions. However, if the ships are brought



Basic Concepts in Adhesion Science, Fig. 6 Box A: represents a vacuum. Box B: A weak radiation field. Box C much stronger field. Box D: All possible radiation

together side by side in a parallel fashion they tend to also create a wave cavity between them where certain wavelengths are preferred and others suppressed. Again, the water waves exert a pressure on the hulls of the ships and since the wave action between the ships is being suppressed by the cavity effect there is more wave pressure occurring from the outside than from within the gap, thus creating a net force that tends to draw the hulls together. This was apparently a serious problem in days of yore when sailing ships had extensive sails and rigging that could become entangled if they got too close.

Adhesion Interactions of Liquids: The Contact Angle Experiment

Moving from the adhesion behavior of solids it is now informative to take up the prolem of the adhesion of fluids. Because of their ability to take on nearly any shape, fluids are able to adhere to various surfaces much more effectively than solids, at least in the form of small droplets or thin films. The adhesion behavior of fluids is most noticeable in capillarity behavior in small tubes and the beading behavior of droplets on surfaces or contact angle phenomena. This section will focus on contact angle behavior.

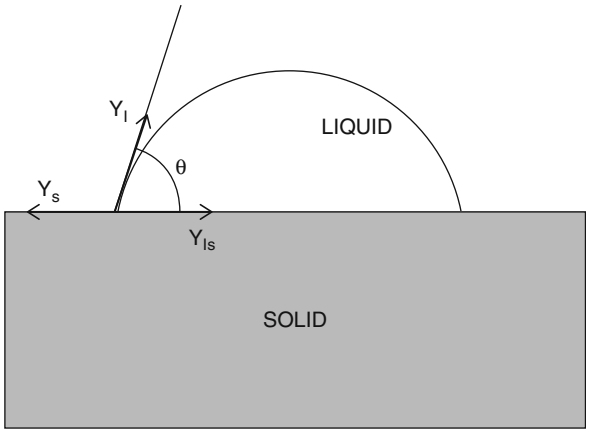
The most elementary theory of contact angle behavior derives from the basic principles of thermodynamics. The contribution of classical thermodynamics to contact angle theory comes in the form of the Dupre equation, which is essentially equivalent to the first postulate of conservation of energy given by

$$W_{ad} = \gamma_s + \gamma_l - \gamma_{sl} \tag{19}$$

where:

W_{ad} = Work of adhesion (J/m^2)
γ_l = Surface tension of liquid (J/m^2)
γ_s = Surface tension of solid (J/m^2)
γ_{sl} = Surface tension of liquid–solid interface (J/m^2)

This relationship states that if you remove a unit area of liquid from the solid surface you gain the surface energy of the solid plus the surface energy of the liquid but simultaneously lose the interaction energy between the two. The difference between what is gained and what is lost then amounts to the net work that must be done to remove the drop, which is defined as the work of adhesion W_{ad}. Equation (19) has four variables of interest, the main one being the work of adhesion. Of the remaining three the surface tension of the liquid γ_l is most readily available



Basic Concepts in Adhesion Science, Fig. 7 Liquid drop on solid surface exhibiting relevant surface tensions and resulting contact angle

from tables of thermodynamic data. This leaves three undetermined quantities to deal with (W_{ad}, γ_s, γ_{sl}). At this point one has to leave the realm of classical thermodynamics and invoke a basic principle of mechanics that states that the sum of all the tensions acting at the edge of the drop must balance to zero for equilibrium to hold. This then leads directly to the well-known relationship of Thomas Young:

$$\gamma_l \cos(\theta) = \gamma_s - \gamma_{ls} \tag{20}$$

where θ is the contact angle the drop makes with the solid surface and all the other quantities are as defined in (19). Figure 7 illustrates the standard geometrical picture.

Equations (19) and (20) contain a total of five quantities of interest, four of which are listed above plus the contact angle. In practice θ is measured using a goniometer and γ_1 is readily measured using independent methods such as the Wilhelmy plate experiment, which relies on the capillarity behavior of the liquid caused by the surface tension. Three quantities now remain to be determined: W_{ad}, γ_s, and γ_{sl}. Since we have only two equations relating these quantities the problem remains indeterminate. However, note that the quantities γ_{sl} and γ_s appear in both equations as the combination $u = \gamma_s - \gamma_{sl}$. Substituting u into (19) and (20) gives

$$W_{ad} = \gamma_l + u \tag{21}$$

Equations (20) and (21) are now readily combined to yield the famous Young-Duprey relation:

$$W_{ad} = \gamma_l \, (1 + \cos\theta) \tag{22}$$

In principle (22) allows one to determine the work of adhesion W_{ad} by measuring the contact angle of a drop of liquid with a known surface tension γ_1. This leaves γ_s and γ_{sl} still at large, and both are difficult quantities to measure while being of great interest for reasons both academic and practical.

From the experimental point of view there are a number of alternate methods for getting at γ_s and γ_{sl}, a short list of which includes:

1. Surface flow calorimetry
2. Inverse gas chromatography
3. Observation of solid–liquid melt boundary

A detailed description of these methods is well beyond the scope of this essay. Needless to say, each method comes with its own particular set of advantages and disadvantages. The interested reader should consult the review article by Etzler (2003) or the edited volume by Neumann and Spelt (See in particular the article: Spelt and Li (1996)).

Measuring the surface energy of a solid has to be one of the most challenging problems of surface science and the one topic where there is little agreement on what the true surface energy of any particular solid might be. Table 2 gives a listing of the most common methods for measuring the surface tension of liquids and solids.

What is directly noticeable is that there are nearly twice as many methods for measuring the surface tension for liquids as opposed to solids. What the table does not show, however, is that, whereas there is a general consensus that the methods used for liquids are reliable and accurate, there is little such a consensus for those methods in the solids column.

The situation has been summed up rather succinctly by Adamson and Gast (1997): "There is a rather limited number of methods for obtaining experimental surface energy and free energy values, and many of them are

Basic Concepts in Adhesion Science, Table 2 Methods for measuring the surface tension (surface free energy) of solids and liquids[a]

Liquids	Solids
Capillary rise method	Contact angle method
Drop shape analysis	Work of cleaving method
Drop weight method	Skapski method
Oscillating jet method	Grain boundary groove analysis
Pulsating bubble method	–
Ring method	–
Wilhelmy slide method	–

[a]Data from: Adamson and Gast (1997)

peculiar to special solids or situations. The only general procedure is the rather empirical one of estimating a solid surface tension from that of the liquid."

The basic problem comes down to the fact that, whereas liquids retain a high degree of mobility at the atomic and molecular level, solids do not. Thus a given liquid has the ability to spontaneously reorganize its surface to minimize the total surface free energy thus giving rise to a well-defined surface configuration that, for all intents and purposes, is smooth down to nearly the atomic level. Thus the surface free energy or surface tension of any liquid is well defined and amenable to precise measurement. The surface configuration of a typical solid is quite another matter, being highly dependent on the thermal-mechanical history that leads to its formation. Nearly every detail of the solidification process comes into play: whether solidification was rapid or slow, what mechanical loading conditions were active during and after the solidification process, what impurities were present, what was their precise concentration and spatial distribution, and so on. Thus any given surface can have a range of defects and structures that affect its surface free energy. In particular, voids and inclusions can induce local tensile and compressive stress fields that will alter the local surface tension and, if widespread, the total surface energy. Grain boundaries emerging at the surface have roughly the same effect. Thus all the details of a polycrystalline solid come into play. Finally, the presence of contaminant layers will disguise the underlying surface and thus make the measurement of the true surface energy nearly impossible even if only a monolayer is present.

Rather than try to use the contact angle method to determine the surface energy of a given solid, it is instructive to go in the reverse direction and, using Hamaker theory as outlined in (17), try to predict the contact angle from the theoretically derived Hamaker constants. The process starts with the interaction energy per unit area for two parallel plates as derived from Hamaker theory:

$$W = -\frac{A}{12\pi z^2} \quad \textit{Two parallel plates} \tag{23}$$

where:

W = Interaction enegy per unit area
z = Separation distance between sheets (m)
A = Hamaker constant (roughly 10^{-19} J)

Equation (23) asserts that the interaction energy between the two solid slabs in Fig. 5 is proportional to the Hamaker constant given by (17) divided by the separation gap squared. Now, letting the gap z shrink to zero produces an infinite interaction energy. Clearly this is impossible since, as the surfaces start to get too close, very strong repulsive forces kick in and prevent any further approach. What happens is an equilibrium separation z_0 is reached where the attractive and repulsive forces just balance one another. A moment's reflection reveals that at this separation the interaction energy predicted by (23) must amount to exactly twice the surface energy γ_1 assuming that both material slabs are made of material 1. This clearly must be the case since performing the inverse operation of separating the surfaces to infinity will yield two free surfaces each of surface energy γ_1. Thus, at the separation z_0 (23) can be written as

$$\gamma_1 = -\frac{A_{11}}{24\pi z_0^2} \quad \textit{Surface energy of material 1} \tag{24}$$

The precise value of z_0 is a matter of some controversy, but it cannot be too far from the atomic spacing in crystalline solids or the mean intermolecular spacing in bulk amorphous materials. A commonly assumed value is $z_0 = 1.65 \times 10^{-10}$ m.

Now going back to the contact angle experiment in Fig. 7 and assuming that the materials are such that the surface energies (surface tensions) are determined predominantly by dispersion forces, then for this system one can write:

$$\begin{aligned} \gamma_l &= -\frac{A}{24\pi z_0^2} \quad \textit{Surface energy of liquid} \\ \gamma_s &= -\frac{A_{ss}}{24\pi z_0^2} \quad \textit{Surface energy of solid} \\ \gamma_{sl} &= -\frac{A_{sl}}{24\pi z_0^2} \quad \textit{Energy of solid liquid interface} \end{aligned} \tag{25}$$

Going back to (20) for the contact angle and inserting the result in (25) assuming a fixed value for z_0 results in

$$A_{ll}\cos\theta = A_{ss} - 2A_{sl} \tag{26}$$

This readily implies that the contact angle is given by

$$\theta = \arccos\left(\frac{A_{ss} - 2A_{sl}}{A}\right) \tag{27}$$

Key Applications

With (27) the whole argument now comes full circle. The argument started off investigating surface energies as the source of contact angle and wettability behavior. This leads directly into the realm of atomic and molecular London–van der Waals interactions as the source for the surface energies. A moment's consideration makes it clear that electromagnetic interactions are the root cause of the interactions and thus the surface energies. One can then follow Hamaker back up to the macroscopic scale with his

deceptively simple formula involving the constant named after him. The electromagnetic interactions then reassert themselves with the work of Lifshitz, who demonstrated that the Hamaker constant is a consequence of quantum electrodynamic interactions between the solids involved, which lead to (17), which determines the Hamaker constant in terms of the dielectric properties of the materials involved. Thus measuring the Hamaker constant involves making index of refraction and ultraviolet absorption measurements. So the spectroscopists have something to say to the lab chemists measuring contact angles and presumably vice versa. To be sure there are a number of loose ends which cannot be addressed in such a short essay, but the overall picture makes for a self-consistent and most curious story with rather profound consequences for a wide range of practical applications including:

1. Flocculation behavior of colloid suspensions
2. Cleaning of particles from surfaces
3. Surface wetting and application of coatings
4. Biomedical applications that depend critically on all of the above
5. Sintering of particle aggregates
6. Friction force of sliding bodies

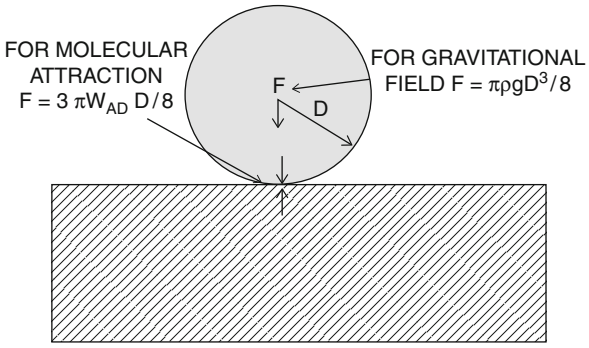
Scaling Considerations of Adhesion Interactions

When dealing with material objects that make physical contact over only a relatively small surface area, the strength of the adhesion interaction depends on the size of the objects in question. Small particles, for example, have a very small area of contact with the surface they rest upon. Similarly, even large objects make contact over a small surface area due to the surface roughness of most objects. In particular, as we go to larger and larger objects the role of atomic/intermolecular adhesion interactions shifts significantly. As an example, consider the spherical object in Fig. 8 that is acted upon by both the force of the earth's gravitational field and by the atomic and molecular interactions at the area of contact with the substrate. The gravitational force can be considered to act at the center of gravity of the sphere and is given by the elementary formula:

$$F_g = \frac{4}{3}\pi\rho g R^3 \tag{28}$$

where:

F_g = Gravitational force
ρ = Density of sphere (kg/m^3)
g = Local gravitational field (m/s^2)
R = Sphere radius (m)



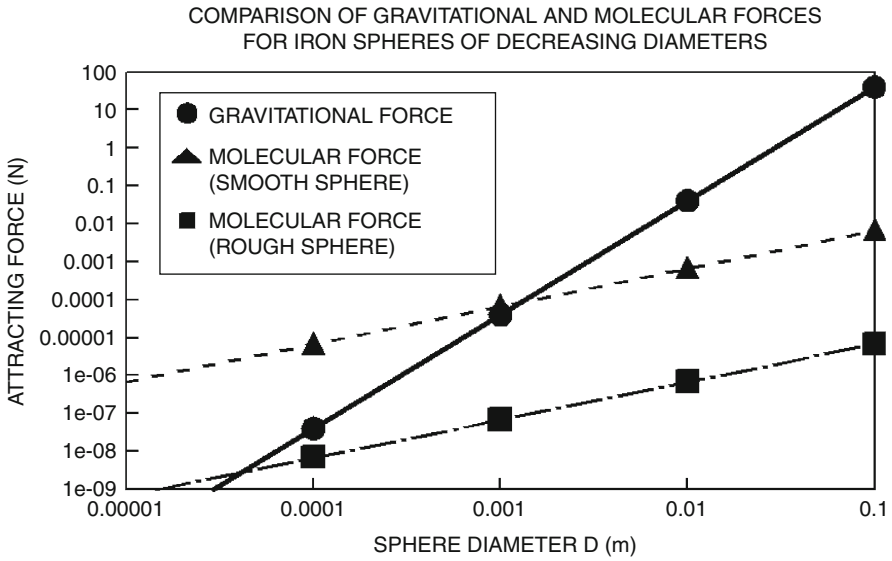
Basic Concepts in Adhesion Science, Fig. 8 Spherical object of some arbitrary material on a flat substrate acted upon by gravity and adhesion forces

The adhesion forces acting at the contact area can be inferred by a simple formula due to Bradley (1932):

$$F_a = 3\pi W_{AD}\frac{R}{4} \tag{29}$$

All symbols in (29) are the same as (28), with F_a being the adhesion force and W_{AD} the work of adhesion as defined in (19). This is a rather remarkable result because it implies that, whereas the gravitational force on the sphere scales as the third power of the radius, the adhesion force scales only linearly. The consequences of this are shown in Fig. 9, which is a log-log plot of the attractive force versus radius for iron spheres. For spheres of 1 cm diameter or greater the gravitational force is at least 10 times greater than the molecular adhesion interaction. However, scaling down to the 1 mm level, we find that in the case of smooth spheres the adhesion force and the gravitational force are closely balanced. Interestingly, for rough spheres the adhesion interaction is still very weak in this size range and becomes competitive only below diameters of 100 μm or less. This phenomenon is indicative of the very short range nature of the adhesion forces. Continuing to yet smaller diameters, on order of 10 μm in diameter the adhesion force for rough spheres is now competitive with the gravitational attraction and is completely dominant for smooth spheres. At the level of 1 μm the adhesion force dominates in all cases. This fact has definite consequences for the problem of removing particles in the 1 μm range from surfaces.

Reviewing what has been covered to this point, the reader should be aware that this brief essay has been something of a flyby overview of the basic field of adhesion science, starting with a look at the fundamental forces involved and ranging up to adhesion behavior of macroscopic entities such as small particles. Much had to be left



Basic Concepts in Adhesion Science, Fig. 9 Plot of attractive force vs diameter for sphere depicted in Fig. 8

out due to the shear enormity of the topic, in particular, the formation of covalent bonds or the weaker hydrogen bonding interactions that can occur in coating and adhesive systems. These interactions are highly specific to particular material systems and need to be carefully discussed on a case-by-case basis. Also, in polymer systems, one can have interchain penetration between two separate bodies that in effect gives rise to covalent binding. It is hoped, however, that enough material of interest has been presented to pique the reader's curiosity to explore further aspects of adhesion behavior that might be of interest.

Cross-References

► Adhesive Contact of Elastic Bodies
► Adhesion Hysteresis
► Contacts Considering Adhesion
► Interfacial Energy

References

A.W. Adamson, A.P. Gast, *Physical Chemistry of Surfaces*, 6th edn. (Wiley, New York, 1997), p. 278

R.S. Bradley, The cohesive force between solid surfaces and the surface energy of solids. Phil. Mag. **13**, 853 (1932)

P.A.M. Dirac, The quantum theory of the electron. Proc. R. Soc. Lond. Ser. A. **117**, 610 (1928)

I.E. Dzyaloshinskii, E.M. Lifshitz, L.P. Piatevskii, van der waals forces in liquid films. Sov. Phys. JETP **37**(10), 161 (1960)

I.E. Dzyaloshinskii, E.M. Lifshitz, L.P. Piatevskii, General theory of van der waals forces. Adv. Phys. **10**(38), 165 (1961)

F. Etzler, Characterization of Surface Free Energies and Surface Chemistry of Solids, in *Contact Angle, Wettability and Adhesion*, ed. by K.L. Mittal, vol. 3 (VSP, Utrecht, 2003), p. 222

R.P. Feynman, R.B. Leighton, M. Sands, *The Feynman Lectures on Physics*, vol. III (Addison-Wesley, Reading, 1963). chapter 19

R.H. French, Origins and applications of London dispersion forces and Hamaker constants in ceramics. J. Am. Ceram. Soc. **83**(9), 2117 (2000)

K. Gottfried, *Quantum Mechanics: Fundamentals*, vol. I (W. A. Benjamin, New York, 1966)

H.C. Hamaker, The London van der Waals attraction between spherical particles. Physica **4**, 1058 (1937)

J.N. Israelachvili, D. Tabor, Direct measurement of long range forces between two mica surfaces in aqueous KNO3 solutions. Nature, **236**, 106 (1972), also Proc. R. Soc. A. **331**, 19 (1972)

J.D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1962)

A. Lambrecht, *The Casimir Effect: a Force from Nothing*. Physics World, September 2002

S.K. Lamoreaux, Demonstration of the Casimir force in the 0.6 to 6 μm range. Phys. Rev. Lett. **78**, 5 (1997)

E.M. Lifshitz, The theory of attractive forces between solids. Sov. Phys. JETP **2**, 73 (1956)

E. Moy, A.W. Neumann, Theoretical Approaches for Estimating Solid–liquid Interfacial Tensions, in *Applied Surface Thermodynamics*, ed. by A.W. Neumann, J.K. Spelt. Surfactant Science Series (Marcel Dekker, New York, 1996)

J.W. Obreimoff, The splitting strength of mica. Proc. R. Soc. A. **127**, 290 (1930)

D. Park, *Introduction to Quantum Theory* (McGraw Hill, New York, 1964)

D.S. Rimai, D.J. Quesnel, *Fundamentals of Particle Adhesion* (Global Press, Moorhead, 2001)

S.S. Schweber, *QED and the Men who Made it: Dyson, Feynman Schwinger and Tomonaga*. Princeton Series in Physics (Princeton University Press, Princeton, 1994)

J.K. Spelt, D. Li, *The Equation of State Approach to Interfacial Tensions*. Surfactant Science Series, vol. 63 (Marcel Dekker, New York, 1996), p. 239

D. Tabor, R.H.S. Winterton, Surface energy and adhesion at metal contacts. Proc. R. Soc. A. **312**, 435 (1969)

Basic Rack

► Rack

BCC – Body Center Cubic

► Brittle-Ductile Transition

BDT – Brittle-Ductile Transition

► Brittle-Ductile Transition

BDTT – Brittle-Ductile Transition Temperature

► Brittle-Ductile Transition

Bearing Analysis

► Finite Element Method for Fluid Film Bearings

Bearing Classifications

► Rolling Element Bearing Selection

Bearing Components

► Function and Structure of Rolling Element Bearings

Bearing Corrosion

► Failure Mechanisms of Rolling Element Bearings

Bearing Endurance Test

► Rolling Bearing Test

Bearing Installation

► Rolling Element Bearing Mounting and Fitting Practice

Bearing Lubrication

► Rolling Bearing Lubrication

Bearing Materials

► Gas Bearing Materials

Bearing Mounting

► Rolling Element Bearing Mounting and Fitting Practice

Bearing Performance

► Function and Structure of Rolling Element Bearings

Bearing Removal

► Rolling Element Bearing Mounting and Fitting Practice

Bearing Selection

► Rolling Element Bearing Selection

Bearing Setting

► Rolling Element Bearing Mounting and Fitting Practice

Bearing Steels

► Rolling Bearing Materials

Bearing Structures

► Function and Structure of Rolling Element Bearings

Bearing Torque

► Rolling Bearing Torque

Bearing Types

► Rolling Element Bearing Selection

Bearing Wear

► Failure Mechanisms of Rolling Element Bearings

Bearing Wear Monitoring Using Electrostatic Charge

T. J. Harvey, M. Craig, R. J. K. Wood
National Centre for Advanced Tribology, Facility of Engineering and the Environment, University of Southampton, Hampshire, Southampton, UK

Synonyms

Electrostatic condition monitoring

Definition

The use of electrostatic charge as a condition monitoring technique in bearing wear.

Electrostatics as a Condition Monitoring Technique

Introduction

Condition monitoring interprets numerous sensors mounted on machines to indicate their health. Information can be extracted and used for fault diagnosis and predictive machinery maintenance. Newer condition monitoring techniques offer both prognostic and diagnostic capabilities. This avoids routine machinery downtime and unnecessary intervention associated with regular maintenance schedules. Condition monitoring will also reduce the risk of catastrophic failure and thus improve safety of machinery operation.

There are numerous techniques available to monitor machines. The four basic methods used in industry are performance monitoring (power, temperature, etc.), vibration monitoring, lubricant and wear debris monitoring, and acoustic emission monitoring. As a new technology, electrostatic sensing has been applied to detect the presence of foreign debris in the gas path of turbine engines (Powrie and McNicholas 1997; Powrie et al. 1994, 1999a, b).

The following sections will detail the fundamentals of electrostatics and its use as a condition monitoring tool.

Principles of Electrostatics

Consider an isolated point charge of Q coulombs. If the charge is enclosed in the center of a spherical Gaussian surface, Gauss' law states that the outward flux of E over any closed surface S is equal to the algebraic sum of the charges enclosed divided by ε, or

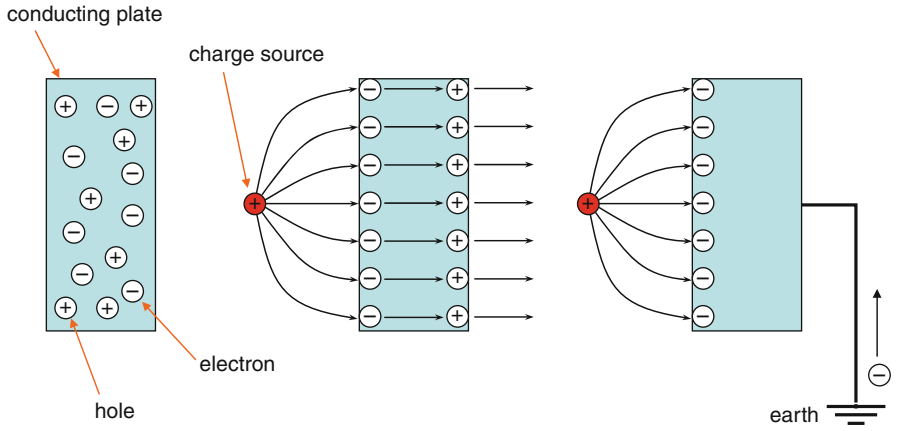
$$\sum Q = \varepsilon \oint_S E.dS \tag{1}$$

The flux of electric field strength, E, from the surface may be deduced by (2):

$$\frac{Q}{\varepsilon} = 4\pi x^2 E \tag{2}$$

where ε is the permittivity and x is the radius of the sphere or the distance between the particle and the surface.

If this Gaussian surface were placed just inside a Faraday cup, then all the charge Q would be induced on the cup surface and hence measurement of the total charge, Q could be made. Consider now a small surface of



Bearing Wear Monitoring Using Electrostatic Charge, Fig. 1 Principles of polarization in electrostatic sensor: (**a**) shows an isolated plate, (**b**) shows an isolated plate in the presence of charge source, and (**c**) shows the same plate connected to earth

area A, within the Gaussian surface. The flux of E from this surface may be approximated by (3):

$$\frac{Q_A}{\varepsilon} \sim AE \tag{3}$$

where Q_A is the charge induced on the surface.

Q_A will be a fraction of Q and can be deduced by dividing (3) by (2):

$$\frac{Q_A}{Q} \sim \frac{A}{x^2} \tag{4}$$

The area A is taken as the surface of the electrostatic sensor and hence its sensitivity to the total charge Q maybe deduced by rearranging (4):

$$Q_A \sim \frac{QA}{x^2} \tag{5}$$

In practice, other factors, such as the presence or proximity of other metal surfaces, will also influence the charge measured at the sensor face, so (4), at best, serves as an approximation of the sensor's response.

For surface charge, the sensor behaves more like a capacitor. The capacitance can be related to charge by (6).

$$C = \frac{Q}{V} = \varepsilon \frac{A}{x} \tag{6}$$

This can be rearranged to give charge:

$$Q = \frac{\varepsilon VA}{x} \tag{7}$$

The potential difference (V) between the sensor and the surface will be proportional to the charge on that surface and thus can be approximated by (8).

$$Q_A \sim \frac{QA}{x} \tag{8}$$

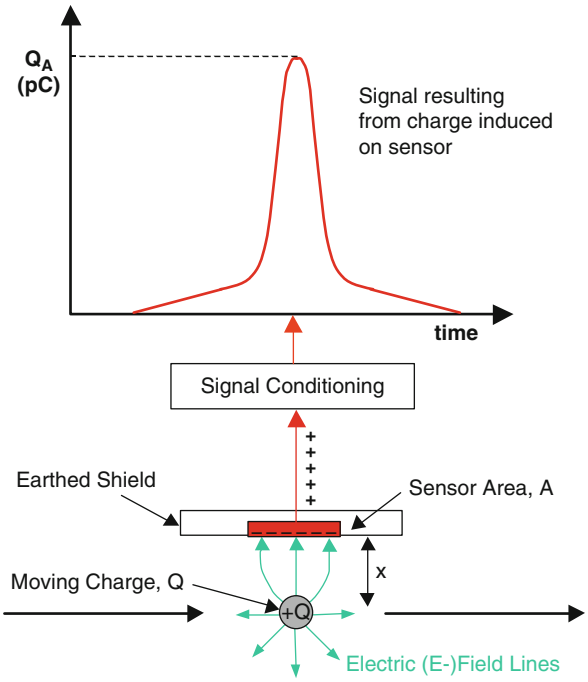
Principles of Electrostatic Sensing

If a charge source (debris particle, surface, etc.) passing in front (or passing the field-of-view) of the sensor field lines will terminate on the sensing face. This will induce free electrons to be drawn to or away from the surface of the sensor (depending on the polarity of the charge source) due to electrostatic attraction, see Fig. 1b. Holes will move in the opposite direction. This phenomenon is known as polarization or induction.

If the sensor is connected to earth, electrons will flow to or from earth (depending on the polarity of the charge source) to counteract the polarization. In an electrostatic system, the sensor (see the following section) is connected to a signal-conditioning unit (a charge amplifier). This signal conditioner will measure the flow of charge (electrons) and produce a measurable output. Figure 2 shows a schematic diagram showing the process of charge detection in an electrostatic sensor as moving charge passes the sensor face.

Electrostatic Sensing Sensors

Electrostatic sensors are passive inductive sensors that can take several geometrical forms, for example planar button or cylindrical ring. The elements/components that all electrostatic sensors have in common are a sensing surface/element, a surround insulator so that the charge does not leak away, an earth shield that minimizes electrostatic and radio frequency interference, and electrical connections to the sensing surface/element and earth shield, which are connected to the measuring equipment,



Bearing Wear Monitoring Using Electrostatic Charge, Fig. 2 Schematic illustration of the electrostatic charge sensing system

generally through a conditioning unit (such as a charge amplifier).

Button-type sensors are especially good for monitoring charge on surfaces (Morris et al. 2002) (see Fig. 3), such as wear sites, and thus are generally referred to as "*wear-site sensors*."

In practice, wear-site sensors are designed to be positioned as close as possible to the surface of interest, as shown in Fig. 4. In existing equipment, such as bearing test housings, modifications to allow penetration of the wear-site sensor face close to the component of interest are needed. Additionally, the sensors are designed to achieve penetration with the least disruption to functionality and structural integrity and thus are generally long and cylindrical-shaped, as shown in Fig. 5. In this case, the sensor face is the whole of the tip of the sensor; in Fig. 5 this can be distinguished by the insulating black layer at the end of the sensor. If the local environment is too noisy (electrostatically), an earthed sheath can be included in the design to effectively focus the sensor face solely to the region of interest.

Ring sensors are employed in pipe flow to detect charge on particles driven by fluid motion to act as a flow meter (Gajewski and Szaynok 1981; Gajewski 1984; 1996; 1997; Gajewski et al. 1993) or to detect the presence of wear debris in a lubrication oil line (Powrie et al. 1999d). The advantage of the ring sensors is that the presence of any charged particle in the flow will be detected and is designed to be non-intrusive to flow. Again, these electrostatic sensors have been referred to previously by their application, in this case oil-line sensor (OLS). In general, they consist of a conductive ring(s) used as the sensing element surrounded by insulating material; the exterior body is usually a conducting body that acts as a shield for electrostatic and radio frequency interference; electrical connections are made with sensing element(s); additionally the shield is electrically connected to earth. For monitoring bearings, the sensor consists of two identical ring sensing elements at a predetermined separation, an example is an OLS (Fig. 6). With a known separation (and fluid flow rate), the output from the rings can be analyzed using a cross-correlation function (Baar and Riess 1997), which is useful in determining the presence of charged entities, such as debris particles, that have drag and are subsequently moving slower than the fluid.

Scientific Fundamentals

Electrostatic charge is measured from four sources:

(a) Tribocharging
(b) Surface charge
(c) Triboemissions and
(d) Wear debris generation

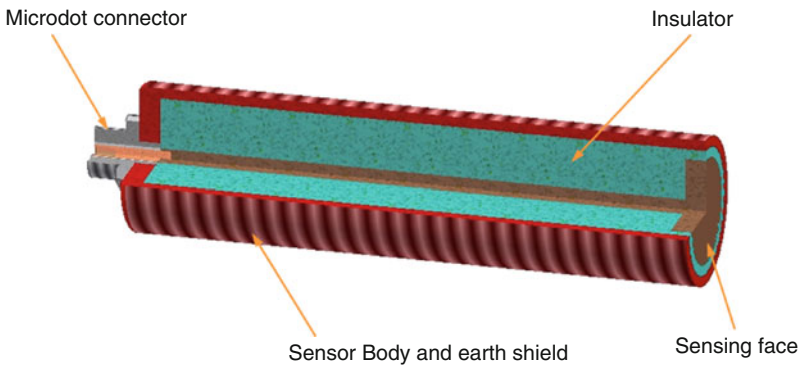
Charge Generation Mechanisms

Introduction

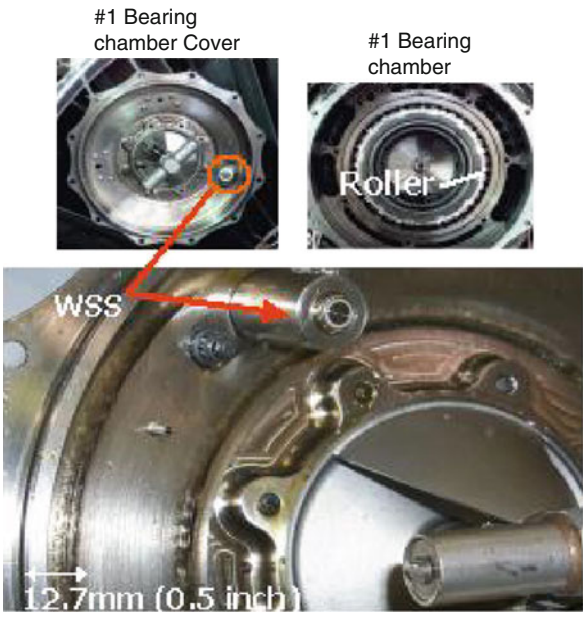
Precursor charge events have been detected prior to the onset of scuffing failure in lubricated sliding contacts (Tasbaz et al. 1999a). The precursor charge was related to one or more charge mechanisms, namely tribocharging, surface charge, triboemissions and wear debris generation. These charge generation mechanisms are complex and affected by chemical and physical properties of the lubricant and contacting materials, as well as any tribochemical changes occurring between the contacting materials. A schematic diagram of the possible charge generation mechanisms as a consequence of frictional work and/or wear is shown in Fig. 7. Details of the different charge mechanisms are discussed below.

Tribocharging

Tribocharging is a term used to describe the charging of a low-conductivity fluid by its relative motion over a surface. An electrical double layer forms naturally at



Bearing Wear Monitoring Using Electrostatic Charge, Fig. 3 Schematic diagram of electrostatic button sensor that is employed at the University of Southampton; the long sensor body is threaded (Booth et al. 2010)



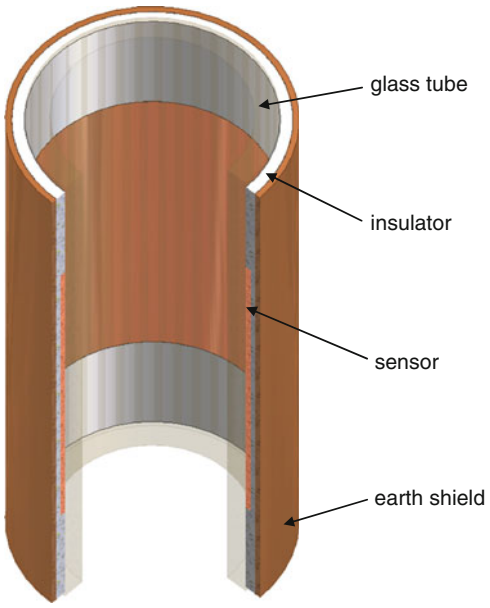
Bearing Wear Monitoring Using Electrostatic Charge, Fig. 4 Wear-site sensor location in F100 seeded fault engine tests, conducted by Pratt and Whitney. Four electrostatic sensors were installed in the oil system of the F100 engine to monitor bearings #1 and #5 seeded fault engine test (Powrie et al. 1999c, 2000)

Bearing Wear Monitoring Using Electrostatic Charge, Fig. 5 Electrostatic Wear-Site Sensors (WSS) used in bearing test facility at University of Southampton

any solid–liquid interface. A charged metal surface attracts opposite charges and repels like charges. Several theories relating to double layer structure and formation have been proposed by Helmholtz (1879), Gouy (1910), Chapman (1913), Stern (1924), and later by Grahame (1947), and have been reviewed more recently by Parsons (1990). When the motion of the fluid shears this double layer at the interface between the solid surface and the liquid phase it results in the liquid carrying part of the electric charge of the double layer, thus generating free electrostatic charges (Harvey et al. 2002a), as shown in Fig. 8. The effect of flow velocity on streaming currents in pipes can be given by (9):

$$I_{\infty} = Kd^{0.88} V_{av}{}^{1.88} \tag{9}$$

Where d is the pipe diameter, I_{∞} is stream current, V_{av} is the average flow velocity, and K is a constant (Tayler and Secker 1994).



Bearing Wear Monitoring Using Electrostatic Charge, Fig. 6 Schematic diagram of electrostatic ring sensor

Tribocharging could be a potential charge source in lubricated wear processes due to the high shear nature of contact (Harvey et al. 2002a). Harvey et al. (2002a) investigated the effect of the oil quality on electrostatic charge generation and electrostatic charging mechanism in oil using a variable speed spinning disc charger. While controlling the oil viscosity, temperature, aging, and disc surface roughness, tribocharging was measured as a function of disc rotation speed. The authors found a direct relationship between the tribocharging levels of base oils and oil viscosity and disc surface roughness, while temperature effects indicated a maximum in chargeability due to transportation-recombination competition, which is more commonly referred to as charge relaxation. Oil aging experiments indicated an increase in conductivity, which would lead to a decrease in charging ability.

Triboemissions

Excitation of the surface by tribological action can lead to the generation of particles and light and is commonly referred to as triboemissions. The particles can be charged, such as electrons, and can have low and high energy. Triboemissions have been extensively examined by Nakayama et al. (Nakayama and Hashimoto 1995a, b; Nakayama 1992, 1994, 1996, 1999; Nakayama et al. 1993), who studied this phenomenon on different solids under various atmospheres and conditions. The work involved the scratching of various metals, semiconductors, and insulators, on a modified pin-on-disc machine, using a diamond stylus with different tip diameters. The general trend obtained from these results indicate the intensity of emitted negative charge increases in the order of conductor<semiconductor<insulator. However, the reverse order was found for the intensity of photon emission (i.e., insulator<semiconductor<conductor).

Kajdas et al. (2001) have reported that triboemission is particularly important in boundary lubrication. It was also found that triboemitted particles include low energy electrons (these are important as they react with the lubricant molecules, leading to degradation and tribochemical reactions).

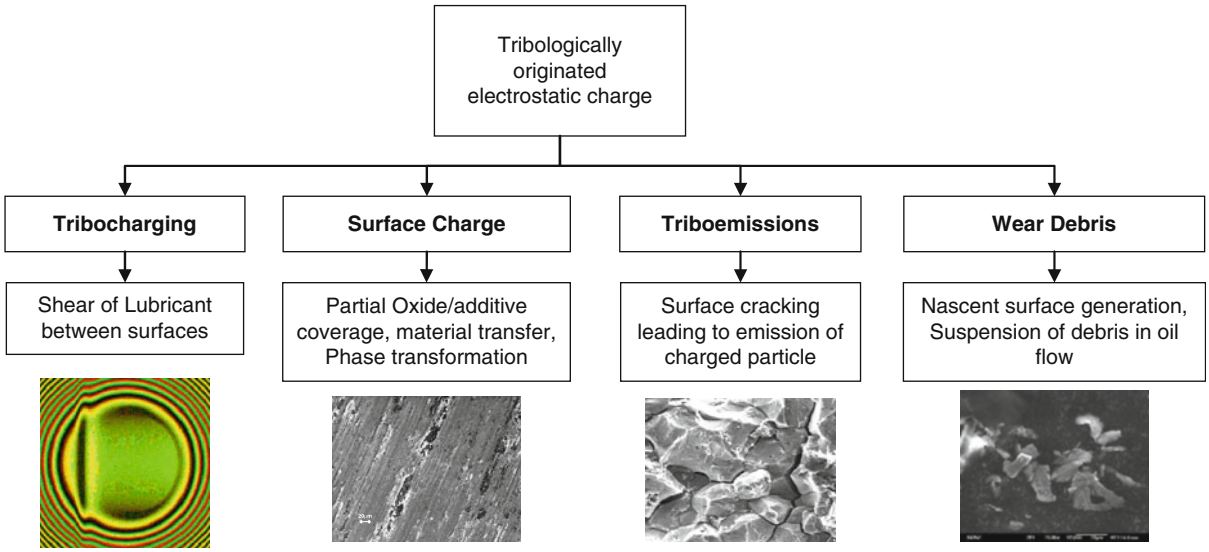
Surface Charge

Surface charge is the imbalance of charge on the surface of the worn material(s) and contact potential differences (CPD) that are caused by different metals or tribologically generated phase transformed regions or oxide stripping (Kasai et al. 1999). Kelvin probe work with a modified atomic force microscope (AFM) by DeVecchio and Bhushan (1998) has detected chemical and structural changes to specimen surfaces caused by nanoscale wear. Zharin and Rigney (1998) have studied wear events with a vibrating Kelvin probe and reported that the probe was sensitive to contact potential difference changes associated with wear. Work function is defined as the minimum work required, in a vacuum, to extract an electron from the Fermi level of a conducting phase through a surface. When two metals, M_a and M_b, are brought into contact, their work functions (ϕ_{Ma} and ϕ_{Ma}) will equalize, see Fig. 9 (Harvey et al. 2002b). This results in a contact potential difference, V_{CPD}, between the two surfaces that is related to the difference in work function, as shown in (10).

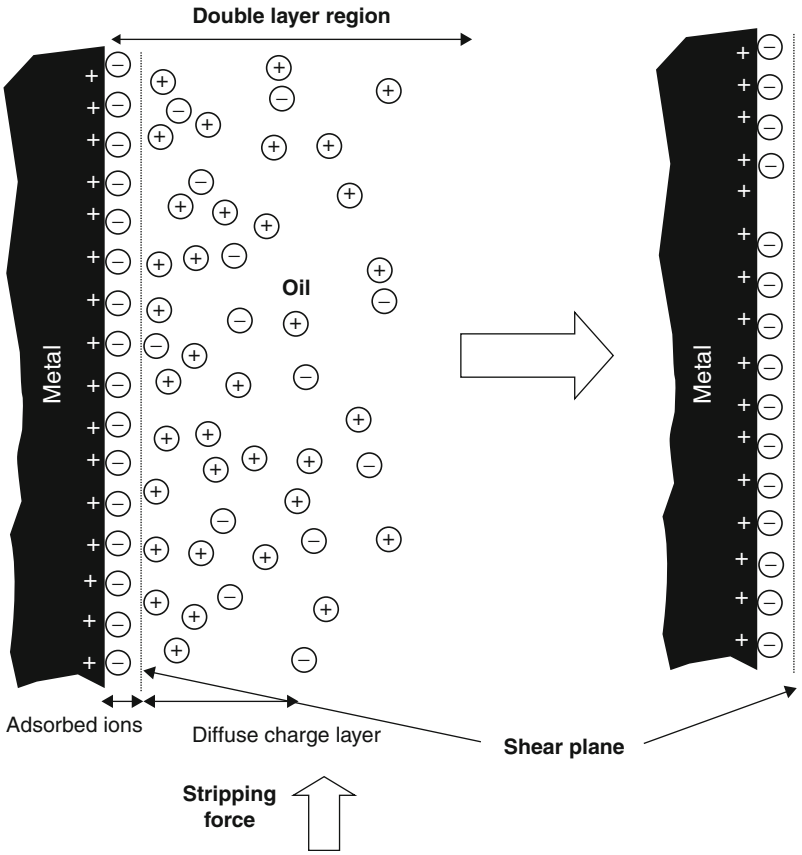
$$eV_{PCD} = \phi_{Ma} - \phi_{Mb} \tag{10}$$

In steel, phase transformations can appear as *white layers* (Tasbaz et al. 1999a, b; Stead 1911–1912; Trent 1941; Rogers 1969; Torrance 1978; Glenn and Leslie 1971; Eyre et al. 1972; Newcombe and Stobbs 1984), namely due to their silvery appearance upon etching, see Fig. 10. These *white layers* are produced during initial asperity-asperity contacts, friction, or cold pressure welding and the resulting high, localized contact temperature transforms the region affected. *White layers* are known to have higher hardness values than the parent bulk material (Tasbaz et al. 1999b).

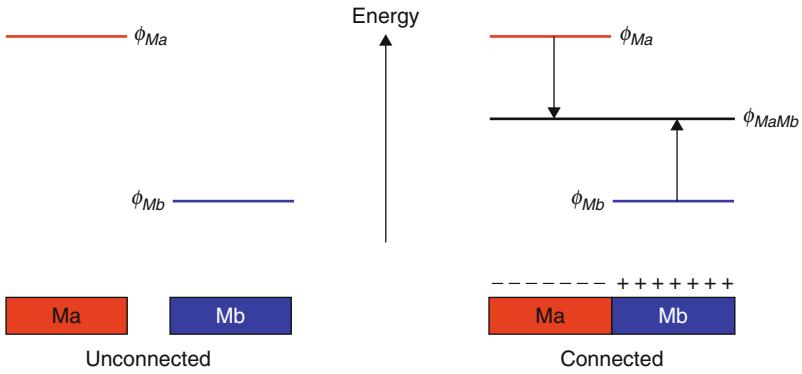
The ability of electrostatic monitoring to detect CPD has been investigated using a bimetallic disc, manufactured by brazing a copper strip into a disc of



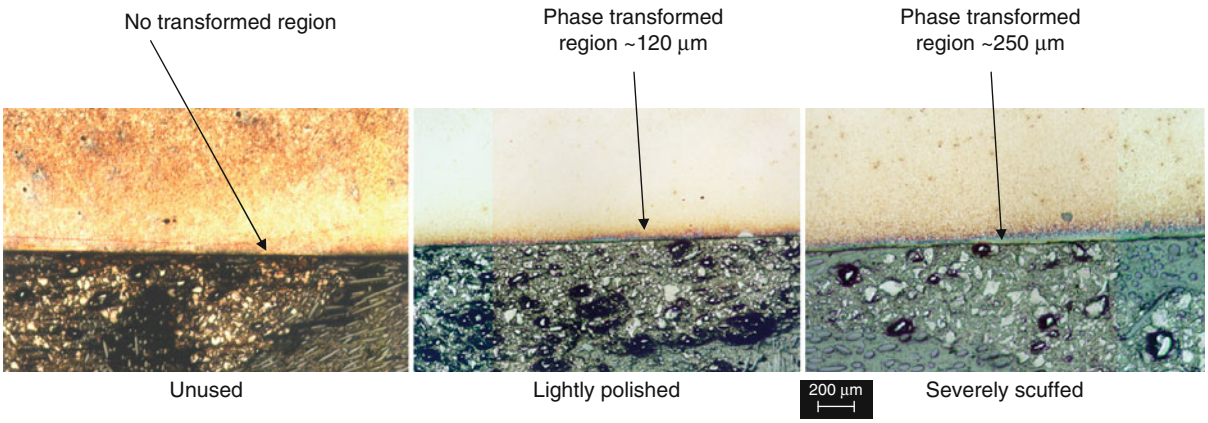
Bearing Wear Monitoring Using Electrostatic Charge, Fig. 7 Charge generation mechanisms within tribological contacts



Bearing Wear Monitoring Using Electrostatic Charge, Fig. 8 Schematic diagram of an electrical double layer, the *dotted line* indicates the shear line (Harvey et al. 2002a)



Bearing Wear Monitoring Using Electrostatic Charge, Fig. 9 Schematic diagram of contact potential difference theory (Harvey et al. 2002b)



Bearing Wear Monitoring Using Electrostatic Charge, Fig. 10 Sectioned gear teeth, from FZG test rig, showing phase (*white* layers) transformation growth as wear severity increases (Tasbaz et al. 1999b)

bearing steel (En31). The experiments produced negative peaks corresponding to the position of the copper strip, indicating that it charges negatively with respect to the steel (Tasbaz et al. 1999b).

Morris et al. (2002) also investigated the detection of contact potential difference using electrostatic sensing technology. The authors carried out non-contact and scuffing tests using a pin-on-disc wear rig. Non-contact tests were carried out to investigate the sensitivity of the electrostatic sensor to variation in surface contact potential difference by inserting copper, aluminum, En31 and carbon steel inserts in a bearing steel disc. These non-contact tests demonstrated that the electrostatic sensors are sensitive to charges generated at the interface when two different metals are present on the disc. The localized surface charge characteristics may be predicted from the work functions of the two metals, which indicate that the electrostatic sensor is detecting the CPD phenomenon. Of particular significance is the carbon steel test, which shows that electrostatic sensor is sensitive to small changes in surface composition.

Wear Debris Generation

Debris formation occurs when touching asperities adhere together and plastic shearing removes the tip of the softer asperities, leaving them adhering to the harder surface. Subsequently, these tips can become detached, giving rise to wear particles (Williams 1994).

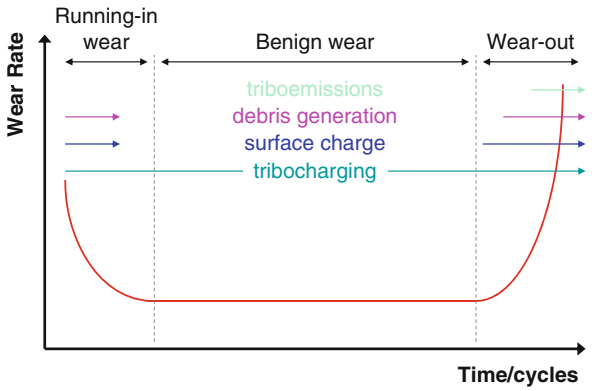
The formation of debris involves the highly energetic shearing of surface asperity tips. The site of asperity detachment will produce fresh metal (oxide free) that is typically work hardened. The debris formed is thought to

be charged and can be detected by an electrostatic sensor at the tribo-contact and at remote locations, such as downstream pipework, should metallic debris become suspended in the oil (and avoid contact with earthed machinery surfaces) or discharge in conducting liquids.

Morris et al. (Harvey et al. 2007a) carried out tests on a modified pin-on-disc rig using a sliding point contact fitted with electrostatic sensors, which monitored the disc wear track and the disc surface just outside the wear track. Baseline tests were performed using mild wear conditions with no seeded particles added to the entrained lubricant, while the high wear tests entrained seeded steel particles into the contact to promote wear. It was observed that the wear debris produced dynamic charge features and the overall charging activity was related to the wear rate. It was also observed that the adhesive wear mechanism was predominantly observed in the normal wear test (non-seeded test). The debris seeding test produced a mixture of two-body and three-body rolling abrasion. Debris produced by adhesive wear was found to be positively charged, while seeding particles and induced wear debris were found to be negatively charged.

Wear Regime and Electrostatic Charge Mechanism

Figure 11 indicates where charge mechanisms are thought to occur during the life of a lubricated tribocontact/component. As can be seen, tribocharging occurs throughout and will always be present if any lubricated is present in a moving contact and will generate a baseline charge level for both wear-site and oil-line electrostatic sensors. Running-in occurs in almost every tribological component where the mating surfaces conform to each other by plastic deforming or shearing asperities. The levels will depend on the levels of conforming needed and the load applied. As the surfaces conform there will be debris generation and surface charging. Once the surfaces have conformed, indicated by a drop in wear rate, the component will experience steady state/benign wear. The final stage of the tribocontact/component after steady-state/benign wear is "wear out." At some point, either due to an accumulation of stress, damage, or lubricant starvation the contact will begin to break down. In bearings, this is generally related to rolling contact fatigue (RCF) and the first sign of degradation is the initial spallation; as the spallation progresses, the contact becomes rougher and leads to collapse of lubricant film, further accelerating wear. Wear debris produced can produce charge either directly or by becoming re-entrained into the contact, where it is generally extruded into the typical platelet/flake debris observed during bearing wear.



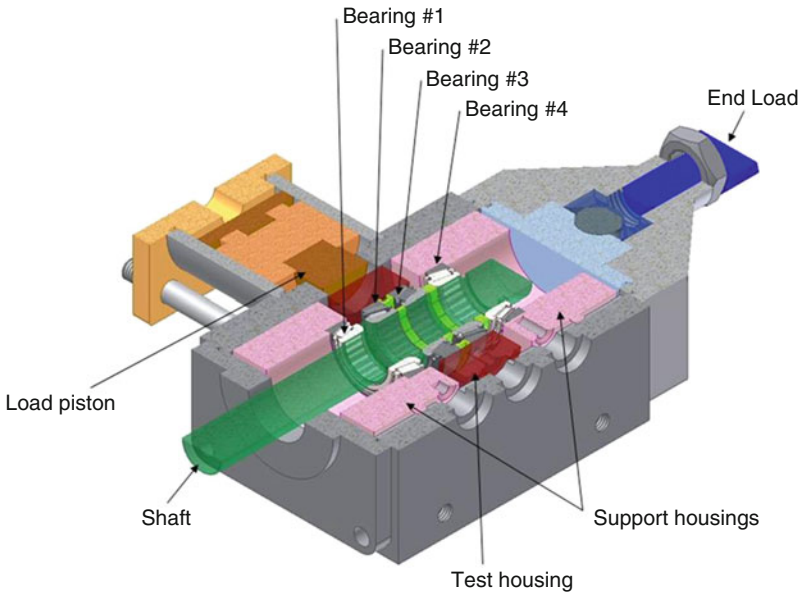
Bearing Wear Monitoring Using Electrostatic Charge, Fig. 11 Idealized "bathtub curve" for predicted regions where charge generation mechanisms are thought to occur

Application of Electrostatics to Tribology

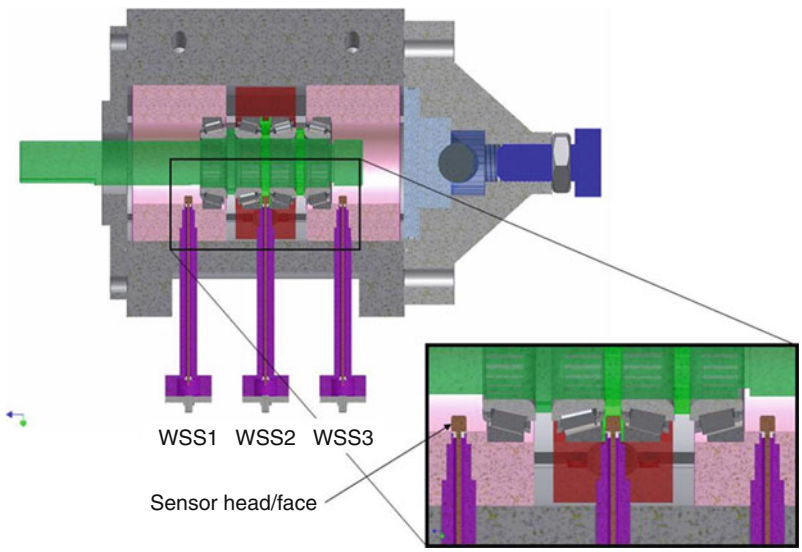
Previous studies at Southampton have shown that electrostatic sensing can successfully measure tribocharging (Harvey et al. 2002a, b) and detect wear debris (Harvey et al. 2007a), surface transformation (Morris et al. 2002; Tasbaz et al. 1999a), ceramic wear (Wang et al. 2003a, b, 2004), and changes in oil chemistry and surface tribofilm (Booth et al. 2006a, b; Wood et al. 2005; Wang and Wood 2007) due to CPD effects. Debris monitoring of steel wear debris shows the magnitude of charge was related to the total wear volume (Morris et al. 2002). Steel-on-steel adhesive scuffing in oil-lubricated sliding gave precursor electrostatic signals that were related to surface phase transformations associated with the initial stages of adhesive failure (Morris et al. 2002; Tasbaz et al. 1999a). Electrostatic sensing of oil chemistry and surface tribofilms has been studied by looking at interactions between oil additives as well as their interactions with metallic surfaces and contaminants (i.e., soot) (Booth et al. 2006a, b; Wood et al. 2005). Surface charge measurement was shown to provide information about the adsorption process, the kinematics of tribofilm growth/decay, and the life compositional changes of these surface films.

Electrostatic Monitoring of Rolling Element Bearings

For an example of how the electrostatic sensing technique can be applied to rolling element bearings, an electrostatic monitoring system was installed onto a taper-roller bearing test rig (Harvey et al. 2006, 2007b), see Fig. 12. The electrostatic system comprised three wear-site sensors (WSS) and an oil-line sensor (OLS). The WSSs were mounted close to the bearings, see Fig. 13, in order to

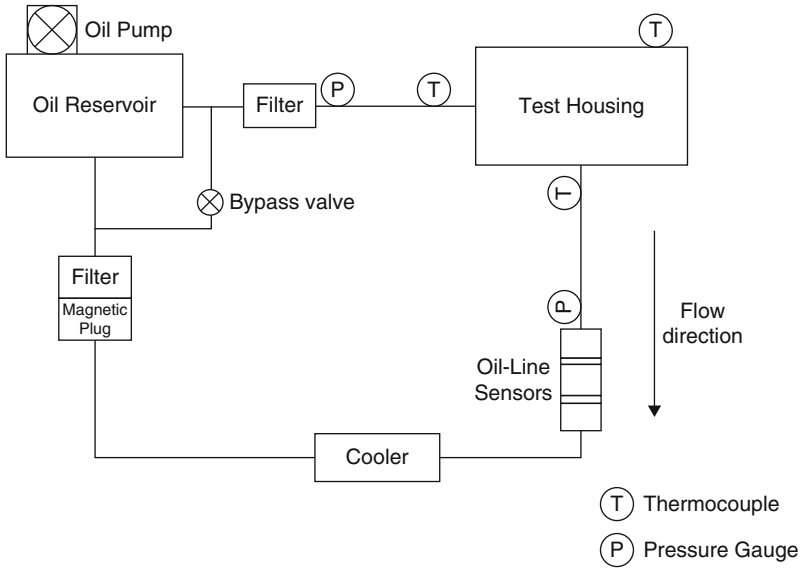


Bearing Wear Monitoring Using Electrostatic Charge, Fig. 12 Schematic showing the loading of the test bearings and the positions of the vibration mounting points is shown by the holes in the front of the rig connecting to the test and support housings

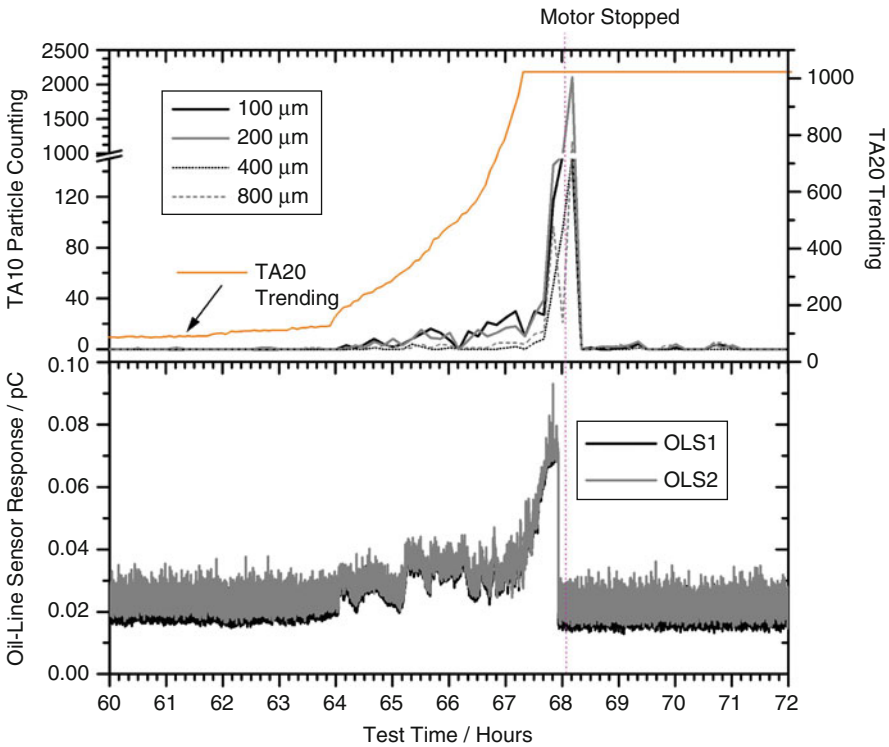


Bearing Wear Monitoring Using Electrostatic Charge, Fig. 13 Schematic diagram showing placement of electrostatic wear-site sensors (WSS) in housings (Harvey et al. 2006)

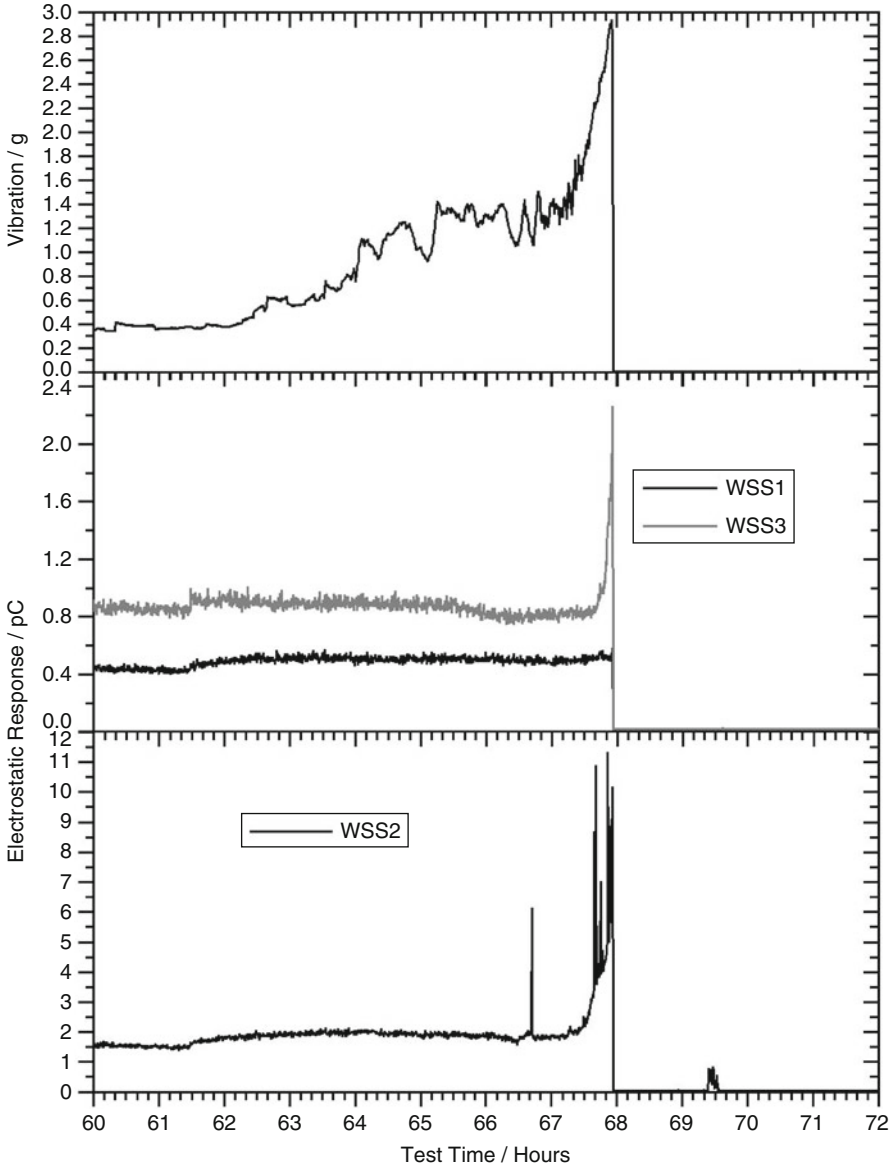
monitor the earliest stages of component surface breakdown and wear, whereas the OLS was placed in the recirculation oil-line to monitor debris generated and transported from the deteriorating component, see Fig. 14. Vibration and lubricant temperature measurements were also made to allow comparison with conventional CM techniques. Additionally, two oil-line monitoring devices, using Eddy current technology and



Bearing Wear Monitoring Using Electrostatic Charge, Fig. 14 Schematic of oil recirculation system (first filter is a 25 μm and the second, next to the magnetic plug, is 6 μm)



Bearing Wear Monitoring Using Electrostatic Charge, Fig. 15 Thermocouple, Macom, and electrostatic oil-line dat, for accelerated bearing test showing final stage of test (60–72 h)

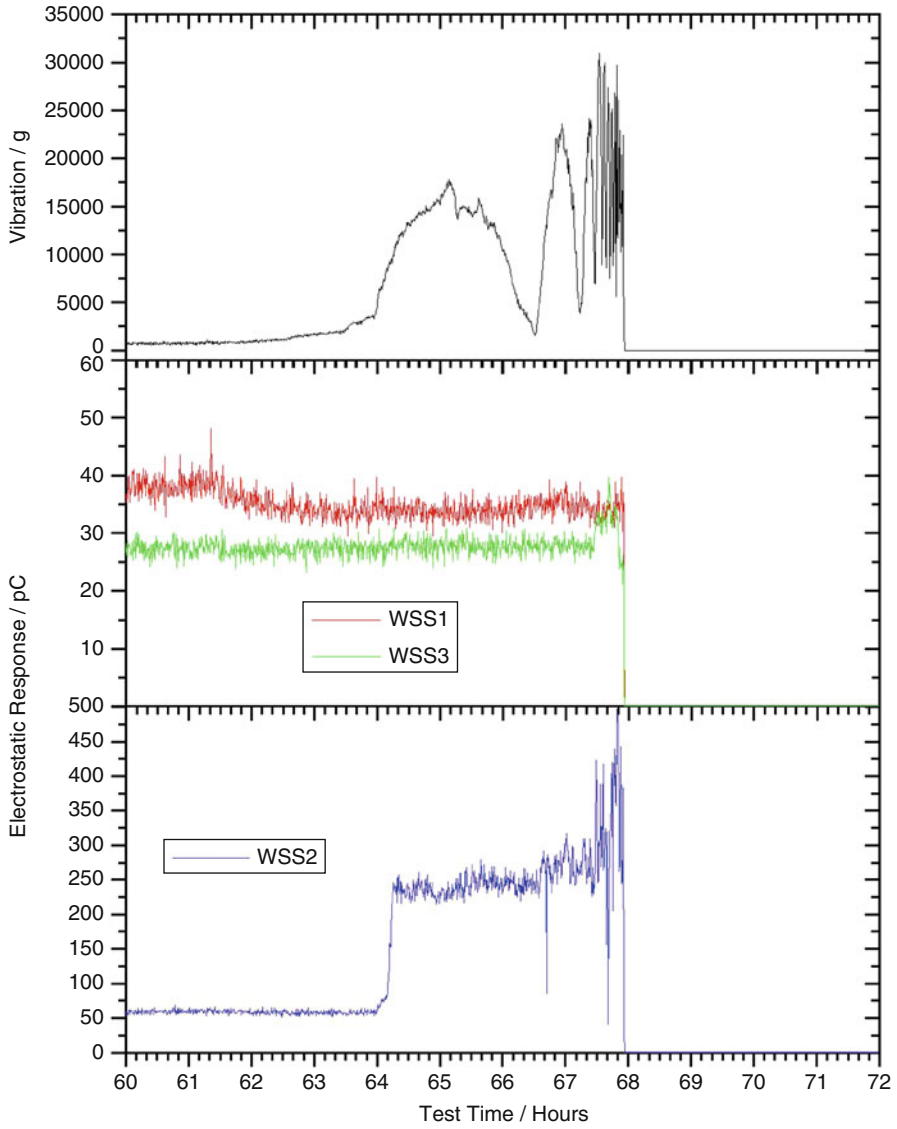


Bearing Wear Monitoring Using Electrostatic Charge, Fig. 16 Vibration and electrostatic wear-site sensor data (RMS), for accelerated bearing test showing final stage of test (60–72 h) (Harvey et al. 2006)

ferromagnetism to sense ferrous debris entrained in the lubricant recirculating system, were employed.

The test was accelerated using an artificial defect on the inner race of the test bearing. The artificial defect was induced by pre-indentation using a Rockwell Hardness testing machine using a "C" type diamond indenter. The test was also accelerated by overloading the test bearings, approximately 200% of the dynamic radial rating of the bearings.

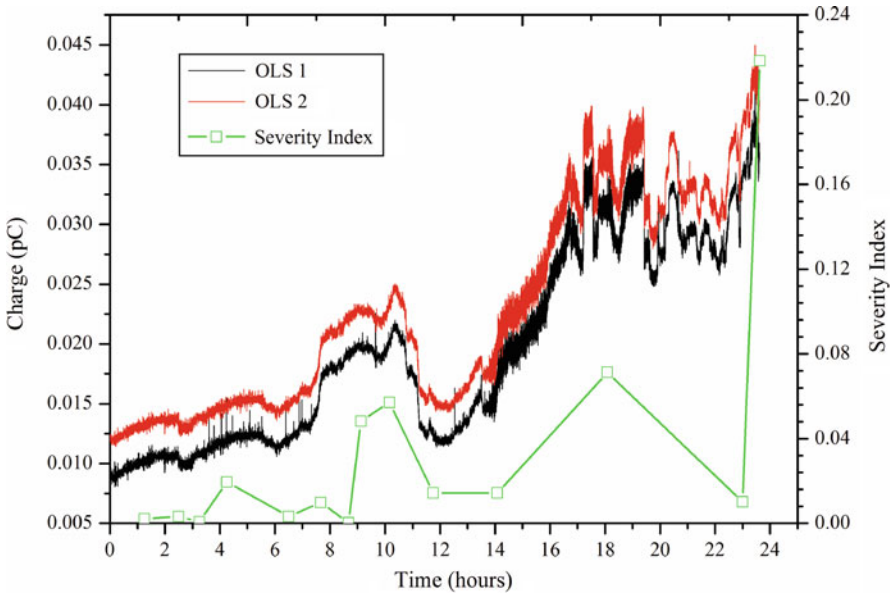
A fatigue-generated failure was detected in bearing deterioration up to 4 h prior to complete seizure. The electrostatic system indicated distress by increased charge levels in both the WSS and OLS signals. These increases coincided with the detection of debris generated by the Eddy current and ferromagnetic detectors, as well as increased vibration levels, see Figs. 15 and 16. Enhanced detection of bearing distress was achieved by normalization by inner race defect frequency in both electrostatic



Bearing Wear Monitoring Using Electrostatic Charge, Fig. 17 Expansion of vibration and electrostatic wear-site sensor data (normalized by inner race defect frequency), for 60–72 h

WSS and vibration signals, see Fig. 17. Defect frequency analysis is performed by a Fourier transform on the data, the subsequent frequency domain data is then windowed around the component's defect frequency (calculated from the rotation speed), and this windowed data is averaged. This enhances sensitivity to damage of the bearing components (inner race, outer race, roller, and cage) and is a common tool for analysis of vibration data. Post-test analysis confirmed extensive fatigue damage to the inner race of test bearing 2 (site of the artificial defect), which corroborated the increased levels of electrostatic activity detected by WSS2 (which monitored bearing 2, and bearing 3).

More recent work has incorporated other condition monitoring techniques to examine the capability of electrostatics for monitoring bearing degradation. Figure 18 shows data from a bearing test where off-line measurements were conducted by taking oil samples periodically



Bearing Wear Monitoring Using Electrostatic Charge, Fig. 18 Data indicating correlation between OLS online measurements with off-line measurements (oil samples analyzed by ferrography)

throughout the test. The oil samples were examined using ferrography[1]; from this the levels of debris particles below 5 μm and above 15 μm were assessed. It is generally considered that debris below 5 μm are produced during normal operation/wear and the larger 15 μm particles are an indication of abnormal wear. Thus, by comparing the relative amounts of these two sizes of particles, expressed as the severity index, an assessment of the system can be obtained. As can be seen from Fig. 18, the off-line ferrography (severity index) measurements follow the online OLS measurements, which gives confidence that the OLS is detecting debris within the oil recirculation system.

Cross-References

- ► Electrostatic Field Effects on Adhesion
- ► Electrostatic Field Effects on Friction
- ► Rolling Bearing Condition Monitoring
- ► Rolling Bearing Contact Fatigue
- ► Rolling Bearing Test
- ► Wear of Bearings

[1]Ferrography is an off-line debris technique that was developed in the 1970's to find a cost effective method of determining the condition of the lubricant from jet engines. This method separates ferrous wear debris according to size and non-magnetic contaminants such as sand or dust

References

R. Baar, W. Riess. Two phase flow velocimetry measurements by conductive-correlative method. Flow Measurement and Instrumentation **8**(1), 1–6 (1997). doi:10.1016/S0955-5986(97)00020-4

J.E. Booth, K.D. Nelson, T.J. Harvey, R.J.K. Wood, L. Wang, H.E.G. Powrie, J.G. Martinez, The feasibility of using electrostatic monitoring to identify diesel lubricant additives and soot contamination interactions by factorial analysis. Tribol. Int. **39**, 1564–1575 (2006a)

J.E. Booth, T.J. Harvey, R.J.K. Wood, H.E.G. Powrie, in *Advances in Electrostatic Health Monitoring of Tribo-contacts*, IMECHE Tribology Group 50th Anniversary Conference. Tribology 2006: Surface Engineering and Tribology for Future Engines and Drivelines, London, 12–13 July 2006b

J.E. Booth, T.J. Harvey, R.J.K. Wood, H.E.G. Powrie, Scuffing detection of TU3 cam-follower contacts by electrostatic charge condition monitoring. Tribol. Int. **43**(1–2), 113–128 (2010). doi:10.1016/j.triboint.2009.05.004

D.L. Chapman, A contribution to the theory of electrocapillarity. Philos. Mag. **1**, 475–481 (1913)

D. DeVecchio, B. Bhushan, Rev. Sci. Instrum. **69**(10), 3618–3624 (1998)

T.S. Eyre, A. Baxter, The formation of white layers at rubbing surfaces. Tribology **5**(6), 256–261 (1972)

J.B. Gajewski, Charge measurement of dust particles in motion 2. J. Electrost. **15**(1), 67–79 (1984)

J.B. Gajewski, Electrostatic, inductive ring probe bandwidth. Meas. Sci. Technol. **7**(12), 1766–1775 (1996)

J.B. Gajewski, Dynamic effect of charged particles on the measuring probe potential. J. Electrost. **40–1**, 437–442 (1997)

J.B. Gajewski, A. Szaynok, Charge measurement of dust particles in motion. J. Electrost. **10**, 229–234 (1981)

J.B. Gajewski, B.J. Glod, W.S. Kala, Electrostatic method for measuring the 2-phase pipe-flow parameters. IEEE Trans. Ind. Appl. **29**(3), 650–655 (1993)

R.C. Glenn, W.C. Leslie, The nature of white streaks in impacted steel armor plate. Met. Trans. **2**, 2945–2947 (1971)

G. Gouy, Constitution of the electric charge at surface of an electrode. J. Phys. **9**(4), 457–67 (1910)

D.C. Grahame, The electrical double layer and the theory of electrocapillarity. Chem. Rev. **41**(2–3), 441–501 (1947)

T.J. Harvey, R.J.K. Wood, G. Denuault, O.D. Tasbaz, H.E.G. Powrie, Effect of oil quality on electrostatic charge generation and transport. J. Electrost. **55**(1), 1–23 (2002a)

T.J. Harvey, R.J.K. Wood, G. Denuault, H.E.G. Powrie, Investigation of electrostatic charging mechanisms in oil lubricated tribo-contacts. Tribol. Int. **35**, 605–614 (2002b)

T.J. Harvey, R.J.K. Wood, H.E.G. Powrie, in *Electrostatic Monitoring of Taper Roller Bearing, Tribology 2006: Surface Engineering and Tribology for the Future Engines and Drivelines*, IMechE Tribology Group – 50th Anniversary Conference, London, July 2006

T.J. Harvey, S. Morris, R.J.K. Wood, H.E.G. Powrie, Real-time monitoring of wear debris using electrostatic sensing techniques. Proc. Inst. Mech. Eng., Part J: J. Eng. Tribol. **221**(J1), 27–40 (2007a)

T.J. Harvey, R.J.K. Wood, H.E.G. Powrie, Electrostatic wear monitoring of rolling element bearings. Wear **263**(7–12), 1492–1501 (2007b)

C. Kajdas, Tribochemistry, In: F. Franek, W.J. Bartz, A. Pauschtz (Eds.), Tribology 2001, Second world tribology congress, The Austrian Tribology Society, Vienna (2001), pp. 39–46

T. Kasai, X.Y. Fu, D.A. Rigney, A.L. Zharin, Applications of a non-contacting Kelvin probe during sliding. Wear **225–229**, 1186–1204 (1999)

S. Morris, R.J.K. Wood, T.J. Harvey, H.E.G. Powrie, Use of electrostatic charge monitoring for early detection of adhesive wear in oil lubricated contacts. ASME J. Tribol. **124**(2), 288–296 (2002)

K. Nakayama, Triboemission of charged particles and photons from solid surfaces during frictional damage. J Phy D: Appl. Phys. **25**, 303–308 (1992)

K. Nakayama, Triboemission of charged particles from various solids under boundary lubrication conditions. Wear **178**(1–2), 61–67 (1994)

K. Nakayama, Tribocharging and friction in insulators in ambient air. Wear **194**(1–2), 185–189 (1996)

K. Nakayama, Triboemission of charged particles and resistivity of solids. Tribol lett **6**, 37–40 (1999)

K. Nakayama, H. Hashimoto, Triboemission of charged particles and photons from ceramic surfaces in various hydrocarbon gases. Wear **185**(1–2), 183–188 (1995a)

K. Nakayama, H. Hashimoto, Effect of surrounding gas pressure on triboemission of charged particles and photons from wearing ceramic surfaces. Tribol. Trans. **38**(1), 35–42 (1995b)

K. Nakayama, H. Hashimoto, Triboemission of charged particles and photons in hydrocarbon atmospheres, in *Proceedings of the 6th International Congress on Tribology, Eurotrib'93* (Budapest, Hungary, 1993)

S.B. Newcombe, W.M. Stobbs, Transmission electron microscopy study of the white etching layer on a rail head. Mater. Sci. Eng. **66**, 195–204 (1984)

R. Parsons, Electrical double layer: recent experimental and theoretical developments. Chem. Rev. **90**(5), 813–826 (1990)

H.E.G. Powrie, C.E. Fisher, J. Nurse, Engine gas path integrity monitoring, in *Aerotech 94* (Aerotech 94 in Birmingham, January 1994), C470/6/052

H.E.G. Powrie, K. McNicholas, Gas path condition monitoring, during accelerated mission testing of demonstrator engine, AIAA-1997-2904, AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit, 33rd, Seattle, WA (1997)

H.E.G. Powrie, C.E. Fisher, Monitoring of foreign objects ingested into the intake of gas turbine aero-engine, in *International Conference on Condition Monitoring Proceedings*, University of Swansea (1999), pp. 175–190

H.E.G. Powrie, C.E. Fisher, Engine health monitoring: towards total prognostics, in *IEEE Aerospace Conference Proceedings 1999* (Snowmass at Aspen, Colorado, March 1999)

H.E.G. Powrie, C.E. Fisher, Engine health monitoring: towards total prognostics, in *IEEE Aerospace Conference Proceedings 1999* (Snowmass at Aspen, Colorado, March 1999), 10.1109/AERO.1999.789759, pp. 11–20

H.E.G. Powrie, O.D. Tasbaz, R.J.K Wood, C.E. Fisher, Performance of an electrostatic oil monitoring systems during an FZG gear scuffing test, in *Proceedings of the International Conference on Condition Monitoring, Swansea* (Coxmore Publishing, Oxford, 1999), ISBN 901892115, pp. 155–174

H.E.G. Powrie, Use of electrostatic technology for aero engine oil system monitoring, in *IEEE Aerospace Conference Proceedings 2000* (Big Sky, Montana), pp. 57–72

M.D. Rogers, Metallographic characterisation of transformation phases on scuffed cast iron diesel engine components. Tribology **2**, 123–127 (1969)

J.W. Stead, *J. West Scotland Iron Steel Inst.*, **19**, 169 (1911–1912)

O.Z. Stern, The theory of the electrolytic double-layer. Elektrochem **30**, 508–16 (1924)

O.D. Tasbaz, R.J.K. Wood, M. Browne, H.E.G. Powrie, G. Denuault, Electrostatic monitoring of oil lubricated sliding point contacts for early detection of scuffing. Wear **230**, 86–97 (1999a)

O.D. Tasbaz, R.J.K. Wood, H.E.G. Powrie, Electrostatic monitoring of oil lubricated contacts for early detection of wear, in *Proceedings of the International Conference on Condition Monitoring, Swansea* (Coxmore Publishing, Oxford, 1999b), ISBN 901892115

D.M. Tayler, P.E. Secker, *Industrial Electrostatics Fundamentals and Measurements* (Research Studies Press, England, 1994)

A.A. Torrance, The metallography of worn surfaces and some theories of wear. Wear **50**, 169–182 (1978)

E.M. Trent, J. Iron Steel Inst. **143**, 401 (1941)

H.L.F. von Helmholtz, Studies of electric boundary layers. Ann. Phys. Chem. **7**, 337–382 (1879)

L. Wang, R.J.K. Wood, The influence of contact conditions on surface reaction layers formed between steel surfaces lubricated by an aviation oil. Tribol. Int. **40**(10–12), 1655–1666 (2007)

L. Wang, R.J.K. Wood, T.J. Harvey, S. Morris, H.E.G. Powrie, I. Care, Wear performance of oil lubricated silicon nitride sliding against various bearing steels. Wear **255**(1–6), 657–668 (2003a)

L. Wang, R.J.K. Wood, T.J. Harvey, S. Morris, H.E.G. Powrie, I. Care, Electrostatic monitoring of oil lubricated ceramic to steel contacts. Tribol. Res. Des. Eng. Syst. **41**, 625–35 (2003b)

L. Wang, R.J.K. Wood, I. Care, H.E.G. Powrie, Electrostatic wear sensing of ceramic-steel lubricated contacts. Transient Process. Tribol. **43**, 711–720 (2004). Tribology Series 43

J.A. Williams, *Engineering Tribology* (Oxford University Press, Oxford, 1994). ISBN 0198565038, 341

R.J.K. Wood, P. Ramkumar, L. Wang, T.J. Harvey, K. Nelson, E.S. Yamaguchi, J.J. Harrison, H.E.G. Powrie, N. Otin, Electrostatic monitoring of the effects of carbon black on lubricated steel/steel sliding contacts. Life cycle Tribol. **48**, 109–121 (2005). Tribology Series 48

A.L. Zharin, D.A. Rigney, Application of the contact potential difference technique for on-line rubbing surface monitoring. Tribol. Lett. **4**, 205–213 (1998)

Bearings Using Water as Lubricant

► Water-Lubricated Rubber Alloy Bearings and Transmission Systems

Beilby Layer

► Self-Mating Metal Articulations in the Hip Joint

Bench Testing of Engine Components

► Simulation of the Tribological Behavior of Engine Components

Bench Tests

► Additive Chemistry Testing Methods

Bending Fatigue

► Gear Bending Fatigue Failure and Bending Life Analysis

Bending Fatigue Resistance

► Gear Bending Fatigue Failure and Bending Life Analysis

Bending Strength

► Gear Bending Fatigue Failure and Bending Life Analysis

Benefits of Eco-friendly Lubricants

► Environmentally Friendly Lubrication Issues

BHA – Bovine Bone Hydroxyapatite

► Modified UHMWPE for the Hip Joint (Particle Filled and Reinforced)

Bingham Fluids

► Newton's Law of Viscosity, Newtonian and Non-Newtonian Fluids

Bio-inspired Surface Texturing

► Tribological Characteristics of Insects' Skin

Biolubrication

► Brush and Hydration Lubrication (Natural Synovial Joints)

Biomimetic Synovial Fluid

► Artificial Synovial Fluid

Bionic Synovial Fluid

► Artificial Synovial Fluid

Bio-reinforced Composites

► Sustainable Technology for Tribological Textiles

Biotribological Joint Simulation System

Fengcai Wang[1,2], Hanyu Wang[3]
[1]School of Mechanical Engineering, Wuhan University of Science and Technology, Wuhan, Hubei, People's Republic of China
[2]National Research Centre of Bearing Technology (ZWZ), Xi'an Jiaotong University, People's Republic of China
[3]Bradford Grammar School, Bradford, West Yorkshire, UK

Synonyms

EDSC – equivalent discrete spherical convolution for joint simulation; ESC-equivalent spherical convolution for joint simulation; FTM – fixed-tracked method for joint simulation; SFFT – spherical fast fourier transform for joint simulation; SGDM – spherical grid data model for joint simulation; SIF – spherical inverse filter method for joint simulation

Definition

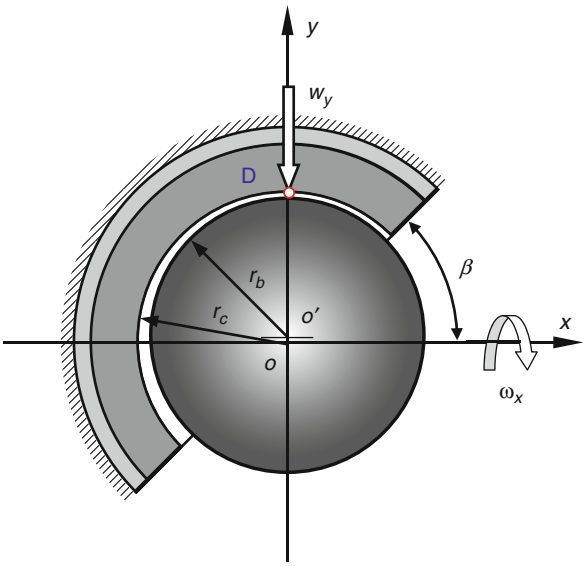
A biotribology simulation system integrates lubrication modeling, deformation evaluation, contact mechanics, friction prediction, and wear modeling in the system for the study of tribological problems in bioengineering and biomechanics. The simulation system is developed from a group of novel methods and advanced numerical techniques of equivalent discrete spherical convolution (EDSC), spherical fast Fourier transform (SFFT), the smooth inverse filter method (SIFM), the fixed-tracked method (FTM), and the spherical grid data model (SGDM). Additionally, the appropriate numerical method and iterative process are designed to obtain the convergent solution of governing equations from the corresponding modeling in the simulation system. Furthermore, the operational working conditions of material property, lubricant property, loading, motion, and so on should be included in the simulation system. The simulation system can be applied not only for the study of tribological problems in bioengineering and biomechanics, particularly for natural or artificial hip joints, but also employed for the analysis of spherical bearing or joints in industrial engineering. The simulation system can be further applied for other complex bearings with the description of the spherical-base geometry, which can be introduced by the overall or the micro-geometric features such as aspheric bearing and machining marks.

Scientific Fundamentals

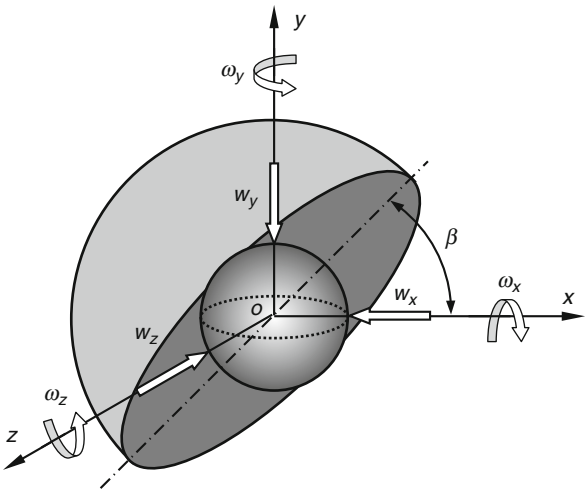
Significant progress in biotribology, such as lubrication of hip joint replacements, has been achieved in the last decade. A group of novel methods and advanced techniques have been developed, for instance, equivalent spherical convolution (ESC) with the spherical fast Fourier transform (SFFT) technique for deformation evaluation on spherical surfaces and the smooth inverse filter method (SIFM) for contact mechanics as well as the fixed-tracked method (FTM) for representation of complex bearing geometry have been proposed. Consequently, such methods and techniques significantly promote the development of lubrication modeling, deformation evaluation, contact mechanics, friction prediction, and wear modeling. Such modeling has been widely applied for the study of tribological problem in bioengineering and biomechanics or industrial engineering, for example, natural or artificial hip joints and spherical joints (Wang et al. 2009b; Flores et al. 2006), and plays an important role in the recognition of their tribological mechanism. Therefore, it is necessary that those related to the modeling and methods as well as relevant techniques are developed as an integrated simulation system in order to accurately show such achievement and application as well as further development of the corresponding tribological modeling and the whole simulation system. It should be pointed out that the methodology for the modeling in the simulation system can be also applied for the development of tribological modeling with complex spherical-based geometry and motion (Wang 2010, 2011; Wang et al. 2010).

Model and Methodology

A biotribology simulation system is mainly developed for the study of lubrication, friction, contact, wear, and the other related tribological subjects in bioengineering. Significant progress in simulation systems has been made since a group of novel methods and advanced numerical techniques has been developed in the last decade. The tribology simulation system in bioengineering has been widely employed for the study of the tribological and related problems of either natural or artificial hip joints either in clinical applications or manufacturing and design. It should be pointed out that the models and methods described in this section are further developed by the spherical grid data model (SGDM) into another form that could be applied for the study of the other



Biotribological Joint Simulation System, Fig. 1 A spherical head against a cup bearing in hip joint replacements with an inclination angle of β



Biotribological Joint Simulation System, Fig. 2 Spherical bearing model with the three-dimensional loading and the multi-directional Euler angular motion

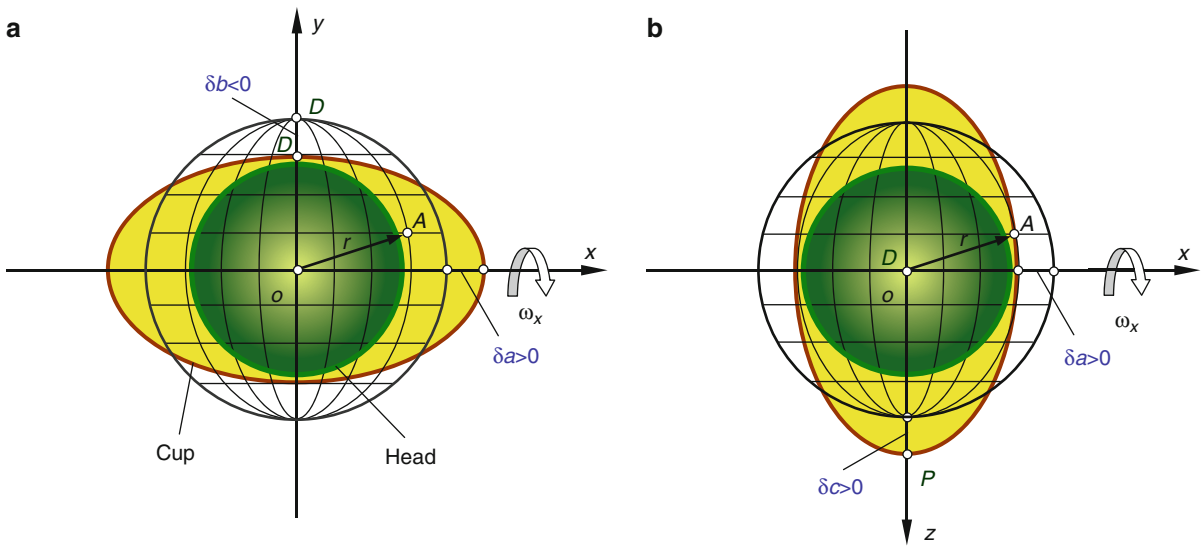
tribological problems such as knee joints in biomechanics or bioengineering and either spherical or aspheric bearings in industry engineering. The general modeling for the simulation system is explained below.

Biotribology Model

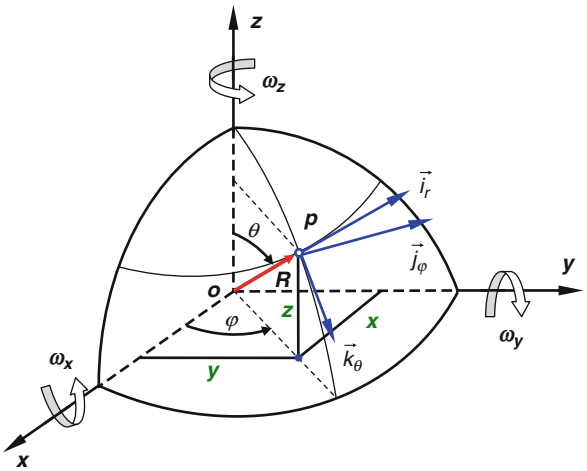
A general bearing model for biotribology simulation systems is introduced as developed previously. A ball-in-socket configuration is selected to describe how the tribological simulation system can be applied for the study of either natural or artificial hip joints in bioengineering, as shown in Fig. 1. An inclination angle is given for the configuration to physiologically describe the positioning of the outer cup bearing in hip joint replacements. The three-dimensional dynamic loading of (w_x, w_y, w_z) and Euler angular motion (ω_x, ω_y, ω_z) is applied to bearings in a spherical coordinate system, as shown in Fig. 2. The bearing geometry may be either spherical or aspheric, which can be described in the different geometric measurement scales, such as roughness on bearing surface or nonsphericity of bearing surface of (δa, δb, δc), as shown in Fig. 3a, b. The complex bearing geometry is described by the deviations of the real spherical or aspheric bearing surface relative to the nominal or perfect spherical bearing surface. Hence, the complex tribological modeling can be carried out on the basis of the spherical bearing, for example, the elastic displacement matrix in deformation evaluation is approximately extracted by the finite element method (FEM) from the corresponding spherical bearing of hip joint replacements. The related modeling for the simulation system will be further explained below.

Coordinate System and Mesh Grid

An appropriate spherical coordinate system with the z-axis through the polar points, used for the biotribology simulation system of hip joints, is selected as shown as Fig. 4. Thus, the inclination angle of a spherical cup bearing, as required by either natural or artificial hip joints, is readily given in modeling by the spherical coordinates of $\varphi = \beta$ for the meshed cup bearing surface, as shown in Fig. 5a, b. In fact, most tribological problems of spherical bearings in hip joints are resolved by the simulation system in the given spherical coordinate system, which consists of contact mechanics, lubrication, friction, deformation, and wear modeling. The fixed coordinate system used for simulating the tribological problem is defined as a space coordinate system, while the regularly meshed grids for description of bearing geometry of hip joints are often designed in the space coordinate system. Similarly, a body coordinate system could be also required for representation of bearing geometry either changing or in motion, for example, wear due to sliding contact with time causes the variations of two bearing surfaces of metal-on-metal hip joint replacements during an operational condition. On the other hand, even without wear, the similar representation of bearing geometry for tribological modeling under the complex transient motion condition of the inner bearing surface relative to the



Biotribological Joint Simulation System, Fig. 3 Aspheric and base-spherical bearing geometry: (**a**) with the variations of semi-axes δa and δb along x and y, and (**b**) with the variations of semi-axes δa and δc along x and z



Biotribological Joint Simulation System, Fig. 4 Spherical coordinate system with the z-axis through the polar points of N and S

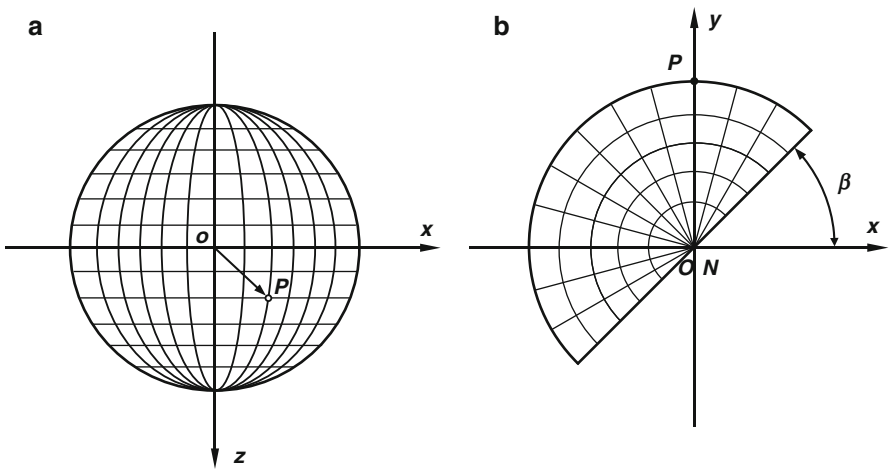
outer bearing surface should be addressed by tracking the objective bearing geometry during the motion condition through a body coordinate system. The details for a space and body coordinate system and the corresponding mesh grids are further given in the geometry and representation section, where a fixed-tracked method (FTM) developed for the geometry representation and complex tribological problem will be introduced.

Operating Conditions

An operating condition for the study of tribological problems of hip joints needs to be input when the biotribology simulation system is run. For example, natural or artificial hip joints could work either under walking or in experimental operating conditions of the given dynamic loading and transient motion, while the lubricant and material properties should be considered during simulating the tribological problem. Therefore, it is necessary to briefly describe how the operating conditions are applied for the modeling of the corresponding tribological simulation.

Material Property

Material property used for biotribology modeling is usually given by the elastic modulus E and Poisson ratioν of the materials of bearing used for hip joints, such as material properties of the acetabular cup and femoral head of hip joint replacements with metal-on-metal or metal-on-ceramic material combinations. Material property plays a significant role in tribological performance of hip joint replacements during running of the biotribology simulation system since the study of material effect on deformation, elastohydrodynamic lubrication, and contact mechanics is carried out by the corresponding modeling on the simulation system. The elastic displacement response matrix [S] of the material property for the study of the tribological problem is often extracted by the finite element method (FEM), as input data ready



Biotribological Joint Simulation System, Fig. 5 Mesh grids in a spherical coordinate system with the *z*-axis through the poles: (**a**) mesh grids and (**b**) inclination angle of β

for the tribology simulation system, and then applied for various tribological simulations such as lubrication or contact mechanics modeling of natural or artificial hip joints as well as other bearings (Wang et al. 2004, 2005). The details of material property used in the corresponding modeling can be found in reports published in recent years (Wang et al. 2008b, 2009a).

Lubricant Property

The rheological property of a lubricant is also to be considered in the study of tribological problems such as lubrication or lubricant film friction prediction of hip joints either during the given walking conditions or the designed experimental measurement (Wang et al. 2008b, 2010). The rheological property may be given by lubricant Newton and non-Newton fluid models that are dependent on viscosity and density and lubricant temperature and pressure. Additionally, lubricant property is dependent on the real operating conditions. For example, the performance of synovial fluid in healthy natural hip joints becomes the periprosthetic synovial fluid after carrying out a hip joint replacement. Thus, the lubricant rheologically behaves with the powerful non-Newtonian fluid characteristics operating under a relatively low shear rate, while Newtonian fluid at high shear rates is similar to the natural synovial fluid. On the other hand, a high shear rate is likely experienced in hip joints during the walking, and so the lubricant may be assumed to be a Newtonian, isoviscous, and incompressible fluid (Wang et al. 2005, 2008a). Therefore, the various models of lubricant property can be employed for the development of a biotribology simulation system in order to resolve the corresponding tribological problem that is concerned with lubrication, wear, and friction.

Loading and Motion

A tribological problem can be resolved by a tribology simulation system either under the given steady or the transient loading and motion condition, such as the transient elastohydrodynamic lubrication of hip joint replacements and the computational wear prediction under walking conditions. The motion and loading condition in tribological modeling can be represented by the three-dimensional dynamic load of $w=(w_x, w_y, w_z)$ and the three-directional time-dependent motion of $\omega=(\omega_x, \omega_y, \omega_z)$, as shown in Fig. 2. It is further explained that the transient motions applied on spherical bearings of hip joints are the three Euler angular velocities of the flexion-extension (FE) ω_x, internal-external rotation (IER) ω_y, and abduction-adduction (AA) ω_z. In fact, the tribological problem is evaluated under the dynamic loading and transient motion condition at each instant within an operational working cycle. During the numerical simulation, the resultant load due to the pressure distributed on a bearing surface should be balanced with the external load. For example, the resultant load can be calculated by an integration of the pressure distributions of either lubricant pressure in a film lubrication regime or asperity contact pressure in a mixed/boundary regime (Wang et al. 2005, 2008a, 2009b). The load evaluation is often finished by the loading balance with the external load during the automatic iteration procedure, where the iterative direction for the loading balance may be judged by the information of loading difference between the

evaluated resultant load and the given external load at the last and the current instants in one operational working cycle; for example, the Newton-Raphson method is applied for load balance evaluation in lubrication problems of hip joint replacements.

Geometry and Representation

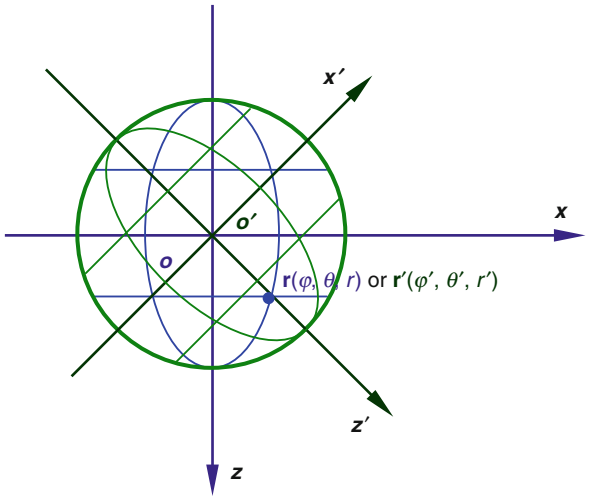
Bearing geometry for the study of a tribological problem needs to accurately be described for the simulation system, which is presented by a given density of mesh grids in an appropriate coordinate system, as shown in Figs. 4 and 5. It can be found that the inclination angle of cup bearing of hip joints is readily considered by the spherical coordinate of $\varphi=\beta$, as shown in Fig. 5. The density of mesh grids is not only dependent on the tribological problem to be resolved, but also on the bearing geometric features and the numerical efficiency. Additionally, the numerical sensitivity analysis should be made by running the simulation system in order to ensure that acceptable results are obtained.

The real bearing geometry may be a simple or complex from, such as spherical or aspheric bearing surfaces in application, design, and manufacturing of hip joint replacements. The bearing geometry can be described in the different topography scales in the geometric measurement, such as spherical or ellipsoidal bearing surface, machining marks on the overall bearing surface, roughness on bearing surface, and so on. Thus, the bearing geometry in an appropriate spherical coordinate system, as shown in Figs. 4 and 5, is determined by

$$r_i = r_{i0}(\varphi,\theta) + \Delta r_i(\varphi,\theta) + \Theta_i(\varphi,\theta) + \delta_i(\varphi,\theta) \quad (1)$$

where r_0 means the radius of nominal spherical bearing surface, $\Delta r_i$is the deviations of the real bearing relative to the nominal spherical surface, Θ_i is the micro-geometric feature such as roughness or machining marks, and δ_i is the deformation. The elastic deformation can be evaluated by the equivalent discrete spherical convolution model (EDSC) and the spherical fast Fourier transform technique (SFFT) in the simulation system (Wang et al. 2005, 2008a, 2009b). The spherical grid data model (SGDM) has been widely applied for the study of tribological problems with the based-spherical geometry (Wang et al. 2009a, 2010).

On the other hand, the variation of bearing geometry due to wear with time during the operational working condition causes the complexity in presenting bearing geometry during numerical simulation, even if spherical bearings are under such complex motion conditions as the three-dimensional Euler angular motion. In fact, such bearing surfaces at each instant within one working cycle



Biotribological Joint Simulation System, Fig. 6 Mesh grids for numerical simulation in space system and for tracking bearing geometry with motion in body system

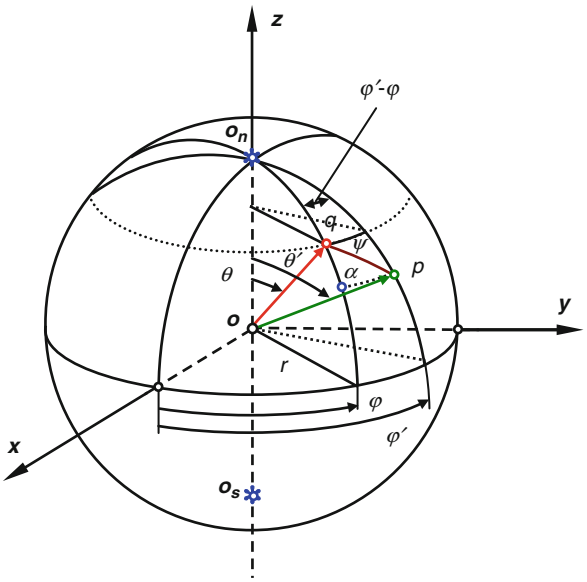
can be described by the geometric representation technique that is employed for the study of the complex tribological problem. The bearing geometry is represented in a body coordinate system by a set of regular mesh grids with the coordinates of $\mathbf{r}'(\varphi, \theta, r)$, which tracks the bearing geometry either in motion or geometric variation due to wear with time, while the corresponding simulation of a tribological problem is carried out in the space coordinate system by another set of regular mesh grids with the coordinates of $\mathbf{r}(\varphi, \theta, r)$, as shown in Fig. 6. Such methodology for representation of bearing geometry is called the fixed-tracked method (FTM). The one-to-one coordinate points in space and body coordinate systems have the following relation:

$$[\mathbf{r}] = [\mathbf{A}]^{-1}[\mathbf{r}'] \quad (2)$$

where [A] means an orthogonal transformation with the directional cosines of an internal body set of axes relative to an external space set of axes, which is expressed as

$$[\mathbf{A}] = [\mathbf{A}]^{-1} = \begin{pmatrix} \cos\beta\cos\gamma & -\sin\beta & \cos\beta\sin\gamma \\ \cos\alpha\sin\beta\cos\gamma & \cos\alpha\cos\beta & \sin\alpha\cos\beta \\ \sin\alpha\sin\beta\cos\gamma & \sin\alpha\cos\beta & \sin\alpha\sin\beta\sin\gamma+\cos\alpha\cos\gamma \end{pmatrix} \quad (3)$$

where (α, β, γ) are the three Euler angles corresponding to the three Euler angular velocities $(\omega_x, \omega_y, \omega_z)$, respectively, as shown in Figs. 2 and 3. The fixed-tracked method for geometric representation can be employed for carrying out the tribological simulation work with complex



Biotribological Joint Simulation System, Fig. 7 Spherical coordinate and distance on a spherical surface in a spherical coordinate system

geometry and motion such as non-sphericity and lubrication and wear modeling in the tribology simulation system (Wang et al. 2009a, 2010).

Deformation

Elastic deformation plays a significant role in tribological problems. Elastic deformation of a bearing surface often needs to be considered by running the tribology simulation system for the study of a tribological problem, for example, the simulation of elastohydrodynamic lubrication and contact mechanics of hip joints includes bearing deformation evaluation in the corresponding film thickness and gap. Most of biological interfaces of the tribological problem in bioengineering can be seen as spherical bearing configuration with the deviations of the real bearing geometry relative to the overall smooth or the perfectly spherical bearing surfaces, such as spherical or aspheric bearings, as shown in Fig. 7. Therefore, the bearing deformation can be evaluated by the equivalent spherical convolution (ESC) if an average latitude is selected on the basis of spherical bearing surfaces as follows (Wang et al. 2005):

$$\delta(\varphi,\theta) = \iint_{\Omega} p(\varphi_\zeta,\theta_\eta)\, S'(\varphi-\varphi_\zeta,\theta-\theta_\eta,\theta) \times R^2 \sin\theta_\eta d\varphi_\zeta d\theta_\eta \tag{4}$$

The integration of the elastic deformation is often calculated by the equivalent discrete spherical convolution (EDSC) in the tribology simulation system as below:

$$\delta_{ij}(\varphi_i,\theta_j) = \sum_{k=1}^{m}\sum_{l=1}^{n}[p_{kl}(\varphi_k,\theta_l)\sin(\theta_l)]\, S_{kl}^{ij} \times (\varphi_i-\varphi_k,\theta_j-\theta_l;\theta_m) \tag{5a}$$

or

$$\left[\tilde{\delta}\left(f_\varphi,f_\theta\right)\right] = \left[\tilde{p}\left(f_\varphi,f_\theta\right)\sin\left(\bar{\theta}\right)\right]\cdot\left[\tilde{S}\left(f_\varphi,f_\theta\right)\right] \tag{5b}$$

where $[\delta]$, $[P]$, and $[S']$ represent the deformation, pressure distribution, and the elastic displacement coefficient matrices, respectively, while $\bar{\theta}$ is seen as the mean value of latitudes on a mesh gird in a spherical coordinate. The elastic deformation evaluation is performed by the fast spherical Fourier transform technique (SFFT) in a spherical coordinate system (Wang et al. 2008a, 2009b). The elastic theory has been widely applied for the study of tribological problems of spherical bearing seither in hip joints or in industrial engineering. The periodic error of the circular convolution model is eliminated by the zero padding technique when the spherical fast Fourier technique is employed in the tribology simulation system.

Before carrying on the tribology simulation system for the evaluation of elastic deformation of a bearing surface, an elastic displacement response on the corresponding spherical bearing surface needs to be extracted by the finite element method (FEM), when a unit pressure is applied at the original point on the spherical bearing surface in spherical coordinates, as shown in Fig. 5a. In fact, one-dimensional deformation response along one longitude on a spherical bearing surface under the given unit pressure is enough to be used for extension of the elastic deformation response on the whole bearing surface by the EDSC model (Wang et al. 2005, 2008a). Thus, the two-dimensional elastic deformation matrix of the bearing surface on the basis of the nominal spherical bearing surface is obtained accordingly and then employed as an input data for the tribology simulation system to carry on further simulation works (Wang et al. 2009, 2010).

Contact Mechanics

The model of contact mechanics in a biotribology simulation system can be employed for the study of contact problems of bearing in hip joints and other bearing interfaces with complex base-spherical geometry. Such a bearing interface brought into contact is often classified as a conformal spherical contact problem, for example, a hip joint replacement comprises a femoral head articulating an acetabular cup, which can be macroscopically

seen as a conformal spherical contact problem. Additionally, those of bearing interfaces either in bioengineering or industrial engineering are also considered as the overall spherical contact problem where the spherical grid data model (SGDM) may be employed to describe the bearing geometry for further contact analysis in the tribology simulation system. Generally, a contact problem needs to satisfy the restriction relationship of the gap between the two surfaces brought into contact and the resultant contact pressure p in contact area Ω_c in spherical coordinates as

$$\begin{cases} g(\varphi,\theta)=0 & (\varphi,\theta)\in\Omega_c \\ p(\varphi,\theta)>0 & (\varphi,\theta)\in\Omega_c \\ g(\varphi,\theta)>0 & (\varphi,\theta)\notin\Omega_c \\ p(\varphi,\theta)=0 & (\varphi,\theta)\notin\Omega_c \end{cases} \tag{6}$$

where the gap of $g(\varphi, \theta)$ between the bearing interface comprises the original gap $g_0(\varphi, \theta)$ and the deformation of $\delta(\varphi, \theta)$ as

$$g(\varphi,\theta)=g_0(\varphi,\theta)+\delta(\varphi,\theta) \tag{7}$$

When the simulation of contact mechanics problem is carried out through the iterative procedure in the tribology simulation system, the evaluation of deformation needs to be made and to satisfy the corresponding contact restriction condition as described above. The deformation can be calculated by the equivalent discrete spherical convolution (EDSC) model and the spherical fast Fourier transform (SFFT) in the simulation system, while contact pressure can be obtained by the smooth inverse filter method (SIFM). The smooth spherical inverse filter method (SIFM) with the smooth function of γ for convergent iterative procedure in frequency domain with the fast Fourier transform is expressed as

$$\tilde{p}_{ij}(f_\varphi,f_\theta)=\frac{\tilde{\delta}_{ij}(f_\varphi,f_\theta)}{\tilde{S}_{ij}(f_\varphi,f_\theta)\sin(\bar{\theta})}\,\gamma_{ij}(f_\varphi,f_\theta) \tag{8}$$

where (f_φ, f_θ) are the corresponding variables in the frequent domain to the variables (φ, θ) in the time domain. The real contact pressure value of $p(\varphi, \theta)$ in the time domain can be obtained when an inverse fast Fourier transform is applied to $\tilde{p}(f_\varphi, f_\theta)$ in the frequent domain (Wang et al. 2005).

Generally, the smooth inverse filter method (IFM) in the tribology simulation system can resolve both macroscopic and microscopic contact mechanics problems. For micro contact problems, a higher density of mesh-grids could be used for evaluation of contact mechanics when the micro-geometric features such as curvature, clearance, structure, and other sophisticated features are considered. Sometimes, the finite element method (FEM) is also applied for the solution of contact mechanics of a bearing, but the limitation of the element number may cause the finite element model to simulate the micro contact problem with difficultly when the overall bearing interface is considered (Wang et al. 2009b). Subsequently, when bearing geometry is varied with time due to wear, the fixed-tracked method (FTM) for geometry representation for numerical simulation in space system will need to be employed. The dynamic contact mechanics can be evaluated by the geometry representation method combined with the corresponding contact mechanic models such as the SIFM method and the FEM model. For instance, the dynamic contact problem in the computational wear prediction of hip joint replacements with the hard-on-hard material combination such as metal-on-metal under the three-dimensional motion and dynamic loading is able to be carried out in the tribology simulation system (Wang et al. 2009b, 2010).

Lubrication and Mixed Lubrication

Lubrication is a significant issue in tribology simulation systems for the study of hip joints under the given operational working conditions, such as elastohydrodynamic lubrication and mixed lubrication problems of natural or artificial hip joints. The lubrication models for governing fluid lubrication, film thickness, asperity contact, mixed or boundary lubrication, and so forth are important parts of the simulation system, including a group of methodologies and techniques for efficient simulations. The general lubrication modeling is applied for the study of fluid film lubrication in the simulation system of hip joints. When the three-dimensional Euler angular motion of $(\omega_x, \omega_y, \omega_z)$ is considered, the Reynolds equation for governing lubrication can be rewritten as

$$\begin{aligned} \sin\theta\frac{\partial}{\partial\theta}\left(h^3\sin\theta\frac{\partial p}{\partial\theta}\right)+\frac{\partial}{\partial\varphi}\left(h^3\frac{\partial p}{\partial\varphi}\right)&=6\mu\, r_{c0}^2\sin\theta \\ \times\left[\begin{array}{l}\left(-\omega_x\sin\varphi+\omega_y\cos\varphi\right)\dfrac{\partial h}{\partial\theta}\\ +\left(-\omega_x\cos\varphi\cos\theta-\omega_y\sin\varphi\cos\theta+\omega_z\sin\theta\right)\dfrac{\partial h}{\partial\varphi}\end{array}\right] \\ +12\mu r_{c0}^2\sin^2\theta\frac{\partial h}{\partial t}& \end{aligned} \tag{9}$$

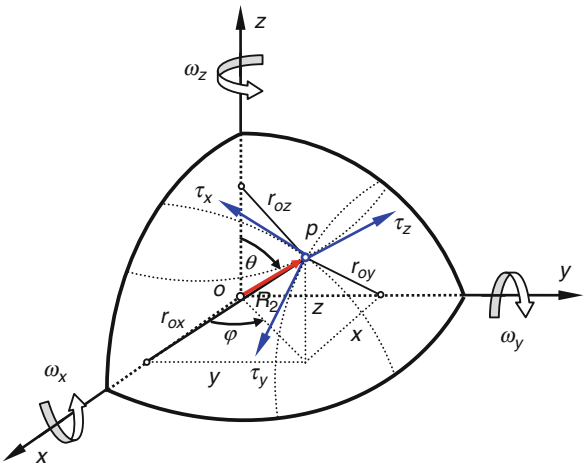
where r_{c0} represents the radius of bearing. It is should be pointed out that the motion, ω_x, should be selected to coincide with the rotational direction along the x-axis. The motion appears with the negative sign in the equation, which corresponds to both the rotational directions of the three angular velocities indicated by the right-hand rule in the appropriate coordinate system, as shown in

Figs. 1 and 4. When the minimum film thickness is less than zero or boundary film thickness, the direct contact or the asperity contact will take place in a mixed lubrication regime. The contact pressure distribution in a mixed lubrication regime needs to be resolved by the corresponding contact mechanics modeling, which is described above. Additionally, a simplified Reynolds equation for the solution of asperity contact pressure, p_c, in mixed lubrication regime can be employed as follows:

$$\varepsilon\left[\begin{array}{l}\left(-\omega_x\sin\varphi+\omega_y\cos\varphi\right)\sin^2\theta\dfrac{\partial h}{\partial\theta}\\ +\left(-\omega_x\cos\varphi\cos\theta-\omega_y\sin\varphi\cos\theta+\omega_z\sin\theta\right)\sin\theta\dfrac{\partial h}{\partial\varphi}\end{array}\right]$$
$$+2\varepsilon\sin^2\theta\frac{\partial h}{\partial t}=0 \text{ at } h\le\varepsilon_b \tag{10}$$

where the constant $\varepsilon=6\mu r_{c0}^2$. Much attention should be paid to the numerical stability of the solution of the asperity pressure caused by the transition between lubricant and asperity contact areas during simulation iteration in a mixed lubrication regime (Wang et al. 2006).

The solution of lubrication problems is performed by running lubrication modeling in the tribology simulation system through a number of iterative cycles only during the solution of the lubrication equation under the given operational condition. The solution of the lubrication equation is only one part of resolving the whole lubrication problem, for example, the solution of the loading balance and deformation as well as possible asperity contact in a mixed lubrication regime also needs to be approached through the whole iterative process. Therefore, once the lubrication pressure is obtained by the discrete Reynolds equation with the difference scheme through the iterative procedure of lubrication in (9), the fluid film thickness can be evaluated by the film thickness equation, where the deformation of bearing surfaces is calculated by the elastic equation as described previously according to the pressure distribution. It is noted that an initial pressure distribution should be given at the beginning of the simulation iterations. Each solution of all governing equations for a lubrication problem is to satisfy the corresponding error control requirements. When a transient lubrication problem is resolved, such as the transient elastohydrodynamic lubrication of hip joint replacements under the walking condition, the position of the center of the inner bearing with the eccentricity of (e_x, e_y, e_z) relative to the outer bearing at each instant within an operational cycle can be automatically determined by the corresponding calculation scheme and applied for the next simulation of lubrication problem in the tribology simulation system. In other words, the solution of the lubrication problem needs to be simultaneously carried out in different levels of simulation iterations and to satisfy the corresponding numerical error requirements.



Biotribological Joint Simulation System, Fig. 8 Shear stress and frictional torque arm in a spherical coordinate system under the three-dimensional motion condition

Friction Prediction

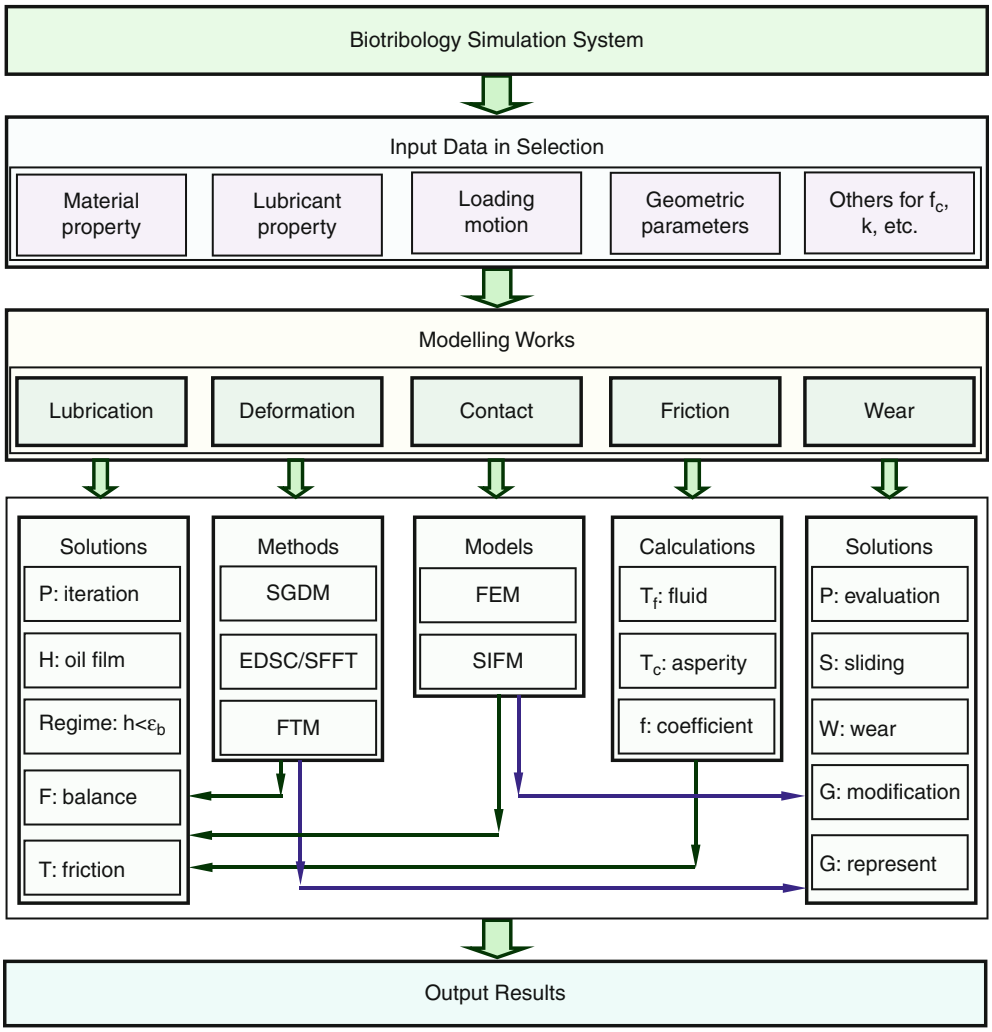
Friction is often caused by shear stress due to fluid shear with viscous effect and asperity contact pressure with a relative motion between the bearing surfaces either in a mixed lubrication regime or dry contact conditions. The shear stress and the friction torque arm about the corresponding axis in spherical coordinate system is given as shown in Fig. 8. The friction within the fluid film thickness is rewritten as

$$[\tau]_f=\left\{\begin{array}{l}\tau_{\theta\varphi}\\ \tau_{\theta r}\\ \tau_{\varphi r}\end{array}\right\}=\left[\begin{array}{c}0\\ -\mu\frac{U_\theta}{h}+\left(\zeta-\frac{h}{2}\right)\frac{1}{R}\frac{\partial p}{\partial\theta}\\ -\mu\frac{U_\varphi}{h}+\left(\zeta-\frac{h}{2}\right)\frac{1}{Rsin\theta}\frac{\partial p}{\partial\varphi}\end{array}\right] \tag{11}$$

where the variable ζ means the local coordinate along the fluid film thickness direction, the variable μ is the viscosity of lubricant g, and the variables of (U_φ, U_θ) represent the velocities of bearing surface. If only friction on bearing surfaces is considered, where $\zeta=0$, the shear stress at a point on the bearing in motion can be obtained by

$$[\tau]_f=\left\{\begin{array}{l}\tau_{\theta\varphi}\\ \tau_{\theta r}\\ \tau_{\varphi r}\end{array}\right\}=\left[\begin{array}{c}0\\ -\mu\frac{U_\theta}{h}-\frac{h}{2}\frac{1}{R}\frac{\partial p}{\partial\theta}\\ -\mu\frac{U_\varphi}{h}-\frac{h}{2}\frac{1}{Rsin\theta}\frac{\partial p}{\partial\varphi}\end{array}\right] \tag{12}$$

Additionally, frictional torque may need to be evaluated by the model of frictional torque in the tribology



Biotribological Joint Simulation System, Fig. 9 Flowchart of tribology simulation system with a group of models, methods, and techniques

simulation system. The total frictional torque consists of two parts, one due to fluid film shear stress and the other due to asperity contact pressure in a mixed lubrication regime. The components of the frictional torque of (T_x, T_y, T_z) about the three axes in a spherical coordinate system, as shown in Fig. 8, can be evaluated by

$$\begin{Bmatrix} T_x \\ T_y \\ T_z \end{Bmatrix} = \iint_{S_f} \begin{bmatrix} (\tau_{\theta r}\sin\varphi + \tau_{\varphi r}\cos\varphi\cos\theta) \\ (\tau_{\theta r}\cos\varphi + \tau_{\varphi r}\sin\varphi\cos\theta) \\ (\tau_{\varphi r}\sin\varphi) \end{bmatrix} R^3 d\theta d\varphi + \iint_{S_b} pf_c \begin{bmatrix} r_{ox}\kappa_x \\ r_{oy}\kappa_y \\ r_{oz}\kappa_z \end{bmatrix} R^2 d\theta d\varphi \tag{13}$$

where the control constants of $(\kappa_x, \kappa_y, \kappa_z)$ with the value of either 1 or 0 represent whether a relative motion exists about the corresponding axis, and the variables of (r_{ox}, r_{oy}, r_{oz}) are the arm lengths of shear stress components to the corresponding axes. It should be pointed out that the friction coefficient for asperity contact, f_c, needs to be given before making the friction prediction in the simulation system, which could be measured from experimental test (Wang et al. 2008b, 2010).

When the frictional torque is obtained, the friction coefficient of a bearing under the operational condition is conveniently calculated by

$$f = \frac{T}{Rw} \tag{14}$$

where the variables T, w, and R represent the frictional torque, load, and radius of the bearing, respectively. The friction prediction of a bearing can be made by the friction and frictional torque modeling in the tribology simulation system after the solution of lubrication and asperity contact in mixed lubrication regime is achieved (Wang et al. 2008b, 2010).

Wear Modeling

Computational wear modeling, for example, wear prediction of hip joint replacements under the multi-directional transient motion and dynamic loading conditions, is another part of the tribology simulation system. When running the simulation system for the study of wear of bearings with hard-on-hard material combination, it is noted that the representation of bearing geometry should be made by the fixed-tracked method (FTM) since wear takes place on the bearing surface couple and causes the variation of bearing geometry at the same time. The linear wear depth along the radius direction of the bearing can be evaluated by the wear modeling in the simulation system according to the Archard's wear law as follows:

$$\Delta r_i(t) = \sum_{t=1}^{n_t} k p_t s_t \tag{15}$$

where the integer n_t represents the working cycles of bearing experienced during an operating condition, k is wear factor that can be obtained from experimental test, and p is contact pressure distribution on bearing surfaces (Archard 1953). During wear simulation, the central angle of sliding distance and the real sliding distance between the bearing interface in a relative motion, as shown in Fig. 7, needs to be calculated by

$$\cos\psi = \cos\theta\cos\theta' + \sin\theta\sin\theta'\cos(\varphi' - \varphi) \tag{16}$$

or

$$s = R\psi \tag{17}$$

The contact pressure distribution of bearings for wear prediction at each instant within one working cycle can be resolved by the contact mechanics model that is either the finite element model (FEM) or the smooth inverse filter method (SIFM). Since the FEM model has a limitation of element number, it is difficult to employed in carrying out wear prediction of bearings with the consideration of micro-geometric features. Currently, the inverse filter method is selected to evaluate the contact pressure distribution because a high density of mesh grids can be used for the wear prediction in the simulation system. Once the linear wear depth of a bearing under an operating condition is obtained, the modification of bearing geometry at each instant during the operating cycle is calculated by

$$r_i^k(t) = r_i^{k-1}(t) \pm \Delta r_i(t) \tag{18}$$

where the sign of either + or − means the geometry modification direction for the bearing couple. Additionally, the volumetric wear can be calculated by the wear modeling in the simulation system after the linear wear prediction is completed (Wang et al. 2009b, 2010). The wear modeling in the tribology simulation system can be well applied for the study of the computational wear prediction of hip joint replacements with a hard-on-hard material combination under three-dimensional Euler angular motion and dynamic loading.

Load Balance

As described previously, the resultant load due to pressure distribution on a bearing surface is required for the balance with the external load of (w_x, w_y, w_z) by the evaluation of the integration of the pressure distribution as follows:

$$\begin{cases} f_x = R\int_{\varphi_1}^{\varphi_2}\int_{\theta_1}^{\theta_2} p\sin\theta\cos\varphi\sin\theta\, d\theta d\varphi = w_x \\ f_x = R\int_{\varphi_1}^{\varphi_2}\int_{\theta_1}^{\theta_2} p\sin\theta\sin\varphi\sin\theta\, d\theta d\varphi = w_y \\ f_x = R\int_{\varphi_1}^{\varphi_2}\int_{\theta_1}^{\theta_2} p\cos\theta\sin\theta\, d\theta d\varphi = w_z \end{cases} \tag{19}$$

where the pressure distribution consists of fluid film pressure and asperity pressure in a mixed lubrication regime. If an iterative method is applied to find the eccentric position of a bearing for the load balance, the load variation of $(\Delta f_x, \Delta f_y, \Delta f_z)$ due to the eccentric perturbation of $(\Delta e_x, \Delta e_y, \Delta e_z)$ needs to be evaluated during the loading balance process within the whole simulation cycle (Wang et al. 2005, 2009b).

Tribology Simulation System

The tribology simulation system is developed from a group of models and methods as well as numerical techniques as described in the above section. The simulation system can be applied for the study of the corresponding tribological problems in bioengineering, such as lubrication of hip joints. Additionally, the simulation system can be also employed for studies in industrial engineering where similar bearing models to thr tribological problem of hip joints are widely found, such as the conformal contact problem of spherical joints in a flexible multi-body system. A further explanation of the tribology simulation system is given below.

System Flowchart

The biotribology simulation system consists of several modeling works of lubrication modeling, deformation evaluation, contact mechanics, friction prediction, and wear simulation, as shown in Fig. 9. A group of novel tribological methodologies and advanced numerical techniques as well as popular methods has been described above. When the input parameters of a bearing under the operational working condition of loading and motion are given, the simulation system can be run for the study of the corresponding tribology problem. The overall configuration of the simulation system is only introduced by the flowchart, and details for each modeling and simulation work can be seen in the reports published in recent years elsewhere (Wang et al. 2008a, 2009b, 2010).

Flowchart Explanations

Although most of words and nomenclatures as well as the corresponding contents, as shown in the flowchart, have been described in detail previously, a further explanation of the flowchart of the simulation system possibly needs to be made in order to describe meanings of each part, abbreviations, and letters either in small or capital form as follows:

- f_c and k as input data represent friction coefficient on asperity contact area and wear factor, respectively.
- P: iteration means that fluid film pressure is iteratively obtained by Reynolds equation in lubrication modeling.
- H: oil film represents that oil film thickness for lubrication or gap for contact between bearing surfaces are evaluated by the oil thickness or gap equations. The elastic deformation needs to be calculated by the pressure obtained from the solution of Reynolds equation. The deformation evaluation can be carried out by the equivalent discrete spherical convolution (EDSC) and the spherical fast Fourier transform (SFFT) technique. If the bearing has complex geometric features and working conditions, the fixed-tracked method (FTM) needs to used. These evaluations are often carried out by the spherical grid data model (SGDM) with the spherical-base geometry. Additionally, a one-dimensional displacement response matrix for deformation evaluation should be obtained by the finite element model prior to lubrication modeling.
- Regime: at $h < \varepsilon_b$ means that there is the regime transition from fluid lubrication to mixed lubrication at the area with film thickness or gap less than boundary film thickness ε_b. At this moment, the asperity contact evaluation is to be made by solution of contact mechanics at the area, and the pressure distribution on mesh gird points is replaced by the corresponding asperity contact pressure. The asperity contact pressure can be resolved by the smooth inverse filter method (SIFM) or the finite element method.
- F: balance is for the load balance calculation. As said previously, once pressure solution at an instant within one operating cycle during dynamic loading and transient motion condition is achieved by the governing equation of either lubrication or mixed lubrication, the resultant load balance evaluation with the external load can be made, and afterwards the simulation of lubrication at the next instant starts. It should be pointed out that the current pressure at this instant will be employed for the iterative solution of the pressure at the next instant by such iterative scheme as $p^{k+1} = p^k + \alpha(p^{k+1} - p^k)$ where α is the iterative factor.
- T: friction means that the friction and frictional torque of bearing can be calculated after lubrication solution is performed at the moment accordingly.
- T_f: fluid and T_c: asperity show that the friction and the corresponding frictional torque are caused either by the viscous shear stress of fluid film in full film lubrication regime or the asperity contact pressure on the bearing interface in a relative motion with a friction factor f_c.
- f represents that the friction factor of bearing can be obtained when the frictional torque evaluation is completed.
- P: evaluation means that the contact pressure of bearing surfaces under the given operational condition of dynamic loading and time-dependent motion is to be resolved by the model of contact mechanics such as the smooth inverse filter method (SIFM) or the finite element model (FEM) in wear modeling.
- S: sliding shows that the sliding distance between bearing surfaces in a sliding process at an instant of one operational cycle under the given loading and motion condition is evaluated.
- W: wear represents that the wear evaluation of bearings such as linear wear depth is able to be carried out under the obtained wear factor after the solution of both pressure distribution and sliding distance are completed.
- G: modification means that the bearing geometry is modified in body system by the obtained linear wear depth at the moment during the wear simulation procedure.
- G: representation shows that the modified bearing geometry on mesh grids due to wear in body system is represented in a space system to resolve the pressure

distribution at the next instant when the newly modified bearing geometry is applied at the last instant. A new model or code of contact mechanics with the modified bearing geometry for the solution of the pressure solution often needs to be produced at that moment.

Key Applications

The biotribology simulation system can be applied not only for the study of hip joints in bioengineering, but also for the study of spherical bearings in industrial engineering. One example of metal-on-metal hip joint replacements with a hard-on-hard material combination and under the walking condition of the dynamic loading and the transient motion is simulated, while the other example of metal-on-polyethylene hip joint replacements under the steady working condition with a load of 2,500 N and a velocity of 2 rad/s is analyzed, by application of the simulation system, respectively. The hip joint replacements comprises a spherical femoral head bearing articulating with a spherical acetebular cup bearing, which is modeled by a ball-in-socket configuration, as shown in Fig. 1. The mechanical properties and the geometric parameters of bearing components are given in Tables 1 and 2, separately. The lubricant viscosity is selected as a typical value of 0.001 Pa.s. Additionally, the acetabular cup is anatomically positioned with an inclination angle of $\beta=45^\circ$ as applied in the clinical implants (Wang et al. 2005, 2008b, 2010). A density of 256×256 mesh-grids in spherical coordinates is adopted. Such high density of mesh grids is reasonable for the analysis of the tribological problem of hip joint replacements, even for the soft elastohydrodynamic lubrication problem at the second case as described above. Convergence criterion for pressure iteration is specified as relative error of 10^{-5} (Wang et al. 2009b).

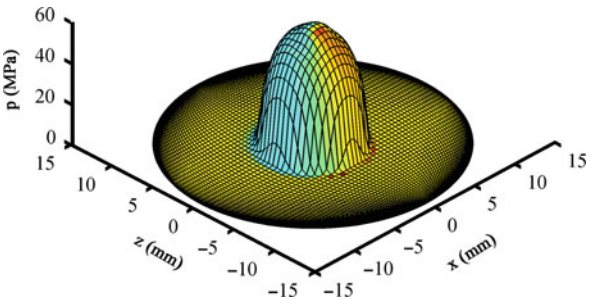
Before the simulation, a one-dimensional elastic displacement response along a longitude on the meshed geometry of spherical bearing of hip joint replacements is extracted by the finite element model (FEM) and then extended into the two-dimensional displacement response matrix on the bearing surfaces by the equivalent discrete spherical convolution (EDSC). The spherical fast Fourier transform (SFFT) technique is employed to facilitate the evaluation of the deformation in spherical coordinate system (Wang et al. 2004, 2005). In the present analysis, the iterative factor for the lubrication pressure iterations is given as 0.3–0.5 for reference, while the initial pressure distribution can be selected as zero. This allows that the solution of the transient elastohydrodynamic lubrication simulation becomes convergent from the second walking

Biotribological Joint Simulation System, Table 1 Mechanical properties of bearing components

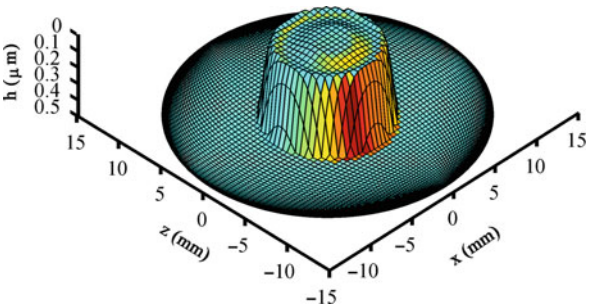
Material	Elastic modulus E	Poisson's ratio ν
CoCrMo	210 (GPa)	0.30
Cement	2.27 (GPa)	0.23

Biotribological Joint Simulation System, Table 2 Geometric parameters of bearing components

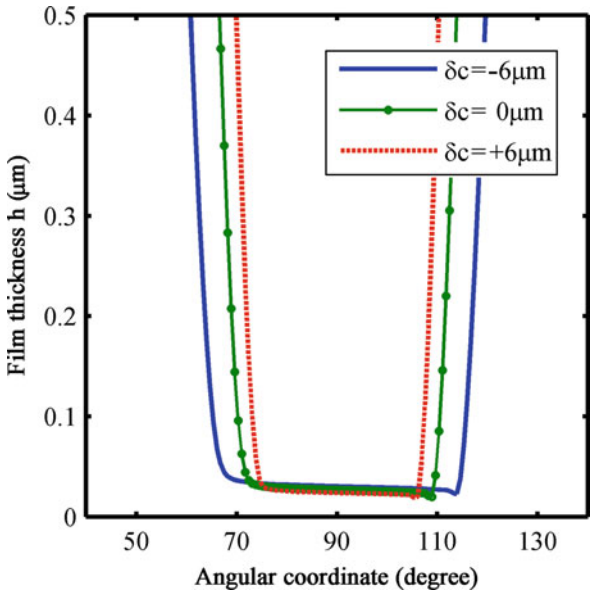
Component	Size
Femoral head radius (R_1)	14.00 (mm)
Acetabular cup radius (R_2)	14.03 (mm)
Cup wall thickness (d)	9.50 (mm)
Cup inclination angle (β)	45 (deg)



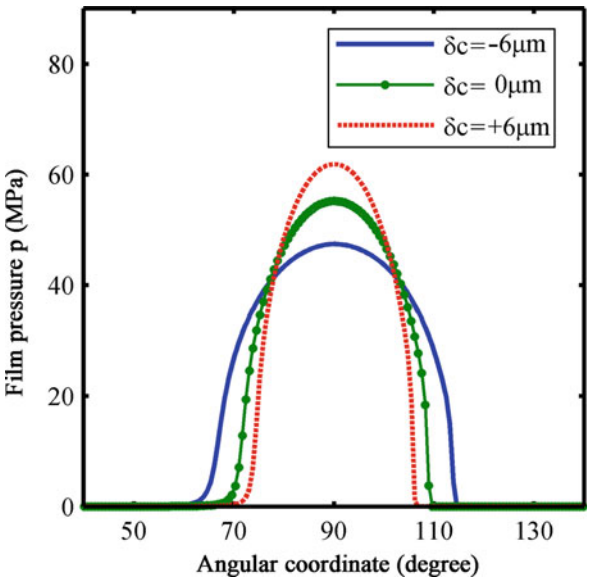
Biotribological Joint Simulation System, Fig. 10 Lubrication pressure distributions at the instant of 0.13 s during one walking cycle



Biotribological Joint Simulation System, Fig. 11 Fluid film distribution at the instant of 0.13 s during one walking cycle



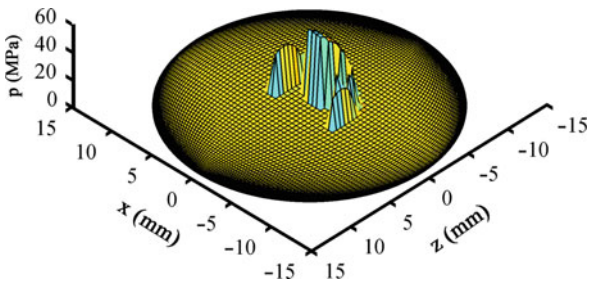
Biotribological Joint Simulation System, Fig. 12 Cross-sectional film thickness profiles through the center of a cup bearing with the non-sphericity of the head bearing along the *z*-axis in θ direction



Biotribological Joint Simulation System, Fig. 13 Cross-sectional film pressure profiles through the center of a cup bearing with the non-sphericity of the head bearing along the *z*-axis in θ direction

cycle (Wang et al. 2005, 2009b). Therefore, the fluid film thickness and pressure distribution of metal-on-metal hip joint replacements under the given operational condition as described above are predicted by the simulation system, respectively, as shown in Figs. 10 and 11. The minimum film thickness of spherical bearing of hip joint replacements under the given walking condition is bigger than 15 nm during the whole walking cycle, which indicates the lubrication condition of the bearing exists in full film lubrication regime (Wang et al. 2005).

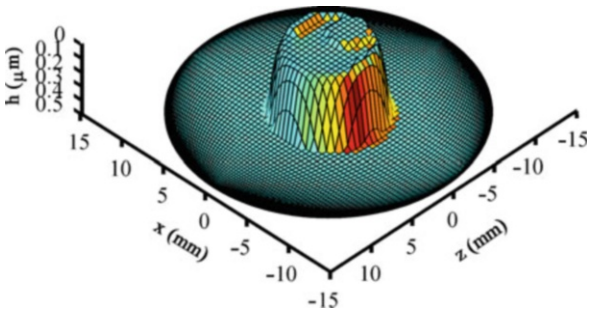
When the lubrication of an aspheric bearing of metal-on-metal hip joint replacements under the walking condition is considered, the bearing geometry under the transient motion causes a more complex simulation than the perfectly spherical bearing due to the representation of bearing geometry for lubrication modeling. The transient elastohydrodynamic lubrication of an ellipsoidal head bearing against a spherical cup bearing under the given working condition can be numerically predicted by the biotribology simulation, where the fixed-tracked method (FTM) is employed for the representation in a space system from bearing geometry in a body system (Wang et al. 2008b, 2009a). Effect of the different non-sphericity of the aspheric bearing on the transient elastohydrodynamic lubrication of metal-on-metal hip joint replacements in motion is evaluated by the simulation system, as shown in



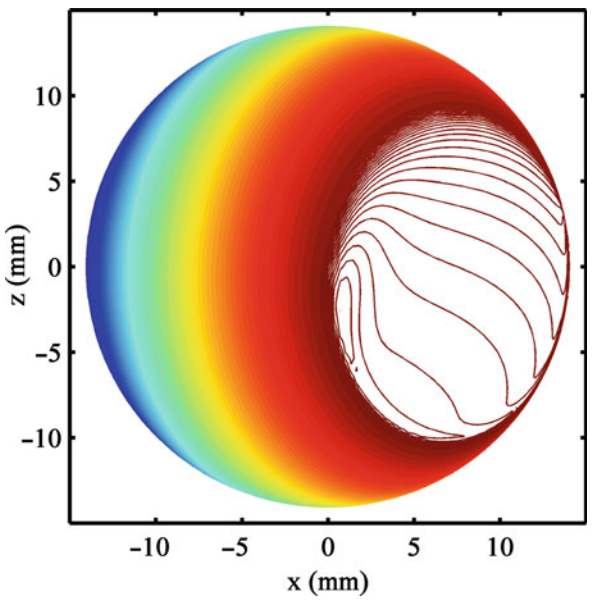
Biotribological Joint Simulation System, Fig. 14 Asperity contact pressure distribution of hip joint replacements in a mixed lubrication regime at the instant of 0.5 s during one working cycle

Figs. 12 and 13. It is found that the biotribology simulation system with the FTM method can be employed for the study of lubrication of bearing with the complex geometric feature and under a given operational condition (Wang et al. 2010).

When the lubrication film thickness or gap at a point between the bearing surface couple is less than a criterion value of either boundary film thickness or asperity contact (i.e., $h(\varphi, \theta) < \varepsilon_b$), the full film lubrication regime is to be transferred to a mixed lubrication regime, as described previously. Consequently, the evaluation of asperity

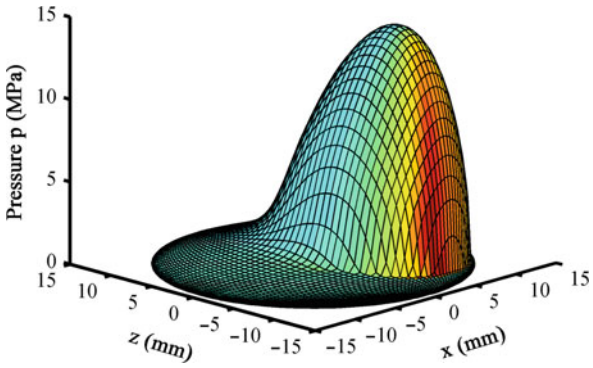


Biotribological Joint Simulation System, Fig. 15 Film thickness distribution of hip joint replacements in a mixed lubrication regime at the instant of 0.5 s during one working cycle

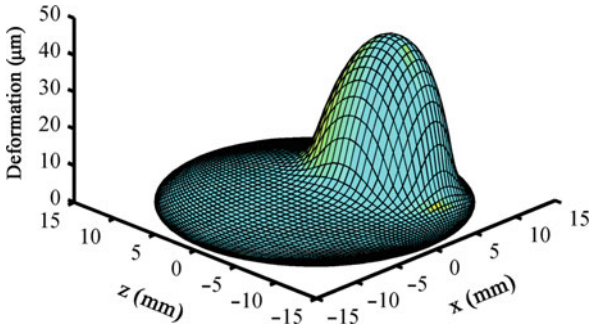


Biotribological Joint Simulation System, Fig. 16 Film thickness contour of metal-on-polyethylene hip joint replacements under the given steady operational condition

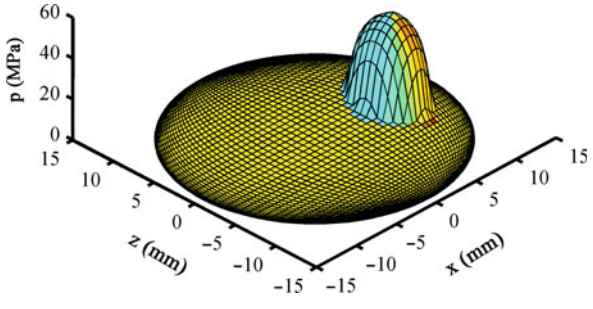
contact pressure distribution in a mixed lubrication regime needs to be made by the model of contact mechanics of either the smooth inverse filter method (SIFM) or the simplified Reynolds equation for the lubrication simulation (Wang et al. 2006, 2008a). The asperity contact pressure distribution and the film thickness of a spherical bearing of metal-on-metal hip joint replacements at an instant under the operational condition is predicted by the mixed lubrication modeling in the simulation system, as shown in Figs. 14 and 15.



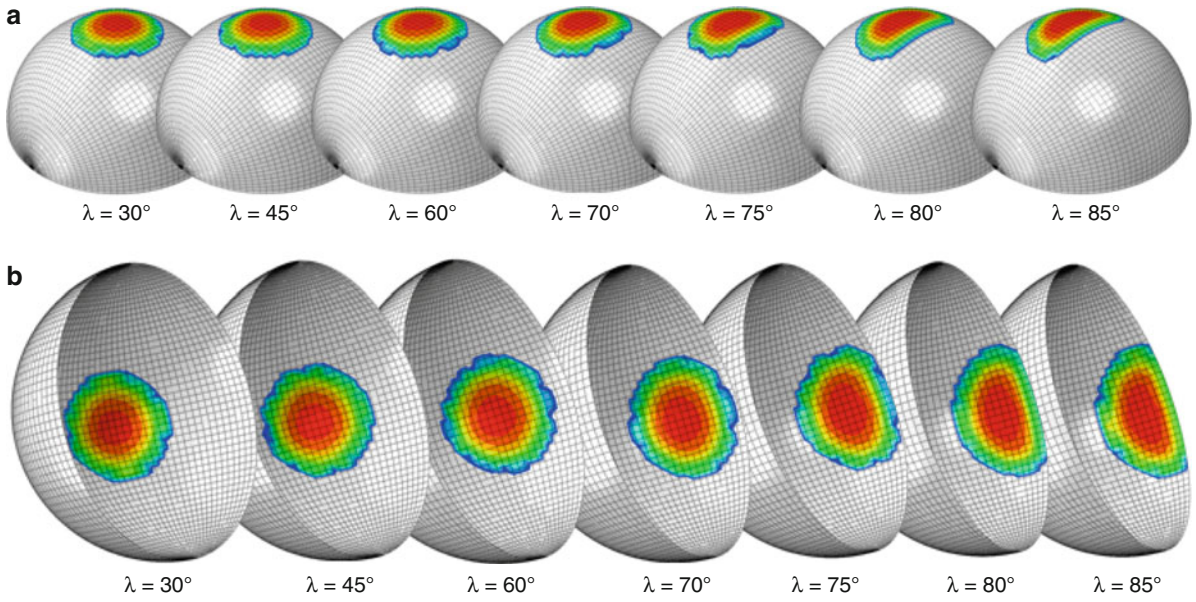
Biotribological Joint Simulation System, Fig. 17 Pressure distribution of metal-on-polyethylene hip joint replacements under the given steady operational condition



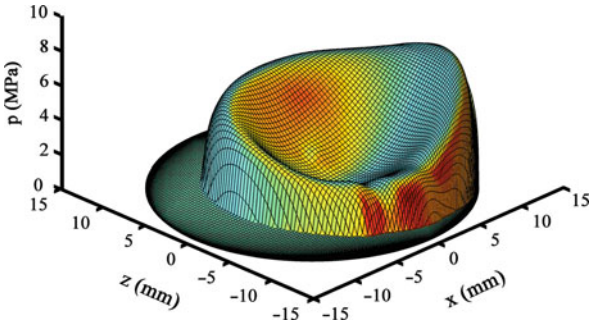
Biotribological Joint Simulation System, Fig. 18 Elastic deformation distribution of cup bearing surface of metal-on polyethylene hip joint replacements



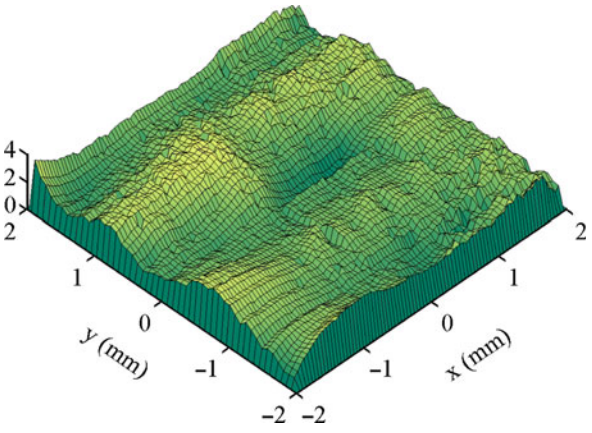
Biotribological Joint Simulation System, Fig. 19 The predicted contact pressure distribution of metal-on-metal hip joint replacements at the instant of 0.2 s during one walking cycle by the finite element model



Biotribological Joint Simulation System, Fig. 20 Comparison of contact areas under the different cup inclination angles: (**a**) case for femoral head, and (**b**) case for acetabular cup

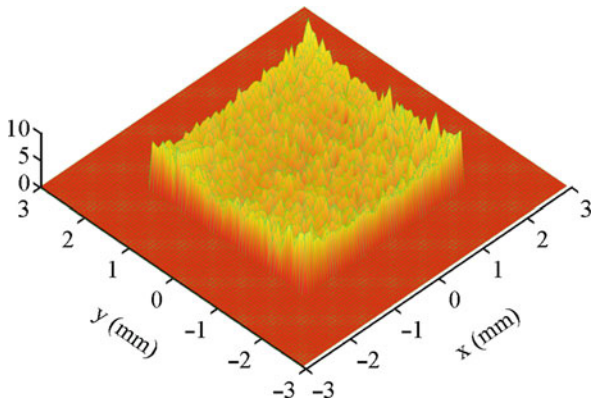


Biotribological Joint Simulation System, Fig. 21 The predicted contact pressure distribution of worn bearing of metal-on-polyethylene hip joint replacements under the loading of 2,005 N by the smooth inverse filter method

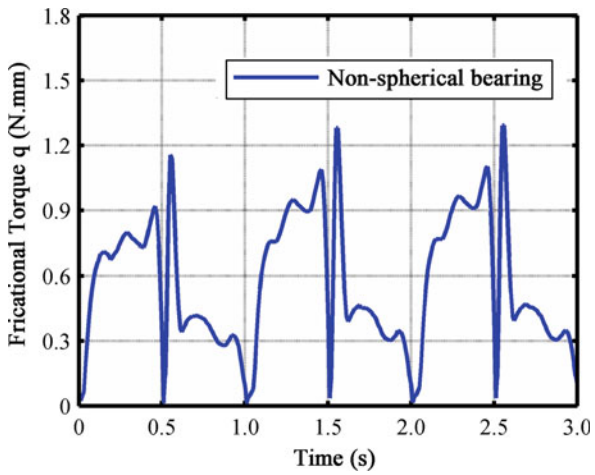


Biotribological Joint Simulation System, Fig. 22 Topography on the local area of worn polyethylene cup bearing

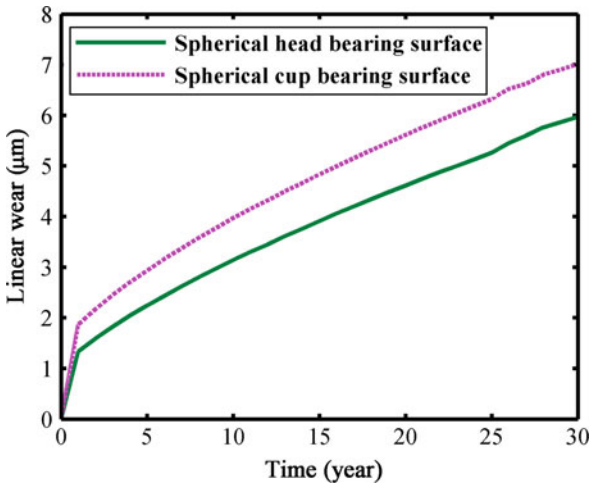
Both the numerical experiment and the physical phenomenon indicate that the lubrication problem of a metallic head bearing against a polyethylene cup bearing of hip joint replacements under the given operational condition often behaves with a bigger contact region of lubricant pressure distribution than that of the metal-on-metal hip joint replacements. The lubrication performance of a bearing due to the bigger deformation can be classified as the soft elastohydrodynamic lubrication (Wang et al. 2005). Therefore, it is seen from lubrication modeling that the convergent process of soft elastohydrodynamic lubrication is numerically different from that of metal-on-metal hip joint replacements. The film thickness and pressure distribution of elastohydrodynamic lubrication of hip joint replacements under the steady operational condition as described above are predicted by the simulation system, respectively, as shown in Figs. 16 and 17. The corresponding elastic deformation of bearing surfaces for the evaluation of film thickness is



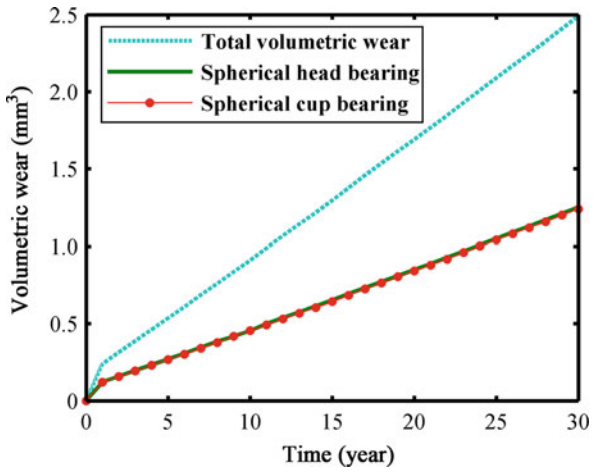
Biotribological Joint Simulation System, Fig. 23 Contact pressure distribution on the local area of worn polyethylene cup bearing



Biotribological Joint Simulation System, Fig. 24 Frictional torque of aspheric bearing of metal-on-metal hip joint replacements in full film lubrication regime during the three walking cycles



Biotribological Joint Simulation System, Fig. 25 Linear wear depth of metal-on-metal hip joint replacements under walking condition



Biotribological Joint Simulation System, Fig. 26 Volumetric wear of metal-on-metal hip joint replacements under walking condition

calculated by the equivalent discrete spherical convolution (EDSC) and the spherical fast Fourier transform (SFFT) technique, as shown in Fig. 18 (Wang et al. 2004; Wang 2010, 2011). It is noted that the initial pressure of p_0 with the zero can be employed to obtain the predicted pressure through iterative process by gradually increasing loading up to the real operational condition.

The biotribology simulation system can be applied for the evaluation of contact mechanics of the bearing surface couple by the finite element method (FEM) or the smooth inverse filter method (SIFM), as described previously. The contact pressure distribution of metal-on-metal hip joint replacements at the instant of 0.2 s during one walking cycle is predicted by the finite element model, as shown in Fig. 19 (Wang et al. 2009b, 2010). As a particular case, the contact areas of head and cup bearing surfaces with the different inclination cup bearing angles relative to head bearing are also given, as shown in Fig. 20a, b (Cheng et al. 2011, 2012a, b). Furthermore, the contact pressure of worn bearing of metal-on-polyethylene hip joint replacements under the vertical loading of 2,500 N is evaluated by the smooth inverse filter method, as shown in Fig. 21 (Wang et al. 2005). Furthermore, a surface topography

on the local area is taken from the polyethylene cup bearing after wear testing, as shown in Fig. 22. The micro-contact mechanics on the area is able to be simulated by the inverse filter method, as shown in Fig. 23 (Wang et al. 2003).

Once the solution of pressure distribution either in full film lubrication or mixed lubrication regime is achieved, the corresponding friction and frictional torque of bearing of hip joint replacements under the given operational condition can be evaluated. The frictional torque of aspheric bearing of metal-on-metal hip joint replacements under the three walking cycles is predicted, as shown in Fig. 24.

The wear prediction of spherical bearings of metal-on-metal hip joint replacements under the walking condition of the three-dimensional transient motion and the dynamic loading can be made by the wear modeling in the biotribology simulation system. In this application case, the wear factor is divided by two stages, which are $k_1 = 0.5 \times 10^{-8}$ and $k_2 = 0.15 \times 10^{-8}$ mm^3N^{-1} m^{-1} for running-in and steady stages by the trial-and-error method (Wang et al. 2010; Hu et al. 2011). The dynamic contact pressure at each instant during the walking cycle is evaluated by the finite element model (FEM), which is produced and controlled by the program code. It is seen that the variation of bearing geometry with wear in motion in body system needs to be presented in a space system for wear simulation work by the fixed-tracked method (FTM) (Wang et al. 2009, 2010). Both the linear wear depth and the volumetric wear rate of metal-on-metal hip joint replacements with time of years are calculated by the wear modeling in the simulation system, separately, as shown in Figs. 25 and 26. The dynamic contact pressure distribution can be evaluated by the finite element model with the density of 60×60 mesh grids on bearing surface, as shown Fig. 19 (Wang et al. 2010).

Cross-References

- ► Biotribological Joint Simulation System
- ► Contact Mechanics for Spherical/Aspheric Bearing
- ► Elasticity Theory for Spherical Bearings
- ► Friction Prediction for Spherical Bearings
- ► Geometry of Spherical/Aspheric Bearings
- ► Lubrication Theory for Spherical Bearings
- ► Wear Modeling of Spherical Bearings

References

J.F. Archard, Contact and rubbing of flat surface. J. Appl. Phys. **24**(8), 981–988 (1953)

Q. Cheng et al., Effect of edge contact behavior of metal-on-metal hip joint replacements. J. Med. Biomechan. **26**(6), 540–545 (2011)

Q. Cheng et al., Dynamic edge contact behavior of metal-on-metal hip joint replacements. J. Med. Biomechan. **27**(2), 29–34 (2012a)

Q. Cheng et al., Contact mechanics of non-spherical hip joint replacements. J. Med. Biomechan. **27**(5), 534–544 (2012b)

P. Flores et al., Dynamics of multibody systems with spherical clearance joints. ASME J. Tribol. **1**(7), 240–247 (2006)

Z.M. Hu et al., Wear prediction of metal-on-metal hip joint replacements. J. Med. Biomechan. **26**(5), 448–453 (2011)

F.C. Wang et al., Prediction of elastic deformation of acetabular cup and femoral head for lubrication analysis of artificial hip joints. Proc. Inst. Mech. Eng. Part J: J. Eng. Tribol. **218**, 201–208 (2004)

F.C. Wang et al., Elastohydrodynamic lubrication modeling of artificial hip joints under steady-state conditions. ASME J. Tribol. **127**(10), 729–739 (2005)

F.C. Wang et al., Microscopic asperity contact and deformation of ultrahigh molecular weight polyethylene bearing surfaces. Proc. Inst. Mech. Eng. Part H: J. Eng. Med **217**, 477–490 (2003)

F.C. Wang et al., An integrated experimental and theoretical contact mechanics study of UHMWPE hip implants tested in a hip simulator, in *ASME Proceedings of World Tribology Congress III*, Washington, 2005, pp. 311–312

F.C. Wang et al., Lubrication modelling of artificial hip joints, in *IUTAM Symposium on Elastohydrodynamic and Micro-Elastohydrodynamic 134: book series: Solid Mechanics and its Applications*, ed. by R.W. Snidle, H.P. Evans (Springer, Dordrecht, 2006), pp. 385–396

F.C. Wang et al., Transient elastohydrodynamic lubrication of hip joint implants. ASME J. Tribol. **130**(1), 011007 (2008a)

F.C. Wang et al., Lubrication and friction prediction in metal-on-metal hip joint implants. IOP Phys. Med. Biol. **53**, 1277–1293 (2008b)

F.C. Wang et al., Non-sphericity of bearing geometry and lubrication in hip joint implants. ASME J. Tribol. **131**(3), p031201 (2009a)

F.C. Wang et al., *Dynamic Contact Mechanics and Wear Modelling of Hip Joint Replacements with Hard-on-Hard Material Combination under Three-Dimensional Loading and Transient Motion*, World Tribology Congress IV, Kyoto, 2009b

F.C. Wang, *Dynamic Contact Behaviour and Evolution of Bearing Interface with Spherical-Base Geometry*, National Natural Science Foundation of China (NSFC) Report No.10972165, pp. 1–30 (2010)

F.C. Wang et al., Tribological modelling of spherical bearing with complex spherical-base geometry and motion. Tribol. Des. WIT Trans. Eng. Sci. **66**, 3–15 (2010)

F.C. Wang, *Dynamic Contact Mechanism and Failure of Rolling Bearing*, National Key Basic Research Program of China (973 Program) Report No. 2011CB706601 (2011)

Biotribology of Ceramic-Ceramic Hip Implant

► Self-Mating Ceramic Applications in the Hip Joint

Biotribology of Contact Lenses

John B. Medley[1], Valentina Ngai[2]
[1]Department of Mech & Mechatronics Engineering, University of Waterloo, Waterloo, ON, Canada
[2]Biomechanical engineering, Robson Forensics Inc, Chicago, USA

Synonyms

Tribology of soft hydrogel contact lenses

Definition

Biotribology is the tribology of a biological system. In the present case, the contact lens itself is not a biological component but the posterior surface of the eyelid, the corneal eye surface, and the tear film are all biological components and thus the overall system is indeed biological.

Scientific Fundamentals

Pre-Lens Interface

Contact lenses are implants with two biotribological interfaces (Fig. 1): the *post-lens* between the posterior surface of the lens and the eye surface (cornea) and the *pre-lens* between the anterior surface of the lens and the eyelid (Dowson 2009). When a contact lens is not in place, the tribology of the cornea interface with the eyelid has been the subject of complex tribological analysis (Jones et al. 2008). Earlier theory (Kamiyama and Khonsari 2000) applied similarly advanced analytical techniques to the post-lens interface of contact lenses. However, it is the pre-lens interface and the associated frictional forces that are likely to be critical for contact lens comfort, and thus it has been the focus of experimental (Rennie et al. 2005; Ngai et al. 2005) and theoretical (Nong and Anderson 2010) studies.

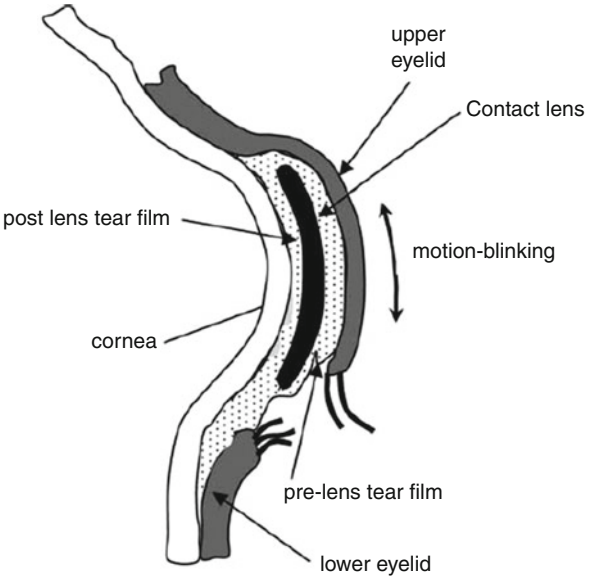
There is a lubricant present: the tear film, which consists of proteins and lipids with an underlying layer of mucin covering the natural tissue surfaces (Dowson 2009). Despite the tear film lubricant, the repeated sliding motion of the upper eyelid over a contact lens during eye blinking may still cause some damage to the posterior eyelid surface and discomfort to the lens wearer.

Contact Lens Materials

The mechanical properties and the interaction of the lens material with tear film molecules are of prime importance. Early contact lenses were made from relatively rigid acrylics, but such hard lenses are rarely used now and the softer hydrogel materials dominate (Efron 2010).

The hydrogel materials, that are used for the overwhelming majority of contact lenses, must be optically transparent, oxygen permeable, hydraulically permeable, durable, resistant to protein deposition, and cost-effective to manufacture in a form that provides vision correction (Jones 2002).

Lens materials can be classified as conventional hydrogel (CH) that can be manufactured at low cost and silicone hydrogel (SH) that allows greater oxygen permeability. In general, SH lenses are more comfortable because they do not starve the eye surface of oxygen. Thus, they permit the convenience of longer wear without removal, although there still some debate regarding how often lenses should be changed. SH has now replaced CH as the dominant type of contact lens material (Pence 2010).



Biotribology of Contact Lenses, Fig. 1 Pre-lens and post-lens biotribological interfaces for contact lenses (Reproduced with permission of Sage Publications Ltd. from Dowson 2009)

Biotribological Issues

Although SH lenses are the dominant type in use, some biotribological issues remain (Jones et al. 2010) related to comfort and protein deposition. Also, there can be problems with microbial keratitis and corneal infiltrates. The dynamic friction encountered between the under-surface of the eyelid and the anterior surface of the lens (the pre-lens interface described in Fig. 1) is suggested as a source of discomfort to lens wearers (Rennie et al. 2005;

Ngai et al. 2005; Nong and Anderson 2010) and thus merits further consideration.

Key Applications

The key applications of the scientific fundamentals that are involved in contact lens biotribology include friction measurement and modeling along with some further consideration for lipid and protein deposition.

Friction

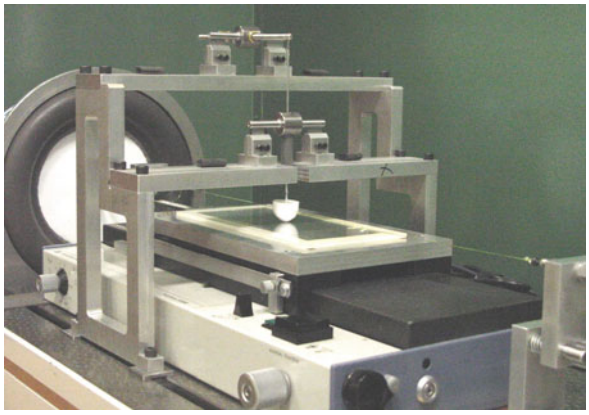
Measurement

Various experimental approaches have been used to measure friction forces associated with the pre-lens interface. In one approach, a sophisticated apparatus (Fig. 2) was developed (Ngai 2004; Ngai et al. 2005). It had a silicone rubber eye-form that retained the contact lens on the 7.8 mm radius of the cornea zone and was loaded against a smooth (~7 nm Ra roughness over small regions) flat glass plate. An electromagnetic drive was used to apply sinusoidal reciprocating motion with an average velocity of 41.2 mm/s. Knife-edge joints were used to reduce internal friction and the small friction forces were measured with a piezoelectric force transducer.

Another similarly sophisticated apparatus was developed (Rennie et al. 2005). In this study, a smooth (~8 nm Ra roughness) glass sphere of 1 mm radius was loaded against a contact lens (with an estimated Ra roughness of ~15 nm) retained on a 7.1 mm radius spherical plastic base. The base moved at velocities up to 6.28 mm/s. Friction and normal forces were measured by optical sensors that detected the small deflections of glass "flexures" (presumably beams).

While neither apparatus simulated the pre-lens interface geometry exactly, they did use actual contact lenses and had the forces, motions, lubricant, and lens support compliance quite close to the physiological range. The initially mentioned apparatus (Ngai 2004) was perhaps a little closer with its use of an eye form. However, both studies applied normal forces as low as 10–20 mN and measured friction forces of about 1 mN. Also, both studies gave some evidence supporting the role of elastohydrodynamic lubrication in contact lens friction.

In the first study mentioned (Ngai et al. 2005), tests were performed in the presence of phosphate buffer solution (PBS) and then repeated using a solution "doped" with two types of protein constituents present in tear films (1.5 ml volume of sterile PBS with 10 mg/ml hen-egg lysozyme and 1 mg/ml of human serum albumin). The friction forces were measured under two normal loads for SH and CH lenses, using the two lubricants (Fig. 3).



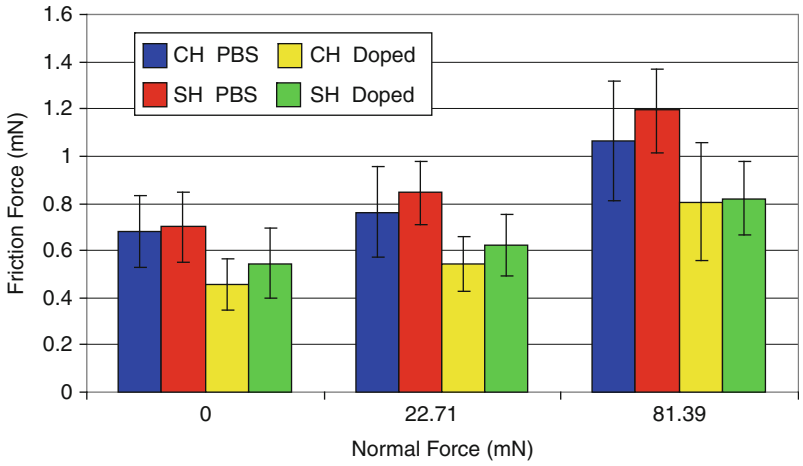
Biotribology of Contact Lenses, Fig. 2 Friction testing apparatus (Reproduced with permission of Elsevier from Ngai et al. 2005)

Friction forces increased significantly with increasing normal force ($p < 0.001$). Protein doping caused a significant decrease in the friction forces for both CH and SH lenses ($p < 0.001$). It was suggested that early stages of protein deposition, prior to accumulation or denaturing, might help lubricate the lens-eyelid interface, thereby reducing friction forces, as had been postulated (Jones 2002). In these experiments, SH lenses did not incur significantly higher friction forces than CH lenses.

Modeling

Complex theoretical approaches (Jones et al. 2008; Nong and Anderson 2010) to the pre-lens contact could be adapted to contact lens lubrication but as yet this has not been done. Earlier research also looked promising (Kamiyama and Khonsari 2000), but it modeled the post-lens contact instead of the pre-lens one that was of more interest with regard to contact lens comfort issues. In the second experimental study mentioned above (Rennie et al. 2005), the friction was modeled but they did not seem to be able to obtain a fully independent comparison of their model predictions with their data. However, they did suggest that a mixed film lubrication occurred (a combination of fluid film and boundary lubrication). This speculation was supported by other researchers (Dowson 2009; Jin and Dowson 2005).

Although a full model for the lubrication of contact lenses has not yet been developed, a simple analytical model can be developed for the friction forces measured in the initially mentioned study (Ngai 2004). This follows a previous approach (Ngai et al. 2005) but only takes the deformation of the eye-form into account. The clinical



Biotribology of Contact Lenses, Fig. 3 Friction force measurements from Ngai (2004)

relevance of this model depends on the fidelity of the contact in representing the actual in vivo pre-lens contact. This is a very important limitation.

The previously described experiments (Ngai 2004) had normal forces (F) of ~0 mN, 22.71 mN, and 81.39 mN applied. The sinusoidal motion of the lower glass surface produced an average entrainment velocity (u_{avg}) of 20.6 mm/s. As already mentioned, the surface deformation was assumed to be only that of the eye-form (made of silicone rubber and considered to be a semi-infinite solid) that had an elastic modulus (E) of 2.6 MPa and an estimated value of 0.5 for Poisson's ratio (υ). The somewhat lower elastic modulus of the 0.1-mm-thick contact lens that covered the eye-form surface in the contact zone would enhance film development but perhaps not all that significantly. In this model, only the PBS solution was considered, and it was assumed to have a constant viscosity (η) equal to that of water at room temperature (~ 1 mPa s). The radius of curvature (R) of the corneal "bump" in the eye-form (where the contact occurred) was 7.8 mm.

The central film thickness (h_C) was calculated using the elastohydrodynamic lubrication (EHL) formula (1) that was specified in a study on hip implant wear (Medley et al. 1997). The simple model assumed that the average entrainment velocity could be used to predict an average central film thickness that would stay fairly constant over the motion cycle at cyclic steady state.

$$h_C = 5.083\ \frac{(\eta u_{avg})^{0.660} R^{0.767}}{(E')^{0.447} F^{0.213}} \qquad (1)$$

where

$$E' = \frac{2E}{1-\upsilon^2}$$

To continue this approximate analysis, the average central film thickness acting over the Hertzian contact zone was assumed to be sheared at the maximum sliding velocity (U_{max}) to obtain the following estimate (2) of the friction force (F_f) encountered at the center of the reciprocating stroke.

$$F_f = \frac{\pi \eta U_{max} a^2}{h_C} \qquad (2)$$

where

$$a = \left(\frac{1.5FR}{E'}\right)^{\frac{1}{3}}$$

For the pin-on-plate configuration under consideration (Ngai 2004), it was possible to derive a rather uncomplicated expression for the coefficient of friction (μ) in terms of a dimensionless Sommerfeld parameter (S) and a dimensionless load parameter (W) using (1) and (2) as follows:

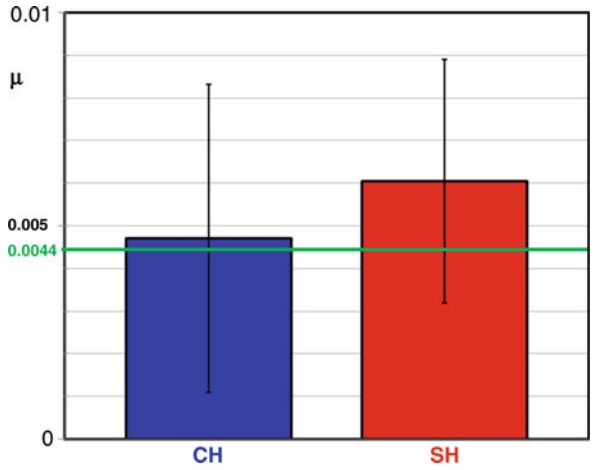
$$\mu = 1.724\ S^{0.340}\ W^{0.220} \qquad (3)$$

where

$$\mu = \frac{F_f}{F} \quad S = \frac{\eta U_{max} R}{F}$$

$$W = \frac{F}{E'R^2} \quad U_{max} = \pi u_{avg}$$

Using (1), the average central film thickness was estimated as 148 nm for a load of 81.39 mN. This film



Biotribology of Contact Lenses, Fig. 4 Coefficients of friction for the CH and SH lenses and the predicted value of 0.0045 for a normal force of 81.39 mN and PBS as a lubricant

thickness was large compared with the roughness of the glass plate that was ~7 nm Ra over small regions and the roughness of a contact lens, which was estimated as ~15 nm Ra (Rennie et al. 2005). Thus it was considered likely that a continuous EHL film separated the lens from the glass surface.

In the previously described experiments (Ngai 2004), it was thought that the tests conducted at nominally zero normal force gave a fairly good estimate of the fluid drag forces acting on the eye-form away from the contact zone. Presumably, the same drag force would also act in the tests conducted at the two values of normal force. In efforts to eliminate this drag (which was not modeled in the preceding simple analysis), the average of the mid-stroke friction forces at zero normal force were subtracted from the average of the mid-stroke friction forces at the two values of normal force. However, the scatter in the friction force measurements was so high that only the values for the higher normal force of 81.39 mN were precise enough to compare with the predictions of the simple analytical model. Thus, these friction values for the CH and SH lenses were divided by the normal force to give coefficients of friction that were then compared with the prediction ($\mu = 0.0045$) of the simple analytical model (Fig. 4). The agreement was surprisingly good, which suggested that in the experiments (Ngai 2004), EHL did indeed occur.

Lipid and Protein Deposition

The role of lipid and protein deposition regarding friction and thus comfort has become more prominent in the recent literature of contact lenses. The protein deposition may impair an EHL film protection or it may enhance it. Recent studies have concentrated on developing methods to study protein deposition (Teichroeb et al. 2008) and efforts to quantify both protein and lipid deposition on lenses retrieved from patients (Boone et al. 2009). While friction studies in the presence of protein-containing fluids have been performed (Ngai 2004), a good understanding of the biotribological mechanisms remains elusive (Jones et al. 2010). It is likely that future research will consider the role of proteins and lipids in the EHL film formation and the friction that develops in the pre-lens contact.

Conclusion

The role of pre-lens friction in contact lens comfort remains a topic of concern. The biotribology of this contact has been examined both theoretically and experimentally. Recent interest has developed a focus on lipid and protein interactions with the contact lens materials in the pre-lens contact. Further research needs to be done to fully understand the biotribology of contact lenses.

Cross-References

► Film Thickness Formulas: Point Contacts
► Hertz Theory: Contact of Spherical Surfaces
► Polymeric Elastomers: Material Aspects of Tribology

References

A. Boone, M. Heynen, E. Joyce, J. Varikooty, L. Jones, Ex-vivo protein deposition on bi-weekly silicone hydrogel contact lenses. Optom. Vis. Sci. **86**(11), 1241–1249 (2009)

D. Dowson, A tribological day. Proc. IMechE. Part J: J. Eng Tribol. **223**, 261–272 (2009)

N. Efron, Obituary – rigid contact lenses. Cont. Lens Anterior Eye **33**, 245–252 (2010)

Z.M. Jin, D. Dowson, Elastohydrodynamic lubrication in biological systems. Proc. IMechE. Part J: J. Eng Tribol. **219**, 367–380 (2005)

L. Jones, Modern contact lens materials: a clinical performance update. Cont. Lens Spectrum **17**(9), 24–35 (2002)

M.B. Jones, G.R. Fulford, C.P. Please, D.L.S. McElwain, M.J. Collins, Elasto-hydrodynamics of the eyelid wiper. Bull. Math. Biol. **70**, 323–343 (2008)

L. Jones, A. Epstein, B. Holden, L.B. Szczotka-Flynn, The genesis of silicone hydrogels. Cont. Lens Spectrum, Issue: October, (2010)

S. Kamiyama, M.M. Khonsari, Hydro-dynamics of a soft contact lens during sliding motion. ASME J. Tribol. **122**, 573–577 (2000)

J.B. Medley, J.J. Krygier, J.D. Bobyn, F.W. Chan, A. Lippincott, M. Tanzer, Kinematics of the MATCO^{TM} hip simulator and issues related to wear testing of metal-metal implants. Proc. IMechE: Part H, J. Eng. Med. **211**, 89–99 (1997)

V. Ngai, Development of a friction simulator for the anterior surface of hydrogel contact lenses. MASc Thesis, Dept of Mechanical Engineering, University of Waterloo, ON, Canada (2004)

V. Ngai, J.B. Medley, L. Jones, J. Forrest, J. Teichroeb, Friction of contact lenses: silicone versus conventional hydrogels. Tribol. Interf. Series **48**, 371–379 (2005)
K. Nong, D.M. Anderson, Thin film evolution over a thin porous layer: modeling a tear film on a contact lens. SIAM J. Appl. Math. **70**(7), 2771–2795 (2010)
N. Pence, Silicone hydrogel lenses: Where do we stand today? Cont. Lens Spectrum, Issue March (2010)
A.C. Rennie, P.L. Dickrell, W.G. Sawyer, Friction coefficient of soft contact lenses: measurement and modeling. Tribol. Lett. **18**(4), 499–504 (2005)
J.H. Teichroeb, J.A. Forrest, V. Ngai, J.W. Martin, L. Jones, J. Medley, Imaging protein deposits on contact lens materials. Optom. Vis. Sci. **85**(12), 1151–1164 (2008)

Biotribology of Tactile Perception: Effect of Mechano-transduction

Hassan Zahouani[1], Roberto Vargiolu[2], Thierry Hoc[2]
[1]Ecole nationale d'Ingénieur de Saint Etienne, Laboratory of Tribology and Dynamic of Systems, UMR 5513 CNRS-ECL-ENISE, University of Lyon, Ecully Cedex, France
[2]Ecole Centrale de Lyon, Laboratory of Tribology and Dynamic of Systems, UMR 5513 CNRS-ECL-ENISE, University of Lyon, Ecully Cedex, France

Synonyms

Friction and vibration of human finger during a perception test

Definition

Biotribology of tactile perception is a rough contact method used to identify the role of the digital print of the human finger and to understand the mechanism of mechano-transduction during a touch test.

Scientific Fundamentals

Perception

Perception by touch is a sensory modality that is essential to the representation of our environment. It also plays an important role in the industrial world, since the touch property is now considered a quality criterion for many everyday products (textiles, cosmetics, leather, food, packaging, paints, "soft feel" soft plastics for computers, phones, and automobile interiors). The development of realistic virtual environments requires integration of the haptic component (science of touch) in devices (robotics, remote manipulation, etc.). The word *haptic* was used in the nineteenth century by the Austrian art historian Riegl. Art progresses by revealing a world of sensitivity of the haptic to the optic.

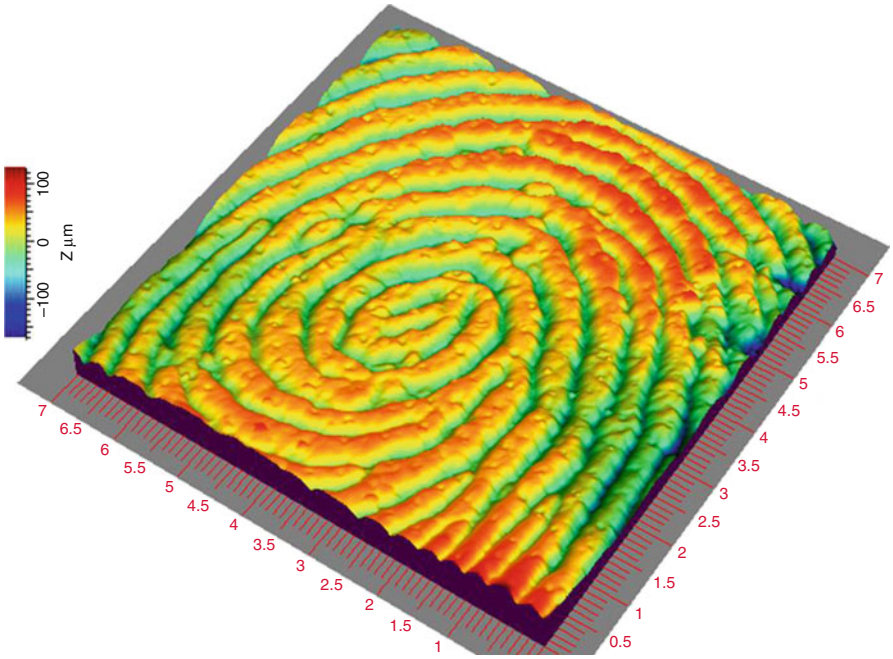
The tactile experience starts very early in humans. From the embryonic stage, the fetus receives the pressure and stimulation of amniotic fluid. These stimuli are translated, transformed, and processed by the somatosensory system, which includes receptors that provide information about temperature, texture, pain, and so on, and are particularly responsible for the reception of tactile messages. These messages are transmitted through the spinal cord to the central nervous system.

In humans, the mechanical skin sensation, or sense of touch, can detect and differentiate multiple environmental mechanical stimuli: vibration, texture, pinching, and stretching skin. The sensory neurons in the somatosensory system control this modality. Every nerve ending has specialized receptors sensitive to mechanical deformations (mechanoreceptors) that convert mechanical deformation into different specific electrical signals to generate action potentials responsible for conducting information to the central nervous system.

Cutaneous mechano-transduction requires the conversion of mechanical deformation into an electrical signal to generate action potentials. The complex mechano-transduction is well defined at the molecular level in the nematode *Caenorhabditis elegans* (Lumpkin et al. 2010). In contrast to mammals, there is no definitive evidence to confirm their molecular identity in the somatosensory system. Among the candidates, ion channels, proteins of the extracellular matrix, and mechanical-sensitive receptor cells of the skin have received particular attention from researchers.

Nerve fibers in the skin, especially at our fingertips, enable human beings to explore their environment through touch; the skin thus allows sensitivity to pressure, temperature, and pain. It has four types of receptors, which react to different stimuli, and that return information to be interpreted by the brain. This information traverses the spinal cord to the thalamus via two channels: one for information about pain and temperature, the other for touch (e.g., texture, hardness, etc.).

The entire surface of our skin is equipped with sensory receptors, however, the greatest number is found at the end of the hand and fingers. The sense of touch is concentrated in the hand, where there are more than 2,000 nerve endings per square centimeter at the fingertips. Neurophysiological function is provided by various tactile sensors converting mechanical stress applied to the surface



Biotribology of Tactile Perception: Effect of Mechano-transduction, Fig. 1 Three-dimensional topography of human finger

by a pulsed electrical signal transmitted to the brain. Each of these receptors is sensitive in a given frequency band. According to biologists (Blake et al. 1997; Bolanowski et al. 1988; Bolanowski and Zwislocki 1984; Cascio and Sathian 2001), the receivers can detect Pacini amplitudes of stress of a few microns at a frequency of 300 Hz (maximum sensitivity) and Meissner's corpuscles respond to lower frequency, in the order of 50 Hz (sensitivity of touch). The slowly adapting sensors (Merkel, Ruffini) determine a loop system with the brain and the depth of penetration of the skin.

When rubbed with the finger, the mechanical stress that stirs our sense receptors will result in tactile sensations analyzed and identified by the brain in real time. These phenomena are systematically accompanied by the generation of vibration of touch, more or less perceptible to our ear, but the acoustic signature is indicative of the quality of touch.

Stresses and Strains Upon Contact Between the Finger and a Rough Surface

Humans use pressure and vibration receptors to evaluate the surfaces they touch and are capable of fine spatial and temporal resolution. The human tactile sensory system provides inspiration and a good starting point in the search for relevant signals that can adequately describe the soft finger–surface interface.

The main underlying hypothesis is that human sense of touch can be decomposed into several quantitative components, similar to the decomposition of color into red, green, and blue with different intensities. The supposed quantitative components are (a) vertical pressure gradient inside the skin, (b) spatial pressure distribution on the skin surface, (c) vibration, (d) skin stretch, and (e) change of skin temperature. This hypothesis is supported by the known facts about the physiology of touch. Before outlining the specific goals of the proposed study, some further explanation about the physiology and neurology of touch is needed.

To identify the role of the digital print of a human finger on the contact, spatial pressure distribution on the skin surface, vertical pressure gradient, and friction, an elastic model was adopted to take full account of interactions between all contact points and to predict the contact geometries of the finger topography under load. The problem is treated in this model by assuming normal effects without the presence of adhesion.

Analysis of the morphology of human fingerprints show morphologies with curves called dermatoglyphics. Figure 1 shows the morphology of the human fingertip obtained by making a replica of the fingerprint of the finger. The three-dimensional topography was measured by a confocal microscope in white light. The furrows of dermatoglyphics serve as flowing channels and for

retention of sebum and sweat. They capture and hold the substances applied to the skin and are therefore the preferred sites of dermal absorption. The mode of deformation and the contact pressures exerted on the dermatoglyphics are responsible for the mechano-transduction and transmission of information of touch, which requires a detailed analysis of the contact mechanics of dermatoglyphics.

Contact Model: Single Smooth Stimulus and Soft Human Finger

In this elastic model, a contact between a single smooth spherical stimulus and soft human finger is considered. The model will consider the three-dimensional character of dermatoglyphics and the geometry of the stimuli to determine the contact pressures, the deformation of dermatoglyphics, and stress distribution in skin layers. In this case, the elastic displacement of a point M of the finger, $u(r)$, is linked to the punctual applied load at a distance r by Boussinesq's equation (Boussinesq 1885):

$$u(r) = \frac{(1-\nu^2)}{\pi E^{*}..}\frac{F_n}{r} \tag{1}$$

Boussinesq's relationship (1) makes it possible to link the displacement $u_i(x,y)$ of a point $M(x, y)$ of a semi-infinite elastic solid with the contact pressure $p(x',y')$ in $M(x',y')$ related to the contact surface, by means of the following Boussinesq's integral equation:

$$u_i(x,y) = \frac{1-\nu_i^2}{\pi E^{*}}\int\int_{A_c}\frac{p(x',y')dx'dy'}{\sqrt{(x-x')^2+(y-y')^2}} \tag{2}$$

To determine the displacement of a point $M(x,y)$ on the surface due to the contact pressure, we must take into account the superposition of all the actions of the different loads in the contact region. The displacement of a point M_{ij} can be written as (Johnson 1985):

$$u_{ij} = C_{ijkl}p_{kl} \tag{3}$$

where C_{ijkl} is the influence matrix. The problem is finding a pressure $p_{ij} > 0$ to solve the following contact equation:

$$C_{ijkl}p_{kl} + h(x,y) - \delta = 0 \tag{4}$$

$$\begin{aligned} u_{ij} &= \delta - h(x,y),\ p_{kl} > 0 \text{ within contact area} \\ \mathrm{u_{ij}} &> \delta - h(x,y),\ p_{kl} = 0 \text{ outside contact area} \end{aligned} \tag{5}$$

Contact pressure must also verify

$$\frac{1}{A_c}\int\int_{A_c}P(x',y')dx'dy' = P_A \tag{6}$$

The contact (7) can be converted into the spatial frequency domain using fast Fourier transform (FFT) techniques. Spectral analysis can be used to determine an algebraic relationship between the surface displacement and the contact pressure on different length scales (Zahouani et al. 2011).

When the problem is converted into the frequency domain, the convolution becomes the product of the input and the transfer function:

$$U(v_x,v_y) = \frac{2(1-\nu^2)}{E^{*}v_{x,y}}P(v_x,v_y) \tag{7}$$

The total load is given as:

$$F_n = \sum_i\sum_j p_{ij}(x,y)\Delta x\Delta y \tag{8}$$

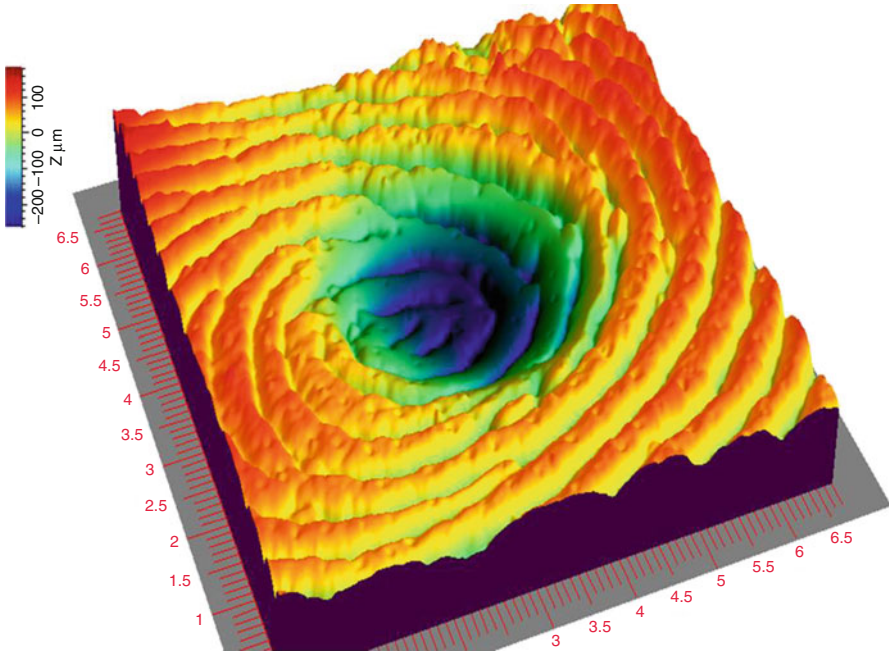
The convergence of the general equation is obtained when the relative difference between the external applied load P and computed load F_n is less than 10^{-4} (Zahouani et al. 2011):

$$\left|\frac{(P-F_n)}{P}\right| \le 10^{-4} \tag{9}$$

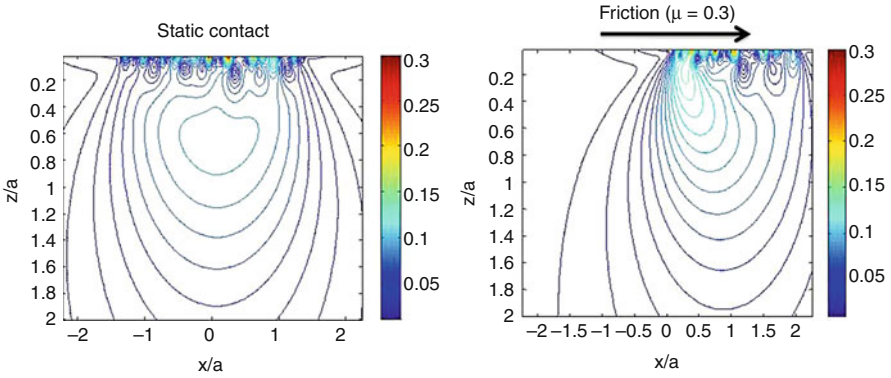
Under this assumption of elastic deformation, in a static and sliding contact, the stress field and mechanical deformation are determined on the surface of dermatoglyphics and throughout the depth of subdermal layers. The elastic model can predict the deformation of the digital pulp of 100 KPa modulus of elasticity in contact with a spherical asperity with an elastic modulus of 210 GPa (Fig. 2). Contours of constant normalized von Mises stresses $\frac{\sigma_{VM}}{p}$ for a rigid spherical indenter acting on the finger for static and sliding contact are presented in Fig. 3, with

$$\begin{aligned}\sigma_{VM} = \frac{1}{\sqrt{2}}\Big[&(\sigma_{xx}-\sigma_{yy})^2+(\sigma_{xx}-\sigma_{zz})^2 \\ &+(\sigma_{zz}-\sigma_{yy})^2+6\left(\tau_{xy}^2+\tau_{xz}^2+\tau_{yz}^2\right)^2\Big]^{1/2},\end{aligned} \tag{10}$$

and p the elastic pressure. The space coordinates are normalized with the contact radius a: $\left(\frac{x}{a},\frac{y}{a},\frac{z}{a}\right)$. Figure 4 shows the effect of friction on normalized von Mises stresses. Compared with a static contact, the results of sliding contact with a friction coefficient $\mu = 0.3$ show the modification of the size of contact area, the contact pressure distribution over the crests of dermatoglyphics, and the stress distribution on surface and sublayers of the skin of the finger containing the tactile sensors.



Biotribology of Tactile Perception: Effect of Mechano-transduction, Fig. 2 Three-dimensional deformation of human finger by a spherical stimulus



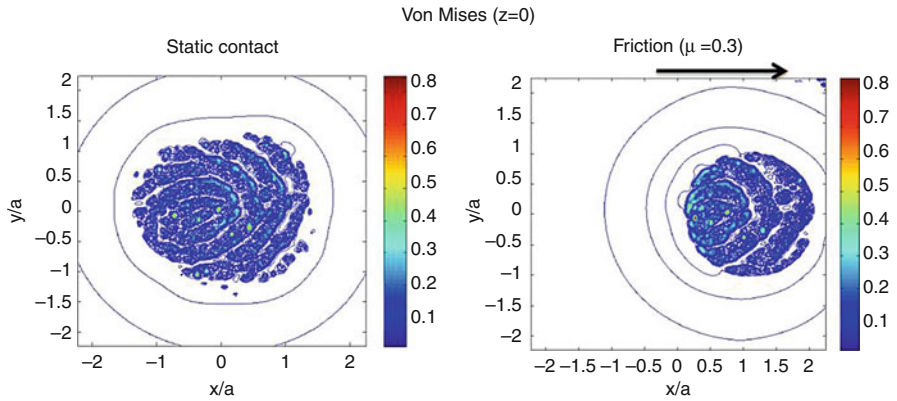
Biotribology of Tactile Perception: Effect of Mechano-transduction, Fig. 3 Von Mises equivalent stress for static and sliding contact

Texture Effect on Finger Deformation

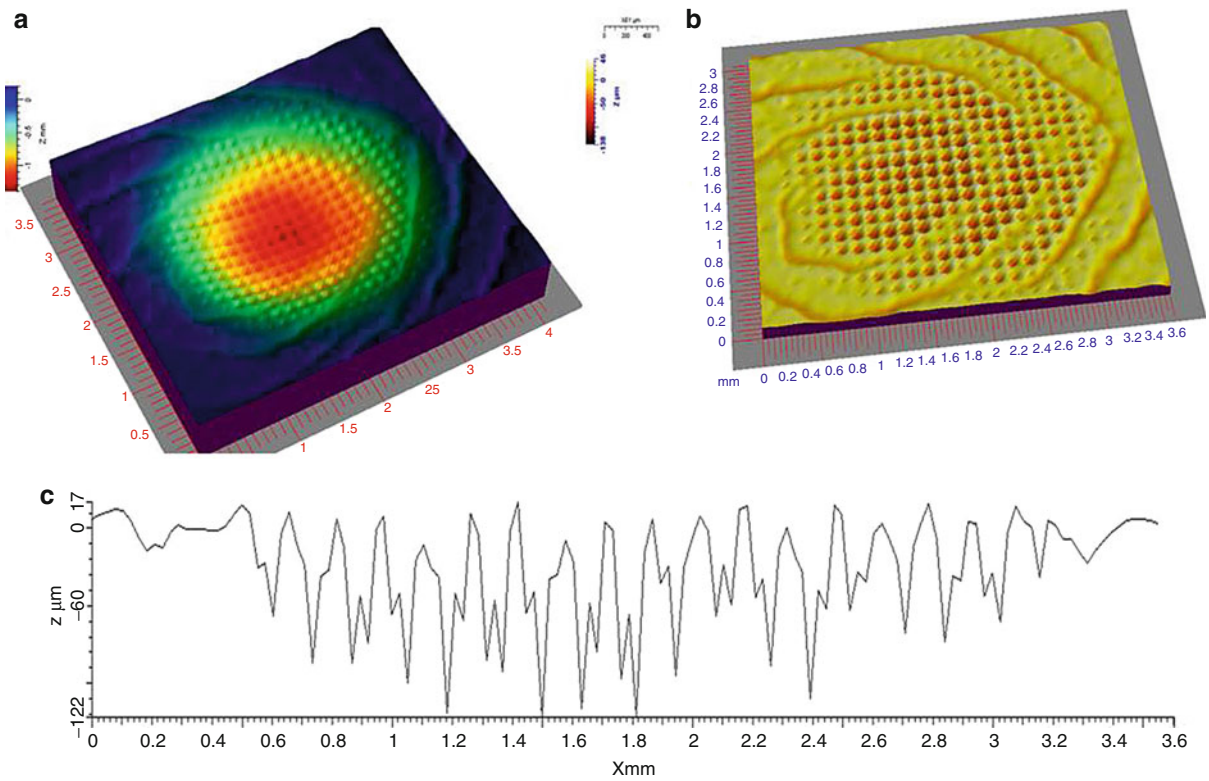
The effect of multiple stimuli was investigated by texturing the surface of a spherical indenter with a periodic roughness (Ra = 40 μm and periodicity of 35 μ). The study of contact in Fig. 5 shows how the stimuli print their forms on the surface of the finger and how the contact area and the distribution of the pressure field are modified by the morphology of the texture of the sphere, which transmits a discrete deformation field consistent with the roughness of the sphere, in contrast to the contact with a smooth sphere, which prints only its shape.

Component Deformation of the Frictional Force

The study of friction forces between two surfaces sliding against each other is a research area undergoing a revival.

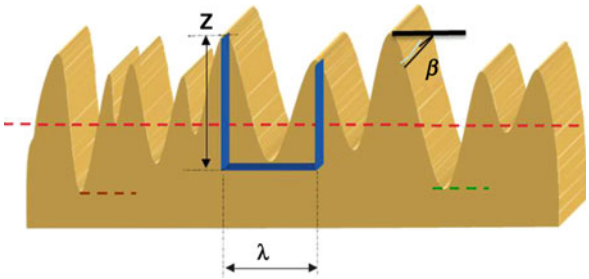


Biotribology of Tactile Perception: Effect of Mechano-transduction, Fig. 4 Von Mises equivalent stress for static and sliding contact for ($z = 0$)



Biotribology of Tactile Perception: Effect of Mechano-transduction, Fig. 5 Effect of roughness on residual deformation: (**a**) macroscopic deformation and roughness transfer to the finger, (**b**) residual roughness printed on the finger, (**c**) profile of roughness printed on the finger

Yet, as long ago as the fifteenth century, Leonardo da Vinci introduced the concept of friction, and two to three centuries later, the Frenchmen Amontons and Coulomb postulated proportionality between friction force and normal force carried by the two surfaces. This simple law, classically taught and used, is not universal. The reality is often more complex and the contribution of the friction force only to the tactile perception has not yet



Biotribology of Tactile Perception: Effect of Mechano-transduction, Fig. 6 Definition of roughness parameters

Biotribology of Tactile Perception: Effect of Mechano-transduction, Table 1 Roughness parameters of abrasive paper

Texture	SPa (μm)	β (°)	(z/λ)
600	21	58.52	1.53
800	14.3	57.41	1.32
1,200	13.3	53.01	1.12
2,400	4.53	35.84	0.45
4,000	3.18	22.64	0.21

Biotribology of Tactile Perception: Effect of Mechano-transduction, Table 2 Contact and friction parameters

Texture	ε_{def}	μ_{def}	$K_T\left(\frac{N}{m}\right)$
600	0.328	1.01	3.25
800	0.312	0.99	3.24
1,200	0.226	0.84	3.81
2,400	0.144	0.45	7
4,000	0.08	0.25	12

been demonstrated. Current research shows that the energy dissipated by friction in the form of vibration or sound wave is a way to access certain parameters related to tactile perception (Zahouani et al. 2009).

To identify the effect of the roughness scale on tactile perception, investigations with surfaces of abrasive paper were performed. The three-dimensional topography of the abrasive paper was measured by confocal microscopy. Roughness parameters that have been quantified to identify their influence on the vibration level of the finger include the average arithmetic roughness parameter of the surface SPa (equivalent to the standard parameter R_a), the mean angle of attack of each peak of the roughness (β), and the ratio between the spatial and amplitude of roughness $\left(\frac{z}{\lambda}\right)$, which is substantially equivalent to the definition of scale of the texture and the density of abrasive paper (Fig. 6). The results of topographic analysis of different abrasive papers are summarized in Table 1.

The friction force can be considered as a combination of forces that resist motion during each energy conversion processes. The total friction force, F_{tot}, may be given by the sum of two non-interacting terms: a deformation term, F_{def}, and an interfacial adhesion term, F_{ad}; thus,

$$F_{tot} = F_{def} + F_{ad} \tag{11}$$

Deformation Component of Friction Force

Friction may be usefully regarded as the energy dissipated per unit sliding distance. The deformation component arises from the work dissipated by the rigid indentor in deforming the material. For viscoelastic materials, such as skin, the deformation–recovery cycle is associated with viscous energy losses. According to the experimental work examined by (Greenwood and Tabor 1957), the deformation component exerted by the conical asperities with a mean attack angle β on a viscoelastic material is defined as

$$F_{def}(\delta) = \left(\frac{2E^*}{\pi}\right)\cot\beta.\delta^2 \tag{12}$$

The mean deformation transmitted to the finger is defined as

$$\varepsilon_{def} = 0.2\ tg\,\beta \tag{13}$$

The normal contact stiffness and the friction coefficient of deformation can be defined, respectively, as

$$k_N = \frac{dF_e}{d\delta} = \left(\frac{4E^*}{\pi}\right)\cot\beta\ \delta \tag{14}$$

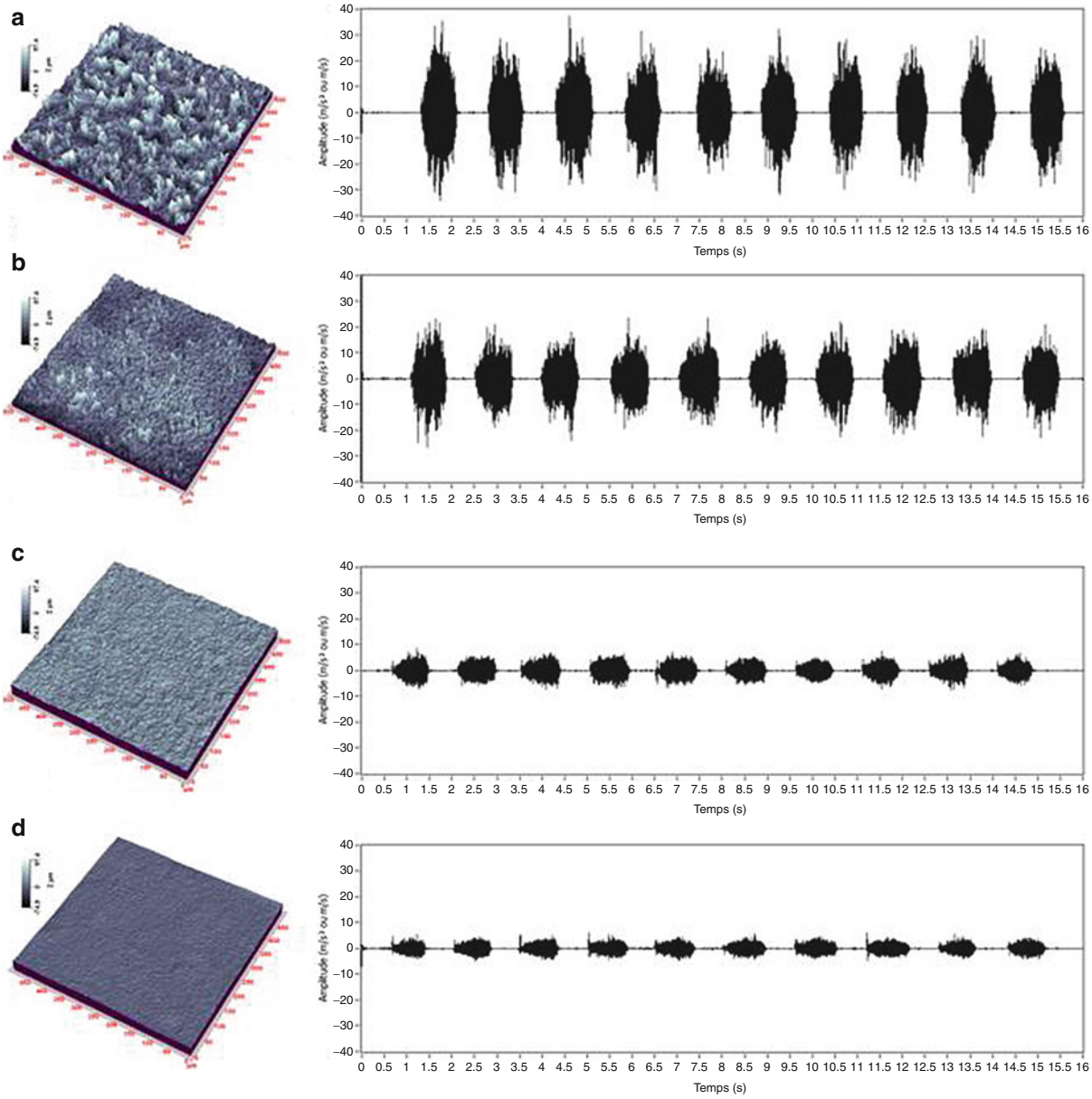
$$\mu_{def} = \frac{2}{\pi}tg\beta \tag{15}$$

In the case of micro-contacts, (Mindlin 1949) defined the ratio of the tangential stiffness and axial stiffness with the same elastic constants as

$$\frac{k_t}{k_n} = \alpha\frac{2(1-\nu)}{(2-\nu)} \tag{16}$$

where α can vary from 0.25 to 0.35. In our study we choose the tangential stiffness as $K_T \approx 0.3K_N$.

Using the values of topographic parameters quantified on abrasive paper (Table 2), one can evaluate the amount of contact and friction associated with the deformation force.

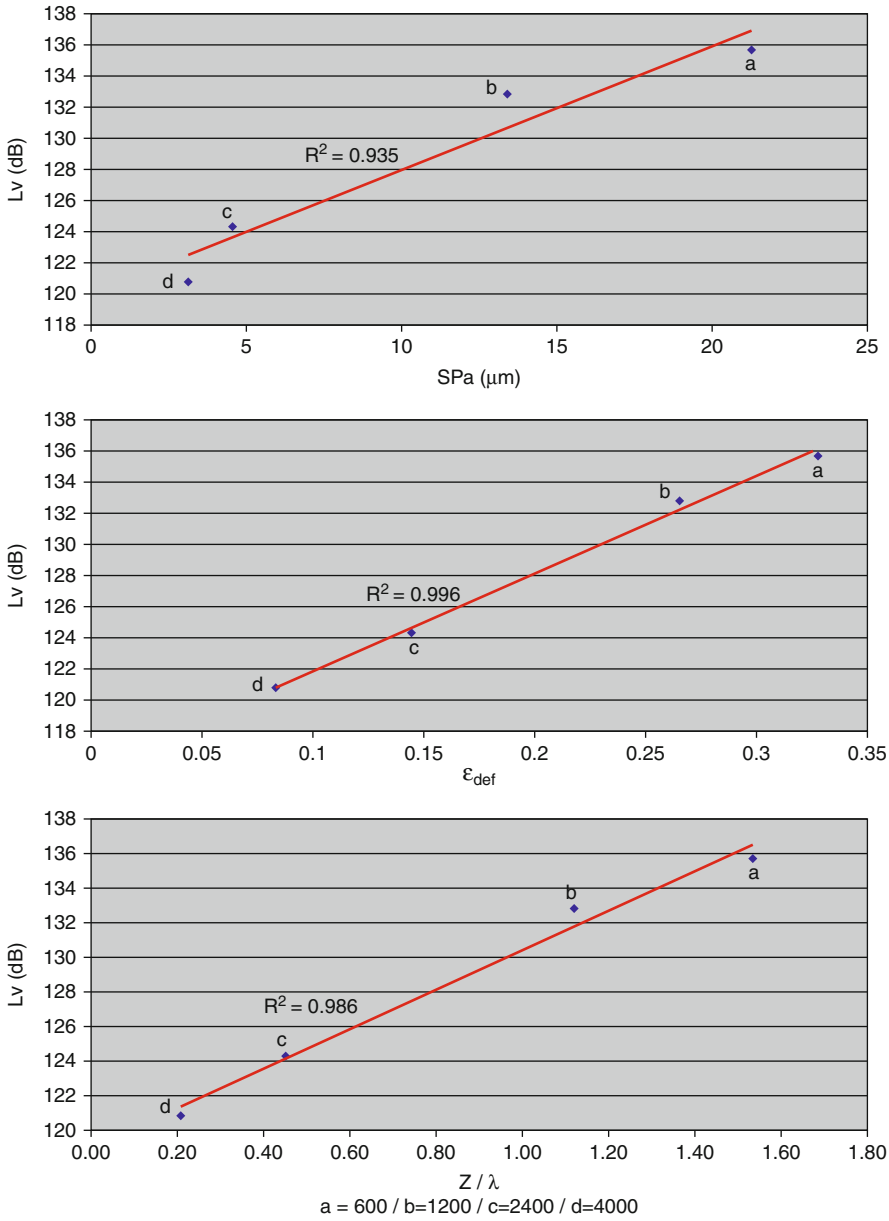


Biotribology of Tactile Perception: Effect of Mechano-transduction, Fig. 7 Three-dimensional topography of abrasive paper and vibration signal transmitted to the finger: *a* = 600, *b* = 1,200, *c* = 2,400, *d* = 4,000

Key Applications

Experimental Study of Perceptible Vibrations by Human Finger During Friction

The objective of the present experimental study was to further evaluate prospective vibrotactile codes for roughness. The vibrations elicited in the skin during the exploration of abrasive papers that systematically vary in the dimensions of their surface features were recorded, and parallels were drawn between aspects of the texture-induced vibrations and their perceived roughness, as measured by means of magnitude estimation. In this experimental study, the frequency content of the textural vibrations varied with roughness parameters (Hollins and Risner 2000; Hollins et al. 2001; Lederman 1974; Miyaoka et al. 1999; Taylor and Lederman 1975).



Biotribology of Tactile Perception: Effect of Mechano-transduction, Fig. 8 Correlation between mean level of vibration (L_v *(dB)*) and roughness parameters

To characterize vibrations transmitted to the finger during a test of tactile perception, the human finger was equipped with a sensor that is sensitive to the vibration generated during a friction test. The sliding speed is that of a precise and silent displacement table (frequency of noise <20 Hz). The human finger plays the role of a resonator and the roughness of the tested material is a source of excitation (Zahouani et al 2009; Ben Abdelounis et al. 2010). When the finger is rubbed on the tested surface, vibrations are generated and excite different tactile receptors in a wide range of frequencies, depending on the roughness and stiffness of the tested surface. The experimental conditions used with the human finger are summarized as follows: normal force: 0.2 N, sliding

speed: 20 mm/s, sampling frequency: 20 KHz, sliding distance: 40 mm, vibration sensor gain: 1 v/m/s^2, and vibration sensor sensitivity is 0.5 pc/m/s^2.

To quantify the transmitted vibrations to the human finger, abrasive papers with different scales of roughness were tested. The determination of the power spectral density allowed identification of the wavebands of vibration transmitted in the finger and the mean level of vibration: L_v(dB). The vibration velocity level in decibels is defined by the following relation:

$$L_v(dB) = 10\log\left(\frac{v}{v_{ref}}\right) \tag{17}$$

with v_{ref} the reference speed, that is, $v_{ref} = 1\text{nm/s} = 10^{-9}\text{m/s}$.

Figure 7 shows the morphology of abrasive paper and vibration signal of human finger. These results show an increase in the average level of vibration depending on the scale of roughness. To identify correlations that may exist between the roughness parameters and the vibration level of the human finger, evolution of the average level vibrations to the three roughness parameters was examined: amplitude (SPa), average deformation (ε_r), and the parameter that reflects the texture density (z/λ).

Figure 8 shows a good correlation between the vibration level L_v and three roughness parameters with a correlation coefficient $R^2 = 0.935$ for the arithmetic average amplitude of the roughness: *SPa*, $R^2 = 0.966$ for the average strain, and $R^2 = 0.986$ for the parameter density of the texture. The spread of vibrational frequencies are consistent with the frequency of Pacini: 257 Hz with a vibration level $L_v = 135$ dB (for abrasive paper 600) and 382 Hz with an average vibration level $L_v = 120$ dB (for abrasive paper 4000).

These vibration levels show that surface roughness plays the role of stimuli that will produce distortions deep into the human finger that will be transformed into mechanical vibration of Pacinian receptors based on their local geometry and roughness of their distribution.

The combination of skin deformation and vibration show that in Nature, humans and other living organisms can make judgments about their physical interaction with the environment, because they possess the sense of touch. Tactile inspections are also common in the manufacturing industry, for example, in automobile body surface inspections.

Real-time rendering of the haptic is more complex than for the visual. Indeed, thanks to the persistence of vision, the feeling of continuity takes place in lower refresh rates (around 25 Hz). For touch, the human being is able to feel vibrations below a refresh rate of 10 kHz, and feels a force varying from 30 to 300 Hz. Motion perception is slower: 1 Hz if the movement is unexpected, 10 Hz for the reflexes. With regard to temperature changes, the man is sensitive to variations in the order of 0.01 °C with a reaction time of between 300 and 900 ms. Developing such measurement capability has implications for a wide field of applications: factory floor quality control, medical diagnostics, tissue engineering, development of new types of prosthesis, and for humanoid and human-interacting robots. At the core of the problem is the understanding of the biomechanics of touch and friction, which involves mechanical interaction at the multiple scales: nano and micro, at the tips of surface peaks, and macro, at the level of the device (or organ) embedding the sensors.

Cross-References

► Friction, Human Body: Skin

References

H. Ben Abdelounis, A. Le Bot, J. Perret-Liaudet, H. Zahouani, An experimental study on roughness noise of dry rough flat surfaces. Wear **268**(1–2), 335–345 (2010)

D.T. Blake, S.S. Hsiao, K.O. Johnson, Neural coding mechanisms in tactile pattern recognition: the relative contributions of slowly and rapidly adapting mechanoreceptors to perceived roughness. J. Neurosci. **17**, 7480–7489 (1997)

S.J. Bolanowski, G.A. Gescheider, R.T. Verrillo, C.M. Checkosky, Four channels mediate the mechanical aspects of touch. J. Acoust. Soc. Am. **84**, 1680–1694 (1988)

S.J. Bolanowski, J.J. Zwislocki, Intensity and frequency characteristics of pacinian corpuscles. I. Action potentials. J. Neurophysiol. **51**, 793–811 (1984)

J. Boussinesq, *Application des potentiels à l'étude de l'équilibre et du mouvement des solides élastiques* (Gauthier –Villars, Paris, 1885), pp. 45–108

C.J. Cascio, K. Sathian, Temporal cues contribute to tactile perception of roughness. J. Neurosci. **21**, 5289–5296 (2001)

J.A. Greenwood, D. Tabor, The friction of hard sliders on lubricated rubber: the importance of deformation loss. Proc. Phys. Soc. **71**, 989–1001 (1957)

M. Hollins, S.J. Bensmaïa, S. Washburn, Vibrotactile adaptation impairs discrimination of fine, but not coarse, textures. Somatosens. Mot. Res. **18**, 253–262 (2001)

M. Hollins, S.R. Risner, Evidence for the duplex theory of tactile texture perception. Percept. Psychophys. **62**, 695–705 (2000)

K.L. Johnson, *Contact mechanics* (Cambridge University Press, Cambridge, 1985). ISBN 0 521 34796 3

S.J. Lederman, Tactile roughness of grooved surfaces: the touching process and effects of macro- and microsurface features. Percept. Psychophys. **16**, 385–395 (1974)

E.A. Lumpkin et al., J. Biol. **191**, 237–248 (2010)

R.D. Mindlin, Compliance of elastic bodies in contact. J. Appl. Mech. **16**, 214–220 (1949)

T. Miyaoka, T. Mano, M. Ohka, Mechanisms of fine-surface-texture discrimination in human tactile sensation. J. Acoust Soc. Am. **105**, 2485–2492 (1999)

M.M. Taylor, S.J. Lederman, Tactile roughness of grooved surfaces: a model and the effect of friction. Percept. Psychophys. **17**, 23–36 (1975)

T. Yoshioka, B. Gibb, A.K. Dorsch, S.S. Hsiao, K.O. Johnson, Neural coding mechanisms underlying perceived roughness of finely textured surfaces. J. Neurosci. **21**, 6905–6916 (2001)

H. Zahouani, M. BenT Kaya, S. Mezghani, C. Pailler-Mattéi, Adhesive contact in the context of multi-asperity interaction. Comptes Rendus Mécanique **339**(7/8), 502–517 (2011)

H. Zahouani, R. Vargiolu, G. Boyer et al., Friction noise of human skin in vivo. Wear **267**(5–8), 1274–1280 (2009)

Biphasic Lubrication

Gerard A. Ateshian
Department of Mechanical Engineering, Columbia University, New York, USA

Synonyms

Lubrication by interstitial fluid pressurization; Self-pressurized hydrostatic bearing

Definition

Biphasic lubrication is a mode of lubrication produced between bearing surfaces where at least one of the two contacting materials consists of a relatively soft, deformable, porous medium. This biphasic medium, which consists of a porous-permeable solid and interstitial fluid, typically exhibits high porosity but low permeability. Biphasic lubrication is effective when the interstitial fluid pressurizes to support a significant fraction of the contact load. The friction force acting on the bearing surfaces is proportional only to the fraction of the load transmitted through solid-on-solid contact. Biphasic lubrication is not critically dependent on the existence of a fluid film between the bearing surfaces. It is also largely insensitive to surface roughness of the biphasic material. This mode of lubrication is most often encountered in biological tissues such as articular cartilage.

Scientific Fundamentals

Mechanics of Biphasic Materials

When a biphasic material contacts another material, the interstitial fluid pressurizes because of the loading-induced deformation and the resulting tendency for pores to change their volume. The magnitude and spatiotemporal distribution of this pressurization is dependent upon the loading history, contact geometry, and properties of the biphasic material. The Cauchy stress tensor $\mathbf{T}$ in a biphasic material represents the contributions of the interstitial fluid pressurization and solid matrix deformation according to

$$\mathbf{T} = -p\mathbf{I} + \mathbf{T}^e, \tag{1}$$

where p is the interstitial fluid pressure, $\mathbf{I}$ is the identity tensor, and $\mathbf{T}^e$ is the stress resulting from the solid matrix deformation and strain. When a biphasic material is loaded and its interstitial fluid pressurizes, the fluid flows from regions of high pressure to regions of low pressure according to

$$\mathbf{w} = -\mathbf{K} \cdot \text{grad}\ p \tag{2}$$

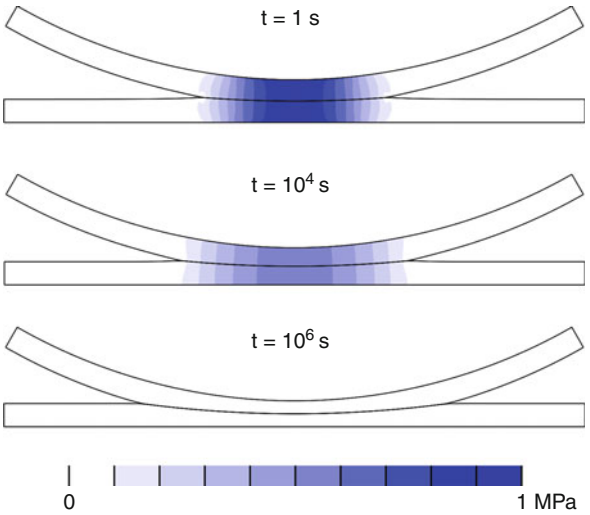
where $\mathbf{w}$ is the volumetric flux (flow rate per total area) of fluid relative to the solid matrix, grad$p$ is the spatial gradient in the pressure, and $\mathbf{K}$ is the hydraulic permeability tensor representing the resistance to interstitial fluid flow within the porous solid. This relation neglects the effects of external body forces such as gravity, which are typically negligible in biphasic lubrication.

The principal material behaviors that need to be characterized by constitutive relations and related to material properties are the dependence of $\mathbf{T}^e$ and $\mathbf{K}$ on solid matrix strain and porosity. For example, the solid matrix may be described using a compressible hyperelastic material behavior to capture pore compressibility under large deformations. The permeability constitutive relation may similarly account for the increased resistance to interstitial fluid flow as pore volume decreases. The interstitial fluid pressure p and solid matrix deformation (displacement vector) $\mathbf{u}$ are obtained by solving the balance of linear momentum and balance of mass equations for the biphasic mixture. For example, when neglecting inertia and external body forces, these balance equations reduce to

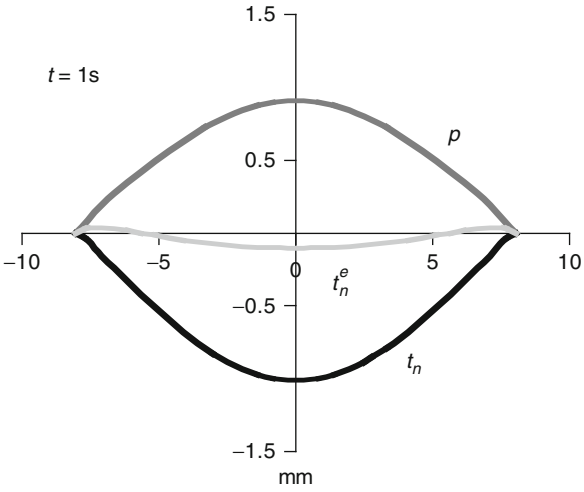
$$\begin{aligned} \text{div}\ \mathbf{T} &= \mathbf{0}, \\ \text{div}\ (\mathbf{v}^s + \mathbf{w}) &= 0, \end{aligned} \tag{3}$$

where div is the divergence operator and $\mathbf{v}^s$ is the solid matrix velocity, equal to the material time derivative of $\mathbf{u}$. Depending on the choices of material constitutive relations and load magnitudes, these governing equations may be solved analytically or numerically, subject to suitable boundary conditions.

A biphasic material exhibits a viscoelastic response because the frictional interaction between the interstitial fluid and porous solid matrix dissipates some of the free energy stored in the solid matrix during loading. Therefore, a biphasic material produces creep and stress-relaxation responses characteristic of viscoelastic materials. This flow-induced viscoelasticity may be further



Biphasic Lubrication, Fig. 1 Contour maps of interstitial fluid pressure distribution under contact loading between cylindrical and flat biphasic layers. Fluid pressurizes immediately following load application and progressively subsides with time



Biphasic Lubrication, Fig. 2 The total contact traction t_n is the superposition of the interstitial fluid pressure, p, and the traction arising from solid matrix strains, t_n^e, according to (4). Immediately upon loading, most of the contact load is supported by the pressurized interstitial fluid

compounded by intrinsic viscoelasticity of the solid matrix, when present.

Interstitial Fluid Pressurization

For biphasic lubrication problems, the most typical loading configuration consists of applying a prescribed load across opposing bearing surfaces, of which at least one consists of a biphasic material. Therefore, consider the canonical case of a cylindrical biphasic layer, of radius R and thickness h, contacting a flat biphasic layer of thickness h, under a constant prescribed load W (Fig. 1). Upon loading, the interstitial fluid immediately pressurizes. With increasing time the pressure progressively subsides until it reduces to zero at steady state. The contact area and creep deformation (normal approach) concomitantly increase.

To better understand the contribution of interstitial fluid pressurization to the load support, let $\mathbf{n}$ represent the unit outward normal vector to the surface of the biphasic layer at various points on the contact interface Γ. The contact traction on Γ is therefore given by $\mathbf{t} = \mathbf{T} \cdot \mathbf{n}$ and its normal component is $t_n = \mathbf{n} \cdot \mathbf{t} = \mathbf{n} \cdot \mathbf{T} \cdot \mathbf{n}$. According to (1), this normal traction may be split into contributions from the interstitial fluid pressure and solid matrix deformation according to

$$t_n = -p + t_n^e, \tag{4}$$

where $t_n^e = \mathbf{n} \cdot \mathbf{T}^e \cdot \mathbf{n}$. From fundamental principles governing the balance of momentum and energy across the contact interface, t_n and p (and thus t_n^e) are continuous across Γ; therefore, when two biphasic materials are in contact, the interstitial fluid pressure at the contact interface must be the same in both bearing surfaces.

A plot of the magnitudes of t_n and p shows that the interstitial fluid pressure is nearly equal to t_n immediately upon loading (Fig. 2). However, over time as interstitial fluid flows away from the loaded region, the load gets transferred to the solid matrix as the relative contribution of p subsides. The total applied load is related to the contact traction by

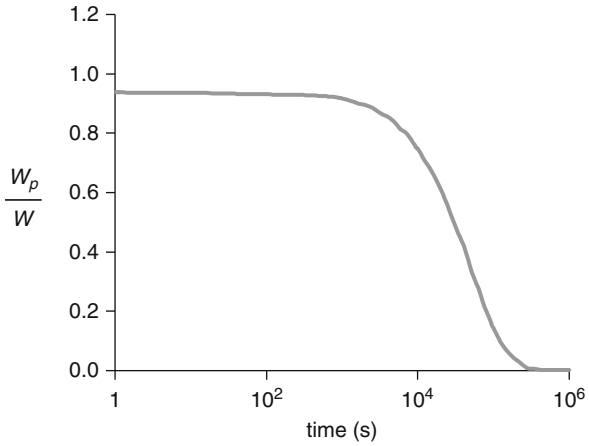
$$W = \int_\Gamma t_n d\Gamma \tag{5}$$

According to (4), the portion of this load that is supported by the interstitial fluid pressure is

$$W_p = \int_\Gamma -p d\Gamma \tag{6}$$

Therefore, the relative contribution of interstitial fluid pressurization to the total load support across the contact interface is W_p/W. For the canonical problem considered here, this interstitial fluid load support is nearly 94% upon loading but subsides to zero over time (Fig. 3).

The time constant for the subsidence of the interstitial fluid load support W_p/W is dependent on factors that



Biphasic Lubrication, Fig. 3 The interstitial fluid load support W_p/W, evaluated from (5, 6), jumps to its peak value immediately upon loading, then slowly subsides back to zero

influence the time required for fluid to flow away from the loaded region. For example, decreasing the hydraulic permeability increases the resistance to flow and delays the loss of pressurization. For an elastic deformable solid matrix, decreasing the elastic modulus reduces the magnitude of the fluid pressure gradient for a given deformation, slowing down fluid flow and pressure subsidence.

Geometric factors also play a critical role in regulating the time constant for interstitial fluid load support. All else remaining constant, greater contact congruence (larger radius R) delays loss of fluid pressurization because the resulting larger contact area increases the path length for fluid to flow from high to low pressure regions (Fig. 4a). Thinner layers (smaller thickness h) similarly constrain the flow to remain parallel to the contact interface, effectively increasing the path length by precluding flow toward deeper, low-pressure regions (Fig. 4b). These geometric factors can be further compounded by the magnitude of the applied load under finite deformation: With increasing W, the contact area increases and the layer thickness decreases, also delaying the subsidence of W_p/W (Fig. 4c).

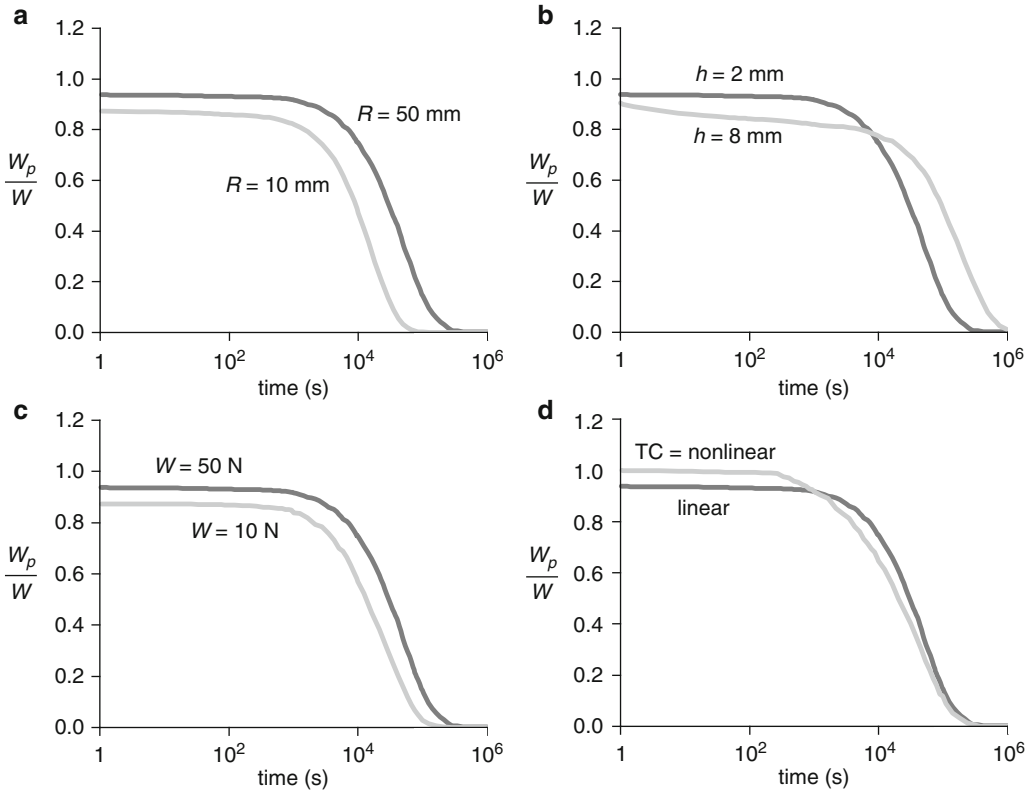
The peak magnitude of interstitial fluid load support similarly depends on the contact geometry and the material behavior. The peak value of W_p/W increases with increasing congruence (Fig. 4a) and decreasing layer thickness (Fig. 4b), for the same fundamental reasons that these factors also slow down the subsidence of fluid pressurization. Furthermore, if the solid matrix of a biphasic material exhibits greater tensile stiffness than compressive stiffness, as often encountered with fibrous materials, the peak value of interstitial fluid load support increases with an increasing ratio of tensile to compressive modulus, approaching 100% (Fig. 4d). The reason for this behavior is that a relatively higher tensile stiffness resists lateral expansion of the biphasic material upon axial compressive loading. Since biphasic materials generally exhibit a nearly isochoric response immediately upon loading, the limited lateral expansion must be compensated by increased interstitial fluid pressurization to support the applied load. However, for a given compressive modulus, increasing the tensile modulus also decreases the time constant for subsidence of interstitial fluid load support (Fig. 4d).

Stationary Versus Migrating Contact Area

Under sliding contact conditions, the temporal response of interstitial fluid pressurization is dependent on whether the contact area migrates along the surface of a biphasic material, and whether the bearing surfaces consist of a biphasic-biphasic pair or biphasic-impermeable pair. Consider the case of a biphasic-impermeable contact pair. In this configuration, the contact area may be either stationary or migrating over the biphasic surface. For example, if the impermeable surface slides horizontally with a speed V under a stationary biphasic material, the latter is being loaded over the same region. Therefore, over time, the fluid flows away from loaded to unloaded regions, allowing the pressure to subside.

However, if the contact region migrates over the biphasic material (Fig. 5a), there may not be sufficient time for interstitial fluid to escape the loaded region before the contact area moves away. In that case, interstitial fluid pressurization may not necessarily subside (Fig. 5b). The condition under which a migrating contact area may sustain elevated pressurization is achieved when the migrating velocity V is significantly greater than the characteristic velocity for interstitial fluid flow, V_c. V_c may be viewed as the characteristic diffusive velocity of fluid within the porous solid matrix, whereas V may be viewed as the convective velocity of the contact load over the biphasic bearing surface. Therefore, the ratio of these velocities represents the Peclet number for this problem, $P_e = V/V_c$. When $P_e \gg 1$, the interstitial fluid load support may be sustained indefinitely; conversely, when $P_e \ll 1$, the fluid load support subsides.

Depending on the choice of constitutive relations for the biphasic material, and the geometry of a contact problem, a theoretical expression for V_c may be formulated in terms of material properties and a characteristic dimension. For example, for a linear elastic isotropic solid matrix and a constant, isotropic permeability, the characteristic



Biphasic Lubrication, Fig. 4 Interstitial fluid load is influenced by the layer geometry, load magnitude, and ratio of tensile to compressive modulus of the solid matrix. (**a**) Effect of initial radius of curvature. (**b**) Effect of layer thickness. (**c**) Effect of load. (**d**) Effect of tension-compression nonlinearity

velocity is given by $V_c = H_A K/L$, where H_A is the aggregate modulus, K is the hydraulic permeability, and L is the contact area radius or half-width.

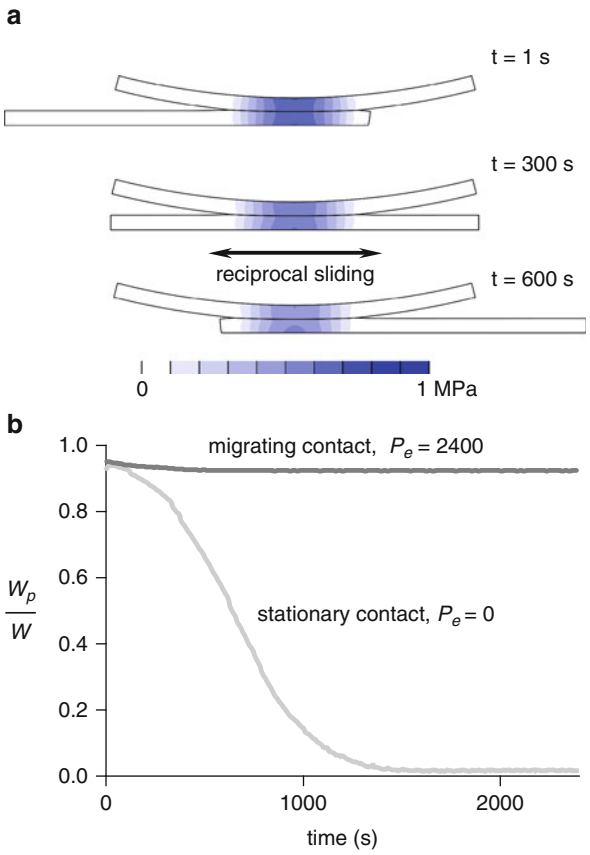
When the bearing consists of two contacting biphasic materials, it only matters whether the contact area migrates on at least one of the surfaces (Fig. 5a). Contact area migration on either bearing surface promotes sustained interstitial fluid pressurization when $P_e \gg 1$. Since the interstitial fluid pressure p must be continuous across the contact interface Γ, sustained pressurization will prevail in the opposing bearing material as well. The magnitude of the interstitial fluid load support remains dependent on all the factors listed above, including contact congruence, and the thickness and properties of both bearing materials.

Load Sharing and Friction

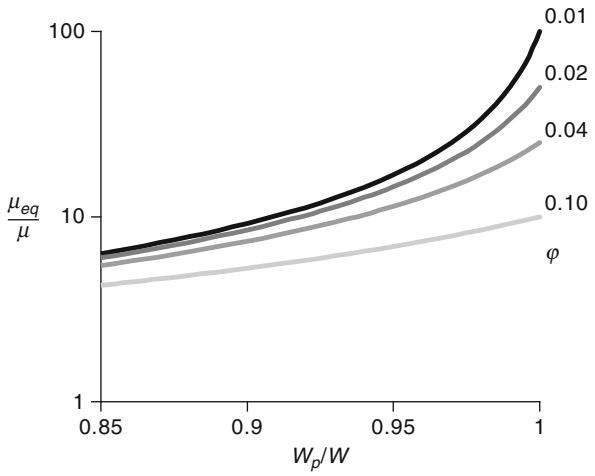
Since biphasic contact promotes load sharing between the deformed solid matrix and pressurized interstitial fluid, similar load sharing occurs with the friction force. In biphasic lubrication, it is assumed that the dominant contribution to the friction force at the contact interface results from rubbing of the solid matrix of one bearing surface against the solid matrix of the opposing surface. The viscous shear of interstitial fluid against either the solid matrix or the interstitial fluid of the opposing bearing surface is considered negligible. The solid-on-solid frictional force is assumed proportional only to the fraction of the contact load transmitted by solid-on-solid contact.

If either of the bearing surfaces is porous, the fraction of the apparent contact area over which solid contacts solid may be denoted by φ. It follows that the fraction of the contact area over which fluid contacts fluid, or fluid contacts solid, is $1 - \varphi$. Therefore, the load supported by fluid-on-fluid and fluid-on-solid contact is $(1 - \varphi)W_p$; the remainder of the applied load W is supported by solid-on-solid contact. In biphasic lubrication, the friction force F is thus given by

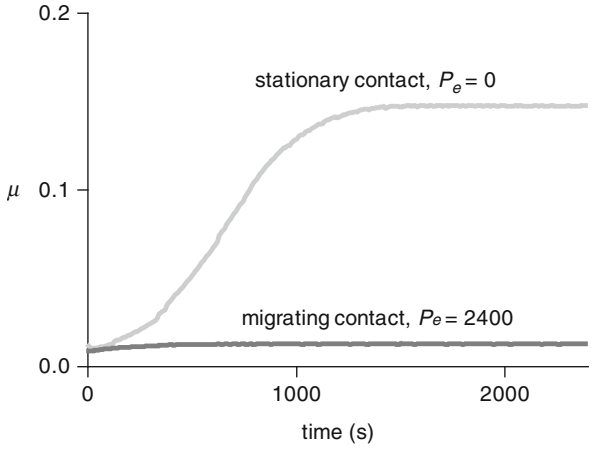
$$F = \mu_{eq}\left[W - (1 - \varphi)W_p\right], \qquad (7)$$



Biphasic Lubrication, Fig. 5 (**a**) When there is relative sliding between biphasic layers, the contact area migrates over at least one of the layers. (**b**) A migrating contact area may sustain elevated interstitial fluid pressurization far longer than a stationary contact area, as long as $P_e \gg 1$



Biphasic Lubrication, Fig. 6 The effectiveness of the biphasic lubrication mechanism is highly dependent on the peak magnitude of interstitial fluid load support, W_p/W, and the solid-to-solid contact area fraction, φ. For example, when $W_p/W \to 1$ and $\varphi = 0.01$, the friction coefficient is 100 times smaller than in the absence of interstitial fluid pressurization



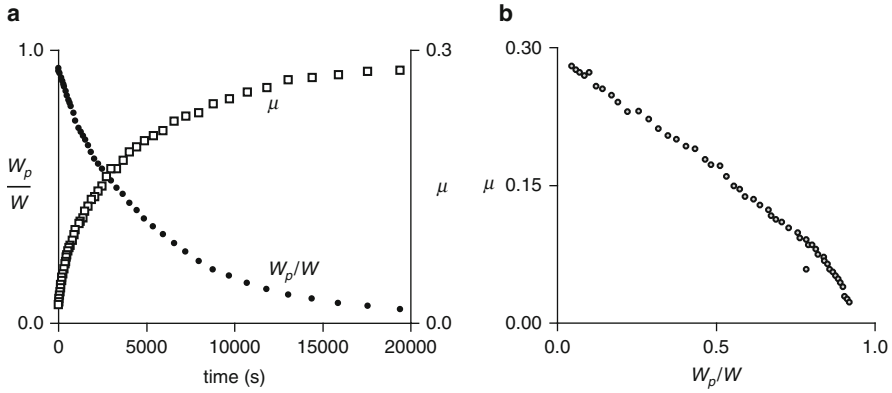
Biphasic Lubrication, Fig. 7 The biphasic friction model of (8), combined with the interstitial fluid load support shown in Fig. 5b, predicts an increasing friction coefficient under stationary contact and a sustained, low friction coefficient under migrating contact. In this illustration, $\mu_{eq} = 0.15$ and $\varphi = 0.01$

where μ_{eq} is the equilibrium friction coefficient achieved when the interstitial fluid pressure has subsided, $W_p = 0$. The friction coefficient under general conditions is defined as usual and reduces to

$$\mu = \frac{F}{W} = \mu_{eq}\left[1 - (1-\varphi)\frac{W_p}{W}\right]. \qquad (8)$$

This expression shows that the friction coefficient μ may be as low as $\varphi\,\mu_{eq}$ when the interstitial fluid load support achieves its highest possible value, $W_p/W = 1$. It increases to μ_{eq} when the interstitial fluid load support subsides completely, $W_p/W = 0$. Therefore, μ_{eq} represents the friction coefficient under boundary lubrication and its value may vary depending on environmental conditions.

The value of φ depends on the porosities of the biphasic bearing surfaces as well as their surface roughness. If the porous bearing surfaces are perfectly smooth, a statistical estimate for φ is given by $\varphi = \varphi^s_+\varphi^s_-$, where φ^s_+ and φ^s_- denote the solid area fractions of the opposing surfaces.

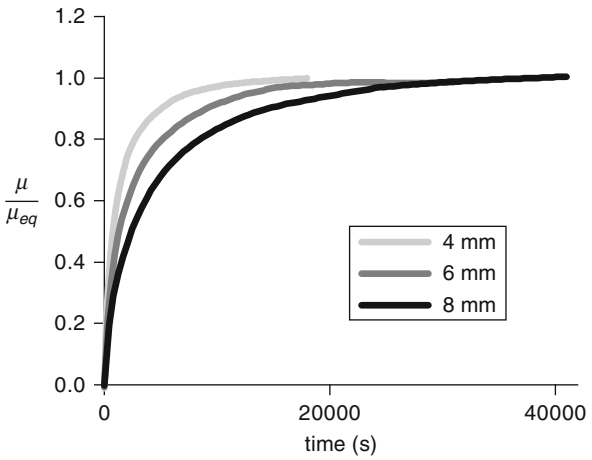


Biphasic Lubrication, Fig. 8 (**a**) Simultaneous experimental measurements of the friction coefficient and interstitial fluid load support of articular cartilage loaded against glass demonstrate that μ rises concomitantly with the decrease in W_p/W. (**b**) A plot of μ versus W_p/W exhibits the linear behavior predicted from the biphasic friction model of (8)

For example, if the biphasic surfaces each have a porosity of 90%, then $\varphi^s_+ = \varphi^s_- = 0.1$ and $\varphi = 0.01$. Therefore, biphasic lubrication may reduce the friction coefficient by a factor up to 100 relative to μ_{eq} in this example ($\mu_{eq}/\mu = 100$). If the surfaces are not perfectly smooth, fluid trapped between surface asperities at the contact interface essentially pressurizes by the same amount as the interstitial fluid of the biphasic material. In this mixed lubrication mode, the formula of (8) remains valid (under the prior assumption of negligible contribution from fluid viscosity); but the presence of asperity contact implies that the effective value of φ is smaller than $\varphi^s_+\varphi^s_-$, since solid-on-solid contact occurs at fewer locations. Therefore, in this mixed lubrication regime, the minimum friction coefficient, $\varphi\,\mu_{eq}$, may be even smaller.

The multifold reduction in friction due to interstitial fluid pressurization, μ_{eq}/μ, is significantly affected by the peak value of the interstitial fluid load support as well as the solid-to-solid contact area fraction. As W_p/W approaches unity and φ approaches zero, subtle changes in these parameters can have large effects on biphasic lubrication (Fig. 6). Therefore, the influence of geometric and material factors on interstitial fluid pressurization, outlined in the previous section, may be very significant. In particular, biphasic lubrication is enhanced with increasing congruence, load, and ratio of tensile to compressive modulus, and with decreasing layer thickness (Fig. 4).

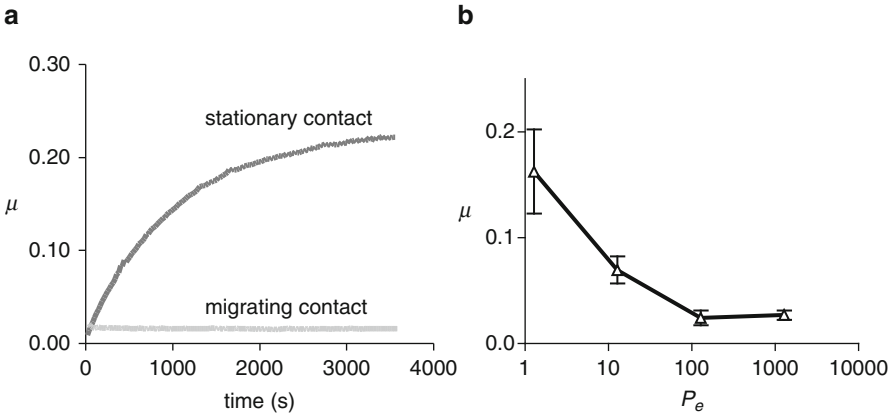
The friction coefficient predicted by (8) for configurations of stationary and migrating contact area are presented in Fig. 7. It is evident that under a constant applied load, the friction coefficient rises as the interstitial fluid load support subsides under a stationary contact area configuration. However, under a migrating contact area, the interstitial fluid load support remains elevated, sustaining a low friction coefficient nearly indefinitely.



Biphasic Lubrication, Fig. 9 Under identical loading conditions of cylindrical cartilage plugs sliding against glass, the friction coefficient increases more slowly with larger diameter plugs because interstitial fluid must travel a longer path to escape from the tissue

Key Applications

The mechanism of biphasic lubrication was first recognized in relation to the lubrication of articular cartilage layers. Articular cartilage is the bearing material of diarthrodial joints, such as the knee, hip and shoulder. It is a connective tissue whose extracellular matrix consists primarily of type II collagen and proteoglycans, and cartilage cells (chondrocytes) that occupy 5–10% of the tissue



Biphasic Lubrication, Fig. 10 (**a**) Experimental friction measurements in articular cartilage confirm that μ rises to an equilibrium value under a stationary contact configuration, but remains low and nearly constant under a migrating contact configuration. (**b**) Experiments also confirm that the magnitude of μ remains low only when $P_e \gg 1$

volume. The interstitial water content of articular cartilage ranges from 65% to 90% of the tissue weight, with the highest water content found near the articular surface. The thickness of articular cartilage ranges from 0.5 to 6 mm in human joints.

Despite its elevated porosity, articular cartilage exhibits a very low hydraulic permeability, indicative of very high resistance to interstitial fluid flow under loading. Proteoglycan molecules contribute significantly to the compressive modulus of articular cartilage, via mechanisms that include Donnan osmotic pressurization arising from the charged nature of these molecules. Because of its fibrillar nature, the collagen matrix of cartilage has a much higher stiffness in tension than in compression, especially near the articular surface. Therefore, the structure of articular cartilage appears optimized to produce very elevated interstitial fluid load support, thereby enabling its function as an effective biphasic bearing material.

Under laboratory testing conditions using a stationary contact area configuration, the friction coefficient of articular cartilage has been shown to increase over time, concomitantly with a decrease in interstitial fluid load support W_p/W (Fig. 8a); when plotting μ versus W_p/W, the relationship is nearly linear, as predicted by (8) (Fig. 8b). The time constant for the increase in friction coefficient has been shown to correlate with the size of the contact area (Fig. 9), consistent with the notion that loss of fluid pressurization is related to the distance required for interstitial fluid to travel from high to low pressure regions.

The characteristic velocity of interstitial fluid flow in articular cartilage is estimated to be on the order of 0.5–5 μm/s. Therefore, sliding velocities on the order of 50 μm/s or greater correspond to $P_e \gg 1$ and produce sustained elevated interstitial fluid load support. Experimental friction measurements of articular cartilage under a migrating contact area configuration show that the friction coefficient remains consistently low under a range of physiologically relevant reciprocating sliding velocities (Fig. 10a). When sub-physiological sliding velocities are used ($P_e \approx 1$), the friction coefficient is observed to increase (Fig. 10b).

Articular cartilage lubrication is not limited to the biphasic lubrication mechanism. Boundary lubricants present in synovial fluid and articular cartilage may contribute significantly to reduce the value of the equilibrium friction coefficient, μ_{eq}. An examination of boundary lubrication may be facilitated by the analysis of the frictional response when interstitial fluid pressurization has subsided.

Cross-References

- ▶ Articular Cartilage as a Bearing Material – An Engineering Perspective
- ▶ Brush and Hydration Lubrication (Natural Synovial Joints)
- ▶ Tribological Design of Natural Joints, an Anatomical Perspective
- ▶ Wear of Natural Joints (Osteoarthritis)

References

G.A. Ateshian, The role of interstitial fluid pressurization in articular cartilage lubrication. J. Biomech. **42**, 1163–1176 (2009)

G.A. Ateshian, V.C. Mow, Lubrication and wear of diarthrodial joints, in *Basic Orthopaedic Biomechanics & Mechano-Biology*, ed. by V.C. Mow, R. Huiskes (Lippincott Williams & Wilkins, Philadelphia/London, 2005)

G.A. Ateshian, H. Wang, A theoretical solution for the frictionless rolling contact of cylindrical biphasic articular cartilage layers. J. Biomech. **28**, 1341–1355 (1995)
G.A. Ateshian, H. Wang, W.M. Lai, The role of interstitial fluid pressurization and surface porosities on the boundary friction of articular cartilage. J. Tribol. **120**, 241–248 (1998)
G.A. Ateshian, S. Maas, J.A. Weiss, Finite element algorithm for frictionless contact of porous permeable media under finite deformation and sliding. J. Biomech. Eng. **132**, 061006 (2010)
M. Caligaris, G.A. Ateshian, Effects of sustained interstitial fluid pressurization under migrating contact area, and boundary lubrication by synovial fluid, on cartilage friction. Osteoarthritis Cartilage **16**, 1220–1227 (2008)
H. Forster, J. Fisher, The influence of loading time and lubricant on the friction of articular cartilage. Proc. Inst. Mech. Eng. H **210**, 109–119 (1996)
H. Forster, J. Fisher, The influence of continuous sliding and subsequent surface wear on the friction of articular cartilage. Proc. Inst. Mech. Eng. H **213**, 329–345 (1999)
R. Krishnan, M. Kopacz, G.A. Ateshian, Experimental verification of the role of interstitial fluid pressurization in cartilage lubrication. J. Orthop. Res. **22**, 565–570 (2004)
C.W. McCutchen, The frictional properties of animal joints. Wear **5**, 1–17 (1962)
V.C. Mow, S.C. Kuei, W.M. Lai, C.G. Armstrong, Biphasic creep and stress relaxation of articular cartilage in compression: theory and experiments. J. Biomech. Eng. **102**, 73–84 (1980)
V.C. Mow, W.Y. Gu, F.H. Chen, Structure and function of articular cartilage and meniscus, in *Basic Orthopaedic Biomechanics & Mechano-Biology*, ed. by V.C. Mow, R. Huiskes (Lippincott Williams & Wilkins, Philadelphia;/London, 2005), pp. 181–258
A. Oloyede, N. Broom, Stress-sharing between the fluid and solid components of articular cartilage under varying rates of compression. Connect. Tissue Res. **30**, 127–141 (1993)
S.S. Pawaskar, Z.M. Jin, J. Fisher, Modelling of fluid support inside articular cartilage during sliding. Proc. Inst. Mech. Eng. [J] **221**, 165–174 (2007)
M.A. Soltz, G.A. Ateshian, Experimental verification and theoretical prediction of cartilage interstitial fluid pressurization at an impermeable contact interface in confined compression. J. Biomech. **31**, 927–934 (1998)

Bit Patterned Media

► Head Disk Interface for Patterned Media

Blend

► Lubricant Formulation

Blood Viscosity

► Tribological Phenomena in Blood Vessels

Bolt Corrosion

► Fastener Failure and Safety Related Issues

Bolt Fatigue

► Fastener Failure and Safety Related Issues

Bolt Loosening

► Fastener Failure and Safety Related Issues

Bolt Strength

► Fastener Failure and Safety Related Issues

Bolted Connections

Sayed A. Nassar, Xianjie Yang
Department of Mechanical Engineering,
Oakland University, Rochester, MI, USA

Synonyms

Bolted joints

Definition

Bolted connections are widely used in almost every mechanical and structural system due to the added flexibility of assembly and disassembly of sub-systems for inspection, replacement, and routine maintenance. A bolted connection often constitutes the weakest link in the design; in many cases, the bolted connection can be responsible for determining the overall reliability and safety of an entire system. A simple bolted connection may just include a threaded fastener (bolt/screw/stud), a nut or a tapped hole, and the parts that would be clamped together by preloading the bolt; mostly by tightening the head or the nut. Preloading the bolt in a bolted connection would allow the transfer of various service

loads through the clamped connection: either directly or through increased frictional resistance at the interface surfaces of the joint. A threaded fastener is tightened to a specific preload (tension) by using one of the commonly used control methods; namely, the torque-only method, torque-turn method, torque-to-yield method, and bolt stretch method (Bickford 1997; Bickford and Nassar 1998). The joint should always be designed and assembled such that the clamp load is never overcome by the forces exerted during the operation of the system. More complicated bolted joints may include gaskets, dissimilar metallic and non-metallic materials, and complicated geometries. Service loads may also be complicated by thermal and mechanical loading, including vibrations, impact, elevated temperature, and thermal cycling. Service conditions may include corrosive, chemical, or radiation environments. In some bio-medical applications, the properties of clamped parts may even change with the age, health condition, or even the food diet of the patient as in the case of the bolted connections used to treat and/or stabilize spine, neck, knee, and hip injuries.

The reliability of bolted connections is mainly determined by both the initial preload level and the rate of decay of clamp load over the life of the joint, which is signified by the residual amount of the clamp load. Hundreds of variables affect the preload level and its rate of decay; these variables are mostly related to material properties of the various components, design and analysis methods, fabrication, surface treatment, tolerances, assembly tools, and process control (Bickford and Nassar 1998).

Scientific Fundamentals

Assembly of Bolted Joints

The safety, reliability, and quality of bolted connections are significantly affected by both the level and the stability of the fastener tension, which is most commonly achieved by the turning of either the bolt head or the nut. The torque-tension relationship in threaded fastener applications is highly sensitive to the friction variations. Moderate friction variations that are particularly common in industrial applications would have a significant impact on both the level and the stability of the clamp force in bolted assemblies (Bickford and Nassar 1998). The frictional torque components depend on the bearing and thread friction coefficients, the thread and fastener geometry, and the fastener tension.

1. Torque-only control bolt tension

   In order to enhance bolted joint reliability, design engineers must use tightening processes that would consistently achieve the desired clamp load with minimum scatter. The torque-tension relationship is often simplified by using a tabulated constant known as the nut factor as follows

$$T = KDF \tag{1}$$

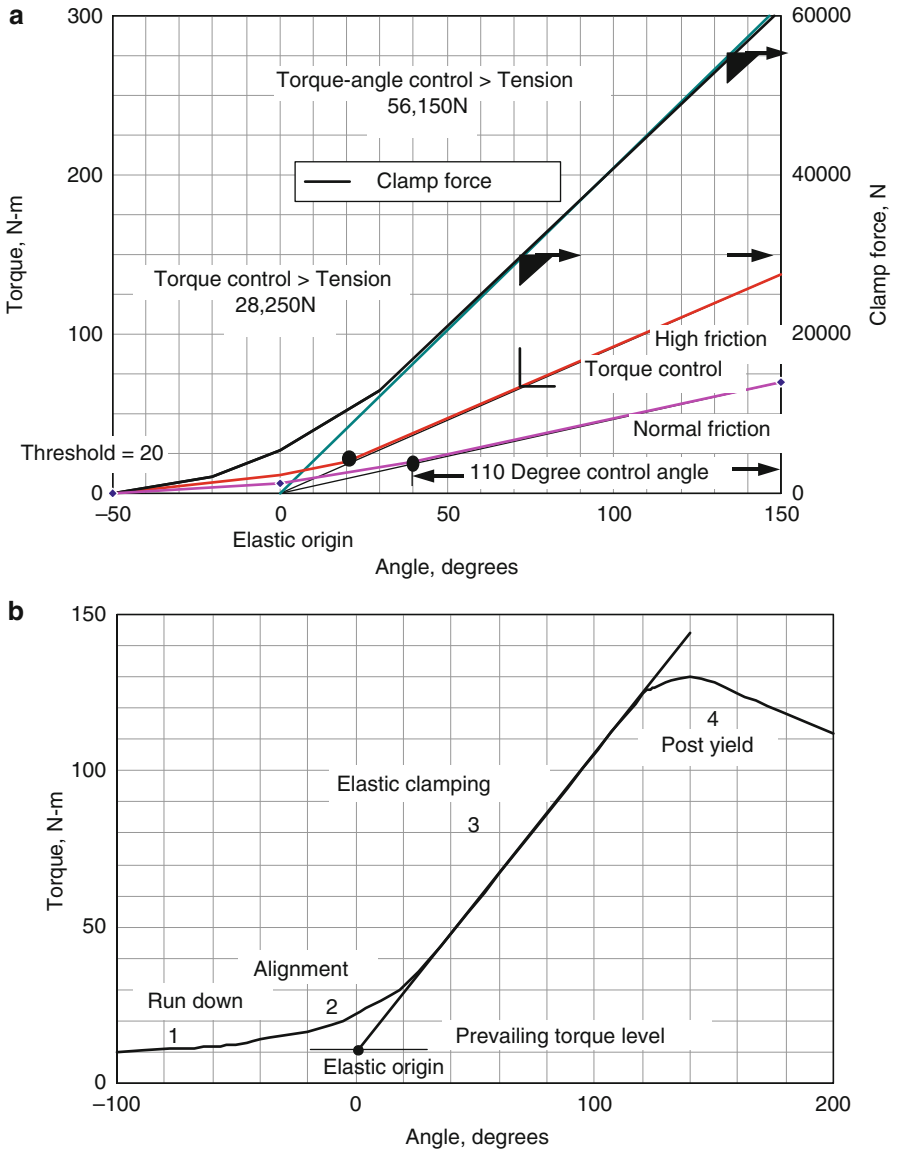
   where T is the tightening torque, D is the nominal bolt diameter, F is the desired bolt preload, and K is a dimensionless constant called the nut factor. Bickford (1997) provided some mean values of the nut factor for various combinations of joint materials and surface conditions. However, the scatter in the nut factor is too large to render it reliable, particularly in critical applications. In the absence of a prevailing torque component, the Motosh model provides the torque-tension relationship as follows:

$$T \approx F\left(\mu_b r_b + \frac{p}{2\pi} + \frac{\mu_t r_t}{\cos\alpha}\right) \tag{2}$$

   where T is the tightening torque, F is clamping force, μ_b is the bearing friction coefficient between the bolt turning underhead and the clamped surface, μ_t is the threaded friction coefficient, r_b is the equivalent bearing radius, r_t is the equivalent thread radius, p is the thread pitch length, and α is half of thread profile angle. The bearing and thread friction coefficients are, in turn, dependent on the coating type and thickness on the contact surface, lubrication, roughness, hardness, material property, radially varying sliding speed, repeated use of the same parts, and so on (Nassar and Zaki 2009). Generally, 90% of the input torque T is consumed in overcoming the bearing friction and thread friction torque components, respectively, the first and third terms in (2); the remaining 10% of T is the useful part – called pitch torque component – that creates the bolt preload/joint clamp load. Even a moderate variation of the friction coefficients and the equivalent bearing and thread radius will lead to significant effect on the torque-tension relationship in (2).

2. Torque-angle control of bolt tension (Shoberg 1998)

   The fundamental tightening procedure for torque-angle tension control is defined as follows: (a) torque is applied until a specified "threshold/snug" level is attained, followed by (b) turning the head or nut by an additional pre-determined angle to achieve the desired bolt preload. A torque-angle signature is experimentally developed for the joint in order to come up with the control values; namely, the threshold torque and the additional angle from the threshold torque. The slope of the linear part of bolt



Bolted Connections, Fig. 1 Principals of torque-turn control

tension-angle of turn signature determines the angle of turn required. If the assembly tool is shut off at a specified angle of turn after the threshold torque is attained, the scatter in bolt preload is much less than that of a torque-only control method.

The curves in Fig. 1 illustrate the basic principle of torque-turn tension control. The "elastic origin" is located by projection of the tangent to the torque-angle signature curve to the zero torque or prevailing torque level. In the elastic tightening, bolt tension is proportional to the angle of turn from the elastic origin. For the determination of the tension-angle coefficient for elastic clamping, a basic assumption is that as the fastener is turned to develop tension in the joint, the fastener stretches and the clamped parts compress elastically according to the effective spring rates of the fastener and the clamped parts. The methods including ultrasonic stretch, strain gauged bolt, force washer, model calculation, and material property-yield can be used to determine the tension-angle coefficient. Once the tightening has been stopped in the elastic clamping zone, a line

tangent to the straight-line portion of the curve $T-\theta$ can be projected backward. If the angle of turn is measured from the elastic origin to the point where tightening was stopped, the tension in the fastener will be determined by using the angle of turn and the tension-angle coefficient.

3. Torque-to-yield control (Bickford 1997; Wallace 1998)

In some critical applications, where high preloads and fewer bolts are desired for weight reduction, bolts may be tightened up to their yield limit using a torque-to-yield control method. Past the threshold torque, only the constant slope of the torque-turn signature is monitored until the slope is reduced at the end of the linear zone signifying material yield; then the tightening is stopped. This control method has much smaller scatter in bolt preload because it directly uses the material property for process control regardless of the thread and bearing friction characteristics. At the yield load, the combined normal stress due to the axial elongation of the bolt and the twisting of the bolt shank by torque, the Von Mises yield criterion is simplified for a torqued bolt (Nassar et al. 2009) as follows

$$\sigma_y = \sqrt{\left(\frac{F}{\pi R^2}\right)^2 + 3\left(\frac{3T_{tp}}{2\pi R^3}\right)^2} \qquad (3)$$

where σ_y is the yield strength of the bolt material, R is the bolt equivalent radius, which is $R = 0.45D$ if the threads dominate the grip length of the bolt and $R = D$ if the shank without threads dominates the grip length of the bolt, D is nominal bolt diameter, $\sigma_y \pi R^2$ is the yield load of the bolt under a uniaxial load, and F_y is the estimated yield load under the combined effect of tensile and torsional loading on the bolt.

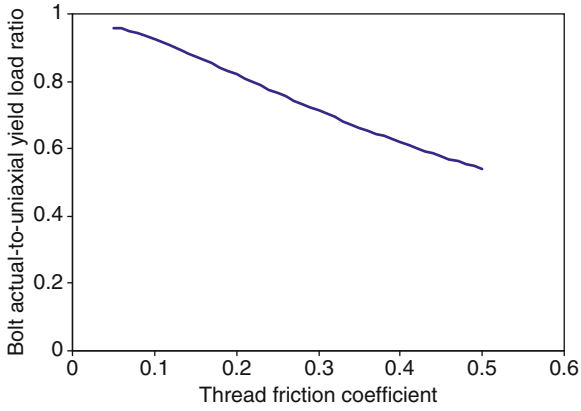
In (3) the reaction torque T_{tp} is simply the sum of the pitch and thread friction torque components expressed as follows from (2)

$$T_{tp} = \left[\frac{p}{2\pi} + \frac{\mu_t r_t}{\cos\alpha}\right] F \qquad (4)$$

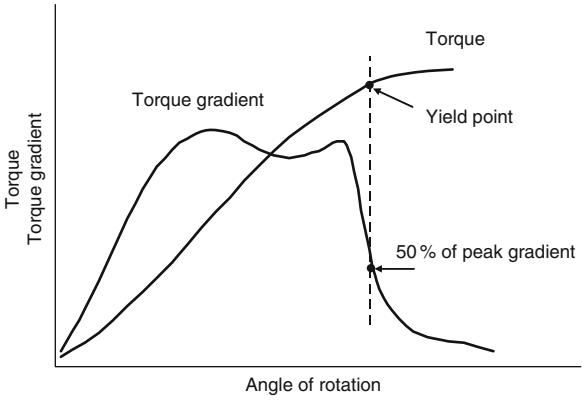
Equations (3) and (4) give the estimated yield load F_y of the bolt as follows:

$$F_y = \sigma_y \pi R^2 \Big/ \sqrt{1 + 3\left(\frac{3\left(\frac{p}{2\pi} + \frac{\mu_t r_t}{\cos\alpha}\right)}{2R}\right)^2} \qquad (5)$$

Figure 2 shows the graphical representation of (5) for the variation in the ratio of actual value of the bolt yield load, as a ratio to its nominal uniaxial yield load, as a function of the thread friction coefficient. During a torque-to-yield approach (Wallace 1998), Fig. 3 shows that after an initial transition, the torque increases linearly with rotation until yield begins, and the slope of torque-angle curve significantly decreases approaching zero. When the torque gradient drops to 50%, the corresponding point at the torque-angle curve is assumed to be the yield point. In some cases, the torque-turn signature may be used for developing a simplified version of the torque-to-yield control method by turning a predetermined angle past a low threshold torque, in order to achieve tightening the fastener to its yield limit.



Bolted Connections, Fig. 2 Variation in bolt actual-to-uniaxial yield load ratio with thread friction coefficient

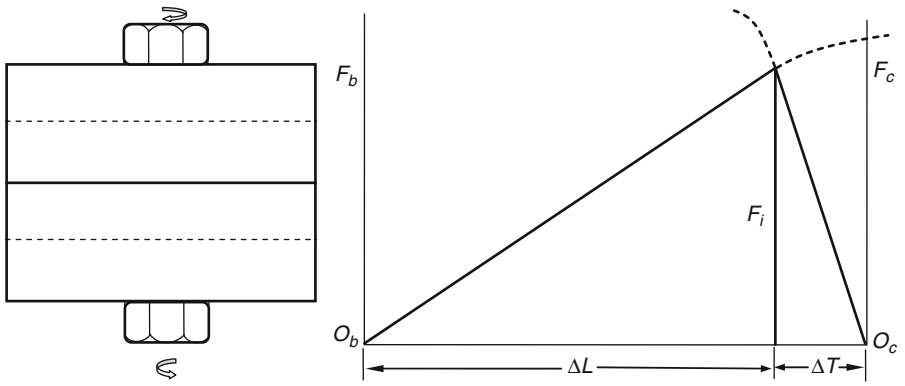


Bolted Connections, Fig. 3 The torque-angle curve and the gradient of the torque-angle curve

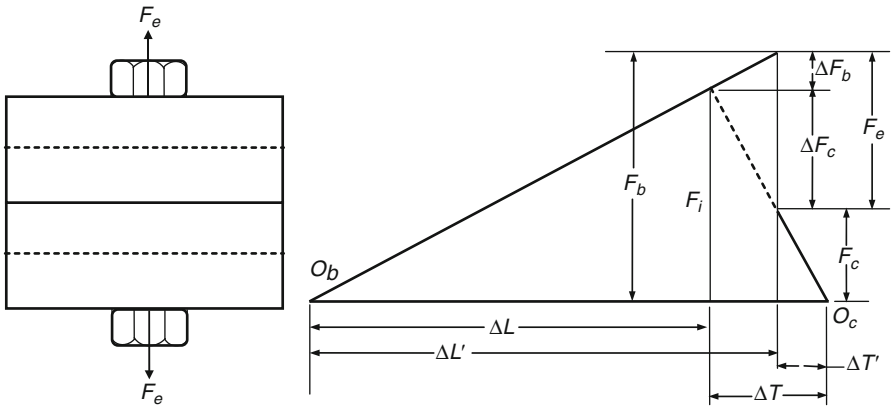
Bolted Joint in Service

1. Joint diagram during assembly

In general, the deformation behavior of the bolted joint during assembly and in service is quite complex. A joint diagram is used to illustrate the load deflection



Bolted Connections, Fig. 4 Joint diagram during assembly



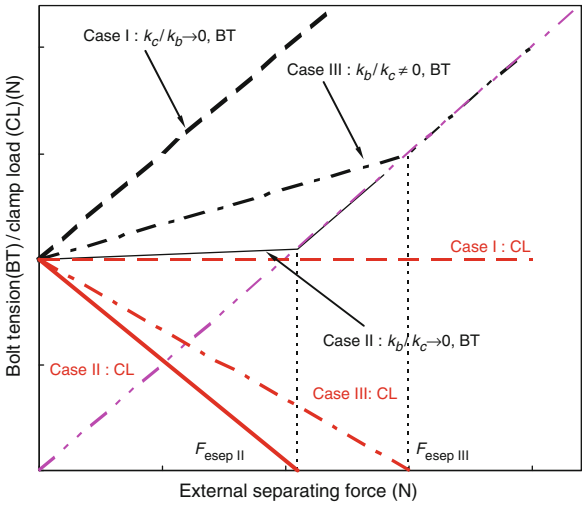
Bolted Connections, Fig. 5 Joint diagram under separating force at bolt center

characteristics of the bolt and the joint that it clamps. Joint diagrams can be used to help analyze the variations that occur when the bolt relaxes and/or external separating loads are applied to the joint-bolt system.

In the elastic range of a bolted joint system the bolt is stretched elastically as the engaged nut is rotated; this is how the bolt tension (preload) is mostly created; the equal reaction to bolt preload creation is the clamp force creation in the joint which compresses the clamped parts. The force–deflection illustration diagram in Fig. 4 shows both the bolt extension and the corresponding joint compression. The slope of the line segment represents the stiffness of the bolt, while the slope of the right segment represents the stiffness of the joint. In general, the bolt shank, bolt head, and nut can be simplified as equivalent one-dimensional springs in series and the method for the determination of the overall stiffness of a fastener is presented in the literature (Barron 1998). For the computation of the joint stiffness, however, Rötscher was the first to propose the conical envelope technique in his analytical solution with an angle α equal to 45°, as reported by Fritsche (1962). Based on the conical envelope technique, some models for the stiffness of joints can be found in the literature (Shigley et al. 2004).

2. Joint Diagram under separating force acting at bolt center

At initial assembly, the bolt preload F_b is equal to the joint clamp load F_c; both are equal to the same preload level F_i. However, when an external separating load F_e is applied to the joint along the bolt axis, as shown in Fig. 5, both the increase in bolt tension and the corresponding reduction in the joint clamp load are determined by the joint-to-bolt stiffness ratio k_c/k_b. Juvinall and Marshek (2000) give the bolt tension and joint clamp load as follows



Bolted Connections, Fig. 6 The bolt tension variation with the separating force (CL means "clamp load")

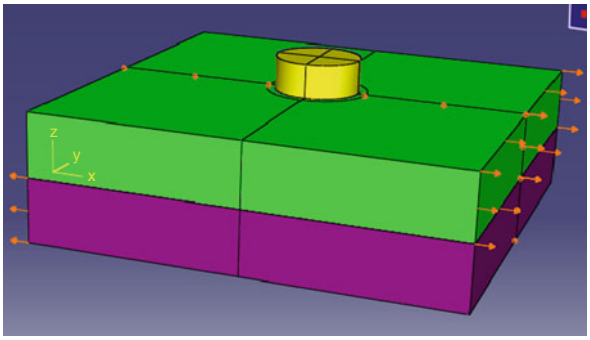
$$F_b = F_i + \left(\frac{k_b}{k_b + k_c}\right) F_e \tag{6}$$

$$F_c = F_i - \left(\frac{k_c}{k_b + k_c}\right) F_e \tag{7}$$

where F_b is the bolt tension, F_c is the joint clamp load, F_i is the initial bolt tension (preload), F_e is the external separating force, k_b is the bolt stiffness, and k_c is the joint stiffness. If the separating force reaches a critical value, the joints will open each other. The critical separating force F_{esep} is presented as

$$F_{esep} = F_i \frac{k_b + k_c}{k_b} \tag{8}$$

Figure 6 shows the bolt tension variation with the separating force under three cases: (1) very stiff bolt in a very soft joint ($k_c/k_b \to 0$); (2) very stiff joint with very soft bolt ($k_b/k_c \to 0$); (3) general case ($k_b/k_c \neq 0$). It must be pointed out that, according to Fig. 6, Case (2), the bolt load remains nearly constant with the separating load, while the clamp load is decreased by almost the full amount of the external load. Under a cyclic external load, this design scenario would improve the bolt fatigue performance by minimizing the alternating stress amplitude in the bolt. That is why head bolts in automotive internal combustion engines are normally longer with a smaller diameter, which reduces the bolt-to-joint stiffness ratio. Obviously, the trade off is that the external force, in this case, would cause the joint clamp load to drop a larger amount and must be carefully investigated in order to avoid joint gapping and potential leakage of engine combustion gases.



Bolted Connections, Fig. 7 Shear joint model

Deformation Behavior of a Shear Joint

After bolts have been tightened in a joint to their initial preload, which is equal to the initial clamp load in the joint, internal friction forces can arise from subsequent external shear forces on the joint. The friction forces are governed by the bolt tension and the interface friction coefficient. On the one hand, if the shear force is not high enough to cause the bolt shank to contact with hole wall, the maximum amount of shear force between two joint members will be proportional to the clamping force between joints and the joints interface friction coefficient. The shear stiffness of the joint can be described as the ratio of the applied external shear force to the joint movement. Two stages of loading can be defined for high strength friction grip bolted joints; they are:

1. Microslip, which takes place when the regions away from the hole experience slip while those close to the hole do not slip.
2. Macroslip, which occurs for tangential loading that results in slip over the entire contact surface.

Wu (2008) conducted a non-linear, three-dimensional finite element analysis of the simple joint model shown in Fig. 7 in order to investigate the shear behavior of single-bolt joint. The focus was on the effect of friction coefficients (under the head/nut and at joint interface), bolt preload level, joint thickness, and joint size. The following conclusions were reached:

1. The effect of friction coefficients on the non-linear shear behavior is very significant. The maximum applied shear force that may be applied increases with increased friction coefficients. The shear displacement to achieve the maximum shear force between the

clamped parts and the one that would achieve the maximum bolt/nut-to-joint shear force are different.
2. The effect of initial level of bolt preload is also very significant to the non-linear shear behavior. The maximum applied shear force linearly increases with the initial bolt preload.
3. The effect of joint thickness on the non-linear shear behavior depends on the total joint thickness.
4. There is a slight effect of the joint size on the non-linear shear behavior of the joint.

On the other hand, if the shear force is sufficiently high to cause the bolt shank to contact with the wall of the hole, the bolt shear strength or the bearing strength, shear out strength, and the net tension of the joint should be considered in the design of joints loaded in shear. If the fastener is simultaneously subjected to both shear and tensile loads, the resulting combined stresses must be calculated and compared against the total strength of the fastener. The details on the strength of the bolt and joint under shear force are given in the literature (Barrett 1998).

Bolt Fatigue

When a bolt is subjected to cyclic loading, which includes the cyclic separating force and cyclic bending caused by the cyclic transverse loading on joints, a bolt will fail in fatigue. Because fatigue failure of a bolt may lead to the function loss of the related structures and machinery or disastrous consequences, bolt fatigue becomes an important consideration in design. As with other machine components, in the evaluation of bolt fatigue life and safety design, a very important aspect is the stress analysis of a bolt, especially on the most dangerous points of the bolt like the first engaged thread root, thread runout, and the bolt head to threads shank fillet. The stress analysis can be done either by design guidelines on bolted joint design (VDI2230 1986) or the numerical simulation methods such as finite element simulation. The more complete and thorough analysis of the stress status of the bolt will yield more reliable fatigue life prediction for the bolted joints. Another important aspect is bolt fatigue strength. In fact, the bolt strength, material, and methods of thread rolling have obvious effect on the bolt fatigue strength. For most bolts in service, the stress status is the fluctuating stress with high mean stress. The mean stress-alternating stress diagram (Goodman diagram) for the infinite fatigue life of fastener is used to evaluate bolt fatigue under the service loading with the preload of the bolt. A more detailed discussion of bolt fatigue is included *in this Encyclopedia of Tribology under* "► *Fastener Failure and Safety Related Issues*."

Bolt Corrosion

All fastened joints are subject to corrosion during their service life. Corrosion will lead to irreversible damage and failure will occur if proper treatment has not been done. Different metals have different standard oxidation potential. Metals with negative potential are anodic to hydrogen while metals with positive potentials are cathodic to hydrogen. When metals of different potentials come into contact, in the presence of an electrolyte, an electrical current flows from the metal of higher potential to that of the lower potential and an electrochemical reaction occurs. The main forms of electrochemical corrosion are galvanic corrosion, stress corrosion, corrosion fatigue, fretting, crevice corrosion, erosion corrosion, and pitting and cavitation (Industrial Fasteners Institute 1988). For bolt design under the corrosive environment, the effects of the galvanic corrosion, stress corrosion, and hydrogen-assisted cracking on the reliability of the bolted joints should be taken into account. The details of these aspects are given *in this Encyclopedia of Tribology under* "► *Fastener Failure and Safety Related Issues*."

Self-loosening of Bolted Joints

Bolted joints may loosen when subjected to cyclic shear service loads. This phenomenon is frequently called vibration-induced loosening or self-loosening of threaded fasteners. The self-loosening leads to the partial loss or complete loss of the clamp load so that the function of the fastener is lost. When the clamp load is completely lost, the bolt will sustain the whole separating force fluctuation under the cyclic separating force. This may result in much easier bolt fatigue failure. If the clamp load of the gasketed bolted joints is lost, the leakage may occur. Therefore, the loss of the clamp load is one of the common failure modes of threaded fasteners. In fact, many fastener parameters have obvious effects on self-loosening under transverse excitation such as bolt length, pitch length, bearing and thread friction coefficients, bearing contact area, bolt preload, stiffness of bolt and joint, transverse cyclic force level, prevailing torque, and hole clearance thread fit. In industry, some fastener locking techniques are used to eliminate or minimize self-loosening. The details on the theoretical analysis on self-loosening and the measures on anti-self-loosening are given *in this Encyclopedia of Tribology under* "► *Fastener Failure and Safety Related Issues*."

Safety and Reliability

The structural integrity of bolted joints is dependent on failure evaluation techniques. If bolted connections lose their function, failure of the bolted connections will occur.

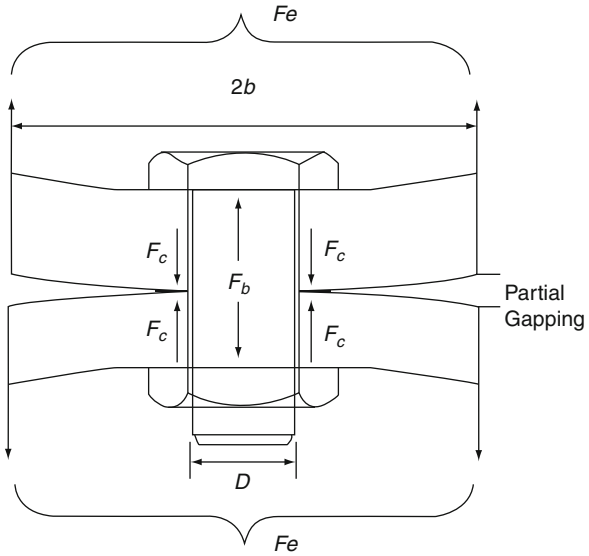
Depending on service condition and the importance of the bolted connections, the design for bolted connections should take into account one of the strengths or all of them, including static strength, fatigue, self-loosening, corrosion, and stress relaxation. Bolted joint failure and the safety of bolted joints may be affected by more than a hundred such variables, including material property, manufacture, processing of fastener and joint, assembly, in-service loading, and environment. A bolted connection must meet the following basic requirements: (1) the bolt must have adequate strength; (2) the bolted connection must have enough safety factors in service loading and environment; (3) the bolted joint must have adequate fracture and fatigue life; and (4) the bolted connection usually requires that no joint separation occur and the contact pressure should be high enough to prevent leakage. Based on different requirements of the bolted connections, different design guidelines and standards are published, such as VDI2230 (1986) and ASME Boiler and Pressure Vessel Code for Unfired Pressure Vessels (1995). To improve the safety and reliability of bolted connections, some measures are used, which are detailed *in this Encyclopedia of Tribology under* "► *Fastener Failure and Safety Related Issues*."

Key Applications

Elastic Interaction of Gasketed Multi-bolted Joint

Elastic interaction can add significant scatter to the bolt-to-bolt preload in flanged connections. This scatter is additional to that caused by variations in the frictional coefficients when bolts are tightened to the same torque value. If the fasteners are preloaded to a precise value with hydraulic tensioners or tightened to a precise preload value by monitoring bolt strains, then only preload reduction and scatter due to elastic interaction are present. For a gasketed bolted joint, clamp load loss is mainly attributed to the elastic interaction between the fasteners, gasket, and joint. Bibel and Ezell (1992) reported that elastic interaction varies greatly with the flange size, joint and gasket stiffness, number of bolts, bolt size and length, and the tightening method and sequence. They experimentally determined the "interaction coefficients" by measuring the effect of tightening various bolts on the previously tightened bolts in the same joint. To overcome the non-uniformity of the bolt preloads of the gasketed multi-bolted joint, the following strategies can be used:

1. Multiple-pass tightening strategy, in which the bolt tension at each bolt is tightened to the target preload.

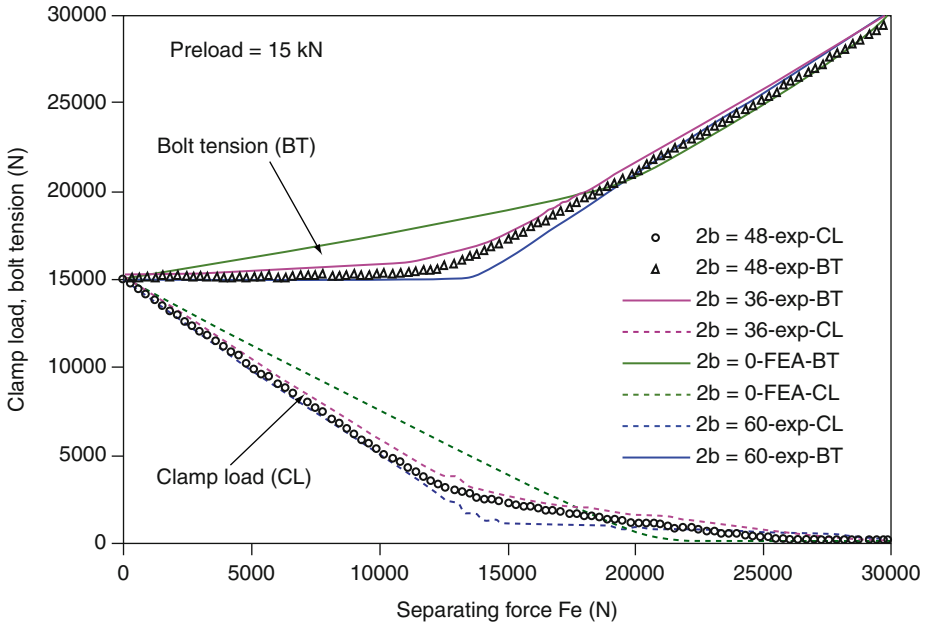


Bolted Connections, Fig. 8 Bolted joint schematic under eccentric separating force

2. One-pass tightening strategy, in which the bolt tension at each bolt is determined based on the elastic interaction factor for the elastic gasket.

Deformation Behavior of Bolted Joint Under Eccentric Separating Force

When the bolted joint is tightened, the deformation behavior of the bolted joint system may be modeled as one-dimensional springs. However, when the bolted joint is subsequently subjected to the external separating force F_e, the joint deformation behavior is quite different and will depend on the form of the separating load, its level, and its location. Even under a pure separating tensile force, the plate bending and shear force distribution in a single-bolt joint will significantly affect the changes in the bolt tension and joint clamp load in a nonlinear fashion. If the separating force is away from the bolt axis, as shown in Fig. 8, the formulation (6) and (7) for the bolt tension and joint clamp load cannot be used under this condition. From the finite element simulation under the separating force away from the bolt axis, the bolt tension increases very little when the separating force is less than 70–80% of the bolt preload while the clamp load decreases very rapidly, nearly the same as the amount of the separating force decrease from the maximum separating force, as shown in Fig. 9. The cyclic amplitude of the bolt tension is a key variable for the bolt fatigue life. When the cyclic amplitude of the bolt tension is small, the fatigue life will be high. If there is no significant opening of the two joint



Bolted Connections, Fig. 9 The variations of bolt tension and clamp load under separating force with different out-off distance from bolt axis

members, the fatigue life of the bolt will be very high under the cyclic separating force applied on the location away from bolt axis.

Joint Design

There are several published guidelines for the design of bolted connections (VDI2230 1986; ASME 1995). The German Society of Engineers published detailed guidelines for steel bolted joints called VDI2230 (1986). Since its publication, VDI2230 has been widely used in bolted joint design. In the guideline, equations based on beam theory are used as a default method for analyzing the behavior of the bolted joint under external force. A multi-bolted joint may be divided into several single-bolt segments for analysis purposes. In the guideline, the bolted joint under external force is checked against the failure criteria, which include (1) insufficient preload, (2) tensile overload of the bolt, (3) fatigue failure of the bolt, (4) yielding of the joint material under the bolt head or nut, and (5) vibration loosening of the bolt.

Even though these guidelines or standards for the design of bolted connections provided a tool for design engineers, too much simplification of the stress calculation of the bolted joints and the qualitative analysis for the self-loosening of the bolted joints cannot guaranty high enough value of the safety factor for the reliability of the bolted joints under some specific loading and environmental conditions. Depending on the applications of the particular joint, a designer must consider the specific factors. When the bolted joints are used in the critical applications such as nuclear power and aircraft, the designer must consider many more factors and make better decisions.

Cross-References

► Contact, Friction, and Wear of Threaded Fasteners

References

ASME, Boiler and Pressure Vessel Code, Section VIII, Division 1, Appendixes 2 and S (1995)

R.T. Barrett, Design of joints loaded in shear, Chapter 21, in *Handbook of Bolts and Bolted Joints*, ed. by J.H. Bickford, S. Nassar (Marcel Dekker, New York, 1998)

J. Barron, Computing the stiffness of a fastener, Chapter 11, in *Handbook of Bolts and Bolted Joints*, ed. by J.H. Bickford, S. Nassar (Marcel Dekker, New York, 1998)

G. Bibel, R. Ezell, An improved flange bolt-up procedure using experimentally determined elastic interaction coefficient. ASME J. Press. Vessel Technol. **114**, 439–443 (1992)

J.H. Bickford, *An Introduction to the Design and Analysis of Bolted Joints*, 3rd edn. (Marcel Dekker, New York, 1997)

J.H. Bickford, S. Nassar, *Handbook of Bolts and Bolted Joints* (Marcel Dekker, New York, 1998)

G. Fritsche, Grundlagen einer Genaueren Berechnung Statisch und Dynamisch Beanspruchter Schraubenverbindungen, Dissertation, Technische Universität Berlin, 1962

Industrial Fastener Institute, Fastener standards, 6th edn. (1988)

R.C. Juvinall, K.M. Marshek, *Fundamentals of Machine Component Design, Chapter 10*, 3rd edn. (Wiley, New York, 2000)
S.A. Nassar, A.M. Zaki, Effect of coating thickness on the friction coefficients and torque-tension relationship in threaded fasteners. ASME J. Tribol. **131**, 021301–021311 (2009)
S.A. Nassar, Y. Xianjie, S. Ganeshmurthy, Analysis on yield tightening of threaded fasteners. SAE Int. J. Mater. Manuf. **1**(1), 249–254 (2009)
J.E. Shigley, C.R. Mischke, R.G. Budynas, *Mechanical Engineering Design*, 7th edn. (McGraw-Hill, New York, 2004)
R.S. Shoberg, Torque-angle tension control, Chapter 32, in *Handbook of Bolts and Bolted Joints*, ed. by J.H. Bickford, S. Nassar (Marcel Dekker, New York, 1998), pp. 603–620
VDI2230 B1.1, Systematische Berechung von Hochbeanspruchten Schraubenverbidungen, Zylindrische Einschraubenverbindungen, VDI, Dusseldorf, Germany, 1986
P.W. Wallace, Increase joint performance by tightening bolts to yield, Chapter 31, in *Handbook of Bolts and Bolted Joints*, ed. by J.H. Bickford, S. Nassar (Marcel Dekker, New York, 1998), pp. 591–602
Z.J. Wu, Non-linear behavior of bolted joints under assembly and service loads, Master Degree Thesis, Oakland University, 2008

Bolted Joints

► Bolted Connections

Bonded Solid Lubrication Coatings, Process, and Applications

Jun Liang
State Key Laboratory of Solid Lubrication, Lanzhou Institute of Chemical Physics, Chinese Academy of Sciences, Lanzhou, People's Republic of China

Synonyms

Solid lubricants and applications

Definition

Bonded solid lubrication coatings are composed of one or more solid lubricants (e.g., MoS_2, graphite, PTFE) and other additives dispersed in an organic or inorganic binder, diluted with a solvent, then sprayed, dipped, or brushed onto a substrate, and cured to form a "bonded" solid film lubricant.

Scientific Fundamentals

Introduction

Bonded solid lubrication coatings came about in the late 1940s with the development of the aircraft industry. Their use accelerated in the early 1950s with the birth of the space program. Almost all of the original bonded solid lubrication coatings were composed of molybdenum disulfide (MoS_2), graphite, phenolic resin (binder), and a suitable solvent, and were cured at a temperature of +150°C. The majority of development work on bonded solid lubrication coatings occurred during the 1950s and 1960s. Since then, there have been many changes or revisions to the original products. The binders have been expanded to include inorganic components such as sodium silicates, phosphates, and ceramics. The lubricating pigments have been expanded to include polytetrafluoroethylene (PTFE), antimony trioxide (Sb_2O_3), silver, and gold. Cure temperatures have been expanded to +1,000°C or higher. In recent decades, bonded solid lubrication coating technology has grown considerably and is now applied in a wide variety of industrial, automotive, military, and aerospace applications (Clauss 1972; Campbell 1972; Fusaro 1978). Today, bonded solid lubrication coatings are an attractive alternative to fluid lubricants for minimizing friction and preventing seizing and galling, especially in harsh environments where conventional oils and greases no longer meet the performance and durability needs of advanced mechanical systems (Lancaster 1984; Gresham 1997).

Mechanisms

Bonded solid lubrication coating has inherent lubricating properties because of the presence of solid lubricants. The solid lubricants are generally composed of lamellar solids (e.g., MoS_2, WS_2, graphite), polymers (e.g., PTFE, phthalocyanine), and soft metals (e.g., In, Sn, Pb, Ag, Au, Pt, Sn) (McMurtrey 1985).

Each type of solid lubricant has different lubricating properties. Lamellar solids crystallize with a layered structure in which interatomic bonding between the layers is weaker than that within them. Thus, these materials have low shear forces between their crystalline lattice layers. When present between sliding surfaces, the crystalline lattice layers can align themselves parallel to the direction of relative motion and slide over one another with relative ease, thus providing low friction. In addition, strong interatomic bonding in each layer is thought to help reduce wear damage (Farr 1975; Erdemir 2001).

PTFEs can also provide lubrication, although they do not have a layered crystal structure. The low friction of

PTFE is attributed to the smooth molecular profile of the polymer chains, which, after orientation in early stages of sliding, can then slip easily along each other, similar to lamellar structures (Pooley and Tabor 1972).

The soft metals can provide low friction on sliding surfaces mainly because of their low shear strengths and rapid recovery as well as recrystallization. Soft metals have crystal structures with multiple slip planes and do not work-harden appreciably during sliding contact. Dislocations and point defects generated during shear deformation are rapidly nullified by the frictional heat produced during sliding contact.

When present at a sliding interface, initial smoothing of the surface of the bonded solid lubrication coating takes place as a result of the relative movement of the opposing surfaces. A partial discharge of materials from the solid film lubricant layer creates a "transfer" film on the opposite surface. The build-up of lubricating film between the two surfaces leads to a reduction in the coefficient of friction.

Characteristics

Bonded solid lubrication coatings have some unique properties compared with conventional fluid lubricants (oils and greases). The major characteristics of bonded solid lubrication coatings are as follows (Chen et al. 1994):

1. The thickness of the bonded solid lubrication coatings is normally thin (several microns). This makes it possible for them to be applied to many tribological parts and components without changing the size.
2. Bonded solid lubrication coatings can work efficiently in many harsh environments, such as high temperatures, ultra-low temperature, ultra-high vacuum, strong radiation, strong redox, and other extreme environmental conditions where conventional fluid lubricants (oils and greases) do not work properly. The load-bearing capacity of a bonded solid lubrication coating with good pressure resistance is several times or even a hundred times higher than that of oils and greases. In addition, they does not pollute the environment like conventional fluid lubricants do.
3. Bonded solid lubrication coatings neither age over time nor are flowable. This makes them suitable for lubricating mechanical parts that are started frequently and function intermittently but reliably between long periods of inactivity. In addition, bonded solid lubrication coatings can be used in combination with oils and greases to guarantee additional safety through emergency lubrication and to optimize the running-in performance of highly stressed components.
4. Solid film lubricants adhering to a workpiece not only provide good corrosion resistance and dynamic sealing performance but also prevent mechanical vibration and the occurrence of "stick-slip" and thus reduce noise.
5. In addition to lubricating metal parts, bonded solid lubrication coatings are suitable for the surface lubrication of a number of other materials (plastics, elastomers, ceramics, glass, etc.), on which oils and greases usually function with difficulty.

Classification

An enormous number of bonded solid lubrication coatings have been developed to date; some of the widely used constituents are listed in Table 1. These coatings can be categorized into several subclasses. Based on the composition of binders, the bonded solid lubrication coatings are commonly categorized into two different classes: organic bonded solid lubrication coatings and inorganic bonded solid lubrication coatings.

Organic Bonded Solid Lubrication Coatings

Organic bonded solid lubrication coatings are most comprehensively and widely used in tribological parts. More than 80% of bonded solid lubricants are organic bonded solid lubrication coatings. Commonly used organic resins include alkyd, polyurethane, polyacrylate, epoxy, phenolics, silicone, polyimide and its modified products, aromatic heterocyclic polymers, and other thermoplastics. With the combination of different solid lubricants and organic binders, bonded solid lubrication coatings with different performances have been successfully developed. In addition, special organic resins can be synthesized for use as binders to satisfy specific requirements. However, most organic resins are only stable below about 300°C. This property limits the application of organic bonded solid lubrication coatings at higher temperatures.

Inorganic Bonded Solid Lubrication Coatings

Using inorganic salts (silicate, phosphate, borate, etc.) and ceramics (silica, B_2O_3, hydrated Al_2O_3, etc.) as binders is classified as inorganic solid lubricant coating. The main advantages of inorganic solid lubricants are their wide temperature range, low vacuum outgassing, and good compatibility with liquid oxygen compared with organic bonded solid lubrication coatings. However, these coatings also have some shortcomings, including high brittleness, poor load-bearing capacity, and relatively inferior tribological properties. Therefore, most inorganic bonded

Bonded Solid Lubrication Coatings, Process, and Applications, Table 1 Typical composition of bonded solid lubrication coatings

Binders	Solid lubricants	Other components
1. Organic resins	MoS_2	Corrosion inhibitors
Room-temperature curing	WS_2	Sodium phosphite
Cellulosics	Graphite	Stannous chloride
Acrylics	$(CF_x)_n$	Lead phosphite
Heat curing	PTFE	
Alkyd	Phthalocyanine	Solvents
Phenolic		
Vinyl butyral		Water
Epoxy	CaF_2/BaF_2	Isopropyl alcohol
Silicone	PbO	Toluene
Polyimide	PbS	Amyl acetate
	Sb_2O_3	Ethyl acetate
		Naphtha
2. Inorganic salts		
	Au	
Sodium silicate	Ag	
Aluminum phosphate	In	
Sodium phosphate	Pb	
Potassium silicate		
Sodium borate		
Titanates		
CaF_2/BaF_2		
3. Inorganic ceramics		
Silica		
B_2O_3		
Hydrated Al_2O_3		

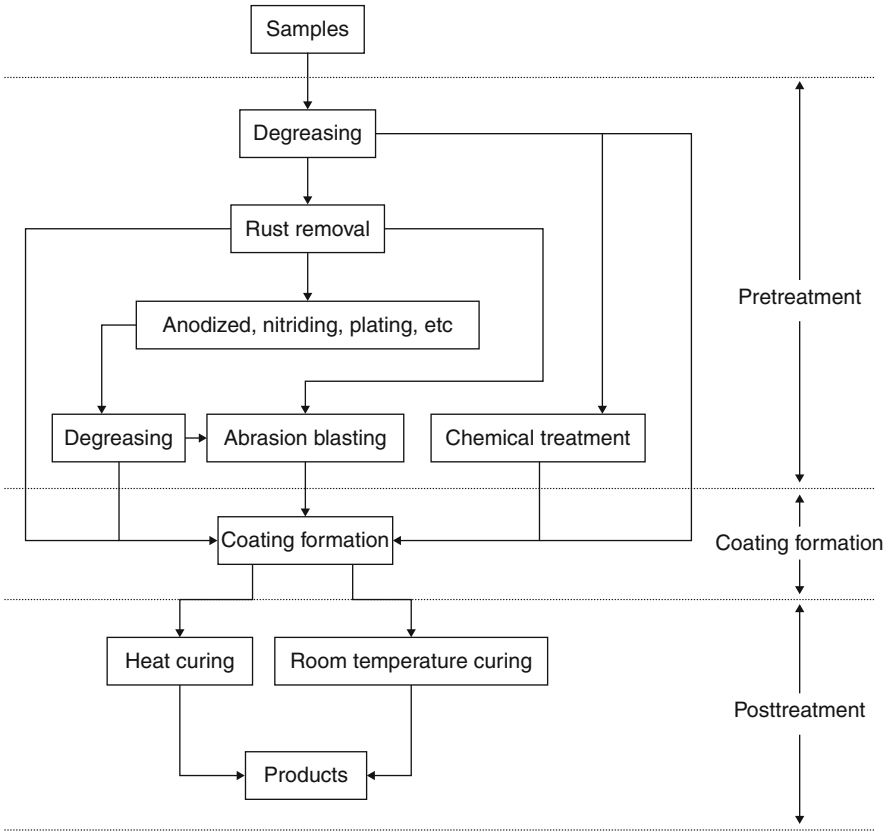
solid lubricant films are currently confined to use in several specific conditions, such as in a liquid oxygen environment, extra-high temperatures, and organic vapor-free satellite machineries.

In addition to organic and inorganic bonded solid lubrication coatings, bonded solid lubrication coatings using organometallic compounds as binders have also been developed. The application performance of this type of bonded coatings is between the organic and inorganic bonded solid lubrication coatings.

Preparation Process

Generally, there are three steps for the preparation process of bonded solid lubricant coatings: pretreatment, coating formation, and post-treatment. The scheme of the preparation process is shown in Fig. 1 (Chen et al. 1994).

Pretreatment of the parts to be coated plays an important role during the preparation process of bonded solid lubrication coating. Good adhesion and, therefore, long lifetime can only be achieved with a pretreatment that is customized for the particular part and solid film lubricant. For non-metallic parts (plastics and ceramics), degreasing or degreasing followed by blasting is the most common pretreatment process. Chemical treatments are also used in some cases. For metals, rust and scale must be chemically removed through pickling in acids or lyes. Further treatments, such as anodizing, plating, nitriding, chemical oxidation, vulcanizing, and phosphating are sometimes



Bonded Solid Lubrication Coatings, Process, and Applications, Fig. 1 Scheme of the preparation process of bonded solid lubrication coatings

necessary to improve adhesion of the solid film lubricant onto the metal surface.

In the pretreatment process, thorough removal of all grease residues, dust, dirt, rust, and scale is a fundamental step in the pretreatment of the workpiece surface. The quality of degreasing significantly affects further pretreatment processes and finally the adhesion between solid lubrication coating and part. The most common degreasing methods include alkali cleaning, pickling, solvent cleaning, emulsion cleaning, and electrolytic cleaning. One or two of these methods can be applied for the specific materials. For many technical standards of preparation of bonded solid lubrication coatings (for example, the US military standard "MIL-L-8937D"), strict regulation of the procedure for degreasing have been created due to its critical effects on final performance (MIL-L-8937D 1982).

The coating formation process of the bonded solid lubrication coatings is similar to the traditional painting process. Spraying, dipping, and brushing are most widely used in the coating formation processes. Electrochemical deposition, powder metallurgy, and other methods can also be applied. In all these methods, dipping is widely used because of its low cost, but spraying is usually the most consistent method. As a result, spraying techniques have been rapidly developed. Many new spraying devices, such as automatic sprayers and painting robots, have been developed to improve further the quality of bonded solid lubrication coatings.

Post-treatment processing of bonded solid lubrication coating includes curing and finishing. Different curing processes, such as room temperature curing, heating curing, light curing, electron beam curing, and water curing, can be used according to the components of the bonded system. In general, overall performance of bonded solid lubrication coatings prepared by heat curing is better than that produced using room-temperature curing.

Key Applications

Bonded solid lubrication coatings are widely used in the following circumstances due to their unique characteristics (Sliney 1993; Miyoshi 1996):

1. Components for which high- or low-temperature lubricants are needed. The application temperature of oils and greases is normally less than 300°C. As illustrated above, one of the major characteristics of the bonded solid lubrication coating is its wide temperature range. As a result, bonded solid lubrication coatings have been widely used to lubricate the sliding parts working in high-temperature conditions, such as rocket, plane, and automobile engines, afterburner and anti-thrust systems in planes, bores of long-range cannons, molds for metal thermal processing, some components in atomic reactors, and high-temperature ablation-resistant screws (Sliney 1993). Typical application examples at low temperatures include the turbo-pump gear of hydrogen-oxygen rocket engines and sliding parts of superconducting devices.
2. Parts that may be stored for long periods of time. Many devices, for example, ejection seats, canopy links, and joysticks in aircrafts, are usually inactive for long periods and are not used frequently. In such cases, bonded solid lubrication coatings can ensure that such devices function reliably and flexibly in an emergency.
3. When sealing is desired, bonded solid lubrication coatings between tribological couples (for example, piston and cylinder) not only provide lubrication but also act as seals to prevent leakage after running-in.
4. When fretting and galling is a problem. In the assembly clearances for some mechanical joint surfaces, such as splines, universal joints, and keyed bearings, fretting and galling can be caused by vibration if lubrication is not sufficient. The use of solid film lubricant can prevent abrasion and prolong service life (Xu et al. 2004).
5. Parts that are subject to frequent disassembly. Some parts, such as trapezoidal nuts, lead screws, and injection molds of PVC are subject to frequent disassembly. Wear and tear on these parts can be effectively avoided when they are pre-coated with solid film lubricants.
6. Parts that will be operated in corrosive atmospheres. The solid lubrication coatings usually provide excellent corrosion protection to the substrate in addition to low friction and wear. In fact, the corrosion resistance of some bonded solid lubrication coatings is similar to some anti-corrosion paints. Many bonded solid lubrication coatings also contain special rust inhibitors that offer exceptional corrosion protection in harsh environments. As an example, rusting of the screws and connectors in chemical machinery can be prevented under heating, oxidation, and corrosion conditions when rust inhibitors containing solid lubrication coating are applied.
7. Where operating pressures exceed the load-bearing capacities of oils and greases. The load-bearing capacity of bonded solid lubricant coatings is much higher than that of oils and greases. Lubrication under high load conditions, for example, the turbine/worm components of torpedo steering rods, bearing pedestals of bridges and flyovers, machine tool chucks, and cold working molds of metals, can be achieved using bonded solid lubrication coatings.
8. Other special circumstances. For example, bonded solid lubrication coatings have been widely used to lubricate devices subject to radiation, vacuum, and other extreme environmental conditions in space programs, such as drive systems of satellite antennas, solar panel systems, drive and temperature-controlled systems of optical instruments, satellite and rocket separation systems, and satellite-carried machineries. Application of bonded solid lubricant coatings in household vacuum machines has also increased (Hiraoka 2001).

Cross-References

- ▶ Graphite Solid Lubrication Materials
- ▶ High-Temperature Solid Lubricating Materials
- ▶ Solid Lubricant: Soft Metal
- ▶ Solid Lubricants
- ▶ Solid Lubricants for Space Mechanisms
- ▶ Solid Lubricants, Layered-Hexagonal Transition Metal Dichalcogenides
- ▶ Solid Lubricants, Polymer-Based Self-Lubricating Materials

References

M.E. Campbell, *Solid Lubricants: A Survey* (Technology Utilization Office, NASA SP-5059, Washington, 1972)

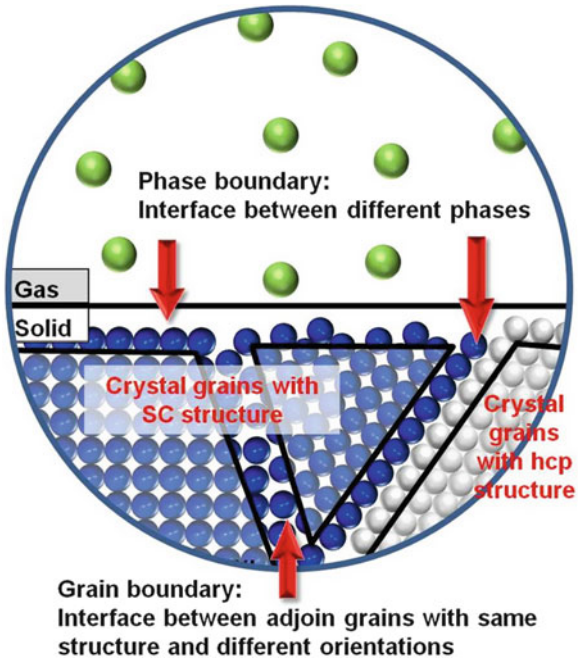
J.M. Chen, Y.P. Ye, H.X. Dang, Developing situation and application of bonded solid lubricant films. Tribology **14**(2), 180–189 (1994)

F.J. Clauss, *Solid Lubricants and Self-Lubricating Solids* (Academic, New York, 1972), pp. 1–260

A. Erdemir, Solid lubricants and self-lubricating films, in *Modern Tribology Handbook, Vol. I, Principles of Tribology*, ed. by B. Bhushan (CRC Press, Boca Raton, 2001), pp. 787–818

L.P.G. Farr, Molybdenum disulfide in lubrication: a review. Wear **35**, 1–22 (1975)

R.L. Fusaro, *Lubrication and Failure Mechanisms of Molybdenum Disulfide Films. Part I. Effect of Atmosphere* (NASA TP-1343, Washington, 1978)
R.M. Gresham, Bonded solid film lubricants, in *Tribology Data Handbook: An Excellent Friction, Lubrication, and Wear Resource*, ed. by E.R. Booser (CRC Press, Boca Raton, 1997), p. 600
N. Hiraoka, Wear life mechanism of journal bearings with bonded MoS_2 film lubricants in air and vacuum. Wear **249**, 1014–1020 (2001)
J.K. Lancaster, Solid lubricants, in *CRC Handbook of Lubrication: Theory and Practice of Tribology, Vol. II, Theory and Design*, ed. by E.R. Booser (CRC Press, Boca Raton, 1984), pp. 269–290
E.L. McMurtrey, *Lubrication Handbook for the Space Industry. Part A. Solid Lubricants.* (NASA TM-86556, George C. Marshall Space Fight Center 1985)
MIL-L-8937D, Military Specification Lubricant, Solid Film, Heat Cured, Corrosion Inhibiting, NATO Code Number S-1738 (29 Mar 1982).
K. Miyoshi, *Solid Lubrication Fundamentals and Applications* (NASA TM-107249, Washington, 1996)
C.M. Pooley, D. Tabor, Friction and molecular structure: the behavior of some thermoplastics. Proc. R. Soc. London Ser. A **329**, 239 (1972)
H.E. Sliney, *Solid Lubricants, Metals Handbook, Vol. 18. Friction, Lubrication, and Wear Technology* (ASM-International, Metals Park, 1993), pp. 115–122
J. Xu, M.H. Zhu, Z.R. Zhou, Fretting wear behavior of PTFE-based bonded solid lubrication coatings. Thin Solid Films **457**, 320–325 (2004)



Bonding at Surfaces/Interfaces, Fig. 1 Interfaces: grain boundaries and phase boundaries

Bonding at Surfaces/Interfaces

Junyan Zhang
State Key Laboratory of Solid Lubrication, Lanzhou Institute of Chemical Physics, Chinese Academy of Sciences, Lanzhou, People's Republic of China

Synonyms

Contact angle for evaluating surfaces and interfaces; Interaction; Interface; Interface bonding; Surface and bonding

Definition

The common boundary between two different phases, or between adjoin crystal grains of different orientations, is generally called an interface (Fig. 1). According to the difference of physical state, an interface can be classified as gas–liquid, liquid-liquid, gas-solid, solid–liquid, and solid-solid. A surface can be defined as a face of a matter presenting freely without any contact with any other matter, which means that surface exists in the condition of absolute vacuum. Therefore, in the open air condition, there is no absolute surface present. Generally, people call the interface contacting with a gas, such as solid/gas or liquid/gas, a surface and the others an interface.

Interfacial bonding creates new bonds between two contacted matters, the core content of an interface reaction, which is related to the interface structure, interface bonding mechanism, and interface stability. Interface stability is directly relevant to the application and is determined by the interface structure. The interface bonding mechanism is the detailed process routine of interface bonding. The bonding mechanism of a special interface is determined by material properties and the technique process. Various techniques have different process conditions and thus different bonding mechanisms.

There are several kinds of bonds at surfaces or interfaces, including covalent bonds, hydrogen bonds, van der Waals forces, and electrostatic forces. As a form of chemical bond, a covalent bond is characterized by the sharing of electrons pairs between atoms. The hydrogen bond is the attractive interaction of a hydrogen atom in one molecule or chemical group with an electronegative atom (such as nitrogen, oxygen, and fluorine) in another molecule or chemical group, which means that hydrogen bonds can occur intra- or inter-molecularly. Van der Waals forces are interactions of dipoles (permanent, induced, and instantaneous dipoles), which exist in atoms and molecules. Electrostatic forces exist in charged particles, which can be in a molecule or between two molecules. In general, the energies of hydrogen bonds

and electrostatic forces are higher than that of van der Waals forces.

The contact angle is the angle at which a liquid/vapor interface meets a solid surface. The contact angle is specific for any given system and is determined by the interactions across the three interfaces. Most often the concept is illustrated with a small liquid droplet resting on a flat horizontal solid surface. The shape of the droplet is determined by the Young-Laplace equation, with the contact angle playing the role of a boundary condition. A contact angle is measured using a contact-angle goniometer. The contact angle is not limited to a liquid/vapor interface; it is equally applicable to the interface of two liquids.

Scientific Fundamentals

Different interfacial bonding mechanisms have been historically described by considering material types (e.g., enamel, powder metallurgy, ceramic sinter, and metal-ceramic) and have assisted with the development of techniques (e.g., heat spray, electrodeposition, and vacuum film deposition) and instruments (e.g., integrated circuits and sensors). In the early 1930s, Dietzel studied the substrate interface bonding mechanism of enamel. He proposed the battery corrosion theory that the battery corrosion effect induced interface morphology undulation and thus formed mechanical bonding (Dietzel 1953). Cole attributed the mechanical bonding to the surface undulation of filling (Simmer and Fincham 1995). O'Brien argued that, on the basis of comparing theoretical simulations and experimental results, the interface wettability force can fully explain the bonding strength (O'Brien 1973). Therefore, the interface bonding of enamel could be attributed to the molecular bond. However, mechanical bonding and the molecular bond theory did not explain the new experiment results, whether using the interface bonding strength or other effects. King attributed the interface combination to the cross-interface bond of metal atoms in metal and metal ions in ceramic (King et al. 1959). Consequently, he first proposed the chemical bond mechanism of an (or the) enamel interface.

Pask proposed an intergrade layer theory of interfacial bonding on the basis of an abundant thermodynamics research of a metal-enamel interface in the 1960s (Fundamentals of glass-to-metal bonding 1953). The basic viewpoint was that, in order to form interface chemical bonding so as to build continuous electron structures of atomic configurations at a cross-link interface, oxidation intergrade layers that keep the equilibrium between the metal and the enamel are needed. Inside the layers, metal ions formed ion bonds or covalent bonds. Simultaneously, the metal ions were bonded to other metal ions in the form of metal bonds on the other side of the layers. The interfacial intergrade layer discussed here was the result of an interfacial chemical reaction, not a deliberately added interfacial intergrade layer. Indeed, the fact is that the intergrade layer could be as thin as a single atomic layer, and the interface is chemically bonded without an intergrade layer.

In 1978, Bene and Walser proposed a dissepiment mode of an interface phase on the basis of studying the interface reaction between transition metals and silicon (Walser and Bené 1976). They claimed that the phase interface could maintain the difference of the atomic concentration, the electron concentration, and the structure; the cross-phase electrochemistry barrier could be adjusted through chemical and electrical methods.

The above two modes were drawn from many experimental results that the intergrade layer mode emphasized the interface structure and bonding mechanism, while the dissepiment mode focused on the relationship between the interface phase and interface stability. Therefore, the two modes are complementary to each other without contradiction. However, although the two modes were built upon numerous experimental results in special material systems, they still lack a direct experimental observation on the interface phase due to the limitation of the experimental condition that time. Now, TEM is one of the most powerful tools for the direct the observation of various interface phases and the analysis of interface bonding modes. More universal laws about interface phases have been proposed on the condition of admitting the validity of the above two modes.

Key Applications

1. Self-Assembled Monolayers

Self-assembled monolayers (SAMs) are ordered molecular assemblies formed by the adsorption of organic molecules on solid surfaces. Due to the change of surface capabilities, for example, wettability, biocompatibility, and chemical stability, the self-assembled monolayer has become a promising candidate for surface modification in many fields, for example, microelectromechanical systems (MEMS). The formation of chemical bonds (covalent bonds) between SAMs and material surfaces are essential to acquire of these surface properties we mentioned above.

2. Adhesion of gecko foot

Geckoes are well known for their extraordinary climbing ability. The amazing capability is due to the gecko's unique setae, with micrometer-size hierarchical structures, on their toes and the powerful adhesive force produced by these setae. It has been demonstrated

that the adhesive force is based on van der Waals force rather than capillary force because the seta pad is superhydrophobic in nature. Therefore, the climbing behavior of gecko tells us that the interface bonding is also important to the movement of some animal species.

References

P.G. de Gennes, Wetting: statics and dynamics. Rev. Mod. Phys. **57**, 827–863 (1985)

A. Dietzel, Theory of adherence of enamel on iron. Ceram. Age (1953)

H.Y. Erbil, *Surface Chemistry of Solid and Liquid Interfaces* (Wiley-Blackwell, Oxford, 2006)

Fundamentals of glass-to-metal bonding: I, Wettability of Some Group I and Group VIII metals by Sodium Silicate Glass. J. Am. Ceram. Soc. **36**, 84–89 (1953)

B.W. King, H.P. Tripp, W.H. Duckworth, Nature of adherence of porcelain enamels to metals. J. Am. Ceram. Soc. **42**, 504 (1959)

A. Nilsson, L.G.M. Pettersson, J.K. Nørskov, *Chemical Bonding at Surfaces and Interfaces* (Elsevier, Amsterdam, 2008)

W.J. O'Brien, Capillary action around dental structures. J. Dental Res. **52**, 544 (1973)

T.N. Rhodin, G. Ertl, *The Nature of the Surface Chemical Bond* (North-Holland, Amsterdam, 1979)

J.P. Simmer, A.G. Fincham, Molecular mechanisms of dental enamel formation. Crit. Rev. Oral Biol. Med. **6**, 84 (1995)

R.M. Walser, R.W. Bené, First phase nucleation in silicon-transition-metal planar interfaces. Appl. Phys. Lett. **28**, 624 (1976)

R.N. Wenzel, Surface roughness and contact angle. J. Phys. Chem. **53**(9), 1466–1467 (1949)

Boriding

Luiz Carlos Casteletti[1], Amadeu Neto Lombardi[2], George Edward Totten[3]

[1]Department of Materials Engineering, São Carlos School of Engineering, University of São Paulo, São Paulo, SP, Brazil

[2]Department of Materials Engineering, Federal University of São Carlos, São Carlos, SP, Brazil

[3]Department of Mechanical Engineering, Texas A&M University, College Station, TX, USA

Synonyms

Boronizing

Definition

Boriding is a thermochemical surface treatment, a diffusion process similar to carburizing and nitriding in that boron is diffused into a metal base. Like other treatments involving diffusion, the substrate to be treated must be in contact with a boronaceous substance in the form of a solid powder, paste, liquid, or gas and held at high temperatures (700–1,200°C). Under these conditions, boron atoms diffuse from the surface into the metal lattice, forming borides with the atoms of the substrate and alloying elements, promoting the formation of the borided layer. Fluidized bed techniques can also be used for boriding. Other means to promote boride layer deposition without using thermochemical activation are plasma vapor deposition (PVD), ion implantation, and plasma spray.

Unlike many other surface treatments, hard boride layers with attractive physical and chemical properties can be developed on most metals and engineering alloys. Exceptions include aluminum and magnesium alloys because of their low melting point and copper, which does not form stable borides. The deposited boride layers have extremely high hardness values – up to 2,000 HV for steels and up to 3,370 HV for TiB_2 – while at the same time exhibiting a low surface coefficient of friction. These outstanding attributes are credited to the bonds between boron atoms, which are mostly covalent in nature. Thus, boriding treatment of steel performs substantially better than carburizing and nitriding with the added bonus of outstanding resistance against molten metals (important in die applications), nonoxidizing dilute acids, and alkalis (Stewart 1997; Sinha 1991; Petrova and Suwattananont 2005).

Scientific Fundamentals

Mechanism of Formation of Borided Layers in Steels

In carbon steels, the boronizing treatment usually forms two boride phases: FeB (more externally) and Fe_2B. The boride layer thickness formed on the steel substrates is governed by treatment time and temperature. Boron diffusion in the FeB and Fe_2B sublayers controls the rate of layer growth as a consequence of boron diffusion perpendicular to the specimen surface. Boride layer thickness varies according a parabolic law as follows:

$$d^2 = \mathrm{k}t \qquad (1)$$

where d and t are the boride layer thickness and treatment time, respectively, and k is boron growth rate constant. The higher the boronizing temperature and/or the longer the treatment time, the thicker the boride layer (Sen et al. 2005).

Influence of Alloying Elements

Mechanical properties of the borided alloys depend on the composition and structure of the boride layers.

The saw-tooth configuration is characteristic of boride layers in pure iron, unalloyed low-carbon steels, and low-alloy steels. With the increase of alloying elements and/or carbon content in the steel, the saw-tooth morphology of the boride layer becomes less pronounced and a smooth interface is formed in high-alloy steels.

With the exception of nickel, cobalt, and manganese, alloying elements retard boride layer growth by forming a diffusion barrier that restricts diffusion of boron into the steel and increases the amount of FeB in the layer. Carbon, silicon, and aluminum are insoluble in FeB-Fe_2B and are displaced ahead of the layer during its growth. Silicon and aluminum promote the formation of brittle structures. Their presence results in the formation of a soft ferrite phase beneath the boride layer, that exhibits a low load-carrying capacity, a phenomenon referred to as the egg-shell effect.

Steels selected for boronizing should not exceed 0.8% aluminum and silicon to avoid ferrite formation. During boriding, carbon is driven from the boride layer to the matrix and forms, together with boron, borocementite $Fe_3(B,C)$ as a separate layer between Fe_2B and the matrix (Stewart 1997; Sinha 1991).

Boride Layers' Properties

Ferrous Materials

The combination of high hardness and low surface friction coefficient of the boronized coating also makes a significant contribution in combating the main wear mechanisms: adhesion, abrasion, and surface fatigue (Petrova et al. 2007). Low-carbon steel and high-strength alloy steel parts develop the phases Fe_2B and FeB, but austenitic stainless steel exhibits only the smooth FeB phase due to the high amount of alloying elements present. Depending on the alloy composition, boride layers frequently develop a saw-tooth morphology and comprise an exterior layer of orthorhombic FeB with 16.3 wt% boron, hardness of 1,900–2,100 HV, and an adjacent interior layer of body-centered tetragonal Fe_2B with 8.83 wt% boron and hardness of 1,800–2,000 HV. The typical irregular interface exhibited by the borided layer and the substrate in low alloy steels is due to the great anisotropy of the diffusion coefficient of the boron in the tetragonal lattice. As the boron atoms diffuse inward from the surface of the substrate, Fe_2B is formed first and grows according to a three step mechanism:

1. Growth starts, preferably at the surface of the substrate in contact with the boride source, forming acicular crystals of Fe_2B, with their [001] axis most likely oriented parallel to the substrate surface, forming a network of crystals with varying orientations.
2. As the surface crystals start to touch each other they are forced to grow inward by the metal lattice, maintaining the acicular shape.
3. As the layer grows, the majority of the crystals turn their axis of growth (direction [001]) perpendicular to the reaction surface to minimize the mechanical resistance of the growing adjacent crystals, which culminates in the formation of a strongly oriented Fe_2B coating along the direction [002] (Martini and Palombarini 2004).

The coefficient of expansion of Fe_2B ($2.9 \times 10^{-8}\ K^{-1}$) is less than that of iron ($5.7 \times 10^{-8}\ K^{-1}$) and hence this phase remains in compression after cooling, while that of FeB ($8.7 \times 10^{-8}\ K^{-1}$) is greater than iron or Fe_2B and therefore remains in tension. This disparity in residual stress can result in the formation of cracks in the region of the FeB-Fe_2B interface, especially when a component is subjected to thermal and/or mechanical shock (Dearnley and Bell 1985; Sinha 1991).

Although small amounts of FeB are present in most boride layers, they are not detrimental if they are not continuous. However, a continuous layer of FeB could lead to spalling. The hardness of the boride layer on plain carbon steel ranges from 1,600 to 2,000 HV and is retained up to 650°C, compared with typical hardness of 650–950 HV for carburized steel. As a result, boronized steels are very resistant to abrasive wear. Furthermore, boride layers have a reduced tendency to cold-weld, resulting in good adhesive wear properties. Boronizing increases corrosion resistance of alloy steels to hydrochloric, sulfuric, and phosphoric acids. Carburized steels can be borided and then rehardened by post-heat treatment without degrading the layer quality. Most tool steels can be rehardened after boriding, provided that the austenitizing temperature is below 1,100°C. Above this temperature, the iron boride eutectic could melt (Stewart 1997; Petrova et al. 2007).

The thickness of boronized coatings varies depending on the substrate. As an example, in a 4 h treatment at 850°C, plain low-carbon steel AISI 1018 exhibits a boron saw-tooth coating structure with an average thickness of 75.76 μm. Similar to AISI 1018, AISI 4340, a high-strength alloy steel, also demonstrated the saw-tooth coating structure with an average depth of 57.40 μm and with FeB and Fe_2B present in the layer. For AISI 304 stainless steel, a flat and thin (average depth of 10.21 μm) layer was observed. Another noted difference between the low alloy steel and 304 stainless steel was the layer composition. AISI 1018

and AISI 4340 layer formed an outer FeB phase with an inner Fe_2B phase. The 304 stainless steel exhibited only a FeB layer. Microhardness test results for the low alloy steels were in the range of 1,400–2,200 HV for AISI 1018, 1,800–2,200 HV for AISI 4340, and 1434 HV for the AISI 304 stainless steel (Stewart 1997; Petrova et al. 2007).

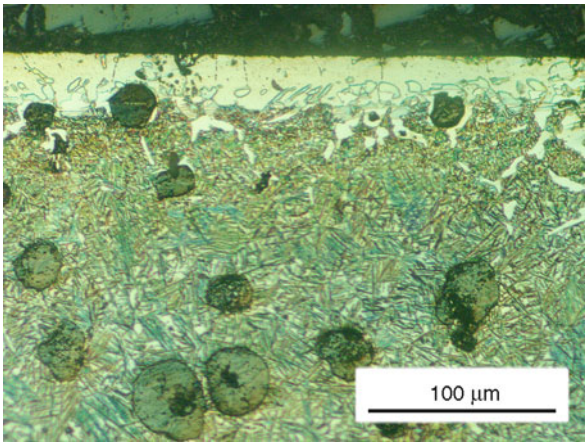
Oxidation resistance of all steels at 600°C was also improved (Petrova and Suwattananont 2005).

Nonferrous Metals

Nonferrous materials such as titanium, nickel, cobalt, and molybdenum-base alloys as well as refractory metals and their alloys and cemented carbides can be borided. Nickel and cobalt alloys are most often borided. Nickel borides range in hardness from 900 to 1,600 HV and exhibit wear and corrosion properties similar to those of iron borides. Nickel alloys are susceptible to adhesive wear. The problem is particularly acute in applications that expose components to hot acidic electrolytes, that can remove the nickel-chrome oxide surface layer. In those situations, the surfaces of nickel alloy components such as gates, balls, and seats, or a threaded linear actuator – even under moderate contact pressure – can seize, gall, or experience fretting wear. These problems are not limited to wet corrosive environments. Any process that can penetrate or remove the very thin nickel-chrome oxide surface layer will create a serious problem in applications that involve contacting components. Abrasions by particles or sliding contact are examples of situations that will break and remove the protective oxide layer. Another means of removing the protective oxide layer (by a chemical attack) is a hot reducing gas from an exhaust. Either process will lead to adhesive wear with a very high coefficient of friction. These problems have been solved, or substantially mitigated, in specific applications by the boriding process that converts a layer of the surface of the nickel alloy component to a complex nickel boride intermetallic compound (Curtis 1993).

The boride layer, which is typically 0.0125–0.05 mm thick on nickel alloys, acts as a high-integrity mechanical barrier to direct metal-to-metal contact. Borided nickel alloy components will not experience the seizure, galling, or fretting wear that is characteristic of nickel alloy components in contact in severe environments (Curtis 1993).

For nickel-based alloys such as Inconel 625 and Inconel 718, a boronized multilayer of Ni_2B, Ni_3B, and Ni_4B_3 are observed. Titanium alloys exhibit a boron-rich layer composed of a single-phase TiB. Salt bath and conventional pack boriding techniques are unsuitable for boriding titanium and refractory metals, because they cause oxidation of the substrate and activator compounds in commercial packs can cause corrosion. Both effects contribute to porosity in the boride layer. For these reasons pack boriding in oxygen-free amorphous boron (in conjunction with a vacuum or an argon atmosphere) or gas boriding with H_2-BCl_3 gas mixtures is preferred.

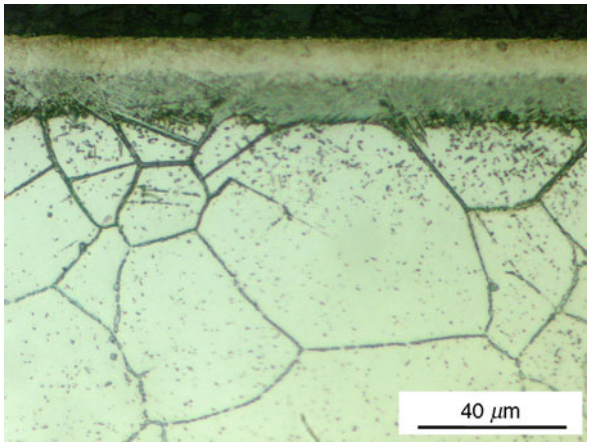


Boriding, Fig. 1 Aspect of the layer deposited by boriding treatment in austempered ductile iron, using the pack boriding process

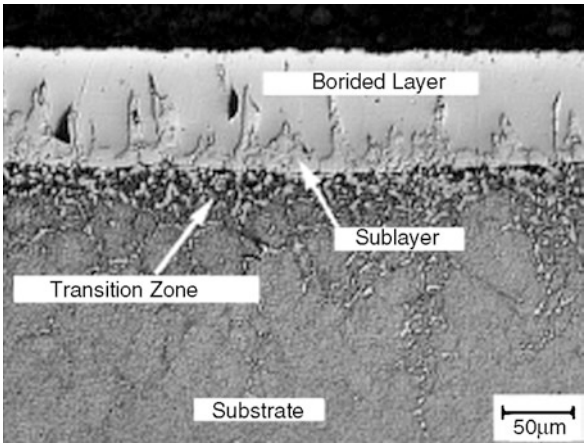
Microhardness of boride layers formed on titanium and refractory metals is very high compared with the microhardness of boride layers formed on cobalt and nickel. The addition of appropriate alloying elements increases the hardness of boride layers formed on all of these metals (Sinha 1991; Petrova et al. 2007).

Alloying additions made to nickel, cobalt, and titanium retards the rate of boride layer growth. Wear properties of sintered carbides that contain at least 6% nickel or cobalt binder can be increased by boriding because of the acceptance of boron by soft cobalt and nickel binders. These elements form borides that surround the tungsten carbide particles, increasing the composite hardness of the alloy without altering the particles. Boride layers formed on tantalum, niobium, tungsten, molybdenum, and nickel do not exhibit tooth-shaped morphology, as with titanium (Sinha 1991; Petrova et al. 2007).

Figures 1, 2, and 3 show the layers deposited by boriding treatment in the following substrates respectively: austempered ductile iron, AISI 304 stainless steel, and AISI H13 tool steel (Sinha 1991; Galibois et al. 1980; Fichtl 1981).



Boriding, Fig. 2 Aspect of the layer deposited by boriding treatment in AISI 304 stainless steel, using the pack boriding process



Boriding, Fig. 3 Aspect of the layer deposited by boriding treatment in AISI H13 tool steel, using the liquid boriding process

Boriding Techniques

Solid Phase Boriding

Pack boriding is the most widely favored boriding technique because of its relative safety and simplicity. The essential benefit is that pack boriding involves only limited equipment. The process may be conducted in practically any tempering shop that is equipped with furnaces for case-hardening operations. The technique involves embedding the components in boriding powders that contain boron carbide (B_4C), ferroboron, or amorphous boron, as the main sources of boron, together with SiC or Al_2O_3 which act as inert diluents, and an activator compound such as KBF_4, $NaBF_4$, $(NH_4)_3BF_4$, NH_4Cl, Na_2CO_3, BaF_2 or $Na_2B_4O_7$. Some materials, formulated in the Institut fur Harterei Technik at Bremen, form the basis of commercial powders such as "Ekabor" or "Durboride," which have a typical composition of 5 wt% B_4C, 5 wt% KBF_4, and 90 wt% SiC.

As the B_4C content is increased, the boriding action becomes stronger and the amount of FeB constituting the layer increases. The considerably lower proportion of the boride FeB formed by the powder technique leads to an absence of cracking due to a decrease in the residual stress (Fe_2B = compressive stress and FeB = tensile stress). This represents a major advantage of this technique over the paste technique (Sinha 1991; Galibois et al. 1980; Fichtl 1981).

Paste Boriding

This technique is commercially practiced only when pack boriding is difficult (e.g., when certain re-entrant features are not accessible to contact with a powder) or when masking off is too expensive or time consuming. Boriding pastes usually comprise a suspension of B_4C and cryolite (Na_3AlF_6) or conventional boriding powder (B_4C-KBF_4-SiC) in an organic binder, such as nitrocellulose dissolved in butyl acetate or an aqueous solution of methyl cellulose. Commercial pastes are supplied in a viscous state and are water soluble, allowing a range of consistencies to be made so that the paste may be brushed or sprayed on. After drying, another coat is applied and the procedure repeated until a layer 1–2 mm thick is obtained. Components are heated inductively, resistively, or in conventional furnaces, to 800–1,000°C and in all cases a protective atmosphere is necessary (Sinha 1991).

Liquid Phase Boriding

There are two techniques, (a) electroless and (b) electrolytic, and both are usually performed at temperatures >900°C. There are restrictions on the size (governed by bath dimension) and complexity of component shape since components must withstand thermal shock on immersion and removal from the bath or distortion and cracking can result. After treatment, excess salt and unreacted boron must be removed, which can be costly and time consuming. Maintenance costs are high for both techniques since they both require regular recharging with salt due to increase in bath viscosity caused by buildup of reaction products, that impair boriding reproducibility (Dearnley and Bell 1985).

Electroless Salt Bath Boriding

Most treatments are based on the reduction of molten borax ($Na_2B_4O_7$) at temperatures of 900–950°C that are necessary to reduce bath viscosity to a level that facilities reasonable heat and mass transfer. The reduction of borax by SiC is not used in practice because it increases the bath viscosity. Instead, additions of B_4C, typically 30 wt%, are preferred. Recent research indicates that low-cost baths containing borax with additions of ferrotitanium and aluminum can be satisfactorily used to produce high-quality layers. Boronizing action can be further improved by replacing up to 20 wt% B_4C with ferroaluminum because it is a more effective reductant. However, superior results have been found by using a salt bath mixture containing 55 wt% borax, 40–50 wt% ferroboron, and 4–5 wt% ferroaluminum. There is also the possibility of using only the addition of aluminum to borax for the formation of boride layers and, in this case, borax is not only the vehicle in which the former elements of the layer are dissolved but is also a supplier of boron, thus becoming an active agent of the process. The layer is then formed by the combination of the iron present in the substrate with the boron arising from the chemical reduction of borax by the aluminum. The action of aluminum facilitates the formation of the boride layer by the reactions Fe + B = FeB or 2Fe + B = Fe_2B (Dearnley and Bell 1985; Sinha 1991).

Electrolytic Boriding

This is the oldest commercial boriding technique and it was devised in the 1930s in the USSR. The container usually of graphite construction, contains molten borax, which is heated by resistance or an induction coil. The component, attached to the cathode, and graphite anode are simultaneously immersed in the molten borax and current is passed between them. In the fused state, borax comprises ions of sodium (Na+) and tetraborate (B_4O_7 2-), which under the influence of the electric field move towards the cathode and anode, respectively. The tetraborate ion is neutralized and reduced to boric acid and nascent oxygen is liberated:

$$B_4O_7 + 2e \rightarrow 2B_2O_3 + O$$

Oxygen reacts with the graphite anode, forming CO. Sodium ions are neutralized in the region of the cathode and react with boric acid:

$$6Na + B_2O_3 \rightarrow 3Na_2O + 2B$$

In this way, a high boriding potential is created in the vicinity of the cathode (Sinha 1991).

Gaseous Phase Boriding

Gas boriding may be conducted by thermal decomposition of diborane (B_2H_6) or by reduction of a boron halide, usually BCl_3 with hydrogen, or using organic boron compounds such as $(CH_3)3B$ and $(C_2H_5)3B$. The use of diborane is not commercially viable in view of its toxicity (the lower toxic limit is 0.1 ppm, compared with the 10 ppm allowed for the extremely dangerous hydrogen cyanide) and is also explosive, which would demand the use of complex equipment. Another major disadvantage is that, even when diluted by hydrogen, diborane is still very expensive (Sinha 1991).

Plasma Boriding

Plasma boriding can be used for ferrous and nonferrous alloys. Mixtures of gases such as BCl_3–H_2–Ar or B_2H_6–H_2 are used in a glow discharge chamber to promote the diffusion of boron into the substrate. Other gases that can be used are $B(OCH_3)_3$ and BF_3. With the BCl_3–H_2–Ar mixture, treatment can be applied at lower temperatures (600°C) when compared with other processes, saving energy and with good control over composition and depth of the borided layer. Also, the treatment times are reduced because of the increased boron potential compared with the pack process. The primary disadvantage of plasma boriding is the extreme toxicity of the gases used. Because of this, this technique has not gained commercial acceptance (Sinha 1991).

Multicomponent Boriding

These techniques involve consecutive diffusion of boron and one or more metallic elements into the surface and near-surface regions of metal substrates, usually steel. Electrolytic salt baths and paste techniques can be used, but most interest has been focused on the pack methods. Multicomponent boriding can be conducted in simultaneous or consecutive processing steps: simultaneous boriding and metalizing (diffusion of the metallic species), boriding followed by metalizing, and metalizing followed by boriding. Although attractive properties are obtained, multicomponent boriding has seen limited application in industry.

The main types of multicomponent boronizing are borochromizing, boroaluminizing, borosiliconizing, and borovanadizing, but new processes continue to be developed. Early work indicates that deposition of borochrome-titanium (borochromotitanizing) can achieve surface hardness of 4,000 HV, making it the boron treatment yielding the hardest surfaces. Multicomponent boriding produces coatings with high wear and oxidation resistance, increased corrosion, and greater fatigue strength. The best layers are obtained when

the boriding treatment is followed by the metalizing treatment. The resulting layer is most likely composed of a solid solution of iron boride in iron with the second metal deposited by the metalizing treatment (Dearnley and Bell 1985).

Key Applications

The high hardness of borided materials makes them suited to wear resistance, particularly to abrasive particles. Components used in sliding wear may have their service life extended by boriding (e.g., steel rings used on textile spinning machines, steel rope textile thread guides, and suspension bogies).

Abrasive wear resistance of borided stainless steel and Ti-6Al-4 V has recently been investigated for potential use as an erosion-resistant leading edge rotor blade cladding for helicopters. Erosion tests, using road dust, yielded a 50-fold improvement in life compared with untreated materials. A common use of boriding is for the protection of steel moulds used for the manufacture of ceramic bricks. This property has also been utilized for the protection of steel moulds and extrusion screws used for the manufacture of plastics containing abrasive fillers-pigments, such as Al_2O_3, TiO_2, B_4C, and mica. Resistance to HCl fume attack, generated by decomposition of PVC, is an added advantage. The corrosion resistance of high-alloy steel pipe elbows used for the transportation of vinyl chloride monomer, which contains traces of HCl vapor, has been quadrupled by boriding. Boride layers developed on plain carbon steel also offer resistance to dilute H_2SO_4 and H_3PO_4. However, the rolling fatigue properties of borided alloy steels are very poor when compared with carburized and nitrided steels at high contact loads (2,000 N), and it is probably for this reason that boriding of gears is restricted to those of a screw design where transverse loading of the gear teeth is minimized (Dearnley and Bell 1985).

Boriding can be used as a multifunctional coating in automotive applications such as steel gears, valves, plungers, and discs and in production machines for extruder screws, bolts, nozzles, gear drives, pumps, and tool dies (Petrova et al. 2007). The following list provides several applications of the use of boriding treatments for several types of steels and mechanical parts (Fichtl 1981)

- AISI 1020 steel – Gear drives, pump shafts
- AISI 1043 steel – Pins, guide rings, grinding disks, bolts, casting inserts, nozzles, handles
- AISI 1138 steel – Shaft protection sleeves, mandrels
- AISI 1042 steel – Swirl elements, nozzles (for oil burners), rollers, bolts, gate plates, gate plates
- AISI W1 steel – Clamping chucks, guide bars
- AISI D3 steel – Bushes, press tools, plates, mandrels, punches, dies
- AISI C2 steel – Drawing dies, ejectors, guides, insert pins, gate plates, bending dies
- AISI H11 steel – Plungers, injection cylinders, sprue
- AISI H13 steel – Orifices, ingot molds, upper and lower dies and matrices for hot forming, disks
- AISI H10 steel – Injection molding dies, fillers, upper and lower dies and matrices for hot forming
- AISI D2 steel – Threaded rollers, shaping and pressing rollers, pressing dies and matrices, engraving rollers
- AISI D6 steel – Straightening rollers
- AISI S1 steel – Press and drawing matrices, mandrels, liners, dies, necking rings
- AISI D2 steel – Drawing dies, rollers for cold mills
- AISI L6 steel – Extrusion dies, bolts, casting inserts, forging dies, drop forges, embossing dies, pressure pad and dies
- AISI 02 steel – Molds, bending dies, press tools, engraving rollers, bushes, drawing dies, guide bars, disks, piercing punches
- AISI E52100 steel – Balls, rollers, guide bars, guides parts for nonferrous metal casting equipment, parts for unmagnetizable tools (heat treatable)
- AISI 4140 steel – Press tools and dies, extruder screws, rollers, extruder barrels, non-return valves
- AISI 4150 steel – Nozzle base plates
- AISI 4317 steel – Bevel gears, screw and wheel gears, shafts, chain components
- AISI 5115 steel – Helical gear wheels, guide bars, guiding columns
- AISI 6152 steel – Thrust plates, clamping devices, valve springs, spring contacts
- AISI 302 steel – Screw cases, bushes
- AISI 316 steel – Perforated or slotted hole screens, parts for the textile and rubber industries
- AISI 410 steel – Valve components, fittings
- AISI 420 steel – Valve components, plunger rods, fittings, guides, parts for chemical plants, shafts, spindles, valves

References

D. Curtis, Case hardening nickel alloys. Mater. Design **14**(6), 349–350 (1993)

P.A. Dearnley, T. Bell, Surf. Eng. **3**, 203–217 (1985)

W. Fichtl, Boronizing and its practical applications. Mater. Eng. **2**, 276–286 (1981)

A. Galibois, O. Boutenko, B. Voyzelle, Mécanisme de Formation des Couches Borurées Sur Les Aciers a Haut. Carbone-II – Technique des Poudres. Acta Metall. **28**, 1765–1771 (1980)

C. Martini, G. Palombarini, Mechanism of thermochemical growth of iron borides on iron. J. Mater. Sci. **39**, 933–937 (2004)
R.S. Petrova, N. Suwattananont, Surface modification of ferrous alloys with boron. J. Electron. Mater. **34**(5), 575–582 (2005)
R.S. Petrova, N. Suwattananont, V. Samardzic, The effect of boronizing on metallic alloys for automotive applications. J. Mater. Eng. Perform. **31**, 1059–9495 (2007). ASM International
S. Sen, U. Sen, C. Bindal, The growth kinetics of borides formed on boronized AISI 4140. VACUUM – Surf. Eng. Surf. Instrum. Vac. Technol. **77**, 195–202 (2005)
A.K. Sinha, *Metals Handbook- Heat Treating-Boriding (boronizing) of Steels*, vol. 4 (ASM International, Materials Park, 1991), pp. 978–999
K. Stewart, Boronizing protects metals against wear. Adv. Mater. Process. **3**, 23–25 (1997)

Boronizing

► Boriding

Boundary Element Methods for Contact Analysis

Chang-Hung Kuo
Department of Civil Engineering, National Chi Nan University, Puli, Nantou, Taiwan

Synonyms

Boundary integral equation methods for contact analysis

Definition

The boundary element method for contact analysis is a numerical method of solving contact problems that have been formulated as boundary integral equations.

Scientific Fundamentals

The boundary element method (BEM) is a numerical method that transforms governing differential equations into integral identities (boundary integral equations) over the surface or boundary of a domain. Instead of full domain discretization, the boundary is divided into small elements where the unknown functions are approximated in terms of nodal values, and the integral equations are discretized and solved numerically. Since all the approximations are limited to the boundary, the BEM has better accuracy in stress calculation than other numerical methods, including the finite element method (FEM). Furthermore, reduction of dimensionality by one order also makes the BEM more efficient in three-dimensional analysis. Since its introduction in the 1960s (Rizzo 1967), the BEM has evolved into a powerful tool that has distinct advantages in various engineering applications including three-dimensional problems with rapidly changing variables, contact analysis, fracture mechanics, and semi-infinite and infinite domains.

A boundary element analysis for the contact problems includes formulation of boundary integral equations, numerical implementation for the integral equations, and numerical iteration to solve the contact area and contact pressure. The formulation of the boundary integral equations has two different types: the indirect formulation and the direct formulation. The indirect formulation employs a set of fictitious density functions as unknown functions in the integral equations. Once the density functions are solved, the physical quantities such as displacements and tractions can be computed. In the direct formulation, the unknown functions appearing in the integral equations are the actual physical variables, such as tractions and displacements. The direct BEM has been increasingly used for various applications and is best suited for contact analysis as the contact tractions and displacements are the direct variables in the formulation.

Boundary Integral Equations

Direct BEM uses either Betti's reciprocal work theorem or the weighted residual method to derive the Somigliana identities, which can be written as:

$$u_i(\xi) + \int_\Gamma T_{ji}(x,\xi)\, u_j(x) d\Gamma(x) = \int_\Gamma G_{ji}(x,\xi)\, t_j(x) d\Gamma(x) \quad \xi \in \Omega \tag{1}$$

Here, u_j and t_j are displacement and traction, respectively; $G_{ij}(x,\xi)$ and $T_{ij}(x,\xi)$ are the fundamental solutions (Green's functions) that represent displacement and traction, respectively, in the x_i -direction at point x due to a unit load applied at point ξ in the x_j-direction. For isotropic materials, the Kelvin fundamental solution is commonly used in the formulation of elasticity problems with a finite domain, such as the contact problems of two finite bodies. The Mindlin fundamental solution for elastic half-space can be used for contact problems of semi-infinite solids. For anisotropic materials, the fundamental solutions for materials with transverse isotropy (Pan and Chou 1976), layered transversely isotropic half-space (Kuo and Keer 1992), and general anisotropy (Ting and Lee 1997) may be used in the formulation of the boundary integral equations.

The Somigliana identity yields the displacements within the interior of the domain given knowledge of the distribution of both tractions and displacements on the

boundary. As the internal point goes to the boundary, the boundary integral equation can be obtained as:

$$c_{ij}\, u_j(\xi) + \int_\Gamma T_{ji}(x,\xi)\, u_j(x) d\Gamma(x) = \int_\Gamma G_{ji}(x,\xi)\, t_j(x) d\Gamma(x) \quad \xi \in \Gamma \tag{2}$$

where c_{ij} is a constant depending on the boundary geometry at the source point ξ. While the parameter c_{ij} can be evaluated analytically for 2-D problems with line boundary, the deduction of a general analytical expression is not straightforward in 3-D (Mantic 1993). An indirect method using the constraint of rigid body displacements has been commonly used to evaluate c_{ij} without having to integrate the strong singularity (Brebbia and Dominguez 1989).

For elastoplastic contact problems, the boundary integral equation is generally expressed in an incremental form. Three approaches are commonly used to derive the boundary integral equations for elastoplastic problems: the initial strain approach, the initial stress approach, and the pseudo body force approach. The initial stress approach is capable of handling perfectly plastic materials, whereas the initial strain approach is more suitable for strongly strain hardening materials. The initial stress approach uses the initial plastic stress rate as a primary domain unknown in the integral equation, which can be written as:

$$c_{ij}\, \dot{u}_j(\xi) + \int_\Gamma \dot{T}_{ji}(x,\xi)\, \dot{u}_j(x) d\Gamma(x) = \int_\Gamma \dot{G}_{ji}(x,\xi)\, \dot{t}_j(x) d\Gamma(x) + \int_\Omega \dot{E}_{mni}(x,\xi)\, \dot{\sigma}^P_{mn}(x) d\Omega(x) \quad \xi \in \Gamma \tag{3}$$

The initial strain formulation uses the plastic strain rate as a primary unknown in the integral equation:

$$c_{ij}\, \dot{u}_j(\xi) + \int_\Gamma \dot{T}_{ji}(x,\xi)\, \dot{u}_j(x) d\Gamma(x) = \int_\Gamma \dot{G}_{ji}(x,\xi)\, \dot{t}_j(x) d\Gamma(x) + \int_\Omega \dot{S}_{mni}(x,\xi)\, \dot{\epsilon}^P_{mn}(x) d\Omega(x) \quad \xi \in \Gamma \tag{4}$$

Here, $\dot{u}_i$, $\dot{t}_j$ are the displacement and traction rates, respectively; $\dot{\sigma}^P_{mn}$, $\dot{\epsilon}^P_{mn}$ are the plastic stress and plastic strain rates; and $\dot{S}_{mni}$, $\dot{E}_{mni}$ are the fundamental stress and strain rates, respectively. The integral equation for elastoplastic problems contains a domain integral and the discretization of full domain is necessary to evaluate the integral. The domain integral can also be converted to a boundary integral using the dual reciprocity method or a particular integral formulation (Henry and Banerjee 1988).

In the boundary integral equations, the boundary of integrals is divided into three parts: the contact boundary Γ_c, the boundary with prescribed displacement Γ_u, and the boundary with prescribed traction Γ_t. Outside the contact boundary, the boundary conditions are prescribed in the form of known displacements or tractions. Inside the contact area, the displacements and tractions (contact pressure and friction) are unknowns and remain to be determined. They are solved by imposing the continuity condition of displacements and the equilibrium of tractions at each contact node-pair within the contact region. The contact conditions in the normal direction require that the contact pressure should be compressive in the contact region and there is no penetration between two contacting bodies. That is,

$$\begin{cases} p_n(x) \geq 0, & \Delta u_n(x) = 0, \quad x \in \Gamma_c \\ p_n(x) = 0, & \Delta u_n(x) \geq 0, \quad x \notin \Gamma_c \end{cases} \tag{5}$$

Here, p_n and Δu_n represent contact pressure and relative displacement between two contact surfaces in the normal direction. In the tangential direction, there are three possible contact modes: frictionless, stick, and slip, and the contact conditions are written as:

$$\begin{cases} |p_t(x)| = 0, & |\Delta u_t| \geq 0, & \text{frictionless contact} \\ |p_t(x)| \leq \mu |p_n(x)|, & |\Delta u_t| = 0, & \text{stick contact} \\ |p_t(x)| = \mu |p_n(x)|, & |\Delta u_t| \geq 0, & \text{slip contact} \end{cases} \tag{6}$$

where μ is the coefficient of friction; p_t and Δu_t are tangential traction and relative displacement in the tangential direction.

Numerical Implementation

To solve boundary integral equations numerically, the boundary is divided into a sufficient number of elements that form a piecewise continuous approximation to the boundary. A number of contact elements and contact node-pairs are defined on the potential contact surfaces of contacting bodies. Both coordinates and the variation of displacements and tractions over elements are generally described in terms of the nodal values with a set of interpolation functions or isoparametric shape functions, $N_s(x)$. The integrations of the fundamental functions are often problematic as they are based on a solution of the system equations subject to a singularity load. For the source point ξ located outside the integration elements, the integral kernels are regular and the Gaussian quadrature can be used to evaluate the integration of the kernels over the elements. When the integration point coincides with the source point, the integral kernels $G_{ji}(x,\xi)$ and $T_{ji}(x,\xi)$ become singular. For the 2-D problems, the integral $\int_\Gamma G_{ji}(x,\xi)\, N_s(x) d\Gamma(x)$ has logarithm singular kernel while the integral $\int_\Gamma T_{ji}(x,\xi)\, N_s(x) d\Gamma(x)$ involves Cauchy singularity. Both logarithm and the principal values of Cauchy singular kernels can be evaluated

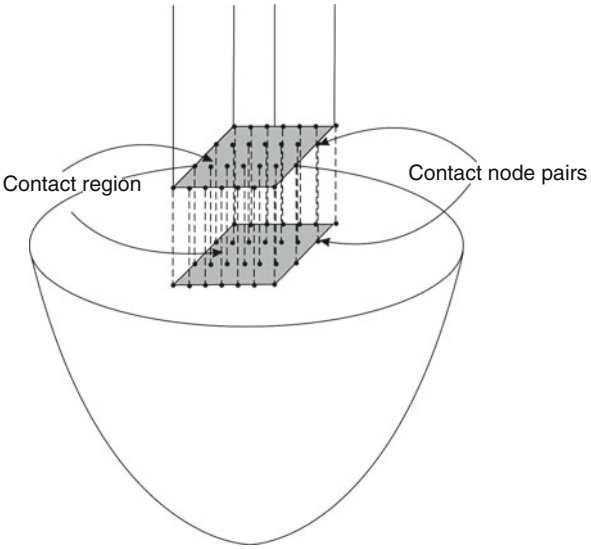
analytically (Kuo 2008). For the 3-D problems, the integral kernel $G_{ji}(x, \xi)$ is of order $1/r$, and the element subdivision technique (Lachat and Waston 1976) can be used to deal with weakly singular integrals. The kernel $T_{ji}(x, \xi)$ is of order $1/r^2$, which is interpreted in the Cauchy principal value. The strong singular integrals together with the constant c_{ij} constitute the diagonal elements of the coefficient matrix that can be evaluated by using the concept of rigid body motion.

Contact Region and Contact Pressure

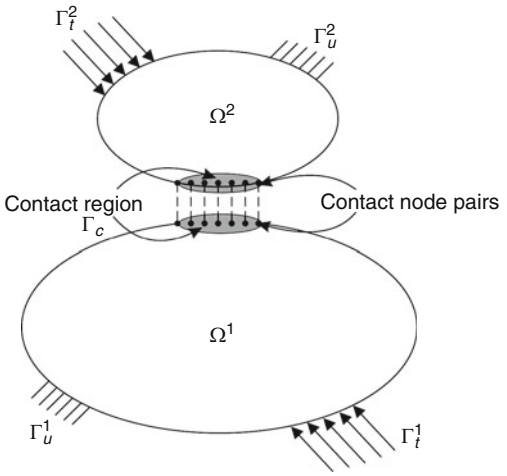
Contact problems are nonlinear because the contact area does not change linearly with the applied load and the friction on the contact surface is load dependent due to the irreversibility of frictional phenomena. Contact pressure and displacement at the interfaces are not prescribed and need to be solved using the contact boundary conditions. Whether the contact area is known a priori or unknown and to be determined depends on the geometric profiles of contact bodies.

For the conforming contact problem, the surfaces of the contacting bodies match each other in the unloaded state. Examples of conforming contact include slider flat bearings, journal bearings, a pin in a conforming hole, and indentation of a flat punch on a flat foundation (Fig. 1). In the conforming contact, the contact area is load independent and known before the application of loads. The contact pressure and displacement can be treated as unknowns and solved directly from the boundary integral equation. However, if the friction is present at the contact interface, the boundary separating the stick and slip regions is unknown and has to be determined iteratively. In friction contact problems, the friction behavior is dependent on the load history and the load is applied in small increments. In the initial step, all the contact node-pairs are assumed to be stuck together, i.e., the ratio of the tangential to normal traction must be smaller than the coefficient of friction. If the ratio exceeds the value of μ, the slip condition is imposed on the tangential traction in the next iteration.

For the non-conforming contact problem, the surfaces of the contacting bodies do not have the same shape, and the contact area is small compared to the sizes of the contacting bodies (Fig. 2). Examples of non-conforming contact include the point contact in a ball bearing, line contact made in a roller bearing, and contact between a pair of gears. In non-conforming contact, the contact area is load dependent and unknown a priori. Numerical iterations are needed to determine the actual contact region. A potential contact area and contact node-pairs are defined in the first solution stage. In addition to the friction slip check, there are two checks performed in the subsequent iterations: overlap check and tensile stress check. The tensile check is performed to ensure that contact stresses are compressive within the contact zone, and the overlap check ensures that no overlap or interpenetration occurs outside the contact area. Contact iterations are terminated when no errors are detected by these checks. In addition to the iterative approach, other formulations of contact problems include linear complementary



Boundary Element Methods for Contact Analysis, Fig. 1
Conforming contact problem: boundary element node pairs in contact region



Boundary Element Methods for Contact Analysis, Fig. 2
Non-conforming contact problem: boundary element node pairs in contact region

formulation, the Lagrange multiplier method, the penalty function method, and the flexibility matrix approach (Faraji 2005).

Key Applications

While the FEM has been well established and commercially developed, the BEM offers the following advantages for contact analysis:

1. Less data preparation time and mesh works. The number of elements and equations can also be reduced by boundary-only discretization.
2. More accurate solutions for contact pressure and contact area with fewer elements. The contact tractions and displacements are directly solved with the same degree of accuracy from the boundary integral equations.
3. High accuracy of contact stresses. Since all the approximations are confined to the boundary, the BEM can yield more accurate results for stress field for both interior and boundary points.
4. No need to define additional contact elements at the interface between two contacting bodies.
5. Ability to treat the singularities of contact tractions on the contact boundaries.

Several key applications of BEM for contact analysis include:

Design and stress analysis of contact components
The BEM provides an efficient and accurate approach for design and analysis of contact elements such as bearings, gears, and rollers. Less data preparation time and subsequent re-meshing works allow modification of geometries and contact loads with minimum efforts in the design process. Unlike FEM, boundary element analysis does not require additional contact elements at the interface, and the contact pressure can be solved with greater accuracy and efficiency.

Contact analysis of semi-infinite solids
Contact analysis of semi-infinite solids (or half-spaces) is commonly used to study such physical phenomena arising from the contact loading as contact fatigue failure, sliding friction, plastic deformation of mechanical components under repeatedly contact loading, and wheel-rail contact modeling. Unlike the FEM, which requires modeling as far as possible for the infinite domain, the BEM can give accurate results without modeling of a remote boundary. The BEM is also very efficient and accurate in three-dimensional analysis and very suitable for the problems with rapidly varying stress gradients such as local plasticity or cracking in materials under contact loading.

Contact analysis of engineered or rough surfaces
For three-dimensional surface topography, vast amounts of data are generated to describe real surface roughness. The micro-contact stress analysis at asperity scale requires ample iterations to determine correct contact state, which is very time-consuming. The BEM allows very fine mesh to model the rough surface and is very efficient and accurate in contact iteration to solve the contact pressure and contact region.

Cross-References

- ▶ Adhesive Contact of Elastic Bodies
- ▶ Adhesive Contact of Inelastic Bodies
- ▶ Contact Elasto-Plasticity
- ▶ Contact of Anisotropic Materials
- ▶ Contact of Layered Materials
- ▶ Contact of Materials with Inclusions
- ▶ Contact of Materials with Inhomogeneities
- ▶ Green's Functions in Contact Mechanics
- ▶ Hertz Theory: Contact of Cylindrical Surfaces
- ▶ Numerical Methods for Elastic Contact Problems

References

C.A. Brebbia, J. Dominguez, *Boundary Elements: An Introductory Course* (McGraw-Hill, Toronto, 1989)

A. Faraji, *Elastic and Elastoplastic Contact Analysis: Using Boundary Elements and Mathematical Programming* (WIT Press, Southampton, 2005)

D.P. Henry, P.K. Banerjee, A new BEM formulation for two and three dimensional elastoplasticity using particular integrals. Int. J. Numer. Methods Eng. **26**, 2079–2096 (1988)

C.H. Kuo, Contact stress analysis of an elastic half-plane containing multiple inclusions. Int. J. Solids Struct. **45**, 4562–4573 (2008)

C.H. Kuo, L.M. Keer, Contact stress analysis of a layered transversely isotropic half-space. ASME J. Tribol. **114**, 253–262 (1992)

J.C. Lachat, J.O. Waston, Effective numerical treatment of boundary integral equation. Int. J. Numer. Methods Eng. **10**, 991–1005 (1976)

V. Mantic, A new formula for the C-matrix in the Somigliana identity. J. Elasticity **33**, 191–201 (1993)

Y.C. Pan, T.W. Chou, Point force solution for an infinite transversely isotropic solid. ASME J. Appl. Mech. **43**, 608–612 (1976)

F.J. Rizzo, An integral equation approach to boundary value problems of classical elastostatics. Q. J. Appl. Math. **25**, 83–95 (1967)

T.C.T. Ting, V.G. Lee, The three-dimensional elastostatic green's function for general anisotropic linear elastic solids. Q. J. Mech. Appl. Math. **50**, 407–426 (1997)

Boundary Integral Equation Methods for Contact Analysis

▶ Boundary Element Methods for Contact Analysis

Boundary Lubricants

► Chemical Vapor Deposition Processes for Boundary Lubricants

Boundary Lubrication and EP Additives

► Tribochemistry of Extreme-Pressure Additives

Boundary Lubrication and Thin Film Lubrication

► Thin Film Lubrication

Boundary Lubrication by Lubricious Coatings

► Solid–Liquid Bi-phase Lubricating Coatings

Boundary Lubrication Mechanisms in Artificial Joints

Philippa M. Cann
Tribology Group, Imperial College, London, UK

Definition

Boundary lubrication is the process whereby a thin protective chemical film is formed on rubbing surfaces in the absence of a fluid film. The boundary layer reduces friction, carries the load, and prevents surface damage. Boundary lubricant additives are polar molecules that react or adsorb onto the surfaces to form the protective film.

Scientific Fundamentals: Tribology of Artificial Joint Implants

Prosthetic joints (hips, knees, ankles, etc.) are increasingly used to restore articular function that has been impaired due to disease, trauma, or genetic condition. Overall, the median age for joint replacement is decreasing, and this places increasing demands on implant function and longevity. In many cases, both short- and long-term failure of artificial joints is due to wear of the articulating surfaces. Artificial implant materials include metal (CrCoMo alloy, titanium, surgical stainless steel), ceramics (alumina, zirconia), and polymers (ultra-high-molecular-weight polyethylene, UHMWPE). Material loss and damage of the surfaces may originate in physical (abrasion or adherence) or chemical (corrosion) wear mechanisms. Surface damage results in the formation of micron (polymer) or nanometer (metal) sized wear debris that is biologically active and often provokes an adverse cellular response (Revell et al. 1997).

Wear of artificial joints is essentially determined both by the properties of the articulating surfaces and the lubricating film. In recent years, there have been significant advances in material properties and implant designs that have improved wear performance and implant life. However, less is known about the properties of the lubricating films and the ways these affect implant wear. At present, there are two recognized mechanisms identified for film formation in artificial joints: fluid film elastohydrodynamic lubrication (EHL) (Dowson 2006) and boundary lubrication (Hills 2000).

Boundary lubrication refers classically to the formation of thin (∼nanometer), adherent films at the rubbing interface, which reduces friction and prevents surface damage in the absence of a hydrodynamically generated fluid film. Thus, the regime is prevalent at low speeds or with low-viscosity fluids. Boundary additives, for example, fatty acids or amines, react or adsorb at the surface to form the load-carrying film. In this regime, the properties of the additives (chemical structure, molecular conformation, polarity) and the surface (chemistry, polarity, surface wetting) determine film formation and lubrication performance.

Surface damage and wear will occur when the lubricant film drops below a critical thickness. EHL analysis, based on isoviscous/elastic assumptions, of articulating implants provides an insight into the level of film formation and the probable lubrication mode. Dowson (2006) reports an average minimum film thickness for metal-on-metal hip joints of ∼30–60 nm. The combined surface roughness of metal hip joints is approximately 20 nm, and this analysis suggests most artificial implants function in the mixed/boundary lubrication regime for at least part of the gait cycle.

Synovial Fluid Properties and Articulation Conditions in Artificial Joints

The mechanisms of film formation in artificial joints depend on both the properties of the lubricating fluid and the contact conditions of the implant. Artificial joints are lubricated by periprosthetic synovial fluid that refills the cavity after the implant operation. Synovial fluid contains a complex mixture of large and surface active molecules and has a shear-thinning rheology (Cooke et al. 1978). Measured viscosity for healthy synovial fluid at physiological shear rates ($>10^3$ s^{-1}) is typically in the range 0.01–0.1 PaS (Cooke et al. 1978). There are significant chemical and physical differences between healthy and diseased or periprosthetic synovial fluid. The effective viscosity of periprosthetic synovial fluid is much less (at physiological shear rates) and can drop to less than 0.001 PaS (Cooke et al. 1978). As one of the most important factors determining EHL formation is the bulk viscosity, it is expected that the predicted film thickness for periprosthetic fluid is less than for healthy synovial fluid.

Synovial fluid chemical composition includes large and surface active molecules including hyaluronan, proteins and glycoproteins, phospholipids, and cholesterol (Kitano et al. 2001). A number of these species, particularly phospholipids and proteins, are thought to contribute to the boundary lubrication of the implant (Schwarz and Hills 1998). Phospholipids are a major component of biological membranes, which form structured mono- and bilayers at interfaces.

Other synovial fluid components have been identified as having boundary lubrication capabilities. In the late 1970s, attention turned to *lubricin* a highly purified glycoprotein fraction of synovial fluid as the primary boundary lubricant of articular cartilage (Schwarz and Hills 1998). Analysis of *lubricin* indicated it contains protein and carbohydrate with a small percentage (~12%) of phospholipid. The phospholipid fraction was identified as the primary boundary lubricant in *lubricin*. The water-soluble glycoprotein component is thought to act as a carrier for the insoluble phospholipid.

Hyaluronan is a high-molecular-weight (10^5–10^6 Da) linear polysaccharide; in diseased fluid, the concentration and molecular weight are decreased. It is thought to contribute to viscoelasticity and viscosity enhancement but to have negligible boundary lubrication function.

Healthy synovial fluid also contains soluble proteins, primarily albumin and γ-globulin (albumin:γ-globulin ratio 2:1) (Kitano et al. 2001). The total protein content is approximately 20 mg/ml, although in diseased or periprosthetic fluid the protein content increases to greater than 30 mg/ml (Kitano et al. 2001). In screening tests it is customary to use diluted bovine calf serum as a periprosthetic synovial fluid substitute. The major constituents of bovine calf serum are proteins (albumin, γ-globulin) at a concentration of approximately 70 mg/ml. To replicate synovial fluid protein concentrations (20–30 mg/ml), bovine serum is diluted to 25% w/w in distilled water for implant simulator tests. Dilute bovine serum has similar rheological properties to synovial fluid, with an effective viscosity of ~0.001 PaS at physiological shear rates.

Artificial joints operate under transient conditions of load and speed, which result in a range of contact conditions and lubrication mechanisms. Typically, these joints function under transient load (~1 Hz) and speed conditions (Dowson 2006). For example, in steady-state, walking hip joints experience two loading peaks per cycle, with loads varying from ~0.1 to 3.5 KN and linear speeds of 0–30 mm/s. The maximum contact pressures (<70 MPa) and speeds are generally low compared with engineering contacts, although this depends on patient biomechanics, component materials, type of joint, and design. The range of speeds and contact pressures experienced by articular implants make it likely different lubrication mechanisms are prevalent at different parts of the gait cycle.

Surface Examination of Explanted Artificial Joints

Direct evidence of boundary film formation by synovial fluid components has come from post-test examination of implant surfaces, from hip simulators, and joint retrieval studies (Wimmer et al. 2003). Analysis of surface layers on explanted hip joints, following revision, provides information on the nature of synovial fluid films formed in vivo. Wimmer et al. (2003) presented evidence of thick, carbon-rich deposited layers on the articulating surfaces of metal-on-metal hip joints. It was concluded that the films were formed by denatured proteins deposited from solution in the high-pressure contact regions and suggested the proteins acted as a solid lubricant reducing adhesion and abrasion.

There is also evidence of phospholipid layers on retrieved implant surfaces (Schwarz and Hills 1998) who analyzed loosely bound surface layers from a series of metal-on-polymer implants (hip and knee). Phospholipids in synovial fluid are predominately phosphatidylcholines, which can contain, for example, saturated dipalmitoyl phosphatidylcholine, or unsaturated fatty acid chains. It was concluded that the unsaturated components were adsorbed at the sliding surfaces, forming a boundary layer that contributed to reduced friction.

Synovial Fluid Tribology Tests and Model Surface Studies

Clearly, proteins and phospholipids are important components of synovial fluid and are found in surface layers on explanted joints. The key question is what role do these molecules play in the lubrication process and surface protection? Insights into synovial fluid boundary lubrication behavior can be obtained from friction and wear tests with different fluid compositions and from film formation measurements on model surfaces.

Tribology Testing of Synovial Fluids

There is an immense library of research literature reporting tribology tests of artificial joints with different materials, lubricants, and test methods. Tribology testing falls into two main categories: implant simulator studies and model laboratory tests. The implant simulators test the artificial joint under realistic levels of gait kinematics and loading. They are usually used to assess the wear performance of new joint designs and materials; however, some studies have also reported the effects of different lubricant composition. These include the effect of protein concentration and phospholipid content on friction and wear.

Most studies have concentrated on UHMWPE wear measurements. In simulator studies, the UHMWPE wear rate is reported to first increase and then decrease, or decrease with increasing protein content (Wang et al. 2003). However, in pin-on-disc tests increasing protein and, in particular, γ-globulin content is reported to increase UHMWPE wear rate. The addition of phospholipids decreases friction and, in some cases, also decreases the wear rate, although this depends on the protein concentration (Sawae et al. 2008). Clearly, the addition of proteins and phospholipids significantly affect both friction and wear of implant materials. However, it is difficult to draw firm conclusions as to the influence of protein concentration on component wear.

There is other more direct evidence for the boundary film formation with synovial fluid components. Imaging methods have been successfully applied to model tests where one of the sliding components has been replaced by a transparent disc, usually glass. Some studies have used optical interferometry (Mavraki and Cann 2009) to measure film formation in a model sliding contact with bovine calf serum and simple protein solutions. The authors reported the thickest films were obtained at low speeds. During rubbing, a deposited surface layer (10–20 nm) rapidly formed, which supported the load even at zero speed. These observations confirm that proteins in bovine serum form adherent films within rubbing contacts and effectively act as boundary lubricants by separating the surfaces. Inspection of the femoral head at the end of the test revealed the presence of gelatinous protein deposits on the inlet side of the contact zone. These films are similar to those reported for organic deposits on explant surfaces (Wimmer et al. 2003).

Nakashima, Sawae, and Murakami (Nakashima et al. 2005) used fluorescence microscopy to probe the contribution of different proteins (albumin and γ-globulin) to boundary film formation on hydrogels. They concluded that for optimum boundary behavior, a multilayer protein composition was required; the γ-globulin providing a coherent surface film that is overlaid by adsorbed albumin molecules. The γ-globulin provided the anti-wear protection and the albumin a low-shear strength film to reduce friction.

Model Surface Studies of Synovial Fluid Component Films

Outside of the biotribology research field there is a substantial body of work concerning protein and phospholipid interactions with model surfaces. These studies usually measure molecular adsorbance from solution under static and solvent flow conditions. Protein (albumin and γ-globulin) adsorbance at surfaces has been extensively studied, particularly with regard to the food science and marine biofouling fields. A wide range of substrates (metals, polymer, and model surfaces), solution chemistries (pH, salt content) and protein concentrations have been reported. In water solution, the bovine serum albumin molecule is reported to adopt an ellipsoidal shape with a 11.6 nm long and 2.7 nm short axis (Pradier et al. 2003) and has a net negative charge at physiological pH (>pH = 4.8). Most studies report a two-stage adsorption process with metal surfaces: an initial fast adsorption with partial unfolding followed by longer-term conformational reorganization. The secondary structure of the adsorbed protein was also reported to change on adsorbance with an increase in the β-sheet content compared with the α-helix (compared with the crystalline form). At higher BSA concentrations (comparable to SF levels), further protein adsorbance occurs to form a second layer that has a more compact helical molecular structure (Pradier et al. 2003).

The formation of adsorbed phospholipid films has been extensively investigated as they are important components of biological membranes. In Langmuir-Blodgett experiments, lipids form structured bilayers with the polar head group adsorbed at the air–water interface. The long methyl chains are oriented away from the interface forming a hydrophobic surface (Hills and Butler 1985).

Summary of Boundary Lubrication Mechanisms in Artificial Implants

This article has considered the evidence of boundary lubrication mechanisms in artificial joints. Two components of periprosthetic synovial fluid, phospholipids and proteins, have been identified as potential boundary additives. Both components have been identified in films formed on explanted joint surfaces. From in vitro testing, there is considerable evidence that phospholipids reduce friction and wear of implant materials. It is suggested that they adsorb to form structured bilayers with a hydrophobic outer surface. Clearly, they function as classical boundary additives.

The boundary lubrication mechanisms of proteins are more complicated. Model studies report albumin adsorbs on metal and polymer surfaces to form mono- and multilayer molecular films. The degree of coverage, molecular conformation, and film thickness depend on the nature of the substrate, solution pH, and protein concentration. The boundary lubrication contribution of these relatively thin (∼nm) protein films has not been clarified. Analysis of implant surface films after testing in hip simulators with bovine serum has shown that much thicker deposits of degraded protein are formed. It is suggested that these act as thick "solid lubricant" layers (Wimmer et al. 2003) rather than as traditional boundary additives.

The boundary role of proteins is only part of the story, however. Wear of implants is the result of both physical (abrasion, adhesion) and chemical surface mechanisms. Organic reacted or deposited layers also appear to play a role in conditioning the alloy subsurface, thus affecting wear (Pourzal et al. 2009). The formation of this layer through mechanical mixing was reported to be the origin of the excellent wear resistance of the CrCoMo alloy. However, proteins also contribute to the corrosive loss through surface film formation. It is suggested that protein films play a dual role in controlling surface reactions; firstly, by forming a negatively charged protective layer, they prevent attack at the CrCoMo interface by corrosive anions. However, the protein layer also contributes through ligand-induced dissolution at the interface, which increases the chromium concentration in the synovial fluid (Lewis et al. 2005).

References

A.F. Cooke, D. Dowson, V. Wright, The rheology of synovial fluid and some potential synthetic lubricants for degenerate synovial joints. Proc. IMechE Eng. Med. **7**, 66–72 (1978)

D. Dowson, Tribological principles in metal-on-metal hip joint design. Proc. IMechE Part H **220**, 161–171 (2006)

B. Hills, Boundary lubrication in vivo. Proc. IMechE Part H **214**, 83–94 (2000)

B.A. Hills, B.D. Butler, Phospholipids identified on the pericardium and their ability to impart boundary lubrication. Ann. Biomed. Eng. **13**, 573–586 (1985)

T. Kitano, G.A. Ateshian, V.C. Mow, Y. Kadoya, Y. Yamano, Constituents and pH changes in protein rich hyaluronan solution affect the biotribological properties of artificial articular joints. J. Biomech **34**, 1031–1037 (2001)

A.C. Lewis, M.R. Kilburn, I. Papageorgiou, G.C. Allen, C.P. Case, The effect of synovial fluid, phosphate buffered saline solution and water on the dissolution and corrosion properties of CoCrMo alloys as used in orthopaedic implants. J. Biomed. Mater. Res. **73A**, 456–467 (2005)

A. Mavraki, P.M. Cann, Friction and lubricant film thickness measurements on simulated synovial fluids. Proc. Inst. Mech. Eng. Part J (J. Eng. Tribol.) **223**, 325–335 (2009)

K. Nakashima, Y. Sawae, T. Murakami, Study on wear reduction mechanisms of artificial cartilage by synergistic protein boundary film formation. JSME Int. J. **48**, 555–561 (2005)

R. Pourzal, R. Theissmann, M. Morlock, A. Fischer, Micro-structural alterations within different areas of articulating surfaces of a metal-on-metal hip resurfacing system. Wear **267**, 689–694 (2009)

C.M. Pradier, F. Karman, J. Telegdi, E. Kalman, P. Marcus, Adsorption of bovine serum albumin on chromium and molybdenum surfaces investigated by fourier-transform infrared reflection-absorption spectroscopy (FT-IRRAS) and X-ray photoelectron spectroscopy. J. Phys. Chem. B **107**, 6766–6773 (2003)

P.A. Revell, N. Al-Saffar, A. Kobayashi, Biological reaction to debris in relation to joint prostheses. Proc. IMechE Part H **211**, 187–197 (1997)

Y. Sawae, A. Yamamoto, T. Murakami, Influence of protein and lipid concentration of the test lubricant on the wear of ultra-high molecular weight polyethylene. Tribol. Int. **41**, 648–656 (2008)

I. Schwarz, B. Hills, Surface-active phospholipid as the lubricating component of lubricin. Rheumatology **37**, 21–26 (1998)

A. Wang, A. Essner, G. Schmidig, The effects of lubricant composition on in vitro wear testing of polymeric acetabular components. J. Biomed. Mater. Res. B Appl. Biomater. **68B**, 45–52 (2003)

M.A. Wimmer, C. Sprecher, R. Hauert, G. Täger, A. Fischer, Tribochemical reaction on metal-on-metal hip joint bearings. A comparison between in-vitro and in-vivo results. Wear **255**, 1007–1014 (2003)

Bovine Serum

► Polymers in Biotribology

Brake Block

► Brake Friction Materials

Brake Friction Materials

Ho Jang
Department of Material Science and Engineering, Friction Materials Research Laboratory, Korea University, Seoul, Korea

Synonyms

Brake block; Brake lining; Brake pad; Brake segment

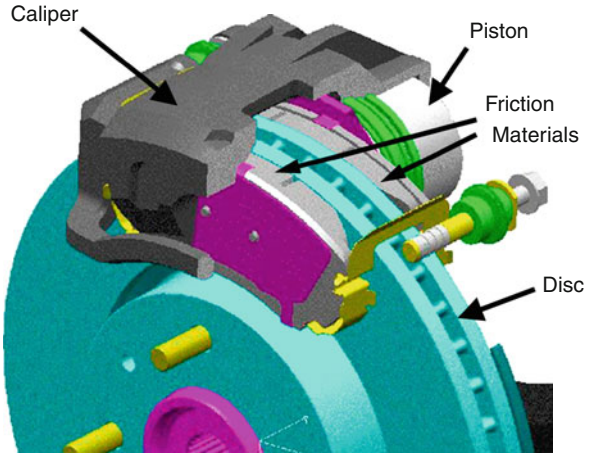
Definition

Brake friction materials are multiphase composites containing numerous ingredients and are used to control the vehicle speed by squeezing a rotating brake disc (or drum). The types and relative amount of ingredients in the friction material and fabrication methodology strongly affect the braking efficiency, wear resistance, and propensity of brake-induced noise, vibration, and harshness.

Scientific Fundamentals

Introduction

The friction material in a brake system is a multiphase composite containing more than ten ingredients and is considered one of the most complicated material systems. This is because the brake friction material must be developed to sustain brake performance in a wide temperature range and designed to moderate friction-induced excitation at the sliding interface by proper material design. A typical set-up for a disc brake system with the friction material in the brake pad is shown in Fig. 1. A pair of brake pads is squeezed against a rotating disc (or a drum) to control the speed of a vehicle by generating friction force at the sliding interface. In the case of a disc brake system, brake pads are enclosed in a caliper to squeeze a rotating disc by the piston actuated by a hydraulic system, while curved linings in a drum brake system expand to slide against the inside surface of a drum. Brake applications comprise the conversion of the kinetic energy to other energy forms such as heat, noise, and vibration. The friction heat increases the temperature on the rubbing surface up to 800°C and the friction-induced excitation in the sliding interface transmits to the brake components and often produces noise and vibration. Gaining a detailed understanding of friction mechanism, however, is a formidable task, since the sliding interface is transient and involves non-equilibrium reactions at high pressure and high temperature.



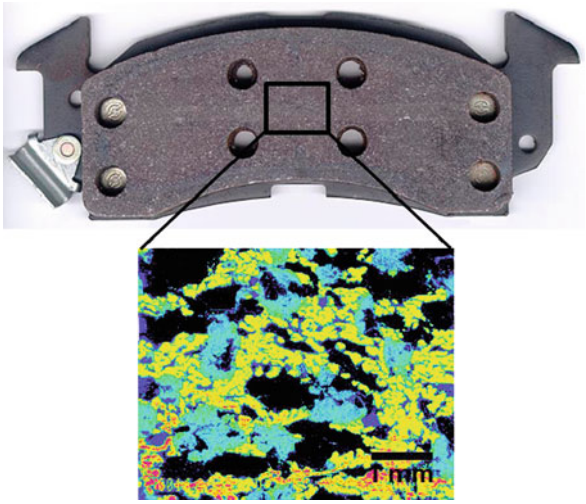
Brake Friction Materials, Fig. 1 A brake assembly showing a pair of brake pads inside a caliper. Each brake pad consists of a friction material, a back plate, and a noise damper (*shim*)

Because a single material or composites with a few ingredients have not fulfilled the numerous performance-related requirements of a brake system, a great deal of effort has gone into the development of brake friction material. This has included modification of ingredients and searching for new ingredients to replace traditional ones. Despite these various efforts, the development of the commercial brake friction material has relied mainly on trial and error and the previous test experience because of insufficient understanding of the friction mechanism at the sliding interface.

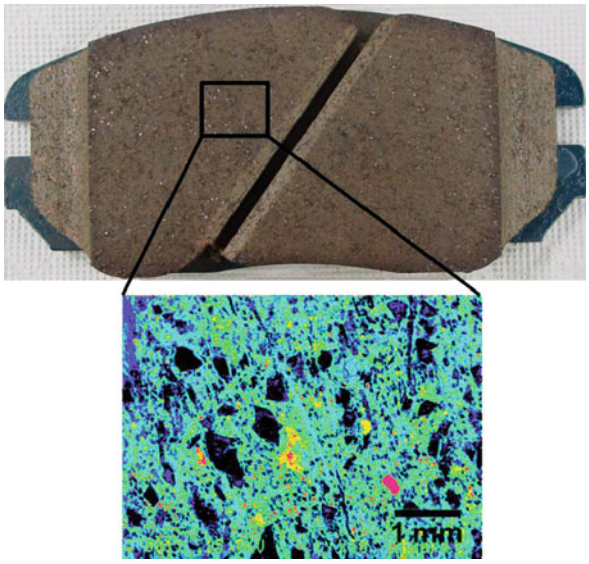
Types of Brake Friction Materials

Brake friction materials are classified as asbestos organic, nonasbestos organic, semi-metallic, sintered metallic, and carbon composite types by the ingredients and their typical friction characteristics.

- *Asbestos organic type*: Asbestos-based organic friction materials were widely used prior to mid-1970 due to its high friction effectiveness, thermal stability, and reinforcing capability. They normally contained 30–40 wt.% of organic ingredients, several property modifiers, and asbestos. The friction materials based on asbestos, however, were banned in 1989 after asbestos was declared a carcinogen (Nolan and Langer 1987). Currently, asbestos is not permitted to be used for automotive components in the US, EU, and most of the developed countries.
- *Semi-metallic type*: This material was introduced in the 1960s and eventually made up approximately 90% of the passenger and light truck market in the 1980s.



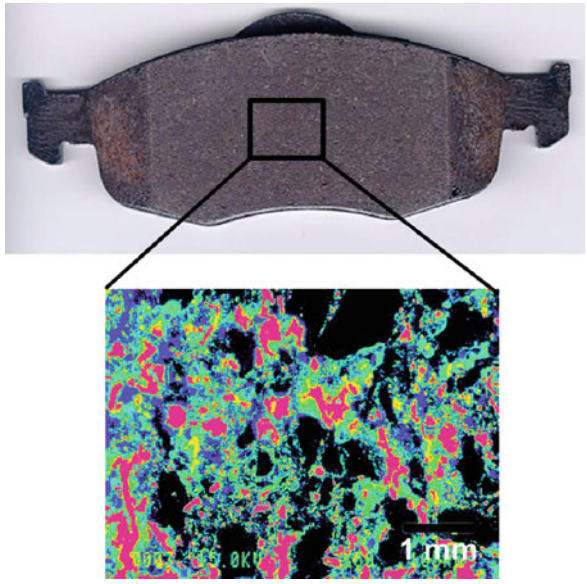
Brake Friction Materials, Fig. 2 A semi-metallic friction material riveted to a *steel back plate*. The micrograph shows ingredients in the friction material at a high magnification, showing large amounts of iron powder and steel fibers (*bright particles*)



Brake Friction Materials, Fig. 3 A non-steel friction material integrate-molded to a *steel back plate*. The micrograph shows organic ingredients (*grey color*) in the friction material at a high magnification

It normally contains more than 50 wt.% of iron powder and steel fibers bound by a phenolic resin (Fig. 2). This type of friction material shows high aggressiveness against gray iron discs while it offers stable friction and improved fade resistance. It has been used for heavy-duty operations.

- *Nonasbestos organic (NAO) type*: NAO-type friction materials were developed to replace asbestos-based friction materials by using nonasbestos fibers and other property modifiers. The asbestos substitutes comprise various inorganic fibers of glass, mineral, ceramic, steel, copper, brass, and organic materials. Platy materials such as mica, vermiculate, and porous metals are also used as an asbestos substitute. No single fiber can replace asbestos, and a mix of several reinforcing fibers is used to meet various performance and processing-related requirements. In general, commercial NAO friction materials are divided into two different types: non-steel (Fig. 3) and low-steel type friction materials (Fig. 4), depending on the existence of the steel fiber (low steel type normally contains less than 10 wt.%) in the friction material, which increases friction effectiveness but often aggravates disc (or drum) wear. Semantic errors are often found in the literature when only the non-steel type is considered an NAO friction material.
- *Sintered metallic type*: This material was developed for high-speed trains and aircraft brakes, which require



Brake Friction Materials, Fig. 4 A low-steel friction material integrate-molded to a *steel back plate*. The micrograph shows typical ingredients of a low-steel friction material such as steel fibers (*grey*) and graphite (*dark*) and cashew particles (*dark*) at a high magnification

Brake Friction Materials, Table 1 Typical ingredients used for brake friction materials

Group	Type/role	Ingredients
Binders	Straight	Phenolic resin, polyimide, resol, COPNA resin
	Modified	Aralkyl modified, boron-modified, cashew modified, silicon-modified, phosphor-modified, acryl-modified, epoxy-modified, rubber modified phenolic resin
Reinforcements	Fibrous	Aramid pulp, steel fiber, bronze fiber,[a] copper fiber,[a] brass fiber,[a] stainless fiber, rock wool, glass fiber, ceramic fiber, potassium titanate (whisker),[a] sepiolite,[a] basalt fiber, cellulose fiber, carbon fiber, acrylic fiber
	Non-fibrous	Mica, vermiculite, potassium titanate (plate), copper powder,[a] iron powder, woolastonite
Property modifiers	Lubricants	Graphite, MoS_2, CaF_2, ZnS, Sb_2S_3,[a] Sb_2O_5,[a] WS_2, CuS, Na_3AlF_6, cokes
	Abrasives	$ZrSiO_4$, quartz, Al_2O_3, MgO, ZrO_2, zeolite, Fe_3O_4, SiC, ZnO, chromite
	Miscellaneous modifiers	$Ca(OH)_2$, cashew, rubber powder, cork, Al, Cu powder,[a] bronze powder,[a] cut copper,[a] Zn powder, Sn powder
	Space fillers	$BaSO_4$, $CaCO_3$

[a]Materials with potential health hazards or environmental issues

thermal stability at extreme braking conditions. It normally contains ceramic materials in a metallic matrix based on iron or copper.

- *Carbon composite type*: This is a carbon fiber–reinforced carbon matrix composite and developed mainly for aircraft and racecar brakes as a substitute for heavy sintered metallic friction materials. Due to its excellent thermal stability and relatively high heat capacity, its application has been expanded to high-end passenger cars.

Ingredients in Brake Friction Material

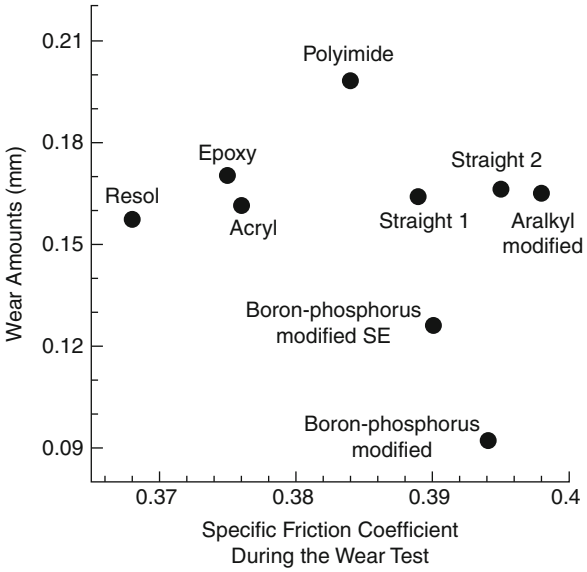
There are more than 200 ingredients used to produce commercial brake friction materials; the typical ones are listed in Table 1. Their tribological properties are affected by size, shape, and chemical composition, and details about the ingredients are normally considered proprietary information, along with the formulation of the friction material. Although little information is available in the literature, the effects of the ingredients on brake performance are summarized by grouping them as binders, reinforcements, and property modifiers, although some play multiple roles.

Binder Resin

Binder resin structurally integrates the ingredients and provides physical strength to maintain friction effectiveness during the braking operation. Phenolic resin is the most commonly used binder for brake friction material, which is produced by the condensation reaction of phenol and formaldehyde by acidic or basic catalyst. Two different types are used according to the manufacturing process. In wet processing, a resol in a viscous state is used to follow the thermoplastic process. This is a one-step resin with sufficient formaldehyde and that uses a basic catalyst. In dry processing, a novolac containing insufficient formaldehyde is normally used. This is a two-step resin and is produced by an acid catalyst. The two-step resin requires a curing agent such as hexa (hexamethylenetetramine). When the phenolic resin is cured at high temperature, it changes the resin from thermoplastic to thermoset by crosslinking.

Wear of the binder resin below 200°C relies on the breaking of weak bonding, such as hydrogen bonding, between chains. On the other hand, wear beyond 350°C is attributed to the random scission of polymer chains, oxidation, and carbonization of phenolic resin. The friction coefficient is considerably affected by the viscoelastic behavior above the glass transition temperature near 280°C followed by a rubbery state up to 320°C, until the resin goes through the post curing of unreacted phenolic resin (Hong et al. 2009).

Various modified phenolic resins have been developed to optimize various tribological properties of the brake friction material. Cashew-modified resins were developed to provide resilience and to increase friction effectiveness. Oil-modified resins were developed to enhance flexibility, although they aggravate fade resistance. Si-modified resin was developed for high friction effectiveness with wear resistance at elevated temperatures. The improved high temperature properties of the Si-modified resin were attributed to the Si-O bonds produced during the reaction of siloxanes with phenolic



Brake Friction Materials, Fig. 5 The wear amount and the coefficient of friction obtained during wear tests using commercial brake pads based on a formulation with different binder resins

compounds (Knop and Scheib 1979). Heat resistance of two different phenolic resins, straight resin and aralkyl modified resin, for non-steel friction materials were compared using a simple formulation (Kim and Jang 2000). It showed that the aralkyl-modified resin had better thermal stability than the straight resin. However, the aralkyl-modified resin exhibited extended heat affected zone on the sliding surface with the wear rate approximately two times higher than that of the straight resin. Figures 5 and 6 show the tribological properties of brake friction materials such as the friction level, wear rate, and fade amounts measured from commercial brake pads based on the same formulation with different binder resins.

Reinforcements

Reinforcing fibers commonly used for brake friction materials are aramid fibers (or pulp), metallic, ceramic, glass, acrylic, carbon, and other natural fibers. In general, commercial NAO friction materials contain 5–25 vol.% of fibrous ingredients and the type and relative amount of the reinforcing fibers affect various aspects of brake performance and wear rate.

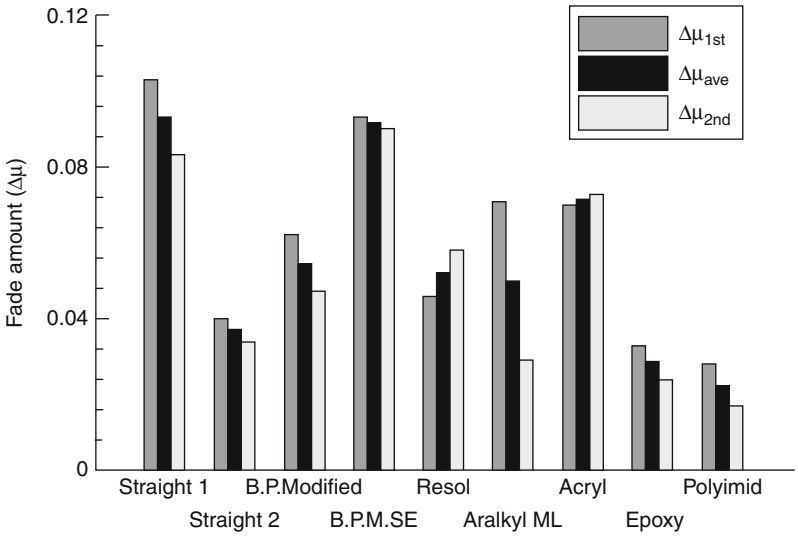
Among them, aramid fibers have been used as an excellent substitute for asbestos since they allow a cold preform process during manufacturing and provide superior wear resistance and anti-fade at elevated temperatures (Nicholson 1995). A study of friction and wear of sliding couples showed that copper and steel fibers are unfavorable in terms of noise and vibration, while they improve fade resistance (Jang et al. 2004). Steel fibers, in particular, have been used to increase friction effectiveness at elevated temperatures. However, a large amount of steel fibers in the friction material can produce excessive disc wear, which often causes brake judder (low-frequency vibration) due to uneven disc thickness. Ceramic fibers are used to improve friction stability at elevated temperature since they have excellent thermal resistance. However, they often produce disc scoring and brake noise when the ceramic fiber contains shots (beads), which are formed during the manufacturing process. Glass fibers have been used as an asbestos substitute because of its high-temperature friction stability. Two different types of glass fibers have been used for commercial friction materials: bundled fibers and open fibers. The fibers with 2–3 mm in length and 5–10 μm in diameter are used and they are normally coated to assure good bonding with a binder resin. Synergistic effects from aramid pulp and whisker fibers were also reported (Fig. 7). The improved friction stability and wear resistance by mixing the two disparate ingredients were closely related to the durability of friction films at the rubbing interface (Kim et al. 2001). When the friction material was produced with aramid pulp and potassium titanate whiskers at a certain ratio, the friction coefficient was well retained at elevated temperatures.

Non-fibrous reinforcements have also been used to improve wear resistance and heat resistance. Various metals and minerals are used in a plate or granular shape. However, they do not increase the strength of the friction material due to a small aspect ratio but are used for other purposes such as wear resistance, thermal conductivity, noise reduction, friction stability, and others (Jacko and Rhee 1992).

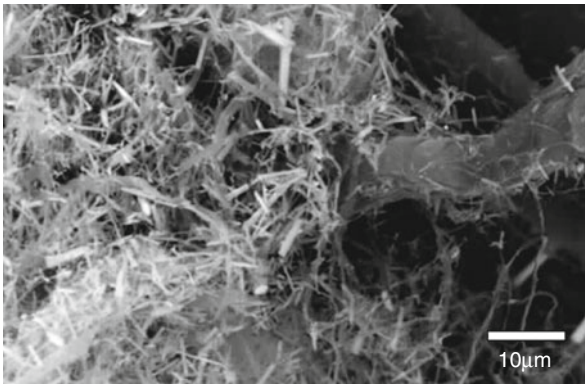
Property Modifiers

Property modifiers are added to the formulation in the initial development stage, not only to secure basic tribological properties such as friction level, fade resistance, and wear rate, but also to adjust the specific friction characteristics such as noise propensity, pedal feel, and aggressiveness against a counter surface. Various modifiers have been used to tailor brake performances at a wide range of braking conditions. They can be classified into lubricants, abrasives, specific modifiers, and space fillers.

Solid lubricants are used to improve wear resistance and reduce noise propensity. Various carbonaceous and chalcogenide materials are used as solid lubricants. Graphite is a typical solid lubricant widely used in the brake friction materials, which produces a self-sustaining



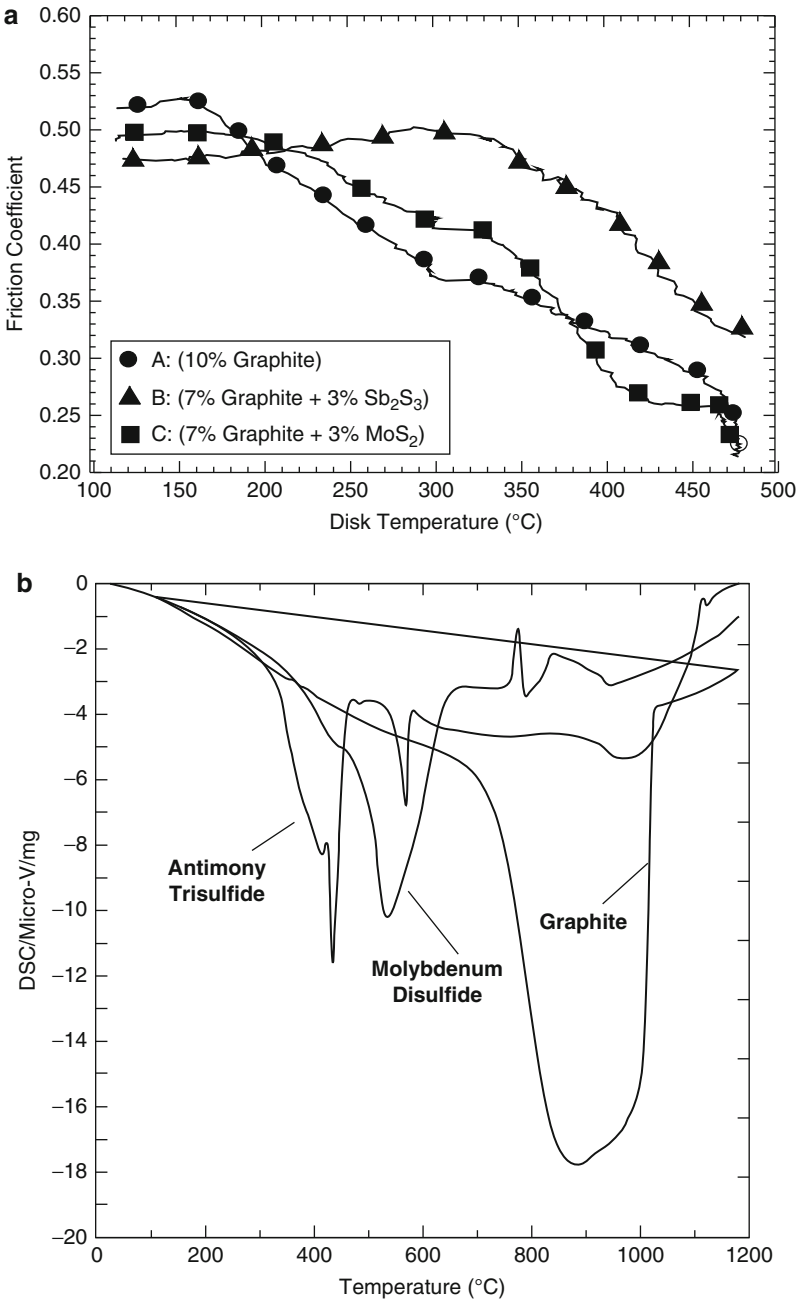
Brake Friction Materials, Fig. 6 The amounts of fade observed during the dynamometer tests measured using commercial brake pads based on a formulation with different binder resins



Brake Friction Materials, Fig. 7 Fibrillated aramid pulp mixed with short ceramic whiskers providing friction stability and wear resistance of brake friction materials due to synergistic effects (Kim et al. 2001)

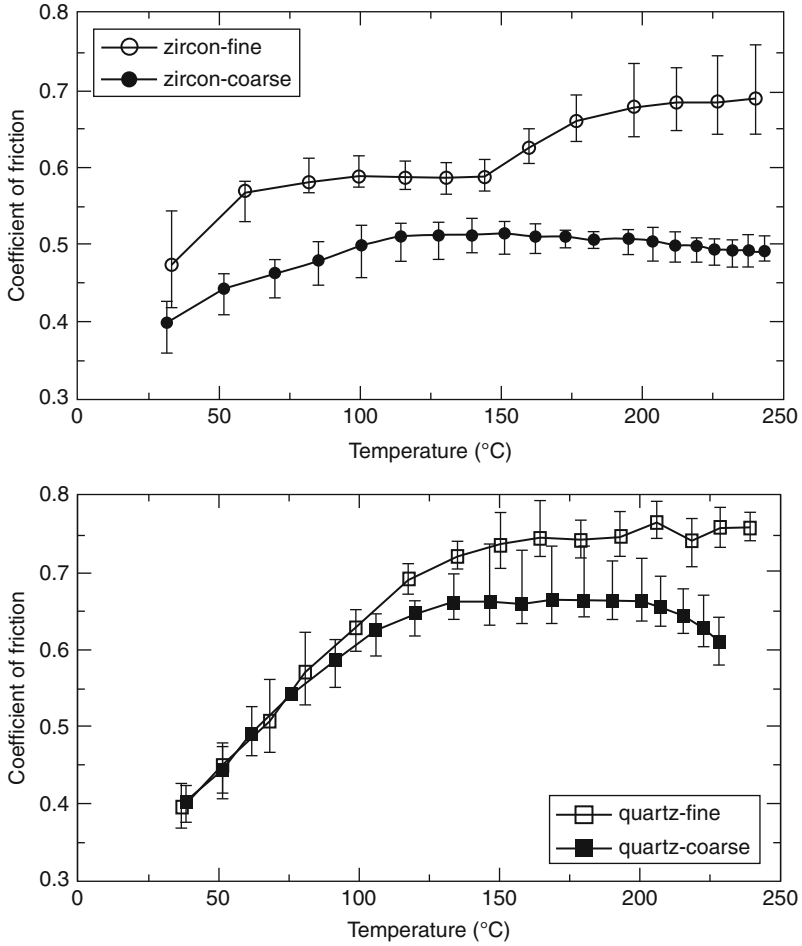
lubricant layer and is sensitive to humidity. Various products such as natural graphite, synthetic graphite, and amorphous graphite are available and they are generally used as a medium-temperature lubricant and sensitive. Molybdenum disulfide is another popular solid lubricant in the friction material, which is used for low-temperature lubrication below 100°C, but it can be oxidized to form MoO_3 above 400°C (Cho et al. 2006), which creates a stable friction transfer layer (Nicholson 1995). Antimony trisulfide is used as a high-temperature solid lubricant due to higher temperature stability and lower thermal conductivity than carbonaceous lubricants. The friction level changed at elevated temperatures and the thermal decomposition of graphite, MoS_2, and Sb_2S_3 obtained from differential scanning calorimetry are shown in Fig. 8. However, Sb_2S_3 is suspected as a carcinogen since it is oxidized to Sb_2O_3, a potential carcinogen, during brake applications (GADS list 2010). As a substitute for antimony trisulfide, tin sulfide and copper sulfide are currently tested for commercial friction materials. A mixture of solid lubricants is often used to maintain lubricity at different temperature ranges because each solid lubricant has its own decomposition temperatures beyond which it does not play a role as a lubricant (Campbell et al. 1966).

Abrasives are used to eliminate pyrolized friction film on the disc surface and to control friction effectiveness (Jacko and Rhee 1992; Nicholson 1995). In general, fine particles are employed when hard materials such as silicon carbide (SiC), alumina (Al_2O_3), and kyanite (Al_2SiO_5) are used, while large particles are used when they are mild abrasives such as quartz and zircon. The size effect of the abrasives on the tribological properties of the friction material was studied considering the friction level and wear of the friction material and gray iron discs (Cho et al. 2008b). They showed that the thermal stability of the friction material was improved and the amplitude of the friction coefficient was decreased when large abrasive particles were used. The friction material with fine abrasives was less aggressive to the counter surface while the friction material wear was relatively large compared with



Brake Friction Materials, Fig. 8 (**a**) The coefficient of friction measured as a function of temperature using the brake pads containing different amounts of solid lubricants and (**b**) the thermal decomposition of graphite, MoS_2 and Sb_2S_3 obtained from differential scanning calorimetry (Cho et al. 2006)

those with large size particles. Fracture toughness of the abrasives was also a crucial factor as the morphology of the particle for friction and wear of the friction material (Lee et al. 2010). They found that the wear rate of the friction material increased when the abrasive with low fracture toughness was used (Fig. 9). They also confirmed that the friction effectiveness of the fine abrasives was higher than coarse ones, and the fracture toughness of abrasive particles affects the friction coefficient. The effect of the relative amount of a solid lubricant (Sb_2S_3) and an



Brake Friction Materials, Fig. 9 The change of the friction coefficient as a function of temperature using the constant interval tests using the brake pads containing two different sizes of zircon and quart particles (Lee et al. 2010)

abrasive ($ZrSiO_4$) on friction characteristics was studied using a full-scale inertial dynamometer (Jang and Kim 2000). They found that the relative amount of the lubricant and abrasive significantly affected the friction material wear and torque variation during brake applications.

Various organic and inorganic materials are added to improve specific tribological properties of a brake friction material. Metallic chips such as copper, brass, and other metal powders are used to remove surface films from discs and to enhance heat dissipation on the rubbing surface (Jacko and Rhee 1992). Zinc and aluminum chips are also known to contribute to recovery from the fade at high temperatures. Friction dust from cashew nutshell liquid and ground rubber are typical non-abrasive modifiers. They are added to obtain friction effectiveness at low and medium pressure and also to reduce noise propensity (Nicholson 1995). However, the cashew dust and rubber particles are prone to aggravate the brake fade at elevated temperatures. Calcium hydroxide is used as a rust inhibitor since it improves corrosion resistance by changing the pH of the friction material. Calcium hydroxide is also known to absorb gases during curing and as a useful additive to reduce swell (Nicholson 1995).

Inexpensive inorganic materials such as limestone ($CaCO_3$) and barite ($BaSO_4$) are included in the friction material as a space filler for cost reduction. Potassium titanate also can be considered a space filler, although a whisker type is often classified as a reinforcing fiber. Strictly speaking, fillers are not inert because their relative contribution to friction and wear is determined by the properties of other ingredients in the friction material. Morphology and particle size of the fillers also change the tribological properties of the friction material. A tribological study of barite-containing friction material

showed a strong negative μ-v relationship, suggesting higher propensity of creep groan (self-excited brake vibration in the range of 200–500 Hz), while the same composite with potassium titanate showed a slightly positive slope (Kim et al. 2004). On the other hand, calcium carbonate has been used as a cheaper alternative to barite. Although the calcium carbonate is inferior to barite in terms of thermal stability, it is still accepted for normal applications.

Environmental Concerns

Health issues concerning the ingredients in the brake friction material began in 1980 when asbestos in the workplace was addressed. After asbestos was classified as a potential carcinogen, most vehicle manufacturers allowed no asbestos in vehicle components and the asbestos ban in the friction material significantly affected their formulation. Various organic and inorganic substitutes were suggested for non-asbestos friction materials that were not classified as carcinogens by IARC (International Agency for Research on Cancer).

Issues around environmentally hazardous ingredients in brake friction material have been addressed again by the EEC (EU regulation) and EPA (Environmental Protection Agency). New concerns have focused on the refractory ceramic fibers containing less than or equal to 18 wt.% of alkaline oxide and alkali earth oxide ($Na_2O+K_2O+CaO+MgO+BaO$) (Directive 2001-41-EC), heavy metals (cadmium, hexavalent chromium, lead and lead compounds, and mercury and mercury compounds), and other health hazard inorganic/organic compounds containing antimony, carbon black, and respirable fibers (GADL lists 2010).

Copper in the brake friction material also has been blamed for the contamination of storm water runoff near metropolitan areas. The regulation of the maximum content in brake friction material was established by the states of Washington (SB346) and California (SSB6557) in 2010. Lead and antimony compounds also attract attention because they are classified as carcinogenic substances. In particular, growing concern has arisen over antimony trisulfide, which has not been classified as a carcinogen. This is because antimony trisulfide changes into carcinogenic antimony oxides when it is exposed to the high temperatures over 300°C during brake applications.

Fibrous ceramic materials have also been suspected as potential human carcinogens. This is because the wear debris from these fibers can be changed to acicular form when airborne during sliding, even if the original fiber is safe. In particular, respirable acicular ceramic fibers such as potassium titanate whiskers are likely to undergo regulation. One study showed that the platy potassium titanate (aspect ratio<1) improved wear resistance with slight degradation of high-temperature properties compared to the whisker-type potassium titanate. However, the synergistic reinforcing capability with other fibrillated fibers was not achieved (Cho et al. 2008a).

Environmental concerns regarding the ingredients in brake friction material have extended from the work environment in the manufacturing plant to public health due to air and water contamination from wear debris produced during brake applications. These concerns around public health and environmental pollution by brake friction material may radically alter the formulation of these materials and significantly impact raw material markets.

Manufacturing Process

The manufacturing process of brake friction material comprises mixing, preforming, hot molding, and post-curing. Although the detailed process is slightly different according to type of ingredients used in the friction material, the following is a brief summary of the manufacturing process for brake pads, linings, and segments:

Brake pads: The mix for NAO-type friction materials requires an intensive mixer to open fibrous ingredients, while the semi-metallic type needs a less intensive blender. The mix is preformed at room temperature and 30–40 MPa of pressure. Then, the preform is hot pressed in a mold at 160–180°C for 8–15 min at 30–50 MPa. During the hot molding, a gas release is carried out several times to remove the reactant gas. Then, the pads are post-cured at 200–250°C for 5–10 h. Depending on the type of ingredients or manufacturing method, the preform can be omitted.

Linings: Linings are curved friction materials used in a drum brake. They are produced from a mix of fibrous reinforcements and friction modifiers in a liquid resin blended at approximately 50°C. During the extrusion process, the mix is heated to 80–100°C and pressed into a flat at 15–30 MPa. In this stage, the mix is fed between two rolls of slightly different speeds for fiber alignment. The green linings are then formed into a segment at 150°C and cured in a curved mold for 4–8 h at 180–260°C.

Segments: Segments are brake friction materials used in drum brakes for buses and medium-sized trucks. They are produced by preforming a green segment from a dry mix at 4–5 MPa. The preforms are then hot pressed for 4–10 min at 150–190°C. At this stage, the preforms are partially cured and becomes thermoplastic so that they can be easily bent. They are cured in a curved mold for 4–8 h at 210–290°C.

Performance Related Important Issues

Overall performance of a brake friction material is determined by various aspects of tribological properties. This is because the brake friction material has to maintain its friction effectiveness in wide ranges of temperature, pressure, and speed, and in different environments. The following list addresses the issues that most concern vehicle manufacturers and drivers during the service of a brake friction material.

Effectiveness: This indicates the stopping efficiency of a friction material and can be obtained as the coefficient of friction after a designated burnish process (normally after 200 burnish stops using a full-size inertial dynamometer according to Federal Motor Vehicle Safety Standard [FMVSS] tests).

Fade: Brake fade refers to a loss of brake effectiveness at elevated temperatures after excessive brake applications. The root cause of thermal fade is attributed to the decomposition of the binder resin and other organic ingredients at high temperatures, which disintegrates the ingredients in the multiphase composite under mechanical and thermal stresses.

Brake noise: Noise, vibration, and harshness (NVH) produced during brake applications are closely associated with friction-induced surface excitation (friction instability) during sliding and the subsequent system response of the brake assembly. The noise is produced in different frequencies and their root causes and preventive methodology are known to be different. In general, the friction materials with the high friction coefficient show high propensity of brake noise generation (Bergman et al. 1999; Eriksson et al. 1999). The low-frequency vibration called judder (or roughness) is caused by uneven disc thickness, and creep groan is related to the stick-slip behavior of the friction material. High-frequency noise is attributed to the sprag-slip (Spurr 1961) behavior of a sliding couple (Crolla and Lang 1991).

Aggressiveness: A brake friction material with high friction effectiveness often produces severe disc wear. This is because the ingredients such as abrasives, metals, and ceramic fibers used to improve the stopping efficiency can aggravate the counter disc. The excessive disc wear often produces uneven disc thickness variation (DTV) causing cold judder. However, the DTV generation during service is also affected by the material properties of the gray iron disc, which are determined by its composition and production control (Jang et al. 2003).

Wear: Wear rate of a friction material is an important property because vehicle manufacturers guarantee product life for a certain period of time. Although the wear of a composite is determined by its constituents, the thermal decomposition of the binder resin (Hong et al. 2009) and the hard reinforcing fibers (Eriksson et al. 2002) are among the most important factors.

Evaluation of the Friction Material

The evaluation of brake friction material has long been considered a difficult task. This is because of the poor data repeatability from different tests and insufficient correlation between simulated test results and actual brake performance in a vehicle. Much effort has been given to the development of test machines and test procedures to achieve analogous friction characteristics compared with the results from a real vehicle test. Various testers are currently used to analyze the tribological property of the friction material. They can be classified as laboratory-scale tribometers, drag mode (Krauss type) tribotesters, small-scale inertial brake dynamometers, full-scale inertial brake dynamometers, and chassis dynamometers. Each tester has its own development history of emphasizing certain tribological properties or assessing overall performance of the friction material.

Laboratory-scale tribometers are mainly used to measure detailed tribological properties of a friction material and are widely used in universities and research labs. They are sensitive to test parameters such as sliding speed, pressure, and temperature and are known as a good assessment tool to obtain the role of ingredients in the friction material. The lab-scale tester has also been used to examine in situ interfacial contact variation during sliding, which is closely related to the microscopic understanding of the friction and wear of the brake friction material. However, the lab-scale tribometer normally uses small size specimens so that the test result often deviates from the test result from a full size dynamometer or a vehicles test results. This is because the small size sample can contain particular ingredients segregated locally on the sliding surface, exhibiting different distribution of contact plateaus. At high temperatures, in particular, the small size specimen often shows dissimilar test results compared with the results from a full-scale dynamometer or in-vehicle test results inasmuch as the heat accumulation in the friction material is different compared to the full size specimen.

Drag-mode tribotesters, often called Krauss-type machines, use relatively large specimens compared with the lab-scale tribometer to avoid errors from local ingredient segregation in the friction material. Several commercial testers are available and they have been used to obtain friction properties in an early stage of the formulation development and for screening the commercial products. One of the first drag mode testers was the FAST

(friction assessment screening test) machine, developed by Ford Motor Company for quality assurance of the friction material. This tester has been used to measure wear rate and fade phenomena at elevated temperatures in a constant drag mode. However, the correlation of the test data with a full-scale dynamometer is poor. Another type of drag-mode tester is the Chase machine, which has been used to determine the edge code of commercial friction materials following the test mode SAE J661. Contrary to the FAST machine, it uses a rotating drum with a pad loaded against the inner surface of a drum. However, the chase machine also has shown discrepancies in the test results when the data have been compared with full-scale inertial dynamometers.

Recently, dynamometer manufacturers have developed inertial type scale dynamometers to overcome inherent drawbacks of the drag mode tribotester. Two different types are commercially available. One of them uses a small size caliper and the other type employs a pad-on-disc type set-up using two pieces of specimen pressed against a rotating disc without a caliper. The scale ratio is approximately one fifth of the full-size pads and the attached inertial moment is also reduced with the same ratio. The scale dynamometer has an advantage of isolating the tribological characteristics of a friction material without serious influence from the caliper assembly since they normally adopt a rigid caliper design. A test procedure such as JASO 406 is available for a scale dynamometer to obtain various performance of the friction material.

The full-scale inertial dynamometer has long been accepted as a standard tester for the development and verification of the commercial friction materials, when it is properly instrumented and operated. It uses a real fixture from a vehicle and is equipped with an inertia that can simulate real stops and drags of a vehicle at a wide range of initial temperature, speeds, and deceleration. Compared with other testers, well-developed test procedures such as FMVSS (Federal Motor Vehicle Safety Standard) tests and JASO (Japanese Association Standards Organization) tests are available and repeatability of the test results is better than the above-mentioned smaller testers. Currently, full-scale inertial dynamometers are well accepted as standard tools to simulate in-vehicle tests, whereas the test procedures and the dynamometer design concept have been continuously improved to duplicate the brake performance of a vehicle.

Key Applications

Automotive brake systems
Brake systems for trains
Aircraft brake systems
Elevator
Wind power generator
Home electronic appliances

Acknowledgments

The author is grateful to Dr. S.J. Kim of Hyundai Motor Company, Dr. M.H. Cho at Hyundai MOBIS, and Dr. K.H. Cho at Mando Corporation for their helpful discussion. The figures in this article were provided by Hyundai Motor Company.

Cross-References

► Band Clutch or Brake
► Cone Clutches and Brake
► Contacts Considering Adhesion
► Friction Modifiers
► Friction Brakes for Automotive and Aircraft

References

F. Bergman, M. Eriksson, S. Jacobson, Influence of disc topography on generation of brake squeal. Wear **225**, 621–633 (1999)

M.E. Campbell, J.B. Loser, E. Sneegas, Solid lubricants, Technology Survey NASA SP-5059 (1966)

M.H. Cho, J.J. Lee, S.J. Kim, H. Jang, Tribological properties of solid lubricants (graphite, Sb_2S_3, MoS_2) for automotive brake friction materials. Wear **260**, 855–860 (2006)

K.H. Cho, M.H. Cho, S.J. Kim, H. Jang, Tribological properties of potassium titanate in the brake friction material; Morphological effects. Tribol. Lett. **32**, 59–66 (2008a)

K.H. Cho, H. Jang, Y.S. Hong, S.J. Kim, R.H. Basch, J.W. Fash, The size effect of zircon particles on the friction characteristics of brake lining materials. Wear **264**, 291–297 (2008b)

D.A. Crolla, A.M. Lang, Brake noise and vibration-the state of art, in *Vehicle Tribology*, ed. by D. Dowson, C.M. Taylor, M. Godet. Tribology Series, vol. 18 (Leeds University, Leeds, 1991), pp. 165–174

Directive 2001-41-EC of the European Parliament and of the Council of 19 June 2001

M. Eriksson, F. Bergman, S. Jacobson, Surface characterization of brake pads after running under silent and squealing conditions. Wear **232**, 163–167 (1999)

M. Eriksson, F. Bergman, S. Jacobson, On the nature of tribological contact in automotive brakes. Wear **252**, 26–36 (2002)

Global Automotive Stakeholder Group (GASG), Global automotive declarable substance (GADS) list (2010)

U.S. Hong, S.L. Jung, K.H. Cho, M.H. Cho, S.J. Kim, H. Jang, Wear mechanism of multiphase friction materials with different phenolic resin matrices. Wear **266**, 739–744 (2009)

M.G. Jacko, S.K. Rhee, Brake linings and clutch facings, in *Kirk-Othmer Encyclopedia of Chemical Technology*, ed. by K. Othmer, vol. 4, 4th edn. (Wiley, Hoboken, 1992), pp. 525–536

H. Jang, S.J. Kim, The effect of antimony trisulphide and zirconium silicate in the automotive brake friction material on friction characteristics. Wear **239**, 229–236 (2000)

H. Jang, J.H. Yoon, S.J. Kim, J.Y. Lee, H.D. Park, The effect of the composition and microstructure of gray cast iron on preferential wear during parasitic drag and on intrinsic damping capacity, SAE Technical Papers 2003-01-3313 (2003), pp. 57–64

H. Jang, K. Ko, S.J. Kim, R.H. Basch, J.W. Fash, The effect of metal fibers on the friction performance of automotive brake friction materials. Wear **256**, 406–414 (2004)
S.J. Kim, H. Jang, Friction and wear of friction materials containing two different phenolic resins reinforced with aramid pulp. Tribol. Int. **33**, 477–484 (2000)
S.J. Kim, M.H. Cho, D.S. Lim, H. Jang, Synergistic effect of aramid pulp and potassium titanate whiskers in the automotive friction material. Wear **251**, 1484–1491 (2001)
S.J. Kim, M.H. Cho, R.H. Basch, J.W. Fash, H. Jang, Tribological properties of polymer composites containing barite and potassium titanate. Tribol. Lett. **17**, 655–661 (2004)
A. Knop, W. Scheib, *Chemistry and Application of Phenolic Resins* (Springer, Heidelberg, 1979)
E.J. Lee, H.J. Hwang, W.G. Lee, K.H. Cho, H. Jang, Morphology and toughness of abrasives particles and their effects on the friction and wear of friction materials: a case study with zircon and quartz. Tribol. Lett. **37**, 637–644 (2010)
G. Nicholson, *Facts About Friction* (P&W Price Enterprises, Croydon, 1995)
R.P. Nolan, A.M. Langer, in *Proceedings of Fibers in Friction Materials Symposium*, The Asbestos Institute, vol. 75 (Quebec, 1987)
R.T. Spurr, A theory of brake squeal. Proc. Instn. Mech. Eng. 33–52, (1961–1962)

Brake Lining

► Brake Friction Materials

Brake Pad

► Brake Friction Materials

Brake Segment

► Brake Friction Materials

Break-in

► Running-in

Break-In of Gear Surfaces

► Gear EHL Film Thickness and Wear Risk Analysis

Bright Field (BF)

► Characterization of Microstructures by Transmission Electron Microscopy

Bristle Seal

► Brush Seals

Brittle Fracture Theory

► Griffith Theory of Fracture

Brittle-Ductile Transition

K. Jimmy Hsia
Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Synonyms

BCC – body center cubic; BDT – brittle-ductile transition; BDTT – brittle-ductile transition temperature; FCC – face center cubic

Definition

Brittle-ductile transition refers to the phenomenon where the fracture mode changes between low energy-absorbing mode and high energy-absorbing mode in solid materials.

Scientific Fundamentals

Basic Phenomenon

Fracture in structures occurs when a crack propagates through the material. During fracture, a certain amount of energy is absorbed (dissipated). The amount of energy absorption during fracture, Γ, is a material property referred to as the *fracture energy* or *fracture toughness*. When the value of Γ is low, the material is considered to be *brittle*; when Γ is high, the material is *ductile*. For example, many ceramic materials such as glass are

brittle with low fracture energy, and many metals such as copper and gold are ductile with high fracture energy. The corresponding fracture modes are referred to as *brittle fracture* and *ductile fracture*, respectively. Upon brittle fracture, the fracture energy of a material can be given by the following expression:

$$\Gamma = \int_0^{\delta_c} \sigma(\delta) d\delta = 2\gamma_0 \tag{1}$$

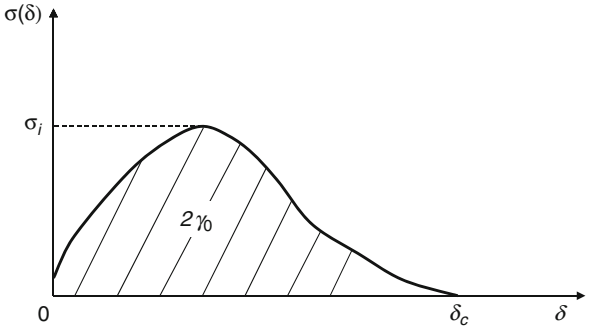
where δ is the separation distance between the fracture surfaces along the crack and $\sigma(\delta)$ is the stress between the two crack surfaces at the given separation distance δ. This function $\sigma(\delta)$ is called the *stress-separation curve*, which usually takes the form shown in Fig. 1, where δ_c is the critical separation when the separation stress is zero, that is, the two surfaces are completely separated and σ_i is the *ideal cohesive strength* of the material. For brittle fracture, the integral in (1) is equal to the surface energy, γ_0, of the two newly created fracture surfaces. A typical value of the intrinsic surface energy is ~ 1 J/m^2.

Upon ductile fracture, however, the energy absorption is several orders of magnitude larger than γ_0. Such high energy absorption during ductile fracture is a result of many different microscopic mechanisms such as plastic deformation, void nucleation and growth, and so on, as categorized by Ashby et al. (1979). Irwin (1948) introduced an additional term, γ_p, in (1) to account for the higher fracture energy during ductile fracture in metals, as

$$\Gamma = 2\left(\gamma_0 + \gamma_p\right) \tag{2}$$

The value of γ_p is usually a few orders of magnitude higher than that of γ_0.

By measuring the fracture energy, one can determine whether the fracture mode is brittle or ductile. It was found that the fracture mode depends on materials and loading rates. For a given material at a fixed loading rate (strain rate), however, many materials exhibit strong dependence of fracture mode on the temperature during the measurement. At low temperatures, the fracture mode is brittle; whereas at high temperatures, the fracture mode becomes ductile. This transition, as temperature increases, from brittle fracture mode to ductile fracture mode is called *brittle-ductile transition* (BDT). In many materials, particularly in certain metallic alloys, the BDT is rather sharp, often occurring within a few degrees Celsius. Such a temperature is thus referred to as the *brittle-ductile transition temperature* or BDTT.



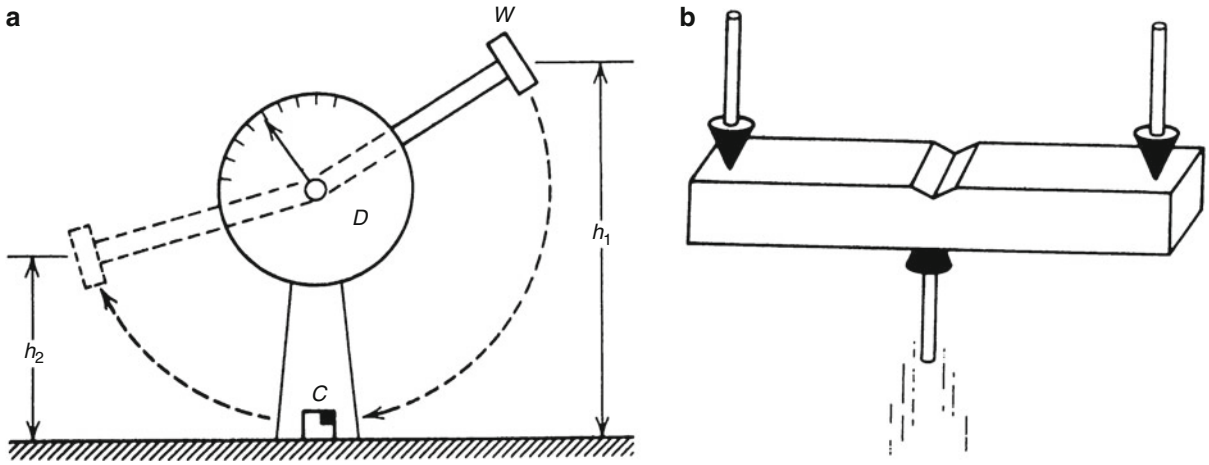
Brittle-Ductile Transition, Fig. 1 Typical stress-separation curve for a solid material

Testing Methods

The Charpy Impact Test. The *Charpy impact test* has been used as a standard test to study the temperature-dependent BDT. The test was named after the French scientist Georges Charpy, who developed the test in 1905. Due to its ease of use, it is still widely used to study BDT in structural materials. The Charpy impact testing machine consists of a pendulum with an impact hammer at the end, as shown in Fig. 2; Figure 2a shows the schematics of the testing machine and Fig. 2b shows the specimen geometry. The testing involves raising the hammer to a certain height h_1 then releasing it. Upon hitting the specimen, the pendulum swings to a height h_2. The height difference h_2–h_1 is measured and used to calculate the amount of energy absorbed during fracture. The detailed requirements for a Charpy impact test can be found in ASTM Standard E23–07ae1 (2007).

Instrumented Impact Test. A more quantitative method is the instrumented impact experiment. Common equipment for performing such an experiment is the Dynatup drop tower, shown schematically in Fig. 3. A critical component of the equipment is a strain-gauged, hardened load cell called the *tup*, which is attached to the falling weight and is in contact with the specimen throughout the impact process. The tup impacts the center of a plate specimen (Fig. 3b) and, given enough driving force, penetrates the specimen and causes fracture. While the velocity of the weight block is continuously monitored by an LVDT sensor, the force at the tip of the tup is continuously measured and recorded. With the measurements of the velocity and the force, one can calculate the total energy absorption during fracture.

Typical results of the testing are illustrated schematically in Fig. 4, which shows the energy absorption during fracture of a specimen as a function of temperature.



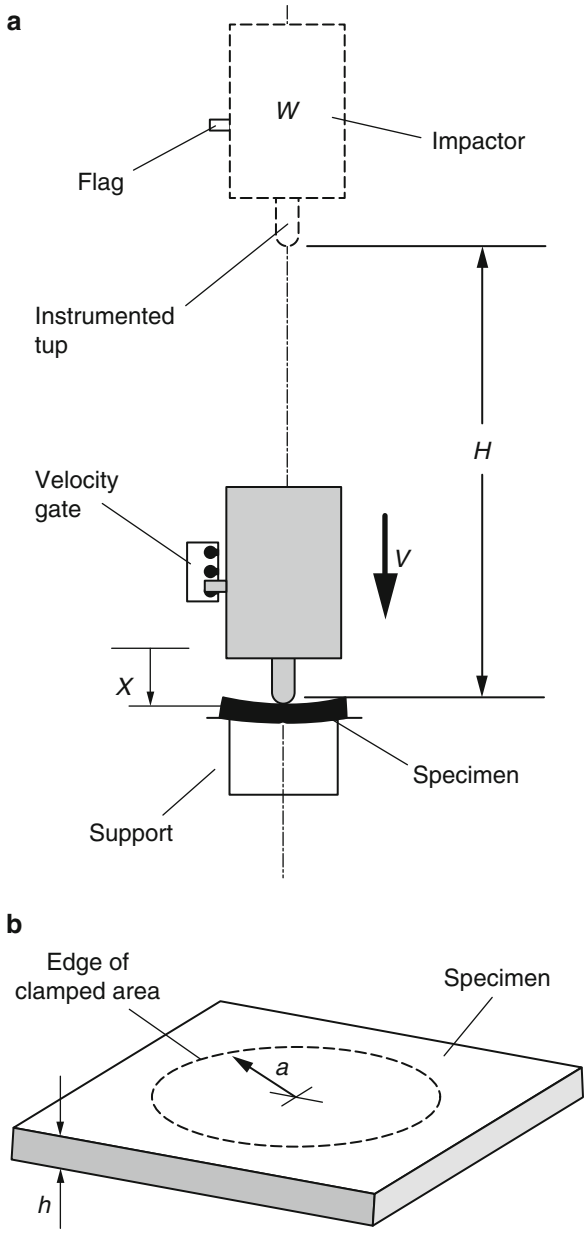
Brittle-Ductile Transition, Fig. 2 (**a**) Schematics of a Charpy impact testing machine, and (**b**) typical impact specimen geometry

Generally, at low temperatures the fracture energy is low, and vice versa. The temperature (or temperature range) at which the fracture mode changes from low energy absorbing mode to high energy absorbing mode is the *brittle-ductile transition temperature* (BDTT). Clearly, different materials exhibit different BDT behavior. Most face-center-cubic (FCC) metals are intrinsically ductile, with high energy absorption during fracture and, when BDT does occur (under special circumstances), exhibit gradual BDT behavior without a clearly defined BDTT. Most body-center-cubic (BCC) metals exhibit relatively sharp BDT behavior with well-defined BDTT. The sharpness of the BDT is strongly influenced by the physical mechanisms (discussed below) by which the BDT occurs. Even in BCC metals and alloys, the sharp BDTT is influenced by many factors, such as variations in their composition and their microstructures (e.g., due to different heat treatments). Furthermore, one extremely important factor controlling the BDTT is the strain rate during loading. The BDTT increases with increasing strain rate, as demonstrated both experimentally (Samuels and Roberts 1989) and through numerical simulations at the continuum level (Nitzsche and Hsia 1994) and discrete dislocation level (Ferney and Hsia 1999).

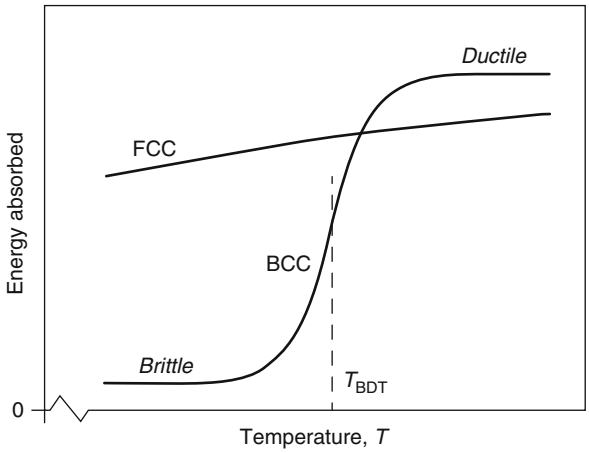
Mechanisms: Crystalline Materials

In crystalline materials, the BDT is controlled by competition between two processes at the crack tip – cleavage fracture and dislocation activities (Rice and Thomson 1974; Argon 1987; Brede et al. 1991). Cleavage fracture occurs on crystallographic planes with the lowest separation energy, characterized by (1). It is a low energy-absorbing mode giving rise to brittle fracture behavior. Dislocation processes at the crack tip, on the other hand, involve two successive steps: dislocation nucleation and dislocation motion. Although dislocation nucleation is a necessary condition for further dislocation activities, it is the dislocation motion and multiplications that dissipate the energy, thus significantly increasing the energy absorption during fracture. Nonetheless, depending on the relative difficulty of dislocation nucleation and dislocation motion, and depending on whether there are ample pre-existing dislocations in the material, the BDT may be dislocation nucleation controlled or dislocation mobility controlled. Dislocation nucleation controlled BDT and dislocation mobility controlled BDT have different characteristics.

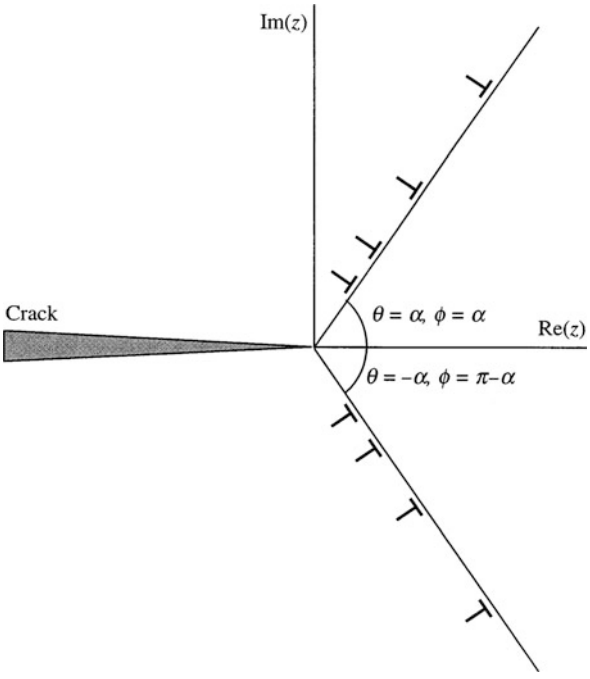
Intrinsically Brittle or Ductile Materials. Rice and Thomson (1974) analyzed the competition between dislocation emission and cleavage fracture at a crack tip by considering the configuration shown in Fig. 5. Their results show that materials can be categorized into two distinct groups: intrinsically brittle materials and intrinsically ductile materials. Ceramic material and semiconductors such as Si are intrinsically brittle. Cleavage fracture occurs in these materials before the conditions for crack tip dislocation emission are met. On the other hand, almost all FCC metals are intrinsically ductile, with dislocation emission significantly easier than cleavage. As a result, fracture in these materials often involves profound plastic deformation through dislocation activities, giving rise to cavity growth and coalescence along the fracture plane. When there is extensive plastic deformation near the crack tip, the mental picture in Fig. 5 is no longer valid. One must take into account the interactions among dislocations in many different slip systems.



Brittle-Ductile Transition, Fig. 3 (**a**) Schematics of a Dynatup drop tower testing machine, and (**b**) the geometry of a specimen

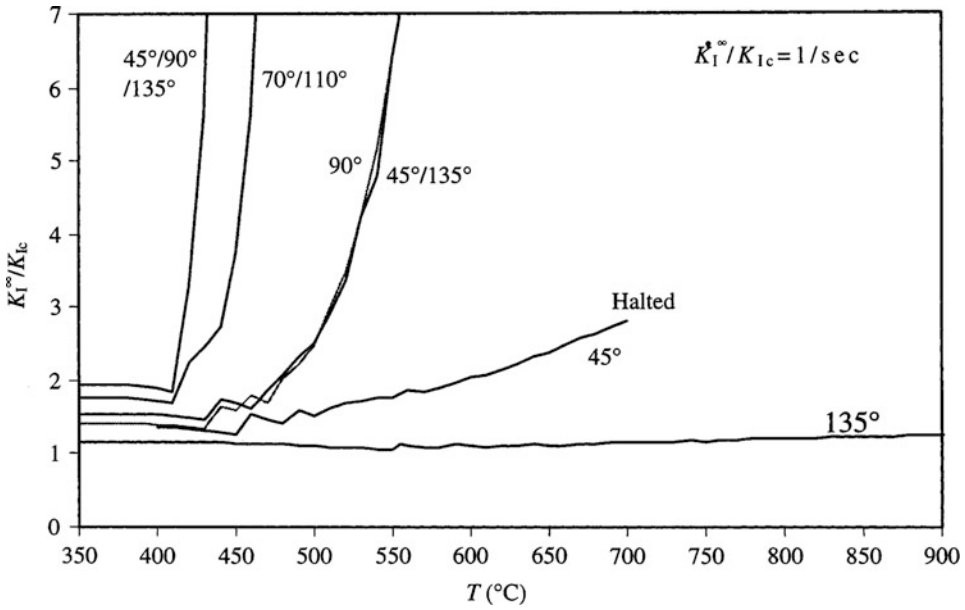


Brittle-Ductile Transition, Fig. 4 Typical measurement results of the impact experiments showing the BDT of a material at a given strain rate



Brittle-Ductile Transition, Fig. 5 A model configuration for analyzing the competition between cleavage fracture and dislocation activities near the crack tip

Dislocation Nucleation versus Dislocation Mobility. One factor affecting the sharpness of the BDT is whether it is dislocation nucleation controlled or dislocation mobility controlled. In a material with very low density of pre-existing dislocations such as Si, the more difficult process will become rating limiting. For example, if dislocation motion is much easier than dislocation nucleation and occurs at much lower temperature, then as soon as the condition for dislocation nucleation is satisfied at a given temperature, dislocations can readily move away from the crack tip, giving rise to crack tip shielding and the BDT (Xin and Hsia 1996). Such an avalanche of dislocation



Brittle-Ductile Transition, Fig. 6 Simulation results of the BDT behavior in Si single crystals demonstrating the importance of orientation and number of available slip systems at the crack tip in determining the sharpness of the BDT (Adapted from Ferney and Hsia 1999)

activities soon after dislocation nucleation results in a very sharp BDTT right at the temperature of dislocation nucleation. On the other hand, if dislocation nucleation is easier than dislocation motion and occurs at lower temperatures than BDTT, there will always be dislocations available near the crack tip at temperatures near or even below the BDTT. Thus, there would be some mild shielding effects to the crack tip even at temperatures below the BDTT and slight elevation of the fracture energy. Dislocation mobility increases gradually with the increase in temperature, and the shielding to the crack tip thus becomes stronger with the increasing temperature. However, since there are always some dislocations available near the crack tip, such enhancement of the shielding effect is gradual rather than abrupt. Based on this argument, it is commonly agreed that the dislocation mobility controlled BDT is gradual.

To explain the extremely sharp BDT observed in Si single crystals (Samuels and Roberts 1989), several theoretical models have been developed. One model is based on the concept of dislocation loop emission along the crack front (Hirsch et al. 1989). They envision that dislocation loops emitting from the crack tip would expand along the crack front until they encounter the specimen surface. This model predicts rather sharp BDT in materials with low density of pre-existing dislocations. Another model is based on the assumption of homogeneous dislocation loop formation within the crystal near the crack tip (Khantha et al. 1994). This model predicts extremely sharp BDT. Although these models provide plausible explanations to the sharp BDT behavior in semiconductors such as Si single crystals, no conclusive experimental evidence has been provided to either validate or invalidate the models.

Other Factors Affecting the BDT. Several other factors influence the BDT behavior of crystalline materials. Besides the controlling dislocation process during the BDT, the sharpness of the transition is affected by the orientation of the crack plane relative to the crystallographic orientation, and by the number of available slip systems at the crack tip (Ferney and Hsia 1999). Using dislocation simulations, Ferney and Hsia (1999) demonstrated that, for the same material under the same loading rate, the BDT behavior could vary from extremely sharp to extremely gradual, depending on which and how many slip systems are active, as shown in Fig. 6. In Fig. 6, the labels of the curves represent the inclination angles of the slip planes relative to the crack plane. Figure 6 shows that the shielding effectiveness of the dislocation activities strongly depends on the orientation of the slip plane. For example, dislocations on the slip planes oriented 135° from the horizontal plane ahead of crack are ineffective in shielding the crack tip, barely raising the fracture energy; while dislocations on slip planes 90° from the horizontal plane are highly effective. In fact, dislocations

on one pair of $\pm 90^\circ$ planes are as effective as those on two pairs of $\pm 45^\circ/\pm 135^\circ$ planes in shielding the crack tip. These results indicate that the local crystallographic orientation near the crack tip plays an important role in the BDT behavior.

Another factor influencing the BDT behavior is the geometrical confinement to dislocation motion. It is known that, when dislocation motion is restricted by obstacles such as grain boundaries, the material's yield strength increases. Meanwhile, the material may become more brittle, fracturing at lower strain level. This factor is especially important for intrinsically ductile materials such as FCC metals in a confined geometry. Under the extreme circumstance when dislocation motion is restricted by obstacles with a spacing of nanoscale, even the most ductile materials may exhibit brittle fracture behavior. Such conditions were analyzed by Hsia et al. (1994), who showed that, in a layered material with alternating metallic and ceramic layers, FCC ductile metals such as copper and aluminum may undergo cleavage fracture with very low energy absorption, even when the volume faction of the metallic phase is significant. It is cautioned, however, that such BDT behavior is not an intrinsic material property, but dependent on material systems and geometric arrangements of different phases.

Mechanisms: Amorphous Materials

Many amorphous materials, especially polymeric materials, also exhibit BDT behavior. Being absent of dislocation-related plastic deformation mechanisms, the BDT in amorphous materials is due to completely different mechanisms. Although thermo-plastic and thermo-set polymeric materials may undergo different physical processes, most exhibit a glass transition temperature T_g above which the Young's modulus of the material decreases quickly. The key phenomenon at T_g is that the material's viscoelastic deformation becomes more dominant compared with its linear elastic deformation. At temperatures significantly higher than T_g, however, the viscous flow becomes more profound in thermoplastic polymers, whereas the polymer hardens (or even burns) in thermoset polymers. It is this softening phenomenon that gives rise to the BDT behavior in polymeric materials. In a given material, if the softening (enhancement of viscous flow) changes gradually with increasing temperature, the BDT is gradual for this material; if the softening changes rapidly, a sharp BDT results. Similar mechanisms also give rise to BDT phenomena in glasses (including metallic glass) at very high temperatures, and are believed to be responsible for the BDT in crystalline ceramic materials in which a glassy intergranular phase exists (Dey et al. 1996). In addition to the viscoelastic deformation, there are other energy dissipation mechanisms in polymeric materials that contribute to the increase of fracture energy (thus BDT) during fracture. One such mechanism is crazing, that is, formation of microscopic crack-like defects called *crazes* ahead of a macroscopic crack tip. The diffused nature of these crazes and the internal microstructures within crazes all contribute to higher fracture energy in polymeric materials. Metallic glasses, another class of amorphous materials first discovered by Klement et al. (1960), exhibit brittle failure in tension at room temperature because of localized deformation such as shear band formation (see Ashby and Greer 2006). Achieving a brittle-to-ductile transition in metallic glasses under compressive loading is a contemporary research topic being actively pursued by researchers (see, e.g., Jang and Greer 2010) with, unfortunately, limited success to date.

Key Applications

Characterization of the BDT behavior in structural materials is crucial to avoiding catastrophic failures. Famous incidences of catastrophic failures include the Boston molasses tank rupture in January of 1919 and fracture of Liberty ships in northern Europe during World War II. It is even suspected that the failure of the *Titanic* (Broad 1993) is due to BDT of the steel plate and welding joints of the ship. The phenomenon of BDT, however, was not well understood at the time of these failures, therefore the materials' fracture properties at different temperatures were not properly characterized.

The main applications of the BDT concept are (1) to establish design criteria based on proper characterization of the BDT behavior of materials such that catastrophic, brittle fracture would not occur under operating conditions; and (2) to improve materials' BDT properties based on the fundamental understanding of the physical mechanisms. The first one addresses the safety issues in structural engineering, whereas the second deals with new materials development.

References

A.S. Argon, Brittle to ductile transition in cleavage fracture. Acta Metall. **35**, 185–196 (1987)

M.F. Ashby, A.L. Greer, Metallic glasses as structural materials. Scr. Mater. **54**, 321–326 (2006)

M.F. Ashby, C. Gandhi, D.M.R. Taplin, Overview No. 3: fracture mechanisms maps and their construction for FCC metals and alloys. Acta Metall. **27**, 699–729 (1979)

ASTM Standard E23-07ae1, *Standard Test Methods for Notched Bar Impact Testing of Metallic Materials* (ASTM International, West Conshohocken, 2007), doi:10.1520/E0023-07AE01, www.astm.org

M. Brede, K.J. Hsia, A.S. Argon, Brittle crack propagation in silicon single crystals. J. Appl. Phys. **70**, 758–771 (1991)
W.J. Broad, New idea on titanic sinking faults steel as main culprit, *New York Times*, 16 Sept 1993
N. Dey, D.F. Socie, K.J. Hsia, Tensile creep behavior of a vitreous-bonded aluminum oxide under static and cyclic loading. J. Am. Ceram. Soc. **79**, 2353–2363 (1996)
B.D. Ferney, K.J. Hsia, The influence of multiple slip systems on the brittle-ductile transition in silicon. Mater. Sci. Eng. **A272**, 422–430 (1999)
P.B. Hirsch, S.G. Roberts, J. Samuels, The brittle ductile transition in silicon, 2: interpretation. Proc. Roy. Soc. Lond. **A421**, 25–53 (1989)
K.J. Hsia, Z. Suo, W. Yang, Cleavage due to dislocation confinement in layered materials. J. Mech. Phys. Solid. **42**, 877–896 (1994)
G.R. Irwin, Fracture dynamics, in *Fracturing of Metals* (American Society for Metals, Cleveland, 1948), pp. 147–166
D. Jang, J.R. Greer, Transition from a strong-yet-brittle to a stronger-and-ductile state by size reduction of metallic glasses. Nat. Mater. **9**, 215–219 (2010)
M. Khantha, D.P. Pope, V. Vitek, The brittle-to-ductile transition, 1. A cooperative dislocation generation instability. Scripta Metall. Et. Mater. **31**, 1349–1354 (1994)
W. Klement, R.H. Willens, P.O.L. Duwez, Non-crystalline structure in solidified gold-silicon alloys. Nature **187**, 869–870 (1960)
V.R. Nitzsche, K.J. Hsia, Modeling of dislocation mobility controlled brittle-to-ductile transition. Mater. Sci. Eng. **A176**, 155–164 (1994)
J.R. Rice, R. Thomson, Ductile versus brittle behavior of crystals. Phil. Mag. **29**, 73–97 (1974)
J. Samuels, S.G. Roberts, The brittle ductile transition in silicon, 1: experiments. Proc. Roy. Soc. Lond. **A421**, 1–8 (1989)
Y. Xin, K.J. Hsia, A technique to generate straight through thickness surface cracks and its application to studying dislocation nucleation in Si. Acta Mater. **44**, 845–853 (1996)

Broaches

► Gear Cutting Tools

Brush and Hydration Lubrication (Natural Synovial Joints)

Jacob Klein
Department of Materials and Interfaces, Weizmann Institute of Science, Rehovot, Israel

Synonyms

Biolubrication; Brush boundary lubrication; Synovial lubrication

Definition

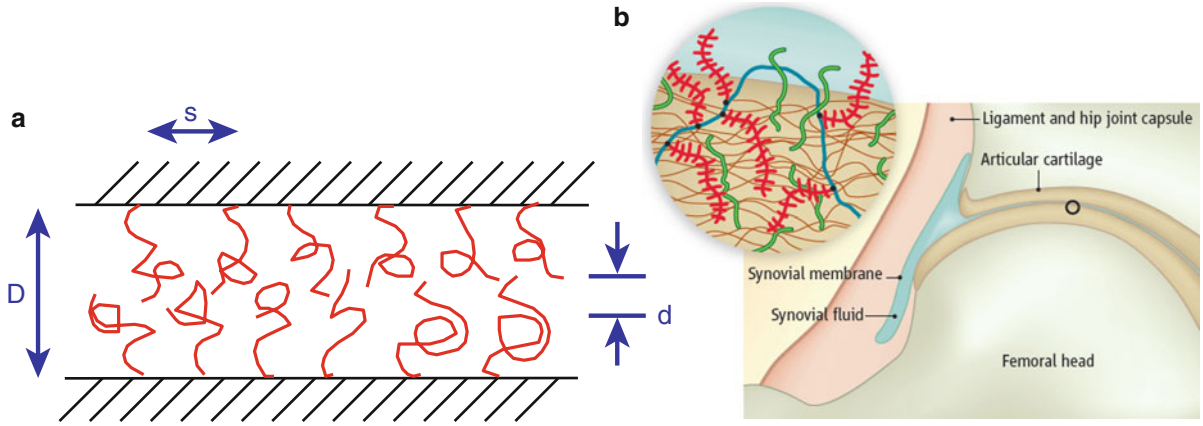
A polymer brush is a layer of non-adsorbing polymer molecules densely attached by one end only (or unusually, by both ends) to a non-adsorbing surface, generally in a good-solvent medium, so that they extend outward away from the surface (like the hairs on a brush). In the context of natural synovial lubrication, the macromolecular surface phase exposed at the outer surface of the articular cartilage is thought to behave like a polymer brush in some respects, including lubrication (Klein et al. 1994; Klein 2009). Brush lubrication is the process whereby polymer brushes reduce friction between sliding surfaces onto which they are attached. The detailed mechanism of brush lubrication in aqueous media is attributed to a combination of factors: These include the weak interpenetration between brush layers up to moderate compression (an effect arising from entropic factors), so that the interpenetration zone between them remains unentangled and thus fluid; the presence of counterions - in the case of charged brushes - that contribute to the osmotic pressure of the brush supporting the load between the surfaces; and at the highest compressions, to hydration layers about the monomers of the brush via the hydration lubrication mechanism. The overall result is that the brushes can sustain large pressures (up to 100 atm and more, comparable or higher than in the major synovial joints) and yet slide with little frictional dissipation past each other, resulting in very low friction coefficients.

Scientific Fundamentals

Chemical and Physical Principles

Two polymer brush layers repel each other when compressed, and may provide extremely efficient boundary lubrication even at high pressures, up to 10^7 N/m^2 (ca. 100 atm) or more, that are comparable or higher than those in the major synovial joints such as hips and knees. This has led to the suggestion that lubrication at the surface of the articular cartilage layers coating these joints be mediated by the macromolecules extending from the cartilage (Klein et al. 1994; Klein 2009), a correspondence indicated in Fig. 1.

Such macromolecules, which include hyaluronic acid (HA), aggrecans, and lubricins, are synthesized by the chondrocyte cells within the cartilage. They make their way by slow diffusion through the cartilage into the synovium, and as they cross the interface from the inside to the outside of the cartilage they are attached as a surface phase that has brush-like properties (schematically shown in the inset to Fig. 1b). A detailed consideration is given in (Klein 2006). In particular, such a brush-like layer may



Brush and Hydration Lubrication (Natural Synovial Joints), Fig. 1 (**a**) Schematic of two polymer brushes attached to solid substrates and compressed against each other. *s* is the mean spacing between the ends of the chains anchored to the substrate, while *d* is the extent of the interpenetration zone when they are compressed. (**b**) The macromolecules emanating from the outer surface of articular cartilage schematically indicated in the inset– such as hyaluronic acid (indicated as *blue*), aggrecans (*red*) or lubricins (*green*) may behave like the brush shown in **a**, especially as regards their lubrication properties (Taken from (Klein 2009))

lubricate in the same way as model polymer brushes, including bottle-brush molecules (Drobek and Spencer 2008): hence the relevance of friction between two compressed brushes (Fig. 1a) to natural joint lubrication (Fig. 1b).

Polymer brushes in a good solvent extend away from the substrate due to osmotic pressure between adjacent chains, such that their uncompressed thickness is given by (de Gennes 1987; Alexander 1977)

$$L \approx Na(a/s)^{2/3} \tag{1}$$

where N is the degree of polymerization of each chain, *a* is the monomer size (freely jointed chains are assumed), and s is the mean interanchor spacing (see Fig. 1a). This thickness varies linearly with N, as opposed to the end-to-end size R_0 of an unstretched chain in a good solvent which varies as $R_0 = N^{3/5}a$: thus, as N is often very large (10^2–10^4 monomers), chains in a brush may be highly stretched. The value of L is determined by the balance between the osmotic repulsion of the chains – which tends to stretch them – and their configurational entropy and consequent chain elasticity that opposes the stretching. For this reason, when two polymer brushes are compressed against each other, they first tend to compress within themselves rather than to interpenetrate, as this reduces the elastic stretching energy, while the osmotic repulsion energy increases by (roughly) the same amount whether interpenetrated or not. A more detailed treatment shows that when two such brushes are compressed to a gap thickness D, as in Fig. 1a, the extent of interpenetration d between them (Fig. 1a) varies only very weakly with the compression (Klein 1996):

$$d \sim D^{-1/3} \tag{2}$$

At the same time, when two such brush layers are compressed, the decrease in surface separation D results in an increased volume fraction ϕ (D) of the monomers in the gap (essentially $\phi \sim (1/D)$), and thus in an increase in the osmotic pressure due to excluded volume, Π_{exc} (ϕ). This varies with monomer volume fraction as (de Gennes 1988)

$$\Pi_{exc}(\phi) \approx \left(kT/a^3\right)\phi^2 \sim \left(1/D^2\right) \tag{3}$$

(ignoring scaling exponents) where k is Boltzmann's constant and T the absolute temperature. Thus, the brushes can support a large load as they are compressed, which is balanced by this osmotic pressure.

The above picture applies for neutral (or zwitterionic) chains. In synovial joints, as in biological systems generally, most biological macromolecules – in particular HA, aggrecans, and lubricins, which are implicated at the cartilage surface (Fig. 1b) (Klein 2006) – are polyelectrolytes and so are highly charged due to their ionization. Indeed, it is this ionization that enables their solubility in the aqueous medium permeating the cartilage and synovial fluid in the first place. This implies that the macromolecular surface phases at the cartilage interface consists of *charged* brushes, and that the brush layer itself is permeated by mobile counterions (to compensate the

charge on the ionized monomers). These counterions exert their own osmotic pressure, $\Pi_{c/i}$, which augments the osmotic pressure Π_{exc} due to the excluded volume of the interacting monomers, given in (3). It can be shown that, for charged brushes, the ratio between the two forms of osmotic pressure is given approximately by

$$(\Pi_{c/i}/\Pi_{exc}) \approx f/\phi \quad (4)$$

where f is the fraction of monomers ionized on each chain. Since for highly ionizable polymers (such as HA or aggrecans) the charged fraction f can readily be as high as 0.2–0.3, the contribution of counterions in supporting the load between the compressed brush-bearing surfaces can exceed that of the excluded volume effects as long as the volume fraction ϕ of monomers is not too high. The overall result is that even strong compression of surfaces bearing brushes, whether charged or otherwise, may indeed be supported by the osmotic pressures exerted within the brush layers – an important consideration when considering the high pressures in the major synovial joints.

The lubricating properties of brushes are also intimately connected to their structure. This is because when two compressed brush-bearing surfaces slide past each other, frictional dissipation occurs only within the sheared, interpenetrated (or overlapped) region of extent d (Fig. 1a). Because of the weak overlap d even at moderate compressions (up to pressures of ca. 5 atm), given in (2), the polymer moieties in the interpenetration region are expected to be short and thus unentangled – i.e., if the length of such a moiety is n(d), then one expects $n(d) < N_e$, the entanglement length of the polymers (beyond which the viscosity increases rapidly) (Klein 1996). If the effective viscosity within the overlap region is $\eta_{overlap}$ (d), then the frictional stress σ (d) arising from shear of this region as the surfaces slide past each other at velocity v_s is given by the Newtonian expression

$$\sigma(d) = \eta_{overlap}(d)[v_s/d] \quad (5)$$

where the term in square brackets corresponds to the shear rate within the overlap region. Since $\eta_{overlap}$ (d) is low – because the polymer moieties n(d) are short and unentangled – the frictional stress, and thus the friction between the sliding brushes, is also low. This is the origin of the low friction observed up to moderate compressions (up to ca. 5–10 atmospheres pressure) of brush-bearing surfaces. At much higher compressions, such that $D <$ ca. $2R_0$, one expects that interpenetration between the opposing layers will extend through most of the intersurface gap, so that the relationship between d and D (2) no longer holds. Under such high pressures the frictional dissipation is controlled both by the entanglements between the opposing brushes, and by the local monomer-monomer friction, and its magnitude varies with specific polymers. These two regimes are considered in more detail in a subsequent section (Key Research Findings). It is especially important to recall here a crucial difference between polymer brushes and adsorbed polymers: For adsorbed chains, in addition to the considerations of excluded volume and counter-ion osmotic pressures – which at strong compressions are similar for both brush-like and adsorbed layers – there is also the issue of the adsorption, which can lead to bridging effects (Klein and Luckham 1984). Such bridging occurs for adsorbed layers but not for brushes (as these by definition consist of chains that do not adsorb along their backbones, but only at their end). Bridging greatly increases the frictional dissipation between compressed, sliding, adsorbed polymer layers, as the bridging moieties from one brush layer are dragged across the opposite surface, and thus can very strongly increase the friction (Klein 2001).

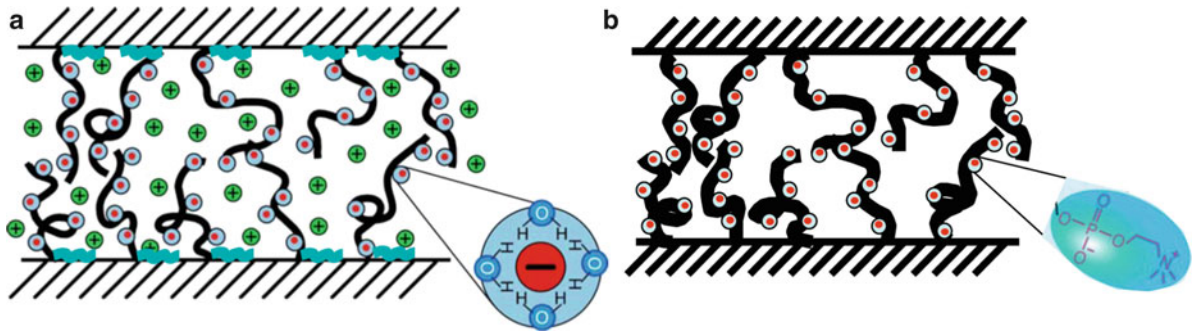
An important implication for the velocity dependence of the friction arises from the fact that the frictional dissipation, as brushes slide past each other, occurs within the overlapped region alone: For a given compression and value of D the region of overlap d becomes thinner at higher sliding velocities v_s. The value of d is determined by a so-called *self-regulation* requirement that the relaxation rate of the moieties n(d) equal the shear rate $[v_s/d]$ within the overlap region. It can be shown that this requirement leads to a very weak sliding velocity v_s dependence of the frictional stress σ (d), so that rather than the linear dependence given in (5), the effect of smaller overlap d at higher v_s leads to

$$\sigma(d) \sim v_s^{\alpha} \quad (6)$$

where the exponent $|\alpha| << 1$, and may be close to zero over a range of sliding velocities (Tadmor et al. 2003). This implies that the friction between rubbing polymer brushes depends only very weakly on the sliding velocity, a feature in common with solid-solid friction but arising from a very different mechanism.

Key Research Findings

Over recent years several new aspects of brush lubrication have been revealed. The early results were on neutral polymer brushes in good organic solvents (Klein et al. 1994), which first demonstrated the remarkable lubrication effect noted above, with effective friction coefficients μ as low as $\mu \approx 0.001$. A difficulty in these early experiments was that the model chains constituting the brushes were polystyrene molecules, which are glassy in the bulk at room temperature: Thus, as the brushes were compressed and the volume fraction ϕ of monomers between the



Brush and Hydration Lubrication (Natural Synovial Joints), Fig. 2 Schematic illustrations of (**a**) polyelectrolyte brushes, anchored by hydrophobic moieties (+ *charges, green*) adsorbed onto the hydrophobized substrate; the counterions (*green*) are indicated, as is the hydration layer about the negatively-charged monomers. (**b**) A polyzwitterionic brush covalently grown from initiation sites strongly bound to the substrate, showing the highly hydrated phosphocholine-like monomer structure. There are no counterions in this brush of overall-neutral zwitterionic chains

surfaces increased, the polymers approached their glass-like state and their viscosity diverged, due to this effect, already at quite low pressures. This was overcome in later studies by using polymers whose glass transition was low relative to the temperature of the experiments, and which consequently exhibited lubricating behavior characteristic of brushes over regimes from low to high compressions (Tadmor et al. 2003). Water-based systems are of more relevance to biological lubrication, and studies using neutral but water-soluble polymers such as poly(ethylene oxide), PEO (or synonymously, PEG, polyethylene glycol) brushes in aqueous media have examined the lubrication of both planar brush (Raviv et al. 2002) and bottle-brush (Drobek and Spencer 2008) configurations.

Macromolecules at the cartilage-synovium interface, including in particular those schematically illustrated in Fig. 1b, are quite highly charged ($f \approx 0.2$ or higher). Thus, insight into the question of synovial lubrication is more properly provided by the lubrication properties of charged brushes. Charged brushes in aqueous media have been examined, using a diblock copolymer with a hydrophobic anchoring moiety at one end adsorbed on a hydrophobized substrate, attached to a charged tail that extends out from the substrate to form the brush (Raviv et al. 2003). Such brushes are shown schematically in Fig. 2a, where the counterions permeating the brush layer are indicated, as are the hydration layers about the ionized monomers.

These charged brushes indeed provide excellent lubrication, with effective friction coefficients μ = ca. 0.001, comparable to those between healthy cartilage surfaces in hips or knees, but only up to quite moderate pressures of ca. 3 atm, very much lower than in the major synovial joints. At higher pressures the friction is sufficient to pull off the chains from the substrate, degrading and rapidly removing the brush: This happens because of the relatively weak attachment (via hydrophobic attraction) of the chains to the substrate in these experiments, together with the friction at the high pressures. Up to the detachment point, however, the efficient lubrication may be viewed as resulting from the non-interpenetration of the chains, up to moderate pressures. The load is borne both by excluded volume effects, contributing Π_{exc}, and by the counterions contributing $\Pi_{c/i}$. At higher compressions stronger interpenetration occurs and the lubrication provided by the hydration layers about the charged monomers (Fig. 2a) is likely to play a role. Nonetheless, the increased friction at the higher pressure is sufficient to remove the weakly attached chains, as noted.

To overcome this limitation, much more robust polymer brushes are required, and these may be grown covalently from initiator groups attached strongly to the surface. Lubrication by such grafted-from brushes was recently studied (Chen et al. 2009): in this case the chains consisted of poly[2- (methacryloyloxy)ethyl phosphorylcholine] (pMPC), and the monomers were zwitterionic rather than charged, and indeed each monomer had the structure of a phosphorylcholine (or phosphocholine) group as in the headgroups of phosphatidylcholine lipids ubiquitous in living systems. Such phosphocholine groups are known to be very strongly hydrated, with some 15 or so water molecules in the primary hydration shell, depending on the method of measurement. It was found that these robust, highly hydrated zwitterionic brushes provided extremely good lubrication, with μ = 0.001 or lower up to pressures of some 75 atm, comparable with those in human hips or knees. The origin of this very low

Brush and Hydration Lubrication (Natural Synovial Joints), Table 1 Some representative brush/good-solvent lubrication systems

Brush system	Pressures (atms)	Friction coefficient μ	Comments
Polystyrene/toluene (Klein et al. 1994)	Up to ca. 3	<ca. 0.001	At higher pressures the brushes become glassy and μ rises sharply
PEO bottle-brush/ buffer solution (Drobek and Spencer 2008)	Up to ca. 100	Ca. 0.003	At higher pressures μ rises sharply
Poly (ethylene propylene)/cyclohexane (Tadmor et al. 2003)	Ca. 10	~0.015	Substantial brush interpenetration, weak velocity dependence
(PMMA-b-PSGMA)/water (Raviv et al. 2003)	Up to ca. 3	<ca. 0.001	At higher pressures the weakly attached brushes are removed by shear
pMPC/water and aqueous salt solutions (Chen et al. 2009)	Up to 75	<ca. 0.001	Robust polyzwitterionic brushes; low friction may extend to even higher pressures

friction is of interest: at these high pressures the suppression of interpenetration of the opposing layers by entropic factors, which is effective up to moderate pressures and is described in the first part of this article, is unlikely to play a role. Likewise, there are no counterions to augment the osmotic pressure supporting the load between the brush-bearing surfaces, as is the case for brushes that are ionized and bear a net charge (Fig. 2a). The origin of the low friction at high pressures must lie elsewhere.

The very efficient lubrication is attributed to the high hydration of the phosphocholine-like monomers on the chains, and operates via the so-called hydration lubrication mechanism: According to this, hydration layers about charges in an aqueous medium may be very tenaciously held, as a result of the dipoles of the water molecules, which comprise them, orienting about the enclosed charge and strongly reducing its Born energy. At the same time, depending on the nature of the charge, these hydration layers may relax very rapidly, and so behave in a fluid-like manner under shear. This combination of being able to support a large normal load – due to the reluctance of the water molecules to be squeezed out – and at the same time to behave in a very fluid manner under shear, leads directly to a very low friction coefficient. This mechanism was first described in a study of the lubrication afforded by hydrated metal ions trapped between charged surfaces (Raviv and Klein 2002), and has since been observed in several different systems. In the case of the robust pMPC brushes (Chen et al. 2009) it results in lubrication efficiency that at these high pressures is the most efficient of all brush boundary lubricants studied so far. This is seen in Table 1, where some representative brush lubrication data are summarized. The high hydration of the monomers that are responsible for this effective friction reduction suggests that in natural synovial joints a major role must also be played by hydration lubrication, part of a boundary lubrication system that is based on the macromolecular surface phases shown schematically in Fig. 1b.

Molecular understanding of synovial joint lubrication along the lines described above is relatively recent, and complements the many models based on extension of engineering tribology concepts to natural joints (Klein 2006). Brush boundary lubrication, which is thought to arise naturally because of the macromolecular surface phases at the articular cartilage surface, has not to date, as far as we know, been implemented externally as a clinical treatment in living joints. Synthetic polymer brushes have, however, been used very effectively in reducing wear in artificial joints: A pMPC brush layer similar to that described above was used to coat the polyethylene acetabular of a prosthetic metal-head/plastic-cup hip implant, and resulted in very substantial reduction in wear, presumably as a result of local lubrication at asperity contacts (Moro et al. 2004). Future developments based on these concepts, in particular that of hydration lubrication, hold great promise for both repair and replacement strategies for synovial joints.

Acknowledgments

This work was supported by the Charles McCutchen Foundation, the European Research Council, and the Israel Science Foundation.

Cross-References

► Hydration Force
► Polymers in Biotribology

References

S. Alexander, Adsorption of chain molecules with a polar head a scaling description. J. Physique **38**, 983 (1977)
M. Chen, W.H. Briscoe, S.P. Armes, J. Klein, Lubrication at physiological pressures by polyzwitterionic brushes. Science **323**, 1698–1702 (2009)
P.G. de Gennes, Polymers at an interface: a simple view. Adv. Colloid Interface Sci. **27**, 189–209 (1987)
P.G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell, Ithaca, 1988)
T. Drobek, N.D. Spencer, Nanotribology of surface-grafted PEG layers in an aqueous environment. Langmuir **24**, 1484–1488 (2008)
J. Klein, Shear, friction, and lubrication forces between polymer-bearing surfaces. Annu. Rev. Mater. Sci. **26**, 581–612 (1996)
J. Klein, Interactions, friction and lubrication between polymer-bearing surfaces, in *Fundamentals of and Bridging the Gap Between Macro- and Micro/Nanoscale Tribology*, ed. by B. Bhushan (Kluwer, Dordrecht, 2001), pp. 177–198
J. Klein, Molecular mechanisms of synovial joint lubrication. Proc. Inst. Mech. Eng., Part J, J. Eng. Tribol. **220**, 691–710 (2006)
J. Klein, Repair or replacement: a joint perspective. Science **323**, 47–48 (2009)
J. Klein, P.F. Luckham, Forces between two adsorbed poly(ethylene oxide) layers in a good aqueous solvent in the range 0–150 nm. Macromolecules **17**, 1041–1048 (1984)
J. Klein, E. Kumacheva, D. Mahalu, D. Perahia, L. Fetters, Reduction of frictional forces between solid surfaces bearing polymer brushes. Nature **370**, 634–636 (1994)
T. Moro, Y. Takatori, K. Ishihara, T. Konno, Y. Takigawa, T. Matsushita, U.-I. Chung, K. Nakamura, H. Kawaguchi, Polymer grafting of artificial joints for preventing periprosthetic osteolysis. Nat. Mater. **3**, 829–836 (2004)
U. Raviv, J. Klein, Fluidity of bound hydration layers. Science **297**, 1540–1543 (2002)
U. Raviv, J. Frey, R. Sak, P. Laurat, R. Tadmor, J. Klein, Properties and interactions of physigrafted end-functionalized poly (ethylene glycol) layers. Langmuir **18**, 7482–7495 (2002)
U. Raviv, S. Giasson, N. Kampf, J.-F. Gohy, R. Jerome, J. Klein, Lubrication by charged polymers. Nature **425**, 163–165 (2003)
R. Tadmor, J. Janik, L.J. Fetters, J. Klein, Sliding friction with polymer brushes. Phys. Rev. Lett. **91**, 115503 (2003)

Brush Boundary Lubrication

► Brush and Hydration Lubrication (Natural Synovial Joints)

Brush Materials

Koichiro Sawa
Department of System Design Engineering, Keio University, Kohoku-ku, Yokohama, Japan

Synonyms

Contact materials for brushes; Sliding contact materials

Definition

Brush materials are those used for electrical brushes to achieve stable current flow between brush and commutator/slip ring. Precious metal brushes are used for micro motors. The carbon-type brushes are classified roughly into carbon type and metal type.

Scientific Fundamentals

Carbon-type brushes are predominantly used for the commutation of DC machines, except for micro DC motors (Shobert 1965; Wu et al. 1995). They are classified into carbon brushes and metal-graphite brushes as follows (Holm 1967; Slade 1999).

Carbon Brushes

Major materials of carbon brushes are lampblack, cokes, pitches, and tars.

Electrographite brushes: produced by a high-temperature process (2,200–2,800°C) of graphitization. They are mainly classified into soot type from lampblack, cokes type from coke or petroleum pitch coke, and electrographite type from natural graphite or artificial graphite. Electrographite grades have the best performance among carbon brushes, that is, they have the best commutating characteristics and the lowest friction coefficients. Generally, "carbon brushes" means electrographite grade brushes.

In addition, electrographite grades are classified into the following groups:

- Soot type is high in resistivity and high in contact voltage drop, so good in commutation.
- Cokes type is medium in resistivity and not so high in contact voltage drop.
- Graphite type is the lowest in resistivity and good in sliding motion.

Carbon-Graphite Brushes

The raw materials such as lampblack, cokes, pitches, and tars are mixed and blended under the same process controls with the electrographite grade. Then they are molded into plates or pills that are baked to the temperature of about 1,200°C to carbonize the binder. These earliest carbon brushes are stronger than the electrographite brushes and have a definite polishing action. Further, they have relatively high coefficient of friction and specific electrical resistivity.

Graphite brushes: Natural and artificial graphite are used in the production with a tar, pitch, or resin binder. The mixed materials are molded into plates or pills, and then cured by heating to about 1,200°C.

The graphite brush has a low coefficient of friction and provides good cleaning action.

Metal-Graphite Brushes

Metal powders and graphite are major raw materials. Metal powders are usually copper, and sometimes bronze or brass. Silver or gold are used as metal powders for special purpose brushes.

Manufacture of Carbon Brushes

Raw materials and production process depends on brush grades and manufacturers. The outline of production process is as follows (Shobert 1965; http://www.aupac.co.jp/e/index.html).

Raw materials:

(a) Soot or lampblack: This made from natural gas and pitch oil, and is used to produce commutation brushes.
(b) Pitch cokes and petroleum cokes: Pitch cokes are produced by carbonizing coal pitch, and petroleum cokes as sub-products of refining oil.
(c) Natural graphite: This is classified from the appearance into flake graphite and amorphous graphite. The former is produced in China and Sri Lanka, and the latter in China, North Korea, and Brazil.
(d) Tars and pitches: These are used in the manufacture process as a binder. Tars are components that are condensed as high boiling temperature gases emitted from carbonized coal. Pitches are remnants after removing distillates from tars.

Figure 1 shows a manufacture process of the carbon-graphite brush. Graphitization is carried out in special furnaces to control accurately temperatures of up to 3,000°C.

Manufacture of Metal-Graphite Brushes

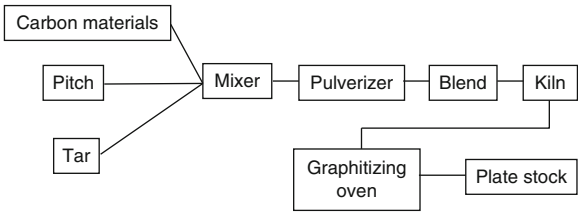
Main materials are copper powders and graphite powders, and silver for special purpose instead of copper (Wilsdorf et al. 1988; Reichner 1979). A small amount of lead, tin, or low-melting-point metals is sometimes added.

A manufacture process is shown in Fig. 2.

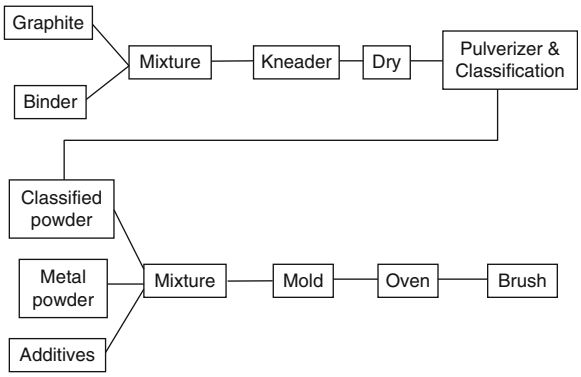
Characteristics of Brush

Physical parameters used to define characteristics of brush are hardness, resistivity, apparent density, true density, bending strength, elasticity modulus, and coefficient of friction (Holm 1967; Shobert 1965).

Hardness: Generally, a brush with high hardness is strong in abrasive ability, but some brushes with low hardness also strong in abrasive ability. Hardness is measured by a Shore Hardness tester.



Brush Materials, Fig. 1 Production process of electrographite brush



Brush Materials, Fig. 2 Production process of metal-graphite brush

Apparent specific gravity: Obtained by dividing mass by the volume calculated from the geometry of a brush.

True specific gravity: Increases with graphitization level and thus used as a measure of graphitization level. The pycnometric technique is used for the measurement.

Table 1 shows physical properties of each brush grade. These values are tentative criteria as they are dependent on raw materials, combination percentage, and production process.

Metal Brushes

Copper system materials were used as metal brushes in the past, but presently precious metals are used for low-voltage and small-current motors, because noise level is low, contact resistance is stable and low, and wear rate is small (Matula 1979; Pentlicki 1971). Materials are precious metals or their alloys.

The materials are divided into ones with spring characteristic (Table 2) and ones with no spring characteristic (Table 3).

Presently, AgPd50%(weight) is often used as a brush materials.

Brush Materials, Table 1 Properties of carbon type brushes

Properties/Grade	Carbon grade	Electrographite (soot type)	Electrographite (pitch coke type)	Graphite grade	Metal-graphite grade
Apprant specific gravity	1.4 ∼ 1.6	1.4 ∼ 1.65	1.6 ∼ 1.75	1.18 ∼ 1.7	2.2 ∼ 6.0
True specific gravity	1.8 ∼ 1.9	1.9 ∼ 2.05	2.05 ∼ 2.15	2.15 ∼ 2.2	–
Resistivity (Ω cm)	0.0025 ∼ 0.008	0.003 ∼ 0.008	0.001 ∼ 0.003	0.001 ∼ 0.0035	0.000009 ∼ 0.001
Bending strength (kg/cm^2)	300 ∼ 400	90 ∼ 300	200 ∼ 500	50 ∼ 250	150 ∼ 500
Elasticity modulus (10^3 kg/cm^2)	–	30 ∼ 90	50 ∼ 120	20 ∼ 150	–
Coefficient of viscosity (CGS)	–	0.7 ∼ 1 × 10^5	0.8 ∼ 12 × 10^5	0.8 ∼ 20 × 10^5	–
Coefficient of friction	0.25 ∼ 0.35	0.20 ∼ 0.30	0.20 ∼ 0.30	0.25 ∼ 0.35	0.15 ∼ 0.27
Hardness (shore)	35 ∼ 50	35 ∼ 45	35 ∼ 55	20 ∼ 30	10 ∼ 18

Brush Materials, Table 2 Precious metals for sliding contacts with spring characteristic

	SP-1[a]	SP-2[a]	SP-3[a]	SP-5[a]	625[b]	625R[b]	Ag-Pd-Cu	Pd-Ru
Compostion[%] Au	10	70	–	20	62.5	62.5	–	–
Pt	10	5	0 ∼ 1	5	–	–	–	–
Pd	35	–	42 ∼ 44	45	–	–	50	90
Ag	30	10	38 ∼ 41	30	7.5	29	20	–
Others	Cu,Zn	Cu,Ni	Cu	–	Cu	Cu	Cu	Ru
Density (g/cm^3)	11.9	15.9	10.3	12.8	13.7	14.4	10.0	12.0
Fusing point (°C)	1,098	955	1,077	1,371	861	1,014	1,115	1,580
Resistivity (μΩ·cm)	31.6	13.3	28.25	39.4	14.0	12.5	27.2	43.0
Young's modulus (kg f/mm^3)	11,900	11,200	11,900	11,200	–	–	–	–
Degree of hardness (HV)	265 ∼ 310	235 ∼ 290	270	95	280 ∼ 300	260 ∼ 290	190 ∼ 320	180 ∼ 280
Tensile strength (kgf/mm^2)	116 ∼ 144	105 ∼ 140	101	37 ∼ 52	–	–	70 ∼ 130	60 ∼ 100
Spring characteristic	Like Cu-Be	Like Cu-Be			Like Cu-Be			
Usage	Brush for micro motor	Slip ring brush Sliding contacts Commutator	Contact for vibrator	Contact for vibrator		Sliding contacts	Brush for micro motor	

[a]Brand name of Tanaka Kikinzoku Kogyo Ltd.
[b]Brand name of Johnson Massey Co.

Key Applications

Materials and applications for carbon brushes are described in the entry ► Electrical Brushes.

Precious metal alloys with spring characteristic and their applications are shown in Tables 2 and 3 (http://www.tanaka.co.jp/english/).

Brush Materials, Table 3 Precious metals for sliding contacts without spring characteristic

	Composition (%)	Melting point (°C)	Hardness (HV)	Electric conductivity I ACS (%)	Density (g/cm^3)	Applications
Au system alloy	Au-Ag8	1,058	30	28.7	18.0	Commutator
	Au-Ag10	1,055	30	25.4	17.9	
	Au-Ag20	1,045	33	18.1	16.6	
	Au-Ag25	1,040	35	16.6	16.0	
	Au-Ag40	1,005	40	15.6	14.5	
	Au-Ag90	970	29	48	11.0	
Pt system alloy	Pt-Pd10	1,550	90	6.2	19.9	Slip ring
	Pt-Pd20	1,560	110	5.7	18.6	
Pd system alloy	Pd-Cu15	1,380	100	4.6	11.2	Potentiometer
	Pd-Cu40	1,223	120	4.9	10.4	Brush for motor
	Pd-Ag20-Cu30	1,150	190	6.3	10.0	Brush for motor
	Pd-Ag30-Cu30	1,066	200	5.0	10.6	
Ag system alloy	Ag-Pd70	1,430	90	4.3	11.5	Brush for micro motor
	Ag-Pd60	1,395	80	4.3	11.4	
	Ag-Pd50	1,350	75	5.7	11.2	
	Ag-Pd40	1,290	65	8.2	11.1	
	Ag-Pd30	1,225	60	11.5	10.9	
	Ag-Cu7.5	799	56	90	10.4	Commutator for micro motor
	Ag-Cu10	778	62	86	10.3	
	Ag-Cu90	778	60	80	9.1	
	Ag-Cu6-Cd2	880	65	43	10.4	Commutator for micro motor
	Ag-Cu24.5-Ni0.5	810	135	68		
	Ag-In18	746	50			Commutator
	Ag-Cd1	959	35	92	10.5	

Various materials are used for brushes and commutators of micro motors, slip rings, and others. Some materials are similar to Cu-Be as a spring material.

Table 3 shows regular contact materials that are used for brush and commutator.

Generally, the shapes are divided into solid types and clad types. In the clad type, a contact material is bonded to a base metal to save a precious metal.

Cross-References

- ► Contact Materials, Silver Graphite, Silver-Refractory-Graphite
- ► Electrical Brushes
- ► Electroplating
- ► Friction Coefficient
- ► Metal Fiber Brushes
- ► Temperature Effect on Fatigue

References

R. Holm, *Electric Contacts* (Springer, New York, 1967)
http://www.aupac.co.jp/e/index.html
http://www.tanaka.co.jp/english/
R.A. Matula, Electrical resistivity of copper, gold, palladium and silver. J. Chem. Ref. Data **8**, 1147–1297 (1979)
C.J. Pentlicki, E.W. Glossbrennerm, The testing of slip rings and brushes for space applications, in *Electric Contacts 1971, Proceedings of the Holm Conference on Electrical Contacts* (Illinois Institute of Technology, Chicago, 1971), pp. 157–172
P. Reichner, Metallic brushes for extremely high current applications, in *Electric Contacts 1979, Proceedings of the Holm Conference on Electrical Contacts* (Illinois Institute of Technology, Chicago, 1979), pp. 191–198

E.I. Shobert II, *Carbon Brushes* (Chemical Publishing, New York, 1965)
P.G. Slade, *Electrical Contacts* (Marcel Dekker, New York, 1999)
D.K. Wilsdorf, D.D. Makel, N.A. Sondargaard, D.W. Maribo, Interfacial temperatures and transition of monolithic silver-graphite brushes on copper, in *Proceedings of the 14th International Conference on Electrical Contracts*, Paris, 1988, pp. 47–53
Y. Wu et al., The research of fiber/graphite composite, in *Electrical Contacts 1995, Proceedings of the IEEE Holm Conference* (Institute of Electrical and Electronics Engineers, Montreal, 1995), p. 315

Brush Seals

Margaret P. Proctor
NASA Glenn Research Center, Cleveland, OH, USA

Synonyms

Bristle seal

Definition

A brush seal is an array of densely packed bristles that has one end fixed and the other end free and that is adjacent to or contacting a surface to be sealed, separating two regions of gas or liquid at different pressures.

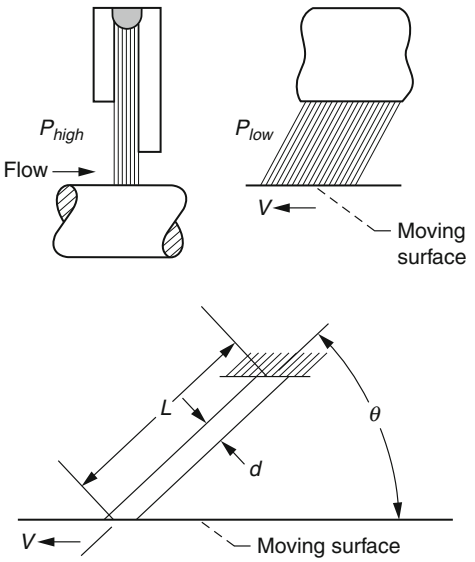
Scientific Fundamentals

Bristles

Brush seals are compliant seals used to seal areas with dynamically changing clearances. Typical brush seals have cylindrical bristles, the axes of which are oriented at an angle to the sealing surface, as shown in Fig. 1. This bristle angle, θ, is typically 35–60°. In this orientation, the bristles act as cantilevered beams, flexing towards the seal surface due to pressure differentials across the bristle pack and flexing away from it when the sealing surface moves towards the bristle pack. The bristle angle determines how much of the bristle load is carried by bending and how much is carried by column compression. Low-angle bristles carry more of the load by bending than column compression. Conversely, high bristle angles result in more column compression than bending of the bristles. Since the bristles usually have a high length to diameter ratio, very high bristle angles tend to result in buckling and breaking of the bristles.

Bristle angle also affects the bristle contact area with the sealing surface. The bristle contact area, $A_{bristle}$, is a function of the bristle diameter, d, and the bristle angle, θ, as shown:

$$A_{bristle} = \frac{\pi d^2}{4 sin\theta} \quad (1)$$



Brush Seals, Fig. 1 Typical brush seal cross-section and bristle angle

Low bristle angles have a higher contact area at the bristle tip, but lower contact pressure per bristle per unit deflection, P/u, as shown in (2), where E is the modulus of elasticity and L is the bristle length.

$$\frac{P}{u} = \frac{E\, d^2 \sin\theta}{5.333\ L^3 \cos^2\theta} \quad (2)$$

Equation (2) is based on simple cantilever beam bending, which is a good starting point, but alone is insufficient to determine brush seal stiffness. The stiffness of a brush seal depends on bristle stiffness, the density of the bristle pack, bristle angle, the pressure differential across the bristle pack, and friction within the seal. The application of fluid pressure to the seal results in a compacting of the bristle array and increased bristle-to-bristle friction. The stiffness of the bristles is a function of their length, diameter, and modulus of elasticity. Larger diameter bristles are stiffer and create larger leakage paths when the bristles move independently of one another and line contact between bristles is lost over the bristles' length. Smaller diameter bristles form a dense barrier against leakage through the bristle pack. Bristle pack density typically ranges from 900 to 1,770 bristles/cm (2,300 to 4,500 bristles/in.).

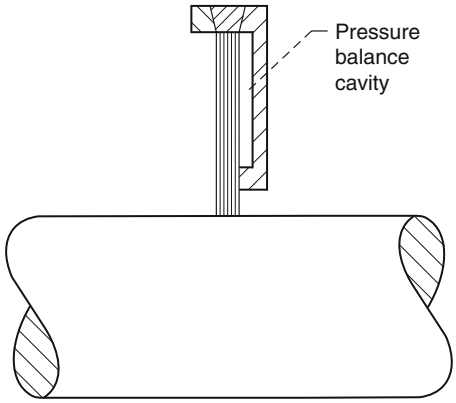
Pressure Loads and Back Plates

When a brush seal is exposed to a pressure differential in a direction perpendicular to the bristle pack, the bristles

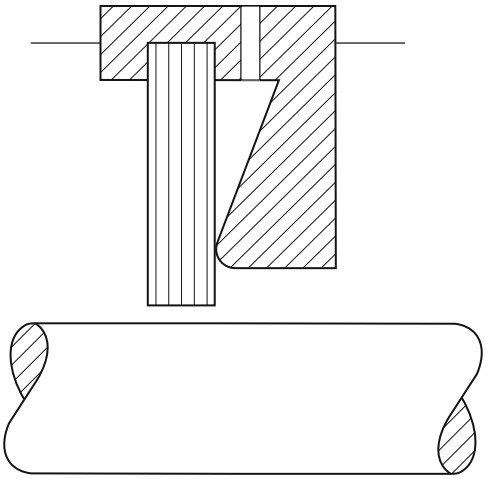
experience loads both parallel and perpendicular to the pressure differential. The loads perpendicular to the pressure differential result in what is often referred to as a "pressure-closing" effect, which draws the bristles into the sealing surface (Fuerguson 1988). This occurs because the pressure drops under the bristle tips while the higher inlet pressure acts above the bristles, resulting in a net closing force on the bristles. When loads parallel to the pressure differential (or perpendicular to the bristle pack) cause sufficient deflection of the bristles in the same direction that a clearance results, then bristle "blow down" has occurred and results in higher leakage rates. There are some occurrences in the brush seal literature in which this phenomena is also referred to as pressure closing.

Generally, brush seals incorporate a back plate on the low-pressure side of the bristle pack to support the bristles and prevent blow down. The distance between the back plate and the sealing surface, sometimes referred to as the fence height, must be large enough to accommodate the expected changes in the sealing surface location. The surface of the back plate that is adjacent to the bristle pack may be, flat, tapered, or have a relief in it to minimize frictional hysteresis between the bristles and the back plate. Designs with a relief in the back plate form a pressure balance cavity (Fig. 2). Some designs provide holes to supply this cavity with upstream pressure (Fig. 3) for pressure balancing.

For brush seals with back plates immediately adjacent to the bristle pack (Fig. 1), bristle deflection is nearly entirely in the plane normal to the sealing surface. Also, as pressure increases the brush will become stiffer in this normal plane due to the frictional forces between the bristle pack and the back plate. This frictional force can also cause hysteresis. Hysteresis occurs when bristles pushed back due to a movement of the sealing surface don't return to their original position once the sealing surface has retreated, resulting in higher leakage rates. For applications with low pressure differentials, hysteresis is a smaller concern. For high-pressure applications it is more important to make use of features in the back plate design to mitigate hysteresis.



Brush Seals, Fig. 2 Brush seal design with recessed back plate forms a pressure balance cavity, which reduces hysteresis



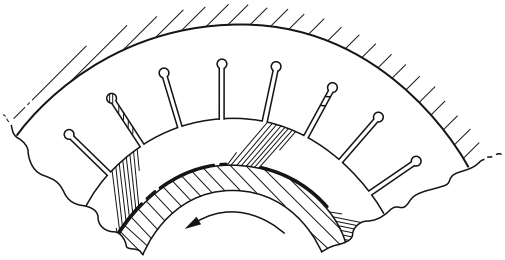
Brush Seals, Fig. 3 Brush seal design with separate supply hole to pressure balance cavity formed by the tapered back plate and bristle pack

Front Plates and Baffles

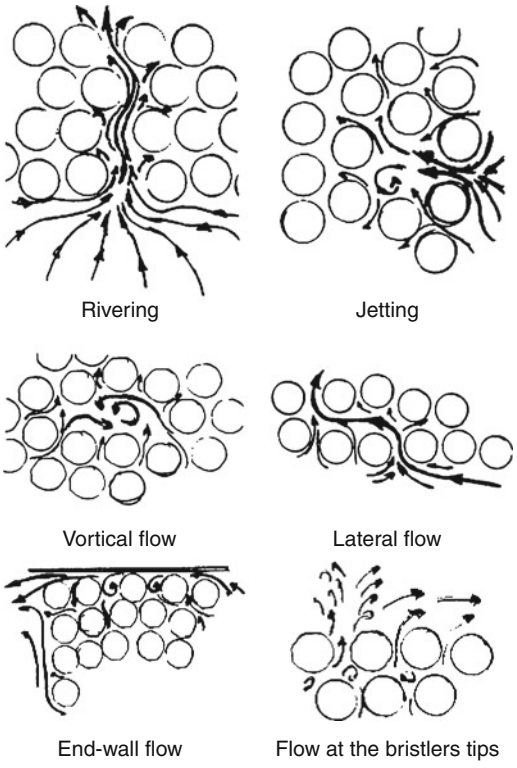
Front plates, similar to the back plate, or baffles (Fig. 4) can be used to protect the bristles from non-uniform flows that jet or tunnel through the bristle pack. (Braun et al. 1990a, b) conducted flow visualization experiments demonstrating these phenomena, which are illustrated in Fig. 5. When used, a baffle or flow deflector should have a lower stiffness in the direction of flow than the bristle pack so that the baffle will follow the bristle pack deflection caused by the pressure differential (Short et al. 1996). Front plates are sturdier than baffles and are used to protect the bristles from the incoming flow and from handling damage during installation. It is advisable to have a slight gap between the front plate and the bristle pack to prevent pinching of the bristles and to allow flow into any pressure balance cavity formed by a relief in the back plate.

Leakage

Models to predict the leakage rate of brush seals have been developed and are based on empirical data. Early bulk flow models (Hendricks et al. 1991) treated the brush seal as



Brush Seals, Fig. 4 Baffle design for use upstream of bristle pack to protect the bristles from non-uniform flows



Brush Seals, Fig. 5 Flows observed in brush seals (Braun et al. 1990b)

a porous media. Brush seal leakage is comprised of three flows: flow that goes directly through the bristles for the area between the free end of the bristles to the fence edge, flow along the bristles from their fixed ends to the fence edge, and flow between the bristle tips and the sealing surface if there is a clearance between them.

More recent models (Chupp and Holle 1996) treat the brush seal as an array of rods perpendicular to a flow and model leakage rate as an empirical function of pressure differential, effective brush thickness, other brush geometry, fluid density, and Reynolds number. The effective brush thickness model can be used to predict brush seal leakage performance at engine conditions based on ambient temperature dynamic-performance leakage tests or speed-corrected static tests. A primary assumption in these models is that all the flow is perpendicular to the bristles across the entire bristle pack. In reality, some flow must be in the direction of the bristles for the flow to get past the back plate. The effective brush thickness accounts for the extra resistance of the back plate.

Leakage comparisons of seal performance are made in terms of flow factor or effective clearance. Flow factor, ϕ, is a function of mass leakage rate, $\dot{m}$, inlet air temperature, T, rotor diameter, D, and the seal upstream pressure, P_u, as shown in (3):

$$\phi = \frac{\dot{m}\sqrt{T}}{DP_u} \tag{3}$$

Effective clearance, C_{eff}, is computed, assuming isentropic, compressible, choked flow through an annular clearance, as:

$$C_{eff} = R\frac{(0.8333)^{0.5}\phi}{\pi(0.5283)\sqrt{\gamma R}} \tag{4}$$

where R is the gas constant and γ is the ratio of specific heats.

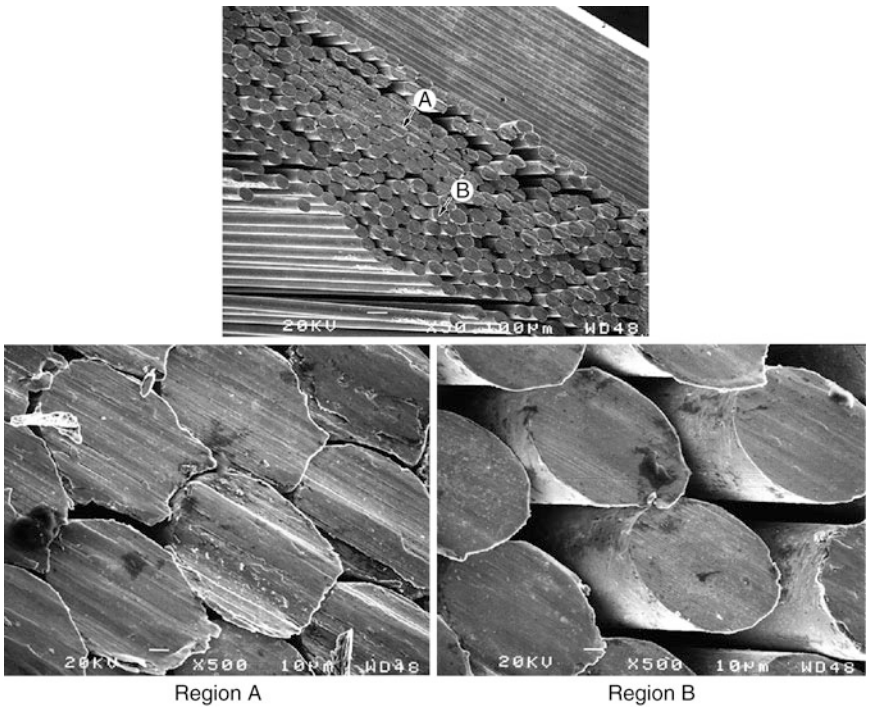
Brush seals have leakage rates that are approximately 50% (or less) of labyrinth seal leakage rates and are similar to those of finger seals. Leakage rate increases as the bristle tips wear away and the clearance increases. This degradation in leakage performance is not seen immediately since the pressure closing force will continue to move the bristles into the sealing surface as the seal wears away.

Power Loss

If valid friction coefficient, μ, normal contact force, n, and surface velocity, V, are available, then frictional power loss, P_{power}, of the seal can be calculated as:

$$P_{power} = \mu \cdot n \cdot V \tag{5}$$

Power loss increases with increased surface speed and pressure differential, which increases the contact force at the bristle tips due to pressure closing. In circular brush seals sealing around a rotating shaft, centrifugal growth of the rotor can further increase the contact pressure at the bristle tip. Proctor and Delgado (2004) measured brush seal power loss of approximately 2 kW at inlet air temperatures of 922 K, surface speeds of 366 m/s, and pressure differential of 517 kPa and found it comparable to the measured power loss of finger seals at the same conditions.



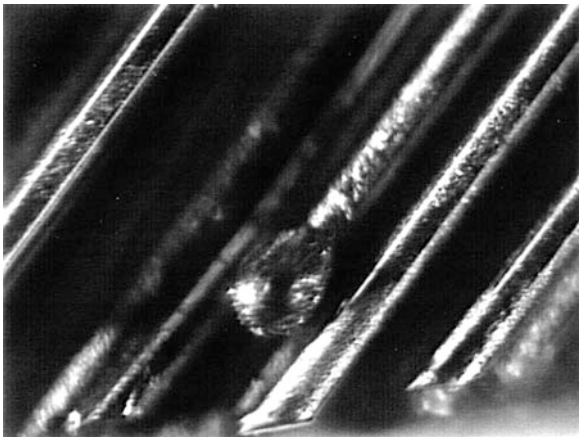
Brush Seals, Fig. 6 Bristle tips after high speed tests in LN_2; substantial smearing wear occurred on downstream bristles (region A), which were closer to back plate (Proctor et al. 1996)

Coating and bristle material combinations with low friction coefficients are desired.

Wear and Surface Tribology

Wear of the bristle tips and sealing surface will occur whenever there is contact while the sealing surface is moving. Typically, it is desirable to use a harder material on the moving surface to protect it and to allow most of the wear to occur on the bristles to avoid a major teardown of a machine. Usually debris from the bristles is small, like a talcum powder, and carried downstream by the flow. However, larger debris can be generated if the brush seal is not designed properly. At very high temperatures and speeds smearing of the bristle tips has been observed (Fig. 6) and even bristle tip melting (Fig. 7) due to localized frictional heating has been seen. In some cases the bristle tips get bent, as shown in Fig. 8.

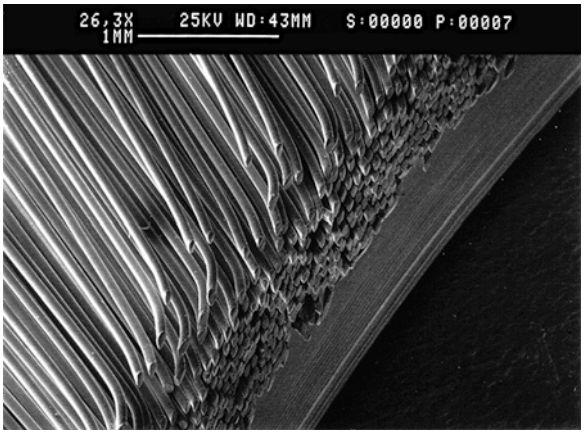
Prior et al. (1998) have empirically modeled brush seal wear rates based on the common understanding that adhesive wear rate is proportional to the normal contact load times sliding velocity divided by hardness. While this is a seemingly simple approach, the modeling is complicated by the fact that hardness changes with temperature and the temperature at the bristle tip depends on the coefficient of friction, contact load, sliding velocity, flow rate and temperature, thermal conductivity, and bristle diameter.



Brush Seals, Fig. 7 Melted bristle tip (Proctor et al. 1996)

Materials

Nylon or polypropylene bristles fixed in a metallic holder are commonly used for door seals in many industrial



Brush Seals, Fig. 8 Outer bristles of brush seal exhibit tip bending in the circumferential direction after testing in LH2 (Proctor et al. 1996)

applications. For gas turbine engine applications, brush seals are made of stainless steel or high-temperature metal alloys. Bristle material is selected for its temperature capability, oxidation resistance, strength, ductility, wear resistance, and ability to be drawn to wire diameters of 51–152 μm (0.002–0.006 in.).

Haynes 25 is often used for bristle material for its good strength properties at high temperature and its good oxidation resistance, which results from the formation of a Cr_2O_3 film. A cobalt, chromium, nickel, tungsten-based superalloy, which derives its strength from solid-solution hardening, Haynes 25 has reasonably low friction at lower temperatures (<400°C), where the hexagonal close-packed structure is stable (Derby and England 1992). Nickel-based alloys, such as Inconel X-750 and Haynes 214, and ceramics may also be considered for bristle material, particularly for higher temperature applications. A major concern with using ceramics for bristles is their fabrication. Hendricks et al. (1993) reported that three ceramics considered feasible for bristle fabrication are silicon carbide, aluminum oxide, and quartz. Active metal hydrides were effective braze materials to withstand 815°C (1,500°F), while not completely wetting the bristles. A silicon carbide bristle brush seal with metal side plates was successfully fabricated. Addy et al. (1995) reported preliminary test results of a silicon carbide brush seal on a zirconia-coated rotor and observed that the bristles are very stable and exhibit significantly less vibration than conventional metallic bristle brush seals. The zirconia coating had significant wear after 9 h of testing.

The sealing surface material is usually metal with a hard coating, typically chrome carbide, for wear resistance. Experimental data show that smoother surface finishes have less bristle wear (Fuerguson 1988). A surface finish of 0.2–0.3 μm (8–12 μ-inch) is recommended. Consideration should be given to matching the coefficients of thermal expansion of the base material and the coating to prevent heat checking in the wear track where frictional heating will occur. Of course, a low coefficient of friction between the bristle and coating material is desirable to minimize heat generation and power loss.

Derby and England (1992) evaluated seven bristle materials and seven coating materials for friction and wear behavior at high temperature and high sliding speed to extend brush seal performance. The aluminum oxide and chromium oxide coatings that they tested exhibited microfracturing and grain pullout which leads to high wear and increased friction. They also found inferior performance in tribopairs that transfer excessive amounts of metal, especially nickel, iron, and tungsten, from the bristle to the coating due to galling from like-on-like metal contact. Bristles made of Alloy A, a nickel, chromium, aluminum-based super-alloy that is solid-solution strengthened (probably Haynes 214), exhibited superior oxidation resistance and excellent low friction against Triboglide® coating at high temperature and high surface speed. Triboglide® is chromium carbide with barium and calcium fluoride solid lubricants. The low friction is attributed to these solid lubricants and the tenacious oxide film on the Alloy A surface.

Kevlar® and carbon fiber are candidate materials for bristles in brush seals used to separate air and oil in bearing sumps. Kevlar is an aramid and is a lightweight, very strong, heat-resistant, synthetic, aromatic, polyamide material. Wear debris produced from these two materials will pass through the bearings without causing damage. The very small diameters of these fibers result in very low leakage rate seals. Ruggiero et al. (2008a) found that heat generation rates of a brush seal with carbon fibers were 66% lower than with Kevlar® due to the low coefficient of friction of the carbon fiber and due to the low bending resistance of the carbon fiber seal design, which produced a low contact force on the rotor. However, the lubricity properties of carbon fiber are highly dependent on the level of humidity in the operating environment.

Key Applications

Door sweeps are a commonplace application of linear brush seals with fiber or polymeric bristles. Wire metal bristle brush seals are more challenging tribologically, yet the benefit of their lower leakage rates have been successfully pursued in gas turbine engines for both power generation and aero propulsion. System studies estimate that

using brush seals in aero propulsion gas turbine engines has potential to reduce specific fuel consumption by 2–3% and to reduce direct operating costs by increasing time between engine overhauls (Steinetz et al. 1998). Chupp et al. (2006) reported typical operating limits for state-of-the-art brush seals of differential pressures up to 2.1 MPa, surface speeds up to 400 m/s, operating temperatures up to 600°C, and diameters up to 3.1 m. Circular brush seals are used in gas turbine engines, replacing labyrinth seals, to seal hot gas leakage along shafts or other rotating components. Segmented circular brush seals have been used in large power generation gas turbine engines. Steinetz and Hendricks (1997) provide a good summary of specific engines in which brush seals have been used or tested.

Dinc et al. (2002) reported that more than 200 brush seals are installed in gas turbine engines and the fleet leader has accumulated about 40,000 h of service. They also reported experimental measurements of brush seal effective leakage less than 20% of a typical labyrinth seal and delineated fundamental design issues and the design process to incorporate brush seals into industrial gas turbine engines.

Although the vast majority of brush seal development has focused on gas turbine engines, brush seals are candidates for other rotating machinery as well. Brush seals were first invented as seals specially suitable for elastic fluid turbines (DeFerranti 1908). Pioneering work testing brush seals in liquid nitrogen and liquid hydrogen by Proctor et al. (1996) and by Walker and Proctor (1995) has lead to brush seal designs being considered for the RS-68 rocket engine liquid hydrogen turbopump. Ruggiero et al. (2008b) reported successful demonstration of a Kevlar® fiber brush seal in a liquid natural gas compressor achieving both oil preclusion and ultra-low leakage performance. Although brush seals are typically used to seal against a sliding surface, they can be used in non-sliding applications as well to accommodate changes in clearances.

Cross-References

- ► Brush Materials
- ► Conformal-Contact Elements and Systems
- ► Finger Seal: A Compliant Seal
- ► Friction in Seals
- ► Labyrinth Seal

References

H.E. Addy, Jr., B.M. Steinetz, R.C. Hendricks, H.W. Howe, J. Flowers, Preliminary results of silicon carbide brush seal testing at NASA Lewis Research Center, in *AIAA 95–2763, 31st AIAA/ASME/SAE/ASEE joint propulsion conference and exhibit*, San Diego, (1995)

M.J. Braun, R.C. Hendricks, V.A. Canacci, Noninstrusive qualitative and quantitative flow characterization and bulk flow model for brush seals. Presented at international tribology conference, Nagoya, (1990a)

M.J. Braun, R.C. Hendricks, V. Canacci, Flow visualization in a simulated brush seal, ASME 90-GT-217. Presented at gas turbine and aeroengine congress and exposition, Brussels, (1990b)

R.E. Chupp, G.F. Holle, Generalizing circular brush seal leakage through a randomly distributed bristle bed. J Turbomach, (1996), 118:153–161

R.E. Chupp, R.C. Hendricks, S.B. Lattime, B.M. Steinetz, Sealing in turbomachinery. NASA TM-2006-214341, (2006)

S.Z. De Ferranti, Fluid packing, U.S. Patent No. 885,032, Issued 21 Apr 1908

J. Derby, R. England, Tribopair evaluation of brush seal applications. AIAA 92–3715, in *AIAA/SAE/ASME/ASEE 28th joint propulsion conference and exhibit*, Nashville, (1992)

S. Dinc, M. Demiroglu, N. Turnquist, J. Mortzheim, G. Goetze, J. Maupin, J. Hopkins, C. Wolfe, M. Florin, Fundamental design issues of brush seals for industrial applications. J Turbomach, (2002) 124:293–300

J.G. Fuerguson, Brushes as high performance gas turbine seals. ASME paper 88-GT-182, American Society of Mechanical Engineers, New York, (1988)

R.C. Hendricks, S. Schlumberger, M.J. Braun, F. Choy, R.L. Mullen, A bulk flow model of a brush seal system, ASME Paper No. 91-GT-325. Presented at the 36th International Gas Turbine and Aeroengine Congress and Exposition, Orlando, (1991)

R.C. Hendricks, R. Flower, H. Howe, Development of a brush seals program leading to ceramic brush seals, Preprint from seals flow code development-93, in *Proceedings of a conference held at NASA Lewis Research Center*, Cleveland, 3–4 Nov (1993)

R. Prior, J. Short, P. Basu, Brush seal wear model, in *AIAA 98–3170, 34th AIAA/ASME/SAE/ASEE joint propulsion conference & exhibit*, Cleveland (1998)

M.P. Proctor, J.F. Walker, H.D. Perkins, J.F. Hoopes, G.S. Williamson, Brush seals for cryogenic applications. Performance, stage effects, and preliminary wear results in LN_2 and LH_2, NASA Technical paper 3536 (1996)

M.P. Proctor, I.R. Delgado, Leakage and power loss test results for competing turbine engine seals. NASA TM-2004-213049, ARL-TR-3157, GT2004-53935, 2004 Turbo Expo, Vienna (2004)

E.J. Ruggiero, J. Allen, R.M. Lusted, Heat generation characteristics of a carbon fiber brush seal, in *AIAA 2008–4508, 44th AIAA/ASME/SAE/ASEE joint propulsion conference & exhibit*, Hartford (2008a)

E.J. Ruggiero, P. Susini, R.M. Lusted, Kevlar® fiber brush seals for LNG compressors. in *AIAA 2008–4623, 44th AIAA/ASME/SAE/ASEE joint propulsion conference & exhibit*, Hartford, (2008b)

J.F. Short, P. Basu, A. Datta, R.G. Loewenthal, R.J. Prior, Advanced brush seal development, in *AIAA 96–2907, 32nd AIAA/ASME/SAE/ASEE Joint Propulsion Conference*, Lake Buena Vista, (1996)

B.M. Steinetz, R.C. Hendricks, in 9.6 Brush seals. ed. by E.V. Zaretsky. Tribology for Aerospace Applications, Society of Tribologists and Lubrication Engineers, Illinois, (1997), pp. 639–658

B.M. Steinetz, R.C. Hendricks, J. Munson, Advanced seal technology role in meeting next generation turbine engine goals. NASA/TM-1998-206961, (1998)

J.F. Walker, M.P. Proctor, Wear characteristics of three rotor coatings for application to brush seals operating in liquid hydrogen. NASA TM 107203 (1995)

Buildup

► Weld Overlay

Bulk Metallic Glass (BMG)

► Crack Growth in Noncrystalline Solids

Bulked Continuous Fiber (BCF)

► Fiber Tribology and the First Man-Made Fiber Production

Buttering

► Weld Overlay