W

Water Jet Cutting

► Gear Manufacturing Machines

Water Lubrication

► Water-Lubricated Rubber Alloy Bearings and Transmission Systems

Water Lubrication-Lubrication of Hard Solid Coatings Sliding Against Mating Materials in Water

► Low-Friction Behavior of Hard Solid Coatings in Water Environment

Water-Lubricated Bearings

► Surface Texture for Water Lubrication

Water-Lubricated Rubber Alloy Bearings and Transmission Systems

Jiaxu Wang
Chongqing University, Chongqing,
People's Republic of China

Synonyms

Bearings using water as lubricant; Water lubrication

Definition

Water-lubricated rubber-alloy bearings and transmission systems utilize natural water as the lubricant and composite materials for the friction pair, replacing mineral oil and metal friction pairs. These systems can be found in marine propulsion, hydraulic and nuclear power generators, mining and construction equipment, ocean exploration, and various watercraft applications. When using water as the lubricant, specifically designed journal bearings made of rubber-alloy composites can help to improve lubrication performance, reliability, efficiency, and environmental compatibility with reduced vibration and noise.

Scientific Fundamentals

Water and Water Lubrication

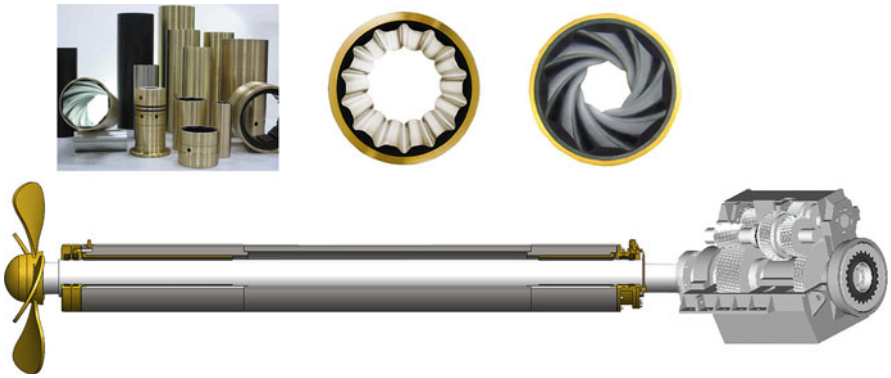
Bearings, sealing devices, and couplings are traditional key metal parts in mechanical transmission systems, especially in marine propulsion systems, and they are usually lubricated with a mineral oil. The drawbacks of the traditional lubrication of these mechanical transmission systems include:

- Complicated structure of sealing device required for preventing lubricating oil leakage;
- Inevitable friction, wearing, vibration, and noise;
- Low reliability and durability;
- Contamination due to mineral oil leakage to water caused by seal failure.

Therefore, the research on innovative design theories and methods for high-performance transmission systems based on resource-saving and eco-friendly technologies has become an urgent and common area for the required sustainable future development.

Water is a liquid most commonly found in nature. Many vehicles, vessels, and machineries operate under water-rich environmental conditions, such as transmission systems for marine propulsion, hydraulic and nuclear power generation, mining and construction, ocean exploration, and various military applications (Orndoff 1985). Bearings in these systems often prefer to use water as lubricant if possible in order to avoid complex sealing systems. However, the absolute viscosity of pure water is

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



Water-Lubricated Rubber Alloy Bearings and Transmission Systems, Fig. 1 A typical propulsion system with water-lubricated bearings

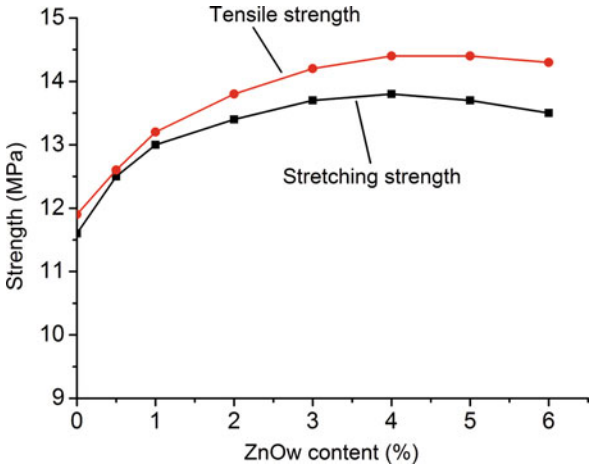
low (e.g., only about 1/65 of a typical engine oil at 50°C). The theory of hydrodynamic lubrication suggests that the load capacity of a fluid-film journal bearing is proportional to the viscosity and inversely proportional to the square of lubricant film thickness. Therefore, in order for a water-lubricated bearing to have the same load capacity, the film thickness is required to be only one eighth of that of the engine oil lubricated bearing under the same conditions. This means that, usually, water-lubricated rubber alloy bearings either have lower load capacity at the same film thickness, or operate at much higher risk of surface contact if the applied loads are the same (Cabrera et al. 2005). Bearing material properties, particularly self-lubricating and anti-corrosion anti-wear tribological properties, are vital to water-lubricated bearing rubber alloy design, performance and efficiency improvement, and failure prevention (Majumdar et al. 2004).

Figure 1 shows a typical propulsion system equipped with water-lubricated rubber alloy bearings. Each bearing consists of a cylindrical shell and a resilient liner. The shell is formed with a metallic material or a hard plastic material, and the liner is traditionally made of a synthetic rubber material. The electromagnetic induction heating process is employed for the liner manufacturing to produce the high bond strength between the liner and the shell. The liner has several internal grooves, often called flutes, on its inner wall surface. The flutes, designed to retain water between liner and shaft surface, can often reduce the coefficient of friction and improve lubrication condition. When the water is not clean, the contaminants such as sand slurry can be carried away by water flows through the flutes. This ensures bearing performance and increases system reliability, especially in a dirty water environment. A thin layer of adhesive is often used between the shell and the liner to prohibit possible relative rotation of the liner in the shell. Meanwhile, the water-lubricated rubber alloy bearings integrated with a seal device, high-elastic coupler, and the gearbox can constitute a water-lubricated mechanical transmission system, which is shown in Fig. 1 (Wang and Wang 2008a–b; Wang et al. 2008a–c).

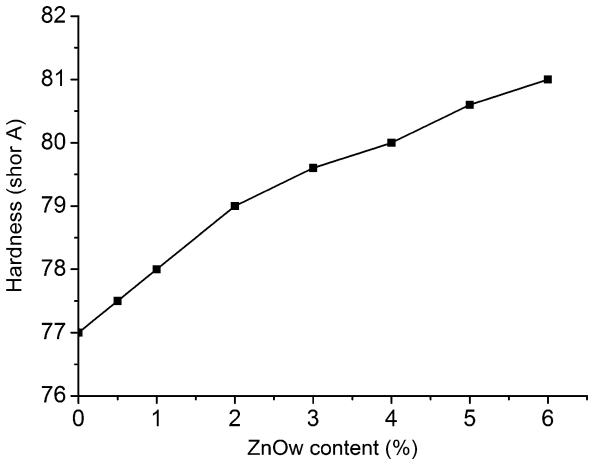
Water-Lubricated Bearing Materials

As mentioned above, water-lubricated bearings usually operate under severe environmental conditions, most likely in the mixed or boundary lubrication regime, in which surface contact between the drive shaft and the bearing sleeve is often significant. This imposes great challenges to bearing design, especially material selection. Conventionally, commonly used materials for journal bearings include white metals (a.k.a. Babbit metals), copper-based alloys, aluminum-based alloys, zinc-based metals, and some nonmetallic materials, such as rubber, graphite, plastics, timber, ceramics, and others. These materials have demonstrated satisfactory performance under normal operating conditions. In the 1960s, timber bearing materials, such as lignum vitae, were widely used. As the awareness of environmental protection and natural resource conservation is enhanced, new materials have increasingly been employed in water-lubricated bearings. These include ceramics, plastics, rubber, nylon, and various types of composites.

Nitrile butadiene rubber (NBR) is a widely used rubber material in bearings, due to its desirable chemical and physical properties. Those properties include excellent airtightness (which plays a directive role in vulcanization) and vibration-absorbing ability. Its glass transition temperature is −22°C, and the stable operating temperature is below 100°C. The basic components of water-lubricated rubber alloy bearings include NBR,



Water-Lubricated Rubber Alloy Bearings and Transmission Systems, Fig. 2 Relationship between tensile strength or stretching strength and ZnOw content



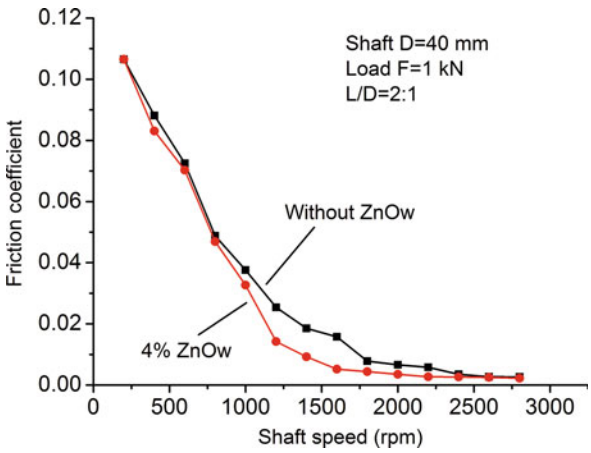
Water-Lubricated Rubber Alloy Bearings and Transmission Systems, Fig. 3 Relationship between hardness and ZnOw content

zinc oxide whisker (ZnOw), sulfur, stearic acid, accelerator, gas black, and others.

ZnOw has outstanding multifunctional properties, including wear-resistance, vibration-absorbing, antiskid, noise reduction, wave absorbing, anti-aging, shock resistance, antistatic property, antisepsis, and others. Proper usage of ZnOw, therefore, can improve the bearing performance and reliability (Xiao et al. 2004). NBR tensile strength and hardness as functions of ZnOw content have been plotted in Figs. 2 and 3. It can be seen that the stretching strength and the tensile strength of the material considerably increases with increasing content of ZnOw in a certain range, and then decreases as the content of ZnOw further increases after reaching a critical point. This means that there is an optimal content of ZnOw, beyond which the material may break and the bearing performance may be worsened. Figure 3 shows that the hardness of rubber increases continuously with the increasing content of ZnOw. Figure 4 demonstrates that the friction decreases with the increasing shaft speed (as predicted by the bearing lubrication theory), and adding ZnOw is beneficial to friction reduction. It is found from Figs. 2 to 4 that 4% of ZnOw appears to be optimal for the NBR bearing material.

The following table shows a performance comparison among several commonly used bearing materials. Note in particular that BTG is a newly invented rubber alloy (Wang et al. 2008d) (Table 1).

Materials composition design is the selection of raw materials and the determination of the relative quantity for each material according to the performance and



Water-Lubricated Rubber Alloy Bearings and Transmission Systems, Fig. 4 Friction coefficient versus shaft speed

process requirements. Pure natural rubbers or synthetic rubbers, whether unvulcanized or vulcanized, often fail to perform satisfactorily. The composition has to be properly modified by adding various additives into the rubber. The quantity of each additive needs to be optimized in order to achieve performance goals and minimize fabrication cost.

As mentioned above, vulcanization is essential and also is the final process in rubber material production (Wang et al. 2008d). It is the key for achieving a desirable rubber performance. Vulcanization is a chemical process

Water-Lubricated Rubber Alloy Bearings and Transmission Systems, Table 1 Performance comparison of materials commonly used in water-lubricated bearings

Materials	Price	Machining	Sensitivity to sand slurry	Required installation accuracy	Wear resistance	Load capacity
Timber	High	Difficult	Sensitive	High	High	Low
Ceramic	High	Difficult	Sensitive	High	High	Depends on toughness
Common compound rubber	Low	Easy	Insensitive	Low	Higher	Lower
UHMWPE	Moderate	More difficult	More sensitive	Lower	High	Higher
BTG rubber alloy	Low	Easy	Insensitive	Low	High	High

converting rubber or related polymers into more durable materials via the addition of sulfur, or other equivalent curatives. These additives modify the polymer by forming cross-links (bridges) between individual polymer chains. The vulcanized material is less sticky and has much improved mechanical properties for engineering applications. Moreover, through the vulcanization process, excellent adhesion between the shell and the rubber alloy liner can be obtained, ensuring satisfactory performance and reliability of the bearings. More importantly, the dimensional precision can be achieved by the vulcanization technology without further machining (such as grinding), thus preventing the thin rubber liner from possible damage or destruction.

Key Applications

Marine Propulsion Systems

Water-lubricated rubber alloy bearings are commonly used in the support propeller shafts and the stern rudders in marine propulsion systems. Adequate design of these bearings can help solve problems associated with bearing friction, wear, corrosion, impact, vibration, and noise, and can improve efficiency and reliability. Recently, through modifying material composition and binders and improving the precision forming process via electromagnetic induction heating, adhesion has been greatly enhanced and tearing resistance between rubber alloy liner and metal substrate significantly increased (Wang et al. 2008d).

With this improved process, the adhesive strength has already exceeded the strength of bearing material itself. In addition, the geometric design has been optimized based on a plasto-elastohydrodynamic lubrication theory (Ren et al. 2010), having an integrated structure with multiply curved surfaces and axial corrugation grooves, so that a more effective water wedge can be formed right at the possible contact location between the bearing and the shaft. Furthermore, the geometric wedge allows the lubricant to be "filtered" so that sand slurry and wear debris do not reach the transmission shaft, preventing the shaft from abrasion. In addition, high reliability of moving parts in water lubrication can be acquired with the deformation-coordination design method (Wang and Wang 2008a–b). For this reason, these bearings are traditionally called non-wear and non-contamination bearings.

Water Pumps

Water pumps are devices used to move liquid water and/or increase hydraulic pressure. A pump displaces a volume by a physical or mechanical action. In industries, many water-based liquids need to be transported through pumping. These include fresh and salt water, water solutions, acids, liquors, lye, emulsions, dispersions, and suspensions, as well as azeotropic mixtures, solids-laden fluids, and others. Guide bearings in pumps play a key role in holding and stabilizing the pump shaft and impeller, most of which are best when water-lubricated while handling water-based liquids. Typically the bearings adopt rubber seals and are lubricated by fresh clean water instead of that being pumped (e.g., river water with sand slurry) in order to avoid severe wear and/or corrosion. For this purpose a water filtration may be needed.

Hydraulic Turbines

Hydraulic turbines are machines that convert potential energy of elevated water into mechanical energy of a rotating shaft. Hydraulic turbines have a row of blades fixed to the rotating shaft or a rotating plate. Flowing water, when passing through the hydraulic turbine, pushes

the blades of the turbine and makes the shaft rotate to generate electricity. Fuide bearings and thrust bearings are present in hydraulic turbines. The former hold and stabilize the shaft and prevent oscillation and vibration in the turbine-generator system, while the latter sustain the axial thrust from the rotating wheel. The guide bearings can be either oil lubricated or water lubricated, depending on the lubricant used. Water-lubricated bearings, usually with rubber bushings, are lubricated by clean water. The filtered lubricating water is imported into the radiator grill on top of the bearings via pipes. There is sealing equipment where lubricating water flows from flutes on bearing bush, and the water is distributed to each part of the bearings to form water films so that the bearings can be lubricated with the rotating spindle. At the same time, the heat generated from the frictional work can be carried away by the water. The rubber bushing, fixed in the bearing housing, has separate pads that can be adjusted from the backside or be replaced if worn out. Usually the bottom part of the rubber bearings do not need to be sealed, so they may be arranged close to the turbine running wheel and may also serve as a vibration absorber, ensuring stable operation. In addition, with the rubber bushing a small amount of foreign particles such as those from sand slurry may be pressed into the rubber and absorbed by the bushing. This may prevent the bearings from quick damage. Rubber bearings have the advantages of simple structure, convenient installation, and easy maintenance. However, they require high-quality clean water, so separate water filtration and supply systems are usually needed.

References

D.L. Cabrera, N.H. Woollet, D.R. Allanson, Y.D. Tridimas, Film pressure distribution in water-lubricated rubber journal bearings. Proc. Inst. Mech. Eng. Part J. J. Eng. Tribol. **219**, 125–132 (2005)

B.C. Majumdar, R. Pai, D.J. Hargreaves, Analysis of water-lubricated journal bearings with multiple axial grooves. Proc. Inst. Mech. Eng. Part J. J. Eng. Tribol. **218**, 135–146 (2004)

R.L.J. Orndoff, Water lubricated rubber bearings, history and new developments. Naval Eng. J. **97**, 39–52 (1985)

N. Ren, D. Zhu, W.W. Chen, Q.J. Wang, Plasto-elastohydrodynamic lubrication (PEHL) in point contacts. ASME J. Tribol. **132**, 031501 (2010)

J.X. Wang, B.C. Wang, Water-lubricated bearing with inner arced grooves, China Patent ZL200510057181.2, 2 July 2008a

J.X. Wang, B.C. Wang, Water-lubricated bearing with inner spiral grooves. China Patent ZL200510057179.5, 2 July 2008b

J.X. Wang, F. Tian, B.C. Wang, Water-lubricated rubber alloys bearing. China Patent ZL200810070089.3, 6 Aug 2008a

J.X. Wang, F. Tian, B.C. Wang, Water-lubricated bearing in mechanical transmission system. China Patent ZL200610054131.3, 21 May 2008b.

J.X. Wang, B.C. Wang, K. Xiao, Water-lubricated dynamic seals bearing. China Patent ZL200510057204.X, 23 Apr 2008c

J.X. Wang, F. Tian, B.C. Wang, K. Xiao, Rubber alloy materials and the methods to manufacture transmission parts using these materials. China Patent ZL200810070080.2, 4 Aug 2008d

K. Xiao, J.X. Wang, Y. Zhang, H.M. Yu, Modification on water-lubricated bearings material with nanometer zinc oxide whisker. Lubr. Eng. **162**, 38–39 (2004)

Water-Soluble Oils

► Lubrication with Emulsions

Waviness

► Filtration of Surface Measurement Data

Way Lube

► Slideway Lubricants

Way Lubricant

► Slideway Lubricants

Way Oil

► Slideway Lubricants

Wear Contact of Homogeneous Materials

Irina Goryacheva
Ishlinsky Institute for Problems in Mechanics of Russian Academy of Sciences, Moscow, Russia

Synonyms

Wear contact problems; Wear modeling for homogeneous materials

Definition

Wear contact problems study the evolution of all the contact characteristics (pressure distribution, position and size of contact region, shape of contacting bodies, approach of the bodies, etc.) in wear processes. The mathematical system of equations describing the wear process includes the wear equation obtained experimentally.

Scientific Fundamentals

One of the principle results of wear is that there are irreversible changes in the shape of the surfaces. These changes are comparable to elastic deformations and thus should be taken into account in the estimation of the contact characteristics of the bodies in sliding contact (distribution of stresses, dimensions of contact areas, etc.).

Problem Formulation

The value of the linear wear w_* (change of the linear dimension of the body in the direction perpendicular to the rubbing surface) is often used to describe the wear quantitatively. Generally, the surfaces are worn nonuniformly, hence the linear wear $w_*(x, y, t)$ should be considered at each point (x, y) of the rubbing surface.

In order to solve the wear contact problems it is necessary to have information about the wear equation that establishes a relation between some characteristics of wear and a set of parameters characterizing the properties of friction surfaces and operating conditions. In general form the wear rate $\partial w_*/\partial t$ can be presented as the function of the contact pressure p and sliding velocity V.

$$\frac{\partial w_*}{\partial t} = F(p,\ V). \tag{1}$$

The form of the function F depends of the type of wear (abrasive, fatigue, etc.), temperature, and so on.

Analysis of a number of wear equations obtained theoretically and experimentally shows that in many cases this dependence can be presented in the form (Kragelsky et al. 1982):

$$\partial w_* \partial t = K_w p^{\beta_1} V^{\beta_2}, \tag{2}$$

where K_w is the wear coefficient, and β_1 and β_2 are parameters that depend on material properties, friction conditions, and temperature.

For many contact problems the assumption that the irreversible surface displacement $w_*(x, y)$ is small and comparable to the surface normal elastic displacement $u(x, y)$ is valid. In that case the boundary conditions are posed at undeformed surfaces, neglecting both the elastic displacement $u(x, y)$ and the surface wear $w_*(x, y)$.

Under this assumption the pressure $p(x, y, t)$ within the contact region and the elastic displacement $u(x, y, t)$ for an arbitrary instant of time t are related by operator A that is analogous to the operator relating the pressure and elastic displacement in the corresponding contact problem when the wear does not occur, that is,

$$u(x, y, t) = A[p(x, y, t)] \tag{3}$$

For example, operator A has the following form for frictionless contact of an axisymmetrical punch an elastic half-space ($0 \le r \le a(t)$):

$$A[p(r, t)] = \frac{4(1 - \nu^2)}{\pi E} \int_0^{a(t)} K\left(\frac{2\sqrt{rs}}{r + s}\right) \frac{sp(s, t)ds}{r + s}, \tag{4}$$

where $K(\rho)$ is the complete elliptical integral of the first kind, $r = (x^2 + y^2)^{1/2}$.

As follows from (4), operator A depends on the size and shape of the contact region Ω. If the size of the contact region varies in the wear process the operator A is time-dependent. The unknown boundary Γ of the contact region is obtained from the following condition:

$$p(x, y, t)|_{(x,y)\in\Gamma} = 0$$

For some wear contact problems, the value of the linear wear $w_*(x, y, t)$ is comparable to the size of the body in contact; then the relation (3) between the elastic displacement and pressure becomes more complex and time dependent. In particular, it can depend on the geometry of the worn body. The method to solve the wear contact problem for elastic body coated by thin elastic layer with variable thickness is described in Goryacheva (1998).

Equations 1 and 3 must be considered with the contact conditions:

$$u(x, y, t) + w_*(x, y, t) = \Phi(x, y, t), \tag{5}$$

where function $\Phi(x, y, t)$ depends on the shapes of contacting bodies and the relative motion features. For example, in wearing of the elastic half-space by the cylindrical punch with the shape $f(x)$ the function $\Phi(x,t)$ has the form $\Phi(x, t) = D(t) - f(x)$ if punch moves along its generatrix. If punch moves in the direction perpendicular to its generatrix the function $\Phi(x, t)$ has the form $\Phi(x, t) = D(t) - f(x - Vt)$, where $D(t)$ is the approach of contacting bodies at instant t.

Equations 1, 3, and 5 provide the complete system of equations for determining the contact pressure $p(x, y, t)$, the shape of the worn surface $w_*(x, y, t)$, and the elastic displacement $u(x, y, t)$. It must be noted that if the

approach function $D(t)$ is not given, but it is known the normal component $F_n(t)$ of the load applied to the contacting bodies, the equilibrium equation is used to complete the system of equations:

$$\int\limits_{\Omega(t)}\int p(x,y,t)\psi(x,y)\,dxdy = F_n(t). \tag{6}$$

where $\psi(x, y)$ is the function determined from the consideration of the geometry of contacting bodies. For example, $\psi(x, y) = 1$ for the punch with flat base, $\psi(x, y) = \cos x$ for the cylindrical punch.

Method of Solution of the Linear Wear Contact Problems

Wear contact problems are reduced to solution of the systems of equations: integral, differential, or integro-differential depending on the contact models using in problem formulation. These systems may be linear or nonlinear.

Linear system arises if the wear equation is linear ($\beta_1 = 1$ and $\beta_2 = 1$ in (2)), and the operator A is linear and time independent. It occurs for some special shapes of contacting bodies and for such types of relative motion providing the constant contact region in wear processes. It must be noted that if $V(x, y) = \text{const}$, it is not necessary to formulate the additional restrictions for the value of integer β_2.

Using the simplified models to describe the compliance of a contacting body (e.g., the Winkler foundation model) reduces the problem to the system of differential equations.

The method of Laplace transform in time was developed in Korovchinsky et al. (1971) to solve some linear wear contact problems.

The general methods of solution of the linear wear contact problems for the elastic bodies based on the analysis of the spectral properties of the integral operators are described in Goryacheva (1998).

If the steady-state (at $t \to \infty$) solution of the system of equations 2 at ($\beta_1 = 1$ and $\beta_2 = 1$), 3, and 5 exists, the contact pressure distribution at any instant of time can be presented in the form (Goryacheva 1998):

$$p(x,y,t) = p_\infty(x,y) + \varphi(x,y,t), \tag{7}$$

where $p_\infty(x, y)$ is the solution of equations 2, 3, 5, and 6 at $t \to \infty$.

The algorithm to find the unknown function $\varphi(x, y, t)$ is illustrated in what follows for the particular case $D(t) = D_\infty t$. Using the method of separation of variables gives:

$$\varphi(x,y,t) = \psi(x,y)T(t). \tag{8}$$

Then it follows from (2), (3), (5), (6), and (7):

$$\dot{T}(t)A[\psi(x,y)] = -K_w(x,y)V_\infty(x,y)\psi(x,y)T(t)$$

or

$$\dot{T}(t) + \lambda T(t) = 0, \tag{9}$$

$$q(x,y) - \lambda A_1[q(x,y)] = 0, \tag{10}$$

where

$$q(x,y) = K_w(x,y)V_\infty(x,y)\psi(x,y),$$
$$A_1[q(x,y)] = A\left[\frac{q(x,y)}{K_w(x,y)V_\infty(x,y)}\right]. \tag{11}$$

If A_1 is a continuous, self-adjoint, and positive definite operator, its eigenvalues λ_n are positive, and the system of its eigenfunctions $U_n(x, y)$ is complete and orthonormalized in the space of continuous functions.

Then, as it follows from (7), (8), and (9), the contact pressure at an arbitrary instant of time is presented by the series:

$$p(x,y,t) = p_\infty(x,y) + \frac{1}{K_w(x,y)V_\infty(x,y)}\sum_{n=1}^{\infty} A_n U_n(x,y)\exp(-\lambda_n t),$$
$$p_\infty(x,y) = \frac{D_\infty}{K_w(x,y)V_\infty(x,y)}. \tag{12}$$

Coefficients A_n are determined by the expansion of the known function for contact pressure at the initial instant of time $t = 0$ in the series of eigenfunctions $U_n(x, y)$:

$$p(x,y,0) = p_\infty(x,y) + \frac{1}{K_w(x,y)V_\infty(x,y)}\sum_{n=1}^{\infty} A_n U_n(x,y).$$

is determined by the equation obtained from (2) for $\alpha = 1$ and (12):

$$w_*(x,y,t) = D_\infty t + \sum_{n=1}^{\infty}\frac{A_n}{\lambda_n}U_n(x,y)\,(1 - \exp(-\lambda_n t)). \tag{13}$$

This method has been used to solve the wear contact problems for the curved beam and the elastic half-plane (Galin 1976), and for initially bent circular beam and the inside surface of a cylinder (Goryacheva 1998). The solution of the last problem was used for the analysis of the wear kinetics and sealing properties of a piston ring, and to evaluate its life-time.

Analysis of the spectral properties of the integral operators in the plane contact problems has been

performed in Aleksandrov and Manzhirov (1987). It was shown that the operator:

$$A[\phi] = \int_{-1}^{1} (-\ln|\xi - x| + C)\phi(\xi)d\xi,$$

of the corresponding contact problem is continuous, self-adjoint, and positive definite acting from the complete Hilbert space $L_2(-1,1)$ in $L_2(-1, 1)$. The wear of the thick elastic layer by a cylindrical punch moving along its generatrix was studied. The different problem formulations were considered: for the given function of punch displacement in time, for the given function of load variation in time, taking into account the roughness of contacting bodies, heating in friction process, and so on.

The axisymmentric wear contact problems for the ring punch ($a < r < b$) rotating at the elastic half-space boundary were considered in Goryacheva (1998). It was shown that the integral operator A_1 for axisymmetric problems:

$$A_1[\phi] = \frac{4(1-\nu^2)}{\pi E}\int_a^b K\left(\frac{2\sqrt{r\rho}}{r+\rho}\right)\frac{\phi(\rho)d\rho}{r+\rho}$$

is continuous, self-adjoint, and positive definite. Asymptotic expression of the contact pressure for $t \to \infty$ has the form:

$$p_\infty(r) = \frac{F_\infty}{2\pi(b-a)r},$$

where F_∞ is the asymptotic value of the load applied to the punch.

Analysis of the wear kinetics in contact of elastic bodies of complex geometry including the wear problems in discrete contact is developed and discussed in Goryacheva (1998).

Methods of Solution of the Nonlinear Wear Contact Problems

Nonlinear wear contact problems arise due to variation of the contact region in wear process, nonlinearity of wear equation, and changes of the mechanical properties and temperature of the subsurface layers in wear properties.

Numerical methods are widely used to solve nonlinear wear contact problems. In what follows, some typical algorithms for numerical study of the problems are described.

One of the traditional approaches to finding the approximate solution of such evolutionary problems is to use a step-by-step method based on time-digitization and using finite difference representation of differential or integral operators of the contact problem. This method of solving nonlinear wear contact problems in different formulations (in application to plane and spherical bearings, coated guides, etc.) was used in Goryacheva and Dobychin (1984), Soldatenkov (1987), Teply (1987), Usov (1985), and Johansson (1994).

The wear contact problems with variable contact region can be solved using mapping of the variable contact region into fixed region (Soldatenkov 1987), that is, if x is the coordinate, $[-a(t), a(t)]$ is the contact region, $\upsilon = da/dt$ is the velocity of the contact region growth, then the wear equation 1 can be written in the form:

$$\frac{\partial W(X,t)}{\partial t} - \upsilon(t)\frac{X}{a(t)}\frac{\partial W(X,t)}{\partial X} = F(P(X,t)) \qquad (14)$$

where $X = x/a$ is the dimensionless coordinate, and $W(X, t) = w_*(x, t)$ and $P(X, t) = p(x, t)$ are the dimensionless wear and contact pressure distributions, respectively. Turning to the fixed contact region simplifies the numerical procedure of the problem solution, since it allows the possibility of using the fixed nodes of a network.

The other idea providing the simplification of the analysis of the contact problem with growing contact region is using the contact radius a instead of the time t (Usov 1985; Aleksandrov and Manzhirov 1987; (Soldatenkov 1987). In this case the wear equation 1 takes the form: $\partial W(X, a)/\partial a = \upsilon^{-1}(a)\ F\ (P(X, a))$, and the relationship (14) is reduced to the following:

$$\frac{\partial W(X,a)}{\partial a} - \frac{X}{a}\frac{\partial W(X,a)}{\partial X} = \frac{1}{\upsilon(a)}F(P(X,t)),$$

Integration of this equation is made using the known characteristics of the differential operator in the left-hand side of the equation.

In wear contact problems for the cylindrical bodies, the effective method uses the approximate representation of the contact pressure in the following form (Usov 1985):

$$p(x,t) = \sum_{k=1}^{N}\lambda_k(t)\sin kx.$$

Substitution of this expression into the system of the integro-differential equations 1, 3, 5, and 6 leads to the solution of the simplified system of ordinary differential and nonlinear algebraic equations. Based on the method, it is possible to study the problems with nonlinear wear equation, with friction, to take into account the simultaneous wear of both contacting bodies.

The asymptotic approach for the solution of the nonlinear wear contact problems was used in Aleksandrov and Kovalenko (1982) to analyze the wear kinetics in plane journal bearings and plane guides for $t \to 0$. It was

established that there is singularity in the rate of the contact region growth at the initial instant of time.

Asymptotic Solution of the Wear Contact Problems for Homogeneous Materials at $T \to \infty$

Some approaches for the wear contact problem analysis are based on the asymptotic properties of the solution.

In the wear process the displacements of the contacting bodies due to their deformation are usually limited, whereas in many cases the linear wear is the monotonically increasing function in time. So for the high values of wear it is possible to neglect by the surface displacements due to deformation in the contact condition (5), and to consider the wear process of rigid bodies in contact (Pronikov 1998). Using this approach makes it possible to calculate the worn shape of contacting bodies and to reduce the analytical expressions for the contact characteristics in wear process. For example, the following dependence of the total wear I of the radial sliding bearing on the half contact angle a was reduced in Usov (1985):

$$I = 2R\Delta(\tan\, a - a)$$

where R and Δ are the radius and the gap of the bearing.

The other feature of the wear contact problems is the existence of the steady-state solution at $t \to \infty$ under definite conditions. This solution is characterized by the uniform linear wear rate within the contact region:

$$\frac{\partial w_*(x,y,t)}{\partial t} = \frac{dD(t)}{dt} = D_\infty.$$

A necessary condition for existence of the steady-state stage of the wear process is stabilization in time of all the external parameters of the problem (the rate of the approach $\lim\limits_{t\to+\infty} \dot{D}(t) = D_\infty$ of contacting bodies or stabilization of the normal load $\lim\limits_{t\to+\infty} F(t) = F_\infty$, velocity of the relative motion of the bodies $\lim\limits_{t\to+\infty} V(x,y,t) = V_\infty(x,y)$, the contact region $\lim\limits_{t\to+\infty} \Omega(t) = \Omega_\infty$).

Under this condition and under the assumption that operator A (3) is time independent, the sufficient conditions for the asymptotic stability of the steady-state solution were formulated in Goryacheva (1998) both for linear and nonlinear wear contact problems.

The pressure distribution also tends to the steady-state pressure p_∞, which depends on the contact geometry, on the relative motion of contacting surface, and the type of the wear equation (Goryacheva 1998):

$$p_\infty(x,y) = \lim_{t\to+\infty} p(x,y,t) = \left[\frac{D_\infty}{K_w V_\infty^\beta(x,y)}\right]^{1/\alpha}.$$

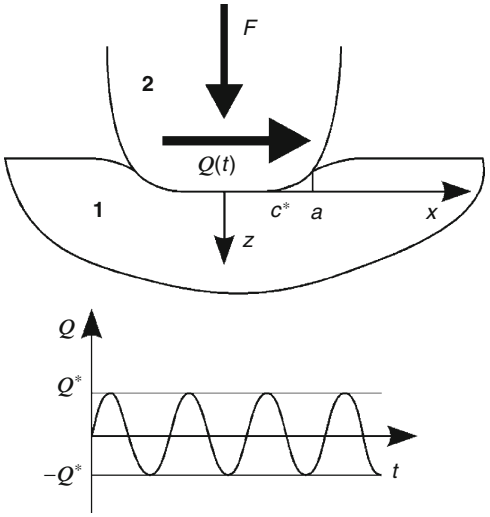
Substituting this equation into the equilibrium equation 6 makes it possible to obtain the formula for determining the steady-state normal load F_∞:

$$F_\infty = \left(\frac{D_\infty}{K_w}\right)^{1/\alpha} \iint\limits_{\Omega_\infty} \frac{dxdy}{V^{\beta/\alpha}(x,y)}. \qquad (15)$$

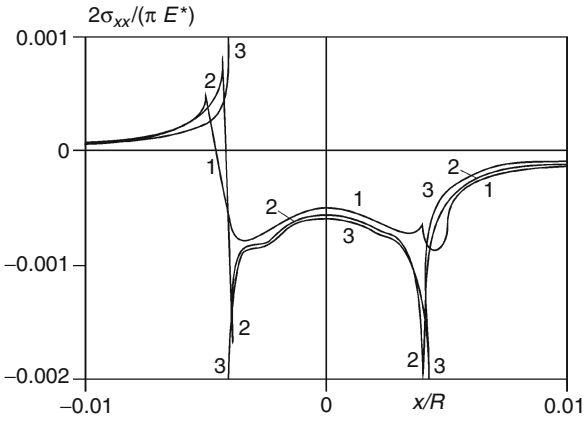
If $D_\infty = 0$ or $f_\infty = 0$, the contact pressure tends to zero (i.e., $p_\infty\,(x, y) = 0$).

Figure 1 illustrates the contact with partial slip of the elastic half-space and the indenter with similar elastic properties. The wear occurs within the slip zones. This problem was studied in Goryacheva et al. (2001) for two types of the initial geometry of the indenter: the parabolic cylinder and for the cylindrical indenter with a flat base and the rounded edges. The bodies are subjected to the constant normal force F and the oscillating tangential force $Q(t)$ so that $-Q^* \le Q(t) \le Q^*$. The following condition is satisfied: $|Q^*| < \mu F$, where μ is the friction coefficient.

It was proved in Goryacheva et al. (2001) that the wear contact problem under consideration has the asymptotic solution. The contact pressure approaches to one which has a singularity at the ends of the stick zone and is zero outside the stick zone. Figure 2 illustrates the redistribution of σ_{xx} component of the stress tensor in wear process. The curve 3 corresponds to the asymptotic solution at $t \to \infty$. The analytical expressions for



Wear Contact of Homogeneous Materials, Fig. 1 Scheme of the contact



Wear Contact of Homogeneous Materials, Fig. 2
Distribution of the σ_{xx} component of the stress tensor for various numbers of cycles: $N = 0$ (*curve 1*), $N = 4{\cdot}10^4$ (*curve 2*), $N \to \infty$ (*curve 3*)

the asymptotic solution (contact pressure and shear stress, tension, shape of the worn surface, etc.) are presented in Goryacheva et al. (2001) for the both cases of contact geometries. In particular, the asymptotic contact characteristics were calculated for the various shapes $f(x)$ of contacting bodies. The following relationships were obtained for the contact pressure $p_\infty(x)$:

$$p_\infty(x) = \frac{E^*}{2\pi\sqrt{c^{*2}-x^2}} \int_{-c^*}^{c^*} \frac{\sqrt{c^{*2}-t^2}\, f'(t)dt}{t-x} + \frac{P}{\pi\sqrt{c^{*2}-x^2}}, \quad |x| < c^*, \qquad p_\infty(x) = 0, \quad c^* < |x| \le a_\infty; \tag{16}$$

and for the shape of the worn surface $w_\infty(x)$:

$$w_\infty(x) = 0, \quad |x| < c^*, \qquad w_\infty(x) = \frac{1}{1-\gamma}\left[f(a_\infty) - f(x) - \frac{2}{\pi E^*}\int_{-c^*}^{c^*} p_\infty(x') \ln\left|\frac{a_\infty - x'}{x - x'}\right| dx'\right], \quad c^* < |x| \le a_\infty. \tag{17}$$

where $-c^*$ and c^* are the ends of the stick zone determined from the equation:

$$-\frac{\mu E^*}{2}\int_{-c^*}^{c^*} \sqrt{\frac{c^*-t}{t+c^*}}\, f'(t)\, dt = \mu P - Q^*, \tag{18}$$

parameter γ characterizes the portion of the worn material removing from the contact region ($0 \le \gamma \le 1$), and $(-a_\infty, a_\infty)$ is the contact region at $t \to \infty$.

The results of this study are used for the analysis of the evolution of the contact characteristics in fretting, and for prediction of the crack initiation in fretting fatigue.

The operating of the junctions at the steady-state regime of wear is more preferable, since the stabilization of pressure and other contact characteristics, and minimization of wear rate and value of wear occur at this stage of the process. The mathematical problems of wear process optimization are formulated and studied in Goryacheva (1998).

Key Applications

The solution of wear contact problems makes it possible to evaluate the running-in and life-time of different tribological junctions, and to control the wear process. Based on the wear contact problem solutions, the contact characteristics' evolution and life-time were studied in the following tribological junctions and units:

- Sliding bearings (Goryacheva and Dobychin 1984; Usov 1985; Aleksandrov and Manzhirov 1987; Teply 1987).
- Rail and wheels (Goryacheva 1998).
- Piston rings (Goryacheva 1998).
- Fretting fatigue process in various junctions (Johansson 1994; Goryacheva et al. 2001).

Cross-References

▶ Sliding Wear

References

V.M. Aleksandrov, E.V. Kovalenko, To the theory of the contact problems in the presence of nonlinear wear. Izv. A.N. SSSR Mech. Solids **4**, 98 (1982)

V.M. Aleksandrov, A.V. Manzhirov, Analytical solution of the contact problem for the plane bearing in the presence of wear. Frict. Wear **8**(4), 985 (1987)

L.A. Galin, Contact problems of the theory of elasticity in the presence of wear. J. Appl. Math. Mech. **40**(6), 981–986 (1976)

I.G. Goryacheva, *Contact Mechanics in Tribology* (Kluwer, Dordrecht/Boston/London, 1998)

I.G. Goryacheva, M.N. Dobychin, Kinetics of the wear process of the solid lubricant in plane bearing. Frict. Wear **5**(4), 581 (1984)

I.G. Goryacheva, R. Pakalapa, T. Farris, Wear in partial slip contacts. ASME J. Tribol. **123**(4), 848 (2001)

L. Johansson, Numerical simulation of contact pressure evolution in fretting. Trans. ASME. J. Tribol. **116**, 247 (1994)

M.V. Korovchinsky, Local contact of elastic bodies with wear of their surface, in *Contact Interaction of Solid Bodies and Calculation of Friction Forces and Wear*, Nauka, Moscow (1971), p. 130

I.V. Kragelsky, M.N. Dobychin, V.S. Kombalov, *Friction and Wear: Calculation Methods* (Pergamon, Oxford, 1982)

A.S. Pronikov, Macrotribology and its goals. Frict. Wear **19**(2), 155 (1998)
I.A. Soldatenkov, Wear of coating in elastic junctions with varying contact region. Sov. J. Frict. Wear **8**(2), 206 (1987)
M.I. Teply, Calculation of contact characteristics and wear in cylindrical sliding bearings. Sov. J. Frict. Wear **8**(5), 895 (1987)
P.P. Usov, Inner contact of cylindrical bodies of near the same radii in the presence of wear. Sov. J. Frict. Wear **6**(3), 404 (1985)

Wear Contact of Inhomogeneous Materials

Irina Goryacheva
Ishlinsky Institute for Problems in Mechanics of Russian, Russian Academy of Sciences, Moscow, Russia

Synonyms

Wear modeling for inhomogeneous materials

Definition

Wear contact problems for inhomogeneous materials study the evolution of all contact characteristics (pressure distribution, position and size of contact region, shape of contacting bodies, approach of the bodies, etc.) in the wear process, taking into account the parameters of inhomogeneity of the contacting bodies.

Scientific Fundamentals

The mathematical formulation and solution of the wear contact problems depend on the type of inhomogeneity of contacting bodies (layered bodies, bodies with inclusions, bodies with nonuniform wear coefficient, composite materials, etc.).

Wear Contact Problems for Coated Bodies

Similar to wear contact problems for homogeneous bodies, mathematical problem formulation for layered elastic bodies includes the wear equation that establishes a relation between some characteristics of wear and a set of parameters characterizing the properties of friction surfaces and operating conditions. In general, the linear wear rate $\partial w_*/\partial t$ can be presented as the function of the contact pressure p and sliding velocity V

$$\frac{\partial w_*}{\partial t} = F(p, V). \tag{1}$$

The form of the function F depends on the type of wear (abrasive, fatigue, etc.), temperature, and so on. For many types of wear the function F can be approximated by the power function, so (1) has the form

$$\partial w_* \partial t = K_w p^{\beta_1} V^{\beta_2}, \tag{2}$$

where K_w is the wear coefficient, and β_1 and β_2 are parameters that depend on material properties, friction conditions, temperature, and so on.

The pressure $p(x, y, t)$ within the contact region and the surface normal displacement $u(x, y, t)$ for an arbitrary instant of time t are related by operator A

$$u(x, y, t) = A_t[p(x, y, t)]. \tag{3}$$

Since the thickness of the surface layer decreases in the wear process, the operator A_t in the relation (3) is time dependent.

Equations (1) and (3) must be considered with the contact conditions:

$$u(x, y, t) + w_*(x, y, t) = \Phi(x, y, t), \tag{4}$$

where function $\Phi(x, y, t)$ depends on the shapes of contacting bodies and the relative approach $D(t)$.

Equations (1), (3), and (4) provide the complete system of equations for determining the contact pressure $p(x, y, t)$, the shape of the worn surface $w_*(x, y, t)$, and the elastic displacement $u(x, y, t)$. It must be noted that if the approach function $D(t)$ is not given, but it is known the normal component $P(t)$ of the load applied to the contacting bodies, the equilibrium equation is used to complete the system of equations (1), (3), and (4):

$$\int\limits_{\Omega(t)} \int p(x, y, t)\ dxdy = P(t). \tag{5}$$

The method to solve the wear contact problem for the elastic half-space coated by a thin elastic layer with variable thickness is described in Goryacheva (1998). It was shown that the worn thin layer compliance can be described by the relationship:

$$u = Bhp, \quad h = h_0 - w_*,$$

where h_0 is the initial layer thickness.

The two-dimensional problem was reduced to the following integral equation to find the unknown contact pressure $p(x, t)$ within the fixed contact region $(-a, a)$ at the arbitrary instant of time t (Goryacheva 1998):

$$\begin{aligned} &h(x, t) - h(0, t) - \\ &- \frac{A}{E_1^*}[h(x, t)p(x, t) - h(0, t)p(0, t)] + \\ &+ \frac{1}{\pi E_2^*} \int\limits_{-a}^{a} p(x', t) \ln\left|\frac{x' - x}{x'}\right| dx' = f(x). \end{aligned} \tag{6}$$

Here, $f(x)$ is the shape of the punch, which is the cylinder moving along its generatrix $0y$, $E_i^* = E_i/2(1-\nu^2)$ ($i = 1$ or $i = 2$ for the layer or for the half-plane, respectively). It is assumed here that the Poisson ratio is the same for the layer and the half-space. The value of the constant A depends on the conditions at the layer-substrate interface: $A = 0.5$ if there is no friction at the boundary between layer and a half-space; if there is complete adhesion at the boundary, then

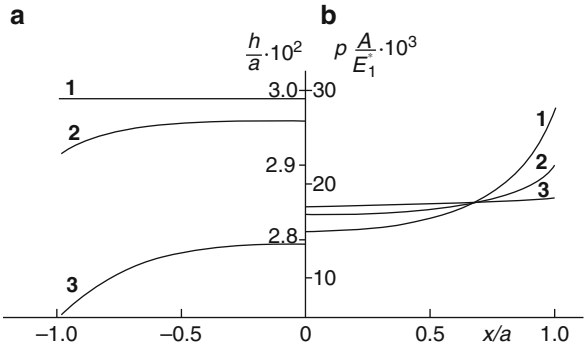
$$A = \frac{2(1-n)(\chi-1)(2-n+n\chi)}{(\chi+1)^2},$$
$$n = \frac{E_1^*}{E_2^*}, \quad \chi = 3-4\nu.$$

The variable layer thickness in the arbitrary instant of time is determined from the relationship that follows from the wear equation (2):

$$h(x,t) = h_0 - K_w \int_0^t \left[\frac{p(x,t')}{p^*}\right]^{\beta_1} dt'. \tag{7}$$

The integral equation (6) is nonlinear. The step-by-step method was used to solve the system of equations (6) and (7).

Figure 1 illustrates the worn profiles and the pressure distributions in contact of the punch with flat base $x \le a$ and the coated elastic half-space at various times t ($t = a\tau/K_w$) of the wear process. Calculations were made for the following dimensionless parameters: $\beta_1 = 1.4$, $\bar{P} = 9 \cdot 10^{-3}$ $\left(\bar{P} = P/\left(aE_2^*\right)\right)$, $n = 0.13$, and under the assumption that the surface layer is bonded to the substrate (Goryacheva and Soldatenkov 1983). This case is applied to study the wear of solid lubricant coatings.



Wear Contact of Inhomogeneous Materials, Fig. 1 Profile of the worn coating surface (**a**) and pressure distribution (**b**) within the contact region during the wear process: $\tau = 0$ (*curve 1*), $\tau = 0.15$ (*curve 2*), $\tau = 0.64$ (*curve 3*)

The results indicate that the contact pressure tends to the uniform steady-state distribution in the wear process for chosen values of parameters. However, the steady-state wear stage may not be realized for very thin coatings, since the complete wear in some points can be reached earlier.

The wear contact problem for the punch and the layered elastic body with monotonically increasing contact area was studied in Soldatenkov (1985) using the iteration method to determine the evolutions of contact pressure and the contact width in the wear process.

The wear of the coating compliance described by the nonlinear function $u = hA(p)$ was analyzed in Soldatenkov (1990). The following relationship for the contact pressure evolution in the wear process was reduced:

$$\frac{\partial p(x,t)}{\partial t} = \frac{\bar{\varphi}(t) - \varphi(p(x,t))}{h(x,t)A'(p(x,t))}. \tag{8}$$

Here, $\varphi(p) = F(p)(1-A(p))$, $\bar{\varphi}(t) = \varphi(p(x,t))$, $\bar{x}$ is some fixed point within the contact region, and $F(p)$ is the wear equation. It follows from (8) that, depending on the function $\varphi(p)$, contact pressure tends to uniformity, or it is in decreasing in some parts of the contact zone, causing catastrophic wear.

The wear contact problem for the multilayered coating with depth-variable compliance $B(y)$ was studied in Soldatenkov (1991) using the Winkler foundation model. The problem was reduced to the wear contact problem for homogeneous body described by the following integral compliance:

$$u(x,t) = \bar{B}(x,t)h(x,t)p(x,t),$$
$$\bar{B}(x,t) = \frac{1}{h(x,t)} \int_0^{h(x,t)} B(y)dy.$$

The similar method was used in Cherskii et al. (1986) to study the wear contact problem for a layered coating in a journal bearing.

The micromechanical finite element models of the pin-on-disc wear test taking into account the plastic strain in the coating during the test are presented in Christofidesa et al. (2002).

The Wear Contact Problems with Surface Inhomogeneities

Different technical methods used for hardening of surfaces change their properties and essentially influence the character of the surface wear during the friction process. Local surface hardening (laser processing, ion implantation, etc.) produces a structural inhomogeneity and, as a consequence, nonuniform wear.

The *wear contact problems with variable wear resistance* were analyzed in Goryacheva (1998). The coefficient K_w in the wear equation (2) is a function of the coordinates (x, y) at the half-space surface, so this equation takes the following form:

$$\frac{\partial w_*(x,y,t)}{\partial t} = K_w(x,y)\left[\frac{p(x,y,t)}{p^*}\right]^{\beta_1}\left[\frac{V(x,y)}{V^*}\right]^{\beta_2}, \tag{9}$$

where p^* and V^* are characteristic values of pressure and velocity.

The mathematical model to study the wear process consists of (3), (4), (5), and (9).

If the operator A in (3) is time independent and either the function dD/dt, which is the rate of surface approach, or the normal load function $P(t)$, possesses an asymptote, that is

$$\lim_{t\to+\infty}\frac{dD}{dt} = \lim_{t\to+\infty}\frac{dw(x,y,z)}{dt} = D_\infty,$$

or

$$\lim_{t\to+\infty} P(t) = P_\infty,$$

then the system of equations (3), (4), (9) (or (3), (4), (5), (9)) permits the stationary solution

$$p_\infty(x,y) = p^*\left[\frac{D_\infty (V^*)^{\beta_2}}{K_w(x,y)V^{\beta_2}(x,y)}\right]^{1/\beta_1}. \tag{10}$$

At the given asymptotic value P_∞ of the normal load, the constant D_∞ is determined from the equilibrium condition (5)

$$D_\infty = \frac{P_\infty^{\beta_1}}{(p^*)^{\beta_1}(V^*)^{\beta_2}}\left[\iint_\Omega \frac{dxdy}{K_w^{1/\beta_1}(x,y)V^{\beta_2/\beta_1}(x,y)}\right]^{-\beta_1}. \tag{11}$$

Sufficient conditions for the asymptotic stability of the solution (10) were established in Goryacheva (1998).

The shape of the worn surface corresponding to the stationary solution (10) can be represented as a sum of a function $f_\infty(x, y)$, which is independent of time (stationary shape) and the time-dependent function $D(t)$.

The steady-state shape of the elastic half-space is described by the function (Goryacheva and Soldatenkov 1983)

$$f_\infty(x,y) = A[p_\infty(x,y)] = A\left[\left(\frac{(p^*)^{\beta_1}(V^*)^{\beta_2}D_\infty}{K_w(x,y)V^{\beta_2}(x,y)}\right)^{1/\beta_1}\right]. \tag{12}$$

Here, operator A takes the form

$$A[p(x,y,t)] = \frac{1-\nu^2}{\pi E}\iint_\Omega p(x',y',t)\varphi(x',y',x,y)\,dx'dy',$$
$$\varphi(x',y',x,y) = \frac{1}{\sqrt{(x-x')^2+(y-y')^2}}. \tag{13}$$

The stationary shape $f_\infty(x, y)$ depends on the geometry of the contacting bodies, wear coefficient $K_w(x, y)$, and the type of the punch motion, i.e., the function $V(x, y)$.

Some wear contact problems for the elastic half-space with variable wear coefficient are considered in Goryacheva (1998). The method is developed there to solve the wear contact problems with the wear coefficient described by the step function $K_w(x, y)$:

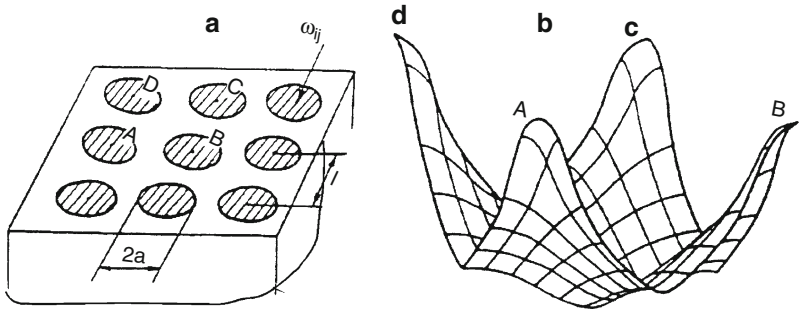
$$K_w(x,y) = \begin{cases} K_{w1,}(x,y) \in \Omega\backslash\omega, \\ K_{w2,}(x,y) \in \omega. \end{cases} \tag{14}$$

Here, ω is the set of the hardened domains (the coefficient K_{w2} characterizes the wear rate within this domain), and $\Omega\backslash\omega$ are the unhardened domain at the half-space surface Ω (the wear rate within this domain is characterized by the wear coefficient K_{w1}).

The method can be applied to study the effect of local hardening on wear process and the shape of the worn surface. It follows from the solution of the *periodic wear contact problems* for the half-space hardened inside the strips or inside the circle domains (these domains form the hardened domain ω), that the half-space surface becomes wavy due to wear. Figure 2 illustrates the scheme of hardening and the shape of the worn surface within the period for the elastic half-space hardened inside the circle domains. Due to wear by the rigid punch (it is assumed that its surface does not wear) with the flat surface moving translationally on the half-space surface in various directions at a constant speed, the initially flat half-space surface becomes wavy (see Fig. 2b). The geometric parameters of the worn surface depend on the dimensionless parameter $m = \frac{K_{w2}}{K_{w1}}$, which characterizes the extent of hardening, and the parameter $\bar{a} = \frac{a}{l}$ (l is the period in both direction x and y), which is a geometric characteristic of hardening. The parameters vary in the ranges $0 < \bar{a} \le 1/2$ and $m_0 < m < 1$, where m_0 is a limit value of m due to processing technology of the half-space surface.

The pressure distribution tends to the step function $p_\infty(x, y)$:

$$\frac{p_\infty(x,y)}{p^*} = \begin{cases} \bar{p}_1, (x,y) \in \Omega\backslash\omega, \\ \bar{p}_2, (x,y) \in \omega, \end{cases} \tag{15}$$



Wear Contact of Inhomogeneous Materials, Fig. 2 Scheme of the hardened domain arrangement at a half-space surface (**a**) and a shape of the worn surface at one period (**b**)

where

$$\bar{p}_1 = \bar{K}_w^{1/\alpha}\left(\frac{P_\infty}{p^* l^2}\right),$$
$$\bar{p}_2 = \left(\frac{\bar{K}_w}{m}\right)^{1/\alpha}\left(\frac{P_\infty}{p^* l^2}\right), \qquad (16)$$
$$m_2 = m^{-1/\alpha} - 1,$$

P_∞ is a load per period. The effective wear coefficient $\bar{K}_w$ in the steady-state stage of the wear process is determined from the expression

$$\bar{K}_w = \frac{D_\infty}{K_{w1}}\left(\frac{p^* l^2}{P_\infty}\right)^\alpha \left(\frac{V^*}{V}\right)^\beta = \frac{1}{1+\pi\bar{a}^2 m_2}. \qquad (17)$$

The effective wear coefficient $\bar{K}_w$ in the steady-state stage of the wear process is equal to 1 for $\bar{a} = 0$ (nonhardened surface) and decreases as the radius $\bar{a}$ of the hardened domain, or the parameter m_2 increases (m decreases).

Thus, the variation of the parameters m and $\bar{a}$ within the limits admissible by technology makes it possible to control the tribological and geometric characteristics of the wavy surfaces generated due to wear.

The same method was used in Chekina and Keer (1998) to study the effect of polishing of the composite material consisting of periodically distributed two phases, on the evolution of the surface shape. The steady-state solution of the periodical wear contact problem for the composite material wearing by the viscoelastic body was developed in Lyubicheva (2006). The results are used to analyze the effect of sliding velocity and wear equation parameters on formation of a steady state shape of inhomogeneous body.

Wear contact problems for inhomogeneous bodies with a bounded contact region possess additional peculiarities due to the edge effects. The steady-state shapes of the worn surface of an annular-end punch hardened inside the concentric rings or inside the sectors were studied in Goryacheva (1998). It was supposed there that the punches are worn by the elastic half-space. Figure 3b illustrates the shape of the worn surface at a radial cross-section for different arrangement of the hardened domain, which are the rings of given width ρ. The dependence $f_2(\xi) = \frac{f(\xi)\pi E}{2(1-\nu^2)p^* \bar{p}_1 (r_N+\rho)}\left(\xi = \frac{r}{r_N+\rho}\right)$ when hardened domains are arranged uniformly along the radius is depicted by the solid line ($N = 5$). Due to the boundedness of the contact region, the hardened domains wear nonuniformly (the edge effect). The location of the hardened domains significantly affects the stationary shape of the surface. By varying the parameters r_i, it is possible to obtain the shape of the worn surface, satisfying, for instance, the condition $f(r_i + \rho/2) = \text{const}$. It is represented by the dashed line in Fig. 3b.

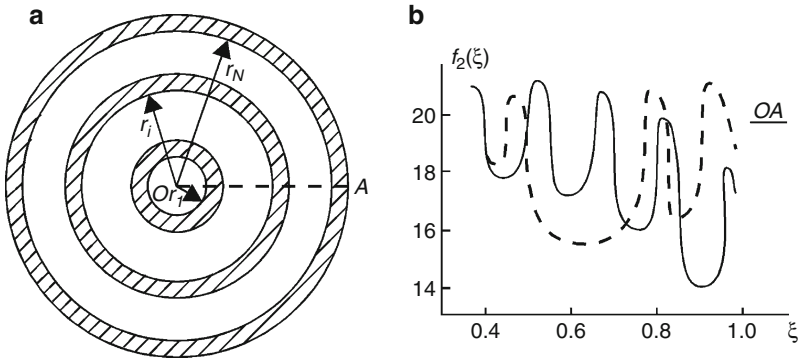
The results open the possibility to control the worn surface shape by using the specific local geometric and tribotechnical hardening parameters.

The model of wear process in discrete contact was developed and used to study the *wear of inhomogeneous bodies with inclusions* (Goryacheva 1998). This model takes into account the interaction between contact spots. It describes the redistribution of the contact pressure during running-in, and the steady-state nominal pressure and the worn surface shape depending on the density and size of inclusions, type of motion, and so on.

Key Applications

Study of wear kinetics makes it possible to predict the durability of the moving parts of machines during operation.

The design of a sliding pair with a protective coating that prevents severe wear and decreases the friction losses



Wear Contact of Inhomogeneous Materials, Fig. 3 Scheme of hardened domain arrangement (**a**) and the shape of the worn surface at the radial cross-section (**b**) for uniform distribution of the hardened annular domains (*solid line*) and for the case $\xi_1 = 0.35$, $\xi_2 = 0.47$, $\xi_3 = 0.78$, $\xi_4 = 0.925$ $\xi_5 = 0.975$ (*dashed line*), $\rho/r_1 = 0.2$

is of interest for engineering. The models described above make it possible *to evaluate the lifetime of the coated friction elements* in different junctions. The wear of a thin antifriction coating in plain journal bearings with the coating either at the bush or at the shaft surface was studied in Goryacheva and Dobychin (1984) and in Goryacheva (1998). The wear process for two-layered coatings in journal bearings was analyzed in Cherskii et al. (1986).

The parameters of surface inhomogeneity, such as relative size and wear coefficient of hardened and unhardened zones and density of different phases or density of inclusions in composite material, influence the shape variation of the surface in the wear process. Based on the solutions of the corresponding wear contact problems it is possible *to predict the evolution of the contact characteristics of friction pairs* with various types of inhomogeneity if these parameters and other characteristics of wear process are known. Based on the model analysis, the evolution of the shape of the composite material in chemical mechanical polishing was predicted in Chekina and Keer (1998).

Study of the wear contact problems makes it also possible to develop the methods of *controlling the wear process* by

- Decrease of the running-in time and provision of savings in material by making the initial surface shape approach the steady-state shape calculated from the models (problem 1)
- Provision of the desired characteristics of junction in the steady-state wear by approach of the steady-state shape of the surface to the optimal shape (problem 2).

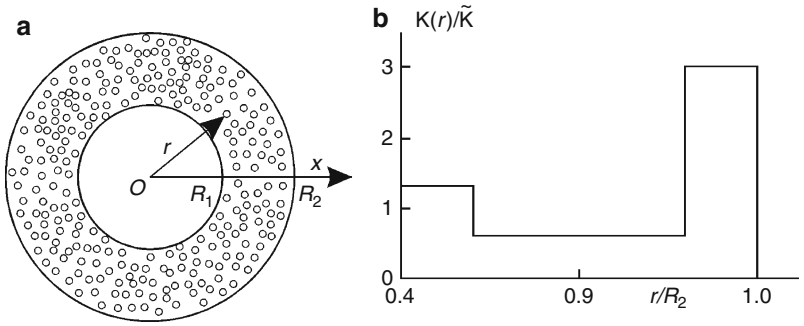
The problem of stabilization of the optimal shape $f_s(x, y)$ of the worn surface (problem 2) can be formulated if there are some parameters $\gamma_i(x, y)$ characterizing, for instance, the inhomogeneity of contacting bodies, that admit the variation within the definite class of functions (Goryacheva and Chekina 1989). The following integral equation in respect of the one or several unknown functions $\gamma_i(x, y)$ must be solved:

$$f_s(x, y) = A[p_\infty(x, y),\ \gamma_i(x, y)]. \qquad (18)$$

Usually the parameters $\gamma_i(x, y)$ have the practical limitations imposed by the technology used in obtaining the inhomogeneous surfaces. Therefore, the problem is reduced to minimization of the functional (Goryacheva 2007).

In wear process of the locally hardened surface discussed above, the governing parameters characterize the hardening intensity and the relative size of the hardened domains. The variation of these parameters admissible from the technological requirements allows control of the tribological and geometrical characteristics of the wave surface formed in the wear process. For instance, by varying the arrangement of the hardened zones, which are the rings of definite thickness, it is possible to get the worn shape $f(r_i + \rho/2) = \text{const}$ that could be considered there as the optimal surface shape (see Fig. 3b).

For discrete contact, the steady-state shape essentially depends on the distribution of the contact spots and their mutual influence. The solution of problem 2 for the abrasive tool surface is obtained in Goryacheva (1998). The method of calculation of the optimal distribution of the abrasive inclusions at the tool surface providing its uniform wear is proposed based on this solution.



Wear Contact of Inhomogeneous Materials, Fig. 4 Abrasive tool surface with inclusions (**a**) and variation of the inclusion density versus radius providing the condition $f_s(r)$ = const (**b**)

The function $\kappa(x, y)$, characterizing the contact density of inclusions, was used to control the wear process. Since it is impossible to manufacture the tool with the density varying continuously, the solution was sought in the class of step functions. The results of calculations for the tool operation surface in the form of the ring, rotating with the constant angular velocity on the boundary of the elastic half-space are presented in Fig. 4b. The function guarantees the surface to be practically flat during the wear process. This function consists of three different parts because the densities different by less than 10% were considered to be indistinguishable for technological reasons.

Wear modeling makes it possible both to predict the main features of the wear process at micro- and macroscales for various friction conditions, and to control this process, providing the optimal surface shape of the elements of friction pairs and minimizing the running-in time. This leads to savings in material used for the friction pairs and to an increase in the lifetime of junctions.

Cross-References

► Contact of Layered Materials
► Contact of Materials with Inhomogeneities
► Wear Contact of Homogeneous Materials

References

O.G. Chekina, L.M. Keer, Wear contact problems and modeling of chemical mechanical polishing. J. Electrochem. Soc. **145**(6), 2100 (1998)
I.N. Cherskii, O.B. Bogatin, L.G. Sokolnikova, Calculation of operation characteristics of antifriction bush and coatings under gross wear. Friction Wear **7**(1), 99 (1986)
C. Christofidesa, P.E. McHugh, A. Fornb, J.A. Picas, Wear of a thin surface coating: modelling and experimental investigations. Comput. Mater. Sci. **25**(1–2), 61 (2002)
I.G. Goryacheva, *Contact Mechanics in Tribology* (Kluwer, Dordrecht, 1998)
I.G. Goryacheva, Modelling of wear of solids at different scales. Phys. Mezomech. **10**(5), 31 (2007)
I.G. Goryacheva, O.G. Chekina, Control of surface deformation during wear. Sov. J. Friction Wear **10**(1), 1 (1989)
I.G. Goryacheva, M.N. Dobychin, Kinetics of the wear process of the solid lubricant in plane bearing. Friction Wear **5**(4), 581 (1984)
I.G. Goryacheva, I.A. Soldatenkov, Theoretical study of running-in and steady-state stage of wear of solid lubricant coatings. Friction Wear **4**(3), 420 (1983)
A.N. Lyubicheva, Steady-state solution of the periodical problem for wear of composite material by viscoelastic body. Friction Wear **27**(5), 465 (2006)
I.A. Soldatenkov, Wear of a thin elastic coating with a varying contact area. Friction Wear **6**(2), 247 (1985)
I.A. Soldatenkov, The peculiarities of the contact pressure in wear of coatings. Friction Wear **11**(6), 973 (1990)
I.A. Soldatenkov, Analysis of a wear process of a multilayer coating. Friction Wear **12**(2), 204 (1991)

Wear Contact Problems

► Wear Contact of Homogeneous Materials

Wear in Contact with Micro Vibration

► Contact with Micro Vibrations for Friction Control and Wear Reduction

Wear in Gears

Ahmet Kahraman[1], Huali Ding[2]
[1]Department of Mechanical Engineering,
The Ohio State University, Columbus, OH, USA
[2]The Ohio State University, Columbus, OH, USA

Synonyms

Abrasive wear of gears; Sliding wear; Sliding wear in mixed EHL contacts

Definition

Surface wear is one of several failure modes experienced at the interface of lubricated gear tooth contacts that must operate under combined sliding and rolling motions. While scuffing and tooth bending fatigue failures have their own mechanisms, contact fatigue failures such as micro-pitting and spalling compete with surface wear as potential failure modes along the lifespan of gear components. Apart from the direct material loss that leads to functional failure, surface wear causes the gear system to change its vibration and noise characteristics significantly as the gear mesh excitations are very sensitive to surface geometry. Surface wear affects the patterns of gear contact in such a way that contact stresses and load distributions are altered to accelerate the occurrence of other failure modes. Hence, a better understanding of mechanisms of gear wear, including its impact on noise and durability, is essential. In addition, knowledge of the sensitivity of design parameters and manufacturing errors on wear becomes crucial for a gear designer who attempts to minimize wear. This chapter provides a brief description of models used to predict gear wear and discusses the influence of gear tooth modifications and lubricant parameters on wear, describing the interactions between dynamic behavior and gear wear.

Other phenomena associated with micro-pitting and scuffing also result in removal of material from the tooth surfaces. In addition, certain manufacturing processes such as gear lapping can be viewed as a form of accelerated surface wear. This chapter will not include such surface changes and focuses solely on long-cycle, mild wear in gears.

Scientific Fundamentals

Gear Contact Parameters Influencing Surface Wear

A large number of parameters must be taken into account to describe wear characteristics of any contacting surfaces accurately. Mechanisms of gear wear are even more complex as the contact parameters such as radii of curvature, sliding velocities, and normal load vary with contact location. Local contact deflections and tooth bending effects also complicate the contact conditions. In addition, other contact parameters influencing the elastohydrodynamic lubrication conditions (lubricant parameters and surface roughness characteristics) are also among the parameters of interest for gear wear. The coupling between the contact conditions and the actual geometry of the contact surfaces that change as they wear can cause inaccurate estimations of contact stresses, resulting in unrealistic life predictions as well. The key to the prediction of the performance of a gear pair accurately over an extended number of wear (load) cycles is the knowledge of evolution of the geometry of the mating tooth surfaces as they wear so that the pressure distribution in the contact zone and, hence, wear rates of mating components can be predicted accurately.

Any sliding wear of a lubricated surface can be described by the initial value problem

$$\frac{dh}{ds} = \Im \tag{1}$$

where h is the wear depth, s is the relative sliding distance, and $\Im$ is a function of a number of parameters influencing wear, including contact pressure p, rolling (V_r) and sliding (V_s) velocities and material hardness, and surface roughness amplitudes and texture, as well as lubricant-related parameters. Wear accumulation can be calculated by integrating (1), provided that the function $\Im$ is known. Archard (1953) proposed a simple model that takes probability of asperity collision into account in the form of a wear coefficient k. Although many researchers (e.g., Williams 1999) proposed, over the years, more advanced wear models using different methodologies and parameter sets, Archard's wear model still remains the most commonly used model for practical applications such as cam-follower contacts, engine piston rings, and gear contacts where the geometry and operating conditions are already complex.

Modeling of Gear Wear

Gear contact involves combined rolling and sliding action of two mating surfaces that have varying geometry and loads. Most recent gear wear models (Wu and Cheng 1993; Flodin and Andersson 1997, 2000; Bajpai et al. 2004) used Archard's wear model for simulating wear conditions of spur and helical gear pairs. With the assumption that the hardness values of the materials of contacting surfaces remain constant throughout the wear process, Archard's

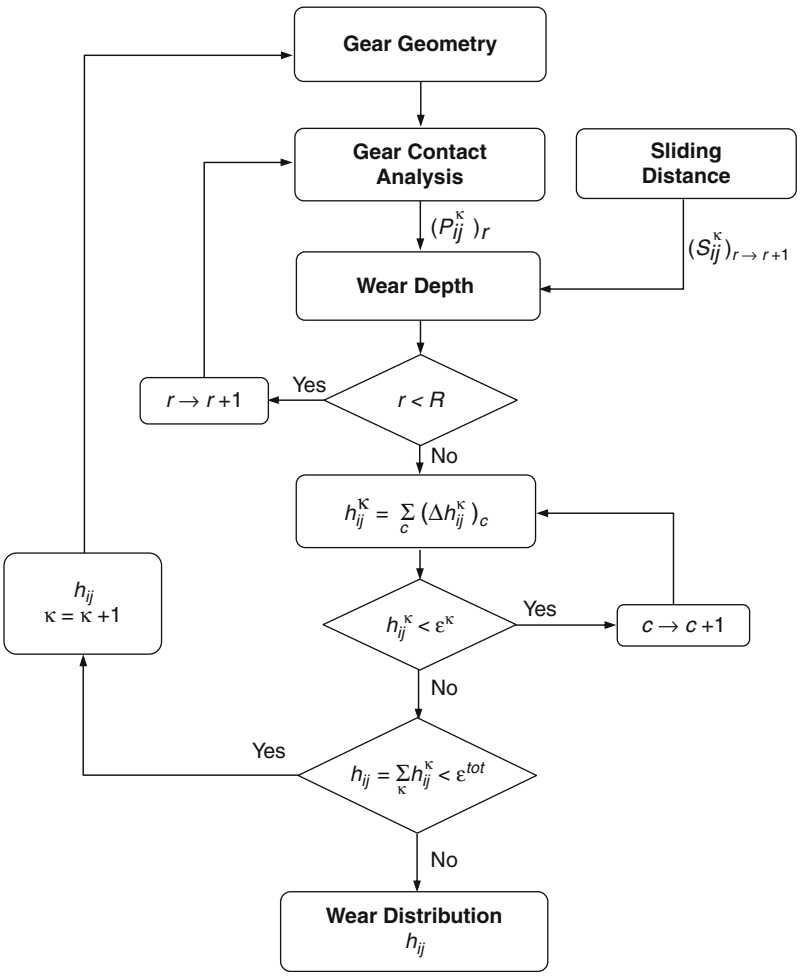
wear equation can be expressed for a local point on one of the contacting surfaces as

$$\frac{dh}{ds} = kp \tag{2}$$

where k is a dimensional wear coefficient. This equation is integrated over s to find the wear depth h of any point on a gear contact surface. Although a number of mathematical relations to compute wear coefficient k for metallic sliding friction have been proposed in the past, k has mostly been defined experimentally (Bajpai et al. 2004) for specific gear contact conditions.

Equation 2 states that prediction of h requires the values of the contact pressure p and the sliding distance s of each point on gear tooth contact surfaces as a function of gear rotation. Most gear wear models cited above employed a common iterative procedure outlined in Fig. 1 to predict the wear distribution of contacting gear surfaces. Here the initial geometric description of the actual gear tooth surfaces that includes manufacturing imperfections and intentional tooth modifications forms the initial state for the wear prediction. Geometries of each gear tooth contact surface are denoted by $(G_{ij}^{\kappa})^{p}$ and $(G_{ij}^{\kappa})^{g}$, representing the deviation of a surface point ij from the perfect involute surface at a wear increment κ for the driving (pinion p) and driven (gear g) gears, respectively. When $\kappa = 0$, the tooth surface is brand new (no wear). By specifying the deviations of points ij from perfect involute at the nodes of a predetermined surface grid, a discretized description of both contacting surfaces is obtained.

The next step in Fig. 1 is the prediction of the contact pressures of each contact point ij at every rotational increment r ($r = 0$ to R) of the gears in mesh. This can be done



Wear in Gears, Fig. 1 Flowchart illustrating a typical methodology to predict gear wear

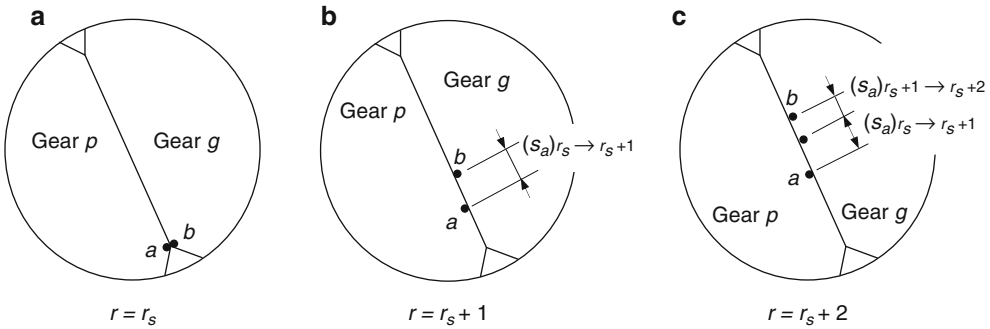
by using finite-element-based gear contact models (Bajpai et al. 2004) or simpler semi-analytical load distribution models (Flodin and Andersson 1997, 2000; Ding and Kahraman 2007). The gear contact model must be sophisticated enough to include initial tooth surface manufacturing deviations and modifications to predict the initial contact pressure distributions $P_{ij}^{\kappa=0}$ as well as the contact pressures P_{ij}^{κ} at various incremental stages of wear. Here, since the contact lines move across the tooth surfaces during the meshing process and width of contact zones is typically very narrow, a large number of rotational increments R must be considered to span an entire tooth loading cycle in order to obtain a smooth contact pressure distribution to be used by the wear formulation.

The sliding distance $\left(s_{ij}^{\kappa}\right)_{r\to r+1}$ experienced by a given contact point ij on a loaded tooth surface of one gear is defined as the distance by which the point ij on one gear slides with respect to its corresponding point on the mating gear as gears rotate from position r to position $r+1$ at the κ-th wear increment. Consider a point a on the active tooth surface of gear p that is at the leading edge of the instantaneous contact at a particular rotational position $r = r_s$ and the contact zone passes through the point a in following incremental rotations. The sliding distance calculation becomes a matter of tracking the relative position of this point with respect to its mating point b on the other gear. Focusing on the cross-section of a gear contact in the direction of sliding on the transverse plane of the gears, Fig. 2a shows the start of the wear cycle of this point a on gear p as well as its mating point b on gear g. At this rotational position r_s, both points a and b are at the leading edge of the contact zone, experiencing non-zero pressure for the first time since the beginning of loading cycle. Position vectors of points a and b at this position relative to a coordinate system located at the center of gear p are the same, i.e., $(\mathbf{X}_a)_{r_s} = (\mathbf{X}_b)_{r_s}$. When the gears are rotated by one incremental rotation to position to $r_s + 1$, point a rotates about the center of gear p by an incremental rotation $\Delta\theta_p$ of gear p, whose new coordinate $(\mathbf{X}_a)_{r_s+1}$ can be obtained via simple rotation transformation. As shown in Fig. 2b at position $r = r_s + 1$, point b on gear g is no longer in contact with point a on gear $p$as gear g rotates about its center by a $\Delta\theta_g = -(N_p/N_g)\Delta\theta_p$, where $N$is the number of teeth of gears. Position vector $(\mathbf{X}_b)_{r_s+1}$ of point $b$is obtained by first translating the coordinate frame from the center of gear p to the center of gear g, then rotating it by $\Delta\theta_g$, and then translating it back to center of gear p. With this the sliding distance experienced by point a between rotations r_s and $r_s + 1$ is

$$(s_a^{\kappa})_{r_s\to r_s+1} = \left\|(\mathbf{X}_b)_{r_s+1} - (\mathbf{X}_a)_{r_s+1}\right\|. \tag{3}$$

In order to generalize the above equation, consider the contact zone at the next position $r = r_s + 2$ as shown in Fig. 2c. In this position, the points of interest a and b move further away from each other. Relative sliding distance increment as gears rotate from position $r_s + 1$ to $r_s + 2$ is equal to the distance between a and b minus $(s_a^{\kappa})_{r_s\to r_s+1}$. If point a remains within the contact zone until position $r = r_e$, the sliding distance that occurs when gears rotate from any position r to $r + 1$ can be given in general terms as (Bajpai et al. 2004).

$$(s_a^{\kappa})_{r\to r+1} = \begin{cases} \left|\left\|(\mathbf{X}_b)_{r+1} - (\mathbf{X}_a)_{r+1}\right\| - \sum_{q=r_s}^{r}(s_a^{\kappa})_{q-1\to q}\right|, & r_s \le r \le r_e, \\ 0, & 0 \le r < r_s \text{ or } r_e < r < R. \end{cases} \tag{4}$$



Wear in Gears, Fig. 2 A graphical illustration of sliding distance experienced by point a on gear p during the first three rotational increments of its wear cycle

The sliding distance calculations must be carried out only for those nodes ij with nonzero $(P_{ij}^{\kappa})_r$ for at least two consecutive rotational positions. Sliding distance calculations for node ij of gear p are continued as r is increased until $(P_{ij}^{\kappa})_r$ becomes zero again. In calculating the sliding distance of point b on gear g with respect to point a on gear p, the same procedure is repeated by applying it now for gear g.

With $(P_{ij}^{\kappa})_r$ and $(s_{ij}^{\kappa})_{r\to r+1}$ in hand, the wear depth occurring at each tooth surface point ij during one wear complete cycle can be computed by discretizing the integrated form of (2) as

$$(\Delta h_{ij}^{\kappa})_c = \sum_{r=0}^{R-1} k(s_{ij}^{\kappa})_{r\to r+1}\left[\frac{(P_{ij}^{\kappa})_r + (P_{ij}^{\kappa})_{r+1}}{2}\right]. \quad (5)$$

Equation 5 is applied continuously C^{κ} times until the maximum wear depth accumulated at any node of either one of the contacting surfaces after the κ-th pressure *update* equals a preset threshold value of ε^{κ}. Then, the wear amount at point ij of gears p and g accumulated after the κ-th pressure update is

$$h_{ij}^{\kappa} = \sum_{c=1}^{C^{\kappa}} (\Delta h_{ij}^{\kappa})_c. \quad (6)$$

In Fig. 1, once the wear depth of any tooth surface point for each κ reaches ε^{κ}, the worn geometries of tooth surfaces are established and input to the gear contact model for an update of pressure distribution $(P_{ij}^{\kappa+1})_r$. This iterative procedure is repeated until the maximum total wear depth on either of the two gears reaches a certain threshold value of ε^{tot}. At this point the total wear depth of point ij is

$$h_{ij} = \sum_{\kappa=1}^{K} h_{ij}^{\kappa}. \quad (7)$$

Finally, the total number of wear cycles resulting in the accumulation of wear shown in (7) becomes

$$C^{tot} = \sum_{\kappa=1}^{K} C^{\kappa}. \quad (8)$$

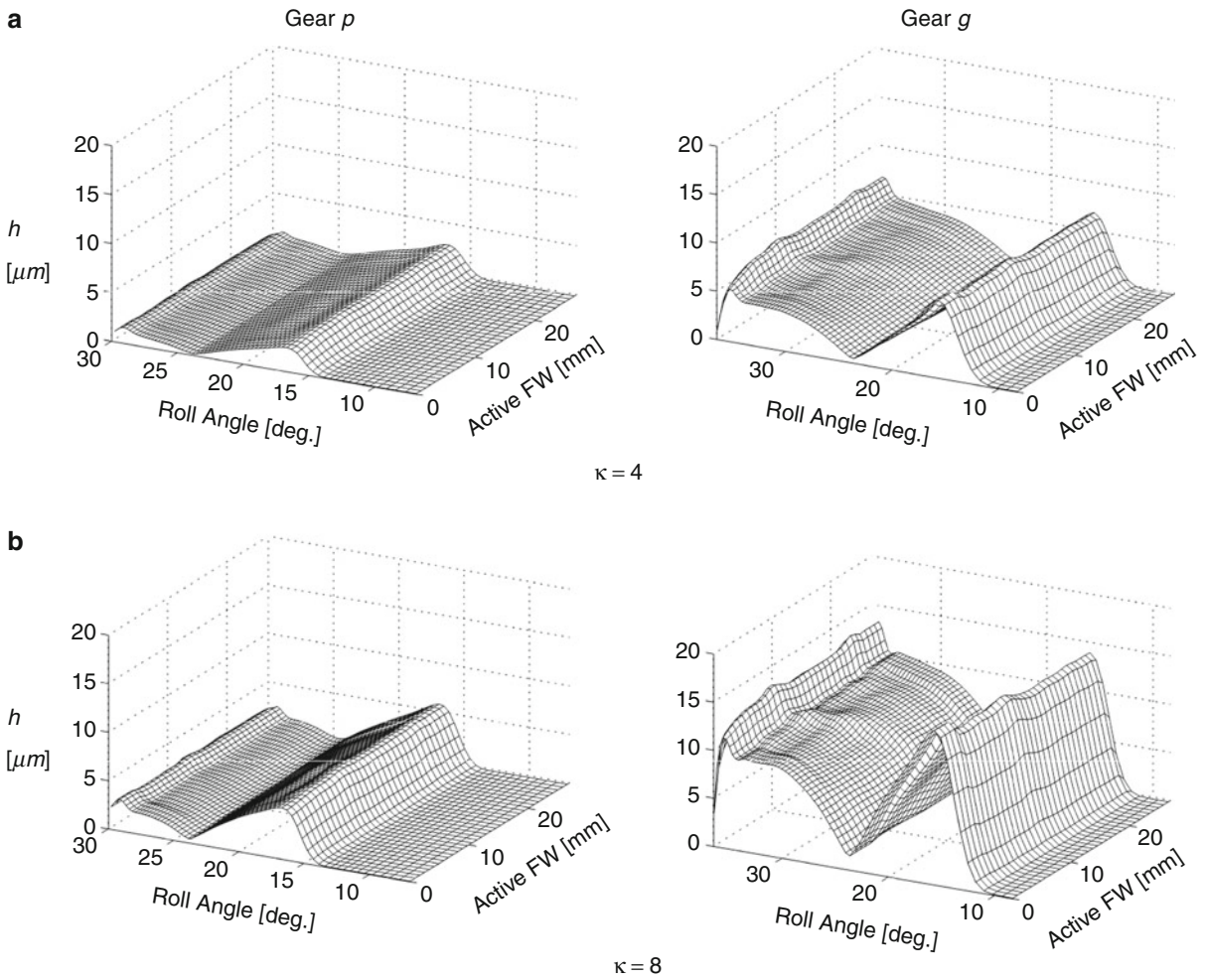
Figure 3 shows simulations of wear in a helical gear pair done by using a basic model similar to the one outlined above. The gear pair is formed by a 34-tooth gear p and a 22-tooth gear g having a normal module of 1.439 mm, active face width of 26.7 mm, and pressure and helix angles of 19° and 20°. These gears have no tooth modifications such that the initial geometries at $\kappa = 0$ are perfect involutes. In this figure, the wear depth distributions after the 4th ($\kappa = 4$) and 8th ($\kappa = 8$) wear geometry updates are shown for both gears p and g, with $\varepsilon^{\kappa} = 2$ μm. It is noted here that both gears experience significant wear in addendum and dedendum regions with no wear along the pitch lines (at about 23° roll angle) where relative sliding is zero. Wear is uniform along the face width direction since the initial surfaces were unmodified. It is also noted that gear g accumulates nearly 50% more wear simply because it has less teeth and hence each of its teeth experiences more cycles. Figure 4 shows the contact pressure distributions of the same gear set at its initial stage ($\kappa = 0$) as well as wear stages of $\kappa = 4$ and 8. Initially, contact pressure is relatively uniform across the tooth surface. Wear accumulation alters this, reducing contact pressures in the dedendum and addendum regions where wear is maximum while increasing pressures in the region near the pitch line. Figure 5 shows wear profiles of the same helical gear pair, now modified using a combination of involute crown, lead crown, and involute slope corrections (Fig. 5a), as it is done commonly in real-life gearing. Changes due to wear are evident in Fig. 5b, c at wear stages of $\kappa = 4$ and 8. Since the contacting surfaces are crowned in both directions, wear causes the surfaces to be flattened.

Such wear models were shown to compare well to the measured wear profiles from gear experiments, provided the wear coefficient is determined empirically. For instance, Bajpai et al. (2004) showed good correlations to a helical gear pair having shaved surfaces operated with a automatic transmission fluid with an empirical wear coefficient of $k = 9.65 \times 10^{-19}\ m^2/N$. Likewise, Ding and Kahraman (2007) obtained a good agreement with their spur gear experiments at different operating conditions with the same lubricant using $k = 2.9 \times 10^{-18}$ m^2/N.

Key Applications

Gear Wear and Gear Vibration Interactions

According to (2), the contact pressure distribution that is determined by force carried by the tooth is one of the main parameters impacting surface wear. The dynamic behavior of the gears amounts to tooth forces that are different from the quasi-static forces in both magnitude (often larger) and shape. These tooth forces also reflect various nonlinear phenomena such as backlash-induced tooth separations. Therefore, surface wear outcome is strongly related to the dynamic behavior of the gear pair during its operation (Wojnarowski and Onishchenko 2003; Ding and Kahraman 2007). In addition, a higher speed operation of the gear pair, while increasing dynamic tooth forces and the resultant contact pressures, also helps increase the film thicknesses at the lubricated tooth contacts due to



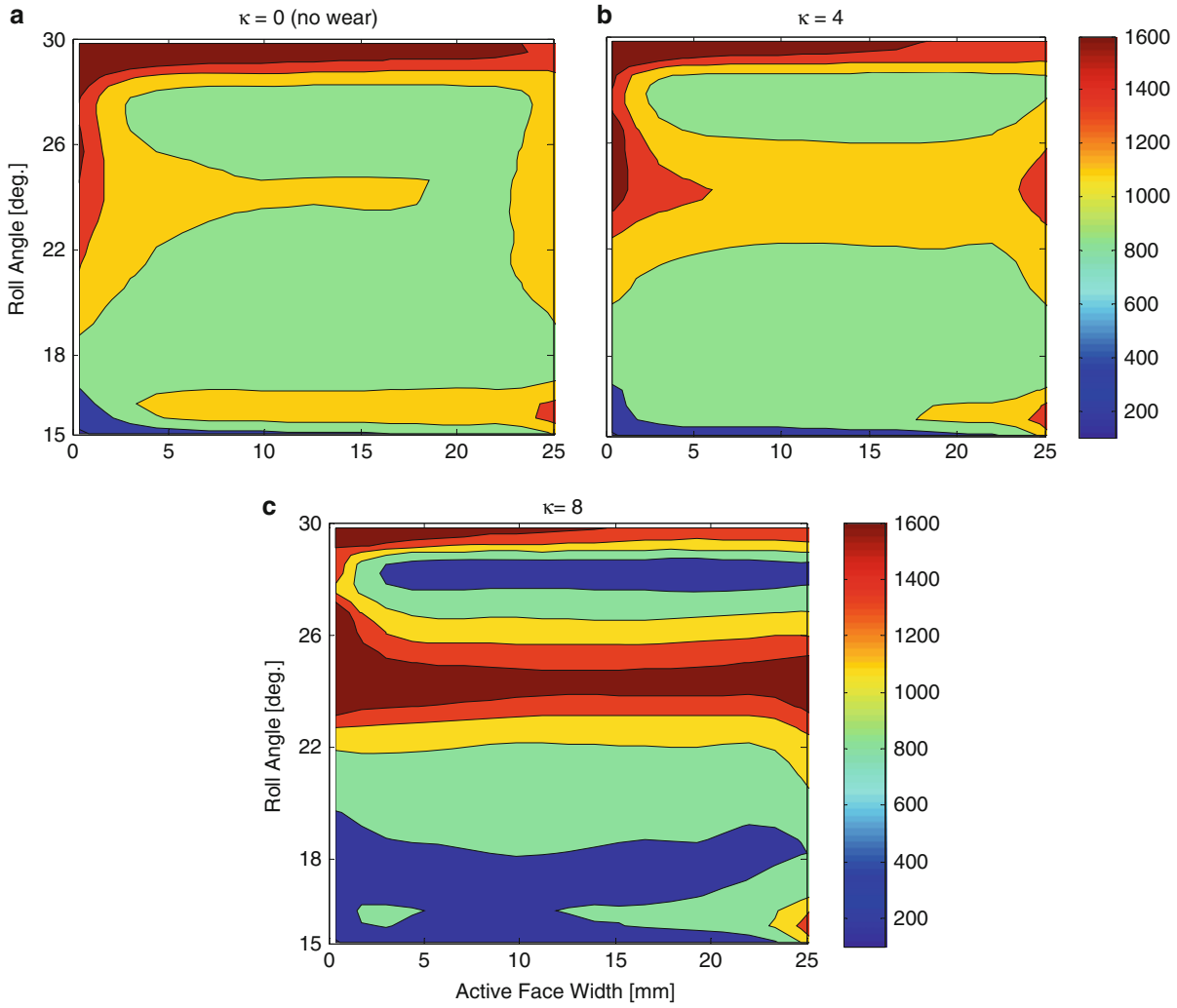
Wear in Gears, Fig. 3 Wear distributions of the gears of an unmodified helical pair after (**a**) 4th and (**b**) 8th wear updates $\varepsilon^{\kappa} = 2$ μm

increased rolling velocities. Priest and Taylor (2000) proposed an approximate wear coefficient model to account for the effect of film thickness on wear profiles, by modifying the wear coefficient k as

$$k = \begin{cases} k_0, & \lambda < \frac{1}{2}, \\ \frac{2}{7}k_0(4-\lambda), & \frac{1}{2} < \lambda < 4, \\ 0, & \lambda > 4. \end{cases} \quad (9)$$

Here, λ is the lambda ratio (the ratio of the minimum film thickness $h_{\min}$ to the composite surface roughness). This equation assumes that the wear coefficient is equal to k_0 for $\lambda < \frac{1}{2}$ that corresponds to the values observed at relatively low speeds such as the ones measured by Bajpai et al. (2007). It also assumes that $k = 0$ for $\lambda > 4$, i.e., no wear will occur if $h_{\min}$ is at least four times more than the composite surface roughness amplitude. In the transition region within $\frac{1}{2} < \lambda < 4$, k reduces linearly from k_0 to zero. This apparent reduction in k was reported to offset some of adverse effects due to dynamic forces while the shape of the wear profiles was shown to be dependent on the shape of the oscillations of the tooth forces (Ding and Kahraman 2007).

On the other hand, dynamic response of a gear pair is very sensitive to deviations of the tooth surface profiles from a perfect involute. Intentional tooth modifications such as the ones shown in Fig. 5a are commonly used to reduce the dynamic forces at a certain design torque. Unavoidable manufacturing errors also influence the dynamic response since they act as a "transmission error" excitation at the gear mesh interface. As surface wear is a material removal process that results in a deviation from the intended tooth profiles, it influences

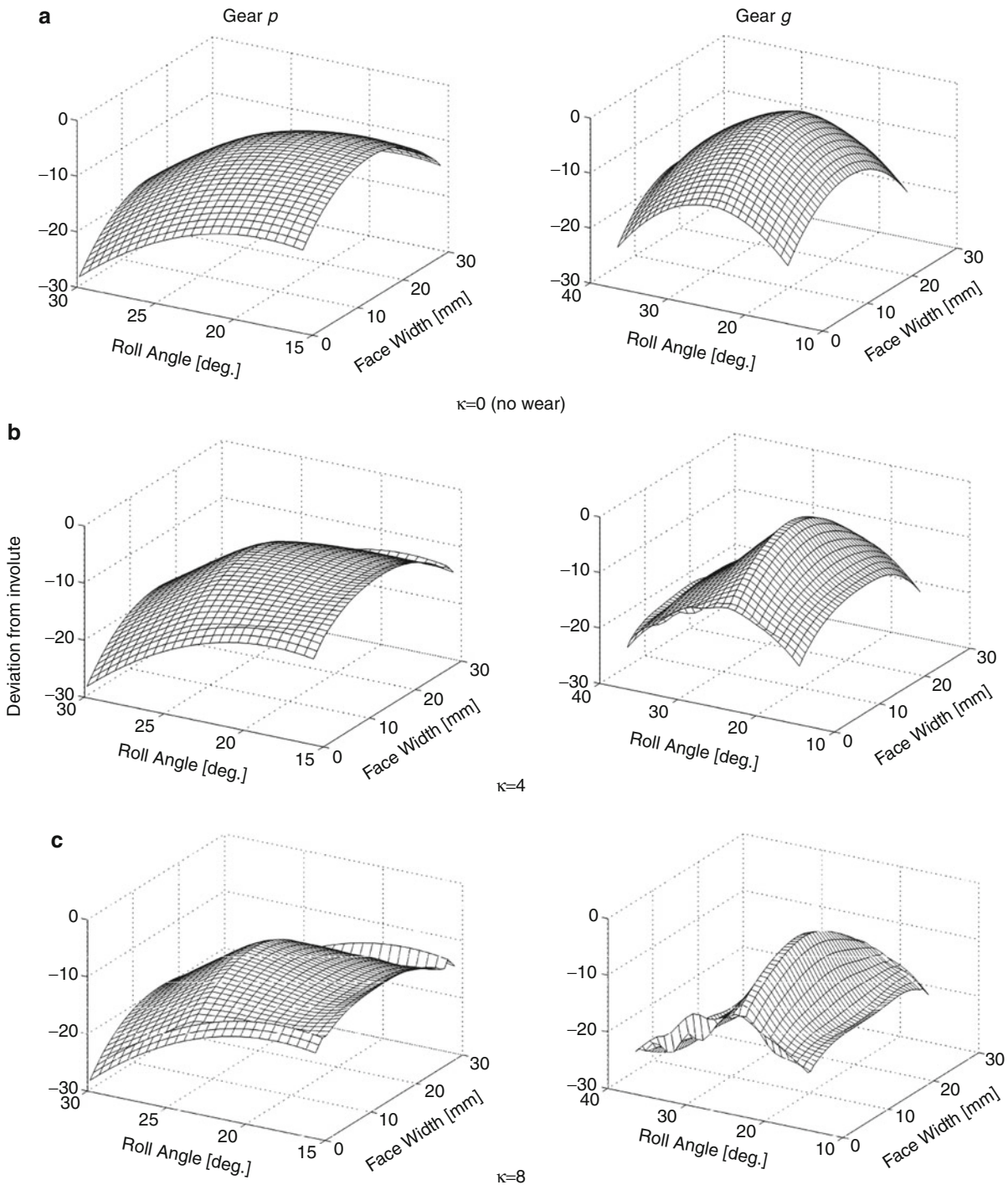


Wear in Gears, Fig. 4 Maximum tooth surface distributions of the gears of an unmodified helical pair. (**a**) Initial pressure distribution, (**b**) and (**c**) pressure distributions after the 4th and 8th wear updates. $\varepsilon^{\kappa} = 2$ μm

vibration amplitudes and the forced frequency response both quantitatively and qualitatively. For instance, when operated near a resonance peak, accumulating surface wear was shown to diminish this particular resonance peak (Ding and Kahraman 2007).

Influence of Tooth Modifications and Errors on Gear Wear

Tooth modifications are often determined based on their ability to neutralize tooth deflections at a certain load level. Yet these modifications also alter the contact pressure distributions to impact the wear characteristics of a gear pair. Kahraman et al. (2005) studied such effects on helical gears by parameterizing common automotive gear modifications as involute slope modifications (or errors) of gears p and g (A_p and A_g) and involute crown of gears p and g (B_p and B_g). They performed a number of simulations to determine that addendum and dedendum wear amplitudes are determined by the total involute slope mismatch $A = A_p - A_g$, and the total involute crown $B = B_p + B_g$. It was shown in this study that the maximum addendum wear rate increases exponentially with A such that $(\bar{h}_{\max})_a = a_1 e^{-b_1 A}$ while the maximum dedendum wear rate decreases exponentially with A according to $(\bar{h}_{\max})_d = a_2 e^{b_2 A}$. Here, the constant coefficients a_1, a_2, b_1 and b_2 are dependent on the gear tooth design parameters and the wear coefficient. Kahraman et al. (2005) further indicated that $(\bar{h}_{\max})_a = (\bar{h}_{\max})_d$



Wear in Gears, Fig. 5 Wear distributions of the gears of a modified helical pair. (**a**) Initial profiles showing tooth modifications, (**b**) and (**c**) tooth profiles after the 4th and 8th wear updates. $\varepsilon^{\kappa} = 2\ \mu\text{m}$

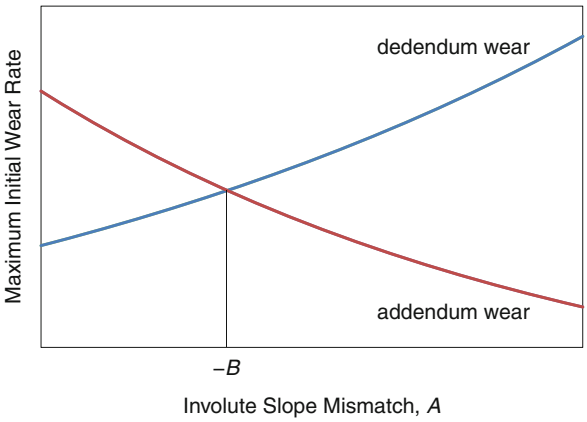
when $A = -B$. With this, influence of tooth profile modifications on maximum wear rate can be written in the form

$$\bar{h}_{\max} = \begin{cases} a_1 e^{-b_1 A}, & A < -B, \\ a_2 e^{b_2 A}, & A \geq -B. \end{cases} \quad (10)$$

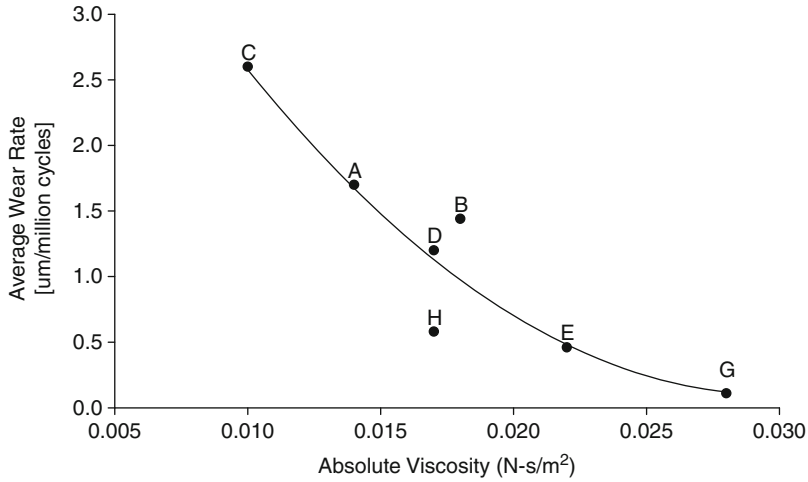
The same relationship is illustrated in Fig. 6 schematically.

Parameters Influencing Wear Coefficient

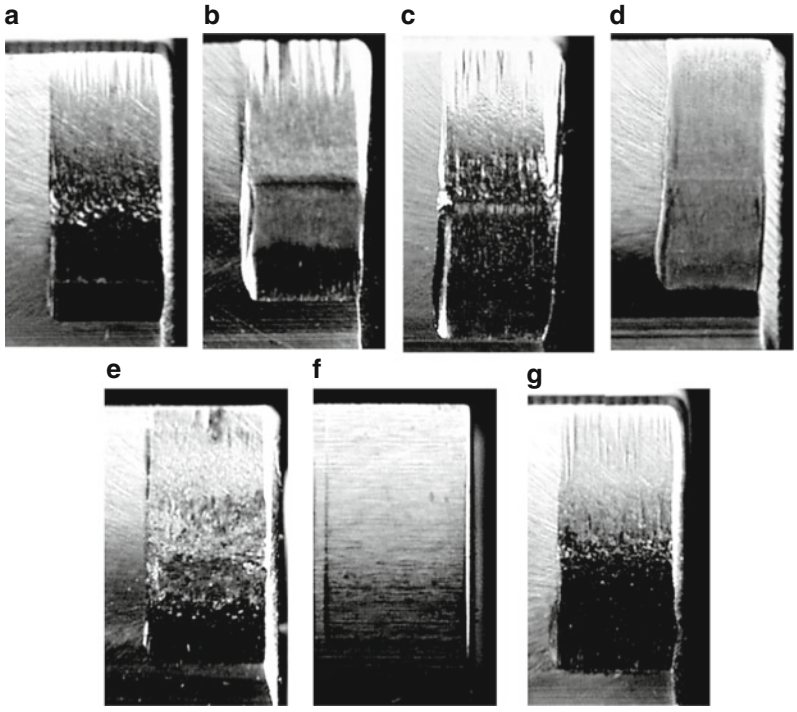
The conventional methodology outlined above for prediction of wear in gears is based on Archard's formula that requires a wear coefficient k that must be determined for the application in hand. This coefficient indeed must take into account all lubricant, surface, and material parameters as well as operating speed and temperature conditions. While this is the main shortcoming of this methodology, it is very practical given the added complexities of a gear contact analysis and sliding distance computations. In industrial applications where surface textures, lubricants, gear materials, and operating conditions are often similar, establishing k empirically from gear experiments or field data would be sufficient. Yet, any improvements to wear problem in gears beyond the ones attainable through design solutions based on these models must focus on how to reduce k. In one such study, Krantz and Kahraman (2004) focused on the influence of lubricant viscosity and additives on gear wear. In this study, seven different rotorcraft gearbox lubricants were used to perform families of high-speed spur gear tests. Figure 7 plots measured average gear wear rates as a function of viscosities of these lubricants to show that a higher viscosity fluid results in less wear due to increased fluid film thickness and reduced asperity contacts. In Fig. 8, images of the teeth of some of these gears are shown. It is clear in this figure that gear teeth tested with lubricant F having the highest viscosity has almost no wear while teeth tested with lubricant C having the lowest viscosity exhibit significant amounts of wear. Also note in Fig. 7 that three of the lubricants considered (B, D, and G) exhibit significantly different wear rates in spite of having very close viscosity values, emphasizing the critical role that lubricant wear additives play on wear in gears.



Wear in Gears, Fig. 6 A graphical representation of involute slope mismatch on initial maximum addendum and dedendum wear rate of a helical gear pair



Wear in Gears, Fig. 7 Variation of measured average wear rates as a function of viscosities of seven different lubricants at the testing temperature. The *solid line* is a trend line representing a quadratic curve fit (From Krantz and Kahraman 2004)



Wear in Gears, Fig. 8 Examples of worn spur gear tooth surfaces tested using the seven lubricants shown in Fig. 7 (From Krantz and Kahraman 2004)

Cross-References

- ► Friction/Traction Behavior of EHL
- ► Gear Lubrication
- ► Gear Sliding
- ► Gear Tooth Modifications
- ► Involute Gear Profiles
- ► Spur Gears

References

J.F. Archard, Contact of rubbing flat surfaces. J. Appl. Phys. **24**, 981 (1953)
P. Bajpai et al., A surface wear model for parallel axis gear pairs. J. Tribol. **126**, 597 (2004)
H. Ding, A. Kahraman, A dynamic wear model for spur gear pairs. J. Sound Vibr. **307**, 662 (2007)
A. Flodin, S. Andersson, Simulation of mild wear in spur gears. Wear **207**, 16 (1997)
A. Flodin, S. Andersson, Simulation of mild wear in helical gears. Wear **241**, 123 (2000)
A. Kahraman et al., Influence of tooth profile deviations on helical gear wear. J. Mech. Des. **127**, 656 (2005)
T. Krantz, A. Kahraman, An experimental investigation of the influence of the lubricant viscosity and additives on gear wear. Tribol. Trans. **47**, 138 (2004)
M. Priest, C.M. Taylor, Automobile engine tribology – approaching the surface. Wear **241**, 193 (2000)
J.A. Williams, Wear modeling: analytical, computational and mapping: a continuum mechanics approach. Wear **225**, 1 (1999)
J. Wojnarowski, V. Onishchenko, Tooth wear effects on spur gear dynamics. Mech. Mach. Theory **38**, 161 (2003)
S. Wu, H.S. Cheng, Sliding wear calculation in spur gears. J. Tribol. **115**, 493 (1993)

Wear in Knee Prostheses: Differences Compared with the Hip

Cassandra J. Wright-Walker, Martine LaBerge
Department of Bioengineering, Clemson University, Clemson, SC, USA

Synonyms

TKR – total knee replacement

Definition

A knee prosthesis, also known as a total knee replacement (TKR), is a medical device used to replace degenerated cartilaginous surfaces in the knee. It is typically comprised of a metal femoral component, polymeric tibial plateau,

and a metal tibial component; TKRs may also contain a patellar resurfacing component.

Scientific Fundamentals

History

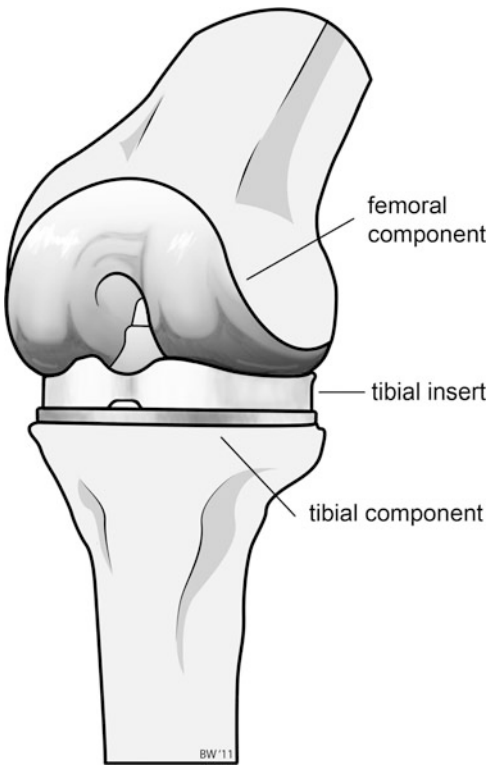
Annually worldwide, there is an increasing trend of patients receiving total joint replacement (TJR) procedures. Over half a million total knee replacements (TKRs) are performed each year in the United States, due in part to revision surgeries but also due to patients receiving TKRs at a younger age. Furthermore, surgeons have begun allowing patients in their 80s and older, even with co-morbid conditions, to receive TJR implants to increase quality of life, which increases the total number of implantations. As arthroplasty-appropriate age decreases, interest increases in implant longevity to reduce the number of potential revisions a patient might undergo.

The first modern TKRs, excluding Gluck's hinge arthroplasty of ivory created in the late 1800s, was the Walldius hinge made in 1958 of cobalt-chrome alloy. The first condylar TKRs, introduced by Gunston in 1971, were cobalt chrome condyles articulating against a UHMWPE tibial insert, without metal backing (Shanbhag et al. 2006). This design is similar to those that are used today. These early TKRs were cemented in place using polymethylmethacrylate (PMMA), which is comprised of PMMA pre-polymerized polymer bead powder that is combined with the liquid monomer of methyl methacrylate. It is injected into the surgical site after it achieves the proper paste consistency. Polymer polymerization is completed in situ between the implant and the prepared bone site, creating a stable support for the implant.

Advances in design to yield better patient outcomes as well as kinematics that more closely parallel the natural knee are still evolving. In the 1970s, these included metal backing for the polyethylene tibial tray and the use of non-cemented components. Although non-cemented implants allowed removal of less bone, many early non-cemented TKRs (up to one third) did not exhibit bone in-growth into the surface, leading to revision (Cook et al. 1988). The key to bone in-growth is lack of micromotion between the prosthesis and the bone that is found via precise surgical skill and patient compliance. Numerous studies of different textures to promote porous in-growth have included sintered beads and fibers, meshes, and hydroxyapatite coatings. While outcomes have been variable among non-cemented knee implants, these have been successful in younger patients with better quality bone stock. Advances in the 1980s included design of revision prostheses with longer stems for both cemented and non-cemented designs.

Materials

TKRs contain four components: femoral component, tibial component, tibial insert (or tibial tray), and patellar button (Fig. 1). Both the femoral and tibial components have a stem of metal that extends into the bone for implant stability. The femoral components are made of cobalt chrome molybdenum (Co-Cr-Mo) alloy due to its mechanical and tribological properties and its biocompatibility. The tibial component may be made of Co-Cr-Mo or titanium alloys such as Ti-6Al-4 V. Femoral and tibial components are either press-fit or cemented over the associated bone (femur or tibia) once it has been surgically prepared. Co-Cr-Mo is an excellent metal for these components: The chromium adds wear resistance as well as allows for a passivating oxidation layer of chromium oxide to form on the surface. Further, the cast Co-Cr-Mo (ASTM F75) allows fabrication of a porous surface for bone in-growth, which is necessary in non-cemented components. Porous coating of the Co-Cr-Mo results in



Wear in Knee Prostheses: Differences Compared with the Hip, Fig. 1 TKR components

decreased overall strength of approximately 200 MPa, which is significantly lower than non-porous Co-Cr-Mo alloy (Shanbhag et al. 2006). The tibial insert, made of ultra-high molecular weight polyethylene (UHMWPE), serves as the sacrificial bearing surface for the joint. UHMWPE has been used for TJR bearing surfaces since the 1950s after the rapid failure of polytetrafluroethylene (PTFE) as a sacrificial bearing surface. UHMWPE is either extruded then machined into its final form or direct molded using compression molding. Historically, sterilization was completed using gamma sterilization in air, but due to oxidative degradation this method was altered. Currently, gamma sterilization is completed in an inert (low oxygen) environment using 25–40 kGy. Moreover, experimentation with UHMWPE and gamma sterilization found that gamma radiation, which caused chain scission and cross-linking, could have a beneficial effect with respect to wear when high radiation doses (up to 100 kGy) were used: Wear resistance of UHMWPE has been shown to increase with cross-linking while decreasing the occurrence of pitting and delamination (Shanbhag et al. 2006).

The patellar button, implanted as needed, consists of a UHMWPE or metal-backed UHMWPE button that is placed against the surgically refinished patella. In spite of advances in design and materials, primary TKRs fail and need to be revised in approximately 10–20 years.

Wear Mechanisms

In general there are three main wear mechanisms observed in TJRs that lead to material removal from a bearing surface: adhesive wear, abrasive wear, and fatigue wear. Adhesion occurs when two parts bond under the exposed load; a portion of the material, one or both surfaces, is removed as relative motion between the two components occurs. Abrasion is primarily due to asperities on the surface of one or both of the materials. The asperities rub against the softer surface and ultimately lead to scratches and thus to material removal. Fatigue occurs when a material is repetitively loaded. The material fails under cyclic loads that may be significantly lower than the ultimate tensile strength or even the yield strength of the material because of the additive nature of the repetitive loading. Fatigue failure begins with an initiation site, often a crack or crevice in the material that can also be termed a local stress concentration. This crack begins to grow due to repetitive loading, which ultimately leads to failure of the material. Predictions of how the material will perform can be estimated from the S-N curve (a graphical representation of stress with respect to number of cycles in which a material will fatigue if the stress for a given number of cycles lies above the curve). However, wear typically is not attributed to one distinct mode, but a combination of one or more of the three types. Therefore, determination of how a joint replacement will fail is most accurate retrospectively, although educated analyses may provide predictions similar to what occurs in actuality. Abrasive and adhesive wear, both surface mechanisms leading to smaller average wear particle size, are predominantly observed in total hip replacements (THR). Unlike THRs, TKRs include all three major modes of wear (adhesive, abrasive, and fatigue). Due to both rolling of the femoral condyles on the tibial insert and sliding during cyclic loading, TKRs experience fatigue wear, which results in delamination and pitting. Fatigue wear generates larger average UHWMPE particle size for TKRs compared to THRs (see Table 1). Additionally, backside wear in which micromotion exists between the tibial insert and tibial component may occur and result in a second wear surface. This backside wear is most prevalent in mobile bearing TKRs; however, it may be observed in both TKRs and THRs between the UHMWPE components and their metal mating component.

As a tribological phenomenon, wear results from contact geometry (conforming or non-conforming surfaces), kinematic loading conditions (velocity, applied load, load distribution), lubrication (fluid viscosity, friction, molecular adsorption), and material properties (hardness, elastic and viscoelastic properties, surface properties). It has been shown using a knee joint simulator that as conformity increases in TKRs, both the wear rate and the size of the wear scar increase (Galvin et al. 2009). Conversely, the average peak contact pressure increases as conformity decreases (Galvin et al. 2009). However, recent clinical retrieval evaluations of coronally flat implants have shown wear scars covering approximately half of the medial and lateral condylar regions (48 ± 16 % and 47 ± 14 %, respectively). Further, it was shown that delamination occurred in only a small percentage (11 %) of cases (Harman et al. 2007).

McKellop and colleagues further described wear for total joint replacements based on specific tribological

Wear in Knee Prostheses: Differences Compared with the Hip, Table 1 Average particle size for THR and TKRs

Average particle size	
Hip	0.43 μm (Schmalzried et al. 1997), 0.53 μm (Shanbhag et al. 1994), 0.694 μm (Mabrey et al. 2002)
Knee	0.52 μm (Schmalzried et al. 1997), 1.190 μm (Mabrey et al. 2002)

contact conditions observed in those bearings (McKellop et al. 2008). Each of the following modes of wear may occur in THRs and TKRs. In Mode 1, which occurs in all TJRs, only intended bearing surfaces are in contact and undergoing wear. In Mode 2, a bearing surface is moving against a nonbearing surface, as in the case of a femoral component articulating on the rim of the tibial insert instead of the intended area. Mode 3 is third-body abrasive wear in which the two primary bearing surfaces are articulating against one another but with abrasive particulate in between the two surfaces. In this case, abrasion results from PMMA, bone particles, or metal particulate entrapped in the UHMWPE tibial insert increasing the wear occurring. Finally, Mode 4 wear occurs when two surfaces that are not designed as primary bearing surfaces are moving against each other such as backside wear between the tibial insert and the tibial component in a fixed bearing TKR. Mode 1 type wear particles will be produced continuously throughout the service life of the implant. However, sometimes a transfer film can occur in which the soft PE is transferred onto the metal surface and alters the wear rate of the implant. If a stable transfer film forms, there may be a reduction of wear over the long term after an initial "wear-in" period shortly after implantation. If the film is not stable, however, film formation may increase wear via abrasive or third-body (Mode 3) type wear (McKellop et al. 2008).

Failure of TKRs

There are many factors that influence the failure rate of primary knee replacements. Patient age and diagnoses at time of implantation significantly influence the service life of an implant (revision is considered the endpoint for the purposes of this review). As patients age, they tend towards lower activity levels. This causes decreased wear on artificial joints and typically lengthens implant service life. A retrospective study of over 11,606 primary knee replacements confirmed that survivorship at 10 years decreases with patient age at implantation (Rand et al. 2003). Further, the survivorship data also showed that overall implant percent survivorship decreases as would be expected at 5, 10, 15, and 20 years. When survivorship is separated by gender, women have a significantly higher implant survivorship at 10 years than men (93% vs. 88%, respectively) (Rand et al. 2003). Rand and colleagues also divided the cause of primary arthroplasty by disease or condition that resulted in significantly higher survivorship of rheumatoid arthritis patients (95%) at 10 years compared with other diagnoses: osteoarthritis (90%), post-traumatic arthritis (86%), osteonecrosis (84%), and other diagnosis (76%). It has been suggested that patients with rheumatoid arthritis would have decreased physical activity compared with osteoarthritic patients and this would result in increased percent survivorship.

Implant design also affects implant service life. It has been shown that posterior stabilized designs had a significantly higher revision rate than posterior cruciate-retaining designs, in which at 10 years after surgery survivorship was 76% (n = 2,994) and 91% (n = 8,052), respectively (Rand et al. 2003). More recent smaller studies have shown higher implant survivorship of both designs. Additionally, lower survivorship may be due to early implant design, such as impingement of UHMWPE, which leads to increased wear, resulting in early failure, as the study by Rand and colleagues contains cases starting in 1978.

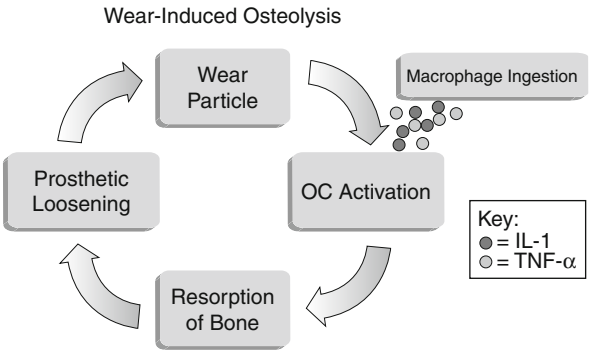
Other non-design-related causes of failure include infection, stress-shielding, periprosthetic osteonecrosis or fracture, implant instability, polyethylene wear-through or fracture, and allergy or hypersensitivity to one of the implant components (Shanbhag et al. 2006). Implant instability causes increased wear as the bearing surfaces will not always retain proper alignment. Additionally, polyethylene wear-through or fracture is a result of Modes 2–4 type wear; the thickness of the tibial insert makes it highly unlikely that the insert will wear through during its service life if Mode 1 wear is the only type wear present.

The most common cause of failure of an implant is implant loosening that results from wear and osteolysis (Shanbhag et al. 2006). Failure rates vary drastically according to design, cemented, hybrid, or non-cemented, implantation time, as well as many other factors.

Repercussions of Wear Particle Generation

Even with advances in factors affecting wear, wear particles (Mode 1 wear) are unavoidable. The particles generated are typically polyethylene, but may also be metal debris or corrosion products; they are released into the synovial fluid, from which they travel into the periprosthetic tissue. This elicits the body's natural defense mechanism to remove foreign material: Macrophages respond by phagocytosing the debris. Because they cannot break down the particles, secreted cytokines cause osteoclast maturation and activation. Once osteoclasts mature, they function to resorb bone in the regions surrounding the prosthesis resulting in loosening (Chiba et al. 1994). Implant loosening causes a destructive positive feedback (see Fig. 2) known as osteolysis or periprosthetic bone loss (Schmalzried et al. 1997).

The average UHMWPE particle sizes generated for THRs and TKRs have previously been shown in Table 1.



Wear in Knee Prostheses: Differences Compared with the Hip, Fig. 2 Wear cycle in bone

The average particle size ranges from approximately a half micron in THRs to just over a micron in TKRs. TKRs tend to produce larger particles due to oxidative embrittlement and delamination. The particle size is important as it determines the biological response. In a murine model, particles from 0.24 to 1.71 μm resulted in activated macrophages and thus bone resorption. Larger particles (7.6–88 μm), such as large delamination-generated particles, did not activate macrophages.

Osteoclast precursors do not resorb bone until they are signaled to become active, a key step in osteolysis. The transcription factor nuclear factor kappa B (NF-κB) must be stimulated to activate the NF-κB pathway, which results in differentiation and maturation of the precursor cells. The two main components of NF-κB pathway are receptor activators of NF-κB: RANK (a receptor found on osteoclast progenitors) and RANK ligand (RANK-L, a molecule secreted by stromal cells that binds and activates the NF-κ B pathway). Osteoprotegerin (OPG), a natural down-regulation mechanism of this pathway, is a protein secreted by osteoblasts that competitively binds to RANK, thus inhibiting osteoclast maturation. TNF-α and IL-1, two cytokines released by macrophages in response to wear particles, are particularly of interest because they up-regulate RANK-L secretion (Jacobs et al. 2001; Wei et al. 2005).

Knee Designs

Ultimately, the goal of TKR design is to reduce wear such that the service life of the implant may be lengthened. However, wear is highly complex. It deals with the interplay of variables including wear resistance, loading conditions, lubrication, surface roughness, implant design (conformity), surgical technique, and so on. Furthermore, all of these will vary during the life of the implant according to accumulated wear at each surface. Recently, modifications to TKR designs have been proposed to enhance the lifetime of the bearing.

Modular Versus Non-modular

The tibial component and insert may be modular or non-modular. Modular implants are systems that allow for several different tibial inserts of varying thickness. This allows a more customized fit for the patient, which may be necessary to provide the best surgical outcome. Excellent results have been seen with both types of implants, and orthopedic manufacturers currently produce both types.

Posterior Stabilized (PS)

A posterior stabilized knee design sacrifices the posterior cruciate ligament (PCL), which functions to restrain the posterior motion of the femur with respect to the tibial component. To correct for the missing PCL, the tibial tray has a raised post in the center that fits within a groove in the femoral component to prevent undesired femoral kinematics.

Cruciate Retaining

Cruciate retaining designs do not have the raised UHMWPE post and groove that posterior-stabilized implants contain. These designs allow for the patient to keep their PCL. However, the patient must have a healthy PCL for this design choice.

"Gender Specific"

One of the recent design alternatives for TKRs are gender-specific knees that are aimed at more closely mimicking the female shape and size. Clinical outcomes have not shown significant differences between standard and gender-specific implants (Young-Hoo et al. 2010) although longer follow-up studies and possibly laboratory simulations must be conducted at 10–15 years post-operatively to determine comparative longer-term wear resistance.

Cemented, Non-cemented, and Hybrid

TKRs may have both components cemented using polymethylmethacrylate (PMMA), or young patients or those with clinically acceptable bone stock may have neither component cemented. Use of non-cemented components requires bone in-growth to achieve adequate fixation. If in-growth does not occur, micromotion will be prevalent that will lead to increased wear and early device failure. Hybrid fixation, in which the femoral component relies on bone in-growth (non-cemented) and the tibial component is cemented, is also an option. A statistically increased survivorship of cemented knees

(92%, n = 11,166) has been shown compared with non-cemented (61%, n = 259) and hybrid (84%, n = 172) at 10 years (Rand et al. 2003). However, the sample size of patients was relatively small with cementless or hybrid fixation designs implanted. The reported survival rate of press-fit (non-cemented) TKR varies by study. Press-fit prostheses require healthy bone stock and are typically implanted in younger thus typically more active patients, which may explain the decrease in survivorship. Further, significant surgical skill is required for the precise implantation techniques used in cementless fixation.

Symmetric Versus Asymmetric Design

Early TKR designs involved symmetric design, in which both condyles are uniform in shape and size; these have shown excellent long term survivorship. Asymmetric knee implants were designed to more closely resemble human knee anatomy in order to increase patellofemoral function by decreasing the occurrence of patella subluxation and instability.

Unicompartmental Knee Replacements

After development of the TKRs, unicompartmental knee replacements (UKRs) were developed using the same materials and design of the total knee at that time except that only the medial or lateral section of the knee was replaced. In 1972, the Sledge prosthesis developed in Germany was one of the earliest reported UKRs with a Co-Cr alloy femoral component bearing on a non-metal backed UHMWPE tibial component. These were designed as a treatment for unilateral osteoarthritis or trauma to only the medial or lateral condyle/meniscus as an alternative to TKRs. UKRs are most often implanted on the medial side, though they may be used for the lateral compartment as well. The disadvantage of UKRs is that they require increased surgical skill for alignment and thus are not used by all orthopedic surgeons. Misalignment may lead to non-ideal kinematic and tribological conditions. Moreover, UKRs may be considered a bridge to a TKR. UKRs often have a predominately flat polyethylene component, which leads to an initially small contact area that will increase over time as the component "wears in." This smaller contact area leads to higher contact stresses that in turn may lead to early fatigue failure of the component.

Fixed Bearing Versus Mobile-Bearing Knees

The majority of TKRs are fixed bearing, in which the metallic tibial tray fits securely into the UHMWPE tibial component. The femoral component then articulates against this fixed UHMWPE surface. However, for younger or active patients, some surgeons may recommend mobile bearing knees. Mobile bearing knees have a metal post that extends upward from the tibial component on which the tibial tray articulates. The tibial tray has a slot in it that does not penetrate through the thickness of the PE to the bearing surface, allowing it to rotate and/or slide short distances with respect to the tibial component. Mobile bearing knees may be divided into three types: anterior-posterior translation only, rotation only, or a combination of translation and rotation (Callahan et al. 2001). The concept of mobile bearings allows more natural joint movement for the patient but also generates a second bearing surface and potentially increases back-side wear. One of the largest advantages of this design type is the transfer of stresses to the soft tissue instead of solely into the surface and subsurface regions of the constrained polyethylene tibial tray and component (Callahan et al. 2001). The drawback of this type implant is that there is additional support required from the soft tissue structures including ligaments to keep the knee aligned and prevent dislocation. Research has not shown yet that these implants are an improvement over the standard fixed bearing knees in performance or longevity.

Revision Components

If the TKR fails during the patient's lifetime, the patient must undergo revision surgery to replace the failed device. Revision TKR components have longer stems than their primary surgery counterparts to allow for contact between the stem and the bone to increase stability; this is because osteolysis is likely to have removed some of the bone surrounding the original implant. Further, revision components may also contain a cam/post system (similar to that described above in the posterior stabilized section) to provide additional stability and less reliance on the patient's ligaments. The major difference between primary and revision components is the length of the stem.

Key Applications

Total knee replacements and total hip replacements.

Cross-References

- ▶ Fatigue
- ▶ Introduction to Biotribology
- ▶ Polymers in Biotribology
- ▶ Testing of Artificial Hip Joints
- ▶ Testing of Artificial Knee Joints

References

J.J. Callahan et al., Mobile-bearing knee replacement: concepts and results. AAOS Instr. Cours. Lec. **50**, 431–442 (2001)

J. Chiba, H.E. Rubash, K.J. Kim, Y. Iwaki, The characterization of cytokines in the interface tissue obtained from failed cementless total hip arthroplasty with and without femoral osteolysis. Clin. Orthop. Relat. Res. **300**, 304–312 (1994)
S.D. Cook, K.A. Thomas, R.J. Haddad Jr., Histologic analysis of retrieved human porous-coated total joint components. Clin. Orthop. Relat. Res. **234**, 90–101 (1988)
A.L. Galvin et al., Effect of conformity and contact stress on wear in fixed-bearing total knee prostheses. J. Biomech. **42.12**, 1898–1902 (2009)
M.K. Harman, S.A. Banks, W.A. Hodge, Backside damage corresponding to articular damage in retrieved tibial polyethylene inserts. Clin. Orthop. Relat. Res. **458**, 137–144 (2007)
J.J. Jacobs et al., Osteolysis: basic science. Clin. Orthop. Relat. Res. **393**, 71–77 (2001)
J.D. Mabrey, A. Afar-Keshmiri, G.A. Engh, C.J. Sychterz, M.A. Wirth, C.A. Rockwood, C.M. Agrawal, Standardized analysis of UHMWPE wear particles from failed total joint arthroplasties. J. Biomed. Mater. Res. B Appl. Biomater. **63**, 475–483 (2002)
J.A. Rand et al., Factors affecting the durability of primary total knee prostheses. J. Bone Joint Surg. **85-A.2**, 259–265 (2003)
T.P. Schmalzried et al., Shape and dimensional characteristics of polyethylene wear particles generated in vivo by total knee replacements compared to total hip replacements. J. Biomed. Mater. Res. B Appl. Biomater. **38**, 203–210 (1997)
A.S. Shanbhag et al., Composition and morphology of wear debris in failed uncemented total hip replacement. J. Bone Joint Surg. **76.1**, 60–67 (1994)
A.S. Shanbhag, H.E. Rubash, J.J. Jacobs, *Joint Replacement and Bone Resorption* (Taylor & Francis Group, New York, 2006)
S. Wei et al., IL-1 mediates TNF-induced osteoclastogenesis. J. Clin. Invest. **115**(2), 282–290 (2005)
K. Young-Hoo, C. Yoowang, K. Jun-Shik, Comparison of a standard and a gender-specific posterior cruciate-substituting high-flexion knee prosthesis: a prospective, randomized, short-term outcome study. J. Bone Joint Surg. **92**, 1911–1920 (2010)

Wear Induced Change of Topography

► Surface Variation in Tribological Processes

Wear Maps

Seh Chun Lim
Department of Mechanical Engineering, National University of Singapore, Singapore, Republic of Singapore

Synonyms

Wear-mechanism maps; Wear-mode maps; Wear-transition maps

Definition

A wear map is a two-dimensional or three-dimensional, user-friendly representation of wear data. With carefully selected parameters as axes for the two- or three-dimensional space, such a map provides a global (or bird's eye) view of the wear behavior of a pair of materials in tribological interactions. The relationship between wear behavior and the condition of interactions can be appreciated at a glance. A wear map can serve as a quick, albeit approximate, guide to a designer in the first-cut selection of suitable operating conditions for components intended for tribological interactions in a larger (engineering) system. A desirable operating condition would be one that promotes a mild form of wear, so that the components interacting with each other will have a sufficiently long service life before the amount of wear sustained would require them to be replaced.

Scientific Fundamentals

Why is There a Need for Wear maps?

The removal of material as a result of wear during the interaction of surfaces in relative motion is a complex phenomenon. While much effort had been made to gain a better understanding of the underlying processes and mechanisms leading to the removal of material during such interactions, precisely how the various parameters and factors interact with each other and contribute to wear (to different degrees) is still not fully understood. However, renewed effort coupled with increasingly capable measurement and observation techniques should help researchers to gain a deeper insight into such issues.

Over time, a large amount of wear data has been amassed and reported in the literature for a variety of material combinations and experimental conditions. However, the information gathered has invariably been "localized," meaning that the data were obtained (experimentally) for a relatively narrow band of sliding/interaction conditions. While these data provide fairly detailed views of the "localized" behavior, it was often difficult to relate the wear data obtained under one experimental condition with those obtained from another, even though the same materials might have been used.

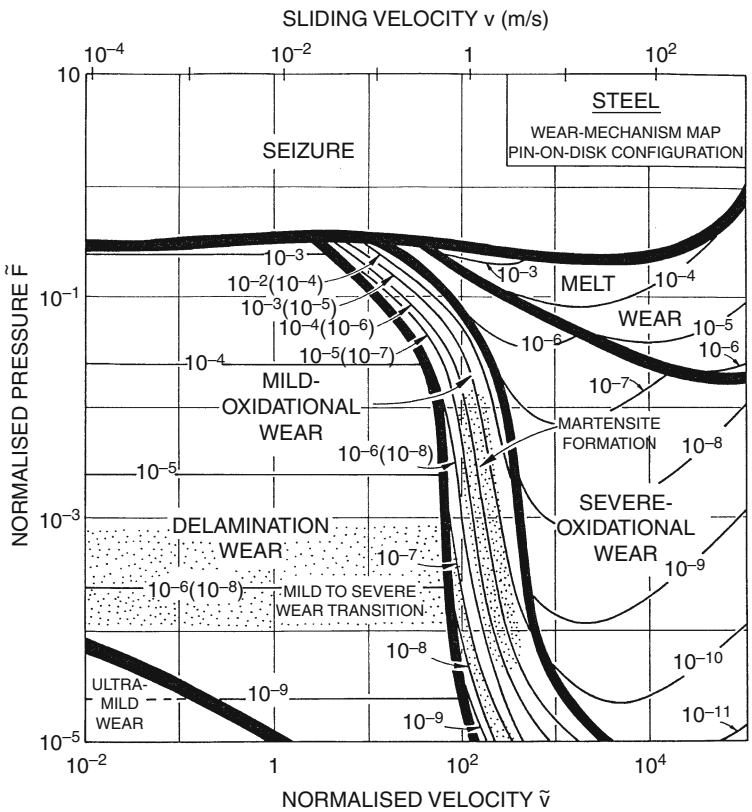
The lack of connectivity among experimental data obtained under different interaction conditions has caused some confusion, especially to those who are new to the field. The notion that there is one wear factor (or coefficient) for a pair of materials further adds to this confusion. The fact remains that wear behavior (both wear rate and wear mechanism) will change with changing interaction conditions.

A properly constructed wear map offers a global (or bird's eye) view of the wear behavior – which usually includes both wear rate and wear mechanism information – of a pair of materials in tribological interactions over a considerable range of contact conditions. It will enable users to appreciate the connectivity between available wear data and to navigate a sea of changing wear behavior.

How Did Wear Maps Evolve?

The idea of presenting wear data in a two- or three-dimensional graphical manner is not new. One of the earliest is the wear-rate surface constructed for the sliding of steel components (Okoshi and Sakai 1941). While a number of such representations have been proposed since then, it was Childs' map for steels (Childs 1980) that took this form of data representation to the next level. Childs used a range of sliding conditions that was far wider than what anyone else had done before. This would be the first time a global wear behavior of soft steels was presented, showing how steels would switch from one mode of wear to another with changing sliding conditions. Changes in wear mode were accompanied by corresponding changes in the mechanisms of wear.

Later, Tabor (1983) remarked that wear might be the result of interacting mechanisms with no single process dominating. He suggested that wear-mechanism maps – summarizing data and models for wear, showing how the wear mechanisms interface, and allowing the dominant wear-mechanism for any given set of conditions to be identified – could be developed to explore this much broader pattern of wear behavior. (Dominant wear-mechanism is that wear mechanism which contributes to the largest portion of the total wear-rate measured during tribological interactions.) Following this line of thought, the wear-mechanism map for steels in unlubricated sliding was constructed based on experimental data (primarily from pin-on-disk tests) and theoretical models culled from a wide range of sources (Lim and Ashby 1987). This map (Fig. 1) describes the unlubricated pin-on-disk wear behavior of steels over a wide range of sliding conditions. It predicts the field of dominance of one wear mechanism and when its contribution will become

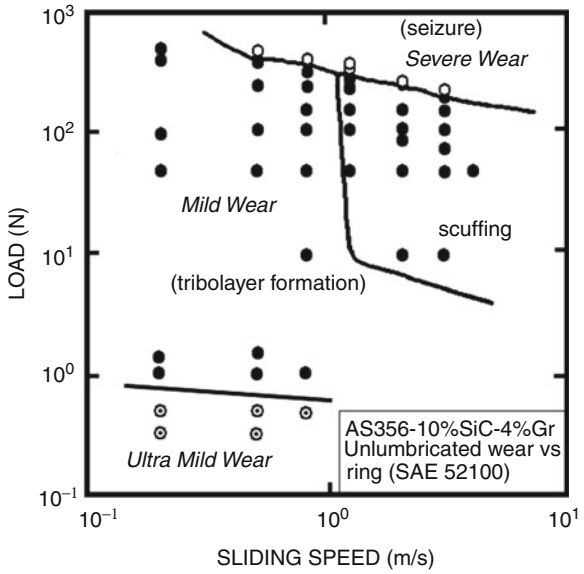


Wear Maps, Fig. 1 A wear map for the unlubricated sliding of steels (Lim and Ashby 1987)

less important. Within each field, contours of predicted normalized wear-rates are superimposed. Companion wear-mode and wear-transition maps for steels were presented to provide additional information that the first map could not conveniently present (Lim et al. 1987). These latter two maps summarize the sliding conditions associated with mild wear and severe wear and where transitions between them occur for steels. The wear-transition map also shows how the various wear transitions reported in the literature could be related and harmonized with one another.

Sometimes, it may be useful to construct a companion friction map to complement an existing wear map. It has been shown that when the sliding becomes more severe, the measured coefficient of friction for a pair of steel components in unlubricated sliding depends on the sliding condition much more than on the surface properties (such as surface roughness), which were found to be more important during slower sliding (Lim et al. 1989).

Wear maps for other groups of materials, ranging from metals, ceramics, and polymers to engineered materials, such as metal-matrix composites, have been proposed. Figure 2 shows the wear map for graphitic aluminum metal-matrix composite sliding against steel in unlubricated condition (Riahi and Alpas 2001). In this example, details of both wear rate and wear mechanism are provided, with the fields of dominance of each wear mechanism clearly demarcated.



Wear Maps, Fig. 2 A wear map for the unlubricated sliding of A356 Al-10%SiC-4%Gr metal-matrix composite against 52100 steel. Three major wear modes, ultra mild wear, mild wear, and severe wear (seizure), are shown together with the sub-regime of scuffing. The associated wear rates next to the data points (as shown in the original diagram) are not reproduced presently (After Riahi and Alpas 2001)

Wear maps for different contact conditions have also been proposed. These include, for example, maps for fretting wear, maps for corrosive wear, maps for erosion and wear maps for tools used in different machining operations. For the last category, a number of maps describing the flank wear and crater wear characteristics of coated and uncoated tools and tool inserts during single-point dry turning operations have been constructed. These maps typically show a region of machining conditions within which tools would experience the lowest rate of wear – the *safety zone*. The effect of the application of a thin layer of coating onto such tools can also be clearly seen; this is the expansion of the range of machining conditions defining each region having comparable rate of tool wear. Figure 3 shows the superposition of the flank wear maps of uncoated carbide tools and TiC-coated carbide inserts used in single-point dry turning operations. The expansion of the safety zone and the least-wear regime due to the presence of the TiC coating can be clearly seen (Lim and Lim 2001).

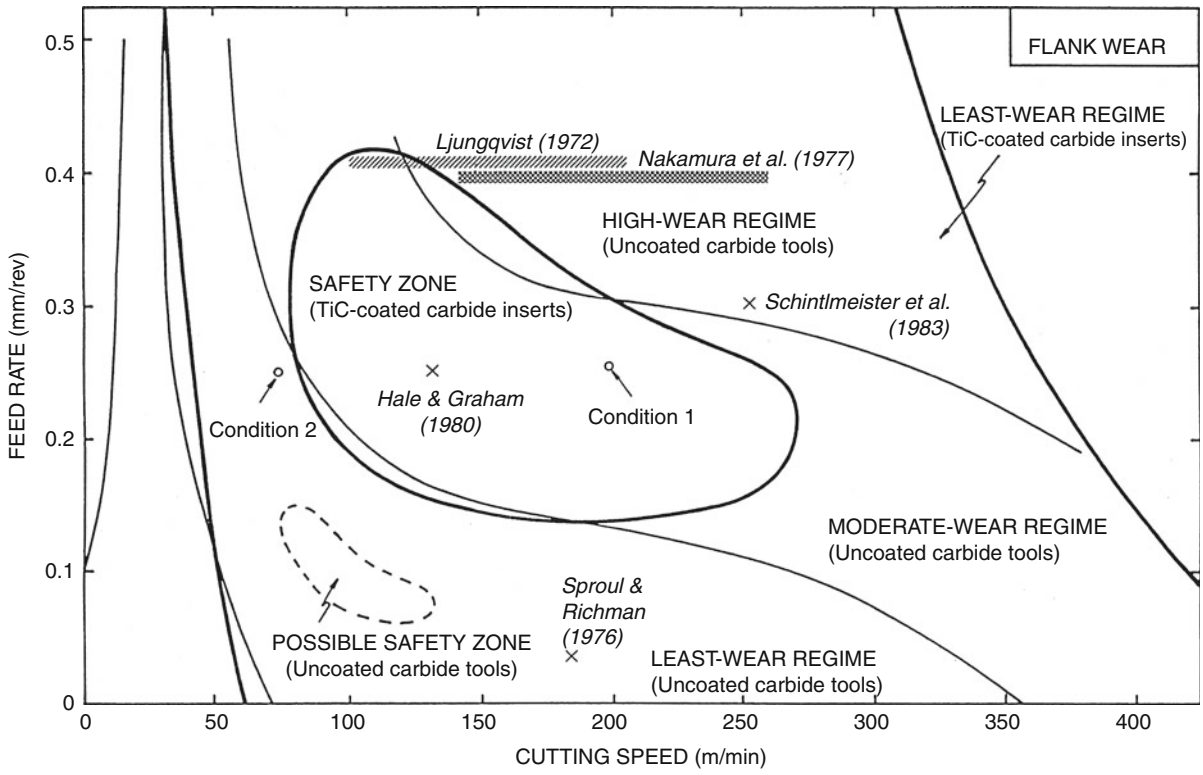
A summary of the early developments of various types of wear maps is given in Lim (1998).

How to Construct a Wear Map?

Much effort is needed to construct a wear map. It invariably involves the collection of a considerable amount of experimental data, both wear rate and wear mechanism, of a particular pair of materials under very similar external environments. These external environments would include lubricated or unlubricated condition (one cannot mix data obtained from lubricated tests with those obtained from unlubricated tests); atmospheric or lower-pressure condition; ambient-temperature or elevated-temperature condition; and clean or "gritty" surrounding, which would suggest the possible presence of abrasive wear.

The key characteristic of a useful wear map is the wide range of operating conditions covered. The operating conditions (which usually would be chosen as the axes of the map) should preferably be those that potential users of such a map could easily replicate in their own experimental apparatus or industrial operations. These include load, pressure, relative velocity, and so on.

Researchers constructing wear maps tend to follow their individually favored approach; the choice of these slightly different approaches is almost always a personal one.



Wear Maps, Fig. 3 A map for the flank-wear characteristics of both coated and uncoated carbide tools during single-point dry turning. The application of TiC coatings leads to the expansion of the safety zone and the least-wear regime (Lim and Lim 2001)

The following describes briefly the steps adopted by the author (Lim 1998).

1. For the pair of materials of interest, their mode of contact (e.g., unidirectional sliding), contact geometry (e.g., pin-on-disk), the environment in which they are to interact in (e.g., atmospheric condition), and lubrication condition will first have to be decided.
2. Gather experimental data from the literature on wear rate and wear mechanism pertaining to this pair measured in conditions exactly like or very close to those specified in step (1). In-house tests will have to be carried out if data is lacking. Mathematical models describing the wear behavior of this pair should be gathered as well. One is expected to look for more than one wear model because more than one dominant wear-mechanism is expected within the wide range of operating conditions covered by the map.
3. Decide on the parameters to be used as axes for the map. One can construct a two- or three-dimensional wear diagram; so far, the majority of wear maps are of the two-dimensional type, and this trend is likely to be maintained. The range of operating conditions to be included in the map will also have to be decided. It is desirable to select as wide a range as possible. For situations such as machining, the ranges of machining conditions should preferably be similar to those recommended by the tool manufacturers for that particular group of tools whose wear behavior is to be mapped.
4. Construct the empirical wear maps. This is done first by grouping the wear data according to the mode and mechanism of wear. The wear-rate and wear-mechanism data, appropriately classified, are then plotted onto the (usually) two-dimensional space defining the map. The field of dominance of each mechanism is then demarcated using field boundaries and the approximate locations of the contours of constant wear-rate are located. At this stage, the wear map is sufficiently informative and it should provide a summary of the global wear behavior of the pair of materials in tribological interactions.
5. The final step is to introduce the appropriate mathematical models available describing the wear behavior of this pair in tribological interactions in different

fields within the map. When these are not available, new models may have to be developed. Noting that temperature usually plays a controlling role at the interface when the relative (sliding) velocity there is high, appropriate temperature calculations need to be incorporated into the wear models used when higher velocities are involved. Parameters used in the wear models are calibrated using the experimental data available. The calibrated model for each field is then used to calculate the projected wear rates for conditions in the field where no experimental data is available. These wear-rate contours are superimposed onto the map. A complete wear map is thus generated.

The construction of a wear map in the manner described above is generally a time-consuming effort.

Key Applications

How to Use a Wear Map

A wear map is like a street map. A street map of a particular city tells a tourist where she is as she walks around in that city; it also tells her the direction(s) she should take to reach her destination(s). A wear map constructed for a particular pair of materials in specific tribological interactions tells a reader the different types of wear behavior to expect at different operating or sliding conditions; it also tells him the most appropriate operating or sliding conditions to be employed if he wants to see little wear damage.

A properly constructed wear map holds much information. Take, for example, the wear map for steels shown in Fig. 1. It was constructed based on mathematical models that predict wear rates derived from the operation of different dominant wear-mechanisms and calibrated against actual experimental data. The resultant wear-rate contours should quite accurately predict the wear-rates to be expected in "gaps" within each field where no actual experimental data were previously available. (The field within a wear map is the region defined by the operating conditions where one particular wear mechanism has dominance and is responsible for most of the wear produced.) Herein lies the usefulness of such a map – its ability to predict wear rates with reasonable accuracy as well as the dominant wear-mechanism for a particular set of sliding conditions (in this case, normalized pressure and sliding velocity) without the users having to conduct experiments to gather wear-rate data or to undertake calculations using known mathematical (wear) models.

At this juncture, it is important to recognize that the majority of tribological systems almost always operate in a well-lubricated environment. When the lubricant does its job properly, separating the contacting surfaces, there is almost no wear. (Some wear, though, is expected to occur during the various start-stop phases of a lubricated system.) However, lubrication starvation can occur during the operation of an engineering system. Even if this lasts only for a short period of time, the resultant direct contact between surfaces (without the lubricant for protection) can lead to considerable wear damages. In this light, a wear map is used to plan for the worst-case scenario when there is a serious or a total lack of lubricant, and for how best to protect the sliding system from excessive wear by designing it (based on the information provided by the wear map) to operate in a mild-wear condition when lubrication fails, i.e., in an unlubricated condition. This is the basis for using a map constructed for an unlubricated-wear situation for a lubricated system.

As an illustration, the wear map given in Fig. 1 is used to determine the optimal sliding condition for a pair of steel components intended for relative sliding motion in an engineering system. This process is interactive in nature.

The designer would have, from other considerations, arrived at a certain range of sliding velocity and contact pressure for this pair of steel components. A quick check with the wear map for steels (Fig. 1) will yield the estimated wear-rate and expected dominant wear-mechanism under the intended range of velocity and contact pressure. If this sliding condition say, falls into the higher-wear region within the "delamination wear" field, it would be useful to see whether the sliding condition could be brought (say, through a reduction in contact pressure by re-designing the contact geometry) to the lower-wear region of this field, moving through the mild-to-severe wear transition region. If this cannot be achieved due to other constraints, one could explore the option of raising the sliding velocity to move the sliding condition into the field of mild oxidational wear. Noting that mild-to-severe wear transition occurs over a narrow band of sliding velocity within this field, great care must be taken to ensure that the chosen sliding condition will continue to produce a mild-wear condition. The process would end once the selected sliding condition brings the components into a mild-wear condition according to the map.

In certain situations, wear maps also provide crucial information concerning the specific environment the tribological system should operate in. For example, if the desired sliding condition falls in the field where mild oxidational wear is the dominant mechanism, not only must the sliding condition be carefully selected to stay in mild wear (because minor fluctuations could

lead to a transition to severe wear), it is also essential that the environment in which the sliding components operate must have a sufficient supply of oxygen. Any interruption to the oxygen supply will cause the dominant wear mechanism to switch from one that is mild (because the oxide formed helps to reduce wear) to one that is akin to a machining process, leading to severe wear damage (Lim 2002). The oxide film (as predicted by the wear map when oxygen is present) would not be in place to protect the components should lubrication fail, resulting in considerable wear damages before a fresh supply of lubricant could again separate the surfaces in contact. The appropriate use of a wear map can help prevent such an environment-induced transition into severe wear.

At the end of this interactive process, a suitable sliding condition should be found for this pair of steel components so that they will always operate in a mild-wear condition, even when lubrication fails. In the event that the desired sliding condition cannot be attained because of other constraints, the components might have to be redesigned. The process to select the optimal sliding condition would have to be repeated for the newly redesigned components.

When the desired sliding condition is finally determined, it is useful to carry out wear tests simulating the chosen sliding condition. This is to ensure that the components will produce the same (or nearly similar) wear rate and wear mechanism as predicted by the wear map.

A similar approach can be adopted when making use of wear maps constructed for tools used in different machining operations. A wear map constructed for a cutting tool used in a particular machining condition should allow the optimized machining condition to be selected. This optimized condition allows the desired rate of material removal (with a certain surface finish of the workpiece) to be attained at an acceptable level of tool wear. While the application of different coatings onto cutting tools reduces tool wear considerably, it has been found that, under certain machining conditions, the wear rate of coated inserts may not be very different from that of their uncoated counterparts. It would therefore be more cost-effective to either switch to uncoated tools when these machining conditions have to be adopted or to select another set of machining conditions that would allow the coated inserts to perform far better than the uncoated ones in terms of tool wear, while still meeting the production and quality targets. Wear maps can help to maximize the cost effectiveness of coated tools and inserts. Readers are referred to Lim and Lim (2001) for more details on this issue.

References

T.H.C. Childs, The sliding wear mechanisms of metals, mainly steels. Tribol. Int. **13**, 285–293 (1980)

S.C. Lim, Recent developments in wear-mechanism maps. Tribol. Int. **31**, 87–97 (1998)

S.C. Lim, The relevance of wear-mechanism maps to mild-oxidational wear. Tribol. Int. **35**, 717–723 (2002)

S.C. Lim, M.F. Ashby, Wear-mechanism maps. Acta Metall. **35**, 1–24 (1987)

S.C. Lim, C.Y.H. Lim, Effective use of coated tools – the wear-map approach. Surf. Coat. Technol. **139**, 127–134 (2001)

S.C. Lim, M.F. Ashby, J.H. Brunton, Wear-rate transitions and their relationship to wear mechanisms. Acta Metall. **35**, 1343–1348 (1987)

S.C. Lim, M.F. Ashby, J.H. Brunton, The effects of sliding conditions on the dry friction of metals. Acta Metall. **37**, 767–772 (1989)

M. Okoshi, H. Sakai, Research on the mechanism of abrasion. Report III, mechanism of abrasion of cast iron and steel. Trans. JSME **7**, 29–47 (1941)

A.R. Riahi, A.T. Alpas, The role of tribo-layers on the sliding wear behavior of graphitic aluminum matrix composites. Wear **251**, 1396–1407 (2001)

D. Tabor, Status and direction of tribology as a science in the 80s, in *Proceedings of International Conference on Tribology in the 80s* (NASA Lewis Research Centre, Cleveland, 1983), pp. 1–17

Wear Mechanisms in Electrical Contacts: Abrasive Wear

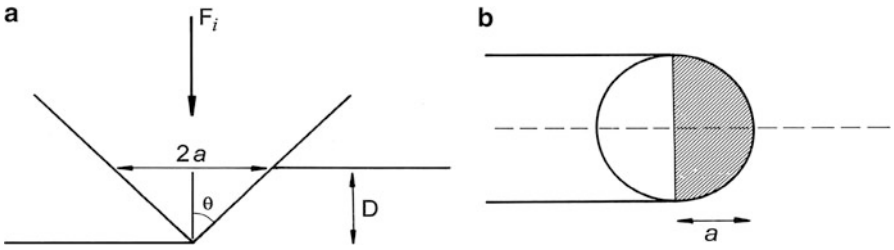
Roland S. Timsit
Timron Advanced Connector Technologies,
A Subsidiary of Timron Scientific Consulting Inc.,
Toronto, ON, Canada

Definition

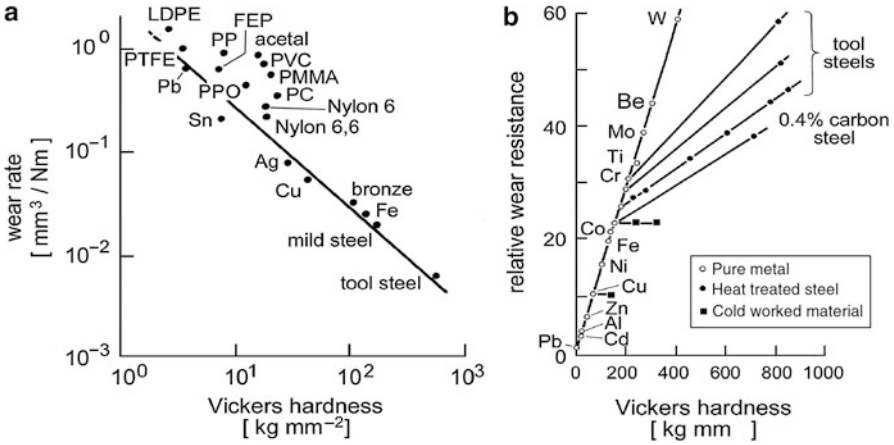
Abrasive wear describes a mechanism whereby material is removed or displaced from a surface by abrasion due to rubbing against hard particles. The mechanical damage generated by abrasive wear generally has a deleterious effect on electrical contact reliability.

Scientific Fundamentals

A simple model for abrasive wear involves the removal of material by plastic deformation, as illustrated in Fig. 1 (Timsit et al. 2008). Consider the action of a large number of abrasive particles with each represented as a cone of semi-angle θ_i dragged across the surface of a ductile material to form a groove. The width of the groove generated by the ith particle is $2a_i$, where a_i is the cone radius at the top of the groove of depth D_i. If F_i is the average normal applied force on the ith particle and σ_{abr} is the maximum mechanical stress on the area πa_i^2, then F_i is estimated as



Wear Mechanisms in Electrical Contacts: Abrasive Wear, Fig. 1 Simple abrasive wear model in which a cone removes material from a surface: (**a**) elevation view of a conical abrasive particle forming a groove in a substrate, (**b**) plan view of the abrasive particle at a distance D from its apex



Wear Mechanisms in Electrical Contacts: Abrasive Wear, Fig. 2 (**a**) Correlation between abrasive wear rate and Vickers hardness of materials ranging from polymers to tool steel. The abrasive material was silicon carbide. These data refer to *two-body* abrasion since the SiC particles were attached to a holding medium, such as a cloth, (**b**) relative resistance to wear (1/wear rate) for pure metals and selected steels, (**c**) relative wear rate as a function of the ratio of the hardness of the abrasive to the hardness of the substrate (Hutchings 1992)

$$F_i = \sigma_{abr}\frac{\pi a_i^2}{2} = \frac{1}{2}\sigma_{abr}\pi D_i^2 \tan^2\theta_i \qquad (1)$$

Since each hard particle is pushed into the surface and deforms it plastically, an estimate for the contact stress is $\sigma_{abr} \approx H$, where H is the hardness of the worn material. In the simple picture in which the volume of material removed from a path of length L is $\varepsilon a_i D_i L = \varepsilon L D_i^2 \tan\theta$, where ε is a probability of removal, the wear rate q_i defined as the material volume displaced per unit sliding distance by the *i*th particle is

$$q_i = \frac{\varepsilon L D_i^2 \tan\theta_i}{L} = \frac{2\varepsilon F_i}{\pi \tan\theta_i H} = \left(\frac{2\varepsilon k_i}{\pi \tan\theta_i}\right)\frac{F}{H} \qquad (2)$$

since F_i is proportional to the total applied force F as $F_i = k_i F$. Summing over all abrasive particles, the total wear rate is $Q = \sum_i q_i$, which reduces to

$$Q = K\frac{F}{H}$$

where $K = \sum_i \frac{2\varepsilon k_i}{\pi \tan\theta_i}$.

The wear rate Q is measured as the volume or mass of material worn away per unit sliding distance. Thus, K is a constant that depends on materials properties (via the parameter ε) and the geometry of the abrasive particles.

Typical abrasive wear rates measured in volume of worn material per unit force and unit travel distance are illustrated in Fig. 2, where the abrasive material is silicon carbide (Hutchings 1992). The wear rate decreases rapidly with increasing Vickers hardness of the worn material. The relative resistance to abrasive wear (1/relative wear rate) of pure metals in comparison with that of selected steels, as a function of the Vickers hardness of the metal,

is shown in Fig. 2b. It is worth noting that electrical contact materials such as silver and copper are not particularly resistant to abrasive wear.

Key Applications

Abrasive particles are intentionally introduced in aluminum connectors, and particularly power connectors, to abrade aluminum oxide surface films and promote the formation of metal-metal contacts in electrical connections involving aluminum (Schindler et al. 1996). In these applications, the abrasive medium often consists of particles of a hard metal such as steel or ceramic powder suspended in a lubricant or mixed into a grease.

In electrical connections where contact surfaces are coated with a soft metal such as tin, for example, in tin-coated copper busbars, the use of abrasive grit to break up tin oxide films requires special precautions. In these applications, the grit must consist of relatively soft material such as zinc to minimize damage to the coating and preclude significant tin loss.

Cross-References

- ► Friction
- ► Friction Coefficient
- ► Sliding Electrical Contact Wear
- ► Tribology

References

I.M. Hutchings, *Tribology: Friction and Wear of Engineering Materials* (CRC Press, Boca Raton, 1992)

J.J. Schindler, R.T. Axon, R.S. Timsit, Mechanical and electrical properties of wedge connectors. IEEE Trans. Compon. Packag. Manuf. Technol. A **19**, 287–294 (1996)

R.S. Timsit, *Course on Performance and Reliability of Power Electrical Connections*, Ft. Lauderdale (2008)

Wear Mechanisms in Electrical Contacts: Adhesive Wear

Roland S. Timsit
Timron Advanced Connector Technologies,
A Subsidiary of Timron Scientific Consulting Inc.,
Toronto, ON, Canada

Definition

Adhesive wear describes a mechanism whereby materials fragments are pulled from one sliding surface and are either transferred over to the mating sliding surface or are dispersed as loose particles over the contacting bodies. The mechanical damage generated by adhesive wear generally has a deleterious effect on electrical contact reliability.

Scientific Fundamentals of Adhesive Wear on Electrical Contacts

The adhesive wear mechanism is illustrated schematically in Fig. 1 (Hutchings 1992). It is initiated when localized adhesion occurs at contact spots between sliding surfaces as illustrated in Fig. 1a. This is followed by repeated transfer of material from one asperity to the other and growth of the transferred material (Fig. 1b–d). Transferred particles eventually detach themselves and are subsequently flattened and dispersed as loose wear debris (Fig. 1e). For non-noble materials, the debris undergoes oxidation to form highly electrically insulating material.

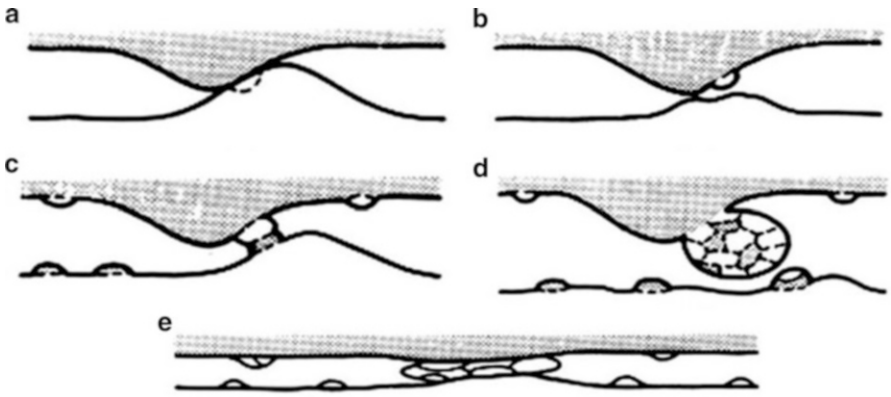
A small amount of adhesive wear can often act to reduce rather than increase contact resistance since this can disrupt and disperse electrically insulating layers such as oxides, sulfides, and other materials formed from chemical reactions with air or environmental pollutants. Such disruption of electrically insulating layers is often labeled as "wiping."

"Wiping" Contaminant and Oxide Layers from Contact Surfaces of Non-noble Materials

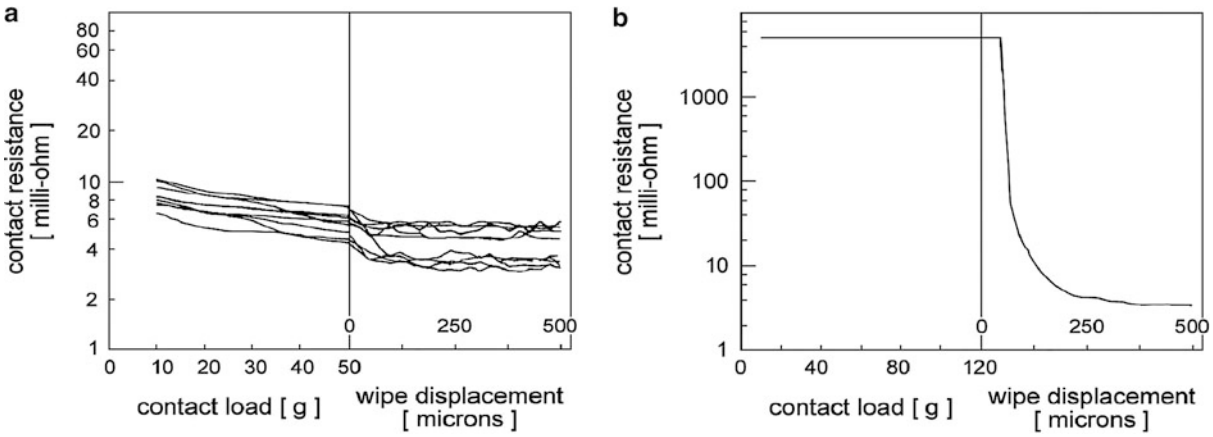
Typical effects of "wiping" are illustrated in Fig. 2a, b (Brockman et al. 1988). "Wiping" is usually recommended as a first step in generating low contact resistance in separable electrical connections such as tin-coated pin/socket contacts, in order to disperse surface oxide layers. This precaution is recommended even for gold-plated contacts to enhance contact reliability, since gold has a tendency to be contaminated by thin carbonaceous layers (Smith 1980).

"Wiping" of Aluminum Electrical Contact Surfaces: A Focus

Aluminum oxide forms almost instantaneously on bare aluminum in air (Natesh and Ansell 1966). The native oxide film consists of amorphous material (Timsit et al. 1985) with a thickness of about 2 nm and is electrically insulating despite its small thickness. A good metal-to-metal electrical contact with aluminum is achieved only if the oxide is fractured or abraded to expose the underlying metal, and electrical contact is made before re-oxidation. The challenge of forming a reliable aluminum-aluminum contact has led to the development of a connector technology (i.e., the "fired-wedge" technology (Broske 1966)), capable of generating a highly reliable high-power electrical connection with aluminum conductor in power distribution circuits.



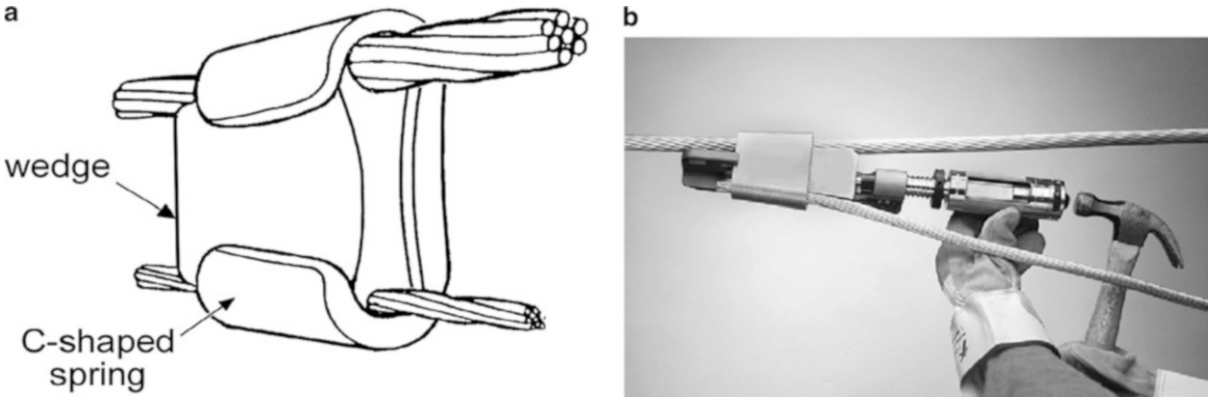
Wear Mechanisms in Electrical Contacts: Adhesive Wear, Fig. 1 Schematic representation of adhesive wear in a sliding interface (From Hutchings 1992): (**a**) localized adhesion between contacting surface asperities during sliding, (**b**) transfer of material from one asperity to the other, (**c**) repetitive adhesion process, (**d**) growth of transferred particles, (**e**) particle detachment and subsequent debris flattening



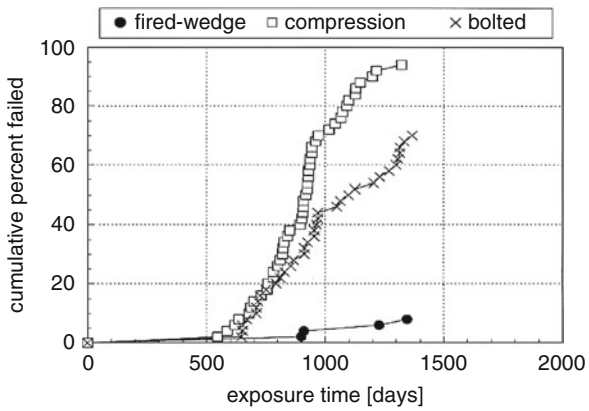
Wear Mechanisms in Electrical Contacts: Adhesive Wear, Fig. 2 Contact resistance versus normal contact load and versus wipe distance for a gold probe sliding on: (**a**) a contaminated gold-plated surface, (**b**) a highly contaminated gold-plated surface (Brockman et al. 1988)

This connector is shown in Fig. 3 and is based on activating a vigorous high-speed "wipe" on the conductor and connector surfaces during installation to disrupt oxide layers effectively (Schindler et al. 1996), leading to outstanding connector service performance. The connector consists of a metal wedge located between the feed and tap cables situated at opposite ends of a C-shaped metal component. The wedge is installed at high speed between the conductors using a tool actuated by a powder cartridge (Broske 1965) and "wipes" the sliding interfaces with the conductors. The "wiping" action displaces the aluminum oxide layers in the contact interfaces, and generates metal-to-metal electrical interfaces of high integrity.

The performance superiority of fired-wedge connectors is evidenced by results of testing. Figure 4 shows results of exposure to a marine environment of populations of the fired-wedge, bolted, and compression connectors shown in Fig. 5 (Callen et al. 2000). The bolted and compression connectors, where "wipe" is minimal, failed much more rapidly than the fired-wedge counterpart. Because power connectors of the type shown in Fig. 5 are widely used in electric distribution systems, the data of



Wear Mechanisms in Electrical Contacts: Adhesive Wear, Fig. 3 Fired-wedge connector: (**a**) schematic view of connector assembly, (**b**) installation of the connector requires the use of a cartridge-operated tool



Wear Mechanisms in Electrical Contacts: Adhesive Wear, Fig. 4 Cumulative percentage of failed connectors versus exposure time, for the populations of the three connectors shown in Fig. 9 (Callen et al. 2000)

Fig. 4 clearly point to potential energy and economic benefits of using connectors with superior performance, and in particular the benefits of "wiping" aluminum oxide surface films in utility-type power connectors. This is worthy of further elucidation.

Example of Energy Savings and Economic Benefits of the Use of Power Connectors of Superior Design

The contact resistance between a conductor and a deteriorating power connector can give rise to an unacceptably large energy loss. The Joule energy E_L lost cumulatively over a time interval t in a deteriorating electrical contact passing an electrical current I is given as (Timsit and Sprecher 1998)

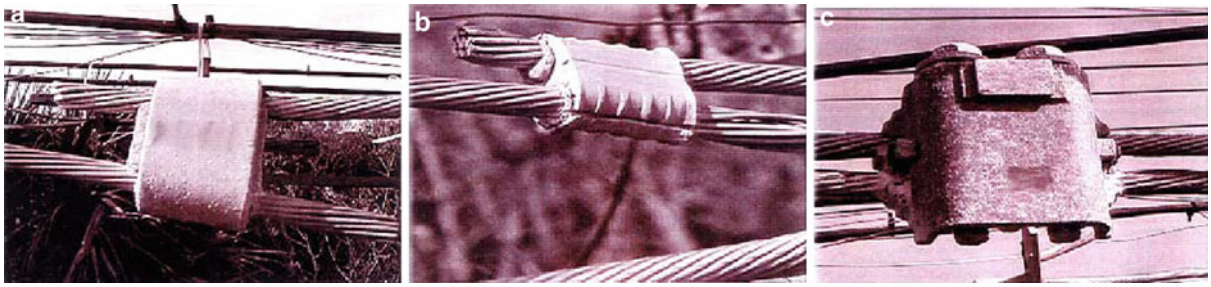
$$E_L = \left(\frac{\Delta R}{\Delta t}\right)\frac{I^2 t^2}{2} F_l \qquad (1)$$

where $(\Delta R/\Delta t)$ is the rate of increase of contact resistance (assumed to be constant) and F_l is a load factor for the line on which the connector is installed. Extensive laboratory measurements on aluminum power connectors, including fired-wedge connectors, indicated that $(\Delta R/\Delta t)$ could vary from a very small value for fired-wedge connectors, to conservative values of about 15 μΩ/year for aluminum-bolted connectors and 75 μΩ/year for compression connectors of the type illustrated in Fig. 5 (Timsit and Sprecher 1998). On the basis of these rates of resistance increase, and a load factor F_l of 0.5, Figs. 6 and 7 show the potential cumulative cost of wasted energy in electrical distribution lines at currents of 100 A, 200 A, and 400 A, respectively, for a bolted and compression connector of the type illustrated in Fig. 5. The generation cost of electricity was assumed to be $0.05/kWHr. The cumulative cost associated with degradation in bolted and compression connectors becomes appreciable after about 2 years of operation on a line carrying a peak current of 400 A. Recently, Braunovic et al. (2007) have followed Timsit's approach to evaluating losses in power connectors and have calculated a far higher cost than illustrated in Figs. 6 and 7.

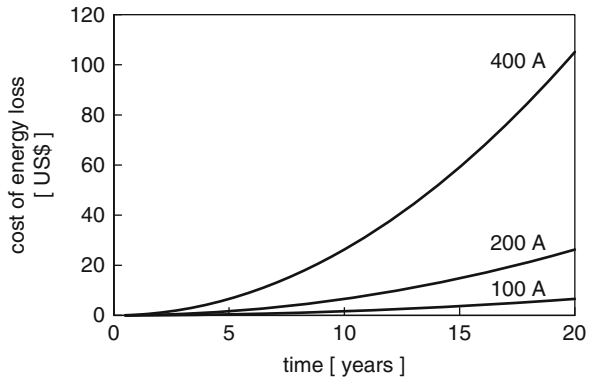
Effects of Adhesive Wear on Electrical Contacts: Noble Metals

Gold Electrodeposits

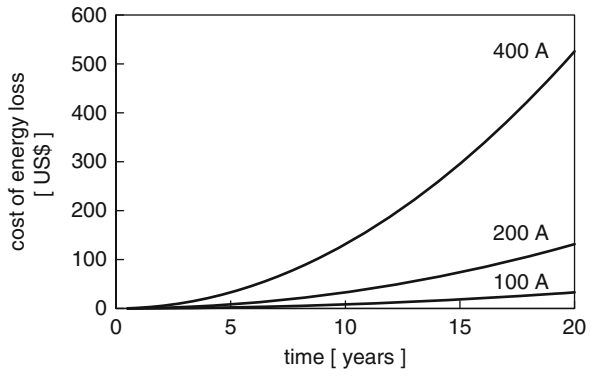
With oxide-free noble metals such as gold, organic material accumulates on the surface after exposure to air. In this case, adhesive wear is effective in dispersing these



Wear Mechanisms in Electrical Contacts: Adhesive Wear, Fig. 5 Photographs of a (**a**) fired wedge, (**b**) compression, and (**c**) bolted connector after approximately 3 years of exposure at the test site described in Callen et al. (2000)



Wear Mechanisms in Electrical Contacts: Adhesive Wear, Fig. 6 Cumulative cost of power loss per bolted connector due to contact degradation (Timsit and Sprecher 1998)



Wear Mechanisms in Electrical Contacts: Adhesive Wear, Fig. 7 Cumulative cost of power loss per compression connector due to contact degradation (Timsit and Sprecher 1998)

contaminants if the contact load exceeds a "transition" load. The transitions loads for gold electroplates are listed in Table 1.

Gold electrodeposits are widely used on contact components in electronic connectors and adhesive mechanisms dominate their mechanical wear properties. Because pure gold is soft and has a tendency to cold-weld in a contact, metals such as Ni or Co are often co-deposited with gold during plating to increase hardness and mitigate mechanical wear.

Over the last decade, the requirement for connectors located under automobile engine hoods to operate reliably has led to investigations of wear induced by slow and rapid reciprocating displacements at elevated temperatures. For gold-plated contacts, these mechanisms include metal softening, possible rapid interdiffusion with substrate materials, and diffusion to the surface of hardeners in gold (Antler 1999).

Effects of Underplates on Wear Properties of Gold Electrodeposits

Hard materials are more resistant to wear than soft materials. On coated electrical contact surfaces such as gold-plated copper alloys, a layer of relatively hard (and usually stiff) material located underneath the gold should act to reinforce the substrate, thus reducing local elastic deflection and reducing "ploughing" action on the gold. This in turn reduces the effective wear rate on the gold layer.

Examples of the benefits of using a hard nickel underplate are given in Fig. 8 for adhesive wear (Antler and Drozdowicz 1979). In these plots, the Wear Index is defined as the ratio of the sum of path lengths over which material transfer occurs (due to wear) divided by the entire length of the track followed by the slider, multiplied by 100 (Antler 1999). Thus a high Wear Index describes high wear rate. The data of Fig. 8 clearly confirm the beneficial effects of the nickel underplate in reducing

wear. The data also provide another example of the higher wear resistance of hard gold over that of pure (soft) gold.

Effects of Surface Roughness on Wear Properties of Gold Electrodeposits

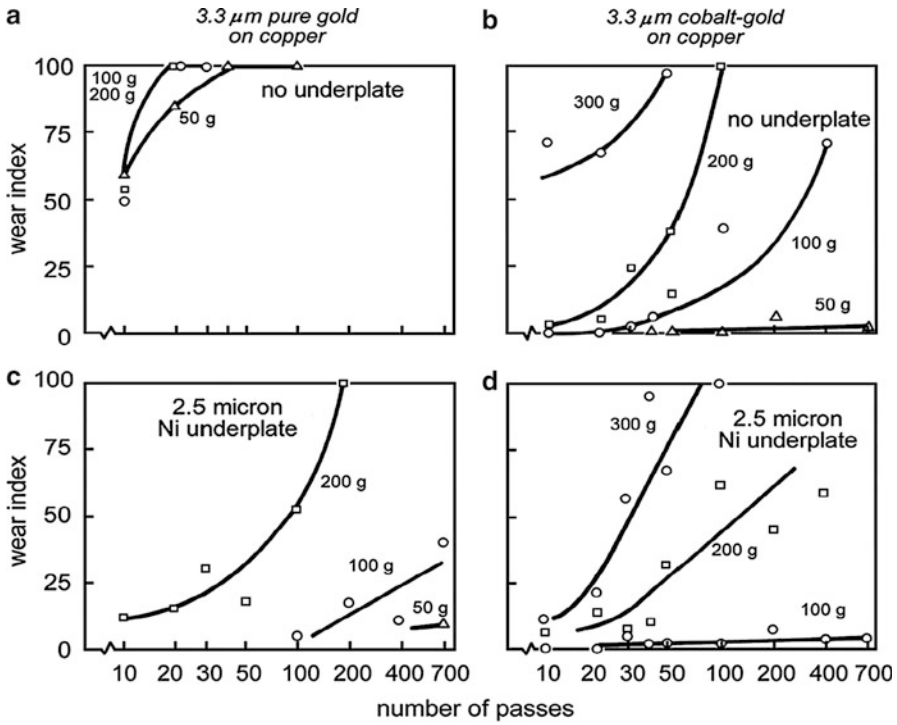
Generally, surface roughness has little effect on the wear rate of *bulk* metals (Rabinowicz 1981). However, similar observations do not generally apply to layered materials, possibly due to non-uniform metallurgical properties or metallurgical defects across surface layers, and to variations in bonding strength with the substrate. Gold-plated contact surfaces are indeed more resistant to adhesive wear or other wear mechanisms when the surface roughness is small than when the roughness is large. This is illustrated by the data in Fig. 9 (Antler 1980). The use of a nickel underplate reduces the dependence of wear rate on surface roughness, as shown in Fig. 9b.

Wear Mechanisms in Electrical Contacts: Adhesive Wear, Table 1 Contact load for transition from mild wear to severe wear for gold electroplates

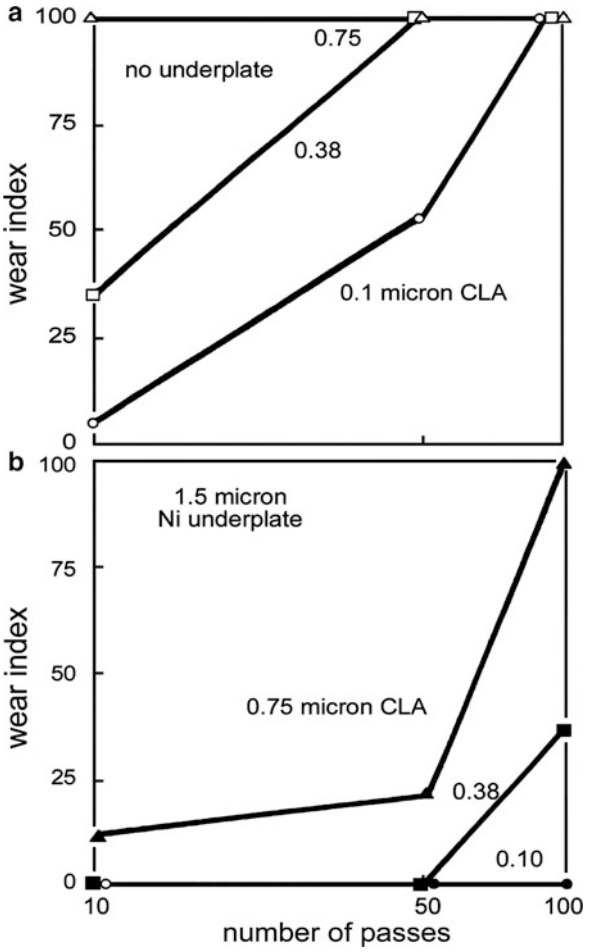
Surface condition	Transition load for contact materials (gf)	
	Pure gold	Hard gold
Clean surface	5	10
Contaminated surface	10–50	25–300
Lubricated surface	100–500	500–2,000

Palladium and Palladium Alloys

One potential substitute for gold is palladium. Palladium is available not only as an electroplate but also as a wrought or a cladding material and is appreciably less costly than gold. One major challenge to the use of unalloyed palladium is its susceptibility to the formation of electrically insulating compounds known as "frictional polymers" in the presence of even small traces of organic contaminants in air (Hermance and Egan 1958). The tendency of palladium to form frictional polymers stems from the catalytic action of the metal on organic deposits to form a variety of insulating organic compounds.



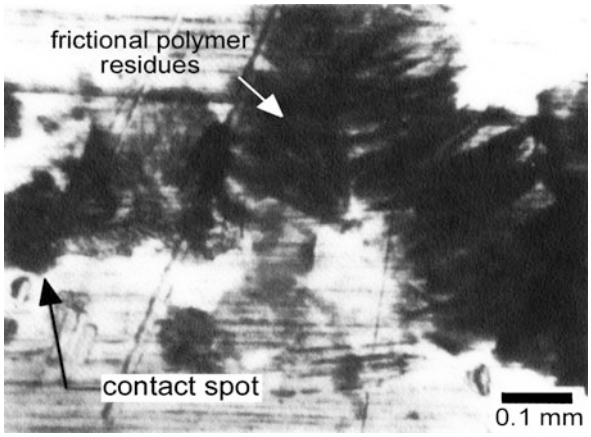
Wear Mechanisms in Electrical Contacts: Adhesive Wear, Fig. 8 Wear Index versus the number of reciprocation passes, from adhesive wear tests on 3.3-μm-thick gold electrodeposits: (**a**) and (**c**) ductile pure gold, with and without 2.5-μm-thick nickel underplate; (**b**) and (**d**) hard cobalt with and without 2.5-μm-thick nickel underplate. The Wear Index is defined as the length of wear track with wear features to the length of the entire slider displacement track, multiplied by 100 (Antler 1999)



Wear Mechanisms in Electrical Contacts: Adhesive Wear, Fig. 9 Variations of Wear Index with repeated passes of a solid gold rider over a 0.6-μm-thick layer of cobalt-gold electrodeposited on a copper flat, for three different surface roughness values: (**a**) no nickel underplate, (**b**) using a 1.5-μm-thick nickel underplate (Antler 1999)

This tendency for frictional polymer formation is greatly enhanced in the presence of small oscillatory interfacial displacements (i.e., fretting wear, which is addressed later). Frictional polymers also form on other Group 10 metals such as rhodium and ruthenium. An optical micrograph of a typical patch of frictional polymer is shown in Fig. 10 (Antler 1999).

One approach to mitigating frictional polymer formation in palladium-palladium contacts is to use a gold flash (thickness of ~0.1 μm) or a thin lubricant layer on the palladium surfaces. A thin lubricant or gold layer has the effect of preventing access to the palladium surface by polymer-forming organic contaminants. Extensive testing



Wear Mechanisms in Electrical Contacts: Adhesive Wear, Fig. 10 Frictional polymer formed after repeated reciprocating sliding in a palladium-palladium contact (Antler 1999)

has indicated that a palladium electroplate deposited on a nickel underplate and flashed with a thin gold layer, performs at least as well as gold over nickel (Grossman et al. 1980). Contacts plated with as little as 0.1 μm gold over 0.6 μm palladium with a 1.5-μm nickel underplate have proven durable for practical uses in ordinary electronic devices.

Promising substitute materials for hard gold as a contact finish include palladium-nickel and palladium-cobalt with nickel and cobalt concentrations ranging from about 10 to 20 wt%. For several years, another promising substitute material for hard gold has been palladium-silver, consisting of 60 wt% palladium and 40 % silver (Pd60Ag40). This material may be electrodeposited on a contact surface but is also available as a wrought alloy (also known as R156) and may be incorporated into a surface as a clad inlay that is then further fabricated into contact components (Nobel 1984). One alloy fabricated to offer a gold-rich surface is a modified version of R156 and has proven to have excellent contact properties when subjected to screening evaluations consisting of corrosion, wear, thermal aging, and reciprocating motion with micron-scale displacement amplitude. This material, called Diffused Gold R156 (DG-Rl56), is characterized by a contact performance that is essentially indistinguishable from that of plated cobalt-hardened gold but with much lower cost (Bader 1982).

Silver

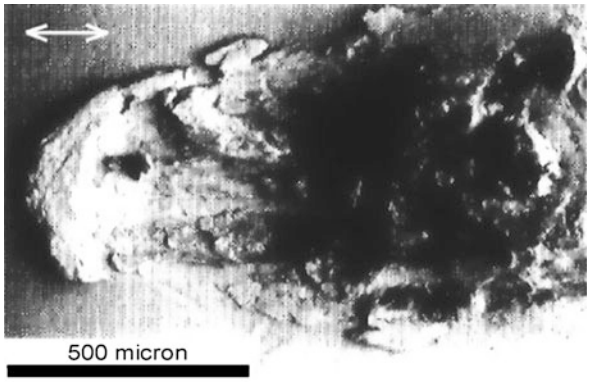
Silver is increasingly regarded as a relatively low-cost alternative to gold. Clean silver contact surfaces have

a tendency to *stick* (i.e., *weld*) under a sufficiently large contact load. This *stick* tendency leads to a high coefficient of friction, which leads to poor durability performance and poor fretting properties where sliding displacements are large. In contrast with gold, silver readily forms surface tarnish films when exposed to sulfur-containing and chlorine-containing atmospheric contaminants. Field exposure studies of silver-plated connectors have shown that tarnish films on silver consist primarily of α silver sulfide (Ag_2S) and to a lesser extent silver chloride (AgCl), which is more insulating and potentially harder to displace than the sulfide (Abbott 1982). These tarnish films are generally semi-conductive at ambient temperatures, mechanically soft, and relatively easily displaced with contact interface wipe at sufficient normal loads. The presence of tarnish films has a significant effect on the wear properties of silver.

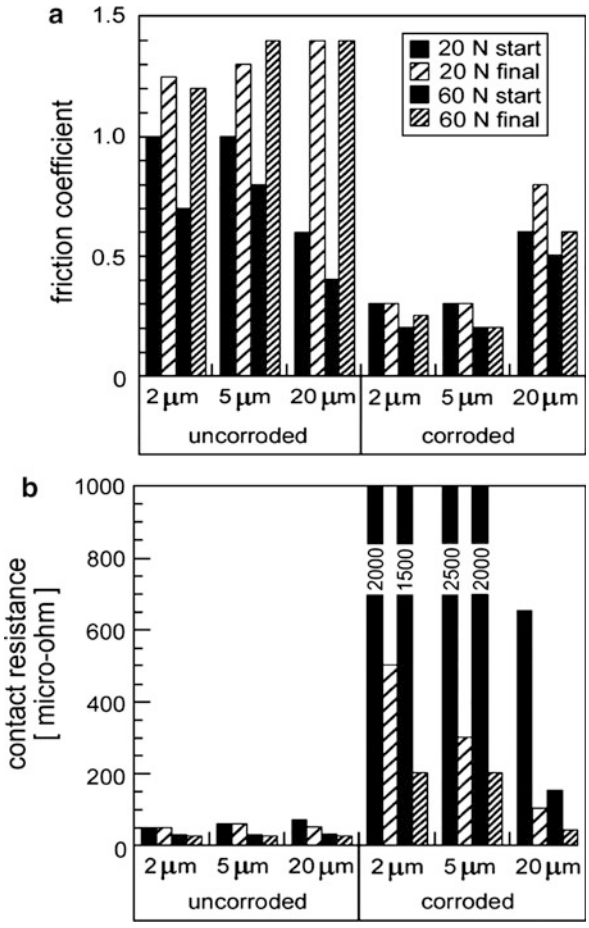
Figure 11 shows an example of the wear sustained in reciprocation motion by clean/untarnished silver surfaces in a sliding couple consisting of crossed copper rods plated with a 4-μm-thick silver coating, at a contact force of 60 N (Kassman Rudolphi and Jacobson 1997). As expected for clean silver at this relatively large contact load, the dominant deterioration mechanism in all untarnished contacts was adhesive wear and mechanical deformation increased with increasing contact force. For tarnished surfaces, the amount of mechanical wear depended strongly on the silver coating thickness. Data on friction coefficient and contact resistance relating to tarnished and untarnished silver contacts are shown in Fig. 12.

The important contact properties of silver coatings are as follows:

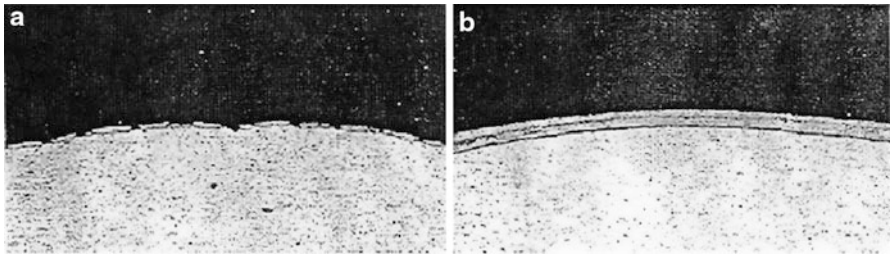
1. Tarnish films on silver are relatively easily displaceable; although the data presented in this section were obtained at relatively large contact loads, similar observations have been reported for contact forces of a few Newtons for electronic connector applications;
2. The presence of tarnish does not necessarily imply high contact resistance, if sufficient relative interfacial displacement is allowed to disrupt the tarnish layers;



Wear Mechanisms in Electrical Contacts: Adhesive Wear, Fig. 11 Example of the appearance of a sliding surface from a clean (untarnished) silver–silver couple after 20 cycles of reciprocating sliding at a frequency of 0.2 Hz and a displacement amplitude of 3 mm. The silver layer was 4 μm thick and the contact force was 60 N. The sliding direction is indicated by the *arrow* (Kassman Rudolphi and Jacobson 1997)



Wear Mechanisms in Electrical Contacts: Adhesive Wear, Fig. 12 (**a**) Contact resistance for tarnished and untarnished silver-silver contacts, (**b**) friction coefficient relating to the experimental data of (**a**) during the first cycle and after about 20 cycles of sliding (0.2 Hz, 3 mm). The data show average values from three separate tests (Kassman Rudolphi and Jacobson 1997)



Wear Mechanisms in Electrical Contacts: Adhesive Wear, Fig. 13 Cross sections showing the surface of contact components (Lees and Greschner 1998): (**a**) electroplated surface showing fractures in the electrodeposit after forming, (**b**) clad surface showing no fractures in the cladding after forming

3. Relatively thick silver coatings generally lead to lower contact resistance than thin coatings, particularly at large contact loads (Imrell 1991); the superior electrical contact properties of thicker coatings is attributed to easier metal deformation in thick metal coatings and hence easier disruption and dispersal of tarnish films on the surfaces of thick silver.

Tribological Properties of Clad Versus Electrodeposited Materials

Wrought materials such as R156 and DG-R156, described earlier, constitute part of a growing family of electrical contact cladding alloys. In the simplest arrangement, a cladding is made by compressing a thin foil of a selected metal alloy against a metal substrate generally thicker than the foil to achieve a metallurgical bond at the compressed interface. A cladding may also involve more than one metal foil. The thickness of a noble-metal inlay for electrical connectors typically varies from 0.75 to 5 μm.

The diversity and composition range of cladding materials contrasts with the narrow materials choice for electrodeposits. Electrodeposits generally comprise two elements selected from a narrow range of elements capable of being deposited from an electrolytic bath. In addition to the advantage of wide alloy range, selected physical properties of claddings such as yield strength and hardness may be controlled by controlling the grain size. Grain size control is achieved relatively easily via metal-working during various stages of the rolling process where the clad is thinned to its final thickness, and by subsequent thermal treatment. Such metallurgical tuning is difficult to achieve in electrodeposits where grain size is generally determined by the magnitude of the electroplating current and the impurity content in the electrolytic bath (Lowenheim 1974). Another advantage of claddings is mechanical formability after bonding to the substrate. Clad components may be deformed significantly without tensile fracture of the cladding, as illustrated in Fig. 13a, in contrast with the generally deleterious effects of deformation on electro-coated surfaces as illustrated in Fig. 13b (Lees and Greschner 1998). Due to factors such as wider composition range and superior "tunability" of metallurgical and mechanical properties, clad materials are generally more resistant to tribological wear than plated layers, provided the clad adheres strongly to the underlying substrate. However, one drawback of claddings is the difficulty of laying them on surfaces that are not either flat or uniformly curved, in clear contrast with electrodeposits. Another potential drawback is cost.

Cross-References

▶ Electrical Contact Materials
▶ Friction
▶ Friction Coefficient
▶ Sliding Electrical Contact Wear
▶ Tribology

References

W.H. Abbott, The effects of test atmosphere conditions on the contact resistance of surface films on silver, in *Proceedings of the 11th International Conference on Electric Contact Phenomena*, West Berlin, pp. 294–296 (1982)

M. Antler, Wear of gold contact finishes: the importance of topography, underplate, and lubricants. Insulation/Circuits **26**, 15–19 (1980)

M. Antler, Tribology of electronic connectors: contact sliding wear, fretting and lubrication, in *Electric Contacts: Theory and Applications*, ed. by P.G. Slade (Marcel Dekker, New York, 1999), pp. 309–402

M. Antler, M.H. Drozdowicz, Wear of gold electrodeposits: Effect of substrate and of nickel underplate, Bell. Syst. Tech. J. **58**(2), 323–349 (1979)

F.E. Bader, Diffused gold R-156: a new inlay contact material for bell system connectors, in *Proceedings of the 11th International Conference on Electric Contact Phenomena*, West Berlin, pp. 133–137 (1982)

M. Braunovic, V.V. Konchits, N.K. Myshkin, *Electrical Contacts: Fundamentals, Applications and Technology* (CRC Press, Boca Raton, 2007), p. 256

I.H. Brockman, C.S. Sieber, R.S. Mroczkowski, A limited study of the effects of contact normal force, contact geometry and wipe distance

on the contact resistance of gold-plated contacts. IEEE Trans. Compon. Hybrid. Manuf. Technol. **CHMT-11**, 393–400 (1988)
W.F. Broske, Explosive device to force a wedge into a clamp for clamping cables, Patent 3 212 534, 19 Oct 1965
W.F. Broske, C. Hill, H.W. Demler, W.H. Knowles, Jr., F.W. Wahl, Method of making an electrical connection to a stranded cable, Patent 3 235 944, 22 Feb 1966
B.W. Callen, B. Johnson, P. King, R.S. Timsit, Environmental degradation of utility power connectors in a harsh environment. IEEE Trans. Compon. Packag. Technol. **23**, 261–270 (2000)
H. Grossman, M. Huck, Tribological studies of electroplated sandwich layers on connectors, in *Proceedings of the 10th International Conference on Electrical Contact Phenomena*, Budapest, pp. 401–411 (1980)
H.W. Hermance, T.F. Egan, Organic deposits on precious metal contacts. Bell Syst. Tech. J. **37**, 739–766 (1958)
I.M. Hutchings, *Tribology: Friction and Wear of Engineering Materials* (CRC Press, Boca Raton, 1992)
T. Imrell, The importance of the thickness of silver coating in the corrosion behaviour of copper contacts, in *Proceedings of the 37th IEEE Holm Conference on Electrical Contacts*, Chicago, pp. 237–243 (1991)
A. Kassman Rudolphi, S. Jacobson, Stationary loading, fretting and sliding of silver coated copper contacts – influence of corrosion films and corrosive atmosphere. Tribol. Int. **30**, 165–175 (1997)
P.W. Lees, U. Greschner, The influence of coating technology on the mechanical performance of thin, high performance connector springs, in *Proceedings of the 31st Annual IICIT Connection and Interconnection Symposium*, Danvers, pp. 147–153 (1998)
E. Lowenheim, *Modern Electroplating* (Wiley, New York, 1974)
R. Natesh, G.S. Ansell, Oxide growth in a Al-Al_2O_3 SAP-type alloy by hot stage transmission electron microscopy. Acta Metall. **14**, 1735–1756 (1966)
F.I. Nobel, Electroplated palladium-silver (60/40 wt%) alloy as a contact material, in: *Proceedings of the 30th Holm Conference on Electrical Contacts*, Chicago, pp. 137–152 (1984)
E. Rabinowicz, *Friction and Wear of Materials*, ASM, p.129 (1981)
J.J. Schindler, R.T. Axon, R.S. Timsit, Mechanical and electrical properties of wedge connectors. IEEE Trans. Compon. Packag. Manuf. Technol. A **19**, 287–294 (1996)
T. Smith, The hydrophilic nature of a clean gold surface. J. Colloid Interf. Sci. **75**, 51–55 (1980)
R.S. Timsit, W.G. Waddington, C.J. Humphreys, J.L. Hutchison, Structure of the Al/Al_2O_3 interface. Appl. Phys. Lett. **45**, 830–832 (1985)
R.S. Timsit, J.D. Sprecher, Energy losses in power tap-connectors, in *Proceedings of the Transmission and Distribution Construction, Operation and Live-Line Maintenance*, ESMO-98, Orlando, pp. 285–296 (1998)

Wear Mechanisms in Electrical Contacts: Fretting Wear

Roland S. Timsit
Timron Advanced Connector Technologies,
A Subsidiary of Timron Scientific Consulting Inc.,
Toronto, ON, Canada

Synonyms

Contact of rough surfaces

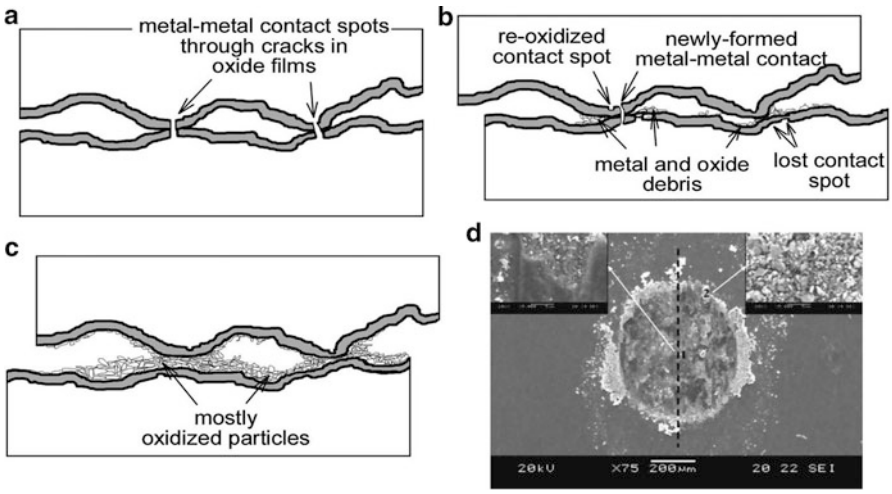
Definition

Fretting wear stems from material transfer across a sliding interface due to small-amplitude reciprocating movement in a contact interface, and the subsequent oxidation of material in that interface. Because fretting wear gives rise to enhanced oxidation of sliding surfaces, it is also described as fretting corrosion. Fretting wear/corrosion generally has a devastating effect on electrical contact properties.

Scientific Fundamentals

Fretting corrosion occurs where at least one of the contact surfaces oxidizes and eventually leads to the accumulation of oxidized wear particles in the interface due to reciprocating interfacial movement. The sequence of events leading to fretting corrosion in an electrical contact consisting of two oxidized surfaces is illustrated schematically in Fig. 1. Electrical contact is established initially through cracks in the oxide layers as shown in Fig. 1a. Subsequently, small reciprocating displacements disrupt some of the metal-to-metal contact spots and simultaneously generate metal debris, as illustrated in Fig. 1b. Finally, sufficient oxidized debris accumulates in the interface to prevent the formation of metal-to-metal contact as shown schematically in Fig. 1c. Figure 1d shows the micrograph of a typical fretting-corrosion scar on a tin surface after reciprocation sliding against another tin surface (Park et al. 2007). Most of the debris consists of oxidized particles.

In fretting corrosion, the amplitude of micromotion displacement giving rise to wear may range from a few micrometers to as much as 100 μm in electronic connectors (Antler and Drozdowicz 1981–1982). Interfacial micromotion is usually caused by external vibrations, by changing temperature that gives rise to differential thermal expansion of contact components and by electromagnetically induced vibrations. Fretting corrosion is observed in a wide variety of electrical and electronic connections in systems including consumer appliances, automotive and other transportation systems, telecommunication networks, and electrical connections in electrical power and distribution lines. One example of the potential disrupting effects of fretting corrosion in electrical contacts relates to the use of connectors in automobiles. Many automobiles use more than 400 connectors with 3,000 separate terminals. It is estimated that more than 60 % of electric problems in automobiles stem from fretting corrosion in these electrical and electronic connections. A comprehensive review of mechanisms of fretting wear and fretting corrosion has been given by Hurricks (1970).

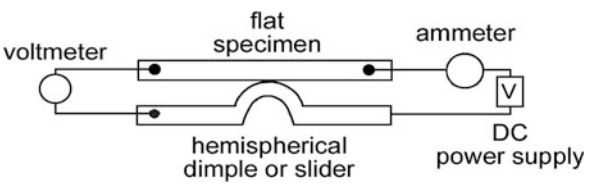


Wear Mechanisms in Electrical Contacts: Fretting Wear, Fig. 1 Schematic illustration of the fretting wear mechanism between oxidized contact surfaces: (**a**) initially, electrical contact is established through cracks in the surface oxide films, (**b**) as micro-reciprocating sliding occurs, metal-to-metal contact spots are lost and metal and oxide debris accumulate in the interface; metal-to-metal may be formed through new cracks in the oxide layers, (**c**) eventually, the accumulation of oxidized debris is so large as to prevent metal-to-metal contact, (**d**) micrograph of a typical example of fretting corrosion on a tin surface after reciprocating sliding with another tin surface. Most of the debris consists of oxidized tin particles (Park et al. 2007)

Gold in the form of a thin coating is a preferred contact material in electronic connectors because it does not oxidize and is therefore not susceptible to fretting corrosion. However, the high cost of gold now precludes its use in many electrical contact applications. Although tin and silver represent low-cost alternatives to gold, the susceptibility of tin-plated contacts to fretting corrosion is often considered a major obstacle to the use of tin in many types of connectors. Because tin and silver are important connector plating materials, the factors that affect fretting corrosion will be described, with emphasis on tin- and silver-plated connections.

Factors Affecting Contact Degradation in Fretting

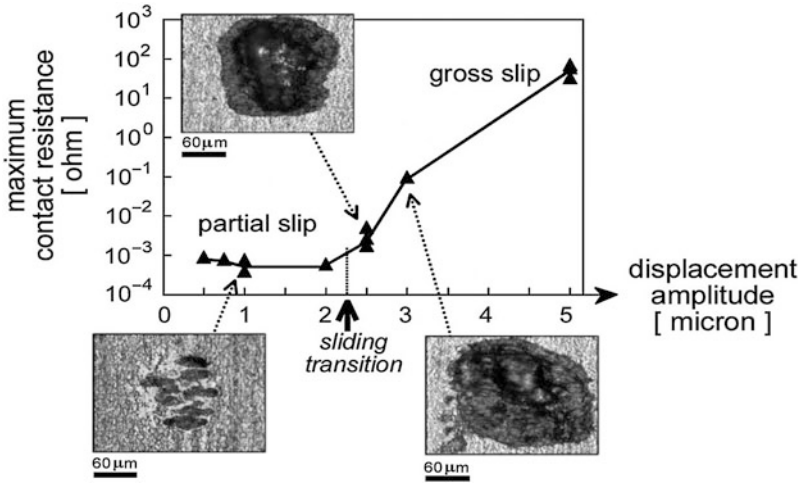
A number of factors, such as the amplitude and frequency of interfacial displacements, temperature, humidity and contact load, influence fretting wear and fretting corrosion rates. Because mechanical contact at a solid-solid interface is made at contacting surface asperities (Holm 1976; Timsit 1999), electrical current affects fretting rates due to the effect of Joule heating on the temperature of contact spots. Investigations of fretting corrosion are usually carried out using an apparatus of the type shown schematically in Fig. 2, in which a contacting slider or dimple is passed over a flat surface in reciprocating motion.



Wear Mechanisms in Electrical Contacts: Fretting Wear, Fig. 2 Schematic diagram of typical apparatus for studying the effects of fretting on electrical contact resistance

Effects of Reciprocating Displacement Amplitude on Fretting Corrosion

There are three broadly defined sliding regimes that are defined by the magnitude of the fretting displacement amplitude (Vingsbo and Soderberg 1988). These regimes correspond, respectively, to (1) *sticking*, whereby the sliding motion is very small; (2) *partial slip*, whereby the sliding surfaces are displaced by a few to several microns and wear damage is relatively small; and (3) *gross slip*, whereby large-scale sliding occurs and fretting damage is the largest (Hannel et al. 2001). Figure 3 illustrates the effects on wear and contact resistance of fretting corrosion in the last two regimes in sliding surfaces of a copper-tin alloy (copper-4 wt% tin or CuSn4), after 10,000 fretting cycles. The data indicate that mechanical damage and electrical contact resistance remain small under conditions



Wear Mechanisms in Electrical Contacts: Fretting Wear, Fig. 3 Evolution of the electrical contact resistance and mechanical wear as a function of the displacement amplitude between surfaces of a copper-tin alloy coated with hot-dipped tin after 10,000 reciprocating cycles at a contact load of 5 N (Hannel et al. 2001)

of *partial slip*, where the maximum displacement remained smaller than about 2 µm but increased rapidly in the *gross slip* regime.

Effects of Reciprocating Motion Frequency on Fretting Corrosion

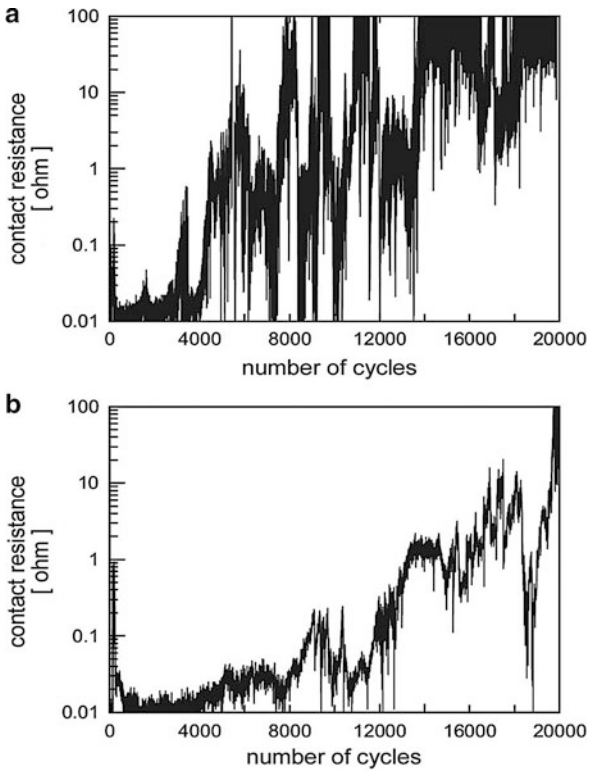
Generally, the rate of contact degradation due to fretting corrosion increases with increasing reciprocation frequency, so that the "time-to-failure" decreases as this frequency increases. This relatively rapid rate of contact degradation is attributed to a larger rate of generation of particulate material at higher reciprocation frequencies. This in turn leads to a quicker accumulation of particulate matter in the contact interface, and to a faster rise of contact resistance. Examples of contact resistance increase with fretting cycles in tin-plated contacts are illustrated in Fig. 4. Although the rise of contact resistance per fretting cycle is clearly slower at the higher frequency in this case, the actual "time-to-failure" is found to decrease with increasing cycling frequency as shown in Fig. 5 (Park et al. 2008).

Effects of Temperature on Fretting Corrosion

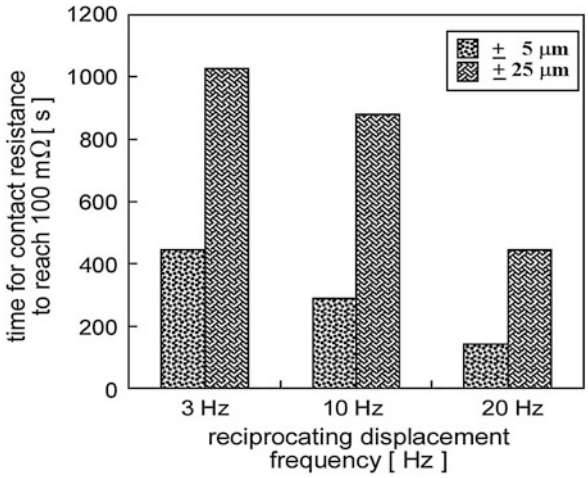
The effects of temperature on fretting corrosion are convoluted. On the one hand, a rise in contact temperature increases surface oxidation rates and therefore increases the probability of oxidized debris formation in a sliding interface. On the other hand, increasing the temperature leads to metal softening and to an ensuing decrease in contact resistance in regions of metal-metal contact. The relative weights of these two (and other) effects can thus lead to an increase or a decrease in contact resistance with continued reciprocation motion at elevated temperatures.

Typical effects of temperature on fretting corrosion in tin-tin sliding interfaces (tin plated on copper on each surface) are summarized in Fig. 6, which shows the time to failure as a function of temperature (Park et al. 2008). The clearly defined increase in the time-to-failure (decrease in fretting corrosion rate) with increase in temperature up to about 85 °C is attributed to softening of tin and to the ensuing increase in metal-metal contact through fractured surface oxide layers. Beyond 85 °C, the degradation rate of the electrical contact increases rapidly and is attributed in large part to the larger oxidation rate of tin at these higher temperatures and to the greater rate of accumulation of debris and corrosion products in the sliding interface.

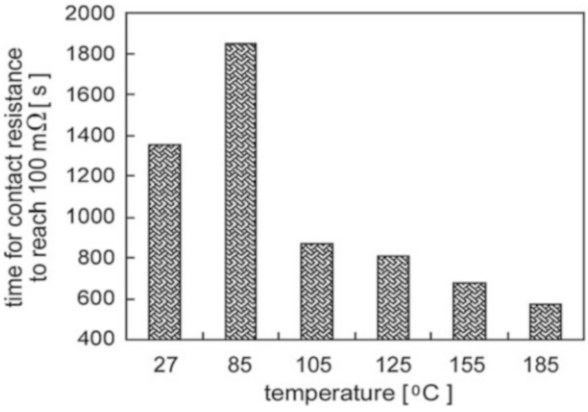
Beside increased oxidation, the rapid formation of layers of Cu-Sn intermetallic compounds (IMCs) at elevated temperatures in tin-plated copper-base contacts has an influence on the performance of the contacts (Hammam 2000). IMC layer formation is undesirable because of the high electrical resistivity and brittleness of the layers. Increased oxidation and increased IMC formation rates at temperatures exceeding about 80°C, suggest that tin-plated copper-base contacts are unsuitable for applications at these temperatures. The introduction of a diffusion barrier such as nickel between the tin and the copper-base substrate mitigates Cu-Sn IMC formation but does not influence tin oxidation rates.



Wear Mechanisms in Electrical Contacts: Fretting Wear, Fig. 4 Change in contact resistance of tin-plated copper contacts as a function of fretting cycles obtained at different frequencies (**a**) 3 Hz and (**b**) 20 Hz. (Tin layer thickness : 3 μm, amplitude: 25 μm, temperature: 22 °C, humidity: 55 % RH, normal load: 0.5 N, current: 0.1 A.) (Park et al. 2008)



Wear Mechanisms in Electrical Contacts: Fretting Wear, Fig. 5 Time required for the contact resistance to reach a threshold value of 100 mΩ for track lengths of 5 and 25 μm at fretting frequencies of 3, 10 and 20 Hz. (Temperature: 22 °C, humidity: 55 % RH, normal load: 0.5 N, current: 0.1 A.) (Park et al. 2008)



Wear Mechanisms in Electrical Contacts: Fretting Wear, Fig. 6 Time required for the contact resistance to reach a threshold value of 100 mΩ at various temperatures. (Displacement amplitude: ±90 μm; frequency: 10 Hz; humidity: 55 % RH; contact load: 0.5 N and current load: 0.5 A.) (Park et al. 2008)

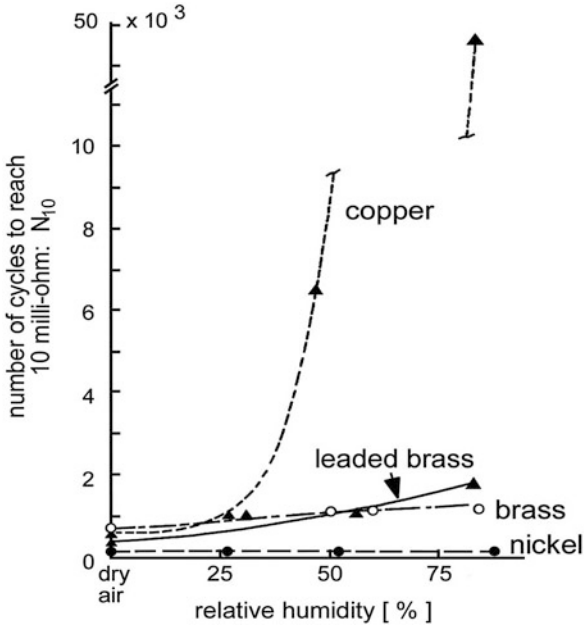
Effects of Humidity in Fretting Corrosion

The presence of moisture has a significant but complicated effect on the fretting behavior of electrical contacts. This stems from the chemical reactivity of many metals with water to form oxides and hydroxides and hence to form electrically insulating particle debris in a sliding interface. On the other hand, moisture films can also act to lubricate and reduce wear. The effects of humidity on the fretting behavior of tin-plated contacts are illustrated from investigation in the range of dry air to 80 % RH by Bruel et al. (1988).

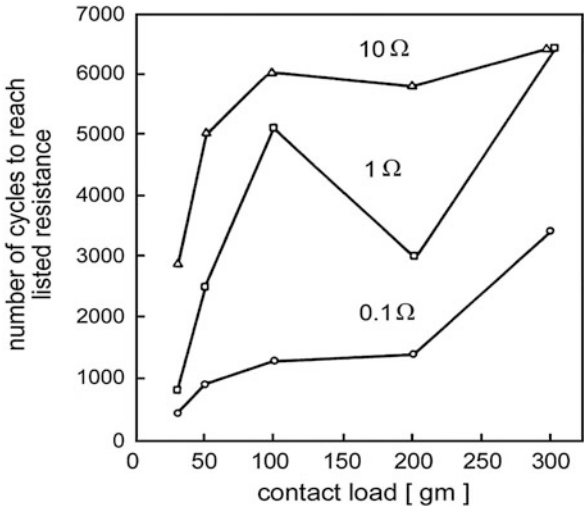
Figure 7 shows the number of reciprocating cycles required to reach a contact resistance of 10 mΩ as a function of RH, for a number of electrical contact materials (Bruel et al. 1988). Nickel was the worst performer in these tests and copper was the best.

Effects of Contact Force in Fretting Corrosion

The rate of contact degradation due to fretting corrosion generally decreases with increasing contact force. This trend is illustrated in Fig. 8 showing the number of fretting cycles to reach a selected value of contact resistance for copper surfaces plated with tin-lead (Antler 1985).



Wear Mechanisms in Electrical Contacts: Fretting Wear, Fig. 7 Number of fretting cycles N_{10} to reach the contact resistance threshold value of 10 mΩ at various values of relative humidity. (Displacement amplitude: 20 μm; frequency: 50 Hz; contact load: 0.3 N) (Bruel et al. 1988)



Wear Mechanisms in Electrical Contacts: Fretting Wear, Fig. 8 Number of fretting cycles for 60Sn-40Pb versus 60Sn-40Pb contacts required to attain a contact resistance of 0.1, 1, and 10 Ω, as a function of the contact load; the wipe amplitude was 20 μm (Antler 1985)

This trend stems in part from increased frictional drag due to a larger normal force, which leads to reduced interfacial motion since some relative displacement is now accommodated by elastic deformation and thus leads to reduced fretting. The decrease in contact degradation with a larger contact force also arises from the increased area of true contact (not just the metallic contact area) generated at larger contact loads. A larger area of mechanical contact subjected to interfacial displacement is more effective in cracking surface insulating layers, and is also less affected by the entrapment of electrically insulating debris than a relatively small contact area at smaller loads.

Effects of Electrical Current in Fretting Corrosion

The effects of electrical current load on fretting corrosion are closely associated with effects due to temperature since the temperature of contact spots is determined by the voltage drop across the contact (Holm 1976; Timsit 1999), and hence by the current. For purposes of illustration, the sequence of events in contact spots as current is increased across an electrical interface is considered.

For a contact initially at room temperature, an increase in current leading to a voltage drop of about 40 mV leads to an *a*-spot temperature of about 50 °C according to the universal temperature-voltage curve (Holm 1976; Timsit 1999). Such a voltage drop across the contact will have a negligible effect on contact resistance for many contact materials but may lead to a slight decrease in contact resistance in interfaces involving metals such as tin and indium since the contact spots in such interfaces may deform relatively rapidly near 50 °C, thus increasing the area of true contact. However, because the oxidation rate of non-noble metals increases rapidly with increasing temperature, the imposition of a steady voltage drop of 40 mV across non-noble metal contacts activates degradation due to oxidation. Under these conditions, the dominance of the creep mechanism over oxidation will lead to acceptable contact performance whereas dominance by oxidation will eventually cause catastrophic contact failure. In practice, the dominance of one mechanism over the other will be determined by external factors such as contact force, fretting displacements, and others.

As current is increased and the voltage drop across a contact exceeds 40 mV, the temperature of *a*-spots quickly approaches the softening point of the materials in contact. Contact asperities of most materials reach their softening temperature at a voltage drop of between 70 and 110 mV, corresponding to contact-spot temperature ranging from about 90 °C to 150 °C (Timsit 1999). In the

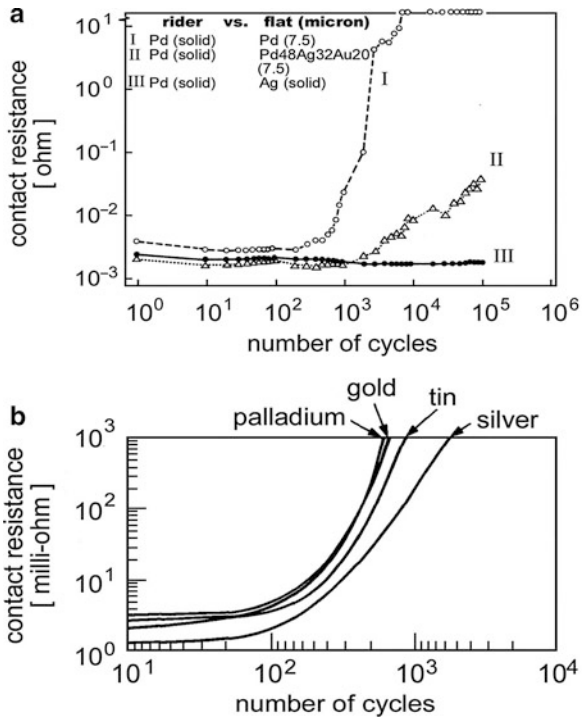
vicinity of the softening temperature, the contact asperities of metals flow rapidly and lead to an increase in area of true contact. However, the nefarious effects of thermally activated mechanisms including oxidation, interdiffusion, galvanic corrosion, and so on are also enhanced. In contacts involving noble or near-noble metals such as gold and platinum, the imposition of a voltage drop corresponding to or exceeding the softening temperature will cause rapid interdiffusion of the noble metal either with the underplate material or with alloying elements (Timsit 1999). For example, the generation of high temperatures in hard gold causes migration to the gold surface of hardeners such as nickel and cobalt (Tompkins and Pinnel 1977). This is deleterious to contact stability since nickel and cobalt oxidize readily and may render the gold surface highly susceptible to fretting corrosion.

As current is further increased to generate a voltage drop exceeding 110 mV, the temperature of *a*-spots approaches the melting point of the materials in contact. Contact asperities of most materials reach their melting temperature at a voltage drop of between 130 mV (tin) and 430 mV (copper and gold) (Timsit 1999). The generation of such large voltage drops in stationary or separable contacts leads to catastrophic failure.

Effects of Surface Finish, Hardness, and Contact Geometry in Fretting Corrosion

Fretting wear may result in penetration of a contact material having small thickness, as is often the case with electrodeposited and clad finishes. In these situations, contact resistance may be affected deleteriously, particularly if the underlying substrate is susceptible to fretting corrosion or frictional polymerization. As mentioned earlier, the durability of a thin coating in sliding over long tracks is affected by the composite hardness of the sliding surfaces (Antler and Drozdowicz 1981–1982). Similar considerations apply to fretting wear.

Fretting wear leads to material transfer across the sliding surfaces. The benefit or drawback of material transfer in reciprocating sliding is illustrated in Fig. 9. The three curves in Fig. 9a show the evolution of contact resistance between solid palladium and various substrates (Antler 1999). Curve I corresponds to solid palladium sliding over a thick palladium-clad on a flat copper strip. Curves II and III correspond, respectively, to solid palladium sliding over the contact alloy Pd48Ag32Au20 and over solid silver. Palladium-palladium contacts are the classic frictional-polymer-forming system (Antler 1999), and the rise of Curve I as discussed earlier, and the rise of Curve I was confirmed to be due to frictional polymers. In contrast, the stability of contact resistance depicted by



Wear Mechanisms in Electrical Contacts: Fretting Wear, Fig. 9 (**a**) Variations of contact resistance with fretting cycles for solid palladium riders sliding over various metals at a contact load of 0.5 N with a 20 μm wipe: Curve I (Pd-Pd), Curve II (Pd-Pd48Ag32Au20), Curve III (Pd-Ag) (Antler 1982). (**b**) Contact resistance with fretting cycles for tin sliding, respectively, on tin, palladium, gold, and silver at a contact load of ~1 N with a 50 μm wipe (Bock 1990)

Curve III stemmed from the transfer of silver to the palladium rider during fretting, thus generating a silver-silver interface not prone to frictional polymer formation. The relatively slow rise of Curve II indicates that the fretting interface was becoming richer in palladium as fretting proceeded, thus promoting the formation of frictional polymers. This suggests that transfer of palladium from the solid palladium rider to the Pd48Ag32Au20 alloy surface proceeded faster than back transfer of silver or gold from the alloy. Similar effects are observed in interfaces involving a soft material sliding over a harder surface. This is illustrated in Fig. 9b, which shows the effects of fretting in interfaces consisting of tin-plated surfaces sliding over harder materials. In all cases, tin is transferred to the mating surface and the effects of fretting are indistinguishable from those in tin-tin contacts (Bock 1990).

Effects of Lubrication on Fretting Corrosion

It has long been known that specific lubricants delay or mitigate the deleterious effects of fretting corrosion in electrical contacts involving base metals (Bock and Whitley 1974). A variety of lubricants, namely, polyphenyl ethers, natural and synthetic hydrocarbons, several types of esters, polyglycols, some fluorinated materials, silicone, and some proprietary formulations have been evaluated for their compatibility with specific connector applications (Freitag 1975; Antler 1984). The ability of lubricants to mitigate frictional polymerization usually depends strongly on the lubricant concentration on surfaces in contact (Antler 1999).

Cross-References

- ► Electrical Contact Materials
- ► Friction
- ► Friction Coefficient
- ► Sliding Electrical Contact Wear
- ► Tribology

References

M. Antler, Fretting of electrical contacts: an investigation of palladium mated to other materials. Wear **81**, 159–173 (1982)

M. Antler, Effect of fretting on the contact resistance of palladium electroplate having a gold flash, cobalt-gold electroplate, and DG R-156. IEEE Trans. Compon. Hybrid. Manuf. Technol. **CHMT-7**, 363–369 (1984)

M. Antler, Survey of contact fretting in electrical connectors. IEEE Trans. Compon. Hybrid. Manuf. Technol. **CHMT-8**, 87–104 (1985)

M. Antler, Tribology of electronic connectors: contact sliding wear, fretting and lubrication, in *Electric Contacts: Theory and Applications*, ed. by P.G. Slade (Marcel Dekker, New York, 1999), pp. 309–402

M. Antler, M. H. Drozdowicz, Fretting corrosion of gold plated connector contacts. Wear, **74**, 27–50 (1981–1982)

E.M. Bock, J.H. Whitley, Fretting corrosion in electric contacts, in *Proceedings of 20th Holm Conference on Electrical Contacts*, Chicago, pp. 128–138 (1974)

E.M. Bock, J.H. Whitley, Fretting corrosion in electric contacts, in *Proceedings of 40th Electronic Components and Technology Conference*, Las Vegas, pp. 840–844 (1990)

J.F. Bruel, P. Smirou, A. Carballeira, Gas environment effect on the fretting corrosion behaviour of contact materials, in *Proceedings of the 14th International Conference on Electrical Contacts*, Paris, pp. 219–223 (1988)

W.O. Freitag, Wear, fretting and the role of lubricants in edge card connectors, in *Proceedings of the 21st Holm Seminar on Electrical Contacts* (Illinois Institute of Technology, Chicago, 1975), pp. 17–23

T. Hammam, The impact of sliding motion and current load on the deterioration of tin-coated contact terminals. IEEE Trans. Compon. Packag. Technol. **23**, 278–285 (2000)

S. Hannel, S. Fouvry, P. Kapsa, L. Vincent, The fretting sliding transition as a criterion for electrical contact performance. Wear **249**, 761–770 (2001)

R. Holm, *Electric Contacts, Theory and Applications* (Springer, Berlin, 1976)

P.L. Hurricks, The mechanism of fretting – a review. Wear **15**, 389–409 (1970)

Y.W. Park, T.S.N. Sankara Narayanan, K.Y. Lee, Fretting corrosion of tin-plated contacts: evaluation of surface characteristics. Tribol. Int. **40**, 548–559 (2007)

Y.W. Park, T.S.N. Sankara Narayanan, K.Y. Lee, Fretting corrosion of tin-plated contacts. Tribol. Int. **41**, 616–628 (2008)

R.S. Timsit, Electrical contact resistance: fundamental principles, in *Electric Contacts: Theory and Applications*, ed. by P.G. Slade (Marcel Dekker, New York, 1999), pp. 1–88

H.G. Tompkins, M.R. Pinnel, Relative rates of nickel diffusion and copper diffusion through gold. J. Appl. Phys. **48**, 3144–3146 (1977)

O. Vingsbo, S. Soderberg, On fretting maps. Wear **126**, 131–147 (1988)

Wear Mechanisms in Small Electrical Contacts

Roland S. Timsit
Timron Advanced Connector Technologies,
A Subsidiary of Timron Scientific Consulting Inc.,
Toronto, ON, Canada

Definition

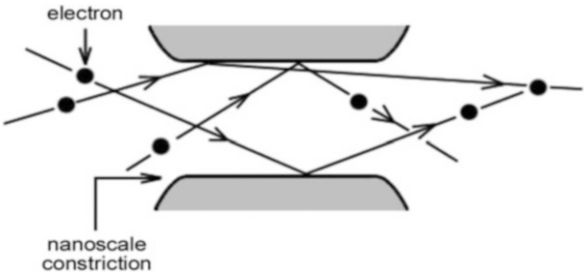
Small electrical contacts are defined as electrical interfaces where the contact force is measured in hundreds of micro-Newtons or less, as in MEMS (Micro Electro-Mechanical Systems) switches and stationary or separable nano-scale electronic devices. In small contacts, electrical contact spots in the contact interface are significantly less than 1 μm.

Scientific Fundamentals

One of the major challenges to the use of a small contact force in an electrical contact in air is the deleterious effects of contaminant surface films, and hence the capability of small contact forces to penetrate through these films. Another challenge associated with small contact forces arises from the relative importance of surface adhesion between metallic microcontacts, as described later. Surface adhesion is often sufficiently large to prevent the opening of a contact.

Contact Resistance in Small Contacts

In small contacts, electrons diffuse ballistically in the constriction, as illustrated schematically in Fig. 1, without losing energy, as proposed by Sharvin (1965). Ballistic conduction occurs because the average constriction opening is smaller than the electronic collision mean-free-path.



Wear Mechanisms in Small Electrical Contacts, Fig. 1 Ballistic motion of electrons through a small horizontal constriction

Wear Mechanisms in Small Electrical Contacts, Table 1 Electrical resistivity and electronic mean free path

Metal	Electrical resistivity [Ω m]	Electronic mean free path [nm]	$C \times 10^{16}$ [Ω m^2]
Al	2.74×10^{-8}	14.5	1.70
Cu	1.70×10^{-8}	38.7	2.79
Ag	1.61×10^{-8}	52.3	3.57
In	8.75×10^{-8}	6.1	2.27
Sn	11.0×10^{-8}	4.1	1.91
Au	2.20×10^{-8}	38.3	3.58

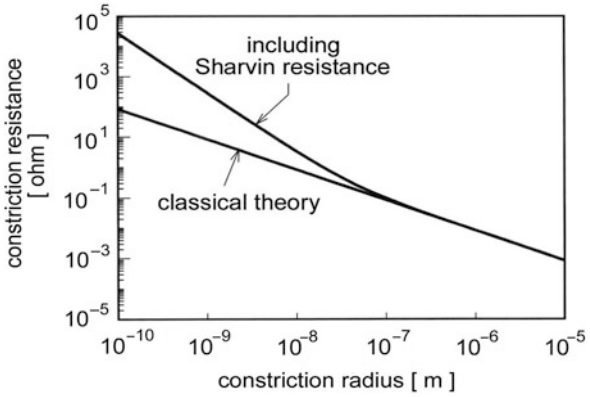
When ballistic conduction prevails, contact resistance is given as Timsit (2004)

$$R_C = \Gamma(K)\frac{\rho}{2a} + \frac{C}{a^2} \qquad (1)$$

where a is the a-spot radius, ρ is the average resistivity of the contacting bodies, and $K = l/a$, where l is the electronic mean free path listed in Table 1; C is also listed in Table 1 for various contact materials and $\Gamma(K)$ is ~0.7 (Timsit 2004). For large contact spots where $l << a$, $\Gamma(K)$ is ~1 and the term C/a^2 is negligible, (1) reduces to the conventional classical relation

$$R_C = \frac{\rho}{2a} \qquad (2)$$

Figure 2 shows the variation of constriction resistance of a copper-copper contact consisting of a single spot at room temperature, as a function of the radius a (Timsit 2004). The two curves correspond, respectively, to the classical constriction resistance evaluated from (1), using the values of ρ and C listed in Table 1. Note that the effect of the Sharvin resistance, giving rise to a larger constriction resistance, begins to be significant for a constriction radius smaller than about 100 nm.



Wear Mechanisms in Small Electrical Contacts, Fig. 2 Constriction resistance for a copper-copper contact at room temperature, as a function of the constriction radius, as calculated from classical contact theory Eq. (2) and from Eq. (1)

Origin of Adhesion in Clean Surfaces

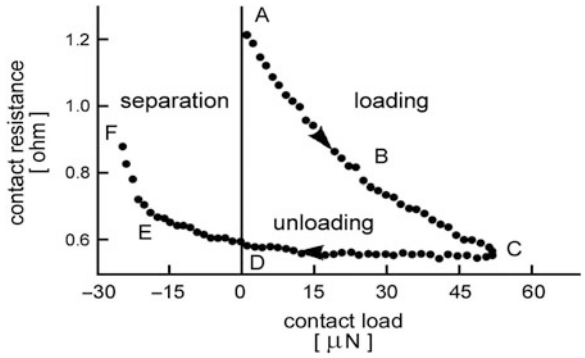
Interfacial surface energy is the mechanical work required to separate two surfaces, and the degree of adhesion is generally characterized by the force required for the separation. Johnson et al. (1971) showed that the pull-off force (adhesion) for a sphere of radius R on an elastic flat surface is approximately $2\pi R\Delta\gamma$, where the surface energy of adhesion $\Delta\gamma$ is given as

$$\Delta\gamma = \gamma_1 + \gamma_2 - \gamma_{12} \qquad (3)$$

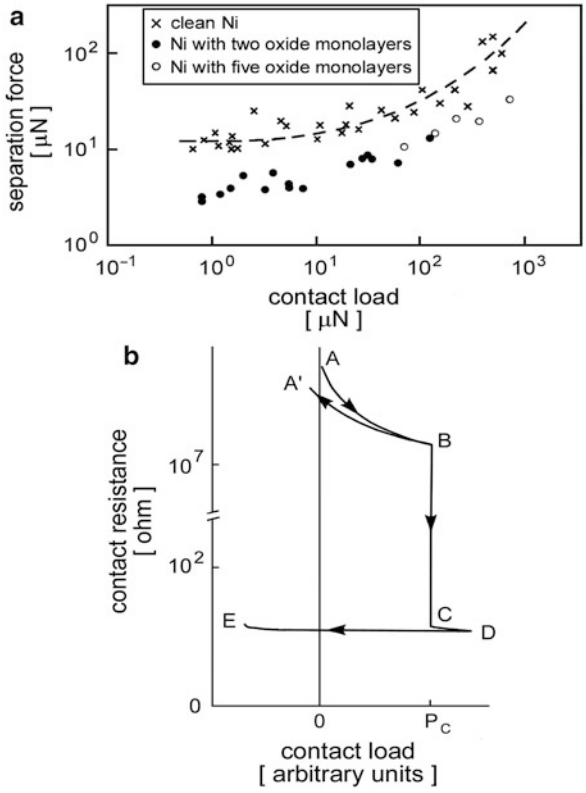
where γ_1 and γ_2 are the surface energies of the two respective surfaces and γ_{12} is the energy of the contact interface. In ductile fracture of the contact, it has been suggested that the "pull-off force" of a circular junction of radius a should be given as $H\pi a^2$, where H is the Vickers' hardness of the softer in the junction (Johnson 1998).

Most reported measurements of adhesion force between contaminant-free metal surfaces have been undertaken under ultra-high vacuum conditions. The effects of organic contamination, or other surface films such as oxide layers, on adhesion force have been determined by deliberately introducing contaminants, and depositing or growing films of selected materials on the surfaces. Examples of adhesion force due to surface energy effects, and the subsequent influence on contact resistance, are shown in Figs. 3 and 4 (Pashley et al. 1984; Pashley and Pethica 1985).

In these investigations, a tungsten indenter was pressed against a flat nickel surface in an ultra-high vacuum environment where the contact surfaces had first been atomically cleaned. The adhesion force and contact resistance were subsequently measured. The measurements were



Wear Mechanisms in Small Electrical Contacts, Fig. 3 Effect of contact force on adhesion force in clean tungsten-nickel contacts (Pashley and Pethica 1985)



Wear Mechanisms in Small Electrical Contacts, Fig. 4 (**a**) Effect of contact force on adhesion force in contaminated tungsten-nickel contacts, (**b**) Dependence of contact resistance on contact load for a clean tungsten indenter in contact with a nickel (111) single crystal surface covered with nickel oxide to a thickness of 5 nm (Pashley et al. 1984)

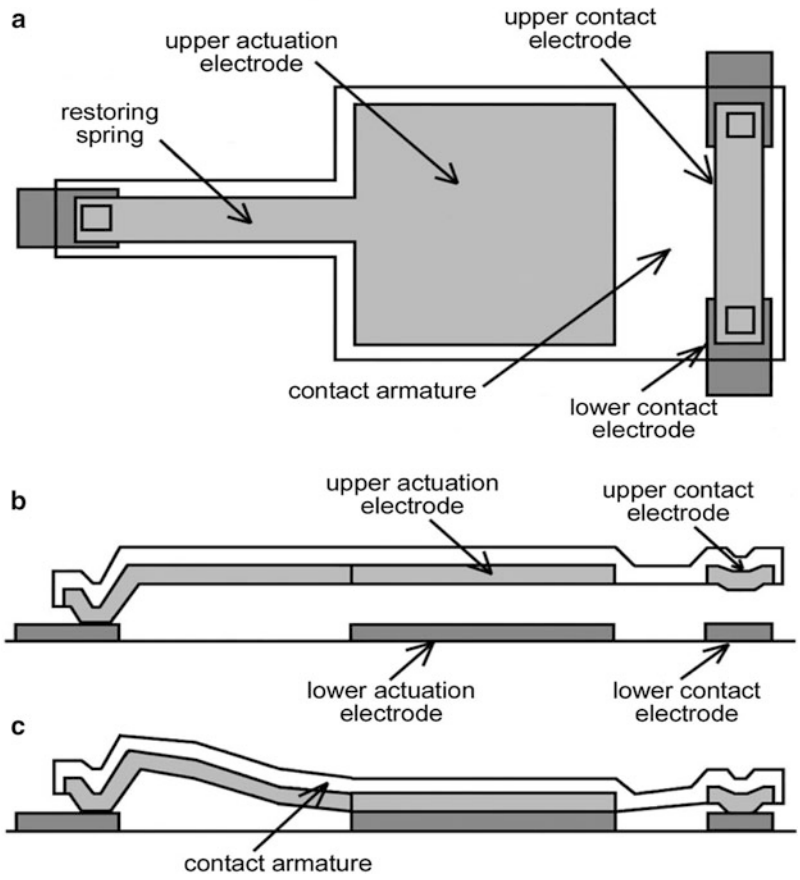
then repeated after the surfaces were oxidized. Figure 3 shows that the contact resistance, and hence the area of true metallic contact, has a finite value at zero load (point A). The observation of a finite contact resistance at zero applied load stems from the action of the adhesion force. Subsequently, the contact resistance then falls as the contact load is increased (curve ABC). On reducing the applied load, the contact resistance (and hence the area of metal-metal contact) remained relatively constant (CD) until the contact load was further decreased to generate a negative force, that is, a tensile load (DEF). As the tensile load was now increased further, the nickel surface necked, and the contact surfaces eventually separated at a tensile load of about 25 μN. Forces of this magnitude or larger are typical of adhesion forces measured between metal surfaces placed under contact loads in MEMS. Such adhesion forces represent a serious concern to separating a contact in a MEMS device.

For the contact couples associated with the data of Fig. 3, contamination of the surfaces had a relatively small mitigating effect on adhesion. Figure 4a shows the effects of the presence of thin oxide layers on the adhesion force (Pashley and Pethica 1985). The presence of 2–5 monolayers of oxide on the contacting surfaces had the effect of reducing the adhesion force by a factor of 2 to 3 over the entire range of contact forces investigated, but led to a large increase in contact resistance, by a factor of about 10^8 over the values measured under clean surface conditions. The effect on contact resistance is shown in Fig. 4b. For these oxidized surfaces, the loading-unloading cycle was fairly reproducible as indicated by the path ABA', with the resistance at A corresponding nearly to an open circuit. The loading-unloading cycle remained reproducible provided that the contact load did not exceed a critical value P_C of approximately 500 μN. Beyond the critical load P_C, the contact resistance dropped precipitously by several orders of magnitude and the loading-unloading cycle (BCDE) was largely irreproducible. The large drop BC in contact resistance stemmed from the formation of cracks in the oxide layers and the micro-extrusion of underlying metal through the fissures to generate metal-metal contact spots and reduce contact resistance.

MEMS

General Requirement

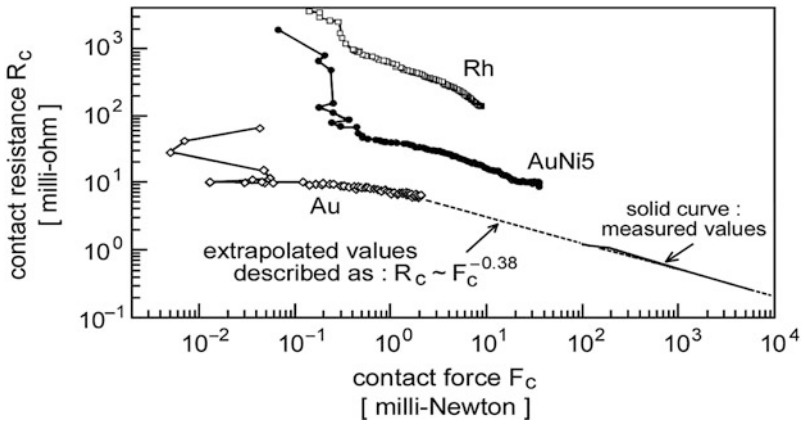
Figure 5 shows the schematic diagram of a typical micromachined MEMS switch, with side views of the device in an open and closed configuration (Larson et al. 1991; Hyman and Mehregany 1999). These devices are



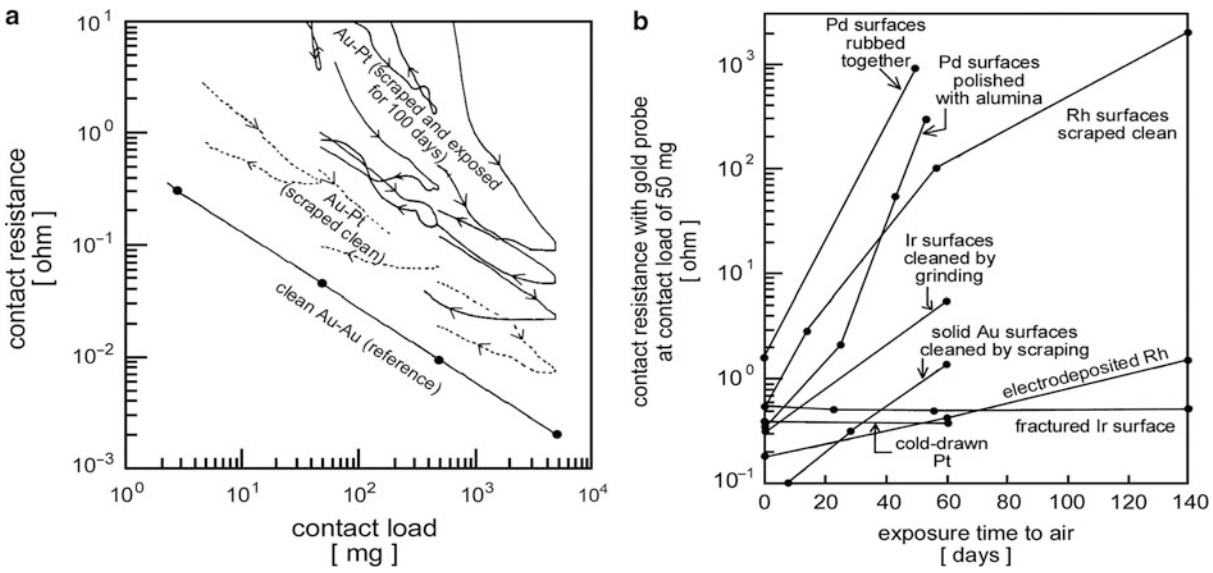
Wear Mechanisms in Small Electrical Contacts, Fig. 5 Schematic illustrations of a typical MEMS switch configuration: (**a**) plan view, (**b**) side view of an open switch, (**c**) side view of a closed switch. The contact force is generated by bending of the armature at the free end due to the dimple height. The typical film thickness is 1–2 μm and typical gaps are 2 μm (Hyman and Mehregany 1999)

often actuated through electrostatic attraction of the armature to substrate electrodes. The typical length of these devices is several hundred micrometers, with a width of about 100 μm, and a thickness of only a few micrometers. Because of the small dimensions and thinness of cantilever springs, the contact force in MEMS ranges typically from tens to hundreds of micro-Newtons. Typically, MEMS structures use gold or a gold alloy on the contact components to take advantage of the high electrical conductivity and chemical inertness of gold to achieve and maintain a low contact resistance.

One of the major challenges to the use of a small contact force in an electrical contact in air is the deleterious effects of contaminant surface films. For example, contact resistance for gold surfaces in air has been reported to vary between 80 and 200 mΩ at 100 μN and to drop to 10–20 mΩ at 1 mN for relatively uncontaminated surfaces (Schimkat 1998; Hyman and Mehregany 1999). Investigations of contact resistance at small contact loads also show that the increase in contact resistance with decreasing contact force deviates appreciably from the $F^{-1/2}$ dependence predicted by classical theory (Timsit 1999) for loads smaller than a few milli-Newtons. This deviation is illustrated in Fig. 6 and stems in large part from the presence of thin residual contaminant layers that insulate electrically some fraction of the regions of contact. The effects of contaminant layers is exacerbated in gold alloys such as Au95Ni5 since any nickel present on the surface oxidizes and further obstructs electrical flow in the contact interface, as illustrated in Fig. 6 and in Fig. 7a for several precious-metal surfaces prepared in a variety of ways. The degree of surface contamination, and the concomitant deleterious effect on contact resistance at low contact loads, increases with increased



Wear Mechanisms in Small Electrical Contacts, Fig. 6 Examples of variations of contact resistance with applied force in Au-Au, Au95Ni5 and Rh-Rh contacts (Schimkat 1998)



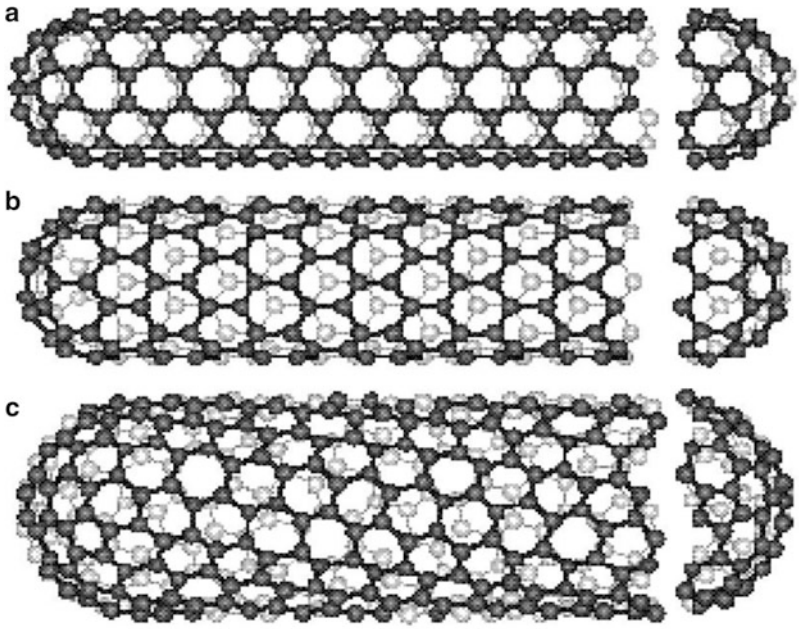
Wear Mechanisms in Small Electrical Contacts, Fig. 7 (**a**) Additional examples of variations of contact resistance with applied force in Au-Au and Rh-Rh contacts, (**b**) Illustrative examples of the deterioration of selected precious-metal surfaces with time (Angus 1962)

exposure to the environment. The data of Fig. 7b illustrate the rise in contact resistance with increased exposure to laboratory air for selected precious-metal contact materials (Angus 1962).

Another challenge associated with small contact forces arises from the relative importance of surface adhesion (also labeled "stiction") between metallic microcontacts as described above and as illustrated in Figs. 3 and 4.

Challenges with Practical MEMS: Exploration of Contact Materials

Although gold is a preferred contact material, the softness of pure gold and its proneness to adhesion constitute drawbacks to its widespread use in MEMS. Adhesion stems both from the action of surfaces forces as described earlier, and from localized melting and subsequent freezing if the voltage across the contact is too high. Typical release forces needed to separate microswitch contacts are



Wear Mechanisms in Small Electrical Contacts, Fig. 8 Examples of SWNTs with different chiralities. The difference in structure is evident at the open end of the tubes: (**a**) armchair structure, (**b**) zigzag structure, (**c**) chiral structure

100–2,700 μN for gold. Despite the challenge of large release forces due to stiction, gold-contact switches have been shown to achieve $>10^9$ switching cycles under conditions of low current, optimum contact force, dry nitrogen environment, and "cold switching" before running into severe mechanical wear. "Cold switching" describes an operation with a small or zero electric field across the contacts as the switch opens or closes.

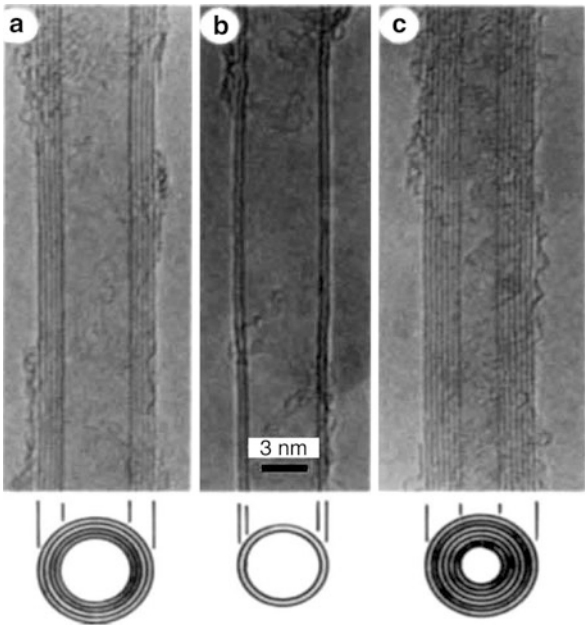
Because there is a need to broaden the range of operating conditions for MEMS switches, a number of investigations of MEMS-compatible contact materials, particularly gold alloys, have been undertaken in recent years. Recent investigations of selected properties and performance of thin film contact materials for MEMS are given by Hosaka et al. (1993), Schimkat (1998), Coutu et al. (2004, 2006), Chen et al. (2007), and Yang et al. (2009).

Carbon Nanotubes (CNTs) in MEMS

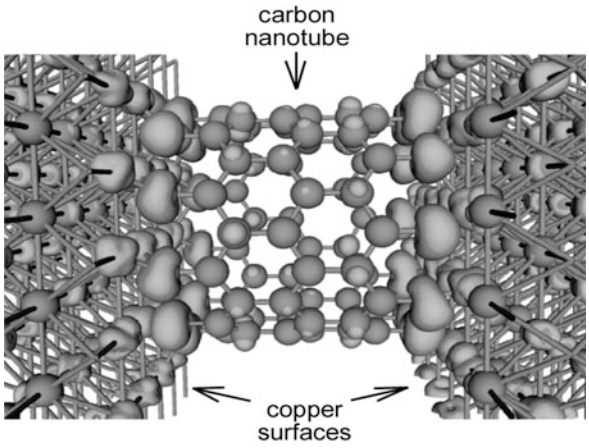
Solids of elemental carbon in the sp^2 bonding state can form a variety of graphitic structures, including graphite filaments and graphitic cylindrical needles (i.e., carbon nanotubes (CNTs)) when these solids are prepared under specified conditions (Iijima 1991). CNTs may consist of single-wall nanotubes (SWNTs) or multiple-wall nanotubes (MWNTs) ranging in diameters from a few nanometers to several tens of nanometers, but with a length of up to hundreds of micrometers. Figure 8 shows examples of well-known structures of CNTs with different chiralities. The difference in structure is evident at the open end of the tubes, where Fig. 8a, b, and c illustrate, respectively, the "armchair structure," the "zigzag structure," and the "chiral structure." Carbon atoms placed in hexagons and pentagons form the end cap structures.

Because nanotubes generally have a length to diameter ratio of 1,000 or larger, they can be considered as one-dimensional structures. A SWNT consists basically of a long wrapped graphene sheet, and an MWST is made up of a collection of concentric SWNTs with different diameters. The length and diameter of MWSTs may differ significantly from those of SWNTs and their properties may also very different. Electron micrographs of SWNT and MWNT structures are illustrated in Fig. 9.

CNTs are characterized by attractive mechanical properties such as a tensile strength of 63 GPa (versus 1.2 GPa for high strength steel) (Yu et al. 2000), an elastic modulus of 1.26 TPa (Wong et al. 1997), and a Vickers hardness probably on the order of that of diamond (~60 GPa) (Kinoshita et al. 2007). It is estimated that a 4–10 μm long single-walled CNT with a diameter of 1.2 nm and free of structural defects should have a resistivity of 0.88×10^{-8} Ωm (Hjortstam et al. 2004). The excellent mechanical properties and potentially outstanding



Wear Mechanisms in Small Electrical Contacts, Fig. 9 Electron micrographs of NCTs. The *parallel dark lines* correspond to lattice images of graphite. A cross-section of each NCT is illustrated as follows: (**a**) five graphitic tubes with an outer diameter of 6.7 nm, (**b**) two tubes with an outer diameter of 5.5 nm, (**c**) seven tubes with a diameter of 6.5 nm, where the smallest tube has a diameter of 2.2 nm (Iijima 1991)



Wear Mechanisms in Small Electrical Contacts, Fig. 10 Atomic structure of a junction consisting of two copper surfaces in contact with the open ends of a CNT (Gao et al. 2010)

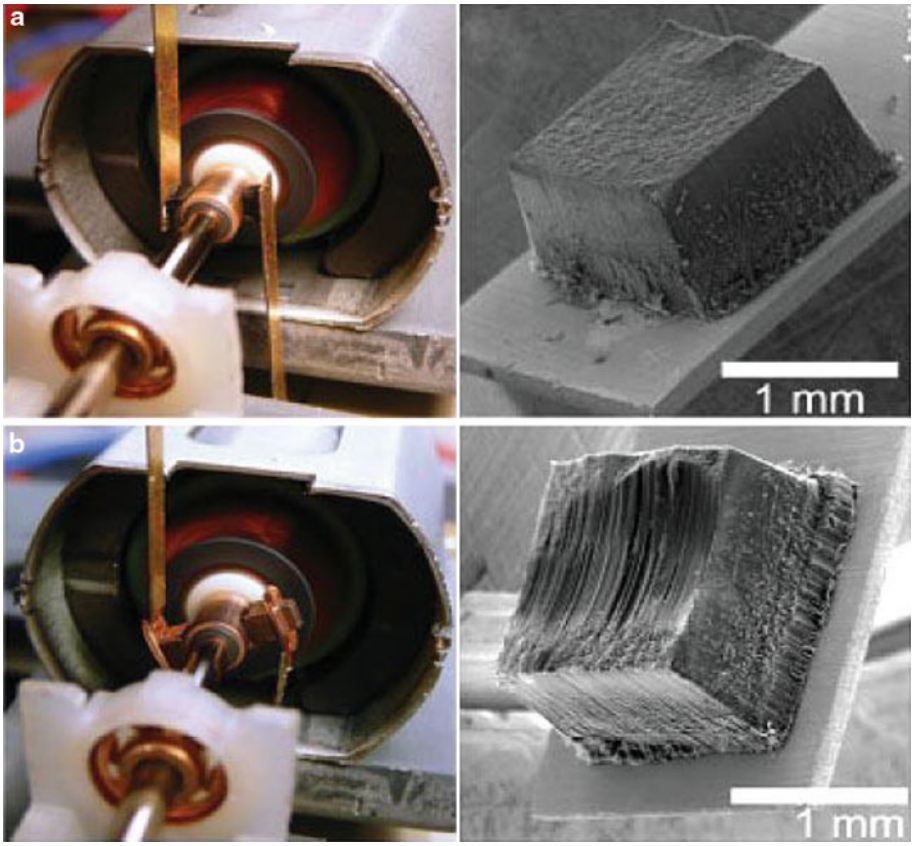
electrical conductivity are obviously highly attractive in terms of the use of CNTs as alternatives to precious metals in electrical contacts in MEMS micro-switches and similar devices. However, measurements have led so far to widely varying values of contact resistance between CNT surfaces and to a wide range of electrical resistivity values for CNTs, depending on the structure and content of the CNT-bearing surface. For example, Yaglioglu et al. (2006) reported a resistivity ranging from about 1×10^{-4} to 1.8×10^{-4} Ωm for tangled SWNTs deposited on gold, and a contact resistance of ~50 Ω for SWNT–SWNT contacts under a mechanical load of ~1 N. This is much larger than the contact resistance of tens of milliohms measured between precious-metal contacts in MEMS. The large resistivity and the high contact resistance were attributed in part to the lack of alignment of the SWNT film.

Over the foreseeable future, CNTs remain highly promising candidates for nanoscale electronic devices owing to their structural strength and excellent thermal and electronic properties. For electronic nanoscale devices, one of the key issues is often the permanent mechanical attachment of the CNT to the metallic electrodes of the device (Bourgoin et al. 2010). In actual devices, SWCNTs must be secured mechanically and electrically to metal electrode surfaces with controllable metal-SWCNT contact resistance. It has been established that the levels of crystallinity and sharpness of the metal electrode-SWCNT interface illustrated in Fig. 10, that is, two factors that influence the width of the contact and its lattice disorder, play a major influence on the current-voltage characteristics of the electrode-SWCNT contact (e.g., see (Gao et al. 2010)). As an illustrative example, in a copper-NCT interface, the Cu–C bond strength is only about one-fourth that of the C–C bond in CNT. The resistance of the interface is also affected by the resistivity of the CNT in the interface. The resistivity values that characterize CNTs range from 5.1×10^{-8} Ω-m to 8×10^{-3} Ω-m. The cause of this wide variation is probably localized structural defects in the CNTs. The wide range of electrical resistivity achieved so far in practical SWNTs and MWSTs may limit their applications in the near future to devices where a very low contact resistance is not essential, such as in selected types of electrical brushes as illustrated in Fig. 11 (Toth et al. 2009) and in nanomechanical switch systems (Dujardin et al. 2005, de Pablo et al. 2000).

Probe Cards

General Requirements

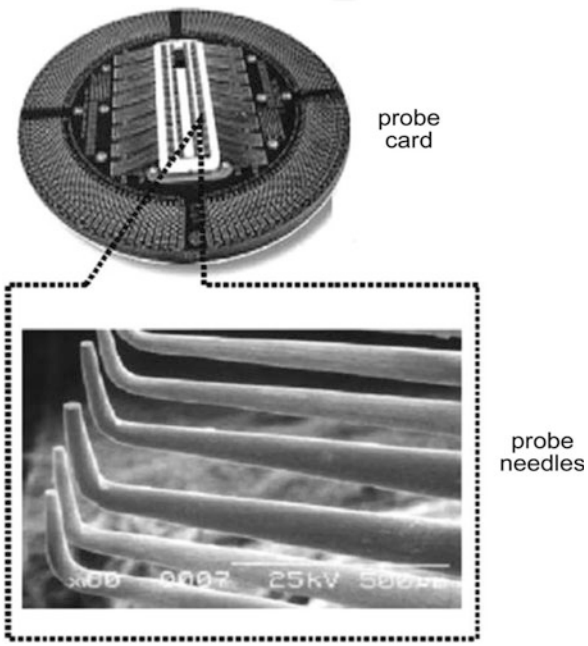
Probe cards have been developed to test the electrical integrity of integrated circuits. Integrated circuits (ICs) consist of a maze of microscopic circuit components and



Wear Mechanisms in Small Electrical Contacts, Fig. 11 (**a**) Experimental arrangement of a MWNT brush on a brass shaft; the SEM micrograph on the right shows the CNT brush before testing, (**b**) Same as (**a**) but showing the wear after several hours of operation (Toth et al. 2009)

electrical pathways that eventually connect to the external world through small metal pads located on the wafer surface. ICs are usually fabricated by the hundreds or thousands on a single thin wafer of semiconducting material, usually silicon, and are adapted with external metal Input/Output (I/O) pads (usually aluminum but sometimes copper) to which tiny gold wires are eventually bonded to connect to microelectronic package leads. Before ICs are cut from the wafer and gold wires are bonded to the external pads, the internal electrical integrity of the circuitry of each IC must be tested. Probe cards are designed for that purpose. They are adapted with small-diameter probe needles that are brought into mechanical contact with the external I/O pads and test each IC by sending diagnostic signals and evaluating IC response. Because the number of probe needles is large (several hundreds or more), they are very thin, on the order of 200 μm in diameter as illustrated in Fig. 12 (Liu et al. 2007).

In practice, a sufficiently large contact force must be applied by the probe needles to displace electrically insulating films and establish electrical contact. This leads to scrub damage on the pads. This damage adds to manufacturing cost by lowering IC assembly yields in manufacturing (Thompson et al. 1994). Figure 13 illustrates the accumulation of aluminum debris on the tips of tungsten needles after increasing number of touchdowns of the needle on I/O aluminum pads. The white rectangles located on the needle tips identify the regions where aluminum flakes were detected (Liu et al. 2007). A typical increase in contact resistance with increasing number of needle touchdowns at a contact load of about 100 mN is illustrated in Fig. 14 (Liu et al. 2007). The increase in contact resistance with increasing number of contact events eventually limits the useful lifetime of probe needles. This limitation has led to investigations of alternatives to large contact loads for establishing a

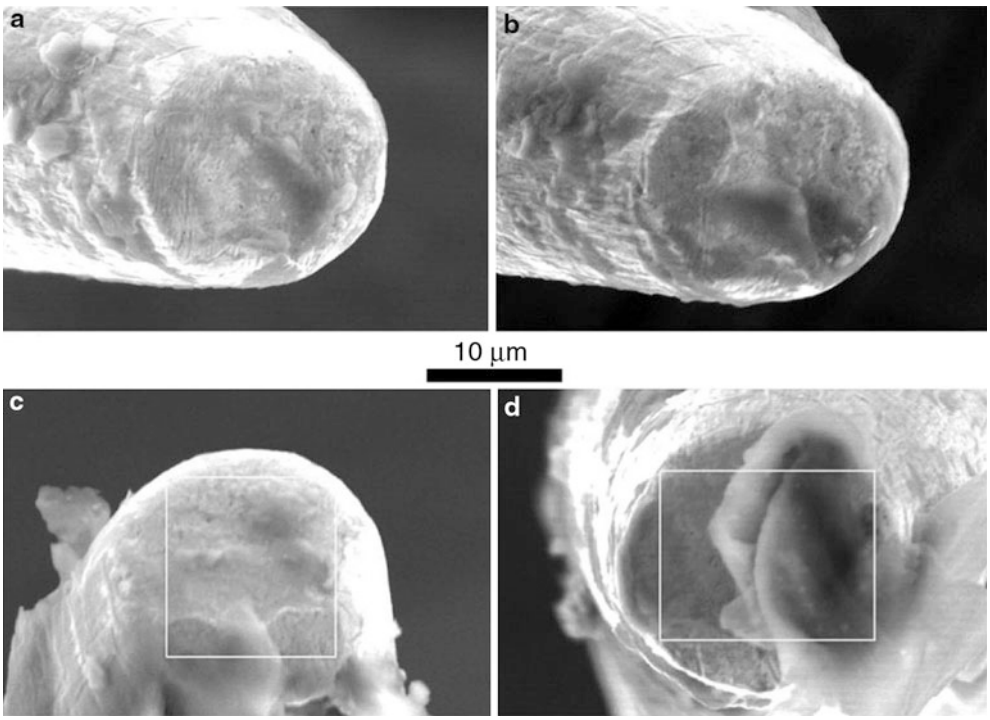


Wear Mechanisms in Small Electrical Contacts, Fig. 12 Photograph of probe card and probe needles (Liu et al. 2007)

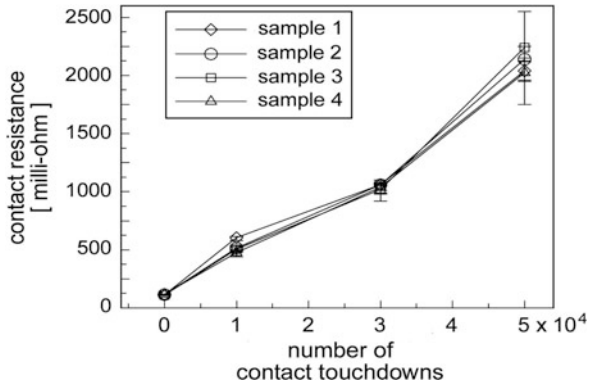
metal-to-metal contact and minimizing the concomitant formation of particulate debris.

One such alternative is based on the mechanism is *fritting* (Holm 1976). In *fritting*, a potential-drop of a few volts is applied between the probe needle and the oxide-covered aluminum surface of the I/O pad to cause electrical breakdown of the oxide layer. This oxide breakdown leads to the subsequent formation of a metal-to-metal contact. If the voltage across the ruptured insulating film is larger than the *melting voltage* of aluminum metal (Holm 1976, Timsit 1999), metal contact is promoted by momentary melting of the metal exposed after film fracture.

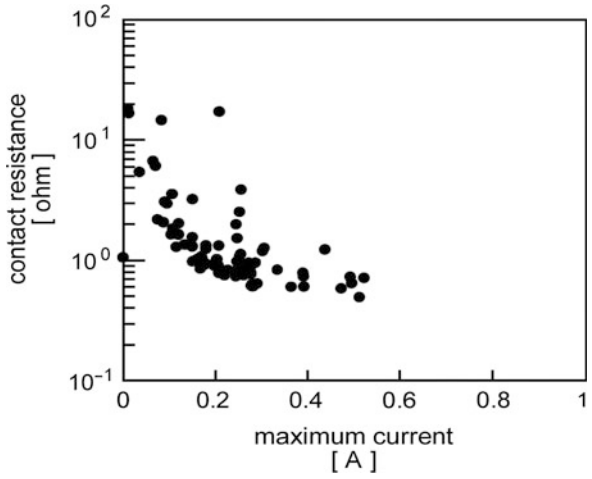
Initial tests have indicated that the contact load for achieving a contact resistance of 2 Ω between a tungsten needle and an oxide-covered aluminum I/O pad could be reduced from about 100 mN in the absence of *fritting* to 0.7 mN with *fritting* at a voltage drop of 1.5 V (Beiley et al. 1995). Later investigations on the use of *fritting* for contacts between Al electrodes and electroplated nickel or gold bumps reported a contact resistance less than 1 Ω without applying any significant external force (Itoh et al. 1999, 2000a, 2000b). Figure 15 illustrates the effect of



Wear Mechanisms in Small Electrical Contacts, Fig. 13 Scanning electron micrographs showing the increased contamination of tungsten probe tips after an increasing number of touchdowns on IC aluminum I/O pads: (**a**) clean tip, (**b**) 10,000 touchdowns, (**c**) 30,000 touchdowns, and (**d**) 50,000 touchdowns (Liu et al. 2007)



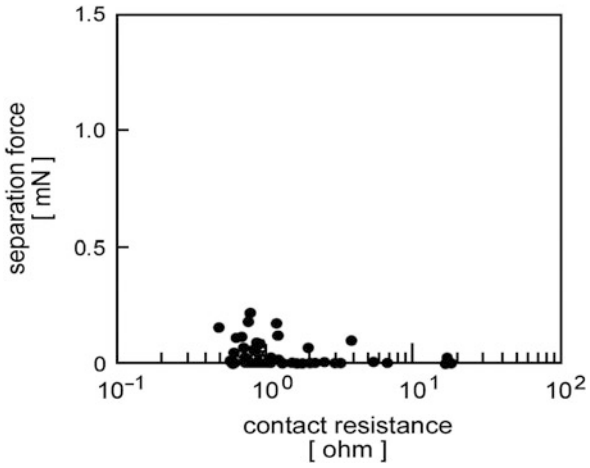
Wear Mechanisms in Small Electrical Contacts, Fig. 14
Variation of contact resistance of a tungsten needle with number of contacts on copper, for a constant contact force on the order of one hundred milli-Newtons (Liu et al. 2007)



Wear Mechanisms in Small Electrical Contacts, Fig. 15
Variation of contact resistance with maximum current passed through an aluminum-tungsten contact after *fritting* of the Al film by a W probe (Kataoka et al. 2003)

current on the contact resistance between a tungsten needle and an oxidized aluminum surface at a contact load of 1 mN, after *fritting* had been achieved using a voltage-drop of 1 V between the needle and the aluminum (Kataoka et al. 2003). The data indicate that a contact resistance of ~0.8 Ω could be achieved at a current of 0.5 A, yielding a voltage-drop across the contact of ~0.4 V.

The data in Fig. 15 are also consistent with the generation of very high contact temperatures and with localized



Wear Mechanisms in Small Electrical Contacts, Fig. 16
Dependence of the force required to separate an aluminum-tungsten contact formed after *fritting*, on the contact resistance shown in Fig. 15 (Kataoka et al. 2003)

melting following *fritting*. Thus, *fritting* at the indicated currents probably led to partial welding on rapid solidification of the contact bridge. This interpretation is confirmed by measurements of a bonding force between the tungsten tip and the aluminum on contact separation, as shown in Fig. 16. Since this separation force amounted only to a few hundred micro-Newtons, it did not materially affect the function of probe card needles.

Cross-References

- ▶ Electrical Contact Materials
- ▶ Sliding Electrical Contact Wear
- ▶ Tribology

References

H.C. Angus, Surface films on precious metal contacts. Br. J. Appl. Phys. **13**, 58–63 (1962)

M. Beiley, J. Leung, S.S. Wong, A micromachined array probe card – characterization. IEEE Trans. Compon. Packag. Manuf. Technol. Part B **18**, 184–191 (1995)

J.P. Bourgoin, S. Campidelli, P. Chenevier, V. Derycke, A. Filoramo, M.F. Goffman, Recent advances in molecular electronics based on carbon nanotubes. CHIMIA Int. J. Chem. **64**, 414–420 (2010)

L. Chen, H. Lee, Z.J. Guo, N.E. McGruer, K.W. Gilbert, S. Mall, K.D. Leedy, G.G. Adams, Contact resistance study of noble metals and alloy films using scanning probe microscope test station. J. Appl. Phys. **102**, 074910-1–074910-7 (2007)

R.A. Coutu Jr., P.E. Kladitis, K.D. Leedy, R.L. Crane, Selecting metal alloy electric contact materials for MEMS switches. J. Micromech. Microeng. **14**, 1157–1164 (2004)

R.A. Coutu Jr., J.R. Reid, R. Cortez, R.E. Strawser, P.E. Kladitis, Microswitches with sputtered Au, AuPd, Au-on-AuPt, and Au-Pt-Cu alloy electric contacts. IEEE Trans. Compon. Packag. Technol. **29**, 341–349 (2006)

P.J. de Pablo, M.T. Martínez, J. Colchero, J. Gómez-Herrero, W.K. Maser, A.M. Benito, E. Munoz, A.M. Baró, Mechanical and electrical properties of single-walled carbon nanotubes. Adv. Mater. **12**, 573–576 (2000)

E. Dujardin, V. Derycke, M.F. Goffman, R. Lefèvre, J.P. Bourgoin, Self assembled switches based on electroactuated multiwalled nanotubes. Appl. Phys. Lett. **87**, 193107/1–193107/3 (2005)

F. Gao, J. Qu, M. Yao, Electronic structure and contact resistance at an open-end carbon nanotube and copper interface. Appl. Phys. Lett. **96**, 102108/1–102108/3 (2010)

O. Hjortstam, P. Isberg, S. Söderholm, H. Dai, Can we achieve ultra-low resistivity in carbon nanotube-based metal composites? Appl. Phys. A Mater. Sci. Process. **78**, 1175–1179 (2004)

R. Holm, *Electric Contacts, Theory and Applications* (Springer, Berlin, 1976)

H. Hosaka, H. Kuwano, K. Yanagisawa, Electromagnetic microrelays: concept and fundamental characteristics, in *Proceedings of the IEEE Workshop on Micro Electro Mechanical Systems, MEMS93, An Investigation of Micro Structures, Sensors, Actuators, Machines and Systems*, Fort Lauderdale, FL, 1993, pp. 12–17

D. Hyman, M. Mehregany, Contact physics of gold microcontacts for MEMS switches. IEEE Trans. Compon. Packag. Technol. **22**, 357–364 (1999)

S. Iijima, Helical microtubules of graphitic carbon. Nature **354**, 56–58 (1991)

T. Itoh, T. Suga, G. Engelmann, J. Wolf, O. Ehrmann, H. Reichl, Applicability of MEMS probe card to wafer-level test, in *Proceedings of InterPACK'99*, Lahaina, HI, vol. 1, 1999, pp. 123–130

T. Itoh, T. Suga, G. Engelmann, J. Wolf, O. Ehrmann, H. Reichl, Characteristics of fritting contacts utilized for micromachined wafer probe cards. Rev. Sci. Instrum. **71**, 2224–2227 (2000a)

T. Itoh, K. Kataoka, G. Engelmann, J. Wolf, O. Ehrmann, H. Reichl, T. Suga, MEMS IC test probe utilizing fritting contacts, in *Proceedings of Symposium on Design, Test, Integration and Packaging of MEMS/MOEMS*, Paris, France, 2000b, pp. 244–249

K.L. Johnson, Mechanics of adhesion. Tribol. Int. **31**, 413–418 (1998)

K.L. Johnson, K. Kendall, A.D. Roberts, Surface energy and the contact of elastic solids. Proc. R. Soc. Lond. A **324**, 301–313 (1971)

K. Kataoka, T. Itoh, T. Suga, Characterization of fritting phenomena on Al electrode for low contact force probe card. IEEE Trans. Compon. Packag. Technol. **26**, 382–387 (2003)

H. Kinoshita, I. Ippei, H. Sakai, N. Ohmae, Synthesis and mechanical properties of carbon nanotube/diamond-like carbon composite films. Diamond Relat. Mater. **16**, 1940–1944 (2007)

L.E. Larson, R.H. Hackett, M.A. Melendes, R.F. Lohr, Micromachined microwave actuator (MIMAC) technology—a new tuning approach for microwave integrated circuits, in *Proceedings of the Microwave and Millimeter-Wave Monolithic Circuits Symposium Digest*, Boston, MA, 1991, pp 27–30

D.S. Liu, M.K. Shih, W.H. Huang, Measurement and analysis of contact resistance in wafer probe testing. Microelectron. Reliab. **47**, 1086–1094 (2007)

M.D. Pashley, J.B. Pethica, The role of surfaces forces in metal-metal contacts. J. Vac. Sci. Technol. A **3**, 757–781 (1985)

M.D. Pashley, J.B. Pethica, D. Tabor, Adhesion and micromechanical properties of metal surfaces. Wear **100**, 7–31 (1984)

J. Schimkat, Contact materials for microrelays, in *Proceedings of the 11th International Workshop on Micro Electro Mechanical Systems, MEMS 98*, Heidelberg, 1998, pp. 190–194

Y.V. Sharvin, A possible method for studying Fermi surfaces. Sov. Phys. **21**, 655 (1965)

P. Thompson, M. Begay, S. Lindsey, D. Vanoverloop, B. Vasquez, S. Walker, B. Williams, Mechanical and electrical evaluation of a bumped-substrate die-level burn-in carrier, in *Proceedings of the 44th Electronic Components and Technology Conference (ECTC)*, Washington, DC, 1994, pp. 700–703

R.S. Timsit, Electrical contact resistance: fundamental principles, in *Electric Contacts: Theory and Applications*, ed. by P.G. Slade (Marcel Dekker, New York, 1999), pp. 1–88

R.S. Timsit, Electrical conduction through small contact spots. IEEE Trans. Compon. Packag. Technol. **CPT-29**, 727–734 (2004)

G. Toth, J. Maklin, N. Halonen, J. Palosaari, J. Juuti, H. Jantunen, K. Kordas, W.G. Sawyer, R. Vajtai, P.M. Ajayan, Carbon-nanotube-based electrical brush contacts. Adv. Mater. **21**, 2054–2058 (2009)

E.W. Wong, P.E. Sheehan, C.M. Lieber, Nanobeam mechanics: elasticity, strength, and toughness of nanorods and nanotubes. Science **277**, 1971–1975 (1997)

O. Yaglioglu, J. Hart, R. Martens, A.H. Slocum, Method of characterizing electrical contact properties of carbon nanotube coated surfaces. Rev. Sci. Instrum. **77**, 095105/1–095105/3 (2006)

Z. Yang, D.J. Lichtenwalner, A.S. Morris III, J. Krim, A.I. Kingon, Comparison of Au and Au–Ni alloys as contact materials for MEMS switches. J. Electromech. Syst. **18**, 287–295 (2009)

M.F. Yu, O. Lourie, M.J. Dyer, K. Moloni, T.F. Kelly, R.S. Ruoff, Strength and breaking mechanism of multiwalled carbon nanotubes under tensile load. Science **287**, 637–640 (2000)

Wear Mechanisms of Polymers

► Polymers for Low-Temperature Tribology Applications

Wear Modeling for Homogeneous Materials

► Wear Contact of Homogeneous Materials

Wear Modeling for Inhomogeneous Materials

► Wear Contact of Inhomogeneous Materials

Wear Modeling in Artificial Knee Joints

Ryan Willing
Hand and Upper Limb Centre, St. Joseph's Health Centre, University of Western Ontario, London, ON, Canada

Synonyms

Knee arthroplasty wear; Knee replacement wear

Definition

Wear modeling in artificial knee joints typically refers to modeling the abrasive/adhesive wear of the ultra-high-molecular-weight polyethylene (UHMWPE) bearing insert of artificial knees using computational methods. Typically, models are limited to the proximal contact surface of the UHMWPE bearing insert, which articulates against the (typically) metal femoral component (Fig. 1). Any model, however, can be extended to study backside wear and wear at the patellofemoral joint. Previous wear models have used different computational analysis methods, input motions, and formulations for actually calculating wear based on simulation results.

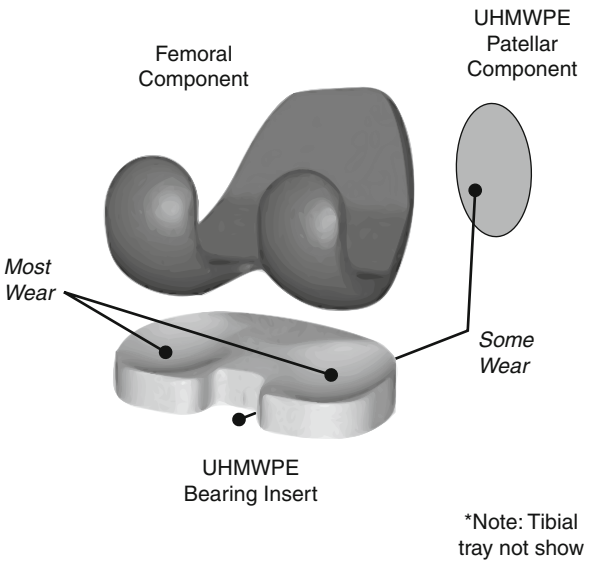
Scientific Fundamentals

Input Motions

During the development and validation of knee replacement wear models, it is important to validate predicted values against experimental results. For this, it is best to use a standard set of knee motions in order to generate results comparable with those of other research groups. A standard set of knee motion data is also beneficial for comparing the performance of different implant designs. For these reasons, most researchers use the standard knee motion protocols established by the International Organization for Standardization (ISO):

ISO 14243-1 (ISO 2009) is a set of load and displacement waveforms for *load* controlled wear simulations. This standard prescribes axial load, anterior-posterior (AP) forces and internal–external (IE) torques. The flexion-extension angle is also prescribed. Actual AP displacements and IE rotations are a function of the design of the implant and spring forces, which are included in order to simulate soft tissue restraint.

ISO 14243-3 (ISO 2004) is a set of load and displacement waveforms for *displacement* controlled wear simulations. The axial load is prescribed by the standard, but no other loads are specified. Instead, AP displacements, IE rotations, and flexion angles are prescribed.



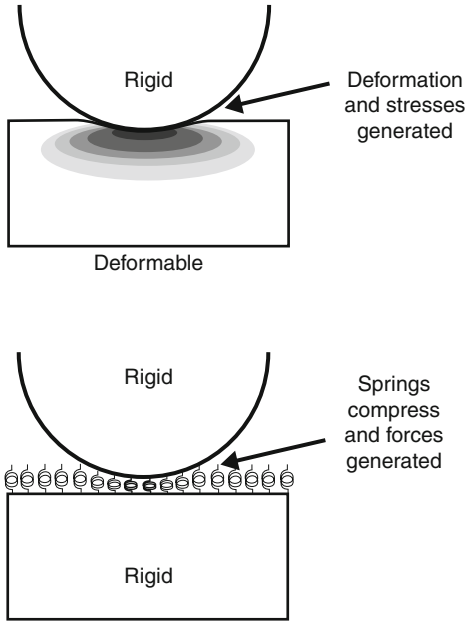
Wear Modeling in Artificial Knee Joints, Fig. 1 Depiction of the femoral component, ultra-high-molecular-weight polyethylene (UHMWPE) bearing insert and UHMWPE patellar component of a typical total knee replacement. Wearing surfaces indicated

The load control standard is preferred, because the actual implant motions become a function of the implant shape, which is a closer match to the physiological condition. The displacement control standard was developed for wear testing apparatuses that do not function in a load control mode. Computational experiments, however, do not have such a constraint and are able to easily implement the load control standard with simulated soft tissue restraint.

Both ISO standards are designed to be close to the conditions at the knee during gait. Gait loading is by far the most frequent loading at the knee, but not necessarily the most critical in terms of wear. For this reason, some wear simulations will use a combination of gait, stair ascending, descending, cutting motions, or other patient-specific loading profiles. The more physiological the motion at the knee, the more likely that any simulation will be able to match the wear patterns seen on retrievals.

Contact Modeling

Two fundamentally different approaches to simulating the contact between knee implant components are used.



Wear Modeling in Artificial Knee Joints, Fig. 2 Contact between components can be modeled as rigid-deformable (*top*), where contact stresses, deformations, and stress gradients through deformable components are calculated. Contact can also be modeled as rigid–rigid (*bottom*), where contact stresses can be calculated using a bed of springs, but material deformation is assumed negligible and stresses are not calculated

In one approach, the UHMWPE insert is modeled as a deformable body, while the femoral component (or other articulating surfaces) can be modeled as deformable or rigid, due to the much higher elastic modulus (Fig. 2, top). The other approach is to consider all components as rigid bodies, neglecting any component deformations, and calculating the contact stresses based on an elastic foundation model (Fig. 2, bottom).

Deformable Models

The benefit of modeling the UHMWPE insert as a deformable body is the ability to account for any small deformations due to loads transferred through the component, as well as determining the resulting internal stress distribution. Depending on the shape of the implant components and the loads being applied, it is possible that the small deformations of the UHMWPE insert will affect the contact pressure and relative motion of the implant components, which would affect the progression of wear. The internal stress distribution, while not used in most wear models, could also provide important information for predicting other damage modes (such as delamination due to subsurface crack initiation). These models will typically rely on quasi-static finite element (FE) contact analysis. During finite element analysis (FEA), the components are discretized into many small elements comprising a FE mesh. Appropriate material properties are assigned for any deformable components, and contact between components, including friction, can be calculated using contact algorithms (e.g., penalty method, augmented Lagrange method, pure Lagrange method), which are typically included with commercially available FEA code. These contact algorithms detect and minimize or eliminate penetration of one contact surface through another by generating repulsive contact forces at the interface, which deforms accordingly. Depending on the refinement of the FE mesh and the nonlinearity of the problem (due to contact, large deformations/rotations, and material properties), computational expense may become a limiting factor.

Elastic Foundation

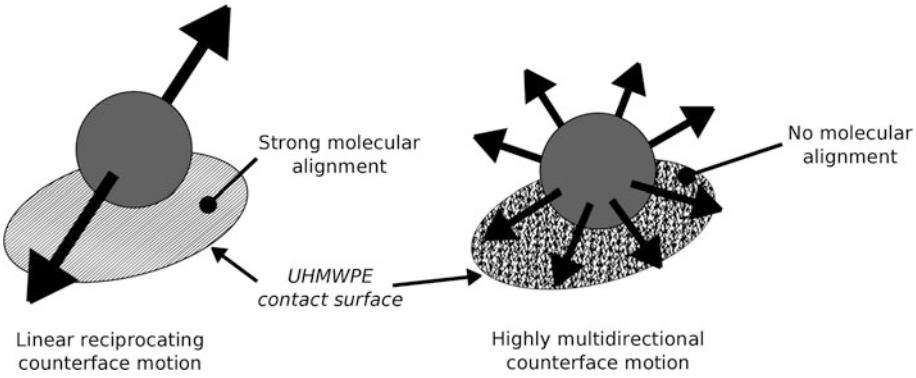
To overcome the limiting computational expense of deformable FE models, researchers have developed reliable elastic foundation models. A bed of springs distributed across the contact surface, the stiffness of which is a function of the relative material properties of the implant components, repels the rigid bodies from one another. The compression of the springs can be used to calculate what the resulting contact stress distribution would be if the parts were deformable. The clear benefit of such models is that contact pressures can be estimated using rigid multibody dynamic simulations, which can be an order of magnitude less computationally expensive than a corresponding FE simulation. The small deformations and stress distributions throughout the UHMWPE insert are neglected, however, so there is a penalty in terms of solution accuracy and the amount of results data which can be extracted.

Using either simulation method, the contact stress (P) and counterface sliding vectors ($\mathbf{s}$) at every point on the contact surface are calculated through the duration of the simulation, which will be a function of the shape of the components, friction, and the driving loads and displacements.

Wear Models

Archard's Wear Model

The original and most widely used wear models are based on Archard's wear model (Archard and Hirst 1956). The simple empirical model was originally derived for



Wear Modeling in Artificial Knee Joints, Fig. 3 Linear reciprocating motion will align surface UHMWPE polymer molecules with the predominant molecular orientation. Highly multidirectional motion will not induce molecular realignment

predicting wear between metal surfaces rubbing under unlubricated conditions. It was adopted for use in simulating the wear of UHMWPE in artificial hips (Maxian et al. 1996), and later for artificial knee systems (Fregly et al. 2005). The model defines wear as a function of contact pressure, sliding distance, and a wear coefficient k, such that

$$\delta_{wear} = k \sum_{i=1}^{n} P_i |\mathbf{s}_i| \tag{1}$$

where P_i and $\mathbf{s}_i$ are the contact pressure and sliding distance during each sub-step i of n total sub-steps representing an entire motion cycle. δ_{wear} is the total damage depth at a single point on the contact surface.

Cross-Shear Sensitive Wear Models

UHMWPE wear rates are sensitive to the amount of cross-shear at the contact surface, evidenced by the dramatic increases in wear rates under situations of multidirectional versus purely linear reciprocating counterface motions. This behavior cannot be accounted for using Archard's wear model, possibly explaining why reported k values range between 2.63×10^{-10} and 1.98×10^{-6} mm^3/Nm.

The strain hardening or orientation softening wear model proposed by Wang 2001 describes the mechanism behind the sensitivity of UHMWPE wear to multidirectional sliding. In short, due to the frictional-work of a counterface motion, the polymer molecules at any point on the UHMWPE contact surface will align in a predominate molecular orientation (PMO, $\bar{\theta}$) (Fig. 3). Frictional work acting parallel with the PMO leads to plastic deformation or further reorientation, whereas frictional work acting perpendicular to the PMO (caused by cross-shear) leads to material removal by intermolecular splitting. This means that only energy released perpendicular to the PMO is responsible for wear. During situations of linear reciprocating counterface motion, the PMO aligns with the direction of counterface motion, so little energy is released perpendicular to the PMO, resulting in little wear. During situations of highly multidirectional counterface motion, up to half of the energy can be released perpendicular to the PMO, resulting in order-of-magnitude increases in wear rates.

This knowledge can be implemented for calculating TKR wear using models of varying complexity. Three different approaches will be described here: the σ^* method, the $k(CS)$ method, and the β-weighted method. Each method will be described in the context of a small finite area of consideration, for which simulation results include sliding vectors ($\mathbf{s}$) and normal contact pressure (P) during $i = 1$ to n sub-steps (Fig. 4).

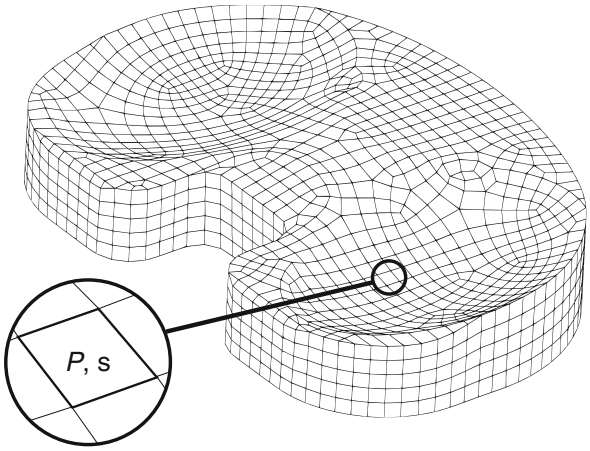
σ^* Method

This model (described further in (Hamilton et al. 2005)) was developed as a method for visualizing the amount of multidirectional motion at every point on the UHMWPE insert contact surface using contour maps. A dimensionless normalized crossing intensity value (σ^*) can be assigned to any location. This method starts with the calculation of tribological intensity vectors ($\boldsymbol{\tau}_i$) for each time increment (i) of a simulation, defined as

$$\boldsymbol{\tau}_i = P_i \mathbf{s}_i \tag{2}$$

which can be decomposed into a scalar τ_i and angle θ_i with respect to a fixed reference angle. The PMO is then calculated by

$$\bar{\theta} = \sum_{i=1}^{n} \tau_i \theta_i \Big/ \sum_{i=1}^{n} \tau_i \tag{3}$$



Wear Modeling in Artificial Knee Joints, Fig. 4 Contact stress (*P*) and counterface sliding vectors (s) are calculated for each finite area at the contact surface, after contact surface descretization. Wear can then be calculated at each finite contact surface location

A crossing intensity value (σ) is calculated based on the root mean square of the deviation of each $\boldsymbol{\tau}_i$ vector from the PMO:

$$\sigma = \sqrt{\sum_{i=1}^{n} \left[\tau_i\left(\bar{\theta} - \theta_i\right)\right]^2 / n} \tag{4}$$

The crossing intensity value is then made nondimensional and normalized by dividing that value by the worst case scenario crossing intensity (σ_o), which is assumed to be caused by circular counterface motion, with the same total tribological intensity magnitude ($|\boldsymbol{\tau}|$). The resulting normalized crossing intensity is

$$\sigma^* = \sigma / \sigma_o \tag{5}$$

and will be a value between 0 (unidirectional or linear reciprocating counterface motion) and 1 (fully multidirectional counterface motion).

While this method does not suggest a direct means of calculating wear while including the normalized crossing intensity value, comparison with previously published experimental results showed that there was a strong relationship between the squared normalized value (σ^{*2}) and wear rates. Possible implementation by simply augmenting Archard's wear model could take the form

$$\delta_{wear} = \sigma^{*2} k \sum_{i=1}^{n} P_i |\mathbf{s}_i| \tag{6}$$

k(CS) Method

Unlike the σ^* method described above, this method (described further in (Kang et al. 2008)) takes frictional work into account, and is based directly on the strain-hardening model proposed by Wang. The essence of these models is to calculate the amount of cross-shear (*CS*), based on frictional work perpendicular to the PMO, and then make the wear coefficient k a function of *CS*.

The amount of frictional work perpendicular to the PMO ($\bar{\theta}$) is calculated as

$$W_{\perp} = \sum_{i=1}^{n} \mu P_i |\mathbf{s}_i| \sin^2\left(\theta_i - \bar{\theta}\right) \tag{7}$$

where μ is the coefficient of friction. The PMO, $\bar{\theta}$, is chosen such that $W_{\perp}$ is minimized. *CS* can then be calculated as the ratio of perpendicular frictional work to the total frictional work

$$W_{total} = \sum_{i=1}^{n} \mu P_i |\mathbf{s}_i| \tag{8}$$

$$CS = \frac{W_{\perp}}{W_{total}} \tag{9}$$

which will be a value between 0 and 0.5.

Once *CS* has been calculated, the most straightforward approach is to augment the Archard model in such a way that the local wear coefficient, k, becomes a function of the amount of cross-shear:

$$\delta_{wear} = k(CS) \sum_{i=1}^{n} P_i |\mathbf{s}_i| \tag{10}$$

Previous work has attempted to find an appropriate *k(CS)* relationship that yields good agreement between the experimentally measured and model predicted wear. The experiments of Kang et al. yielded the power function relationship $k = 8.0 \times 10^{-7} CS^{0.2071}$, which had an *R*-squared value of 0.92 for conventional UHMWPE. For cross-linked UHMWPE (10 MRad), a logarithmic expression was used (R^2=0.72) such that $k = 1.0 \times 10^{-10} \ln(CS) + 1.0 \times 10^{-7}$.

β-Weighted Method

The *k(CS)* method described above assumes that all polymer molecules have aligned completely with the PMO. The β-weighted method (described further in (Willing and Kim 2009)), however, assumes that any point on the surface can instead be described by a fraction that has been completely aligned with the PMO, and a fraction that has not undergone any molecular realignment (no preferential direction). This means that this model is able to take into account the intensity of the PMO, or the amount of alignment. Still relying on the frictional work theory proposed by Wang, this model calculates the wear expected

from the fully aligned fraction and the multidirectional fraction. The model then calculates the actual wear as the weighted sum of the two fractions, which is dependent on the predicted amount of molecular alignment.

This model again calculates the amount of cross-shear based on the ratio of frictional work perpendicular to the PMO. The ratio has been doubled, however, in order to normalize the value between 0 and 1 (like the crossing intensity value from the σ^* method) and is denoted σ such that

$$\sigma = 2CS = \frac{2W_{\perp}}{W_{total}} \tag{11}$$

The PMO in the aligned fraction is determined in the same manner as the *k(CS)* method, and it is the angle $\bar{\theta}$ that minimizes $W_{\perp}$ (which will also minimize *CS* and σ). In this method, it is assumed that the amount of molecular alignment (β) is a function of the amount of cross-shear or the crossing intensity, and is calculated

$$\beta = 1 - \sigma = 1 - 2CS = \frac{W_{total} - 2W_{\perp}}{W_{total}} \tag{12}$$

In the multidirectional, nonaligned polymer molecule fraction, there is no PMO. In this fraction, exactly half of the work would contribute to intermolecular splitting (wear), and is proportional to a wear coefficient *k*. The wear damage in this fraction can then be calculated as

$$\delta_{multi} = \frac{1}{2} kW_{tot} = \frac{1}{2} k\mu \sum_{i=1}^{n} P_i |\mathbf{s}_i| \tag{13}$$

Note that the coefficient of friction (μ) can be absorbed into *k*, or handled separately as shown in the equation above.

In the aligned polymer molecule fraction, wear only occurs due to frictional work released perpendicular to the PMO, but is also proportional to the wear coefficient *k*. The wear damage in this fraction is calculated as

$$\delta_{aligned} = kW_{\perp} = k\mu \sum_{i=1}^{n} P_i |\mathbf{s}_i| \sin^2(\theta_i - \bar{\theta}) \tag{14}$$

Since this model assumes a varying amount of alignment (β) at any location, the resulting wear (δ_{wear}) is the weighted sum of the two wear calculations (δ_{multi} and $\delta_{aligned}$), such that

$$\delta_{wear} = \beta\delta_{aligned} + (1 - \beta)\delta_{multi} \tag{15}$$

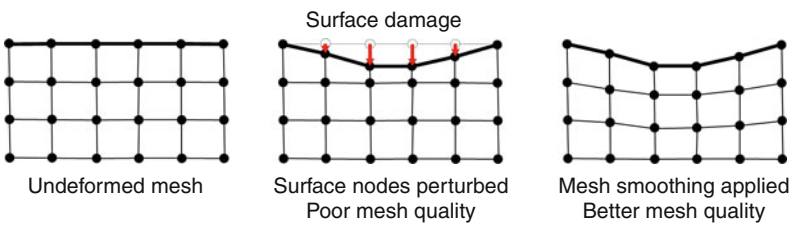
Each of the three methods has its own strengths and weaknesses in terms of computational cost, ease of implementation, and level of complexity. The important thing is that cross-shear is considered, at the very least.

Surface Evolution

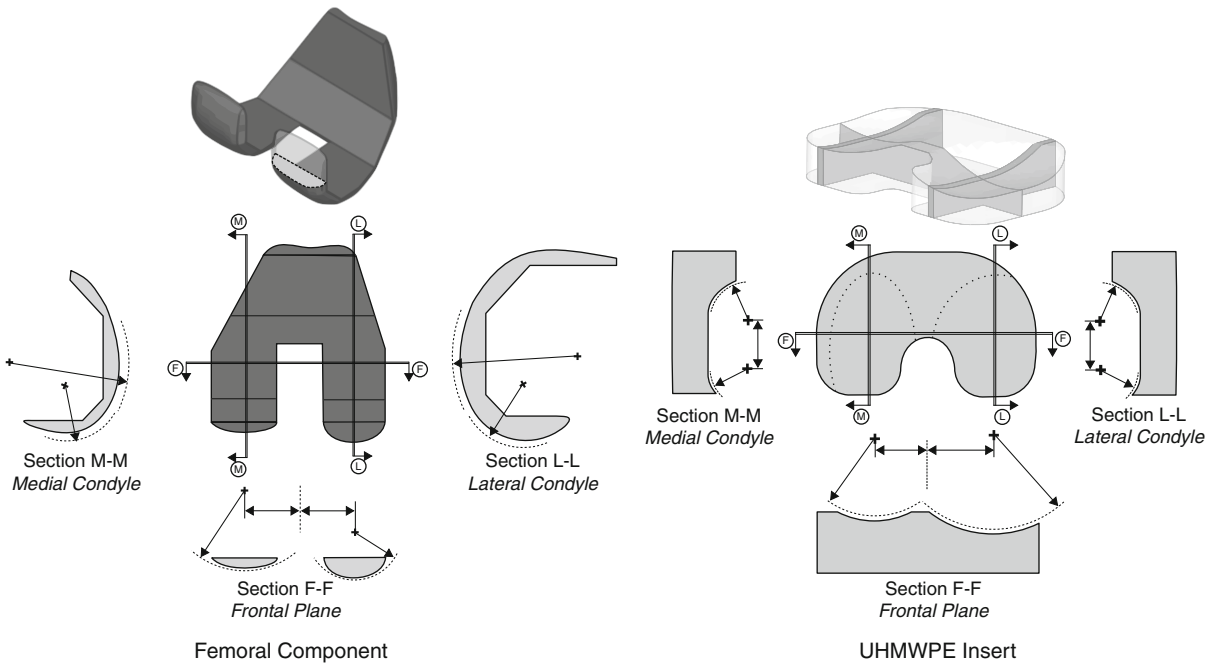
Long-term wear simulations must account for the change in geometry at the contact surface as a result of material removal and creep. Creep, or cold-flow, is the time-dependant deformation of a material subjected to a load. UHMWPE is a viscoelastic and viscoplastic material that will deform under load in a time-dependant manner. Creep should be included in any TKR wear simulation, the magnitude of which depends on the applied loads, duration of applied loads, and the thickness of the UHMWPE insert. While the wear rates at the UHMWPE insert contact surface are typically linear, the creep deformation rates are highly nonlinear, with most creep damage occurring early in the duty cycle of the component. The wear simulation should include surface evolution iterations where the damage due to wear and creep at the contact surface is calculated, and the contact surface is modified to represent that damage. Subsequent iterations should then be based on this "damaged" model. Depending on the mesh resolution and the amount of damage at the contact surface, the FE mesh (when considering deformable components) may need to be smoothened to ensure model accuracy (Fig. 5).

Fluid Film Lubrication

Fluid film lubrication is known to occur in artificial joints, and previous studies have investigated the importance of



Wear Modeling in Artificial Knee Joints, Fig. 5 Long-term surface damage can be approximated by deforming the contact surface. If a deformable FE mesh is being used, the underlying mesh may need smoothing in order to ensure model accuracy



Wear Modeling in Artificial Knee Joints, Fig. 6 Parametric TKR models can be used during parameter studies or for design optimization of implant components for reduced wear. Example parameters include the different radii of curvature of implant features in different planes, which affect surface conformity and contact characteristics

the film thickness in artificial knee joints using experimental and numerical approaches. Previous models have suggested that changes in counterface sliding direction at the contact surface may reduce the film thickness and increase the amount of wear damage locally. To the author's knowledge, however, there have not been any published artificial knee wear models that include the effects of fluid film lubrication. As the complexity of numerical models increases and the capability of computer systems expand, it will be possible to consider these effects and better understand the relationship between wear and lubrication.

Key Applications

Complement to Experiments

Wear experiments for total knee replacements will typically run between 3 and 5 million cycles in order to observe the long-term wear characteristics of a given implant design under a given set of driving motions. Such long-term experiments are costly in terms of time and money. Computational wear simulation can be completed in a fraction of that time. Experimental in vitro testing is still an essential stage of implant development, but the amount of experiments required can be greatly reduced. Reliable computational wear simulations can be used to screen out undesirable implant designs early, before the prototype and experiment stage.

Parameter Studies and Design Optimization

The ability to calculate the wear characteristics of a candidate implant design *in silico* introduces an opportunity to perform parameter studies on implant shapes (using parametric models, as in Fig. 6), driving motions, and implant positioning. Such studies can be used to quickly identify what factors truly contribute to changes in wear characteristics. Taking this notion a step further, these wear simulations can be included as an objective function for total knee replacement (TKR) design optimization. A properly developed optimization framework can quickly and efficiently examine a large design space of candidate TKR designs and identify the most optimal shapes with respect to wear performance. Pending the availability of other performance measurement

simulations, such as kinematic analyses, one can consider multiple objective functions simultaneously using multiobjective design optimization methods. Such methods can help implant developers better understand the relationship between wear and other TKR performance measures.

Other Contact Surfaces

While few examples exist in the literature, it is certainly possible to apply the same wear modeling approaches to other contact surfaces in TKR. One example is the patellofemoral joint. The patellar component is normally resurfaced with UHMWPE and articulates against the femoral component. Wear modeling can also be applied to the distal (bottom) surface of the UHMWPE insert in mobile or fixed bearing designs. Patellar component and backside wear has been observed clinically in both mobile and fixed bearing TKR designs.

Cross-References

► Contact Conditions in the Artificial Knee
► Testing of Artificial Knee Joints
► Wear Modeling of Artificial Hip Joints

References

J.F. Archard, W. Hirst, The wear of metals under unlubricated conditions. Proc. R. Soc. London Ser. A **236**, 397–410 (1956)

B.J. Fregly, W.G. Sawyer, M.K. Harman, S.A. Banks, Computational wear prediction of a total knee replacement from in vivo kinematics. J. Biomech. **38**, 621–626 (2005)

M.A. Hamilton, M.C. Sucec, B.J. Fregly, S.A. Banks, W.G. Sawyer, Quantifying multidirectional sliding motions in total knee replacements. J. Tribol. **127**, 280–286 (2005)

L. Kang, A.L. Galvin, T.D. Brown, Z. Jin, J. Fisher, Quantification of the effect of cross-shear on the wear of conventional and highly cross-linked UHMWPE. J. Biomech. **41**, 340–346 (2008)

T.A. Maxian, T.D. Brown, D.R. Pedersen, J.J. Callaghan, A sliding-distance-coupled finite element formulation for polyethylene wear in total hip arthroplasty. J. Biomech. **29**, 687–692 (1996)

ISO Standard 14243-3. Wear of total knee-joint prostheses – Part 3: Loading and displacement parameters for wear-testing machines with displacement control and corresponding environmental conditions for test (International Organization for Standardization, Geneva, 2004)

ISO Standard 14243-1. Wear of total knee-joint prostheses – Part 1: Loading and displacement parameters for wear-testing machines with load control and corresponding environmental conditions for test (International Organization for Standardization, Geneva, 2009)

A. Wang, A unified theory of wear for ultra-high molecular weight polyethylene in multi-directional sliding. Wear **248**, 38–47 (2001)

R.T. Willing, I.Y. Kim, A holistic numerical model to predict strain hardening and damage of UHMWPE under multiple total knee replacement kinematics and experimental validation. J. Biomech. **42**, 2520–2527 (2009)

Wear Modeling of Artificial Hip Joints

Feng Liu
School of Mechanical Engineering, Institute of Medical and Biological Engineering, University of Leeds, Leeds, UK

Synonyms

Computational wear simulation of artificial hip joints; Computer wear modeling

Definition

The wear of bearing surfaces caused by mechanical wear processes such as abrasive and adhesive wear accounts for the majority of the wear debris generated in artificial hip joints. Wear debris-induced late implant loosening has been recognized as the primary factor affecting the long-term performance of artificial hip joints. Computational wear modeling has been increasingly considered to complement laboratory simulator wear tests for wear assessment and implant design optimization. Here, the focus is on the wear modeling of artificial hip joints, and the methodologies developed may also be relevant to other joints such as knee, spine, and ankle, as covered separately in ► Wear modeling in artificial knee Joints; ► Tribology of ankle joints. Computational wear modeling provides a quick evaluation tool for wear simulation of artificial hip joints, particularly in parametric studies.

Computational wear modeling mainly deals with the numerical simulation of the wear laws for ultra-high-molecular-weight polyethylene (UHMWPE) bearings and metallic bearings used for artificial joints and its application in artificial hip joints. The main wear modeling procedures are generally described here. The wear models are Archard's law and a newly proposed alternative law for metallic and UHMWPE hip bearings, respectively. The wear factors or coefficients considered by the above wear models are experimentally determined either from the physical hip simulators and clinical measurements or from the simple wear tests, for example, using a pin-on-plate tester. In contact modeling, finite element (FE) methods are largely used through quasi-static contact analyses in commercially available FE software to obtain contact stress distribution on the bearing surfaces. In the wear calculation, the linear wear depths derived as the

functions of contact pressure and/or sliding distance are numerically evaluated. The resulting wear is incorporated through geometry modification to the FE contact model and the above solution process is iteratively carried out until the required simulation cycles are completed.

Scientific Fundamentals

Wear Models

The main wear model used for artificial hip joints is the classical Archard wear law:

$$V = KFL \tag{1}$$

where V is the volumetric wear, F is the normal load, L is the sliding distance, and K is the constant referred to as *specific wear rate* in engineering or *wear factor* in orthopedic biotribology (Maxian et al. 1996). This equation is derived from experimental observations and the assumption that the surface asperities in contact are deformed plastically and the real contact area is directly proportional to the load applied (Arachard 1953). The constant (K), defined as a wear factor, represents the wear volume per unit sliding distance per unit normal load. This factor is incorporated to reflect the overall tribological influences of the bearing surfaces at work, such as material combinations, surface roughness, lubrication, and so on, and its values are generally experimentally determined. For the computational modeling of artificial hip joints, the wear factor is selected from a wide range of data to match clinical or experimental simulator measurements. The computational prediction thus obtained is not generic and independent but relevant to the specific set of conditions.

A linear wear depth equation can be further obtained by dividing both sides of (1) with the apparent contact area associated with the load (F) as follows:

$$\delta = KpL \tag{2}$$

where δ is the linear wear depth and p is the contact pressure. This equation is normally used to numerically calculate the wear instead of the volumetric wear equation.

Another wear law especially for the UHMWPE articulating against metal or ceramic bearings in total hip replacements (THRs) is based on the hypothesis that the volumetric wear (V) is explicitly proportional to the apparent contact area (A) and the sliding distance (L) under a certain range of contact pressure, given by

$$V = CAL \tag{3}$$

where C is a dimensionless constant referred to as a *wear coefficient*. In association with (3), the linear wear depth equation is written as

$$\delta = CL \tag{4}$$

For polymeric bearing materials used in artificial joints, due to their relatively low elastic modulus, the deformation of the asperity contacts is largely elastic, and the real contact area of the polyethylene bearing is not markedly different from its apparent contact area. The use of Archard's law for polyethylene leads to the wear factor (K) being a strong function of contact pressure, while with the use of the new law, the wear coefficient (C) is largely constant under the pressures considered for the UHMWPE bearings in artificial joints (Liu et al. 2011).

A major factor affecting UHMWPE wear rates is recognized as the cross-shear resulting from the multidirectional motion of the UHMWPE bearings. Cross-shear motion refers to the local counterface motion that is transverse to the direction of strain hardening of the polyethylene. With linear tracking motion, the molecules of UHMWPE are stretched along the sliding direction, leading to significant strain hardening and increase in wear resistance. With multidirectional motion, surface molecules align preferentially in the principal direction of sliding, strengthening in one particular direction and weakening in the transverse direction. This weakening is known as *orientation softening*, which increases wear generation (Wang 2001).

The cross-shear can be quantified using the ratio of the frictional work in the direction perpendicular to the principal molecular orientation (PMO) to the total frictional work (CSR) in the form of

$$CSR = \sum W_{ti} / \sum (W_{ti} + W_{pi}) \tag{5}$$

where W_{ti} and W_{pi} are the frictional work components perpendicular to and along with the PMO direction, respectively. The frictional work can be generally calculated as the product of the contact pressure and the sliding distance, and the PMO direction is the direction along which the maximum amount of frictional work is released. Equation (5) can be used in an attempt to develop a more generic approach. By doing so, the coefficient is determined for the polyethylene hip joints as $C = V/(AL)$ based on laboratory pin-on-plate wear tests. In the test, a rotating polyethylene pin is used against a translating metal plate under lubricated conditions, and different cross-shear motions are achieved by combining different rotation angles of the pins and translation lengths of the plates during tests (Kang et al. 2008). The CSR is calculated to range between 0 and 0.5, accordingly.

For metal-on-metal bearings, wear factors or coefficients can be expected to be highly dependent on lubrication regimes. However, directly incorporating the

lubrication regimes into the equations of wear factors or coefficients has not been easy (Harun et al. 2009). Instead, different wear factors have been used to reflect different lubrication regimes, for example, the running and steady-state phases respectively (Liu et al. 2008).

Contact Models

Contact mechanics of artificial hip joints are largely modeled with FE methods, as addressed elsewhere *in Encyclopedia of Tribology*. The contact models with smooth bearing surfaces without incorporation of asperities and under dry contact conditions are used to obtain the contact pressure distribution on the bearing surfaces. Such an approximation is considered to be valid because the pressure distribution under lubricated contacts is close to that of dry contacts, and the smooth surface assumption has a small effect on the global, average pressure distribution modeled with the above wear equations. The quasi-static contact analyses are modeled at each time instant during a motion cycle instead of a complete transient simulation. FE contact analyses are widely available in many commercial FE software packages. With the FE software used, material properties can be readily incorporated, for example, the plastic deformation of UHMWPE can be modeled using true strain-stress data.

Another simplified model used is the elastic foundation model or column model for an elastic solution (Bartel et al. 1986). It assumes that the contact pressure at a point on the surface is only dependent on the deformation at that point, and the pressure is given by

$$p = \left[\frac{1}{1-2v} + \frac{2}{1+v}\left(\frac{R_3}{R_2}\right)^3\right] \cdot \left[\left(\frac{R_3}{R_2}\right)^3 - 1\right]^{-1} \cdot \frac{E}{R_2} u \tag{6}$$

where u is deformation, p is contact pressure, E is Young's modulus, v is Poisson's ratio, and R_2 and R_3 are the inner and outer radius of the bearing. If R_3/R_2 tends to 1, it can be further simplified

$$p = \frac{E}{d}\left[1 - \frac{2v^2}{1-v}\right]^{-1} u \tag{7}$$

where d is the thickness of the bearing.

Numerical Wear Calculation

The wear depth equations are used and numerically solved in the forms

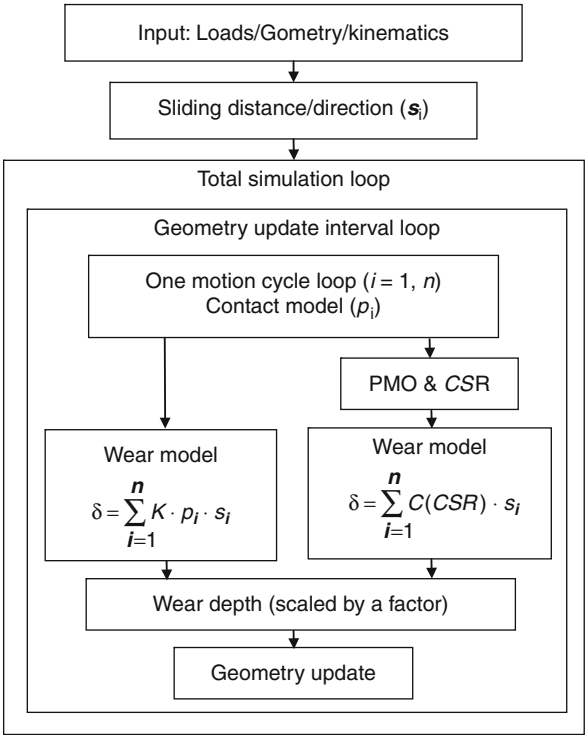
$$\delta = \sum_{i=1}^{n} K \cdot p_i \cdot s_i \tag{8}$$

and

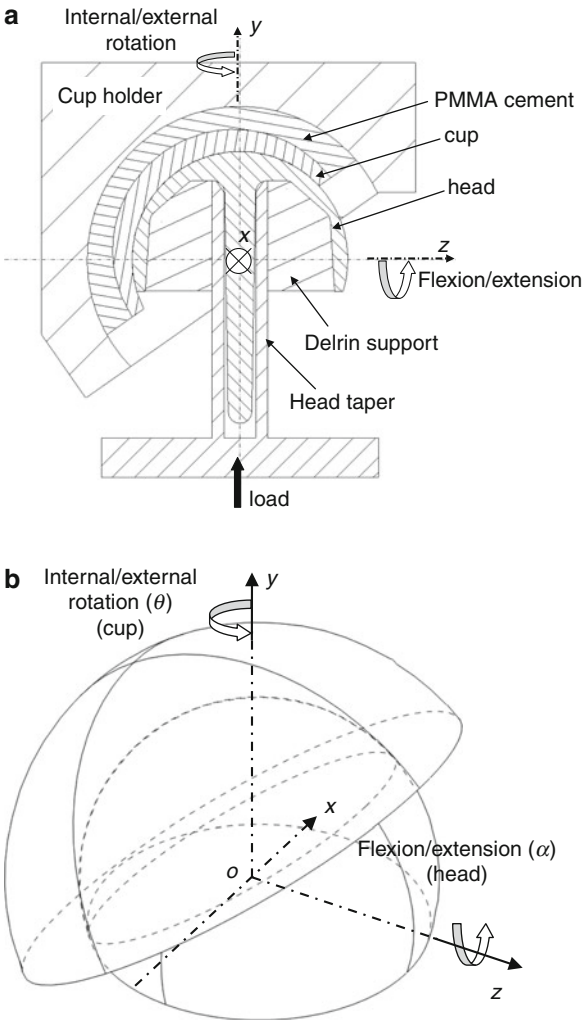
$$\delta = \sum_{i=1}^{n} C(CSR) \cdot s_i \tag{9}$$

where n is the discretized time intervals over a motion cycle, p_i and s_i denote the instantaneous contact pressure and sliding distance, respectively, and C is the function of cross-shear ratio (CSR) as determined by using (5) for the hip joint bearing surface.

In (8) and (9), the FE contact models are evaluated by quasi-static solutions and the corresponding sliding distances are calculated using the coordinate transformation for the relative motions of the cup and head. The geometry modification due to the wear is considered at a larger time interval, such as 250,000 motion cycles for an example hip joint model, and, therefore, the linear wear depths are calculated for a motion cycle and multiplied by the above factor to obtain the wear over the geometry update interval. The contact pressure, the PMO direction, and the cross-shear ratio are calculated accordingly after each geometry modification interval. The linear wear depths calculated are then converted to the volumetric wear for



Wear Modeling of Artificial Hip Joints, Fig. 1 Wear modeling flowchart



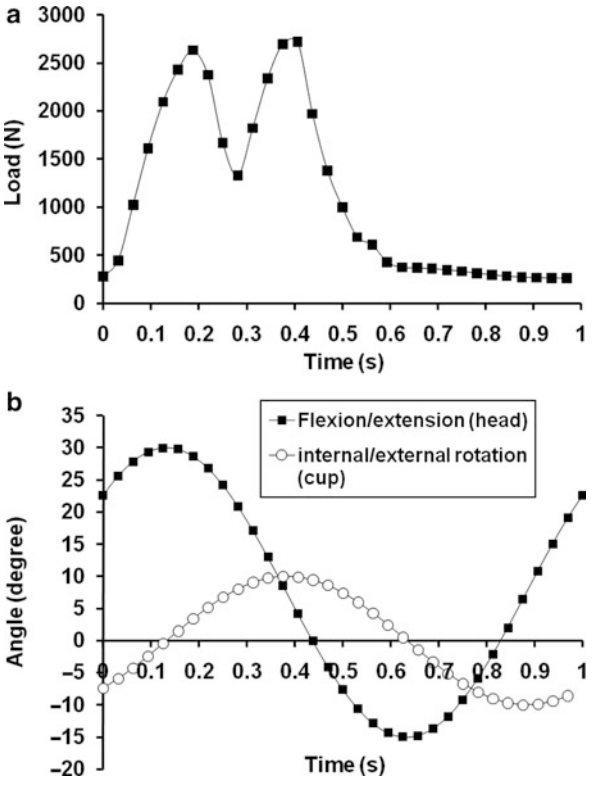
Wear Modeling of Artificial Hip Joints, Fig. 2 (**a**) Schematic cross-section of a hip joint resurfacing replacement tested in a hip joint simulator showing the bearing geometry, articulation, fixation, motion, and loading. (**b**) Three-dimensional representation of the bearings under the motion and loading conditions (Taken from Liu et al. 2008)

the joint bearings. Figure 1 shows a flowchart for the wear modeling.

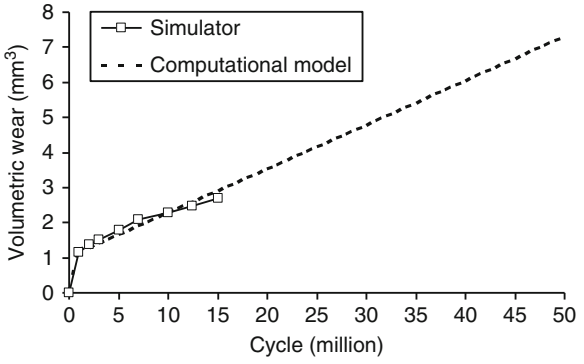
Key Applications

Metal-on-Metal Bearings

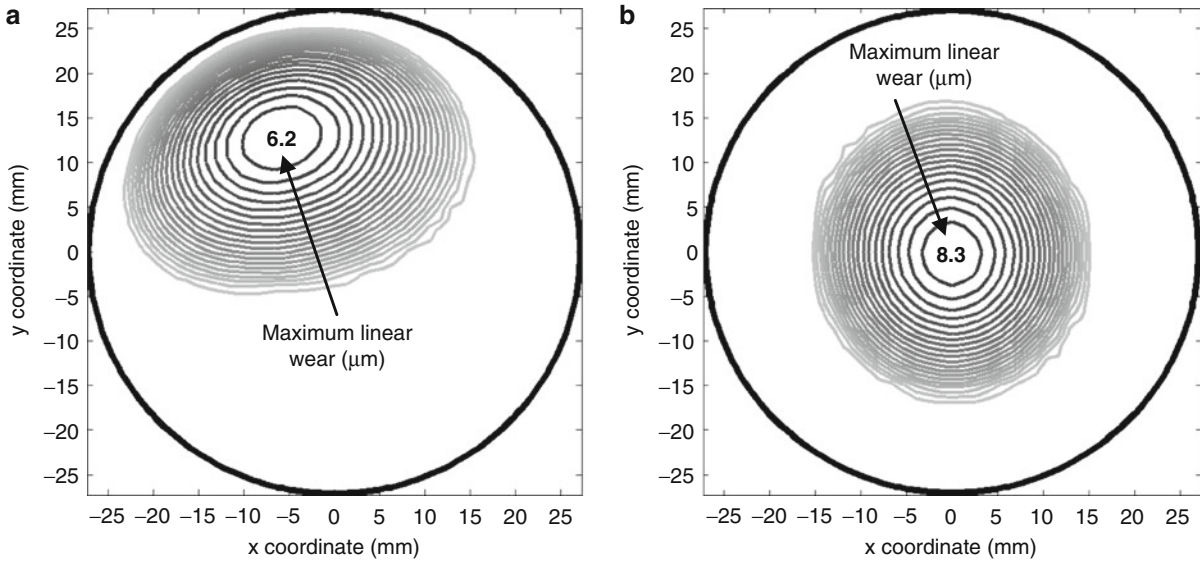
Artificial hip joints consisting of metallic/ceramic femoral heads articulating against metallic acetabular cup are used as low-wear bearings and an alternative to UHMWPE bearings (Fisher et al. 2006). One of the



Wear Modeling of Artificial Hip Joints, Fig. 3 (**a**) Loading force and (**b**) motions of a simulated gait cycle corresponding to the hip joint resurfacing replacement shown in Fig. 1 (**a**) and (**b**) (Taken from Liu et al. 2008)



Wear Modeling of Artificial Hip Joints, Fig. 4 Comparison of the predicted volumetric wear between the hip simulator testing and computational modeling (Taken from Liu et al. 2008)

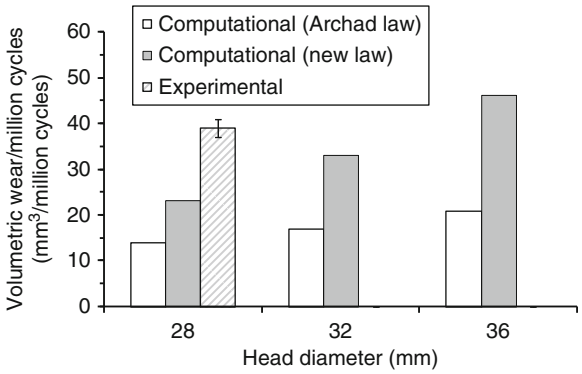


Wear Modeling of Artificial Hip Joints, Fig. 5 Computationally predicted wear depth distributions on (**a**) the cup bearing surface and (**b**) the head bearing surface of the hip joint resurfacing replacement under simulated gait cycle conditions after 50 million cycles (Taken from Liu et al. 2008)

metal-on-metal (MOM) bearings is resurfacing replacements, with the femoral bearing surface being replaced rather than the whole head. This bearing is developed for longer term use, particularly for younger, more active and demanding patients. The simulator testing of artificial hip joints is generally carried out for three to five million cycles, representing clinical follow-up of 3–5 years. The resurfacing joints are expected to last for up to 50 years, equivalent to 30–50 million cycles, which is time-consuming and impractical. The computational model is more suitable for long-term simulation (Liu et al. 2008). Figure 2a shows the resurfacing joints tested in a hip simulator and the 3D representation of the bearings with motion and loading applied in Fig. 2b.

The corresponding profiles of the load and motions of a simulated walking gait cycle are given in Fig. 3a, b, respectively.

The wear of the MOM resurfacing bearings is modeled with Archard's law using different wear factors. The FE contact model created in a commercial FE package, ABAQUS (version 6.5–1, ABAQUS Inc., Rhode Island) is used to obtain the contact pressures on both the cup and head bearing surfaces. Wear factors of 1.13×10^{-8} and 1.20×10^{-9} $mm^3N^{-1}m^{-1}$ are determined based on the simulator results and used for the bedding-in and steady-state wear phases in the modeling, respectively. The results of volumetric wear and linear wear depth distributions on the bearings of the simulation extended to 50 million cycles are shown in Figs. 4 and 5, respectively. Both the volumetric wear and wear depth distributions computationally predicted agree with those of the simulator testing results.



Wear Modeling of Artificial Hip Joints, Fig. 6 Wear rates predicted using different wear laws, simulator testing, and different femoral head sizes of the UHMWPE hip joint bearings under simulated gait conditions

Metal-on-UHMWPE Bearings

The use of the Archard equation for UHMWPE hip joints results in the wear factor being a strong function of

contact pressure and hence a wide range of wear factor values (Barbour et al. 1997). This has been further complicated by incorporating the cross-shear effect of the multidirectional motion of the hip joints into the wear factor (Kang et al. 2009), the wear factor being curve fitted as a function of both the *CSR* and contact pressure:

$$K(CSR, p) = e^{(-13.1+0.19\ln(CSR)-0.29p)} \quad (10)$$

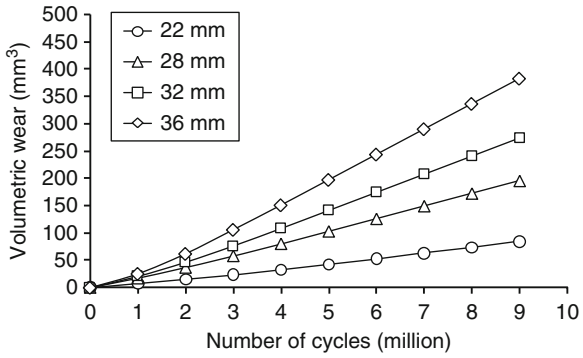
The use of the alternative wear equation leads to a relatively constant wear coefficient over the contact pressure considered for UHMWPE hip joints, and hence the wear coefficient being a function of *CSR* only. For example, a representative wear coefficient can be determined only as an explicit function of *CSR* based on the result of the representative pin-on-plate tests given by curve-fitted function as follows:

$$C = (32.0CSR + 0.3) \times 10^{-9} \text{ when } CSR < 0.04 \quad (11a)$$

$$C = (1.9CSR + 1.6) \times 10^{-9} \text{ when } 0.04 < CSR < 0.5 \quad (11b)$$

Figure 6 shows the volumetric wear rates predicted using Archard's law and the new law as well as a representative simulator testing for different head sizes. The prediction with the new wear law improves the agreement with that of simulator testing and with that of the clinical wear measurement, particularly for the larger head sizes. Figure 7 shows the increase in volumetric wear associated with the larger head size resulting from the larger sliding distance and the contact area.

Future modeling may include the effects of lubrication on the wear factors, different materials such as cross-linked UHMWPE and ceramics, patient-specific models, and more diverse wear scenarios such as edge loading occurring during hip joint microseparations and other joint application such as spinal disc and ankle replacements. The dynamic loading may be necessary in determining the wear coefficients for UHMWPE instead of constant loading used in the above formulation. Some other factors that may affect the surface deformation and lead to differences in UHMEPE wear prediction, such as the effect of polyethylene creep known to increase the surface worn area, material properties such as Young's modulus, and heat effect, should also be considered in future modeling work.



Wear Modeling of Artificial Hip Joints, Fig. 7 Volumetric wear predicted for the UHMWPE hip joint bearings for different femoral head sizes (head diameters 22 ~ 36 mm) under simulated gait conditions over nine million cycles

Cross-References

- ▶ Biotribological Joint Simulation System
- ▶ Introduction to Biotribology
- ▶ Sliding Wear
- ▶ Testing of Artificial Hip Joints
- ▶ Wear Modeling in Artificial Knee Joints

References

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

P.S. Barbour et al., The influence of stress conditions on the wear of UHMWPE for total joint replacements. J. Mater. Sci. Mater. Med. **8**(10), 603–611 (1997)

D.L. Bartel et al., The effect of conformity, thickness, and material on stresses in ultra-high molecular weight components for total joint replacement. J. Bone Jt. Surg. (Am) **68**(7), 1041–1051 (1986)

J. Fisher et al., PRESIDENTIAL GUEST LECTURE: tribology of alternative bearings. Clin. Orthop. Relat. Res. **453**, 25–34 (2006)

M.N. Harun et al., Long-term contact-coupled wear prediction for metal-on-metal total hip joint replacement. Proc. IMechE Part J J. Eng. Tribol. **223**(7), 993–1001 (2009)

L. Kang et al., Quantification of the effect of cross-shear on the wear of conventional and highly cross-linked UHMWPE. J. Biomech. **41**(2), 340–346 (2008)

L. Kang et al., Enhanced computational prediction of polyethylene wear in hip joints by incorporating cross-shear and contact pressure in additional to load and sliding distance: effect of head diameter. J. Biomech. **42**(7), 912–918 (2009)

F. Liu et al., Development of computational wear simulation of metal-on-metal hip resurfacing replacements. J. Biomech. **41**(3), 686–694 (2008)

F. Liu et al., A new formulation for the prediction of polyethylene wear in artificial hip joints. Proc. IMechE Part H: J. Eng. Med. **225**, 6–24 (2011)

T.A. Maxian et al., The Frank Stinchfield Award. 3-dimensional sliding/contact computational simulation of total hip wear. Clin. Orthop. Relat. Res. **333**, 41 (1996)

A. Wang, A unified theory of wear for ultra-high molecular weight polyethylene in multi-directional sliding. Wear **248**(1–2), p38–p47 (2001)

Wear Modeling of Spherical Bearings

Fengcai Wang[1,2], Lynn Niel[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]Edlan Research Centre, Bingley, UK

Synonyms

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

Definition

Wear modeling of spherical bearings allows effective computational prediction of wear of bearings under given operating using a set of methods and advanced simulation techniques. General wear modeling is helpful understanding the wear mechanisms of both spherical and aspheric bearings. Wear modeling of spherical bearings also helps to develop methodologies and techniques that are employed to simulate the sliding wear process and related tribological problems. Wear modeling of real spherical bearings is a complex issue in tribology; for example, the complex or changing geometry of spherical bearings with either macroscopic or microscopic bearing geometry features needs to be represented by the fixed-tracked method (FTM), developed on the basis of the spherical grid data model (SGDM). On the other hand, a number of factors, such as lubricant, asperity contact, lubrication, geometry, material mechanical property, operating conditions, and so forth, play a significant role in the sliding wear process of spherical bearings in motion. Therefore, further study of effects of the factors related to the sliding wear process is also addressed by the development of wear modeling. The general methods and advanced numerical techniques for wear modeling of spherical bearings can be applied not only for wear prediction in integrated tribology simulation systems for spherical bearings but also for the study of other tribological problems.

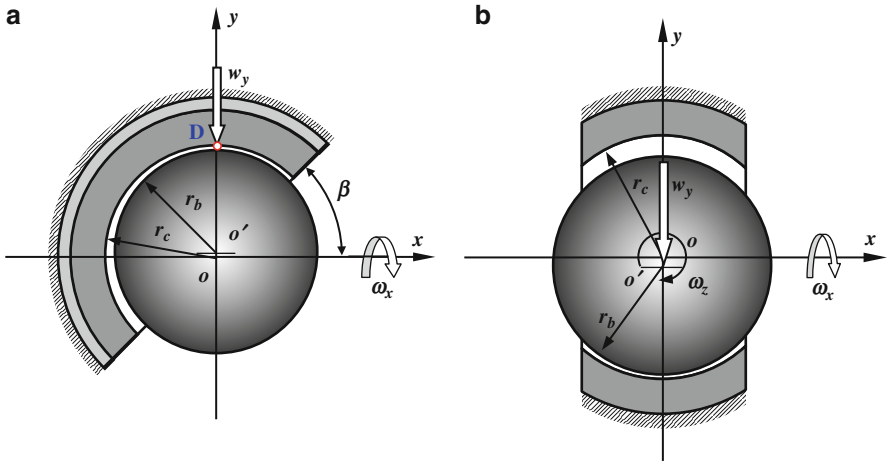
Scientific Fundamentals

Wear often takes place in the interface of bearing in motion under given operational conditions. Spherical bearings are widely applied in industrial engineering and bioengineering, for example, spherical bushings, hip joint replacements, and spherical joints (Flores et al. 2006; Scholes et al. 2000, Wang et al. 2009a, b). Significant progress in tribological modeling of spherical bearings involving friction, wear, and lubrication has been achieved in the last decade. Some subjects related to this progress, such as contact mechanics, deformation, and corresponding methodology and advanced numerical techniques, have been significantly developed. Throughout this period, wear modeling of spherical bearings with hard-on-hard bearing material combinations and under multi-directional transient motion and dynamic loading conditions has also been built with the support of a group of novel methodologies and advanced simulation techniques such as the fixed-tracked method (FTM) and the smooth spherical inverse filter method (SIFM) (Wang et al. 2005, 2008b, 2009a, b). These play an important role in the development of the tribology system of spherical bearings. For example, both dynamic contact mechanics and computational wear prediction of spherical bearings with geometrical modifications of two bearing surfaces due to wear with time can be simulated. Furthermore, an aspheric bearing with complex geometry and motion can be resolved by the regular spherical grid data model (SGDM) (Wang et al. 2010; Wang 2010, 2011). More details of wear modeling and methodology of spherical bearings is explained below.

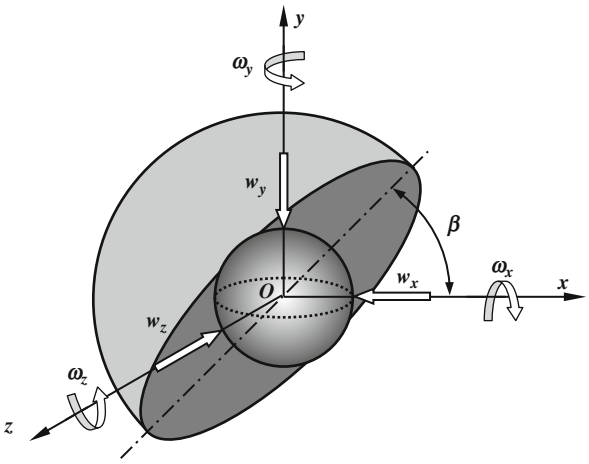
Wear Model and Methodology

Wear Model

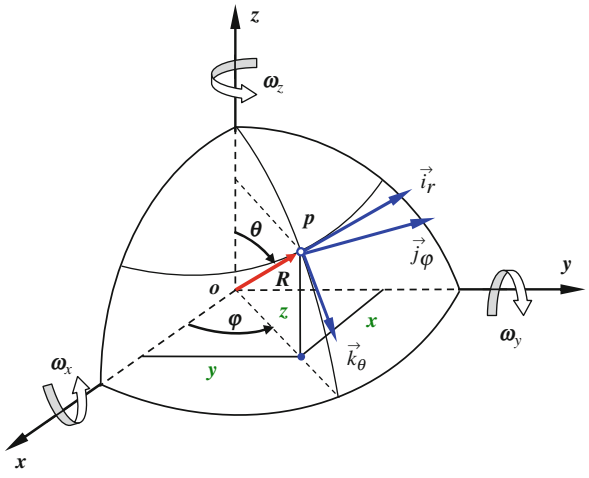
Wear may often occur in spherical bearings with various bearing structures or bearing surfaces, including aspheric bearings and spherical bushings. A ball-in-socket configuration for representing the spherical bearing, either with full or partial spherical bearing surface, can be widely found in engineering and bioengineering, as shown in Fig. 1a, b (Wang et al. 2005, 2008a, b, 2009a, b, 2010). For example, the bearing configuration as shown in Fig. 1a is often used to describe the spherical head against the spherical cup in hip joint implants, while the spherical bushing shown in Fig. 1b represents the spherical joints or spherical bearing. A general spherical bearing model under three-dimensional loading and Euler angular motion is used for the study of wear modeling and methodology, as shown in Fig. 2 (Wang et al. 2008a, b). An appropriate spherical coordinate system with the z-axis



Wear Modeling of Spherical Bearings, Fig. 1 Spherical bearing model: (**a**) hemi-spherical bearing against spherical bearing and (**b**) spherical bushing against spherical bearing



Wear Modeling of Spherical Bearings, Fig. 2 Spherical bearing model with three-dimensional loading and multi-directional motion



Wear Modeling of Spherical Bearings, Fig. 3 Spherical coordinate system with the *z*-axis through the polar points of N and S

through the polar points is selected, as shown in Fig. 3. Thus, an inclination angle of spherical bearings, such as the physiological positioning of hip joint implants, is readily given in the corresponding wear modeling by the spherical coordinate $\varphi = \beta$ in the meshed bearing surfaces, as shown as Fig. 4a, b (Wang et al. 2005). The mesh grids of spherical bearing surfaces are also adopted for the study of tribology of spherical bearings, such as contact mechanics, lubrication, and wear modeling, as described in the following sections.

The wear problem of spherical bearings due to the sliding process on bearing surfaces under the loading condition can be described by the Archard wear equation as follows (Archard 1953; Wang et al. 2010):

$$V = kFs \tag{1}$$

where V is the volumetric wear (mm^3), k is the wear factor ($mm^3N^{-1}\ m^{-1}$), F is the overall loading (N) on the interface between the contact surface couple brought into contact, and s is the sliding distance (m) of the surfaces in contact. The original Archard wear equation, involving the two asperities in contact, can be also given by the simple formulation with regards to a point on the contact area as

$$w = kps \tag{2}$$

where w represents the linear wear or the linear wear depth and p is the contact stress or contact pressure. This wear equation can predict wear accurately for most of materials, but still depends on the given wear factor. The wear factor can be determined through laboratory or test, where the simple wear model is employed to evaluate it as follows:

$$k = \frac{V}{\sum Fs} \tag{3}$$

In fact, the wear factor used for wear prediction of spherical bearings is often evaluated by matching the volumetric wear obtained from testing with the predicted results from several iterations of wear modeling, which is called the trial and error method (Hu et al. 2011; Wang et al. 2010).

Representation of Bearing Geometry

A challenging problem in wear modeling of spherical bearings is how the bearing geometry changes due to wear with time under the multi-directional motion condition, such as in the three-dimensional Euler angular motion, which is effectively described as required for the numerical simulation space in the spherical coordinate system at each instant during the operational cycles (Wang et al. 2010). For example, a spherical bearing with a hard-on-hard material combination, such as metal-on-metal bearing hip joint implants or steel-on-steel spherical bearing joints, often work under the multi-directional transient motion and dynamic loading condition; thus, wear always takes place on the two bearing surface couple. In the wear simulation process, the meshed bearing geometries in motion are represented accurately with the regular mesh grids in the simulation space where each simulation can be carried out (Wang et al. 2008a, 2009b, 2010).

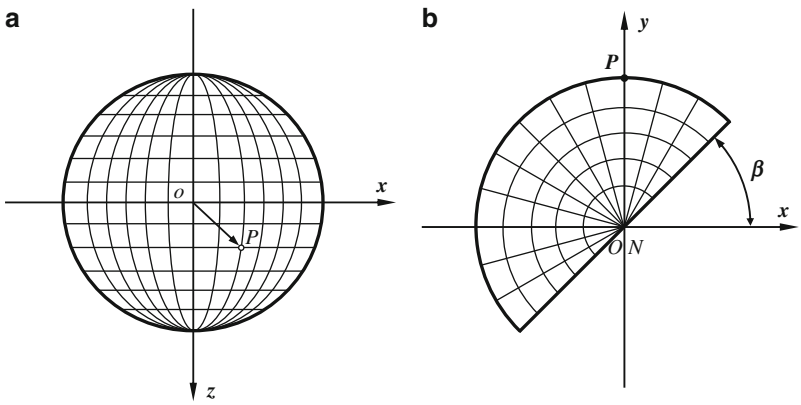
The fixed-tracked method (FTM), which is widely employed for the study of tribological problems with complex bearing geometry or combined with complex motion, had made significant progress in complex tribological modeling due to representation of interfacial geometry in motion (Wang et al. 2009a, b; Wang 2010). The method is explained in the following section.

The geometry of spherical bearings in an appropriate spherical coordinate system, as shown in Figs. 2 and 3 (Wang et al. 2009a, b, 2010), is given by

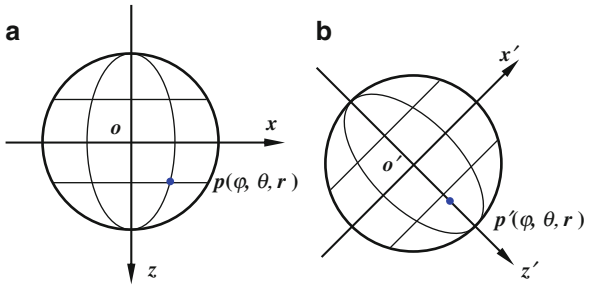
$$r_i = r_{i0}(\varphi, \theta) + \Delta r_i(\varphi, \theta) + \Theta_i(\varphi, \theta) + \delta_i(\varphi, \theta) \tag{4}$$

where r_{i0} represents the radius of a nominal spherical bearing surface, Δr_i is the deviations of an aspheric bearing relative to the nominal spherical surface, Θ_i is the micro-geometrical change, such as roughness or machining marks, and δ_i is the deformation. Elastic deformation is evaluated by the equivalent discrete spherical convolution model (EDSC) and the spherical fast Fourier transform technique (SFFT) (Wang 2004; Wang et al. 2005, 2009a).

To carrying out numerical simulations in the space coordinates, a point on the regularly meshed geometry of a spherical bearing, as shown in Fig. 4, can be described by the spherical coordinate $p(\varphi, \theta, r)$, while in the body coordinates for tracking the geometry or the changing geometry in any motion, the set of regularly meshed grids is still adopted with the corresponding spherical coordinate $p'(\varphi, \theta, r)$ to describe the geometry. The points on the meshed geometry in the space and body coordinates is one-to-one correspondence between the two sets



Wear Modeling of Spherical Bearings, Fig. 4 Spherical coordinate system with the z-axis through the poles: (**a**) mesh grids, and (**b**) inclination angle β



Wear Modeling of Spherical Bearings, Fig. 5 Diagram of spherical mesh grids: (**a**) the fixed mesh grids in space coordinates for wear simulations, and (**b**) mesh grids in body coordinates for tracking geometry in motion

of mesh grids because of mesh grid motion, as shown in Fig. 5. It should be pointed out that the two mesh points on the two sets of mesh grids of spherical bearings have the same or similar coordinate value either in space or body coordinates, respectively. However, the coordinate values at the same point on the spherical bearing surface at each instant during motion for wear modelling are different in space and body coordinate systems, as shown in Fig. 6 (Wang et al. 2009a, b, 2010). Therefore, a general methodology for representing bearing geometry in motion must be developed in order to build the relation of the two points as below.

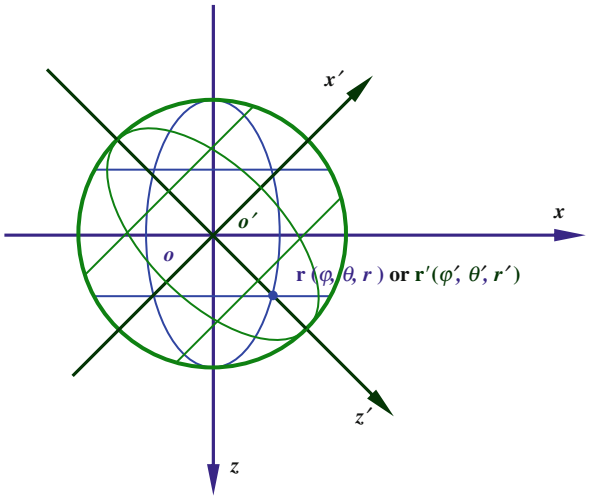
Generally, the transition from data information on the meshed geometry of bearings in body coordinates to that in space with common origin of coordinate systems can be accomplished by means of an orthogonal transformation of [A] (Wang et al. 2009a, b, 2010). The orthogonal transformation has the directional cosines of an internal body set of axes relative to an external space set of axes, as shown in Fig. 7. Thus, the location vectors of $\mathbf{r}(\varphi,\ \theta,\ r)$ and $\mathbf{r}'(\varphi',\ \theta',\ r')$ at a point on the bearing surface in space and body coordinate systems, as shown in Fig. 5 (Wang et al. 2009a, b), can be expressed as

$$\mathbf{r} = [\mathbf{A}]^{-1}\mathbf{r}' \tag{5}$$

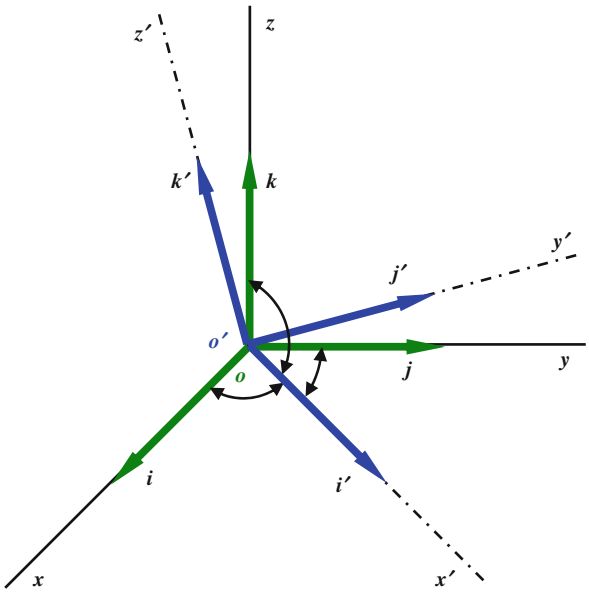
The two sets of mesh grids can be designed in space and body coordinates, one for the numerical simulation of tribological problems and the other for the representation of moving bearing geometry. Therefore, the relations between the two sets of mesh grids can be rewritten as

$$[\mathbf{r}] = [\mathbf{A}]^{-1}[\mathbf{r}'] \tag{6}$$

where $[\mathbf{r}]$ and $[\mathbf{r}']$ represent the local matrices in space and body coordinates, separately. The transformation matrix of $[\mathbf{A}]$ is given by the transposed matrix as



Wear Modeling of Spherical Bearings, Fig. 6 Coordinates of mesh grids in space and body coordinate systems



Wear Modeling of Spherical Bearings, Fig. 7 Directional cosines of the internal body set of axes relative to an external space set of axes with the common origin of coordinates

$$[\mathbf{A}] = [\mathbf{A}]^{-1} = \begin{pmatrix} \cos\beta\cos\gamma & -\sin\beta & \cos\beta\sin\gamma \\ \cos\alpha\sin\beta\cos\gamma & \cos\alpha\cos\beta & \sin\alpha\cos\beta \\ \sin\alpha\sin\beta\cos\gamma & \sin\alpha\cos\beta & \sin\alpha\sin\beta\sin\gamma + \cos\alpha\cos\gamma \end{pmatrix} \tag{7}$$

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. It is noted that the vectors of $\mathbf{r}$ and $\mathbf{r}'$ in the wear modelling process, as described in (5) and (6), respectively, should be given by their Cartesian coordinates as

$$\mathbf{r} = \begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = \begin{Bmatrix} r\sin\theta\cos\varphi \\ r\sin\theta\sin\varphi \\ r\cos\theta \end{Bmatrix} \tag{8}$$

and

$$\mathbf{r}' = \begin{Bmatrix} x' \\ y' \\ z' \end{Bmatrix} = \begin{Bmatrix} r'\sin\theta'\cos\varphi' \\ r'\sin\theta'\sin\varphi' \\ r'\cos\theta' \end{Bmatrix} \tag{9}$$

The general method of representing both the bearing geometry fixed in space coordinates and tracking the bearing geometry in body coordinates, as described in (6), is the so-called fixed-tracked method (FTM). The method has been widely employed for complex tribological simulations in spherical bearings, aspheric bearings, lubrication, friction, and contact mechanics (Wang 2010, 2011). The complex tribological problem of both real spherical and aspheric bearings with complex bearing geometry is resolved with a spherical grid data model, where a perfectly spherical bearing is used to describe the deviations between real spherical and aspheric bearings (Wang et al. 2009a, b). Therefore, the fixed-tracked method is applied not only to the study of the wear modeling involving the changing bearing geometry of spherical bearings in motion due to wear with time, but also to other tribological problems of aspheric bearing due to complexity of geometry.

Sliding Distance

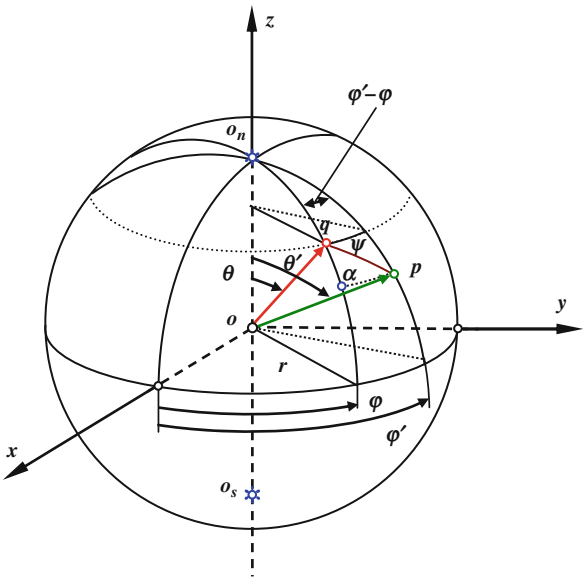
Wear of spherical bearings is related to the sliding distance in the sliding process (Maxian et al. 1996; Harun et al. 2009). If ψ shows the spherical distance between the two points of $p(\varphi, \theta)$ and $q(\varphi', \theta')$ on a unit spherical surface, as shown in Fig. 8 (Wang et al. 2010), then the sliding distance on the spherical surface is expressed as

$$\cos\psi = \cos\theta\cos\theta' + \sin\theta\sin\theta'\cos(\varphi' - \varphi) \tag{10}$$

where ψ is also equal to the central angle of arc between p and q points on the unit spherical surface. Thus, the spherical sliding distance produced between two spherical bearing couples due to the relative motion is evaluated by

$$s = R\psi \tag{11}$$

where R represents the nominal radius of the spherical bearing surface in motion, such as r_{b0} or r_{c0}, as shown in Fig. 1. The sliding distance matrix $[\mathbf{S}]$ can be completed by the evaluations of spherical sliding distances of all mesh grids on the spherical bearing surface in motion relative to the other bearing surface in wear modeling practice (Wang et al. 2009b, 2010). A case for wear modeling of spherical bearings of hip joint implants under the given operating conditions of the three-dimensional Euler motion and dynamic loading will be discussed in the Key Applications section.



Wear Modeling of Spherical Bearings, Fig. 8 Sliding distance on the unit spherical bearing surface

Contact Mechanics Model

The evaluation of contact mechanics of spherical bearings under the given operational conditions is required for the wear modeling, as described in (2). The contact mechanics simulations can be carried out either macroscopically or microscopically, depending on the requirement for the computational wear prediction. The contact problem of bearings can be made by the restriction relationship of the gap between the two surfaces brought into contact and the resultant contact pressure p in contact area Ω_c (Wang et al. 2003, 2005, 2006, 2008b). A general constrain condition for description of contact mechanics in spherical coordinates is expressed by

$$\begin{cases} g(\varphi,\theta) = 0 & (\varphi,\theta) \in \Omega_c \\ p(\varphi,\theta) > 0 & (\varphi,\theta) \in \Omega_c \\ g(\varphi,\theta) > 0 & (\varphi,\theta) \notin \Omega_c \\ p(\varphi,\theta) = 0 & (\varphi,\theta) \notin \Omega_c \end{cases} \tag{12}$$

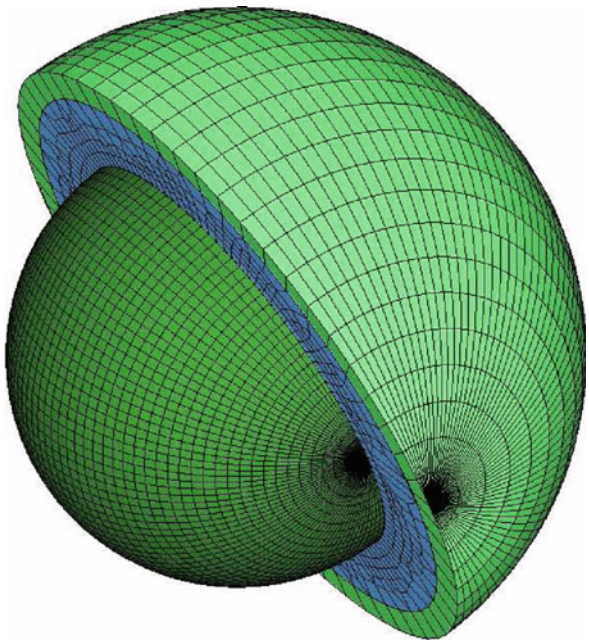
where the gap g comprises the original unreformed gap g_0 without contact pressure and the deformation caused by contact pressure (Wang et al. 2005, 2006, 2009b). The gap between the spherical bearing surface couple in contact can be described as

$$\begin{aligned} g(\varphi,\theta,t) = {} & (r_c - r_b) - e_x(t)\sin\theta\cos\varphi \\ & - e_y(t)\sin\theta\sin\varphi - e_z(t)\cos\theta \\ & + \Theta_{cb}(\varphi,\theta) + \delta_{cb}(\varphi,\theta,t) \end{aligned} \tag{13}$$

where r_b and r_c are the radial distance from a point on the inner or outer bearing surface of the real spherical or aspheric bearing to the center of the nominal spherical bearing, respectively. The variables of (e_x, e_y, e_z) are the position of the center of the inner bearing surface with reference to the center of the outer bearing surface. The variables of Θ_{cb} and δ_{cb} are the combined micro-geometrical variation and deformation of two bearing surfaces separately. The elastic deformation evaluation is made by the equivalent discrete spherical convolution model (EDSC) combined with spherical fast Fourier transform (SFFT), which can be found elsewhere (Wang et al. 2005, 2008a, b, 2009a).

The finite element method is usually employed for contact mechanics of spherical bearing, as shown in Fig. 9. Due to a limitation of the number of mesh elements, the finite element model is often applied for the analysis of contact mechanics of spherical bearings with a macroscopically geometrical scale. To evaluate the contact mechanics of spherical bearings with a microscopically geometrical scale, such as curvature, clearance, structure, and other sophisticated features, an efficient methodology with a high density of mesh grids has to be considered (Wang et al. 2005).

Significant progress has been made in contact mechanics since the smooth inverse filter method (IFM) was developed. The method can resolve contact problems macroscopically and microscopically (Wang et al. 2010). The method has been also developed into the smooth spherical inverse filter model (SIFM) of spherical contact, under the support of the elastic theory of spherical surfaces, involving the equivalent discrete spherical convolution (EDSC) model and the spherical fast Fourier transform (SFFT) (Wang et al. 2008a, b, 2009a). The spherical inverse filter method has been widely employed for the evaluation of contact mechanics of spherical contact problems, both in spherical and aspheric bearings. A high density of mesh-grids for contact mechanics can be adopted by incorporating the efficient method with the fast numerical technique to facilitate the numerically iterative process of the contact problem. Thus, the contact



Wear Modeling of Spherical Bearings, Fig. 9 Finite element model of contact mechanics of a spherical bearing

pressure of spherical bearings can be resolved by the smooth spherical inverse filter method (SIFM) with the smooth function of γ for convergent iterative procedures (Wang et al. 2005, 2009b). Contact pressure in the frequency domain with the fast Fourier transform is evaluated by

$$\tilde{p}_{ij}(f_\varphi, f_\theta) = \frac{\tilde{\delta}_{ij}(f_\varphi, f_\theta)}{\tilde{S}_{ij}(f_\varphi, f_\theta)\sin(\bar{\theta})}\ \gamma_{ij}(f_\varphi, f_\theta) \tag{14}$$

where (f_φ, f_θ) are variables in the frequency domain, corresponding to variables (φ, θ) in the time domain. The contact pressure $p(\varphi, \theta)$ is achieved after the inverse fast Fourier transform during iteration of the contact solution (Wang et al. 2005).

The resultant load due to contact pressure distribution on spherical surfaces of spherical bearings in contact is required for balance with the external load (Wang et al. 2005, 2008a, b). The resultant load balance is given by

$$\begin{cases} f_x = R_2^2 \int_{\varphi_1}^{\varphi_2} \int_{\theta_1}^{\theta_2} p\sin\theta\cos\varphi\sin\theta d\theta d\varphi = w_x \\ f_y = R_2^2 \int_{\varphi_1}^{\varphi_2} \int_{\theta_1}^{\theta_2} p\sin\theta\sin\varphi\sin\theta d\theta d\varphi = w_y \\ f_z = R_2^2 \int_{\varphi_1}^{\varphi_2} \int_{\theta_1}^{\theta_2} p\cos\theta\sin\theta d\theta d\varphi \quad = w_z \end{cases} \tag{15}$$

where (w_x,w_y,w_z) is the loading on the spherical bearing. The numerical integration is made for the evaluation of the resultant load. Once the contact pressure distribution

and the corresponding sliding distance are obtained, wear simulation can be performed, as described in the next section.

Numerical Wear Prediction

Wear modeling of spherical bearings during multi-directional transient motion and dynamic loading conditions, involving the representation of bearing geometry by the fixed-tracked method and the contact pressure distribution by the corresponding contact mechanic model as well as the sliding distance of spherical bearings, can be carried on. The linear wear depth in the radius direction of spherical bearing, evaluated by Archard's wear law, as given in (2), is rewritten as

$$\Delta r_i(t) = \sum_{t=1}^{n_t} k p_t s_t \tag{16}$$

where n_t is cycles experienced during the given operating condition. During the operating condition, the bearing geometry is modified with time due to wear (Hu et al. 2011). It is noted that modification of bearing surfaces may occur on the two bearing surfaces with a hard-on-hard material combination. The modifications at each instant during the operating cycle are found by

$$r_i^k(t) = r_i^{k-1}(t) \pm \Delta r_i(t) \tag{17}$$

where either the + or − sign presents the direction of geometry modification of bearings during wear simulation (Wang et al. 2010). Thus, when the modification caused by the variations of geometry of spherical bearings due to wear with time is performed at each instant during one operating cycle, the worn geometry of spherical bearings is projected by the fixed-tracked method in space coordinates, and the next cycle of wear simulations starts. The wear modeling is carried out by the developed in-house software. A tribology simulation system by the in-house software incorporating a group of methods and numerical techniques has been developed for solutions of tribological problems of spherical or aspheric bearings under the given operating conditions, including friction, lubrication, deformation, contact mechanics, and wear (Wang et al. 2005, 2008a, b, 2009a, b, 2010). The details of wear modeling are further explained in the next section.

Key Applications

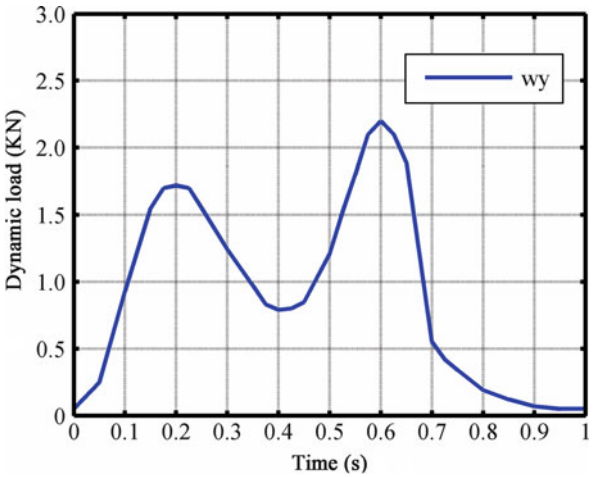
Wear modeling of spherical bearings has been applied for the study of computational wear simulation and related tribological problems in both spherical and aspheric bearings, including the corresponding methodology and advanced numerical techniques. An example of wear simulation of spherical bearings of metal-on-metal hip joint implants is run by the wear modeling in this section. The hip joint implant consists of a spherical femoral head articulating with a spherical acetebular cup and is modeled by a ball-in-socket configuration, as shown in Fig. 1 (Wang et al. 2005, 2008a, b). The mechanical properties of the metal-on-metal bearing components used in hip joint implants are shown in Table 1, while the geometric parameters of the metal-on-metal bearing components are given in Table 2. The spherical acetabular cup is

Wear Modeling of Spherical Bearings, Table 1 Mechanical properties of the metal-on-metal bearing components used in hip joint implants

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

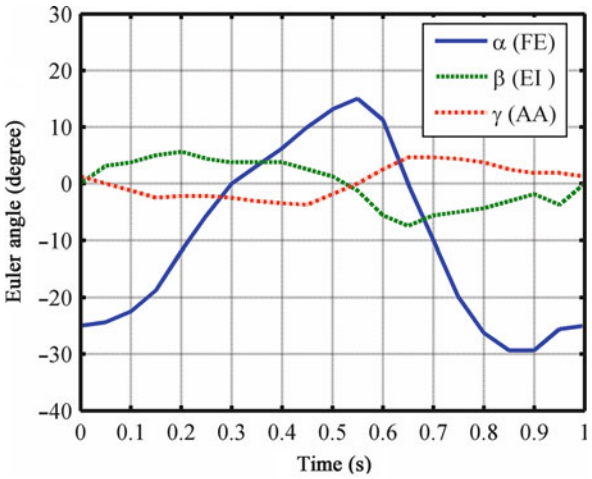
Wear Modeling of Spherical Bearings, Table 2 Geometric parameters of the metal-on-metal bearing components used in hip joint implants

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)



Wear Modeling of Spherical Bearings, Fig. 10 Dynamical load applied to spherical bearing of hip joint implants

anatomically positioned with an inclination angle of $\beta = 45°$ in clinical implants (Wang et al. 2009a). Spherical bearings in bioengineering are operated under the given walking condition of the three-dimensional Euler motion and dynamic loading, as shown in Figs. 10 and 11.

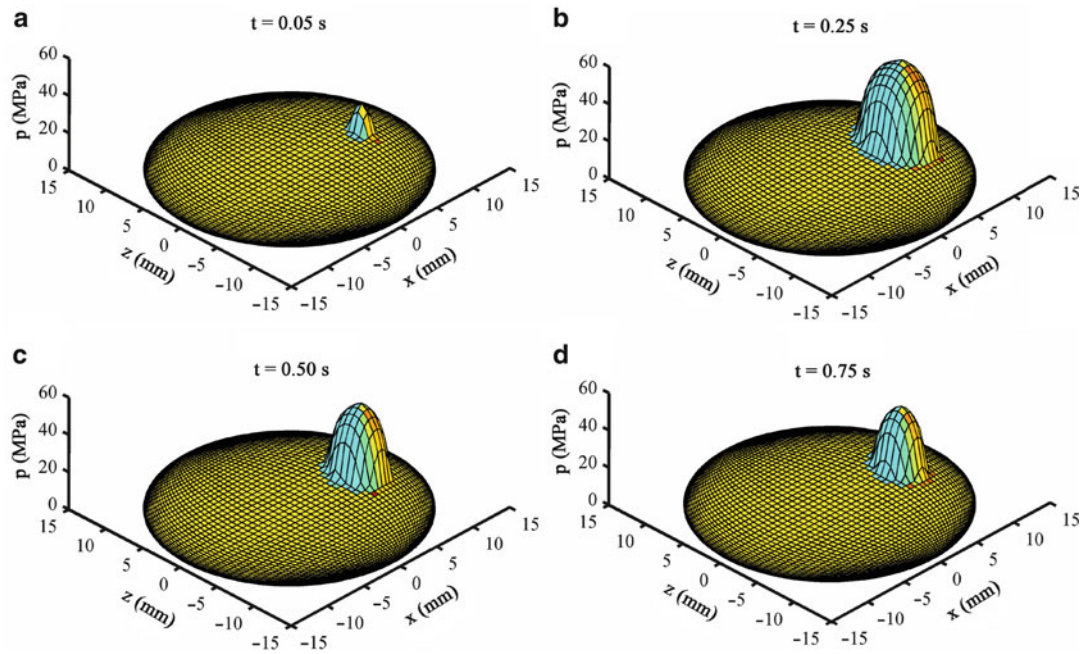


Wear Modeling of Spherical Bearings, Fig. 11 Three-dimensional Euler motion and angular positions of spherical bearing of hip joint implants

One walking cycle consists of both the stance phase and the swing phase. In this application case, this is divided into 20 instants for both the time-dependent loading and motion (Hu et al. 2011).

The wear volume of 28 mm diameter metal-on-metal bearing experienced during an experimental measurement by wear simulator is employed to obtain the wear factor for wear modeling of spherical bearings (Chan et al. 1999), as explained in the previous section. The total wear volume of both spherical head and cup bearing is classified into two parts, one for the running-in stage and the other for steady stage for the 1st year and above, respectively. Consequently, the wear factor is divided by the corresponding ones, 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. The method is employed through matching the volumetric wear within the partial stages, which is obtained by experimental simulator, with the volumetric wear by wear modeling, respectively (Harun et al. 2009; Hu et al. 2011; Wang et al. 2009b, 2010).

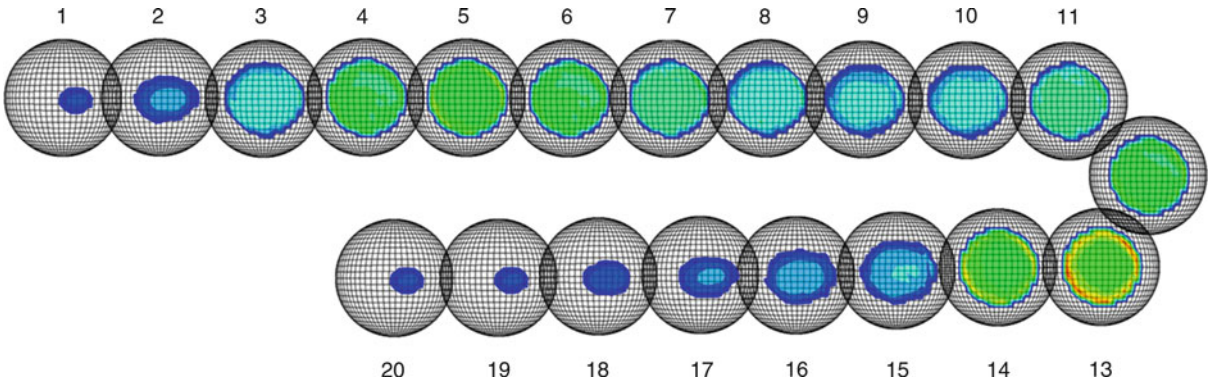
The contact pressure distribution of spherical bearings of hip joint implants is solved by the finite element model (FEM) with the density of 60×60 mesh grids in the given spherical coordinate system, as shown in Figs. 3, 4a, b, and 6. Running and updating of the FEM model in code is



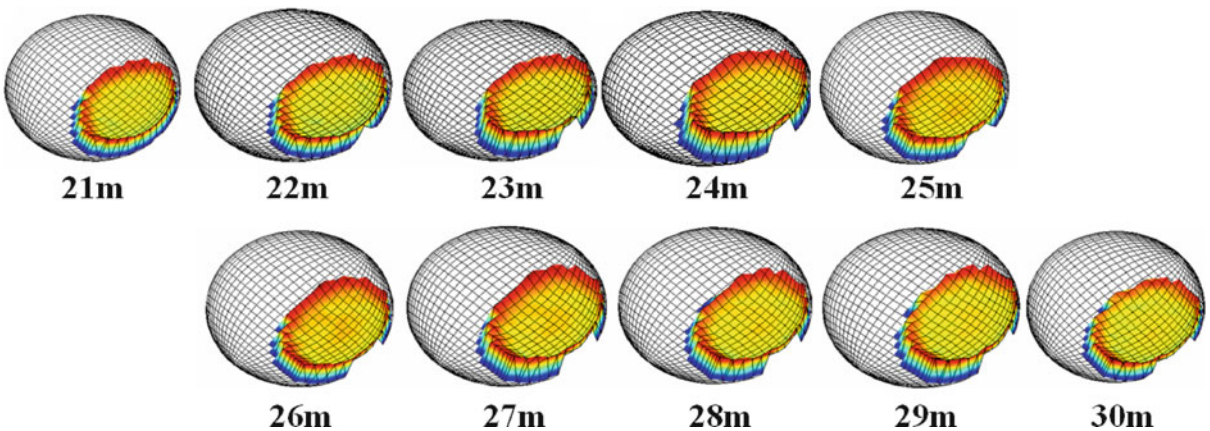
Wear Modeling of Spherical Bearings, Fig. 12 Dynamic contact pressure distributions on the spherical cup bearing surface during one walking cycle at the instants: (**a**) 0.05 s, (**b**) 0.25 s, (**c**) 0.5 s, and (**d**) 0.75 s

controlled by the wear modeling system developed by a computing program. Data communication between the fixed mesh-grids for simulating wear prediction in space coordinates and the tracked mesh-grids for describing bearing geometry in motion in body coordinates can be made by the fixed-tracked method automatically in the wear simulation system. All wear simulation works are carried out by the wear modeling of the integrated tribology simulation system by a group of methodologies and advanced numerical techniques. The calculation of the linear wear depth for both the spherical head and cup bearing surfaces is carried out after evaluation of sliding distance is performed.

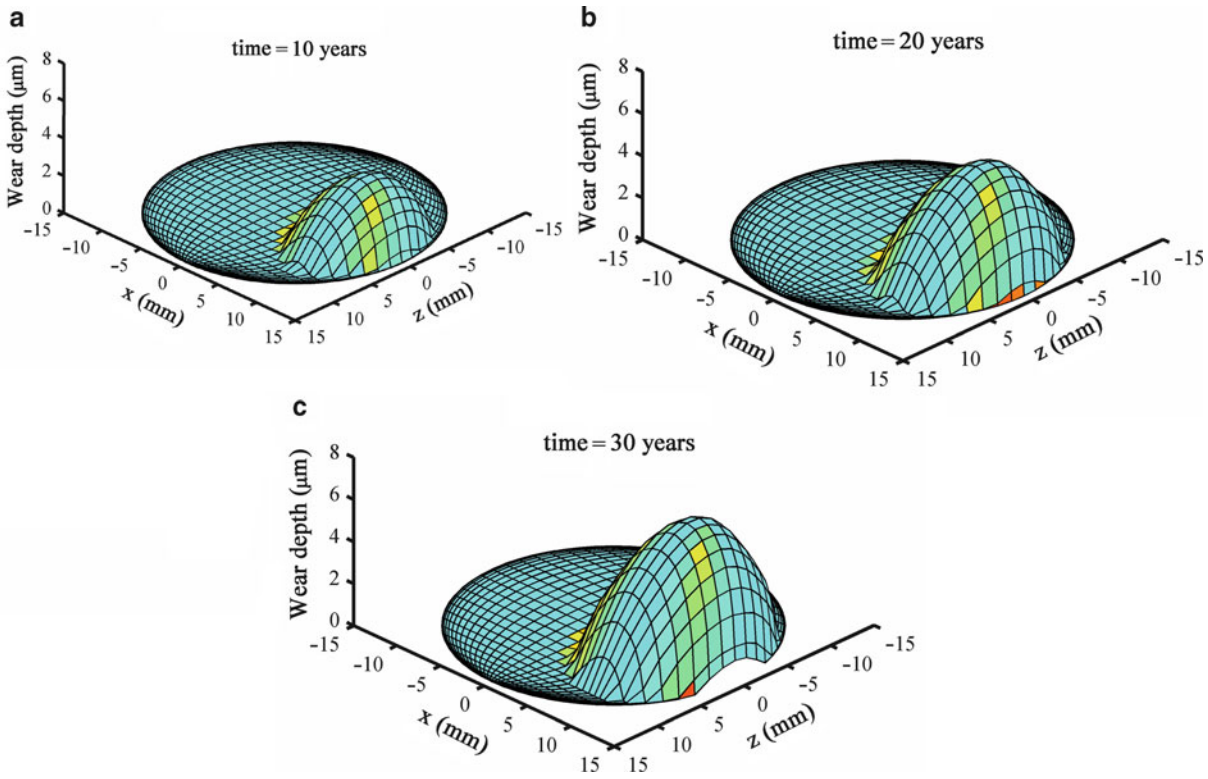
The various results produced in the computational wear simulation process are predicted by wear modeling, as described above. Some of them are given in the following descriptions. The dynamic contact pressure distributions on the cup bearing surface of spherical bearings of hip joint implants are obtained by the corresponding finite element model for contact mechanic solutions of contact stress in the wear model. Several results at the instants of 0.05 s, 0.25 s, 0.5 s, and 0.75 s during one walking cycle are presented as shown in Fig. 12a–d (Wang et al. 2009a, b, 2010). The finite element model can be carried out during the dynamic contact mechanics process in wear modeling. A micro-contact mechanics model had to be used for wear prediction of spherical bearings with micro-geometrical variations such as roughness with a high density of meshes for simulation due to the limitation of the number of elements of the element model. To resolve the problem, the smooth inverse filter method can be employed, as explained previously (Wang et al. 2005). Additionally, the contour of contact pressure distributions on head bearing surfaces of



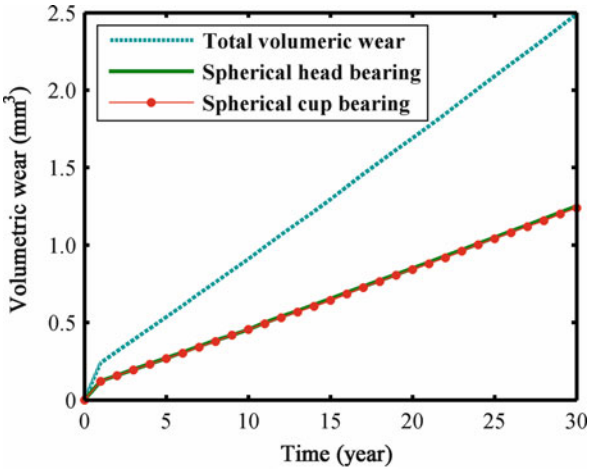
Wear Modeling of Spherical Bearings, Fig. 13 Contour of contact pressure on a spherical head bearing surface at 20 instants during walking condition of 10 million cycles



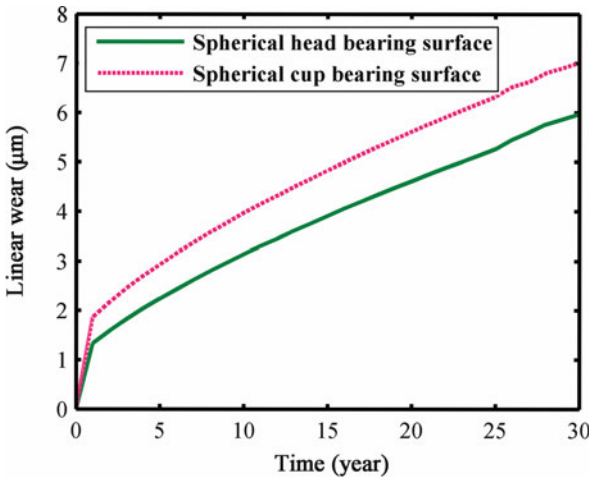
Wear Modeling of Spherical Bearings, Fig. 14 Contact pressure distributions on a spherical cup bearing surface during walking condition between 21 million and 30 million cycles



Wear Modeling of Spherical Bearings, Fig. 15 Computational wear distributions on a spherical cup bearing surface at the operating times: (**a**) 10 years, (**b**) 20 years, and (**c**) 30 years



Wear Modeling of Spherical Bearings, Fig. 16 Volumetric wear on a spherical head and cup bearing surface and total volumetric wear with time



Wear Modeling of Spherical Bearings, Fig. 17 Linear wear depth on a spherical head and cup bearing surface with time

spherical bearing of hip joint implants at 20 instants during the walking condition of 10 million cycles, in which wear occurs with time, is shown in Fig. 13. The contact pressure distributions on the spherical cup bearing surface during the walking condition between 21 million and 30 million cycles are shown in Fig. 14, where wear with cycle time approach to the edge of the cup bearing. (Hu et al. 2011; Wang et al. 2009b).

Furthermore, the computational wear prediction of spherical bearings of hip joint implants under the three-dimensional time-dependent motion and dynamic loading condition is carried out by wear modeling (Wang et al. 2009b, 2010). The linear wear distributions on the cup bearing surface in operating cycles of 10 years, 20 years, and 30 years are given, as shown in Fig. 15a, b, c. Here, one million cycles are equivalent to 1 year operating time for hip joint implants. At the same time, the volumetric wear of the head and the cup bearings, and the total volumetric wear with operating time of walking cycles within 30 years are presented, respectively, as shown in Fig. 16. Wear has obviously taken place on the edge of the cup bearing surface (Hu et al. 2011). Subsequently, the linear wear depth on the head and cup bearing surfaces and the total linear wear depth on both bearing surfaces are given, as shown in Fig. 17. Therefore, it is found that the wear modeling incorporated in the tribology simulation system was able to carry out the wear prediction (Wang et al. 2010). Further details have been published elsewhere (Wang et al. 2008b, 2009b, 2010).

For specific details of wear modeling for the study of tribological mechanisms of both spherical and aspheric bearing and the tribology simulation system, readers can refer to the following entries: ► Geometry of Spherical/Aspheric Bearings, ► Elasticity Theory for Spherical Bearings, and ► Lubrication Theory for Spherical Bearings.

Cross-References

► 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
► Sliding Wear
► Wear Modeling of Spherical Bearings

References

J.F. Archard, Contact and rubbing of flat surface. J. Appl. Phis. **24**(8), 981–988 (1953)
F.W. Chan et al., Wear and lubrication of metal-on-metal hip implants. Clin. Orthop. **369**, 10–24 (1999)
P. Flores et al., Dynamics of multi-body systems with spherical clearance joints. ASME J. Tribol. **1**(7), 240–247 (2006)
M.N. Harun, F.C. Wang et al., Long-term contact-coupled wear prediction for metal-on-metal total hip joint replacement. Proc. Inst. Mech. Eng. Part J J. Eng. Tribol. **223**(7), 993–1001 (2009)
Z.M. Hu, F.C. Wang, et al., Wear prediction of metal-on-metal hip joint replacements. J. Med. Biomech. Chinese **26**(5), 448–453 (2011)
T.A. Maxian et al., A sliding distance-coupled finite element weight polyethylene. J. Biomech. **29**(1996), 687–692 (1996)
S.C. Scholes et al., A frictional study of total hip joint replacement. Phys. Med. Biol. **45**, 3721–3735 (2000)
F.C. Wang, 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., 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*, eds. by R.W. Snidle, H.P. Evans. Solid Mechanics and its Applications, vol. 134 (Springer, Dordrecht, The Netherlands, 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, in *World Tribology Congress IV*, Kyoto (2009b)
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 Behaviour and Its Evolution of Bearing Interface with Spherical-Base Geometry*, National Natural Science Foundation of China (NSFC) Report No. 10972165, 2010, pp. 1–30
F.C. Wang, *"Dynamic Contact Mechanism and Failure of Rolling Bearing", National Key Basic Research Program of China (973 Program)* Report No. 2011CB706601, 2011

Wear of Bearings

Ryan D. Evans
The Timken Company, Timken Tech Center – North Canton, Canton, OH, USA

W

Synonyms

Damage of bearings due to wear

Definition

Wear in bearings is the removal of component surface material as a result of non-ideal operating conditions, usually attributable to the presence of contamination, inadequate lubrication, or general misuse. The removal of surface material may be mild on an asperity length scale by adhesive (two-body) or abrasive (three-body) mechanisms (Tallian 1992), or it may be macroscopic in the form of spalling or pitting damage. Although bearing design life is based typically on an expected material or subsurface fatigue limit, the effect of unanticipated wear may be to initiate damage that leads to premature failure.

Scientific Fundamentals

The majority of bearing damage in the field is attributable to wear rather than contact stress-related subsurface material fatigue. Approximately 10% of bearings achieve predicted fatigue life in application (Ludema 1996). Even if a bearing is designed, manufactured, and specified properly for an application, wear may still occur unless the potential for contamination, inadequate lubrication, and/or misuse is managed. Specific examples of these common wear modes and their causes are presented here with emphasis on wear of steel rolling element bearings, including ball, cylindrical roller, tapered roller, needle roller, and spherical roller bearings. However, the general categories of wear and examples may be applicable to plain or fluid film bearings as well. It is noted also that bearings may continue to perform satisfactorily in the presence of certain types or degrees of mild wear. However, if wear is allowed to progress without corrective action, the consequences can include excessive deflection, loss of design clearances, excessive noise and/or vibration, high torque and/or temperature, and bearing seizure (Harris and Kotzalas 2007).

Common Damage Modes

Although wear in bearings is generally defined as the undesired removal of surface material, this rarely happens as one discrete step. Instead, wear processes may require a sequence of events under specific conditions to occur before the full measure of bearing damage is detected. The wear process is usually initiated by an event that produces a surface defect site such as localized plastic deformation, a pit, a crack, a scratch/gouge, an adhesion junction between asperities, or abrasive third-body particles, for example. In rolling element bearings, continued cyclic contact stress over these defective surface regions may initiate cracks that can then propagate by fracture, leading to the removal of surface material. A few common damage modes by which the removal of surface material occurs in rolling element bearings are defined below before focusing on examples of mechanisms that may initiate them.

Spalling is a type of contact fatigue failure that occurs on contact surfaces by propagation of near-surface cracks that ultimately lead to macroscopic surface material removal and thereby the formation of craters (Tallian 1992). Spalling on contact surfaces that are subjected to cyclic Hertzian contact can be initiated by surface defects and cracks or by subsurface cracks associated with imperfections or material inclusions. Prediction of spalling initiated by the latter in the absence of surface flaws is the focus of design fatigue life models. On the other hand, surface-initiated spalling in rolling element bearings (such as point-surface origin (PSO) spalling) is mostly attributable to localized surface damage and can be much harder to predict. Careful microscopic analysis of the fracture surfaces may be required to distinguish between surface and subsurface-initiated spalling (Widner 1986).

Like spalling, pitting is the formation of surface craters or voids by loss of material. However, pits are usually smaller than spalls and have steep side walls, sharp edges, and they may be accompanied by local micro-cracks (Tallian 1992). Inherently, small pits are defects that may initiate larger spalls on surfaces in cyclic Hertzian contact. Flaking occurs when large groups of pits join together in lines or networks to deteriorate bearing contact surfaces (Widner 1986).

Peeling or micro-pitting occurs when thin flakes of material are removed from contact surfaces. The damage area is usually shallow on a microscopic scale, giving the appearance of "gray staining." Peeling or micro-pitting damage can also initiate cracks that later propagate into macroscopic fatigue spalls (Widner 1986).

Wear Related to Contamination

Bearing design algorithms, as well as many bearing performance tests, assume a clean environment to predict or assess performance. In the field, solid, usually particulate, foreign material may enter the bearing and initiate wear of critical contact surfaces. The foreign material may be residual particles from bearing manufacturing processes or it may enter into the bearing from the outside environment during operation. Particles from manufacturing can be eliminated in post-finishing cleaning prior to bearing packaging. Proper selection and use of seals can prevent ingress of debris particles from the environment around the bearing. However, occasionally particles may be generated elsewhere in an application (e.g., a gearbox) due to wear of adjacent components and it may not be possible to

seal them out. In this situation, lubricant filtration is required.

The size and type of particulate contamination in bearings directly affects the wear damage that can result. Fine hard particle contamination can remove material from bearing contact surfaces by excessive polishing like a lapping compound, changing internal bearing geometry by abrasive wear (Widner 1986). Large hard particles can become embedded into softer cage material and create groove marks in rolling element body surfaces. Bruising and denting of contact surfaces can occur when ductile particles are entrained through heavily loaded bearing contacts and plastically deform the surfaces. When a dent passes through a rolling element–raceway contact, high stress concentration may occur at the dent shoulder (or outer ridge) that can lead to crack initiation. Experimental work has shown that dents produced by large ductile particles may lead to pitting and spalling, while dents attributed to small ductile particles can lead to micro-pitting or peeling of bearing raceways (Ai and Nixon 2000). Furthermore, the extent of bearing life reduction is influenced by the slopes and area densities of surface indentations produced by debris in rolling element bearings (Ai and Nixon 2000). Depending on the size and shape of the dent, lubricant entrained into the contact may collapse into the dent depression and effectively lower the lubricant film thickness in that location (Harris and Kotzalas 2007).

Non-particulate chemical contamination of bearings can produce wear and damage by corrosion or etching. If a steel bearing is exposed to water or moisture through damaged packaging or seal failure, corrosion pits or rust can form on critical surfaces that may lead to increased friction (overheating), initiate spalling, or create particles that cause abrasive damage when the bearing is operated. Other chemical contaminants such as products from lubricant degradation or halogens from metalworking fluids can produce etching damage if they are in persistent contact with bearing components. It is critical that bearing assemblies and components be properly protected by rust preventatives in storage, and that a lubricant with appropriate rust and oxidation inhibitors is used when the bearing is in service.

Wear Related to Inadequate Lubrication

Bearings are designed to require lubrication. Whether it is by oil, grease, air, solid lubricant materials, lubricious wear debris, or transfer films, the presence of an appropriate quantity of lubricant in the proper location is a fundamental component of design life predictions. If lubricant supply is frequently interrupted, or if the lubricant properties are degraded by overheating, excessive oxidation, or poor sealing, bearing wear may occur and it is unlikely that full design life will be achieved.

Boundary lubrication results when the conditions for formation of a lubricant film to separate bearing contacting surfaces are inadequate, usually at low speeds, high loads, high temperatures, inadequate lubricant viscosity, high slide-to-roll ratio, or under lubricant starvation. Under these conditions, frequent contact between opposing surface asperities produces regions of high stress concentration localized at the surfaces, which may result in mechanical damage. Full surface separation by a lubricant film is achieved in the elastohydrodynamic lubrication (EHL) regime for heavily loaded contacts, and the transition between boundary and EHL conditions is described as the mixed lubrication regime. Prolonged bearing operation under mixed or boundary lubrication conditions is associated with peeling or micro-pitting damage (Widner 1986). Micro-pitting damage has also been observed to initiate or accompany PSO spalling in rolling element bearings. If peeling is observed uniformly across a bearing contact surface, marginal lubrication film thickness for the whole bearing is to be suspected. On the other hand, if peeling or micro-pitting is non-uniform on a contact surface, it is possible that localized boundary lubrication around a scratch or some other defect is the cause (Widner 1986). The potential for micro-pitting and peeling wear may be minimized by ensuring proper lubricant film separation of contacting surfaces in bearings.

A severe wear mode in bearings associated with operation in the boundary lubrication regime is adhesive wear. Adhesive wear occurs when contacting asperities bond in intimate contact (similar to cold welding) and tear. It is usually accompanied by sliding contact and high temperatures, and results in plastic deformation of the surfaces and the transfer of material from one surface to the other. A milder form of adhesive wear is smearing, in which roller-raceway slip in boundary-lubricated rolling element bearings produces plastic deformation at the asperity scale that may be accompanied by material transfer from one surface to the other. More severe wear modes that may involve adhesion mechanisms include scoring, scuffing, or galling. Scoring is adhesive wear that may occur upon loss of lubrication at contacts with a large sliding component, such as the rib-roller end contact in tapered roller bearings. Heat damage or discoloration of the wear zone is frequently observed to accompany scoring. The term “scoring” has also been used to describe macroscopic

gouging by an abrasive wear process (Tallian 1992), so caution must be exercised in determining whether adhesive or abrasive wear occurred if damage is described by that term. Scuffing is associated with plastic deformation of asperities and surface roughening, potentially including localized adhesive wear damage as well (Ludema 1996). Like the smearing and scoring cases, the extent of adhesion in scuffing is not so great that two scuffed components in contact cannot be separated. Scuffing may occur in a plain bearing if the fluid film is interrupted and the bearing surfaces come in sliding contact under boundary lubrication conditions, for example. Galling is described as severe adhesive wear with significant plastic deformation of surfaces and material transfer, which may result from gross roller sliding in the presence of a cage failure, for example. Often, two components that have galled together are not easily separated due to significant welding of the worn surfaces. Adhesive wear can usually be prevented by operating bearings under appropriate fluid-film lubrication conditions, but as in the galling example given above, it may occur secondarily as a consequence of some other damage mechanism and therefore complicate post mortem root cause analysis.

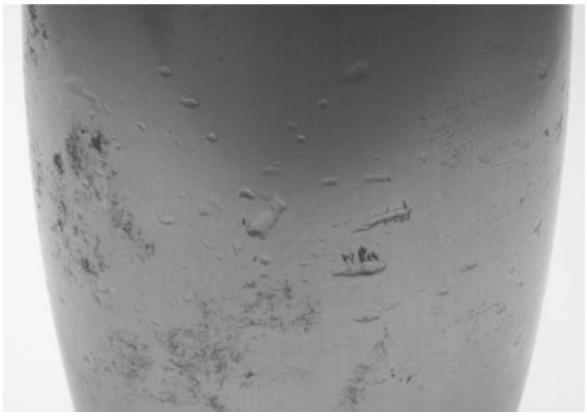
False brinelling can occur in rolling element bearings at rolling element–raceway contacts that are subjected to low-amplitude oscillating motion, usually produced by sustained machine vibration or during machine transportation. Under these conditions, a lubricant film is not maintained between contacting surface asperities and localized high friction and adhesive wear interactions lead to material oxidation. Subsequent micro-sliding removes or disrupts this oxide layer, and the cycle repeats. In this regard, false brinelling is a type of friction oxidation or fretting wear (Widner 1986). It is so-called because the resultant damage at roller-spaced increments on the raceway can appear to be similar to true brinelling, which is produced when rolling element–raceway contacts are overloaded beyond the material yield limit and dents spaced according to the rolling element spacing are produced in the raceways. True brinelling is distinguished from false brinelling in that raceway finish marks are preserved in the dented areas, indicating purely plastic flow. On the other hand, false brinelling is usually accompanied by loss of surface finish marks in the worn areas and the presence of reddish or black particulate debris (iron oxides, in steel bearings). False brinelling wear may be minimized or prevented by proper bearing preloading practices (if applicable), isolation from machine vibrations, modification of operating clearances, or the use of lubricants containing additives that provide fretting damage protection (Harris and Kotzalas 2007).

Wear Related to Misuse

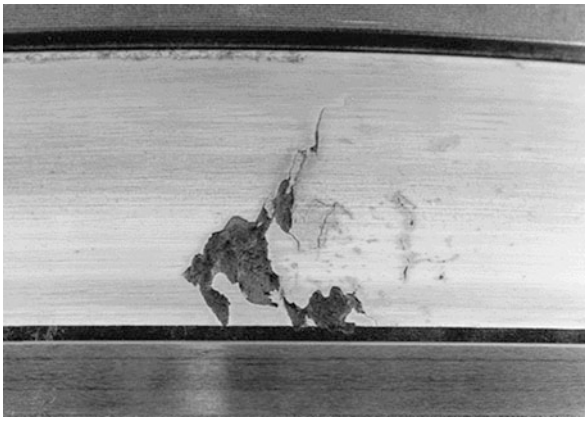
Bearings are designed to accommodate specific load and speed ranges. However, if the bearing is installed with excessive pre-load or is overloaded during operation, loss of surface integrity and cracking (i.e., peeling) may occur at minimum or gross fracture of bearing components may occur at worst. In case-carburized steel components, subcase fracture may occur under overload conditions. Under impact or shock loading conditions, bearing rolling elements can plastically deform the raceway surfaces at the roller-spacing intervals via true brinelling. Subsequent bearing operation and rolling element contact over true brinelling indentations can initiate raceway spalling. To minimize the potential for overloading, thorough application load analysis must be accompanied by proper bearing selection (Ludema 1996).

Wear damage to bearings can also result from poor installation, misalignment, high spots, and incorrect fitting practices. Misalignment of bearing rings on a shaft or within a housing can produce evidence of irregular roller path or wear tracks on the raceways and/or geometric stress concentration spalling (GSC). GSC damage usually has the appearance of spalls that start at a raceway edge and propagate inward toward the center of the raceway. High spots may result from improper bearing, shaft, or housing manufacturing or installation and can lead to uneven wear tracks on bearing raceways and ultimately spalling damage. Incorrect fitting may lead to improper preload conditions and result in spalling or other evidence of wear. Fretting can occur between bearing rings and shafts or housings as a result of insufficient press fit. For example, if the inner ring fit on a shaft is insufficient to prevent intermittent spin motion of the ring on the shaft with bearing operation, fretting damage between the inner ring bore and shaft outer diameter surface may result. Wear attributed to improper installation and fit may be prevented by proper bearing design and/or selection and proper installation practices.

If a bearing is improperly exposed to electric current, such as is possible in an electric motor, pitting, burning, or fluting damage may occur. Fluting damage has the appearance of corrugation marks on the raceway with a spacing that is dependent on the electric current frequency and internal bearing dynamics (Harris and Kotzalas 2007). Other potential sources for electric current exposure are improper grounding of equipment, welding damage, or static discharge issues. Material removal and pitting damage resulting from electric current damage may lead to subsequent spalling and failure. This wear mode can be prevented by proper electrical isolation or grounding of bearings in application.



Wear of Bearings, Fig. 1 Example debris-dented roller body due to particle contamination in the bearing



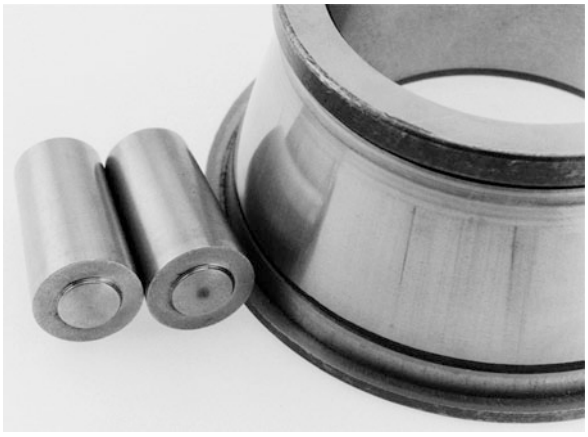
Wear of Bearings, Fig. 2 Example point-surface origin (PSO) spall on a tapered roller bearing raceway. The rolling direction was left to right, and the arrowhead-shaped leading edge is apparent on the left side of the spall

Key Applications

Photographs of example bearing wear and damage are given in this section, again with an emphasis on steel rolling element bearings (The Timken Company 1994). However, many of the general principles can be applied directly to plain or fluid-film bearings as well.

Examples of Wear by Contamination

An example of denting or bruising damage on a roller body is shown in Fig. 1 that may be attributed to hard ductile particle contamination. Although the denting process deforms rather than removes surface material, each dent is a potential initiation site for cracks that may initiate point-surface origin (PSO) spalling wear with



Wear of Bearings, Fig. 3 Excessive abrasive wear at the rib-roller end contact due to hard fine particle contamination

Wear of Bearings, Fig. 4 Corrosive wear in a tapered roller bearing due to contact with water or moisture

Wear of Bearings, Fig. 5 Over temperature damage of a tapered roller bearing resulting from inadequate lubrication

continued bearing operation. The example PSO spall shown in Fig. 2 was not initiated by debris denting, but it is included to show the characteristic arrowhead leading-edge appearance for these types of surface-initiated spalls.

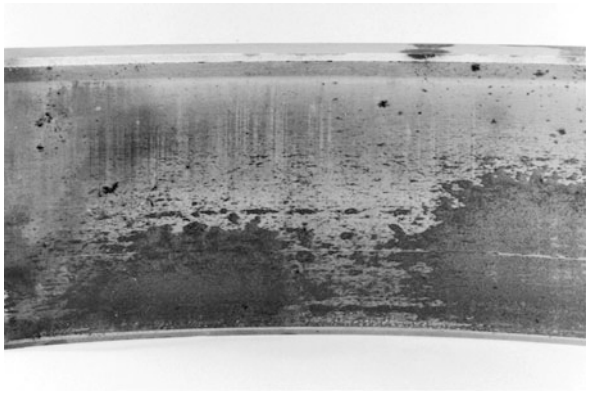
When hard fine particle contamination occurs, abrasive wear may result that produces a lapping or excessive polishing effect. Figure 3 shows an extreme example for a tapered roller bearing, in which the original roller end rib geometry was abrasively removed to the degree that it appears both surfaces were machined off. Whereas the center of the tapered roller end was originally a recessed area, it is pronounced in Fig. 3 after the contact area for the roller large end was polished down. Likewise, the original rib was mostly removed by aggressive polishing, resulting in the rollers effectively shifting down the cone body during operation and creating excessive axial clearance within the bearing.

Unintended exposure of a bearing to water or moisture may produce rust (abrasive debris) and corrosion pitting as shown in the extreme case of Fig. 4.

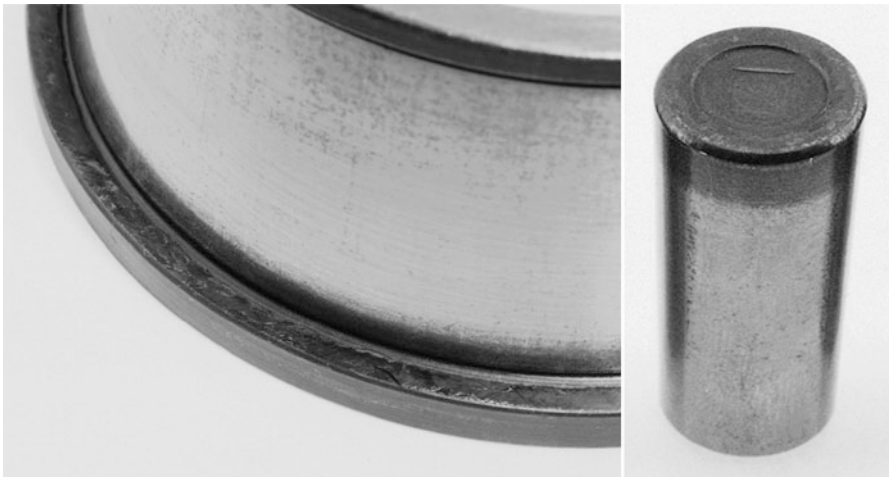
Examples of Wear by Inadequate Lubrication

Lubricant loss or degradation in a bearing application may promote high friction and excessive operating temperatures. If high temperatures are localized within a bearing, thermal expansion mismatches among the components may occur and cause the inner ring to expand more than the outer ring, for example. Inconsistent thermal expansion may reduce internal clearances and increase friction even more. This over-temperature scenario initiated by insufficient lubrication may lead to bearing component material property degradation, including softening and yielding of components, or severe adhesive wear damage. An example of this for an overheated tapered roller bearing is shown in Fig. 5, where the rib, roller bodies, and cage are severely distorted. The onset of an over-temperature situation may be identifiable by lubricant staining of components.

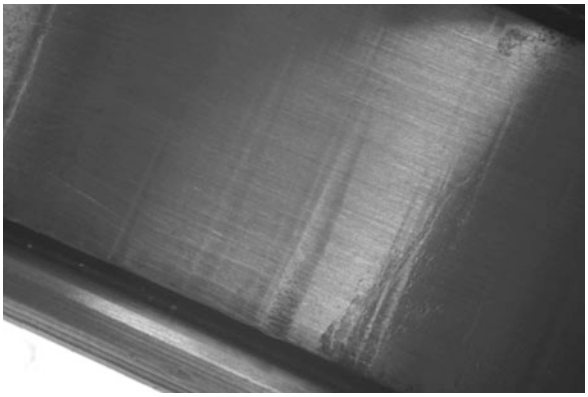
Peeling damage as shown in Fig. 6 is associated with lack of sufficient lubricant film thickness in a bearing contact, that is, under boundary or mixed lubrication conditions. The damage areas are very shallow and have a gray, frosted appearance. Insufficient lubricant film thickness in a bearing may result from high temperatures,



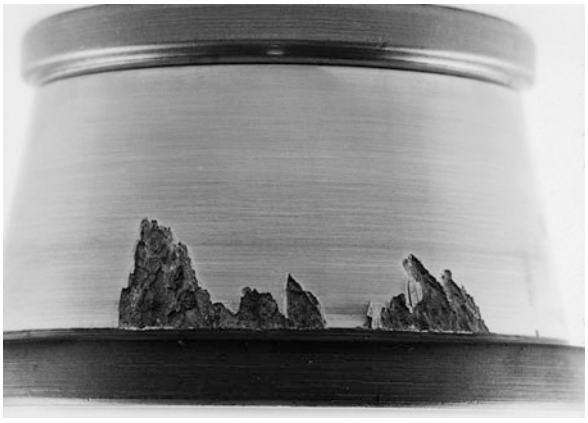
Wear of Bearings, Fig. 6 Peeling damage on a raceway attributed to insufficient lubricant film thickness



Wear of Bearings, Fig. 7 Rib-roller end scoring damage in a tapered roller bearing attributed to loss of lubricant, or oil-off conditions



Wear of Bearings, Fig. 8 True brinelling damage on a raceway surface due to overloading



Wear of Bearings, Fig. 9 Geometric stress concentration (GSC) spalling on a raceway surface attributed to excessive edge loading from misalignment

heavily loaded conditions, or incorrectly selected or applied lubricant.

Tapered roller bearings are susceptible to rib-roller end scoring or adhesive wear if lubrication is lost during operation. As shown in Fig. 7, adhesive wear damage is observed on both the rib face and roller end, with transferred metal and high heat discoloration apparent. The rib-roller end contact is especially susceptible to this damage mode when lubrication is lost because of the significant sliding component of contact at that location. In addition, adhesive wear or smearing may persist in cylindrical or spherical roller bearings under marginal lubrication conditions when significant roller slip against the raceways is possible.

Wear of Bearings, Fig. 10 Fluting damage on a bearing raceway from improper exposure to electric current

Example Bearing Wear due to Misuse

Long dents in the axial direction of a raceway from true brinelling damage are shown in Fig. 8, most likely attributable to overloading or impact loading. Notice that finishing marks are visible in the bottom of the dent, indicating plastic deformation and distinguishing it from false brinelling damage. As with debris denting, true brinelling is not technically wear because it plastically deforms rather than removes surface material. However, the dents can initiate spalling with continued bearing operation. If true brinelling occurs while the bearing is static, spalls may initiate at the brinell marks at equal roller-spaced intervals around the raceways.

If a bearing is misaligned during installation, an excessive edge loading condition may result that produces geometric stress concentration spalling (GSC) as shown in Fig. 9. The spalled region extends from the edge toward the center of the raceway. GSC spalling may also occur as a result of uneven loading from shaft or housing deflections.

Electric current damage to a bearing raceway is shown in Fig. 10. As current passes through the bearing, burning and arcing can occur at the contact surfaces. If this happens while the bearing is rotating, fluting damage can result, as shown in Fig. 10, with a corrugated-type appearance. Proper grounding and electrical isolation design are required to prevent this wear mode.

References

X. Ai, H.P. Nixon, Fatigue life reduction of roller bearings due to debris denting: Part II – experimental validation. Tribol. Trans. **43**(2), 311–317 (2000)

T.A. Harris, M.N. Kotzalas, *Rolling Bearing Analysis*. Advanced Concepts of Bearing Technology, vol. 2 of 2, 5th edn. (CRC Press/Taylor & Francis, Boca Raton, 2007)

K.C. Ludema, *Friction, Wear, Lubrication: A Textbook in Tribology* (CRC Press, Boca Raton, 1996)

T.E. Tallian, *Failure Atlas for Hertz Contact Machine Elements* (ASME Press, New York, 1992)
The Timken Company, Material and photos in this section based on: The Timken Company, *Bearing Maintenance Manual for Transportation Applications*, Tapered Roller Bearing Tech Series (The Timken Company, Canton, 1994). www.timken.com
R.L. Widner, Failures of rolling element bearings, in *Metals Handbook*. Failure Analysis and Prevention, vol. 11, 9th edn. (American Society for Metals, Metals Park, 1986)

Wear of Coatings

► Solid–Liquid Bi-phase Lubricating Coatings

Wear of Dental Materials

► The Tribology of Dental Materials

Wear of Fastener

► Contact, Friction, and Wear of Threaded Fasteners

Wear of Natural Joints (Osteoarthritis)

Susan Chubinskaya[1], Markus A. Wimmer[2]
[1]Department of Biochemistry, Rush University Medical Center, Chicago, IL, USA
[2]Department of Orthopedics, Rush University Medical Center, Chicago, IL, USA

Synonyms

Degeneration of articular joints; Degeneration of weight-bearing joints; Degenerative osteoarthritis; Idiopathic osteoarthritis

Definition

Osteoarthritis (OA) is a progressive joint disease with a heterogeneous set of conditions of varied etiologies associated with altered biomechanics, which may initiate in cartilage or bone (Abramson and Attur 2009). It is a significant medical problem in today's aging population and has an impact on the lives of a growing number of patients as the life expectancy of both males and females increases. Due to partial correlation of disease progression to symptoms, it is complicated to obtain an accurate assessment of the number of people suffering from OA. According to the Arthritis Foundation, in 2010, the lives of more than 27 million Americans were affected by this serious, painful, and potentially life-altering joint disease. Diagnosis is usually based on joint pain, visible signs of joint deformity, radiographic changes, and biochemical tests that detect inflammation.

The weight-bearing joints and the spine are particularly vulnerable, and among articulating joints the knee and the hip have the highest prevalence of OA. OA develops as a result of poor intrinsic regeneration of hyaline cartilage. It is a multifactorial disease and the factors that contribute to disease onset and progression are age, genetics, injury, obesity, occupation, bone density, metabolic diseases, inflammation, and abnormal biomechanics (Abramson and Attur 2009).

Scientific Fundamentals

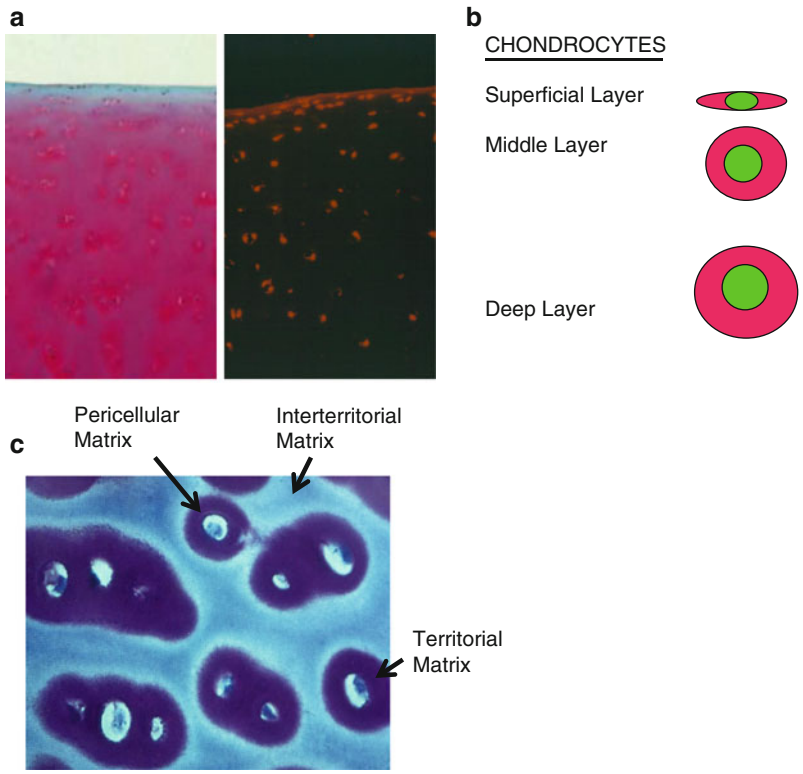
Mechanisms

Articular Cartilage Morphology and Structure

The integrity of the articular joint is dependent upon its architecture (geometry) and the health of the cartilage, the bone, and supporting structures, like menisci, ligaments, and tendons (Bullough 2004). The articular cartilage is a highly complex stratified tissue originated from mesenchymal stem cells. It covers the articulating surfaces of long bones. The main functions of cartilage are to transmit applied load across surfaces, to permit smooth articulation within the joint with lubrication and low friction, and to perform a shock absorption. In order to perform its main functions cartilage must have the high tensile strength and elasticity that is provided by the composition of the extracellular matrix (ECM) (Guilak et al. 2000). The main constituents of the ECM are water (about 65–80%) and solid components (20–35%), primarily composed of cross-linked collagen type II (10–20%), proteoglycans (4–7%), glycoproteins, and cellular components (1–10%). In adult human cartilage, the resident cells, called chondrocytes, represent about 5% of the volume. The chondrocytes are embedded in an abundant extracellular matrix rich in hydrophilic proteoglycans entrapped in a cross-linked network of collagen fibrils and matrix proteins. This solid matrix has intrinsic mechanical properties

that enable it to undergo rapid reversible deformation during loading. Under normal homeostasis, hyaline cartilage undergoes controlled turnover, with the chondrocytes being responsible for the production, organization, and maintenance of the extensive extracellular matrix. Adult articular cartilage is avascular, aneural, and alymphatic tissue, and therefore has a limited ability for self-repair and regeneration. It receives its nutrients by diffusion from the synovial fluid (Guilak et al. 2000).

Morphologically, human articular cartilage can be described as tissue with heterogenic matrix and cells (Fig. 1). There are four zones identified horizontally in adult cartilage: superficial zone representing about 10–20% of the full thickness; the middle zone, also called transitional, representing 40–60% of the tissue; the deep zone that constitutes about 30% of the entire thickness; and the calcified zone that separates the cartilage tissue from the underlying subchondral bone. The junction between the deep and calcified layers is defined by a smoothly undulating tidemark, which is critical for the transmission of load from the cartilage to the bone. Each cartilage zone is distinguished by the orientation of collagen fibers, the composition of matrix components, and by the properties of residing chondrocytes (Fig. 1). In the superficial zone, the collagen fibrils are arranged parallel to the articular surface and the amount of proteoglycans is reduced in comparison with other zones. Chondrocytes residing in the superficial zone have elongated shape; they are small in size, present in higher density, and are distributed parallel to the surface as single cells (Fig. 1). In addition, their phenotype and metabolism are distinct from other chondrocytes by expressing genes that are specific to the superficial zone, for example, superficial zone protein (Schumacher et al. 1994), and by synthesizing high levels of collagen and low levels of proteoglycans. The middle zone of cartilage is characterized by randomly arranged collagen fibrils that are less densely packed, and by reduced fibril concentration and high levels of aggrecan and water. Middle zone chondrocytes are randomly distributed throughout the matrix as single rounded cells.



Wear of Natural Joints (Osteoarthritis), Fig. 1 Structural organization of human normal adult articular cartilage. (**a**) Safranin O (*left panel*) and picrocirius red (*right panel*) staining indicate cell orientation in the superficial, middle, and deep cartilage zones. (**b**) Schematic depiction of chondrocytes shape within each of the cartilage zones. (**c**) Toluidin Blue staining of human adult articular cartilage showing deep cartilage zone indicating pericellular, territorial, and interterritorial matrices

In the deep zone, the collagen fibrils are arranged radially, perpendicular to the joint surface, and cross the tidemark to enter the calcified zone (for general reference, see Guilak et al. (2000)) allowing a stable anchorage between soft and hard tissues, specifically cartilage deep layer and hard, calcified cartilage and subchondral bone. The water content is the lowest in the deep zone, while the concentration and density of proteoglycans is the highest. The deep layer is characterized by rounded chondrocytes arranged in columns in the same vertical direction as collagen fibers. The deep layer chondrocytes form functional structural units called chondrons (Poole 1997).

In addition to horizontal matrix heterogeneity, extracellular matrix around each chondrocyte can be differentiated circumferentially into pericellular (the closest to the cell), territorial, and interterritorial (the furthest-removed from the cell or cell chondron; Fig. 1) (Guilak et al. 2000; Poole 1997). These matrices are distinct in the composition, concentration, and organization of proteoglycans and collagens. Collagen organization is also differentially arranged within these concentric subdivisions. Collagen fibers adjacent to the cell form a tightly woven (consisted exclusively of type VI collagen), densely compacted, nest-like enclosure around each chondrocyte. In the territorial matrix, collagen fibers are thicker and they form radial bundles. The interterritorial matrix is characterized by the largest collagen fibers, where their compact organization and radial alignment defines the "collagen arcades" classically described by Benninghoff (for general reference Poole (1997)). Thus, articular chondrocytes are surrounded by a complex pericellular microenvironment that, in the middle and deep layers, is integrated with a territorial matrix, separated from adjacent territories by the interterritorial matrix. This structural organization of adult articular cartilage defines biomechanical properties of the tissue and recognizes chondrocytes as key regulators of both catabolic and anabolic events necessary for cartilage homeostasis.

The Role of Biomechanics in Normal Joint Homeostasis

Diarthrodial joints are subjected to combinations of static and dynamic stresses during the day. Thereby articular cartilage acts as a cushion and shock absorber for the joint. Typical mean contact stresses are around 1.5 MPa, as can be concluded from pressure readings of cadaveric human hips that were loaded at three times body weight in the laboratory (Day et al. 1975). Two and a half times body weight is about the load expected during walking. Stair climbing and jogging can load the cartilage surface with up to three to four times the body weight. In vivo experiments with a hemi-arthroplasty have shown that contact stresses on articular cartilage surface in the hip are in the range of 5–8 MPa for level walking (Rydell 1966). However, these peak stresses are short in duration and last only a few milliseconds. Accompanying stress rates were estimated to be 6–25 MPa/s for walking, 50–210 MPa/s for jogging, and 140–250 MPa/s for jumping. In addition, diarthrodial joints undergo complex motions during loading that are a combination of rolling, sliding, and spin. The various rotations and translations cause the point of contact between opposing surfaces to move. This movement varies with joint type as well as location within the joint and plays a vital role in regulating the biosynthetic and functional responses of articular cartilage. In particular, the fluctuation of the loaded cartilage area helps to keep the interstitial fluid pressure at high levels, as outlined below.

The interstitial fluid pressure supports a majority (up to 95%) of the loads applied to the tissue. The pressurization affects many different physical, electrical, and biochemical phenomena at the cellular level, including fluid flow, streaming potentials, nutrient, and ion concentration gradients (Gray et al. 1988). This interstitial fluid support, however, is short term and decreases quickly over time (Soltz and Ateshian 1998). This may cause an increased loading of the chondrocytes and the ECM components. Prolonged, excessive loads can result in cell death, which alters the structural, functional, and synthetic properties of articular cartilage. Disease or damage of the tissue can alter the interstitial fluid support and thus the capacity of the tissue to support the load.

Limited Intrinsic Repair of Hyaline Cartilage

Articular cartilage is a metabolically active tissue. However, due to low mitotic activity of its fully differentiated chondrocytes and the structure and composition of the extracellular matrix it has a limited intrinsic reparative capacity, especially when matrix integrity is compromised. A number of studies have documented that in vitro isolated chondrocytes or chondrocytes residing in cartilage explants can produce various matrix components specific for hyaline-like cartilage, especially when stimulated with pro-anabolic mediators or physiological load (reviewed in Chubinskaya et al. (2007)). However, on the joint level or in vivo, spontaneous or induced cartilage regeneration almost does not exist or fails to reproduce tissue that resembles characteristics of native cartilage. The nature of repaired chondral tissue is fibro-cartilage, whose extracellular matrix is generally composed of a higher ratio of collagen type I to collagen type II and a lower content of proteoglycans (Khan et al. 2008). Its biomechanical

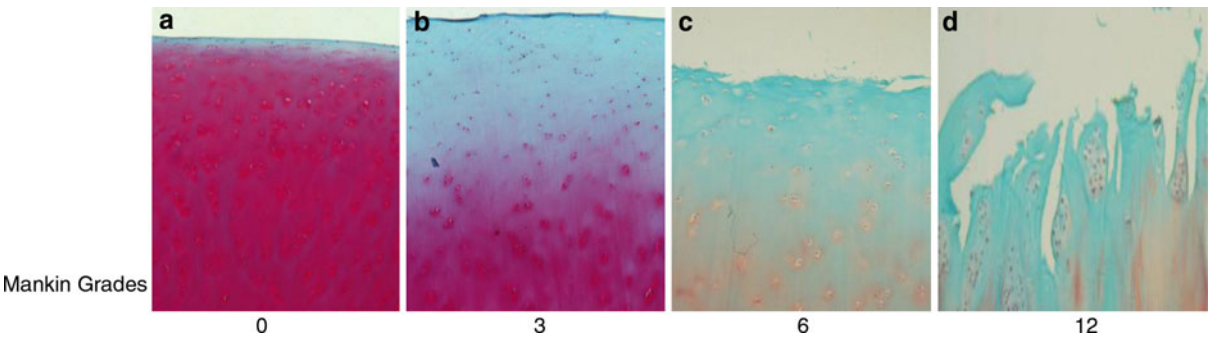
function and properties are inferior to those in native cartilage and it has little morphological or structural identity to hyaline cartilage. Though initially, repaired and host tissue may appear to be integrated, inherent biomechanical weakness of the fibro-cartilage makes it more susceptible to wear or re-injury in both the junction area or within newly produced cartilage. This eventually evolves into full-thickness fissures, leading to the loss of the repaired tissue. Further, it sets the stage for progressive cartilage degeneration and predisposes the joint to the development of OA.

Osteoarthritis as Not a Simple Wear of Articular Joint

OA is a degenerative disease that affects the entire joint. It is characterized by an imbalance between catabolic and anabolic activities of chondrocytes, suggesting that alterations in cellular metabolism contribute to the onset and progression of the disease. Chondrocytes regulate their metabolic activity in response to mechanical, chemical, or electrical stimuli by altering cartilage structure and composition to accommodate the physical demands of the body. The solid matrix of articular cartilage has intrinsic mechanical properties that enable it to undergo rapid reversible deformation during loading. In combination with mechanical factors, local cellular responses, including specific phenotypic changes, contribute significantly to loss of cartilage matrix structure and function in OA, concurrent with changes in other joint tissues. Once the OA process has been initiated, the forces borne by the joints during motion and support (both aberrant loads and physiological loads) may accelerate the progress of structural degeneration. Ultimately, this process proceeds through degeneration of the articular hyaline and fibrocartilage and substantial periarticular bony changes.

Because cartilage is a soft-tissue spacer in the joint, loss of cartilage can cause joint laxity, even in the absence of other changes. Cartilage loss typically proceeds gradually (Fig. 2) and is accompanied by extended periods where the loss is stabilized or partially reversed by a fibrocartilaginous repair process. The typical pattern of joint degeneration in OA involves (1) the progressive loss of the articular cartilage surface, with sporadic attempts at repair; (2) remodeling of the subchondral bone with sclerosis and, possibly, microfractures; (3) synovial and capsular thickening, often with local inflammation, which is common and has increasingly been recognized as one of the drivers of joint destruction; (4) the growth of osteophytes at the joint margins; and (5), especially in end-stage disease, sclerotic bone cysts (Abramson and Attur 2009; Guilak et al. 2000).

Though in OA the biomechanical properties are compromised, it is still not understood whether this precedes the onset of the disease or is a result of it (Loeser 2009). It has been widely documented that the prevalence of idiopathic OA rises directly with age and it is the most common cause of chronic disability in older adults. However, it is important to note that OA is not an inevitable consequence of aging; it is not a simple "wearing out" of the joints; and aging-related changes in the joint can be distinguished from those due to disease. Not all older adults develop OA and not all joints are equally affected. It is more apparent that aging changes in the musculoskeletal system contribute to the development of OA by working in conjunction with other factors such as obesity, joint injury, genetics, and of course biomechanics. It is also



Wear of Natural Joints (Osteoarthritis), Fig. 2 Stages in human adult articular cartilage degeneration illustrated by Safranin O staining and assessed by Mankin score. (**a**) Grade 0, normal appearance of Safranin O staining; (**b**) grade 3, indicates the loss of staining primarily in the upper part of cartilage; (**c**) grade 6, indicates the loss of superficial layer, surface fissures, and substantial loss of staining throughout cartilage depth; (**d**) grade 12, indicates fissures into the deep layer, formation of large clones of chondrocytes, loss of staining, and matrix disorganization

clear that OA-like changes in the joint can develop without contribution of aging, which is especially true for post-traumatic OA. Thus, aging and OA are inter-related but not inter-dependent (Loeser 2009).

The Role of Biomechanics in the Development and Progression of OA

Normal joint function depends on the low joint friction and absorption and transmission of loads provided by healthy cartilage. Even relatively small changes in the integrity, composition, or organization of the cartilage matrix will alter its mechanical properties and compromise its functionality. Since adult articular cartilage has a limited capacity for natural regeneration, even small injuries or small changes in mechanical properties can predispose the onset of OA. The factors initiating such a process are manifold and include major trauma, repetitive overuse, aging, and genetics. The progression of the already established disease is largely biomechanically mediated. The disease progresses more rapidly with increased load (Andriacchi et al. 2004).

Biochemical factors in cartilage degradation are well known (Cawston and Wilson 2006). OA involves the proteolytic cleavage of key structural matrix components and aggrecan loss from the cartilage matrix is an early and sustained feature of OA. For example, aggrecanases (a Disintegrin And Metalloproteinase with Thrombospondin Motifs [ADAMTS]-4 and -5) play a big role in the cleavage of aggrecan, while collagenases, and especially collagenase 13, initiate the cleavage of collagens. The fragments of matrix constituents elute into the synovial fluid and the matrix loses its function of water attraction. The cartilage loses its mechanical damping characteristics and becomes "softer" over time and more prone to overload and tear. Less is known about the mechanical factors of tissue wear. In pure mechanical terms, wear is defined as the removal of material from the contact surfaces due to mechanical action between them. At this point there is no proof whether adhesive, abrasive, or fatigue wear of the collagen matrix can occur without chemical cleavage.

Beyond the joint, clinical OA of the lower extremities proceeds in a manner that is heavily influenced by biomechanics. Evidence of this comes from the previous studies of OA progression. If OA were primarily mechanically mediated, then it should progress in a non-random, mechanically predictable manner. Shakoor et al. (2002) evaluated the outcome in the other joints of the lower extremities in a group of patients who had end-stage OA in a single lower extremity joint. A comparable group undergoing single-joint arthroplasty for rheumatoid arthritis (RA), a systemic inflammatory disease that is not primarily mechanically mediated, was studied as a control. Using a database of ~6,000 patients undergoing hip or knee arthroplasty, it was found that patients undergoing primary hip replacement for OA who then had evidence of OA progression in the knees, progressed primarily in the contralateral knee rather than in the ipsilateral knee. This was in stark contrast to RA patients, for whom progression in the knees was random. Confirmation of the mechanical contribution to lower extremity OA progression was provided by an analysis of gait in a large cohort of subjects undergoing hip replacement. Thus, in contrast to the systemic arthritis RA, which progresses randomly in the lower extremities, OA progresses in a non-random manner that is directly related to asymmetric dynamic loading of the involved joints.

Cross-References

- ▶ Articular Cartilage as Bearing Materials-An Engineering Prospective
- ▶ Artificial Synovial Fluid

References

S.B. Abramson, M. Attur, Developments in the scientific understanding of osteoarthritis. Arthritis Res. Ther. **11**(3), 227 (2009)

T.P. Andriacchi, A. Mündermann, R.L. Smith, E.J. Alexander, C.O. Dyrby, S. Koo, A framework for the in vivo pathomechanics of osteoarthritis at the knee. Ann. Biomed. Eng. **32**(3), 447–457 (2004)

P.G. Bullough, The role of joint architecture in the etiology of arthritis. Osteoarthritis Cartilage **12**(Suppl A), S2–S9 (2004)

T.E. Cawston, A.J. Wilson, Understanding the role of tissue degrading enzymes and their inhibitors in development and disease. Best Pract. Res. Clin. Rheumatol. **20**, 983–1002 (2006)

S. Chubinskaya, M. Hurtig, D.C. Rueger, OP-1/BMP-7 in cartilage repair. Int. Orthop. **31**(6), 773–781 (2007)

W.H. Day, S.A. Swanson, M.A. Freeman, Contact pressures in the loaded human cadaver hip. J. Bone Joint Surg. Br. **57**(3), 302–313 (1975)

M.L. Gray, A.M. Pizzanelli, A.J. Grodzinsky, R.C. Lee, Mechanical and physiochemical determinants of the chondrocyte biosynthetic response. J. Orthop. Res. **6**, 777–792 (1988)

F. Guilak, L.A. Setton, V.B. Kraus, Structure and function of articular cartilage, in *Principles and Practice of Orthopaedic Sports Medicine*, ed. by W.E.J. Garrett, K.P. Speer, D.T. Kirkendall (Lippincott, Philadelphia, 2000), pp. 53–73

I.M. Khan et al., Cartilage integration: evaluation of the reasons for failure of integration during cartilage repair. A review. Eur. Cell Mater. **16**, 26–39 (2008)

R.F. Loeser, Aging and osteoarthritis: the role of chondrocyte senescence and aging changes in the cartilage matrix. Osteoarthritis Cartilage **17**(8), 971–979 (2009)

C.A. Poole, Articular cartilage chondrons: form, function and failure. J. Anat. **191**(Pt 1), 1–13 (1997)

N.W. Rydell, Forces acting on the femoral head-prosthesis. A study on strain gauge supplied prostheses in living persons. Acta Orthop. Scand. **37**(1) (1966)

B.L. Schumacher et al., A novel proteoglycan synthesized and secreted by chondrocytes of the superficial zone of articular cartilage. Arch. Biochem. Biophys. **311**(1), 144–152 (1994)

N. Shakoor, J.A. Block, S. Shott, J.P. Case, Nonrandom evolution of end-stage osteoarthritis of the lower limbs. Arthritis Rheum. **46**, 3185–3189 (2002)

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)

Wear of Tires

David Manas
Department of Production Engineering, Tomas Bata University, Zlin, Czech Republic

Definition

Wear is a phenomenon that can never be completely eliminated. It is present everywhere, in all aspects of life. While the problem of wear in metallic materials has been described in detail, the problem of wear in rubber products requires further investigation. It is common for rubber products (tires, conveyor belts, etc.) to show signs of wear. Tire production is the largest sector among products made of rubber. Their wear is an important factor both for manufacturers and buyers. Manufacturers aim to produce tires with better resistance to wear than those of their competitors, while buyers look for a tire with the best resistance to wear. This chapter will introduce you to the wear of standard tires used on roads and highways as well as off-road tires used in harsh terrain conditions such as quarries, construction sites, or forests.

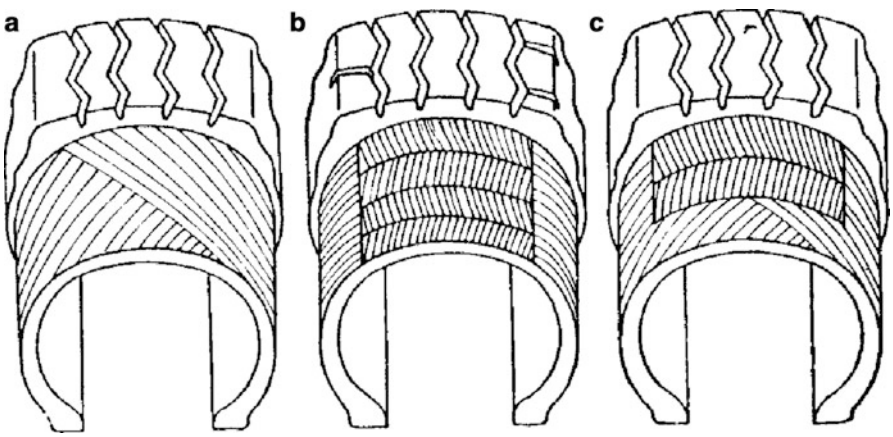
Wear is usually considered in terms of abrasion, which is defined as the loss of material resulting from mechanical action on a rubber surface. Abrasion resistance is a complicated phenomenon and is dependent on many things, such as residence, stiffness, thermal stability, and resistance to cutting and tearing, and different applications require a wide range of these properties.

The definition of cutting and chipping, for the purpose of this essay, is as follows: Cutting takes place when the tire strikes a sharp object with enough force that the surface is penetrated or cut. Chipping can follow cutting by the effect of reactive, braking, or other forces on rough or sharp surfaces, causing tearing of the rubber compound, usually at 90° to the direction of the cut. Chipping can also occur in an early stage if the relative motion between the tread and the cutting object can gouge out a piece of rubber.

Scientific Fundamentals

A typical tire consists of several parts: tire carcass, tread, sidewall, bead, breaker, inner rubber, and other parts. A tire provides an immediate contact between the vehicle and the road. It must transmit the load of the vehicle, help transmit the driving torque, and ensure satisfactory driving conditions. The tire design is divided into radial or cross-ply (Fig. 1).

In the case of radial tires, the cords of the carcass run across the direction of travel from bead to bead. This arrangement is called "radial." Cords do not ensure sufficient stability of the casing. This is ensured by a belt located around the circumference under the tread, which consists of several plies of rubber-coated steel cords.

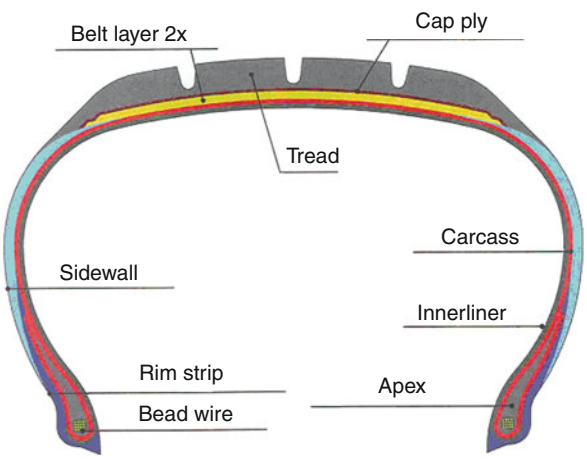


Wear of Tires, Fig. 1 Basic design types of tires (**a**) cross-ply tire, (**b**) radial tire, (**c**) cross-ply tire with a belt

The individual belt plies are crossed over and positioned at an acute angle to the direction of travel. In the case of modern tires, both of these reinforcing elements are made of steel cords. These tires are referred to as "ALLSTEEL." Initially, these tires were manufactured for special uses on a 5° tapered divided rim (including tube and flap). The latest radial ALLSTEEL truck tires are designed for 15° tapered bead seat rims. These tires are always tubeless. The rim is one-piece (drop-center rim).

Unlike the radial design, the carcass of cross-ply tires consists of several rubber-coated textile cord plies. Cord plies are positioned at approximately 40° to the direction of travel (crossing from the ply). This arrangement is called "cross-ply" in Europe and "bias-ply" in the US. A breaker can be inserted between the tread and the carcass to reduce the risk of damage to the carcass (www.conti-online.com).

In the rubber industry, the wear of rubber parts is a frequent problem. Some kinds of wear, in particular wear of tire treads or conveyor belts, are very similar to machining. The tire tread is the part of the tire that provides the contact between the vehicle and the road and is directly involved in the transmission of the driving force (Fig. 2). Friction between the tire tread and the road surface has a direct influence on the ability to transfer the driving forces but also in the ability to brake the car on a required track. It depends on the properties of the rubber compound the tire tread is made of, as well as on the properties of the road on which it is moving. In the case of the tire carcass of passenger cars and trucks using standard roads, the wear of the tread is characterized by its abradability. The tread of a tire is exposed to the abrasive effect of the road on which the car is driving.



Wear of Tires, Fig. 2 Cross-section of a passenger car tire

Resistance to wear is a very complicated property. In laboratory testing it is subject to a protocol testing method, although in practice it is influenced by many factors and is significantly affected by the way in which the rubber product is to be used. In some cases, it is necessary to use rubber with great elasticity (e.g., for tire treads, soles, and heels), which provides resistance to the separation of particles during deformation, i.e., when in contact with sharp edges of stones and rough terrain. In other cases, the best performance can be attained by rubber with little resistance to deformation at high deformation speeds or rubber that can flex to avoid sharp edges of stones and terrain irregularities. This phenomenon can be seen to a certain extent in the case of tire treads – harder rubber is more suitable for highways with a smooth surface and softer rubber is better for stony surfaces. Good resistance to wear depends on both the composition of the tread compound and the uniform dispersion of the filler used. Seeking good resistance to wear of rubber parts subjected to high stress usually involves a combination of resistance to flex, cracking, and dynamic stress (Schätz et al. 1979).

It is important to consider the influence of elasticity and hardness in the wear-resistant materials. Wear resistance is significantly influenced by the type of elastomer and filler used. To a lesser extent, it is also influenced by the vulcanizing agents and softeners used.

Wear Tests

Wear tests are carried out in order to determine the resistance of different materials to wear and consequently the changes in the material surfaces and the loss of material under the test conditions. These tests evaluate the wear resistance (to abrasion, rubbing, friction) of rubber products, such as tire treads, shoemaking material, or conveyor belts, which is important in determining the service life of a product.

Wear tests are of great importance, taking into consideration that field tests of products such as tires, usually carried out on roads or on a test circuit, are time consuming and very expensive. Tire manufacturers usually check the tread compound of tires during production to guarantee minimal resistance to wear (Grosh and Schallamach 1961).

Selecting the right testing method for a given application is very important, as different types of tires are subjected to different stresses. In most cases (tires for passenger cars and trucks driving on standard roads), friction of a tested material for rough to abrasive surfaces is important, and the following values are used – friction force, normal pressure force, and coefficient of shear friction.

$$F = \mu \times N[N] \quad (1)$$

where F is the friction force, N is the normal pressure force, and μ is the coefficient of shear friction.

According to the speed of the wear process, the coefficient of shear friction is divided into static (μ_{stat}), in which case the friction force necessary to set the system in motion is inserted in the formula, and dynamic (μ_{dyn}), in which case the formula includes values necessary to maintain the system at constant speed. In some cases, the friction force is reduced to the effective friction force. However, as the effective friction area is difficult to establish, this value is rarely used. When evaluating tread compounds, depending on experimental arrangement and the condition of the surface, it is possible to separate the process of friction into deformation and adhesive friction, and hence divide the friction force into deformation and adhesive force (Grosh and Schallamach 1961).

$$F = Fd + Fa[N] \quad (2)$$

where F is friction force, Fd is deformation force, and Fa is adhesive force.

Deformation friction relates to the transformation of surface unevenness of a softer element by a harder friction element. This process should be considered as any friction process from the point of energy. The supplied external energy is used for shape deformation of surface projections to the direction of motion while at the same time elastic energy is released together with deformation. The ratio of supplied and released energy should be proportional to the hysteresis of losses related to viscoelastic properties of the tested tread compounds and it should dissipate as heat energy (heating). That means that the center of energy losses will be under the surface layer, where the maximum shear forces are present.

Adhesive friction is smaller than deformation friction and is present in every surface contact (even smooth) of two bodies. The origin of these forces is in low molecular forces (electrostatic forces, van der Waals forces, hydrogen bridges). The degree of approximation is determined by surface irregularities (effective distance is 5–10 μm), where adhesive friction is shown by a continuous sequence of creation and dissipation of adhesive "link."

To sum up, smooth surfaces and loads support adhesive forces, while great surface irregularities or rough friction elements and high normal pressure forces cause deformation friction. The mechanism of tread that compounds abrasion (tire wear) is as follows: First, smooth surfaces are rubbed by abrasive edges, which cause the formation of parallel cuts perpendicular to the direction of abrasion. These cuts are asymmetrical in profile. A set of these cuts, a sort of abrasive marks, are also visible on tires in real life and are very similar to those made by laboratory abrasive equipment (sandpaper or wheel). After the abrasion, projected bits that flex in the direction of motion occur. One of its sides is protected while the other side is exposed to more rubbing. The edges of these projections get thinner and thinner until they are ripped off and the surface becomes smoother. In the meantime, new edges from the bottom layer are formed, new abrasive marks are formed, and the whole process repeats itself.

There are a number of laboratory devices for testing tire tread wear. Although each one works on its own principle, their task is to model practical behavior of tested tread compounds (tires). This article mentions only a few of them.

Wear tests are performed both in the laboratory and in the field. Laboratory tests focus almost exclusively on one phenomenon and a special device and methods are proposed for each factor. Field tests not only facilitate establishing wear but also a set of properties important for the operation of the given vehicle (Schätz et al. 1979).

Laboratory Tests for Wear

Laboratory Tests for Wear of Passenger Car Tires

Resistance to wear cannot be given by a positive value but by a negative one, i.e., loss during frictional stress to which rubber is subjected. The value of wear is not a given property of rubber because it depends on how it occurs. Wear is a very complicated phenomenon subject due to many factors. There have been many studies and papers designed to establish the different contributions to wear, but so far with little success. It is known that hardness, strength, tear strength, and resilience have some effect on wear, but this relationship is indirect and complicated. Part of the problem is that wear caused by frictional stress occurs at high temperatures, while the static tests are performed at normal temperatures.

In a laboratory test, wear is very often determined as follows: a sample of a surface is rubbed under a certain pressure against moving standard abrasive material, which can be sandpaper or a carborundum disc. A variety of scenarios are given by different devices, differences in the shape of the sample, the way of pressing, moving direction, the character and shape of the abrasive material, given pressure, and speed of movement.

Results are usually given as the mass loss in a unit of time, after a given number of revolutions, or after a given distance. To eliminate some experimental mistakes caused by factors such as unevenness of the abrasive material,

the loss is reduced to the loss for standard rubber with known values tested under the same conditions. The standard loss equals 100; the loss of the test is expressed by a dimensionless number, which shows the ratio of the standard loss and the tested loss. A number higher than 100, e.g., 125, indicates rubber of a poorer quality than standard rubber, a number lower than 100, e.g., 85, indicates rubber of a higher than standard quality.

Given the great variety of arrangements of these devices and different test conditions, it is not possible to compare the results from the different devices directly. Despite these drawbacks, laboratory tests represent a good tool for checking and for development of compounds.

Resistance to wear is an important property for tire treads, heels, soles, and conveyor belts. There is plenty of literature that gives many examples of testing treads.

Due to the many factors influencing wear, and also to the nature of this property itself, laboratory results do not fully correspond with the field test results, and, moreover, the results are often contradictory.

Wear is affected by the composition of the compound – more precisely, by the type and amount of carbon black used. In recent years, carbon black has been substituted with silica. Based on the present knowledge, increased resistance to wear, for elastomers, can be ranked as follows: styrene butadiene rubber, natural rubber, and others (Schätz et al. 1979).

A number of methods and testing devices have been developed for tire wear tests. The devices can be divided into two groups. The first group contains testing devices that measure wear of standard tires used on standard roads and highways. Wear is characterized by abradability. In the other group, the devices can test off-road tires (rubber products subject to high stress). National and international standards for abrasion resistance are shown in Table 1.

Wear of Tires, Table. 1 National and international standards for abrasion resistance

ISO 4649 (ISO (DIN) abrader)
ISO 5470 (Taber)
ASTM D394 (Du Pont)
ASTM D1630 (NBS)
ASTM D2228 (Pico)
ASTM 3389 (Taber)
BS 903: Part A9
DIN 53516

The following examples relate to some of the devices and methods used:

The *Du Pont methodology using constant friction* is suitable for evaluation of tread compounds and, to a certain extent, correlates with the road tests. The mass loss of the samples is established by weight. The samples have square sections and are fitted with fixtures. They are attached to the lever of the device and during the test they are forced against a circular ring of abrasive material with a constant pressure.

The *Bussen–Schlobach methodology* measures wear on rubber with a variable size of friction.

The *method of Akron–Croydon* uses a small disc sample that rotates at a defined speed and at constant pressure when pressed against a carbon corundum disc. The resistance to wear is given as a volume loss after a defined number of revolutions.

With the *Dunlop–Lambourn methodology*, the test sample rotates and is forced against a carbocorundum wheel with a separate drive. Resistance to wear is established by mass loss per 1 km of track.

The *dry sand–rubber wheel test* (*DSRW*) uses a device designed for laboratory test of metal wear, which is adjusted to measure elastomer samples. Sand acting as an abrasive is spread between the test sample and the wheel.

The *National Bureau of Standards* (*NBS*) abrader is used for testing soft vulcanized rubber compounds, such as those used for shoe soles.

The *Pico abrader* uses a pair of blunt tungsten carbide knives that rub the test piece while it rotates on a turntable. The force on the test piece and the speed of rotation is reserved at intervals throughout the test and a dusting powder is fed to the test piece surface (Smith 1993).

Effective Wear of Tires on Standard Roads and Highways

When a vehicle is moving, the tread of the tire is in contact with the road surface. The tread is the part of the tire that ensures contact of the vehicle with the road and helps to transmit the driving force. The tire rubs against the road surface. This is evident from the tires' diminishing tread depth. Wear is caused mainly by friction, which enables the transmission of driving forces on the road surface, as in the case of acceleration or braking or cornering in the case of lateral forces. Slip is a relative movement between the road and the tire that occurs when force is transmitted. Slip means that the vehicle speed is greater or smaller than the wheel's circumferential speed. In other words, the distance the vehicle covers is longer or shorter than the rolling circumference of the tire. Wear, either slow or quick, is unavoidable.

The degree of wear and thus mileage performance of a passenger car tire depends, among other things, on the degree of slip. While slip, as such, is a necessary condition of driving, the degree of slip is greatly dependent on the motorist's driving style. The degree of wear is directly linked to the degree of slip. Depending on the driving style – whether it is economical or high performance – comparable tires can attain mileage performance of anywhere between 5,000 and 40,000 km. Every type of excessive wear shortens the potential mileage performance of a tire. The causes of premature wear are evident on the basis of characteristics of wear pattern displayed by the worn tread. Besides design, material, and driving style, wheel alignment and maintenance affect the wear of tires. This effect implicates, for example, side wear, wear in the tread center, diagonal spot wear, heel and toe wear, flat spot wear, excessive one-side wear, and many other problems (www.conti-online.com).

Laboratory Test of Off-Road Tire Wear

For tires that are used in harsh terrain conditions (Fig. 3), the mechanism of wear is totally different. Sharp edges of stones and terrain irregularities cause cuts in the tread, which can be compared to the mechanism of machining. There is a certain similarity, e.g., with milling, although very specific conditions are necessary. The mechanism of wear of tire treads operating in harsh terrain conditions is referred to as the chip-chunk effect and it can be compared to "machining" of a rubber surface.

Thirteen kinds of tread compounds were selected for a wear test. These tread compounds are designed for the production of off-road tires and are industrially produced and processed. Treads were produced on the basis of NR and BR and carbon black was used as filler. The selected tread compounds are used for the production of agricultural, multifunctional, high performance, and motocross tires.

The tests for tire wear are both time consuming and expensive. They are carried out using tires in testing rooms or in the terrain during driving tests. This is one of the reasons why a new wear test method is needed, one that would allow comparing different kinds of tread compounds using small sample sizes in a very short time.

Based on these requirements a device was designed. Its principle is illustrated in Fig. 4. A device for testing chip–chunk wear (Batty 1989) was selected for basal measurements. A new device enabling changing the test parameters, providing true simulation of the process conditions, was designed (Manas 2005; Manas et al. 2008a).

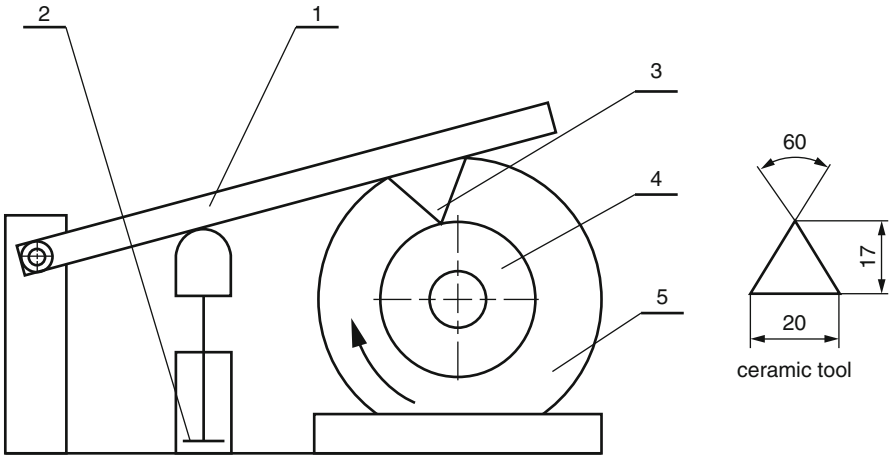
Arm *1*, revolving around a pivot, is lifted by a lifting device (piston of the pneumatic roller) *2*. Once lifted, the arm, to the end of which is attached a special ceramic tool *3*, is dropped on the rotating wheel *4* (test sample) powered by an electric motor *5*. The tool, when it drops on the rotating wheel, gradually chips the material and creates a groove on the wheel. The size of the groove created by the tool in a given time is the scale of wear.

Dimensions of the Test Sample

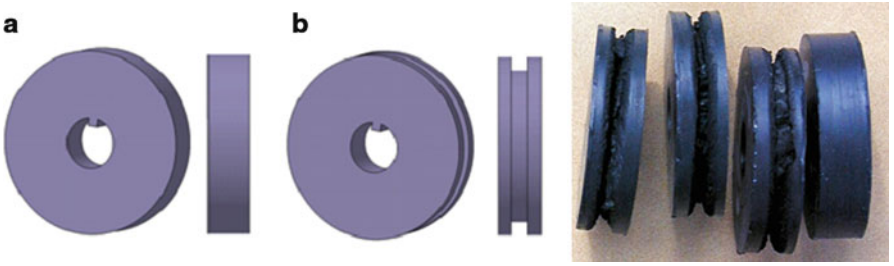
Test samples were prepared according to Fig. 5. The diameter of the test sample was 55 mm and the width 13 mm, the same as the test sample for establishing Luepke rebound resilience (Fig. 5).

In the course of the test, a groove was made (cut) in the test sample using a ceramic tool. As experienced with

Wear of Tires, Fig. 3 Off-road tire wear



Wear of Tires, Fig. 4 Machine for testing chipping and chunking effect *1* – arm, *2* – pneumatic roller, *3* – ceramic tool, *4* – test sample, *5* – electric motor



Wear of Tires, Fig. 5 Sample for the test of quick wear (**a**) before testing (**b**) after testing

tooling other materials such as metals, wood, plastics, and others, the shape of the groove was expected to be regular. However, due to the properties of the machined rubber that showed its elasticity, the material of the rotating wheel was torn the moment the ceramic tool dropped on the rotating wheel. Therefore, the original intention of evaluation of wear by measuring the diameter of the groove was dismissed and a gravimetric evaluation was performed (Manas 2005; Manas et al. 2008a).

Analysis of Wear

The influence of the energy of dropping the ceramic tool on the surface of the test sample is crucial for wear. In the case of a rigid sample, it would be easy to calculate the effect of the dropping force. However, for elastic test samples, after the first drop of the ceramic tool on the surface of a test sample, a series of other effects with less intensity, like "jumping" on the surface, took place. The main effects of the ceramic tool have only partial influence on the total wear. It has been found that to establish the total work necessary for wear (i.e., creation of a groove in a test sample) considering only the energy of the main drop would be biased. For the first testing of the experimental equipment, it was considered that the results of the given series of measurements would be comparable provided that the experiment was done under the same conditions. The design of the test sample with a key fitting into a groove on the shaft and clamps with teeth will prevent slipping of the test sample (Manas 2005; Manas et al. 2008a).

Conditions of Experiment

The tests of quick wear were performed on the experimental equipment under the following conditions:

Revolutions of the test sample	500, 750, 910 rpm
Frequency of dropping of a ceramic tool	1 Hz
Stroke of the ceramic tool	60 mm
Ambient temperature	21 °C
Time of the experiment	270 s

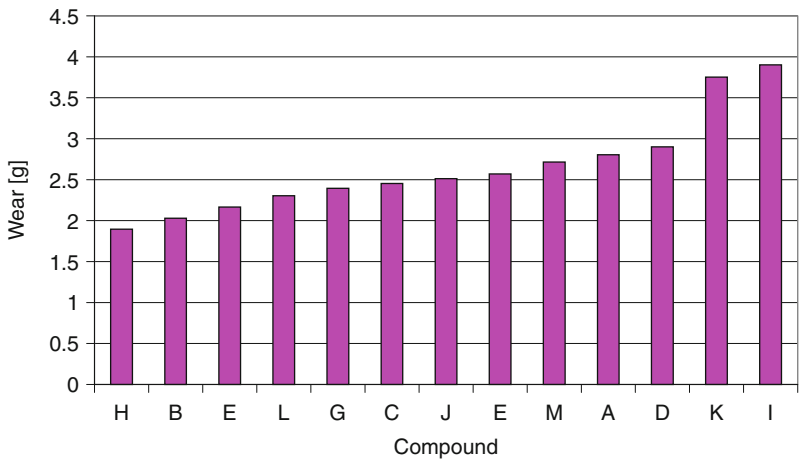
The test sample was clamped between the wrists of the machine to prevent slips and was set into rotation. At the

same time, a lifting device used for lifting the arm with the ceramic tool was started. The time was measured from the first contact of the tool with the test sample. During the experiment, ten test samples from each compound were used. Upon completion of the experiment, the mass loss was established by weighing using analytic balances. The values measured were graphically processed and evaluated (Fig. 6) (Manas 2005; Manas et al. 2008a).

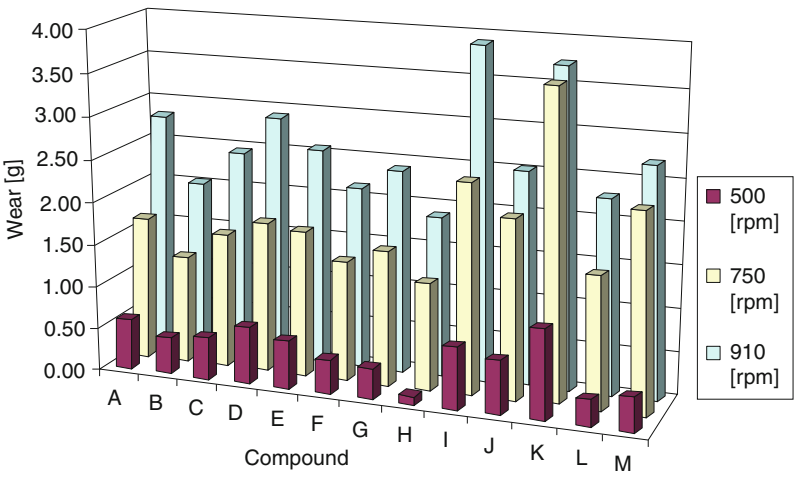
The greatest wear was observed with compounds I, K, and D. The best properties were reported with test samples prepared from H, B, and E compounds (Fig. 6).

Dependence on the Running Conditions

Vehicles move at different speeds in the terrain, in running conditions that can be characterized by the circumferential speed of the tread. Therefore, more experiments were carried out to establish the degree of wear in different conditions. The wear test was run at the following test revolutions: n_1 = 910 rpm, n_2 = 750 rpm, n_3 = 500 rpm. The other conditions of the experiment remained the same. The results show a tendency of increased wear at higher revolutions (Fig. 7) (Manas 2005; Manas et al. 2008b).



Wear of Tires, Fig. 6 Comparing wear of the compounds



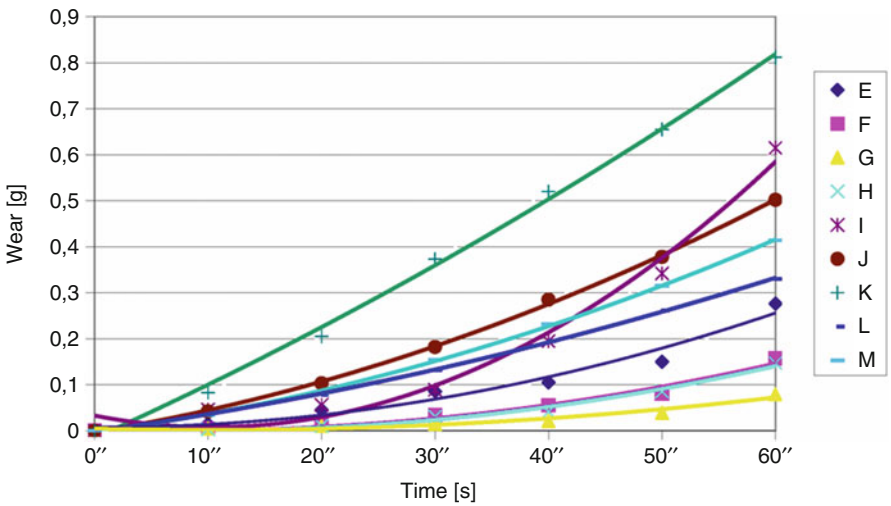
Wear of Tires, Fig. 7 Comparing the mass loss (wear) at different running conditions

The Course of Wear

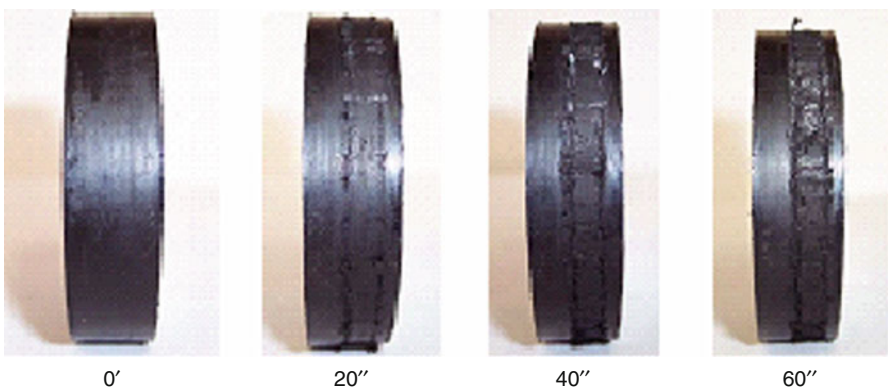
Figure 6 shows that, from the point of view of wear (cut, chipping and chunking resistance), it is possible to compare various compounds with each other.

The mass of samples was measured every 30 s from the beginning to the end of the experiment (270 s). Due to the different behaviors of tested compounds at the initial phase of the test, special attention was paid to the phase of experiment in the interval from 0 to 60 s, in which the weight of the test sample was measured every 10 s (Figs. 8, 9).

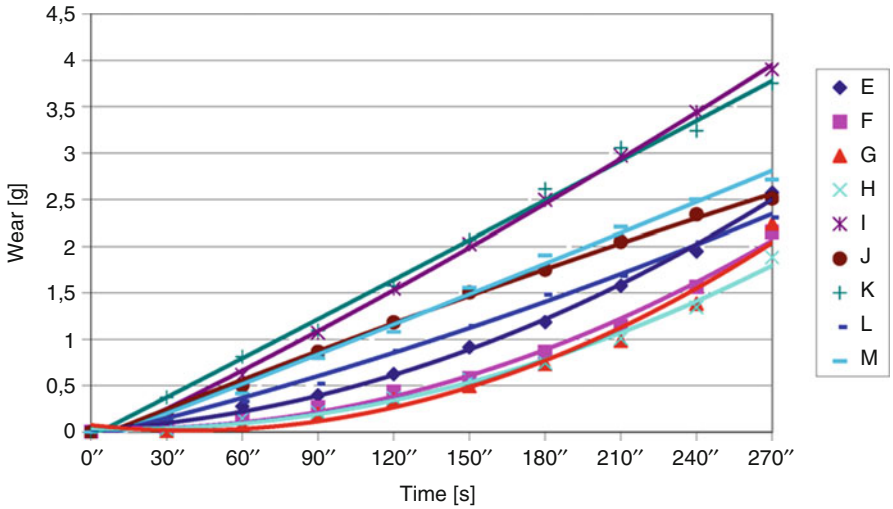
Most of the test samples showed a gradual increase of wear in the initial phase of the experiment (Figs. 8, 10). A more pronounced wear occurred only after the formation of cracks. This means that wear is minimal until the first manifestations of surface damage. It is caused by the elastic nature of the rubber testing wheel. The tool has a kinetic energy at the drop of the ceramic tool on the rotating test sample that can cause deformation (Figs. 12, 13). Deformation work consists of the elastic component of the deformation work and work that is reduced by friction during molecule regrouping. The rebound of a ceramic tool is formed by an elastic element of the deformation work, which is diminished by the hysteresis of the loss. This process goes on until the moment of creation of first cracks in the test sample. Increased wear from the beginning of the test can be seen in compounds with low resistance to wear (Fig. 10) (Manas 2005; Manas et al. 2008).



Wear of Tires, Fig. 8 Gradual mass loss (wear) of all compounds in the time period (0–60 s)



Wear of Tires, Fig. 9 Gradual mass loss (wear) in the test samples in the time period (0–60 s)

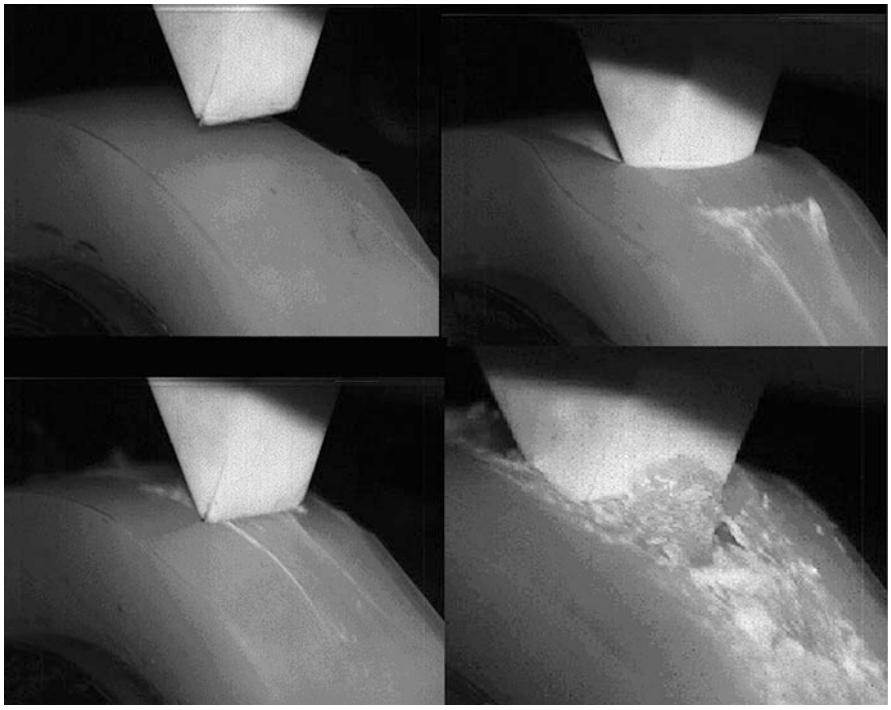


Wear of Tires, Fig. 10 Gradual mass loss (wear) in all compounds in the time period (0–270 s)

Wear of Tires, Fig. 11 Gradual mass loss (wear) in the test samples in the time period (0–270 s)

There is a gradual tendency for wear to accelerate with proceeding time in most cases. This means that, before the creation of the first rips on the surface of the tire tread while driving on harsh terrain conditions (sharp stone edges, etc.), the wear is quite small. The first damage to the tire tread starts the "avalanche effect" of other damage and the wear increases faster (Figs. 10, 11).

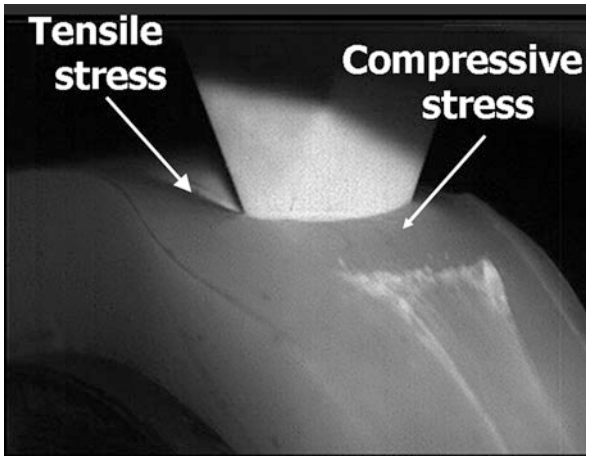
The actual contact surface between the surface of the tested rubber sample and sharp edges of stones and terrain irregularities is very small during the process of wear. Stress develops in this spot during rotation (rotary movement) of the test sample. When the ceramic tool is dropped on the circumference of the test sample, the tool is forced against the surface layer of the rubber, which causes tensile stress behind the head of the deformation on the sides of the groove (Figs. 12, 13). If the tensile stress exceeds the mechanical strength of the rubber material, a part of the rubber is ripped off, either completely or partly. Stress and deformation in the area around the ceramic tool are reduced and the process can repeat a bit further away. A ceramic tool machines the surface of the rubber test sample (by turning or milling). The way in which individual parts are ripped off (Fig. 12) is similar to the way of ripping during a test of tear strength. Thus, it can be expected that the factors influencing the process of wear are deformation energy needed for ripping and the size of a ceramic tool (in real conditions they are sharp edges of stones and terrain irregularities). This is determinant for the size of deformed micro–macro areas and ripped particles. The process during which a given rubber particle is ripped off can be seen as an elementary tensile test, where the speed of ripping is much bigger than the speed of the movement of the test sample. Resistance to wear will be closely linked to the deformation energy needed for ripping the given rubber particle.



Wear of Tires, Fig. 12 Damage of the surface of the test sample

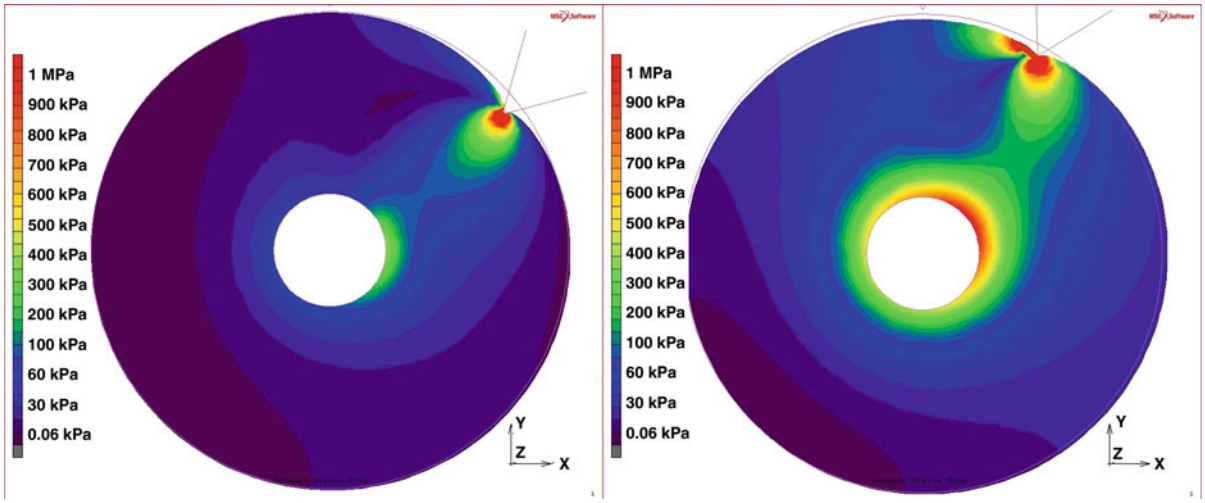
Falling of the cutting edge onto the rotating specimen was simulated. At first, the sliding of the cutting edge along the specimen surface without material destruction was analyzed. Von Mises stress in the specimen during counter-clockwise rotation is shown in Fig. 14. Then, a hack of cutting edge into the specimen during rotation was simulated and is shown in the Fig. 14 (Manas 2005; Manas et al. 2007a, b, 2008b).

A high-speed camera system Olympus i-SPEED 2" was used to examine properties of the test samples. The camera system was used to visualize the behavior of the test sample at the drop of the ceramic tool (Figs. 12, 13). Deformation of the test sample was observed at the time of drop of the ceramic tool. At the same time, the creation of first rips at the drop of the ceramic tool on an undamaged surface of the test sample was observed, which starts the avalanche effect of wear and hence destruction of the test sample. The high-speed camera enables a better understanding of the process of tire wear. Next to the observation of deformation of the test samples, it is possible to determine the degree of pressure on the sample at the drop of the ceramic tool, the size of the ripped particles, and their speed of movement. It is also possible to observe the character of damage and the creation of rips on the test samples (Manas 2005; Manas et al. 2008).

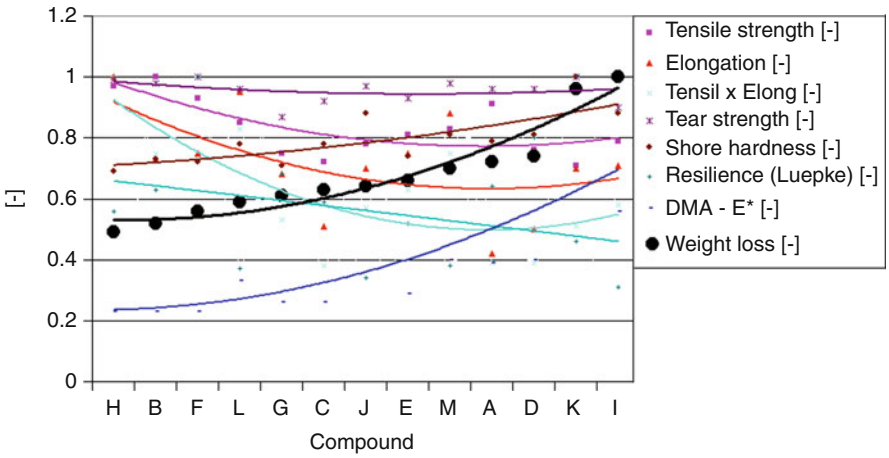


Wear of Tires, Fig. 13 Development of tension in the place of the drop of the ceramic tool

In the rubber industry, mechanical properties such as tensile strength, elongation, tear strength, and rebound resilience are usually tested. Also, a complex dynamic model is usually used. The producers seek to find correlation between these mechanical properties and wear. These mechanical properties play an important role in tire wear.



Wear of Tires, Fig. 14 Stain in the rotating sample by sliding and hack of the cutting tool



Wear of Tires, Fig. 15 Comparison of all the observed properties

The sets of measured values were processed and the results are illustrated in graphs (Fig. 15). For easier and faster comparison of the measured values, dimensionless values expressed as the ratio of separate measurements to the maximum value reached during the given measurements were used. The influence of mechanical properties on wear of rubber parts (tire tread compounds) subjected to high stress was investigated. This influence was statistically processed using the correlation coefficient (r) and the p value (the extent of relevance).

The experimental results of the experiment show that the resistance of heavily stressed rubber parts to wear depends on mechanical properties of rubber as well as on a given deformation energy. Substantial influence is recorded by defects (and especially their size) that were present in the rubber testing sample from the beginning of the test, or were created during the deformation. Significant correlation between individual mechanical characteristics should also be noticed.

Apart from mechanical characteristics, the impact of chemical activity of ozone and oxygen present in air has considerable effect.

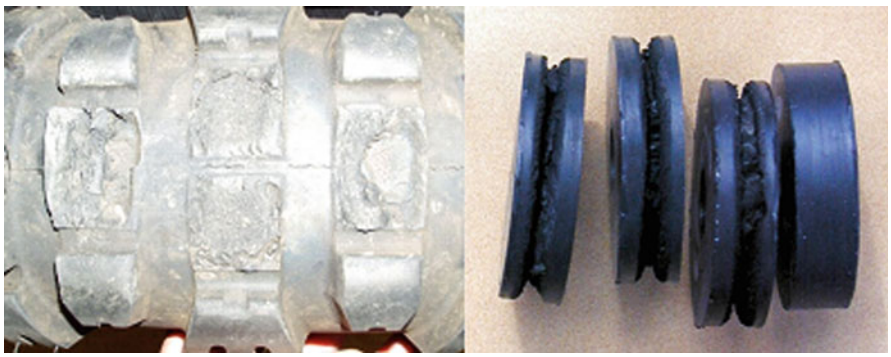
Tread compounds of off-road tires are subjected to harsh terrain conditions (construction sites, quarries) being stressed by sharp edges and terrain irregularities. Their high load contributes greatly to the deformation

when the tires move on a stony surface. If a high resistance to wear is required, the tires must resist cutting tools such as sharp edges of stones and terrain irregularities as tools for the surface of the tread compound. The compounds with high value of hardness, resilience, and dynamic complex module will easily manage to roll around the sharp bits on the surface. Their greater values of elongation strength will ensure better toughness and thus a high resistance against damage of the tread compound surface what would cause micro and macro cracks, leading sources for avalanche effect of wear. Hard compounds with smaller elongation strength and high values of resilience and dynamic complex module E^* will be more susceptible to crumbling when in contact with irregular surfaces. This leads to damage on the surface of the compound, which creates ideal conditions for more cracks, resulting in the avalanche effect of wear.

The character of wear shows that laboratory conditions for testing tread compounds used for the production of off-road tires are very close to real conditions. Figure 16 shows the comparison of a test sample with an actual tire. The character of wear is very similar (Manas 2005; Manas et al. 2007a, b, 2008b, 2009a).

Correlation Matrix

Correlation matrix (Table 2) shows the linear dependence of mechanic characteristics on work of heavily stressed rubber parts. Strong dependence of wear on hardness



Wear of Tires, Fig. 16 Comparison of actual wear with a test sample

Wear of Tires, Table 2 Correlation matrix

Correlation matrix	X1	X2	X1xX2	X3	X4	X5	X6	Y
X1	1.00	0.35	0.64	0.48	−0.57	0.35	−0.52	−0.59
X2	p = .23	1.00	0.93	0.27	−0.22	−0.27	−0.13	−0.32
X1xX2	p = .01	p = .00	1.00	0.40	−0.41	−0.07	−0.31	−0.49
X3	p = .09	p = .36	p = .17	1.00	0.13	0.12	0.20	−0.20
X4	p = .04	p = .46	p = .16	p = .65	1.00	−0.62	0.96	0.83
X5	p = .23	p = .36	p = .81	p = .69	p = .02	1.00	−0.60	−0.48
X6	p = .06	p = .66	p = .29	p = .50	p = .00	p = .03	1.00	0.78
Y	p = .03	p = .28	p = .08	p = .49	p = .00	p = .09	p = .00	1.00

where Y = Wear [g]
Xi = Measured value of properties of rubber compounds, i.e.,
X1 = Tensile strength [MPa]
X2 = Elongation [%]
X3 = Tear strength [N/mm]
X4 = Shore hardness [ShA]
X5 = Resilience luepke [%]
X6 = DMA (complex tensile modulus E^*) [MPa]

and complex module E^* as well as mutual correlation of mechanical characteristics are recorded. Correlation coefficient values are stated in the matrix above the diagonal, the importance of (p-factor) is stated below the diagonal (Manas 2005; Manas et al. 2007c, 2008b).

Actual Wear of Off-Road Tires

Circumferential Scoring (Cuts)

Cuts (Fig. 17) or a similar wear pattern along the circumference of the tire are caused by projecting vehicle parts or foreign objects trapped in the tire when the vehicle is in motion (www.conti-online.com).

Mechanical Damage of the Tread

Mechanical damage of the tire tread is caused by stony roads or terrain irregularities. This wear pattern is also closely connected to the damage caused by sharp objects such as glass or metal. The tire damage is accentuated by moisture and overinflation (Manas 2005; Manas et al. 2008, 2009) (Fig. 18).

Key Applications

Tire wear is a rather complicated matter. It is not entirely clear which tire tread compound is the most effective and which will show poorer qualities with respect to wear. The character of wear should always be taken into account and the right properties of a given compound should be selected accordingly. For movements on standard roads and highways where the tread is exposed to abrasion, it is more suitable to select a compound that is harder and more resistant to the abrasive effect of the road. On the other hand, a softer compound is a better choice for tires used in harsh terrain conditions.

It is also necessary to pay attention to the correct inflation pressure of the tires. A properly inflated tire can play an important role in the final wear of the tires used. Finally, it should not be forgotten that the proper alignment of a vehicle is crucial for the reduction of tire wear; misalignment can cause excessive wear. Tire wear can also be affected by subjective factors, such as driving style.

Excessive wear is a serious safety and economic problem. Suitable laboratory methods for testing tire tread wear and finding correlations between routinely measured rubber compounds can play a significant role in the prediction of tire tread behavior, because testing tires in the running conditions is both time consuming and expensive (Manas et al. 2009a, b; Stanek 2010).

Wear of tires used on standard roads is characterized by their abradability. As the running conditions are varied, laboratory tests can only give approximate results and they also depend on the method used for measuring abradability and its evaluation.

Laboratory testing methods for tire treads used in harsh terrain conditions are well worked out.

Wear of Tires, Fig. 17 Cuts along the circumference of the tire

Wear of Tires, Fig. 18 Chipping and chunking, tire damage

Special equipment known as a chip-chunk tester is able to compare wear (by chip-chunk resistance) of different compounds or to make a comparison with etalon. The test is not time consuming and uses minimal material. The correlation found between wear and routinely measured properties makes it possible to predict behavior of the tested product during its operation.

When observing the process of wear, other unconventional methods prove very useful, such as observing the process using a high-speed camera. This leads to new, useful findings and the problem of wear will continue to be examined with the objective of finding new methods, describing the process further, and helping to design tires with better resistance to a given wear pattern.

Cross References

► Friction of Polymers
► Polymeric Elastomers: Material Aspects of Tribology
► Polymeric Elastomers: Tribological Behavior and Engineering Components

Acknowledgment

This article is financially supported by the Ministry of Education, Youth and Sports of the Czech Republic under the Research Plan No. MSM 7088352102 and by the European Regional Development Fund under the project CEBIA-Tech No. CZ.1.05/2.1.00/03.0089.

References

J.R. Batty, B.J. Miksh, A laboratory cutting and chipping tester for evaluating off-the-road and heavy-duty tire treads. Rubber Chem. Technol. **55**, 1531 (1989)

M. Schätz, P. Vondráček, Polymer testing, textbook VŠCHT Prague (1979)

K.A. Grosh, A. Schallamach, Wear **4**, 356 (1961)

J. Javořík, Z. Dvořák, Equibiaxial test of elastomers. Kautsch. Gummi Kunstst. **60**(11), 608–610 (2007), ISSN 0948–3276

D. Manas, *Rubber Workability and Wear of Rubber Parts* (VUT FSI, Brno, 2005). ISBN 80-214-3026-5

D. Manas, M. Manas, M. Stanek, T. Drga, Wear of rubber components. Chemické listy, Ročník 101, 36–37, Vydavatel: Česká společnost chemická, Novotného lávka 5, 116 68 Praha 1 (2007a), ISSN 0009–2770

D. Manas, M. Manas, M. Stanek, V. Pata, Wear of tire treads. Arch. Mater. Sci. **28**(1–4), 81–89 (2007b), Published quarterly by the Committee of Materials Science of the Polish Academy of Sciences, ul. Śniadeckich 8, 00–656 Warszawa, Poland, ISSN 1734-9885

D. Manas, M. Stanek, M. Manas, Workability and wear of rubber parts, in *DAAAM International Scientific Book* (DAAAM International, Vienna, 2007c), pp. 611–626, ISSN 1726–9687, doi:10.2507/daaam.scibook.2007.54. ISBN 3-901509-60-7

D. Manas, M. Stanek, M. Manas, T. Drga, Tire wear measurements, in *ICPM 2007, International Congress on Precision Machining*, 25–28 Sept 2007, Sandomierz–Kielce, vol. II, pp. 153–156. ISBN 978-83-88906-91-6

D. Manas, M. Manas, M. Stanek, V. Pata, New investigation in chip-chunk resistance of tire treads, in *International Rubber Conference – IRC 2008*, 20–23 Oct 2008a, Kuala Lumpur, pp. 37–48, CD

D. Manas, V. Pata, M. Manas, M. Stanek, New investigation in wear of rubber components, in *Eighth Fall Rubber Colloquium 2008*, 26–28 Nov 2008b, Hannover, pp. 93–94 [+CD Proceedings]

D. Manas, V. Pata, M. Manas, M. Stanek, A. Sivtseva, New investigation in wear of rubber components. Vestnik iževskogo gosudarstvennogo techničeskogo universiteta. Rusko **2**(42), 39–40 (2009a), ISSN 1813-7903

D. Manas, M. Stanek, M. Manas, V. Pata, J. Javorik, Influence of mechanical properties on wear of heavily stressed rubber parts. KGK Kautsch. Gummi Kunstst. **62**, 240–245 (2009b), ISSN 0948-3276

L.P. Smith, *The language of Rubber* (Butterworth-Heinemann Ltd, Oxford, 1993)

M. Stanek, Influence of surface quality of injection molds on fluidity of polymers. Plast. Rubber **XLVII**(5–6), 142–144 (2010), ISSN 0322-7340, FT UTB in Zlin

www.conti-online.com

Wear Prevention and Fuel Economy

► Fuel Economy: Lubricant Factors

Wear Quantification by Comparison of Surface Topography Data

G. Vorlaufer, S. Ilincic, F. Franek, A. Pauschitz
Austrian Center of Competence for Tribology AC^{2}T research GmbH, Wiener Neustadt, Austria

Synonyms

Point-wise measurement of surface topography changes

Definition

The surface topography of solids is changed due to removal or at least displacement of material from tribologically interacting surfaces. The amount of wear may be quantified by measuring the local change of topography of the same region of the worn and unworn surfaces.

Scientific Fundamentals

Tribological interactions between a solid surface and another solid, liquid, or gaseous media normally causes surface wear. The processes leading to wear are adhesion, abrasion, erosion, fretting, and others that are described elsewhere (Kato 2005). In general, the result of a wearing process is the removal of material from the tribologically stressed surface and subsurface layers. The exception is the case of material transfer to the surface from a counterpart (Czichos 1978; Bhushan 1999).

Usually the amount of wear is given in terms of mass or volume loss ("wear volume") of the wearing body or as the change of some characteristic linear dimension ("wear height," "wear depth," or diameter). However, the amount of worn material is usually very small compared to the total amount of material and the dimensional changes of wearing bodies of technical interest are typically in the range of micrometers and below. Thus, quantification of wear is often difficult in practical applications.

In terms of surface metrology, surface topography is often assessed in terms of form, waviness, and roughness of a surface (Whitehouse 2002). However, in the context of this article, the term "surface topography" simply denotes the set of points defining the geometric locus of the surface. A topography data set is a digital representation, i.e., a discretely sampled subset of the surface topography, obtained by a suitable measuring device. These devices include among others the atomic force microscope, confocal white-light microscope, or laser interferometer. Hence, a surface topography data set (S) is given as the elevation (z_{ij}) of the surface above a regular plane grid (x_j,y_i), see Fig. 1, i.e.,

$$S = \left\{ \vec{r}_{ij} = \left(x_j, y_i, z_{ij}\right)^T = \left(x_0 + j\Delta x, y_0 + i\Delta y, z\left(x_j, y_i\right)\right)^T \right\} \tag{1}$$

with the integer index set

$$I_S = \{(i,j) | (1 \le i \le I) \wedge (1 \le j \le J)\} \tag{2}$$

A representation of the surface topography in the form $z = z(x, y)$ is also known as the relief of the surface and its discrete form (1) as the height map of the surface. An illustration of these definitions is given in Fig. 1.

As stated above, wear manifests itself as a change of surface topography. Therefore, one possibility to quantify wear is the accurate detection of those changes, i.e., the point-wise comparison of topography data sets of corresponding regions of worn and unworn surfaces. Slightly different implementations of this idea may be found in (Corniani et al. 2009; Monetti et al. 2009; Heintze et al. 2006; Harris and Krauss 2001; Borodich et al. 2003; Schöfer and Santner 1998; Gåhlin and Jacobson 1998).

A surface topography data set of a worn surface is suitable for wear quantification only if this data set also contains unworn regions of the surface that is used as reference and either the nominal shape of the unworn surface may be described by an analytical function (e.g., plane, sphere, or cylinder) or a surface topography data set of the unworn surface is available.

Method 1: Quantification of Wear in Case of a Known Nominal Shape

The nominal shape of the unworn surface topography is given as

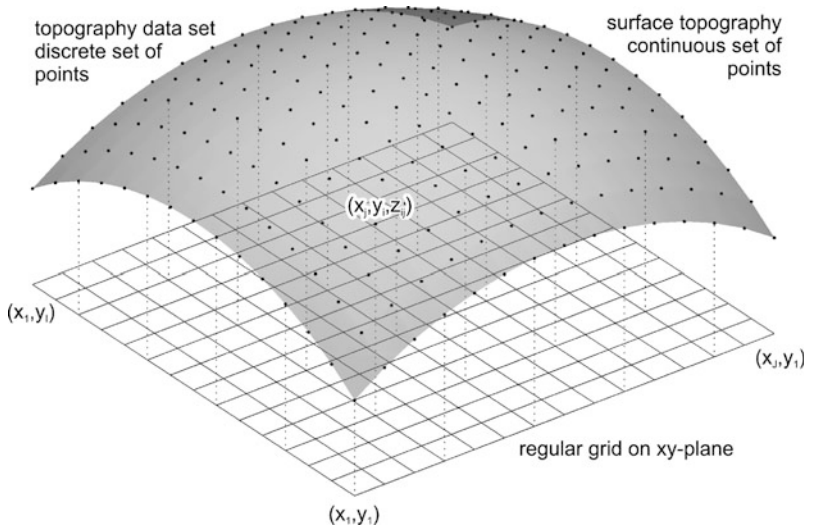
$$z = z^*(x, y; \vec{a}) \tag{3}$$

where $\vec{a} = (a_1, \ldots, a_K)$ is a vector of shape parameters. For example, a plane may be specified analytically with $\vec{a} = (a, b, c)$ as

$$z^*(x, y; \vec{a}) = a\,x + b\,y + c \tag{4}$$

or a sphere with $\vec{a} = (x_C, y_C, z_C, R)$ as

$$z^*(x, y; \vec{a}) = z_C + \sqrt{R^2 - (x - x_C)^2 - (y - y_C)^2} \tag{5}$$



Wear Quantification by Comparison of Surface Topography Data, Fig. 1 Scheme illustrating the definitions of "surface topography" and "topography data set"

If I_R is the set of indices of points lying on the reference regions, i.e., the unworn part of the surface, the nominal shape of the unworn surface can be recovered by chi-square minimization, i.e., by minimizing

$$\chi^2 = \sum_{(i,j)\in I_R} \left(\frac{z_{ij} - z^*\left(x_j, y_i; \vec{a}\right)}{\sigma_{ij}} \right)^2 \qquad (6)$$

with respect to the shape parameters, i.e., by solving a set of K (possibly non-linear) equations

$$\frac{\partial \chi^2}{\partial a_k} = -2 \sum_{(i,j)\in I_R} \frac{\left(z_{ij} - z^*\left(x_j, y_i; \vec{a}\right)\right) \frac{\partial z^*}{\partial a_k}}{{\sigma_{ij}}^2} = 0 \qquad (7)$$

with $k = 1, \ldots, K$. In (6) and (7), σ_{ij} are the standard deviations of the elevation values z_{ij}. Hence, the values σ_{ij} characterize the measurement uncertainties. The solution of (7) is a vector of shape parameters $\vec{a} = \vec{a}_{opt}$ leading to a nominal surface shape that best fits the topography data set. The deviation of the data set from the nominal surface can be calculated on a point-by-point basis as

$$h_{ij} = z_{ij} - z^*\left(x_j, y_i; \vec{a}_{opt}\right) \qquad (8)$$

Negative values of h indicate material loss, whereas positive values indicate material build-up. Thus, the wear height can be effectively quantified on a point-by-point basis, i.e., in a spatially resolved way, by mapping h_{ij} onto the grid (x_j,y_i). Actually, the wear height is usually defined as being positive where material is removed, i.e.,

$$w_{ij} = -h_{ij} \qquad (9)$$

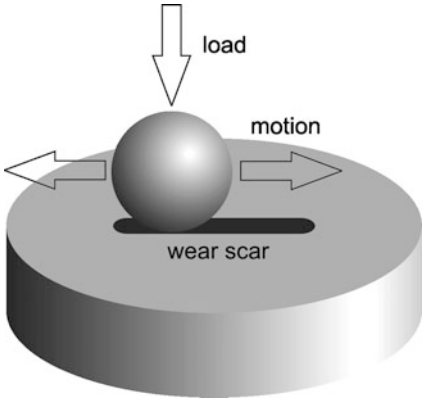
One factor limiting the applicability of the above-described method is the fact that no real surface is following a mathematically described nominal shape. Following standard surface metrology terms, the topography of a real surface will have deviations in form, waviness, and roughness. Therefore, the uncertainty in the calculation of the wear height is given by the average height of these deviations, given, for example, by the R_a or R_z value.

Method 2: Quantification of Wear in Case a Topography Data Set of the Unworn Surface is Available

If the nominal shape of the original surface is not known and/or the wear height is in the same order of magnitude as the typical height of the roughness asperities (see concluding remarks of the above section), wear can still be quantified if there is available a topography data set of the same region of the unworn surface, i.e., one has a reference data set. Denote the reference data set by

$$S_0 = \left\{ \vec{r}^{\,0}{}_{ij} = \left({x^0}_j, {y^0}_i, {z^0}_{ij}\right)^T \right\} \qquad (10)$$

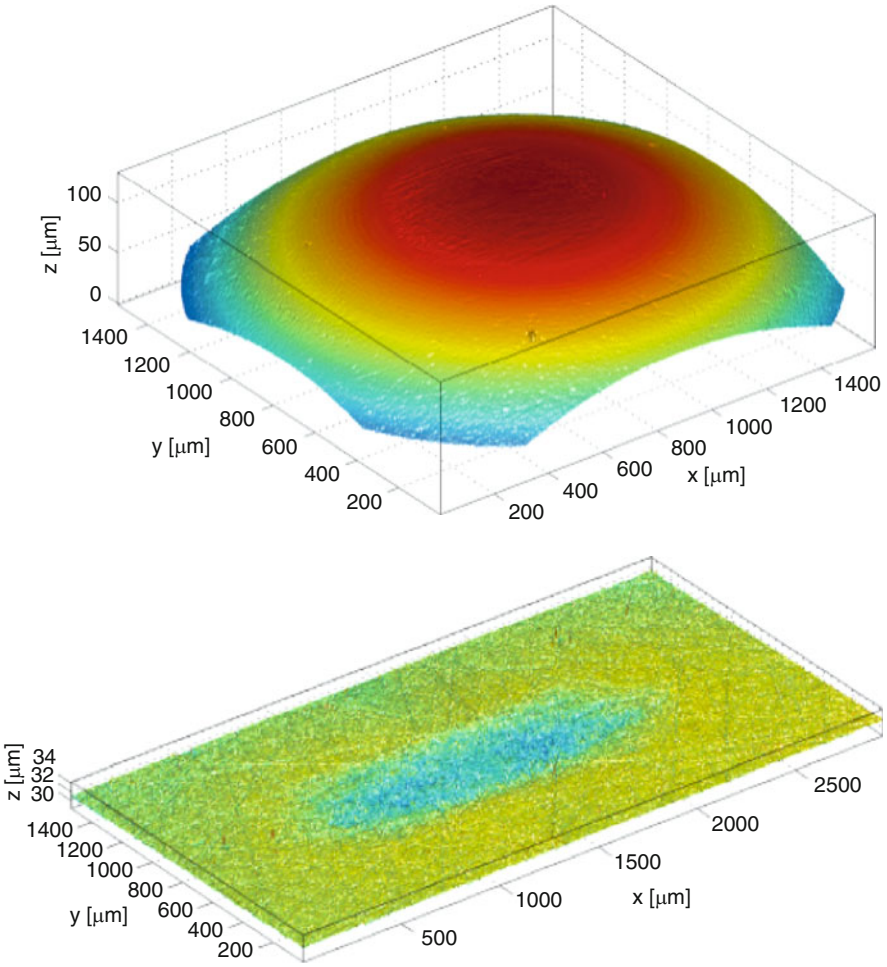
At first sight one could assume that the grid arrangement $({x^0}_j, {y^0}_i)$ is identical to (x_j, y_i) in order to calculate the point-by-point difference as $z_{ij} - {z^0}_{ij}$. However, this naïve approach is not correct since consecutive measurements of the same surface region will lead to a certain amount of misalignment, especially if the sample has been manipulated between those measurements.



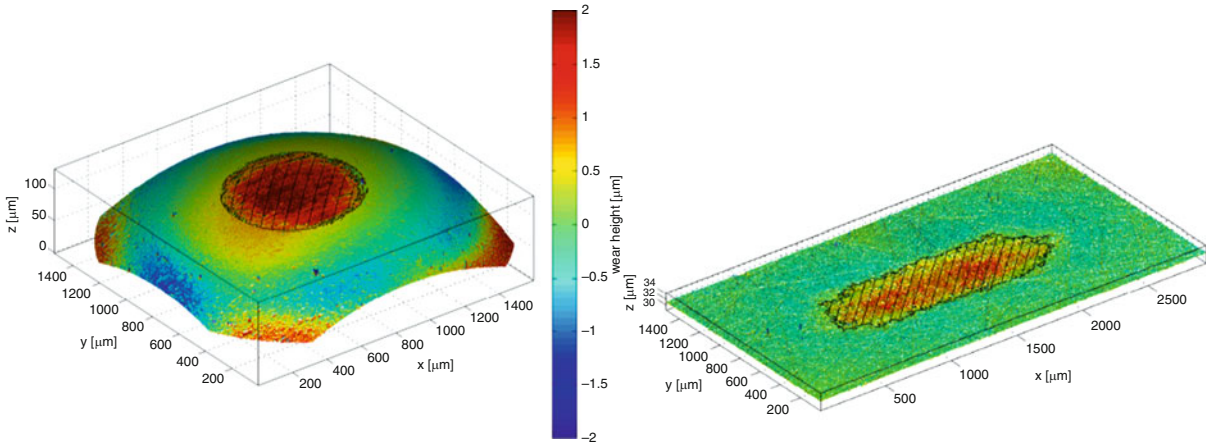
Wear Quantification by Comparison of Surface Topography Data, Fig. 2 Contact situation during a ball-on-plate linear reciprocating friction and wear experiment

Such misalignments are usually the case when performing some wearing experiment. As a consequence, since the underlying co-ordinate system is affixed to the measuring device and not to the surface, even identical values of (x^0_j, y^0_i) and (x_j, y_i) will in general not refer to the same points on the surface. Therefore, an alignment procedure has to be employed.

The goal of the alignment procedure is to match reference regions of the original topography data set to those of the worn one. The degrees of freedom for the alignment procedure are given by three angles of rigid body rotation $(\alpha_x, \alpha_y, \alpha_z)$ and three components of the rigid body translation vector (t_x, t_y, t_z). These components may be given in many different ways. In particular, taking the rotation matrix



Wear Quantification by Comparison of Surface Topography Data, Fig. 3 Topography data of wear scars on ball (*top*) and plate (*bottom*)



Wear Quantification by Comparison of Surface Topography Data, Fig. 4 Wear height (*false color*) and wear scar (*hatched region*) calculated from topography data shown in Fig. 3

$$R = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha_x & -\sin\alpha_x \\ 0 & \sin\alpha_x & \cos\alpha_x \end{pmatrix} \cdot \begin{pmatrix} \cos\alpha_y & 0 & \sin\alpha_y \\ 0 & 1 & 0 \\ -\sin\alpha_y & 0 & \cos\alpha_x \end{pmatrix} \cdot \begin{pmatrix} \cos\alpha_z & -\sin\alpha_z & 0 \\ \sin\alpha_z & \cos\alpha_z & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{11}$$

and the translation vector

$$\vec{t} = \left(t_x, t_y, t_z\right)^T \tag{12}$$

the transformation of a point $\vec{r} \in S$ is given by

$$\vec{r}'_{ij} = R \cdot \vec{r}_{ij} + \vec{t} = \vec{r}'_{ij}\left(\vec{r}_{ij}; \alpha_x, \alpha_y, \alpha_z, t_x, t_y, t_z\right) \tag{13}$$

The match is achieved by χ^2-minimization (6) with respect to the parameters $\vec{a} = \left(\alpha_x, \alpha_y, \alpha_z, t_x, t_y, t_z\right)$, which will be written as

$$\chi^2 = \sum_{(i,j)\in I_R} \left(\frac{z'_{ij} - z^0\left(x'_{ij}, y'_{ij}\right)}{\sigma_{ij}} \right)^2 \tag{14}$$

$z^0(x, y)$ must be a continuously differentiable function interpolating the reference data set S_0, i.e., ${z^0}_{ij} = z^0\left({x^0}_j, {y^0}_i\right)$. Any suitable interpolation scheme may be chosen, e.g., using cubic b-splines. The result of the minimization procedure is a set of optimal fitting parameters, leading to

$$\vec{r}'_{ij,opt} = R\left(\alpha_{x,opt}, \alpha_{y,opt}, \alpha_{z,opt}\right) \cdot \vec{r}_{ij} + \vec{t}_{opt} \tag{15}$$

and the wear height can subsequently be calculated as a point-by-point deviation of topography data as

Wear Quantification by Comparison of Surface Topography Data, Table 1 Results of wear quantification procedure applied to ball and plate

	Mean wear height	Worn area	Wear volume
Ball	1.4 µm	0.32 mm^2	4.5×10^{-4} mm^3
Plate	0.8 µm	0.71 mm^2	5.8×10^{-4} mm^3

$$h_{ij} = z'_{ij,opt} - z^0\left(x'_{ij,opt}, y'_{ij,opt}\right) \tag{16}$$

The advantage of this method, as compared to the one described previously, is given by the fact that the shape of the unworn surface is derived from measured data and does not rely on assumptions about a nominal shape. Therefore, form, waviness, and roughness are accurately captured down to the lateral resolution of the measuring device. Thus, the uncertainty of wear height values is determined by the point-wise uncertainties, i.e., the vertical resolution of the measuring device itself.

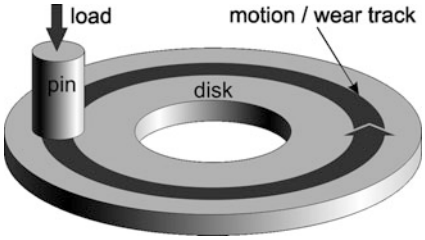
Key Applications

Ball-On-Plate Wear Volume Determination

One of the most common contact principles used for model friction and wear experiments is a ball sliding against a plate in a linear reciprocating motion with lateral oscillation amplitudes in the range of few millimeters (see Fig. 2).

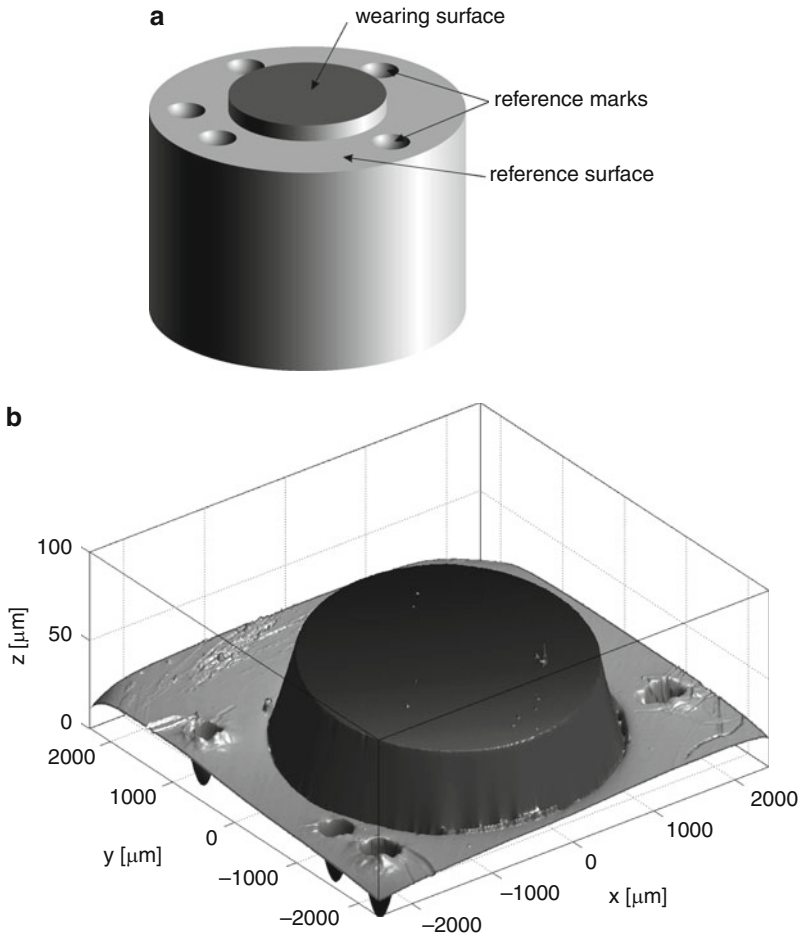
One advantage of this experimental setup is that the regions of the counteracting bodies affected by wear,

i.e., the wear scars, are very well localized and small enough to be entirely captured by a surface topography measuring device. Furthermore, the nominal shape of the test specimen is well known and the amount of wear is typically much larger than the average roughness of the test specimen, such that method 1 is suitable for wear volume calculation. Typical data sets of worn ball and plate surface topography are shown in Fig. 3 and will be used to illustrate the calculation procedure.



Wear Quantification by Comparison of Surface Topography Data, Fig. 5 Contact situation of a pin-on-disk tribometer

At first, the optimal nominal geometries, i.e., a sphere (5) and a plane (4), are determined solving (7). Subsequently the wear height is calculated on a point-by-point basis using (8) and (9). In Fig. 4, wear height is shown as false color image mapped onto the surface topography data. Finally, the wear volume is obtained by integrating the wear height over the wear scar region (depicted as black hatch in Fig. 4). The results of wear quantification of by comparison of surface topography data are summarized in Table 1.

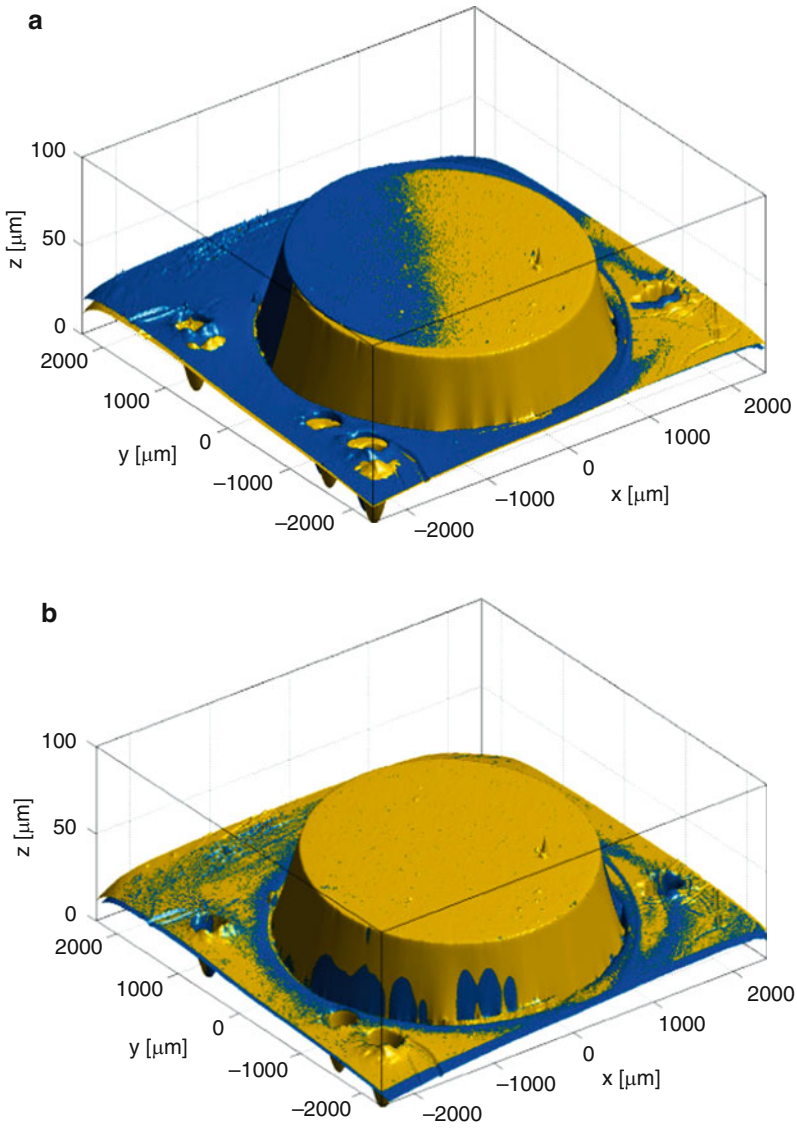


Wear Quantification by Comparison of Surface Topography Data, Fig. 6 Schematic view (**a**) and topography (**b**) of cylindrical pin test specimen

Sub-micrometer Wear Height Determination on a Cylindrical Pin

A pin-on-disk rig is one of the most widely used tribometers for model friction and wear tests. A cylindrical pin is pressed against a rotating disk, as shown schematically in Fig. 5. The disk is wearing on a circular track according to its motion, whereas the pin (see Fig. 6) is wearing on its complete contacting surface.

For accurate determination of the wear height of the pin, method 2 described above is applied. For this purpose a slightly recessed ring is machined into the contacting surface (see Fig. 6a), which does not wear during the test and can therefore be used as reference surface. Additionally, reference marks are created on the surface using a spherical indenter. These reference marks allow for better initial positioning and faster convergence of the



Wear Quantification by Comparison of Surface Topography Data, Fig. 7 Superposition of the worn (*blue*) and unworn (*yellow*) surface topography data sets (**a**) before alignment and (**b**) after alignment

alignment procedure. Figure 6b shows the topography data set of the unworn surface.

By superimposing the topography data sets of the worn and unworn surfaces in the same diagram, as shown in Fig. 7, the effect of the alignment procedure (11, 12, 13, 14, 15) becomes clearly visible. Figure 7a shows the two datasets using an initial guess for the fit parameters α_x, α_y, α_z, and $\vec{t}$ whereas for Fig. 7b the best fit parameters were used. The alignment procedure involves solving an unconstrained, non-linear minimization problem (14), which must be done numerically. For this particular application, the BFGS algorithm (Press et al. 2007) has been used. Following (16), the point-wise height difference can be calculated using the optimally aligned topography data sets. With the example data shown in Fig. 7, the average wear height on the top pin surface (wearing surface) amounts to 0.37 ± 0.08 µm.

Cross-References

- ► Asperities
- ► Confocal Microscopy
- ► Optical Interferometry
- ► Surface Roughness
- ► Surface Variation in Tribological Processes

References

B. Bhushan, *Principles and Applications of Tribology* (Wiley, New York, 1999)

F.M. Borodich, L.M. Keer, S.J. Harris, Self-similarity in abrasiveness of hard carbon-containing coatings. J. Tribol. **125**(1), 1–7 (2003)

E. Corniani, M. Jech, F. Ditroi, T. Wopelka, F. Franek, TLA and wear quantification of an aluminium-silicon-copper alloy for the car industry. Wear **267**(5–8), 828–832 (2009)

H. Czichos, *Tribology: A Systems Approach to the Science and Technology of Friction, Lubrication and Wear* (Elsevier, Amsterdam, 1978)

R. Gåhlin, S. Jacobson, A novel method to map and quantify wear on a micro-scale. Wear **222**(1998), 93–102 (1998)

S.J. Harris, G.G. Krauss, Improved technique for measuring the ball volume removed in a ball-on-disk test. Tribol. Lett. **10**(3), 187–188 (2001)

S.D. Heintze, A. Cavalleri, M. Forjanic, G. Zellweger, V. Rousson, A comparison of three different methods for the quantification of the in vitro wear of dental materials. Dent. Mater. **22**(2006), 1051–1062 (2006)

K. Kato, Classification of wear mechanisms/models, in *Wear – Materials, Mechanisms and Practice*, ed. by G.W. Stachowiak (Wiley, Chichester, 2005), p. 9. Chap. 2

C. Monetti, S. Ilo, T. Lebersorger, Characterization of lubricants for metal forming by means of an oscillating tribo-test-rig. Proc. IMechE Pt. J: J. Eng. Tribol. **223**(5), 817–826 (2009)

W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes*, 3rd edn. (Cambridge University Press, Cambridge, 2007), p. 521 ff

J. Schöfer, E. Santner, Quantitative wear analysis using atomic force microscopy. Wear **222**(1998), 74–83 (1998)

D. Whitehouse, *Surfaces and Their Measurement* (Hermes Penton, London, 2002)

Wear Reduction via Micro Vibration

► Contact with Micro Vibrations for Friction Control and Wear Reduction

Wear Resistance of Materials After Accelerated Quenching

► Tribological Properties of Intensively Quenched Materials

Wear-In

► Running-in

Wear-Mechanism Maps

► Wear Maps

Wear-Mode Maps

► Wear Maps

Wear-Resistant Materials for Gas Bearings

► Gas Bearing Materials

Wear-Transition Maps

► Wear Maps

Weld Cladding

► Weld Overlay

Weld Overlay

Cleiton C. Silva
Department of Materials and Metallurgical Engineering, Welding Engineering Labor, Federal University of Ceará, Fortaleza, Ceará, Brazil

Synonyms

Buildup; Buttering; Cladding; Hardfacing and weld overlay; Weld cladding

Definition

Traditionally, the term *weld overlay* is used to define applications of welding processes to deposit one or more layers of metal with specific characteristics on a base metal to improve desirable properties that are not inherent to the base metal or to restore the original dimension of the component.

The procedure consists of the deposition of several weld beads arranged side by side, leading to the formation of a continuous surface layer.

The weld overlay technique can be classified according to the specific objective of the coating. Weld cladding corresponds to deposition of a corrosion-resistant material, in general, measuring at least 3 mm (1/8 in.).

Hardfacing is a form of surfacing used when the weld deposit is done using a hard and wear-resistant material to reduce the loss of material by abrasion, erosion, galling, cavitation, and other forms of wearing mechanisms.

Buildup is a specific term to denote the deposition of weld metal on a base metal surface to restore dimensional requirements of a component. The term *buttering* refers to the deposition of one or more layer of weld metal on the face of the joint with the purpose of assuring some metallurgical aspects before weld filling.

Scientific Fundamentals

The weld overlay procedure consists of the deposition of one or more layers of filler metal holding specific characteristics on a base metal to improve desirable properties not inherent to the base metal. The post-weld desirable properties are related to the application of the equipment, tool, or machining parts and also depend on the characteristics of the material to be deposited.

When the layer is deposited with corrosion-resistant materials to protect against corrosive environments, the term used is *weld cladding*. When the weld deposit is made using hard and wear-resistant materials, with the objective of reducing the loss of material by abrasion, erosion, galling, cavitation, and other forms of wearing, the specific term is *hardfacing*. *Buildup* is a specific term to denote the deposition of weld metal on a base metal surface to restore dimensional requirements of component. The term *buttering* refers to the deposition of one or more layers of filler metal on the face of the joint with purpose of assuring some metallurgical aspects before weld filling (Davis 1993).

This procedure is a common practice in the manufacturing or maintenance of equipment and components for a wide range of industries, including agriculture, mining, steel, paper, food processing, chemical, petrochemical, petroleum and gas, and nuclear. One of the most beneficial aspects of this procedure is the reduction of cost due to application of an elevated hardness, wear- and/or corrosion-resistant coating on a base metal whose function is strictly structural, when compared with solid noble alloy construction.

An important aspect of welding overlay is the dissimilarity of materials employed in the cladding process, in other words, a great difference of chemical composition, physical, mechanical and metallurgical properties between filler and base metal.

The main materials used as filler metals to hardfacing applications include iron-based, cobalt-based, nickel-based, and copper-based alloys, high-chromium white irons, carbides, and low-alloy ferrous materials (Davis 1993). The corrosion-resistant materials commonly used for cladding are austenitic stainless steels and Ni-based alloys. As for base metal, these are typically carbon, C-Mn or low-alloy steels, stainless steel, cast-iron, nickel-based alloys, and others.

With regard to dissimilar welding, some special characteristics of the weld bead should be achieved to guarantee high-quality, defect-free coating, and also with high performance in service. One of the most important characteristics is the deposition of weld with a minimal dilution level.

Dilution is the term used to define a change in the chemical composition of filler metal that was deposited due to mixture with base metal melted. Mathematically, dilution is defined as the percentage of the base metal that has melted during welding and has been incorporated to the fusion zone to compose the final weld metal chemical composition (AWS 2004).

Then, depending on dilution, weld metals with different chemical compositions than the original ones can be obtained. Dilution can be evaluated directly through chemical analysis of weld metal or estimated by the measurement of the cross-sectional area of the weld bead or layer (Fig. 1). In such a case, (1) is used to determine the percentage of moisture present in the melted base metal added with filler metal.

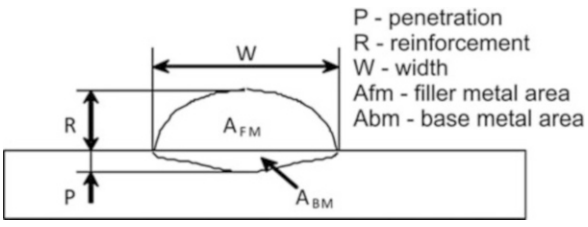
The A_{BM} corresponds to the base metal area melted during welding and A_{FM} corresponds to the area added by deposition of melted filler metal.

$$D = \frac{A_{BM}}{A_{BM} + A_{FM}} \times 100 \quad (1)$$

The importance of controlling the dilution level in overlay operations is attributed to the fact that the mixture between base and filler metals can significantly alter metallurgical characteristics such as microstructure, susceptibility to cracking and embrittlement, wear and corrosion resistance (Balasubramanian et al. 2009; Day et al. 2004).

The allowable dilution level depends on specific service requirements. For hardfacing applications, it is important to control dilution level less than 20 % to guarantee a harder layer with appropriated wear resistance.

For cladding applied as protective coating to petroleum and gas industry equipment, the dilution level is indirectly indicated by the standards, which specify the maximum iron content in weld overlay. The iron contents are separated in two categories: FE5, for iron contents equal or below to 5 %, and FE10, for coatings with 10 % or less in iron content. These considerations refer to coatings done with alloy 625 measuring 3 mm from the original surface of base metal to the surface of the coating (ISO 2009). Higher dilution levels are acceptable for other types of weld overlay as buildup and buttering.



Weld Overlay, Fig. 1 Weld bead geometry

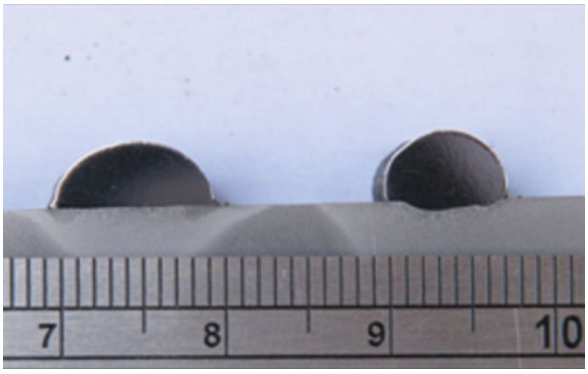
Geometric characteristics of weld bead are an important factor to guarantee the success of overlay operation. The most important geometric parameters to be evaluated are penetration, which corresponds to the maximum depth of melted metal into the base metal; and reinforcement, which corresponds to height of weld bead measured from base metal surface and width of weld bead measured on surface level.

The penetration of the weld bead should be as small as possible to avoid dilution-related problems. The reinforcement should be high enough to assure appropriate thickness according to welding procedure or standard specifications (ISO 2009; ASME Boiler and Pressure Vessel Code Committee 2004).

When post-welding finishing machining is necessary, the procedure should foresee the deposit of a thicker layer.

Wide and flat weld beads are preferable to increase the coated area and to minimize the surface irregularities, contributing to productivity increase and quality of the overlay. Another aspect to be observed is the reinforcement/width ratio (R/W). This geometric parameter indirectly evaluates the convexity of the weld bead. Weld beads that are very convex may hinder the overlap of subsequent passes, causing lack of fusion in the crossing among passes.

Figure 2 shows two examples of weld bead with different geometries. In the first case, the weld bead had low penetration, adequate reinforcement, and reasonable width. This geometry resulted in a 0.83 % dilution level, which was considered extremely low.



Weld Overlay, Fig. 2 Examples of single-weld bead of Ni-based alloy deposited on C-Mn steel. With arc oscillation (*left*). Without arc oscillation (*right*). Scale in mm. GTAW process with cold wire feed

The R/W ratio also was considered good, 0.38, which corresponds to a good convexity for a weld bead designated to overlay. In general, values of R/W ratios greater than 0.5 are indicative of weld beads that are exaggeratedly convex.

In fact, the R/W ratio of the second weld bead was 0.58, which exhibited a high convexity. Such behavior was attributed to a deeper penetration, minor width, and an increase in reinforcement, resulting in large dilution. This problem can be solved with the correct choice of welding parameters, as shall be discussed later.

An inappropriate selection of welding parameters leads to deposition of poor-quality weld beads. The formation of layers through the deposition of weld beads with inadequate geometry can result in internal defects, as observed in Fig. 3. The macrograph showed some defects identified as lack of fusion due to high convexity of weld beads. Although the defects are into the weld metal, their presence can be critical in some applications because of their elongated morphology that can concentrate stress on the tip and propagate a crack, especially when components are subject to cyclic loads.

As mentioned before, the welding parameters must be carefully adjusted to guarantee both a correct weld bead geometry and low dilution level.

Unfortunately, each welding process has its particularities and, sometimes, specific parameters. However, the main parameters to be adjusted are common to all processes that use electric arc as the heat source. These parameters are welding current, voltage, and travel speed. In addition, the product current versus voltage correspond to arc power. When the arc power is divided by travel speed, the heat input can be determined by the equation (2):

$$E = \frac{I \cdot V}{S} \cdot \eta \tag{2}$$

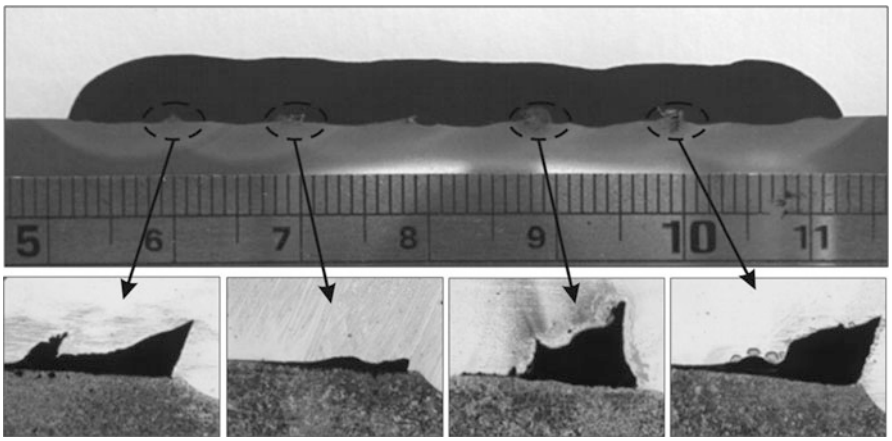
Where

I – welding current
V – voltage
S – travel speed
η – arc efficiency

The welding heat input is an essential factor to be selected due to its large influence on the quality of the overlay. Low heat input levels result in a poor deposition rate and make the process unproductive. Nevertheless, the increase in the heat input can possibly lead to increases in productivity, but the change can have disastrous effects on dilution. The welding heat input can be adjusted mainly by changing the current or travel speed.

As already mentioned, the welding current is an important parameter because of the strong influence it causes on heat input. On one side, the increase in current provides high fusion and deposition rates and, consequently, contributes to an increase in the amount of material deposited, resulting in productivity gain. On the other hand, the higher the welding current, the larger the amount of heat available to melt the base metal has to be.

In addition, high current levels make the actuation of electromagnetic forces influencing the flow of liquid metal on the weld pool more intense. Depending on the flow of the liquid metal on the weld pool, the weld bead penetration can be more or less deep. The arc voltage is the electric potential established between the electrode and workpiece. In general, this is associated with the arc length. In the welding processes with shielding gas, the arc



Weld Overlay, Fig. 3 Lack of fusion among passes due to high convexity

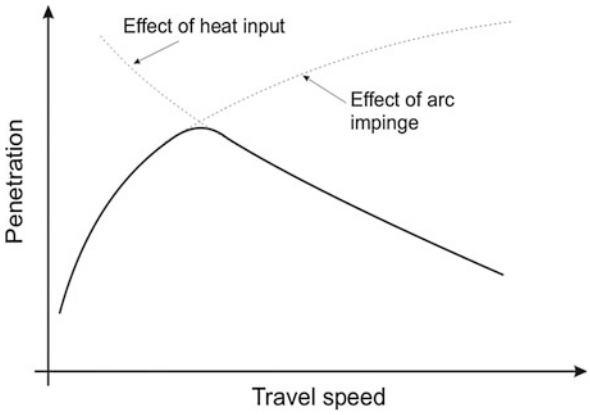
voltage is influenced by the type of gas or moisture used. However, due to the small variation caused, this is less effective for controlling the heat input as compared with current and travel speed.

Travel speed can be defined as the velocity at which the arc is moved along its trajectory. As mentioned, the travel speed has a strong effect on the heat input. The combination of moderate to high current levels with low travel speed may result in overheating, causing excessive melting of base metal and increasing the dilution. In addition, this parameter influences the weld bead penetration and profile.

When travel speed is sufficiently slowed to impinge the arc on the weld pool, the metal deposited acts as a barrier to the action of the arc on the base metal, diminishing the penetration. However, if the travel speed is excessively high, the arc impinge occurs on the base metal ahead of the weld pool, eliminating the barrier caused by the reinforcement. Such behavior transfers more heat to the base metal, resulting in a deeper penetration and, consequently, in a highly diluted weld bead (Scotti and Ponomarev 2008).

In addition, with an increase in the travel speed, gradual reduction of the heat input is observed. The combination of these effects is shown in Fig. 4. As can be seen in the figure, there is a critical travel speed in which the combined effects of impinge and heat input are maximized. Before the critical travel speed, the arc impinge on the weld bead is the most important effect on the penetration. After this point, the effect of the heat input exerts stronger influence on the penetration of weld bead than the former.

Welding with high travel speed decreases the penetration and, consequently, the dilution, and produces in general a small weld bead with inadequate geometry.



Weld Overlay, Fig. 4 Combined effects of arc impinge and heat input on penetration as a function of travel speed

For the above-mentioned reasons, the use of low travel speed in welding processes is recommended.

Another control parameter of great interest in the weld overlay is the arc length, which is defined as the distance from electrode tip to base metal. Short arc length causes a compression of the arc on a small anodic region (surface of base metal), which implicates it in a high density of current and strong pressure of the arc on the weld pool, causing an increase in penetration and dilution.

Arcs with longer lengths may become more extended, increasing the surface area of weld pool and reducing the density of current. Such fact leads to a better distribution of heat on the surface of the workpiece, therefore implicating it in a shallower and wider weld pool, with low penetration and dilution.

Mechanical arc oscillation is another parameter that can be adopted to increase the width of weld bead and to better distribute the heat from the arc on the base metal surface, avoiding excessive dilution.

The procedure consists of applying relative motion to the arc from side to side, orthogonal to the direction of travel. Several types of trajectories can be employed, for example, triangular, spiral, or double-8. Figure 2 shows an example of a single-pass weld where arc oscillation was the only parameter modified.

Research Findings

Welding Process

The selection of the welding process is an important step in the overlay operation and must be based on the engineering factors. For example, the type and composition of the filler metal, required properties of hardfacing or cladding, quality requirements of the deposited layer, metallurgical aspects of the base metal, metallurgical compatibility between filler and base metal, geometry, size and weight of workpiece to be coated, and operational cost.

Several welding processes have been improved for overlay applications, each with its own advantages and particularities. The main welding process used in overlay operations will be discussed later.

Shielded metal arc welding (SMAW) is typically a manual process in which an arc is established between the covered electrode and the workpiece, producing a pool of liquid metal.

The electrode is basically a metal core, which consists of a wire, covered by a mixture of mineral and metal powders and/or organic compounds. The functions of the covering ingredients are, among others, to produce gas to shield the arc and weld pool, to contribute for the

plasma arc ionization, to produce a slag cover on the weld metal conferring form and protecting the melted metal, to introduce alloy elements to confer desirable metallurgical and mechanical properties, and in some cases to enhance the deposition rate.

In this process, the welder is responsible for feeding the electrode into the weld pool and for moving it. Because of dilution, it is recommended to use a welding current lower than half of the amperage range specified by the manufacturer (AWS 2004).

The high versatility and flexibility of this process, combined with low cost due to simple, inexpensive, and portable power sources, absence of auxiliary shielding gas or flux, and the capacity of welding a large variety of metals and its alloys can be regarded as great advantages of SMAW process (Davis 1993).

Nonetheless, the low deposition rate due to limited maximum useful welding current, low operator factor (i.e., open arc time/total labor time to weld) because of the time needed to reload new electrodes and to clean the previous weld bead, and removing the slag before resuming the weld operation, make the SMAW process slow and expensive for welding overlay, except when low volume of deposition is required or in maintenance operations.

The *submerged arc welding* (SAW) process has been used to weld overlay deposition on carbon and low-alloy steel. The fusion of the metal is produced by an arc established between a bare metal electrode and the workpiece. The arc is enveloped by a blanket of granular fusible flux, which functions to protect the weld pool against atmospheric contamination, to guarantee the stability of the voltaic arc, to avoid spattering and sparks, and to prevent fumes and radiation emission. The flux also produces changes in chemical, metallurgical, and mechanical properties in the weld metal.

Advantages of the SAW process are a high deposition rate due to high current levels and large dimension of the electrode. In some instances, multiple wires or metallic strips can be simultaneously utilized instead of a single wire as filler metal, improving the productivity.

The operational configuration of the SAW process can be semiautomated, mechanized, and automated. A great inconvenience of this process is the high dilution level, as much as 50 %, demanding, sometimes, the deposition of three or more layers to promote a gradual reduction of dilution until an acceptable chemical composition is achieved.

Gas metal arc welding (GMAW) is another process that has been used to weld overlay operations. This process uses an arc established between a continuously fed wire electrode and the workpiece. The shielding of the weld pool is obtained by an external supply of a shielding gas. In general, the most commonly used shielding gases for GMAW are argon, helium, CO_2, and mixtures.

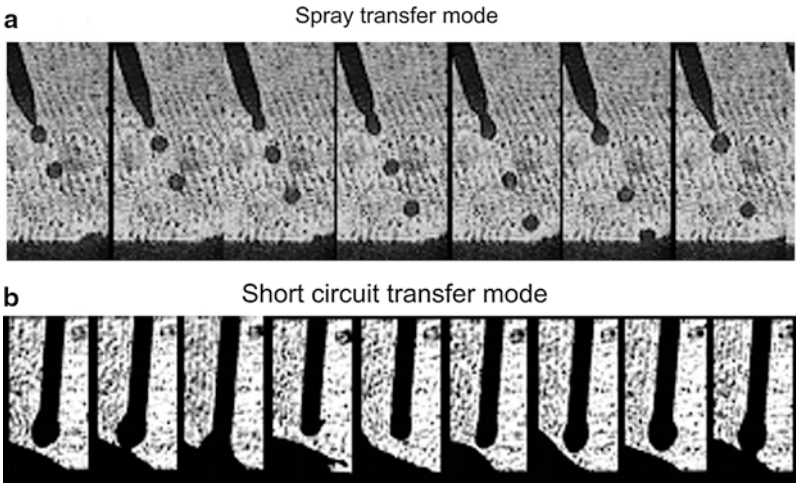
The continuous wire feeding awards to this process a high deposition rate, significantly higher than SMAW, and a high productivity due to deposition of long weld beads without the need of pauses for reloading and cleaning.

Another beneficial feature of wire continuous feeding is the possibility of utilization of higher travel speeds, notably increasing the productivity. In addition, the weld is very clean due to the absence of slag, requiring a minimal cleaning after weld. A smaller weld pool, in contrast to SAW, makes it possible to weld in all positions. Other characteristic is that the process can be easily automated.

Although this process has shown several advantages as compared with the others presented thus far, certain limitations can restrict its utilization. The first is the complexity and cost of the equipment compared with SMAW. Another factor is that many filler metals or materials for hardfacing cannot be drawn into wire or rod forms because of low ductility or brittle nature, limiting the utilization to a smaller variety of alloys. In addition, for outdoor applications under the action of strong airflow a larger supplement of shielding gas should be supplied to avoid contamination.

For the GMAW process, the parameters can be adjusted to obtain spray metal transfer, characterized by small droplets of melted filler metal propelled axially across the arc to the base metal (Fig. 5a). This transfer mode produces a very stable process, absence of spatter, and a high-quality weld bead, however, dilution levels are high. Short circuit transfer is another transfer mode where the liquid metal at the wire tip is transferred during a fast contact with weld pool (Fig. 5b). Although this transfer mode causes instability of the arc and spattering, the dilution level is smaller than spray transfer and has been a good option to overlay applications when a strict control of dilution is needed. In some cases, pulsed current is preferred to reduce the dilution level. In this case, a one droplet detachment per pulse transfer mode is the desired condition.

The *flux cored arc welding* (FCAW) process is similar to GMAW. Both processes use the same set of equipment, including a continuous wire feeding system. A metallic tubular electrode with an inner fusible flux is used as filler metal in the FCAW process.



Weld Overlay, Fig. 5 Filler metal transfer mode. (**a**) Spray mode. (**b**) Short-circuit mode (Miranda et al. 2004)

Similar to the SAW and SMAW processes, the flux present inside the tubular electrode is of great importance in the welding process and introduces metallurgical characteristics to the weld metal that could not be done in the case of the GMAW.

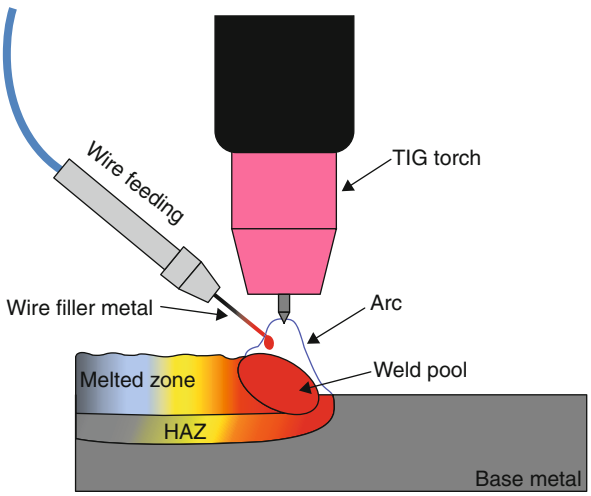
The shielding of weld pool and arc stability are achieved using a gas supply (gas-shielded) or by the fusion of flux (self-shielded) forming a slag cover and some gases that maintain the arc stable.

The most notable advantage of the FCAW process is the high productivity due to the continuous feeding of filler metal along with the opportunity of manipulating the chemical composition of the weld, making it possible to deposit several alloys, which is not possible with the GMAW process. FCAW, therefore, reaches a higher deposition rate and productivity when compared with SMAW.

Gas tungsten arc welding (GTAW) is a process where the arc is established between a nonconsumable electrode and the workpiece. The protection of the weld pool is accomplished by an inert gas flux. In this case there is not deposition of material and the unique material melted is the base metal (autogenous welding).

The electrode is made using pure tungsten or this element doped with cerium, lanthanum, thorium, or zirconium to improve the electron emissivity. The use of tungsten is due to its high melting point (3,410 °C), which allows the thermionic emission.

For hardfacing or cladding operations, where addition of material is needed, an extra supply of filler metal is assisted by a wire feeding system. The wire feeding is possible in two configurations, cold wire, when the filler metal is added at room temperature, and hot, wire when the filler metal is preheated using electric resistance.



Weld Overlay, Fig. 6 Configuration of the GTAW process with cold wire feed

When cold wire is used, the filler metal should be fed into the weld pool, crossing the arc welding to absorb the maximum amount of heat available (Fig. 6). The hot wire can be fed in front of or behind the arc, not necessarily being fed into the arc.

The great advantage of this process is associated with the independence between the heat source and filler metal. Such independent settings allowthe adjustment of arc parameters and wire feeding speed to ensure that the arc heat melts the filler metal as much as possible, resulting in little heat input to melt the base metal and producing a weld bead or layer with very low dilution level. Remember that with hardfacing or cladding the wire feed speed should be adjusted as necessary.

The success of the weld overlay deposited by this process is dependent on several parameters besides current and heat input, such as arc length, geometry of tungsten electrode tip, and configuration and geometry of the wire feeder system. When correctly adjusted, the result is a high-quality weld bead deposited with dilution level as little as 1 %, as can be observed in Fig. 2.

Other advantage of GTAW are the high quality of the weld bead, and versatility as it can be used in either the manual or automated version. Some of the disadvantages of this process are the low rate of deposition compared with other processes, the relative complexity of adjusting a large number of minor parameters, and difficulties of welding in outside environments or windy areas.

Plasma arc welding (PAW) is considered an evolution of GTAW process because they use the same principle (an arc electric opened between an nonconsumable electrode and the workpiece). However, in the PAW process the arc is constricted by an orifice, causing a contraction of the arc diameter and a higher density of energy (AWS 2004). Two distinct fluxes of inert gas are necessary to promote the process. The first is the plasma gas, where a flux of gas is fed through the constrictor nozzle surrounding the electrode, promoting the ionization of the environment and maintenance of arc, and producing a hotter and stronger gas jet. The second flux of gas is responsible for shielding the weld pool against atmospheric contamination.

Two variations of the PAW process are found, depending on how the arc plasma is established. When the arc is opened between the electrode and workpiece, it is called a *plasma-transferred arc*. However, if the arc is established between the electrode and the nozzle, the process is called a *plasma-nontransferred arc*. The first kind of process is used to melt electric conductor materials such as metals and alloys. The second kind is usually applied to weld nonconductor materials such as glass or polymers. In this case, the heat is transferred only by the plasma jet gas (AWS 2004). For weld overlay applications the first technique establishes the arc is directly on the base metal.

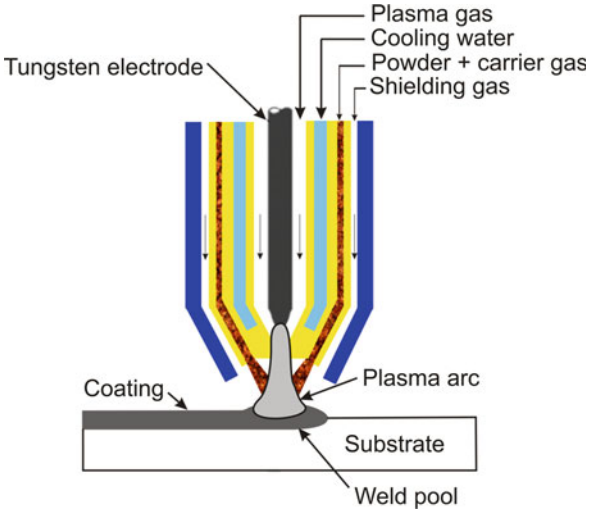
The deposition of material in this process is similar to that described for the GTAW process, for which additional equipment is required to feed the filler metal into the arc to form the weld pool. The same configurations (cold wire and hot wire) are used for the plasma arc welding overlay.

Among all characteristics of this process, the most important are the high arc stability, high density of energy, and excellent control of the heat source.

Plasma transferred arc (PTA) is one of the most notable processes used for overlay applications today. Although PTA is also used to characterize the mode of formation of the arc in the plasma arc welding process, industrially PTA is recognized as special surfacing technique that uses the material in powder form to deposit layers for specific applications such as hardfacing or cladding (AWS 2004).

As already mentioned, the heat source is a plasma transferred arc and a flux of powdered material is introduced into the arc, which melts the powder and deposits it on a surface promoting a weld bead. Three fluxes of inert gas are used in this process: the first is plasma gas flux, the second carries the powder and injects it into the arc, and the third flux shields the weld pool against atmospheric contamination (Fig. 7).

An important aspect of the PTA process is the of use the filler materials (which can be metal, alloys, carbides, or other hard particles) in powder form. This eliminates



Weld Overlay, Fig. 7 Plasma-transferred arc hardfacing process (Cheng et al. 2008)

problems with the production of the filler metal in wire or rod, which becomes difficult for some hardfacing materials due to intrinsic mechanical properties. In addition, the use of powdered metal results in more spreading of melted material on the surface, increasing the area covered. As the heat of the arc is in great part used to melt the powder material, the result is a shallow weld pool, contributing to obtain a reduced dilution level, as low as 5 %, which is another attractive aspect of PTA process for weld overlay applications.

Key Applications

Weld cladding and hardfacing are a common practice for improving or restoring a specific surface property in equipment and components for a wide range of industries, including agriculture, mining, steel, paper, food processing, chemical, petrochemical, petroleum and gas, and nuclear.

In mining industries, where severe abrasive wear is encountered due to the superficial contact during extraction of rocks, crushing and pulverizing of minerals, and transportation. In this case, a hard and wear-resistant material is applied on the surface of the tractor parts and equipment that has direct contact with the abrasive material.

Another industrial sector in which the hardfacing is applied to reduce wear is the sugar and alcohol industry. In this industrial segment, there are some machines and equipment that squeeze and crush the sugarcane to produce a cane juice, which is later processed to obtain sugar or alcohol (ethanol fuel). Due to high loads and movement of the crushed cane, intense wear of the machine parts is observed during service. The wear is associated with several factors such as corrosion of the parts due to acidic cane juice, sliding contact among parts, and sliding of the bagasse (the fibrous part of the sugarcane) against the roller surface (Buchanan et al. 2007). In most cases, to guarantee a wear-resistant surface, layers are deposited by weld overlay. Iron-based alloys are the most popular group of hardfacing materials used for this application in the sugar and alcohol industry due to their relatively low cost and ease of application (Buchanan 2009).

In transformation industries such as steelmaking and metalworking, some activities such as heavy forging and rolling are subject to severe service conditions because of high temperature and high loads applied. In these cases, hot forging dies and rolling mill rolls undergo several forms of damage associated with abrasive wear, oxidation wear, thermal and mechanical fatigue, and plastic deformation, among others (Gulenc and Kahraman 2003; Fouilland et al. 2009). To improve or restore a wear-resistant surface, the weld overlay is considered a good alternative, from the point of view of economy and quality, to extend the service life of the tools. Co- and Ni-based alloys are the most common materials used as filler metal for overlay applications because of their strength and resistance to oxidation at high temperatures (Fouilland et al. 2009).

Cross-References

► Laser Cladding
► Laser Welding

References

ASME Boiler and Pressure Vessel Code Committee, *ASME Boiler and Pressure Vessel Code*. Section III, Division I. (The American Society of Mechanical Engineers, New York, 2004)

AWS, *Welding Process – Part 1*, vol. 2 (American Welding Society, Miami, 2004)

V. Balasubramanian, R. Varahamoorthy, C.S. Ramachandran, C. Muralidharan, Selection of welding process for hardfacing on carbon steels based on quantitative and qualitative factors. Int. J. Adv. Manuf. Technol. **40**, 887–897 (2009)

V.E. Buchanan, Solidification and microstructural characterization of iron–chromium based hardfaced coatings deposited by SMAW and electric arc spraying. Surf. Coat.Technol. **203**, 3638–3646 (2009)

V.E. Buchanan, P.H. Shipway, D.G. McCartney, Microstructure and abrasive wear behavior of shielded metal arc welding hardfacings used in the sugarcane industry. Wear **263**, 99–110 (2007)

J.B. Cheng, B.S. Xu, X.B. Liang, Y.X. Wu, Microstructure and mechanical characteristics of iron-based coating prepared by plasma transferred arc cladding process. Mater. Sci. Eng. A **492**, 407–412 (2008)

J.R. Davis, Hardfacing, weld cladding, and dissimilar metal joining, in *Welding, Brazing and Soldering*, vol. 6, 10th edn. (ASM International, Metals Park, 1993), pp. 1967–592

J. Day, X. Huang, M. Yao, Study on composition-induced microstructural variation in the interface between co-based hardfacing alloys and IN738 Ni-based superalloy. J. Mater. Eng. Perform. **13**(2), 158–166 (2004)

L. Fouilland, M. El Mansori, A. Massaq, Friction-induced work hardening of cobalt-base hardfacing deposits for hot forging tools. J. Mater. Process. Technol. **209**, 3366–3373 (2009)

B. Gulenc, N. Kahraman, Wear behavior of bulldozer rollers welded using a submerged arc welding process. Mater. Design **24**, 537–542 (2003)

H.C. Miranda, V. Ponomarev, A. Scotti, V.A. Ferraresi, Optimisation of MAG-CO2 welding short-circuiting metal transfer. Int. J. Joining Mater. **16**(2), 43–48 (2004)

ISO 10423, *Petroleum and Natural Gas Industries – Drilling and Production Equipment – Well Head and Christmas Tree Equipment*. 3 rd edn. (International Standard, Geneva, Switzerland, 2009)

A. Scotti, V. Ponomarev, *MIG/MAG Welding. Better Understanding, Better Performance* (Ed. Artliber, São Paulo, 2008) [In Portuguese]

Wet Clutch Friction Material: The Surfaced Groove Effect

J. Y. Jang, M. M. Khonsari
Department of Mechanical Engineering, Louisiana State University, Baton Rouge, LA, USA

Synonyms

Automotive Transmission; Band Clutch or Brake; Clutch Friction Materials; Cone Clutches and Brake

Definition

Grooves are either machined or molded into the friction material to improve the performance of a wet clutch by enhancing the ATF (automotive transmission fluids) flow rate, thus providing a means for better dissipation of heat during engagement. Grooves can also improve friction and wear characteristics as well as durability.

Scientific Fundamentals

A disk-type wet clutch is an important component of an automatic transmission. Wet disk clutches are designed to provide adequate torque for the transmission of power from the driving member to the driven member during operation. A wet clutch pack consists of a series of separator and core disks to which friction-lining materials are bonded. The core and the separator disks are arranged alternately to a clutch hub that accommodates axial actuation and hydraulically presses the disks together. A wet clutch is immersed in an automatic transmission fluid (ATF) that lubricates and cools the surfaces and provides smoother performance with longer life. The surface features of the friction-lining are much rougher than the mating separator disk and, as such, they offer a fairly high frictional coefficient with relatively low wear-rate and stable operating characteristic. The wet clutch connects the driving and the driven members so that they may be locked together and rotate at the same speed (engagement phase), locked together and rotating at different speed (slipping phase), or unlocked and rotating at the different speed (disengagement phase).

In the first stage of engagement, pressure is applied hydraulically by means of a piston. Hydrodynamic pressure is developed in the ATF as a result of squeeze action, which supports most of the applied load. At this stage, the surfaces are separated by a relatively thick film of fluid and thus the behavior of the clutch is governed by the theory of hydrodynamic lubrication. During the engagement process, the film thickness is reduced to the extent that the surface asperities come into contact. As a result, the contact pressure at the asperity level begins to support a major portion of the imposed load, thus significantly influencing the behavior of the wet clutch. The film thickness undergoes a further reduction as the friction-lining material is compressed and deforms elastically. When the relative velocity reduces to zero, the two disks are locked together, and the engagement process is complete. It follows, therefore, that in a typical engagement cycle, the lubrication regime undergoes a transition from hydrodynamic to mixed or boundary lubrication regime. The time scale of the first stage of the engagement process is typically on the order of 1 s (Khonsari and Booser 2008).

During the engagement, a large amount of frictional heat is generated both due to viscous dissipation and also as a result of interaction of surfaces. Clearly, the viscosity of the ATF and, therefore, the entire operating performance of a wet clutch is influenced. In the absence of adequate cooling, high interfacial temperature rise can lead to thermal failure of the clutch face as a result of oxidation and decomposition of ATF. Various groove patterns on the friction material have been employed to provide more effective cooling and to facilitate the flow of ATF across the faces.

Types of Friction Material

There are four basic types of friction material available for use in wet clutches:

- Paper-based friction material is made of cellulose fibers with friction modifiers saturated in a resin to improve its performance. This material is widely used for ATF-immersed clutches. It has strong abrasion resistance, but it is susceptible to damage because of its relatively soft nature. This type of material offers good torque capacity, smooth and stable transmission, and low noise. Porosity of the paper-based friction material is considered to be an important property for controlling the friction characteristics.
- Sintered bronze is made by pressing tiny bronze powders tightly together, making it highly porous, tough, and durable. It can be used in applications that require high temperature resistance. However, it is susceptible to chatter noise.
- Graphitic material is manufactured by molding graphite and resin binders. Graphitic material has good thermal capacity and can be used in high-energy applications. However, in time it becomes susceptible to noise.

- Carbon material is ideal for applications that demand high load-bearing capacity and low wear. However, their use is limited because carbon material can be costly.

Types of Groove Pattern

Grooves on the face of the friction material provide oil flow channels from the inner radius to the outer radius and accommodate in dissipating heat during the engagement process. Typically, grooves are machined in the friction material after the friction material is bonded to the clutch plate. This manufacturing method is rather time consuming and expensive. The limited groove depth, which is due to the necessary machining tolerances, reduces the oil flow rate through the grooves and, thus, decreases the cooling capacity of the clutch plate. Alternatively, grooves can be molded into the friction material. However, molding method also limits the groove depth. Groove shapes, patterns, and manufacturing methods can influence shudder during the engagement.

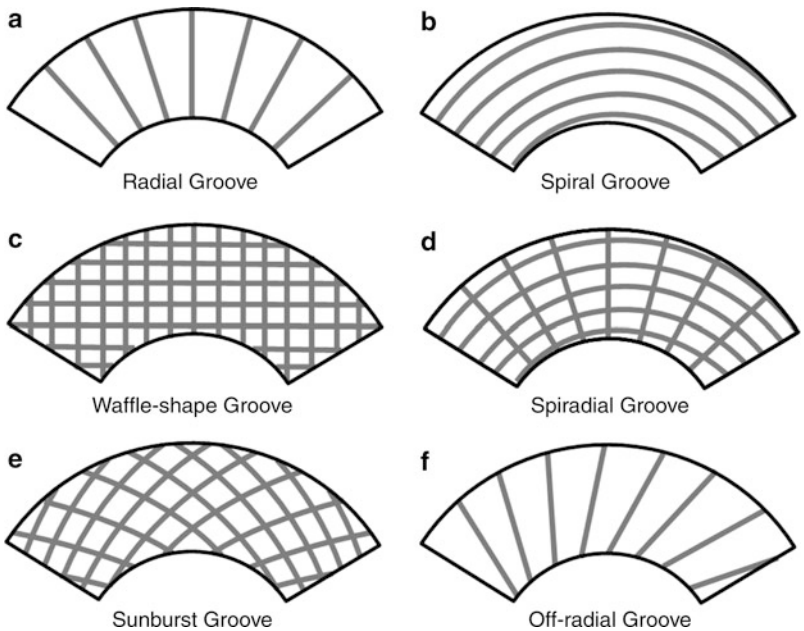
Most groove patterns are designed to force the lubricant to flow from the inner radius to the outer radius through the grooves by centrifugal forces. Therefore, when the clutch plate is stationary or operating at a relatively low speed, insufficient oil flows through the grooves, leading to inadequate cooling. In the absence of oil flow through the grooves, the clutch face temperature can rise and result in formation of macroscopic hot spots or thermal failure of the clutch facing (Jang and Khonsari 2002). Excessive frictional heat due to high temperature will also cause decomposition of the oil. A desirable groove pattern equalizes the surface temperature of the friction facings and thus increases the thermal capacity of the clutch. Numerous groove patterns exist depending upon the needs of each application. Typical groove patterns, shown in Fig. 1, are:

Radial grooving pattern: By far the simplest pattern, this type of grooving easily provides pathways for ATF to flow radially outward during the engagement phase. It is thus suitable if plenty of lubricant is available. Clearly with high flow rate, heat is dissipated quickly, keeping viscous drag low. One has to keep in mind that, similar to thrust bearings, excessive lubricant between the grooves can increase the parasitic drag losses (see Khonsari and Booser 2008).

Spiral grooving pattern: This type of grooving distributes cooling oil quickly but has the disadvantage of developing relatively high drag losses.

Waffle-shape grooving pattern: This type of grooving offers less drag loss compared with spiral groove pattern and demands less flow requirement than radial groove pattern.

Spiradial grooving pattern: This type of grooving may be used in wet clutch applications where there is limited oil available.



Wet Clutch Friction Material: The Surfaced Groove Effect, Fig. 1 Typical groove patterns

Sunburst grooving pattern: This type of grooving is a relatively common design and generally suitable for ordinary applications in which sufficient oil is available. Sunburst groove patterns operate at lower drag levels.

Off-radial grooving pattern: This type of grooving is a variation of the radial grooving pattern. The orientation of off-radial grooves significantly affects the squeezing process and, therefore, the torque.

There are many properties of the wet clutch that influence the friction characteristics. They include material parameters, lubricant, surface roughness, elasticity, and permeability of the friction material. Permeability, the capability of the fluid flow inside the friction material, influences both the engagement time and the temperature profile due to the change of lubricant flow. The engagement time decreases due to a smaller squeezing action and the temperature decreases due to the fluid flow in the material with increasing the permeability. When the groove pattern is manufactured by pressing, there is a possibility that the groove walls become very dense. Therefore, dense groove walls can seal the friction material and friction material becomes almost impermeable.

Some of the pertinent analytical studies for investigation of wet clutch performance are available in the published literature. Payvar (1991) developed a numerical solution of the heat transfer equations with simplified assumptions for computing the heat transfer coefficient in the grooved wet clutch. Yang et al. (1995) developed an axisymmetric model to describe the heat transfer in a wet clutch. Their predictions of interface temperature compared favorably with experimental measurements. Berger et al. (1996) developed a finite element model and later on reported an analytical solution (Berger et al. 1997) for modeling the torque response of the grooved wet clutch during the engagement. Jang and Khonsari (1999) presented a comprehensive model capable of considering the thermal effect in wet clutches with radial grooves. Razzaque and Kato (1999a) also developed an analytical model for the radial and off-radial groove patterns by applying the narrow groove theory to investigate the transient engagement characteristic of the wet clutch. The narrow groove theory is based on the assumption that the number of groove is large (i.e., the grooves are narrow enough). They found the orientation angle to be one of the most important parameters.

Key Application

Automotive Transmission

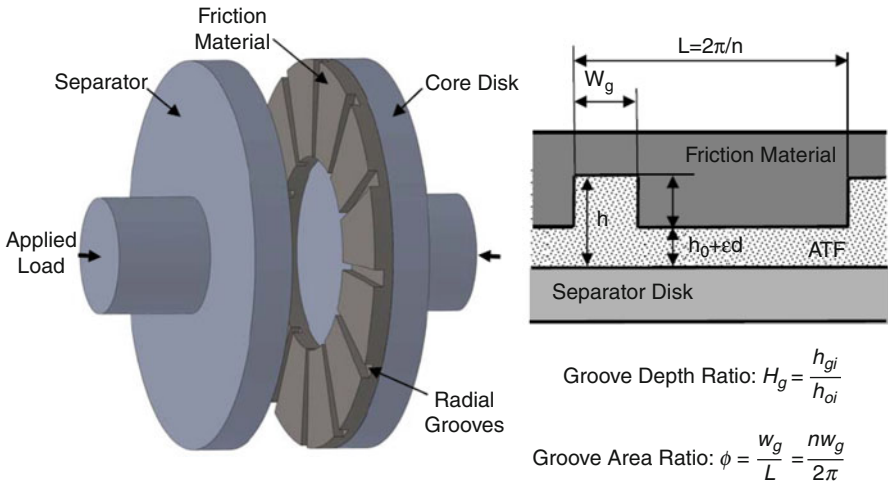
An automotive transmission is a device that provides different gear ratios, as well as neutral and reverse between the engine and drive wheels. It enables the vehicle to accelerate to a desired speed by converting the torque. The transmission input shaft connects through the disk-type wet clutch and flywheel to the engine crankshaft. To get the vehicle in motion, the transmission allows the drive wheels to turn slowly with large torque. When the vehicle is accelerated, the transmission provides less torque but the drive wheels turn at a higher speed. Four-wheel drive (4WD) vehicles transfer the torque from the engine to all four wheels simultaneously and provide better control on rough roads.

Orientation of Radial and Off-Radial Grooves

Razzaque and Kato (1999b) presented a model for investigating the effect of radial and off-radial grooves on the wet clutch in the disengaged state by neglecting the squeeze action and focusing only on the hydrodynamic behavior of the wet clutch. They considered straight grooves with rounded, trapezoidal, and V-section at different angular orientations. Their results show that compared with rounded and trapezoidal grooves, V-section grooves produce higher torque and lower flow-rate due to their smaller cross-sectional area (typically 3/4 of the cross-sectional area of rounded grooves). They also showed that rounded grooves produce the smallest viscous torque and that the flow rate is almost identical to that of trapezoidal grooves because the cross-sectional area is the same. Razzaque and Kato showed that the viscous torque becomes maximum at $\beta = -8^\circ$, where β is the orientation angle. A large number of grooves produce larger viscous torque. The effect of groove numbers on the viscous torque disappears at a large β value, either positive or negative. The flow rate is maximum when $\beta = 50^\circ$, where the radial velocity is high. At a higher radius ratio, defined as the ratio of the inner radius to the outer radius, the radial span decreases and the interior area is influenced more by the flows, which causes a drop of the viscous torque. However, at a large β value, either positive or negative, flows become directional and the effect of the radius ratio disappears.

Thermal Effect in a Wet Clutch with Radial Grooves

Jang and Khonsari (2011) developed a comprehensive model to investigate the effect of radial grooves and waffle-shaped grooves on the performance of a wet clutch. The model was based on the three-dimensional formulation of the governing equations with provisions for the thermal aspects. Roughness, centrifugal force, deformability, and permeability of the friction material with grooves were taken into account. They showed that



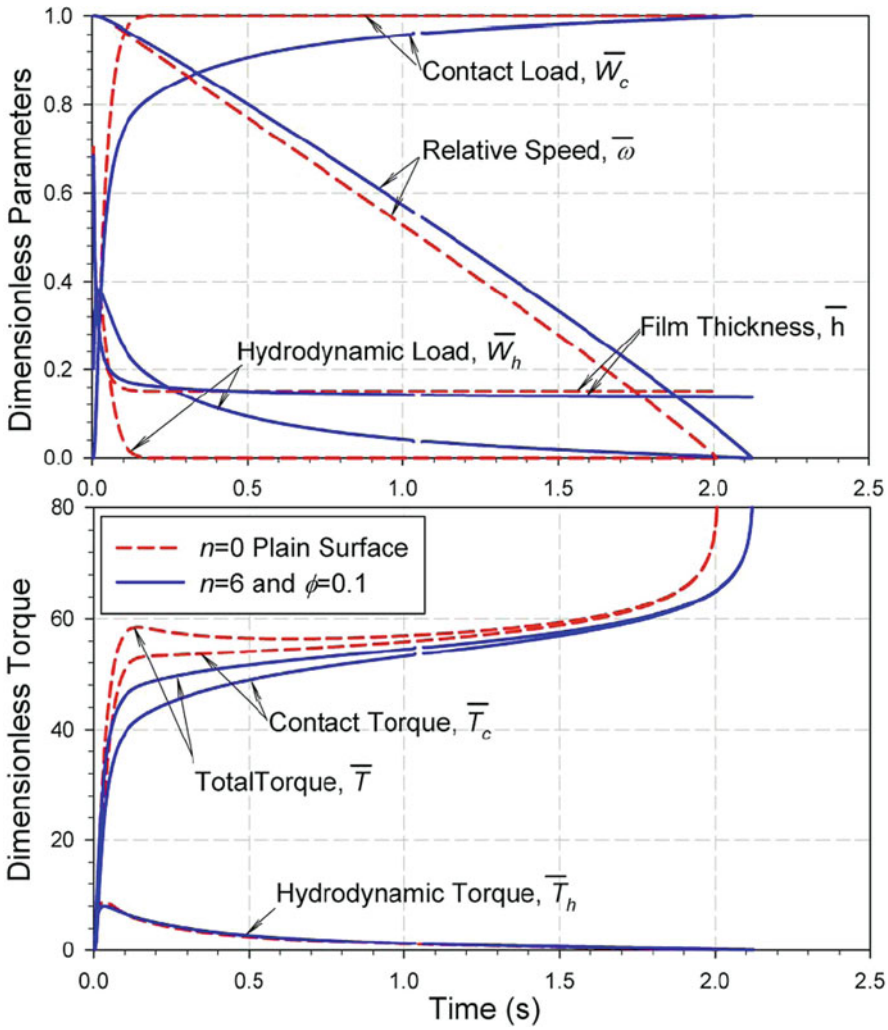
Wet Clutch Friction Material: The Surfaced Groove Effect, Fig. 2 Analytical model of clutch with radial grooves

effects of groove geometry such as groove depth, grooved area, and number of grooves on the engagement characteristic of a wet clutch play an important role. Figure 2 shows the schematic of a clutch with radial grooves, where h_{gi} represent the initial groove depth, h_{oi} represents the initial film thickness, w_g denotes the groove width in radian, d represents the thickness of friction material, and ε represents the compressive strain and n is the number of radial grooves.

Figure 3 shows how the hydrodynamic film thickness, relative speed, load, and torque at 2,000 rpm vary with time. The solid lines represent the thermohydrodynamic (THD) results with six radial grooves while the dashed lines represent the THD results without grooves. Thermohydrodynamic (THD) analysis takes into account the influence of temperature on the hydrodynamic lubrication process, especially on the viscosity of the lubricant. Therefore, the energy equation is coupled with the rheological equations through the temperature-viscosity relationship. The groove depth ratio is $H_g = 1$ and the grooved area ratio is $\phi = 0.1$. The input data is listed in the paper by Jang and Khonsari (2011). The dimensionless torque is defined as $\overline{T} = h_{oi}T/\mu_i b^4 \omega_i$, the dimensionless film thickness is $\overline{h} = h/h_{oi}$, the speed is $\overline{\omega} = \omega/\omega_i$, and the dimensionless load per unit area is $\overline{W} = W/P_o$, where μ_i is the initial viscosity, b is the outer radius, ω_i is the initial speed, and P_o is the applied pressure. At the beginning of the engagement process, the gap between surfaces is filled with ATF and completely separated without any asperity contact. When the load is applied, the film thickness drops very rapidly at the early stage of the engagement process. As the separator and the friction surface come into contact, asperities begin to support the load and the portion of the load supported by the hydrodynamic load decreases. For a plain friction surface without grooves, the hydrodynamic load-carrying capacity becomes nil at about 0.2 s, the film thickness reaches its minimum, and the entire load is taken up by the asperity contacts. The minimum film thickness remains constant until the end of the engagement process, where the relative speed between the separator and the friction disks becomes nil. The relative speed decreases at an approximately linear rate. For the radial grooves, the hydrodynamic load decreases continuously until the end of the engagement since the hydrodynamic pressure is generated within each groove. Consequently, the engagement time with radial grooves is longer than that of plain friction surface.

Figure 3 also shows the variation of viscous torque, contact torque, and total torque. At the beginning of the engagement, the contact torque increases very rapidly until the asperity contact starts to support the applied load.

The contact torque for the plain friction surface without grooves is higher than that of the radial grooved friction surface since there is less contact area involved. For a plain friction surface without grooves, it is shown that the total torque reaches its peak when the film thickness is at its minimum. After reaching its peak, the total torque remains relatively flat for a length of time, and then begins to rise gradually as the relative speed between the two disks decreases. The increase in the torque is a consequence of the change in the friction coefficient corresponding to the reduction in the speed, as the lubrication regime undergoes a transition from fully



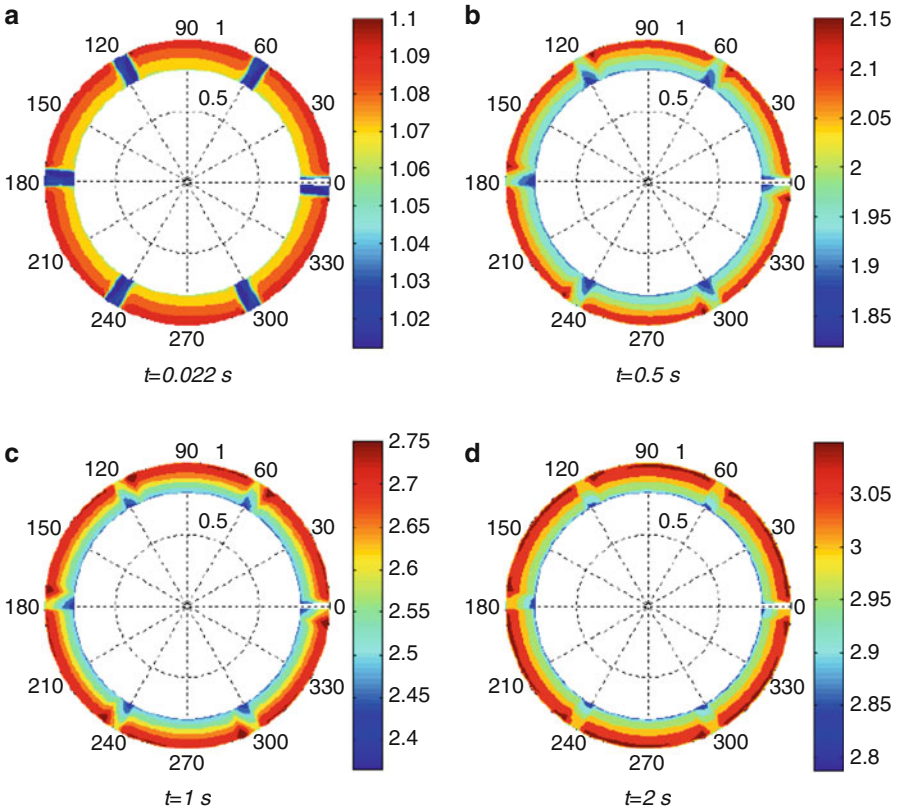
Wet Clutch Friction Material: The Surfaced Groove Effect, Fig. 3 Variation of dimensionless parameters at 2,000 rpm (ϕ = 0.1 and H_g = 1)

hydrodynamic to boundary lubrication (Khonsari and Booser 2010). At the end of the engagement the total torque tends to suddenly increase, which is sometimes referred to as the lockup torque or rooster's tail. This highly undesirable sudden spike of torque can be reduced by altering the friction behavior.

Figure 4 shows the dimensionless temperature contour at the surface of the friction material at four instances at 2,000 rpm. The dimensionless temperature $\overline{\Theta} = \Theta/\Theta_i$, where the initial temperature Θ_i = 100°C is used in this simulation. At t = 0.022 s as shown in Fig. 4a, where the hydrodynamic squeeze action is maximum, the temperature inside the grooves is clearly distinguishable from the temperature in the ungrooved area since the flow rate in grooves is large at this instant in time. It is shown that the cooling effect in grooves is reduced as the engagement proceeds since the film thickness and the flow-rate decrease after the squeezing motion disappears. Note that the temperature near the outer radius is higher than that near the inner radius due to the higher linear velocity. In the ungrooved area, the temperature remains relatively constant in circumferential direction due to the rotation of the disk.

Effect of Radial Groove Depth Ratio on Engagement Time

Figure 5 shows the engagement time with variation of groove depth ratio, H_g, grooved area ratio, ϕ, and radial



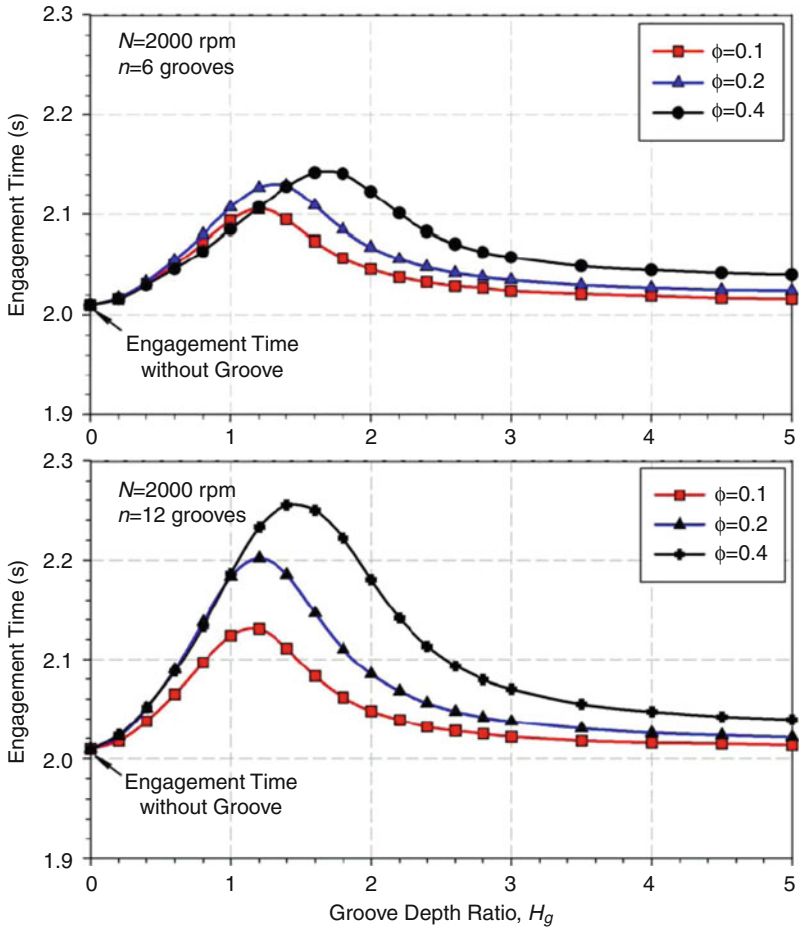
Wet Clutch Friction Material: The Surfaced Groove Effect, Fig. 4 Dimensionless temperature contour of radial grooves at 2,000 rpm ($\phi = 0.1$ and $H_g = 1$)

groove number at 2,000 rpm with consideration of the viscosity change due to the temperature rise. Also shown for baseline comparison is the engagement time for the plain friction pair without the groove, represented by $H_g = 0$. The engagement time increases as the number of grooves, n, increases since the hydrodynamic pressure generated in each groove contributes to an additional load-carrying capacity. At the end of the engagement, the hydrodynamic load becomes nil since the relative speed is zero and friction pair is locked up. Therefore, at the end of the engagement the entire load is supported by the contact load and the film thickness is minimal. The film thickness with grooves is less than that without grooves due to the reduction of the contact area.

For shallow-depth grooves, the engagement time increases with increasing the groove depth ratio, H_g, due to the higher hydrodynamic pressure generated within the grooves. As the hydrodynamic pressure within the grooves increases, the hydrodynamic load-carrying capacity increases and, consequently, the contact load decreases during the engagement to satisfy the force balance. The net result is a smaller torque and the engagement time increases to complete the engagement time. However, beyond a certain value of the groove depth ratio the squeezing velocity becomes larger and reduction in the film thickness is faster with increasing the groove depth ratio due to the less resistance to the squeezing motion. Therefore, the film thickness approaches its minimum quickly. Consequently, the contact load becomes larger and the engagement time decreases.

The engagement time is also influenced by the grooved area ratio, ϕ. The hydrodynamic load-carrying capacity increases with increasing the grooved area ratio, which results in a reduction in the contact load and the associated torque. Therefore, the engagement time becomes longer. However, the opposite trend is observed for the shallow-depth grooves.

Compared with the isoviscous results, the THD simulation results predict that the engagement time at very shallow and very deep grooves is much greater. Considering thermal effects, the engagement time is influenced by the viscosity drop, yielding a smaller hydrodynamic



Wet Clutch Friction Material: The Surfaced Groove Effect, Fig. 5 Effect of radial groove depth on engagement time based on THD analysis

torque, leading to a larger engagement time. However, in the middle range of the groove depth, the engagement time by the THD simulation is smaller.

Cross-References

► Band Clutch or Brake
► Clutch Friction Materials
► Cone Clutches and Brake

References

E.J. Berger, F. Sadeghi, C.M. Krousgrill, Finite element modeling of engagement of rough and grooved wet clutches. ASME J. Tribol. **118**, 137–146 (1996)

E.J. Berger, F. Sadeghi, C.M. Krousgrill, Analytical and numerical modeling of engagement of rough, permeable, grooved wet clutches. ASME J. Tribol. **119**, 143–148 (1997)

J.Y. Jang, M.M. Khonsari, Thermal characteristics of a wet clutch. ASME J. Tribol. **121**, 610–617 (1999)

J.Y. Jang, M.M. Khonsari, On the formation of hot spots in wet clutch systems. ASME J. Tribol. **124**, 336–345 (2002)

J.Y. Jang, M.M. Khonsari, Three-dimensional thermohydrodynamic analysis of a wet clutch with consideration of grooved friction surfaces. ASME J. Tribol. **133**, 011703-1–011703-12 (2011)

M.M. Khonsari, E.R. Booser, *Applied Tribology: Bearing Design and Lubrication*, 2nd edn. (Wiley, West Sussex, UK, 2008)

M.M. Khonsari, E.R. Booser, On the stribeck curve, in *Recent Developments in Wear Prevention, Friction and Lubrication*, ed. by G.K. Nikas (Old City Publishing, Philadelphia, PA, 2010), pp. 263–278

P. Payvar, Laminar heat transfer in the oil groove of a wet clutch. J. Heat Mass Trans. **34**, 1791–1798 (1991)

M.M. Razzaque, T. Kato, Effect of a groove on the behavior of a squeeze film between a grooved and a plain rotating annular disk. ASME J. Tribol. **121**, 808–815 (1999a)

M.M. Razzaque, T. Kato, Effect of a groove orientation on hydrodynamic behavior of wet clutch coolant films. ASME J. Tribol. **121**, 56–60 (1999b)

Y. Yang, R.C. Lam, Y.F. Chen, H. Yabe, Modeling of heat transfer and fluid hydrodynamics for a multidisk wet clutch, *SAE 950898* (Society of Automotive Engineers, Warrendale, 1995). doi: 10.4271/950898

Wet Clutch, an Introduction

Jaewon Choi, Michael D. Bryant
Department of Mechanical Engineering, The University of Texas at Austin, Austin, TX, USA

Synonyms

Fluid clutch; Hele-Shaw clutch; Viscous clutch

Definition

A wet clutch transfers power through mechanical and fluid couplings by mating rotating friction plates immersed in lubricant. What differentiates wet clutches from dry clutches is lubricant between the plates. The inter-plate fluid plays multiple roles, including torque transfer via fluid coupling, lubrication, separation of plates without actuator input applied, flushing particles, and cooling. A wet clutch solely based on fluid coupling is rare (Hele-Shaw et al. 1923). Most wet clutches transfer torque via mechanical friction and fluid interaction. Wet clutches are typically used in a compact and lubricant-rich environment, such as automatic transmission (Deur et al. 2005) because a wet clutch is typically smaller than a dry clutch, maximum transmittable torque per plate pair is limited for wet clutches. Multiple plates are often employed to compensate the limited torque per plate pair. Actuators are also embedded in the package. Hydraulic or electromagnetic actuators are common due to the size advantage (Ross et al. 2007). Synchronizing the timing of clutch engagement is important in many applications, for example, smooth gear shifting in automatic transmission.

Scientific Fundamentals

Principles of Operation

Wet clutches engage in two distinct stages depending on lubrication regimes: hydrodynamic lubrication and boundary lubrication. At the beginning of clutch engagement, the inter-plate lubricant fully accommodates speed differences of rotating plates. The velocity profile and the external pressure induce normal and shear stresses in the fluid, which result in thrust and tangential forces on the plates. In this stage, fluid coupling transfers torque, with operation similar to hydrodynamic thrust bearings. Unlike a thrust bearing, the mating surfaces are purposefully rough and flat. Some lubricant becomes trapped in the valleys of the rough surfaces and moves with the rotating plates. Because friction surfaces are made of porous materials, surface pressure and centrifugal force cause leakage through the circumference of the friction lining. As the plates are further pressed together by an actuator, microscopic asperities on mating surfaces contact and transmit torque. Meanwhile the contribution of fluid coupling becomes small to nil. After engagement, the maximum torque is largely transferred by contact. Statistical properties of random and complex friction surface profiles such as centerline average and roughness characterize the relationship between maximum torque and contact pressure. Lubrication evolves from hydrodynamic, through mixed, to boundary lubrication. Engagement of wet clutches differs from dry clutches because of the transient torque from fluid coupling, and an inherent time delay due to a squeeze film effect between plates, influenced by lubricant and surface properties.

Wet Clutch Model Backgrounds

Friction surfaces of wet clutches are flat and rough unlike hydrodynamic thrust bearings. Patir and Cheng's average flow model (Patir and Cheng 1979) corrects Reynolds equation for surface roughness effects on sliding surfaces, using empirical flow and stress correction factors. Effects of porous friction linings mandate other correction factors (Natsumeda and Miyoshi 1994) in the modified Reynolds equation for most wet clutch models (Berger et al. 1996; Deur et al. 2005). While the transient response is influenced by fluid interactions, performance at steady state depends on contact pressure. The dynamics of film thickness are affected by rough surfaces and the squeeze film effects.

Average Flow Model

Consider two smooth parallel plates 1 and 2 moving at speeds U_1 and U_2 in the x direction. Assuming constant density ρ, the Reynolds equation simplifies to

$$\frac{\partial}{\partial x}\left(\frac{h^3}{12\eta}\frac{\partial p}{\partial x}\right)+\frac{\partial}{\partial y}\left(\frac{h^3}{12\eta}\frac{\partial p}{\partial y}\right)=\frac{U_1+U_2}{2}\frac{\partial h}{\partial x}+\frac{\partial h}{\partial t} \quad (1)$$

Average flow $\bar{q}$ incorporates statistical surface properties into the Reynolds equation via empirical correction factors ϕ_k, calculated by regressing results of computer simulations wherein surface profiles were numerically generated in a random fashion based on specified statistical surface settings. Average flows in x and y are defined as

$$\bar{q}_x=-\phi_x\frac{h^3}{12\eta}\frac{\partial p}{\partial x}+\frac{U_1+U_2}{2}\bar{h}_T+\frac{U_1-U_2}{2}\sigma\phi_s \quad (2)$$

$$\bar{q}_y=\phi_y\frac{h^3}{12\eta}\frac{\partial p}{\partial y} \quad (3)$$

where ϕ_x and ϕ_y are the pressure flow factors in x and y, ϕ_s is a correction factor for shear flow, $\bar{h}_T$ is the average film thickness, and $\sigma = (\sigma_1{}^2 + \sigma_2{}^2)^{0.5}$ is the composite standard deviation of roughness of two sliding surfaces 1 and 2 having standard deviations of roughness $\sigma 1$ and $\sigma 2$ respectively.

Pressure flow factors ϕ_x and ϕ_y relate pressure-driven flows on rough sliding surfaces to that on smooth sliding surfaces. To illustrate shear flow factor ϕ_s (Patir and Cheng 1979), suppose a rough surface slides against a smooth stationary surface. Fluid caught in numerous valleys on the rough surface move with the rough surface, increasing average flow, giving $\phi_s > 0$. If the rough surface is stationary and the smooth surface slides, stagnant fluid in the valleys of the rough surface impedes average flow, giving $\phi_s < 0$. Considering these corrected flows, the extended Reynolds equation becomes

$$\frac{\partial}{\partial x}\left(\phi_x \frac{h^3}{12\eta}\frac{\partial \bar{p}}{\partial x}\right) + \frac{\partial}{\partial y}\left(\phi_y \frac{h^3}{12\eta}\frac{\partial \bar{p}}{\partial y}\right) = \frac{U_1 + U_2}{2}\frac{\partial \bar{h}_T}{\partial x} + \frac{U_1 - U_2}{2}\sigma\frac{\partial \phi_s}{\partial x} + \frac{\partial \bar{h}_T}{\partial t} \tag{4}$$

where $\bar{p}$ is the average pressure on sliding surfaces, and Patir and Cheng's flow factors are

$$\phi_x = 1 - Ce^{-rH} \text{ for } \gamma \le 1 \tag{5}$$

$$\phi_x = 1 + CH^{-r} \text{ for } \gamma > 1 \tag{6}$$

$$\phi_y(H,\gamma) = \phi_x(H, 1/\gamma) \tag{7}$$

$$\phi_s = V_{r1}\Phi_s(H,\gamma_1) - V_{r2}\Phi_s(H,\gamma_2) \tag{8}$$

$$\Phi_s = A_1 H^{\alpha_1}\exp\left(-\alpha_2 H + \alpha_3 H^2\right) \text{ for } H \le 5 \tag{9}$$

$$\Phi_s = A_2\exp(-0.25H) \text{ for } H > 5. \tag{10}$$

In the foregoing, $H = h/\sigma$, surface pattern parameter $\gamma = \gamma_{0.5x}/\gamma_{0.5y}$, $\gamma_{0.5x}$ and $\gamma_{0.5y}$ are 0.5 correlation lengths of surface profiles along x and y directions (characteristic length at 25% standard deviation), γ_1 and γ_2 are the γ of surfaces 1 and 2, and $V_{r1} = (\sigma_1/\sigma)^2 = 1 - V_{r2}$. The empirical flow factor coefficients C, r, A_1, α_1 α_2 α_3 and A_2 values are tabulated in Table 1. For an isotropic surface, the surface pattern parameter becomes unity ($\gamma = 1$). As a surface becomes smooth ($H \to \infty$), the pressure flow factor becomes unity ($\phi_{x,y} \to 1$) and the shear flow factor becomes zero ($\phi_s \to 0$).

Assuming a uniform pressure distribution along the moving direction (i.e., $\partial p/\partial x = 0$), the average shear stress is expressed as

$$\bar{\sigma}_{xy} = \frac{\eta(U_2 - U_1)}{h}\left(\phi_f \pm \phi_{fs}\right) \tag{11}$$

where shear stress correction factor ϕ_f accounting for averaging of local shear stress, is given by

For $H \le 3$,

$$\phi_f = \frac{35}{32}z\left[\left(1-z^2\right)^3\ln\left(\frac{z+1}{\varepsilon/3\sigma}\right) + \left\{\frac{-55 + z(132 + z(345 + z(-160 + z(-405 + z(60 + 147z)))))}{60}\right\}\right] \tag{12}$$

For $H > 3$,

$$\phi_f = \frac{35}{32}z\left[\left(1-z^2\right)^3\ln\left(\frac{z+1}{z-1}\right) + \frac{z}{15}\left\{66 + z^2\left(30z^2 - 80\right)\right\}\right] \tag{13}$$

In (12) and (13), where $z = H/3$, the empirical shear stress correction factor ϕ_{fs} due to effects of surface roughness is defined as

$$\phi_{fs} = V_{r1}\Phi_{fs}(H,\gamma_1) - V_{r2}\Phi_{fs}(H,\gamma_2) \tag{14}$$

$$\Phi_{fs} = A_3 H^{\alpha_4}\exp\left(-\alpha_5 H + \alpha_6 H^2\right) \tag{15}$$

The coefficients of (15) are tabulated in Table 2.

Wet Clutch, an Introduction, Table 1 Empirical flow factor coefficients (Patir and Cheng 1979)

	ϕ_x			ϕ_s					
γ	C	r	Range	A_1	α_1	α_2	α_3	A_2	Range
1/9	1.48	0.42	$H > 1$	2.046	1.12	0.78	0.03	1.856	$H > 0.5$
1/6	1.38	0.42	$H > 1$	1.962	1.08	0.77	0.03	1.754	$H > 0.5$
1/3	1.18	0.42	$H > 0.75$	1.858	1.01	0.76	0.03	1.561	$H > 0.5$
1	0.90	0.56	$H > 0.5$	1.899	0.98	0.92	0.05	1.126	$H > 0.5$
3	0.225	1.5	$H > 0.5$	1.560	0.85	1.13	0.08	0.556	$H > 0.5$
6	0.520	1.5	$H > 0.5$	1.290	0.62	1.09	0.08	0.388	$H > 0.5$
9	0.870	1.5	$H > 0.5$	1.011	0.54	1.07	0.08	0.295	$H > 0.5$

Wet Clutch, an Introduction, Table 2 Emprical shear stress factor coefficients (Patir and Cheng 1979)

	ϕ_{fs}				
γ	A_3	α_4	α_5	α_6	Range
1/9	14.1	2.45	2.30	0.10	$7 > H > 0.5$
1/6	13.4	2.42	2.30	0.10	$7 > H > 0.5$
1/3	12.3	2.32	2.30	0.10	$7 > H > 0.5$
1	11.1	2.31	2.38	0.11	$7 > H > 0.5$
3	9.8	2.25	2.80	0.18	$7 > H > 0.5$
6	10.1	2.25	2.90	0.18	$7 > H > 0.5$
9	8.7	2.15	2.97	0.18	$7 > H > 0.5$

Wet Clutch Model

Consider two circular disks immersed in lubricant, one stationary and the other rotating at angular speed ω. A friction lining is attached on the stationary plate, and an external force is applied on the other plate, as shown in Fig. 1. The diagram shows one pair in a multiple clutch disk pack, often employed in real applications. Each pair of a separate plate and a core plate with a friction lining can be modeled based on this simple two-plate model.

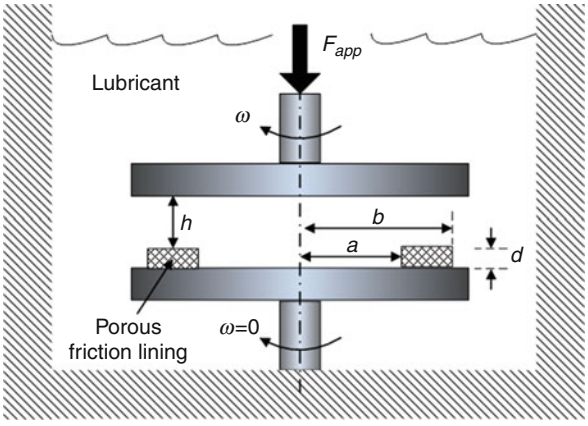
Assuming fluid density ρ constant, the extended Reynolds equation in cylindrical coordinates can be stated as

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\phi\, h^3}{\eta_s}\frac{\partial p}{\partial r}\right)+\frac{1}{r^2}\frac{\partial}{\partial \theta}\left(\frac{\phi\, h^3}{\eta_s}\frac{\partial p}{\partial \theta}\right) = -6\omega\left(\frac{\partial \bar{h}_T}{\partial \theta}+\sigma\frac{\partial \phi_s}{\partial \theta}\right)+12\frac{\partial \bar{h}_T}{\partial t} \tag{16}$$

Considering the centrifugal force acting on the fluid, permeability β of paper facing, strain ε of paper facing along the lining thickness d, and the exuding velocity u_z of fluid from the surface of porous and compressible paper facing, (16) revises to (Natsumeda and Miyoshi 1994)

$$\begin{aligned}&\frac{1}{r}\frac{\partial}{\partial r}\left[r\left\{\frac{\phi\, h^3}{\eta_s}+\frac{12\beta d(1-\varepsilon)}{\eta_m}\right\}\frac{\partial p}{\partial r}\right]\\&+\frac{1}{r^2}\frac{\partial}{\partial \theta}\left[\left\{\frac{\phi\, h^3}{\eta_s}+\frac{12\beta d(1-\varepsilon)}{\eta_m}\right\}\frac{\partial p}{\partial \theta}\right]\\&=\frac{3}{10}\phi\rho\omega^2 h^3\frac{1}{r}\frac{\partial}{\partial r}\left(\frac{r^2}{\eta_s}\right)-6\omega\left(\frac{\partial \bar{h}_T}{\partial \theta}-\sigma\frac{\partial \phi_s}{\partial \theta}\right)\\&+12\left(\frac{\partial \bar{h}_T}{\partial t}-d\frac{\partial \epsilon}{\partial t}\right)\end{aligned} \tag{17}$$

where u_z is the Darcian flow velocity on the small sector-shaped column having the same height as that of paper facing,



Wet Clutch, an Introduction, Fig. 1 A schematic diagram of wet clutch model (Note that the clutch is submerged in the lubricant)

$$u_z=\beta\left[\frac{1}{r}\frac{\partial}{\partial r}\left\{\frac{rd(1-\varepsilon)}{\eta_m}\right\}\frac{\partial p}{\partial r}+\frac{1}{r^2}\frac{\partial}{\partial \theta}\left\{\frac{d(1-\varepsilon)}{\eta_m}\frac{\partial p}{\partial \theta}\right\}\right]+d\frac{\partial \varepsilon}{\partial t}\ \text{and}\ \frac{1}{\eta_m}=\frac{1}{d}\int_0^d\frac{dz}{\eta} \tag{18}$$

Deur et al.'s wet clutch model will simplify (17) into (25), based on the following assumptions:

1. Film thickness h is constant over a normal clutch, giving

$$\frac{\partial h}{\partial \theta}=\frac{\partial h}{\partial r}=0\ \text{and}\ \frac{\partial \phi_s}{\partial \theta}=0 \tag{19}$$

2. Average film thickness in tangential and radial directions (Patir and Cheng 1979)

$$\bar{h}_T=hg(h)+\frac{\sigma}{\sqrt{2\pi}}e^{-\frac{h^2}{2\sigma^2}}\ \text{and}\ g(h)=\frac{1}{2}\left[1+erf\left(\frac{h}{\sqrt{2}\sigma}\right)\right] \tag{20}$$

is constant, and thus

$$\frac{\partial \bar{h}_T}{\partial t} = g(h)\frac{\partial h}{\partial t}, \frac{\partial \bar{h}_T}{\partial \theta} = g(h)\frac{\partial h}{\partial \theta} = 0, \frac{\partial \bar{h}_T}{\partial r} = g(h)\frac{\partial h}{\partial r} = 0 \tag{21}$$

3. Lubricant viscosity is uniform over normal clutch area, giving

$$\eta = \eta_m = \eta_s \tag{22}$$

4. Compressibility of paper facing can be neglected, giving

$$\varepsilon \approx 0 \text{ and } \frac{\partial \varepsilon}{\partial t} \approx 0 \tag{23}$$

5. The centrifugal term $\rho\omega^2$ can be neglected because of a minuscule amount of fluid trapped between clutch plates. Instead add the Beavars and Joseph velocity slip term for radial flow at the permeable boundary. The Beavars and Joseph factor with slip coefficient Σ is defined as

$$\eta_{BJ}(h) = \frac{1}{1 + \Sigma h/\sqrt{\beta}} \tag{24}$$

The modified Reynolds equation simplifies to

$$\frac{\partial h}{\partial t} = \frac{\phi h^3}{12\eta g(h)}\left[\frac{1}{r}\frac{\partial}{\partial r}\left\{r\left(1 + 3\eta_{BJ}(h) + \frac{12\beta d}{h^3}\right)\frac{\partial p}{\partial r}\right\} + \frac{1}{r^2}\frac{\partial}{\partial \theta}\left\{\left(1 + 3\eta_{BJ}(h) + \frac{12\beta d}{h^3}\right)\frac{\partial p}{\partial \theta}\right\}\right] \tag{25}$$

The external force F_{app} balances with the sum of hydrodynamic pressure force and the asperity contact force, giving

$$F_{app} = \iint_{A_n} p dA + \iint_{A_n} p_c dA \tag{26}$$

The mean contact pressure on the paper facing is defined as

$$p_c = E\varepsilon = E_r \frac{A_r}{A_n} \tag{27}$$

where A_r is the real contact area, $A_n = \pi(b^2 - a^2)$ is the nominal contact area, ε is the strain in the thickness direction, and E_r is the Young's modulus of the paper facing. According to Kimura and Ohtani (1991), typical values are $E = 31$ MPa and $E_r = 27$ MPa.

The contact pressure can be calculated based on the Greenwood and Williamson model (Natsumeda and Miyoshi 1994):

$$p_c(h) = \sqrt{\frac{\pi}{2}} ENr_a\sigma\left[e^{-h^2/2\sigma^2} - \frac{h}{\sigma}\sqrt{\frac{\pi}{2}} erfc\left(\frac{h}{\sqrt{2}\sigma}\right)\right] \tag{28}$$

where E is the Young's modulus, N is the asperity density, and r_a is the asperity tip radius.

For an ungrooved clutch, the pressure distribution on tangential direction is assumed to be uniform. After applying $\partial p/\partial\theta = 0$, the Reynolds equation (25) further simplifies to

$$\frac{\partial h}{\partial t} = \frac{\phi h^3}{12\eta g(h)}\left(1 + 3\eta_{BJ}(h) + \frac{12\beta d}{h^3}\right)\left\{\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial p}{\partial r}\right)\right\} \tag{29}$$

Accompanying (29) are boundary conditions on pressure at inner and outer radii

$$p(r = a) = 0 \text{ and } p(r = b) = 0. \tag{30}$$

Solving the above boundary value problem for pressure $p(r)$ yields

$$p(r) = \left[\frac{1}{g(h)}\frac{\phi(h)h^3}{12\eta}\left\{1 + 3\eta_{BJ}(h) + \frac{12\beta d}{h^3}\right\}\right]^{-1} \frac{dh}{dt}\frac{1}{4}\left(r^2 + \left(b^2 - a^2\right)\frac{\ln(r/b)}{\ln(r/a)} - b^2\right) \tag{31}$$

The pressure in the radial direction has a slightly asymmetric parabolic distribution. Applying the pressure distribution to (29), the dynamics of the squeeze film becomes

$$\begin{aligned}\frac{dh}{dt} &= \left[\frac{\phi(h)}{g(h)}\left(1 + 3\eta_{BJ}(h) + \frac{12\beta d}{h^3}\right)\right]\left(\frac{h^3}{12\eta Q}\right)\iint_{A_n} p dA \\ &= \left[\frac{\phi(h)}{g(h)}\left(1 + 3\eta_{BJ}(h) + \frac{12\beta d}{h^3}\right)\right]\left(\frac{h^3}{12\eta Q}\right) \times \left(F_{app} - A_n p_c(h)\right)\end{aligned} \tag{32}$$

where

$$Q = 2\pi\left(\frac{b^4 - a^4}{16} - \frac{a^2(b^2 - a^2)}{8} - \frac{(b^2 - a^2)^2}{16\ln(a/b)} - \frac{b^4 - b^2a^2}{8}\right) \tag{33}$$

Film thickness can be estimated by solving the initial value problem with initial condition:

$$h(t = 0) = h_0 \tag{34}$$

The clutch torque transfer per pair is the sum of the asperity contact torque $\boldsymbol{\tau}_{contact}$ and the viscous torque $\boldsymbol{\tau}_{viscous}$, giving

$$\boldsymbol{\tau}_{clutch} = \boldsymbol{\tau}_{contact} + \boldsymbol{\tau}_{viscous} = \text{sgn}(\omega)\int_0^{2\pi}\int_a^b r\mu p_c r dr d\theta + \int_0^{2\pi}\int_a^b r\eta\left(\phi_f - \phi_{fs}\right)\frac{\omega r}{h} r dr d\theta \tag{35}$$

After the integration of (31), the total torque from N_f identical clutch pairs is

$$\tau_{clutch} = -N_f \left[\mathrm{sgn}(\omega_d)(2\pi)\mu p_c \frac{b^3 - a^3}{3} - (2\pi)\frac{b^4 - a^4}{4}\eta\omega_d \frac{\phi_f - \phi_{fs}}{h} \right] \quad (36)$$

The external force in (26) can be calculated with the actuator dynamics typically approximated by a first-order equation

$$t_p \dot{P}_{app} + P_{app} = P_u \quad (37)$$

The time constant t_p determines the bandwidth of the actuator.

Simulation Using the Wet Clutch Model

Key equations for numerical simulation of the wet clutch model were provided in previous sections, including (32) for film thickness and (37) for applied pressure. Not introduced was the rotor dynamics, which depends on the application. To illustrate the nonlinear characteristics of wet clutch engagement, the wet clutch model in Fig. 1 was simulated. Surface properties and initial clearance h_0 between clutch plates were assumed known before simulation. The detailed parameter values can be found at (Choi et al. 2010). Film thickness h, applied pressure P_{app} and clutch slip speed ω_d are to be determined. The simulation procedures with relevant equations are:

1. Given the initial film thickness h_0 and surface properties, calculate the correction factors:

$$\phi(h) = 1 - Ce^{-rH} \text{ for } \gamma = 1 \text{ and}$$
$$g(h) = \frac{1}{2}\left[1 + erf\left(\frac{h}{\sqrt{2}\sigma}\right)\right] \text{where } H = h/\sigma$$

For $H \le 3$

$$\phi_f = \frac{35}{32}z\left[(1-z^2)^3 \ln\left(\frac{z+1}{\varepsilon/3\sigma}\right) + \left\{\frac{-55 + z(132 + z(345 + z(-160 + z(-405 + z(60 + 147)))))}{60}\right\}\right]$$

For $H > 3$

$$\phi_f = \frac{35}{32}z\left[(1-z^2)^3 \ln\left(\frac{z+1}{z-1}\right) + \frac{z}{15}\left\{66 + z^2(30z^2 - 80)\right\}\right]$$

where z = H/3

$$\phi_{fs} = (V_{r1} - V_{r2})\,\Phi_{fs}(H, \gamma = 1) \text{ where}$$
$$\Phi_{fs} = A_3 H^{\alpha_4} \exp\left(-\alpha_5 H + \alpha_6 H^2\right)$$

$$\eta_{BJ}(h) = \frac{1}{1 + \Sigma h/\sqrt{\beta}}$$

2. Calculate the clutch torque

$$p_c(h) = \sqrt{\frac{\pi}{2}} ENr_a\sigma\left[e^{-h^2/2\sigma} - \frac{h}{\sigma}\sqrt{\frac{\pi}{2}} erfc\left(\frac{h}{\sqrt{2}\sigma}\right)\right]$$

$$\tau_{clutch} = -N_f \left[\mathrm{sgn}(\omega_d)(2\pi)\mu p_c \frac{b^3 - a^3}{3} - (2\pi)\frac{b^4 - a^4}{4}\eta\omega_d \frac{\phi_f - \phi_{fs}}{h}\right]$$

3. Given initial h and P_{app} and pressure input command P_u integrate the following equations:

$$t_p \dot{P}_{app} + P_{app} = P_u$$

$$\frac{dh}{dt} = \left[\frac{\phi(h)}{g(h)}\left(1 + 3\eta_{BJ}(h) + \frac{12\beta d}{h^3}\right)\right] \times \left(\frac{h^3}{12\eta Q}\right)\left(A_p P_{app} - A_n p_c(h)\right)$$

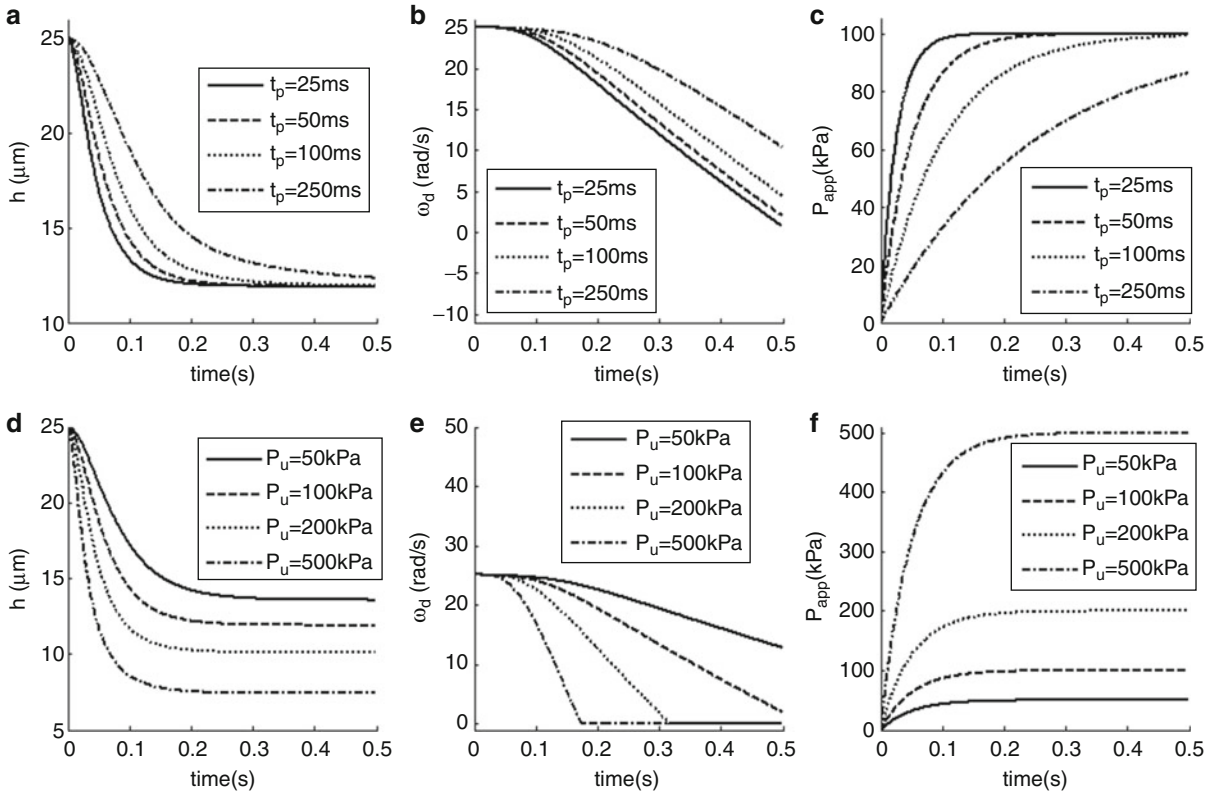
$$\text{where } Q = 2\pi\left(\frac{b^4 - a^4}{16} - \frac{a^2(b^2 - a^2)}{8} - \frac{(b^2 - a^2)^2}{16\ln(a/b)} - \frac{b^4 - b^2a^2}{8}\right)$$

4. Integrate the rotor dynamics for the slip speed ω_d (note f(τ) depends on the application)

$$J\dot{\omega}_d = f(\tau_c)$$

5. Repeat steps 1–4 until the final time is reached.

Figure 2 presents simulations of clutch engagement actuated by step inputs in pressure. Figures 2a–c show transient responses of film thickness, slip speed, and applied pressure to a step of 100 kPa (i.e., P_u = 100 kPa) applied at $t = 0$; different curves pertain to different values of time constant t_p. The pressure rises in Fig. 2c are typical of a first-order linear time invariant system, with different time constants. As the applied pressure pushes the two plates together, initially the film thickness rapidly decreases, but then slows due to squeeze film effect between plates. Meanwhile, the difference in rotational speeds between the two plates steadily diminishes. The speeds of the various transient responses are not proportional to the corresponding time constants of the linear pressure



Wet Clutch, an Introduction, Fig. 2 Wet clutch engagement simulation

dynamics, because the dynamics governing film thickness are highly nonlinear. Figures 2d–f show the effects of magnitude of input pressure P_u on film thickness and slip speed. Here t_p = 50 μs only. Figure 2d shows that a larger input pressure magnitude renders a smaller terminal film thickness. Note that the rate of slip accommodation increases as the pressure magnitude increases. However, the relations are not proportional to the magnitude of input pressure, due to the nonlinear behavior of film thickness.

Key Applications

As shown in the previous section, wet clutch engagement is a nonlinear process affected by film thickness, slip speed, lubricant property, and contact surface conditions. Because of the complexity, control of wet clutch engagement for most applications has been limited to a torque switch or a torque limiter. However, affordable digital computers have enabled recent torque vectoring mechanisms such as active differential (Ross et al. 2007), which requires the precise control of clutch torque and slip speed. Typical applications of wet clutch include torque limiter, motorcycle clutch, shifting mechanism in automatic transmission, primary clutches in dual clutch transmission, torque vectoring mechanism for 4WD vehicles, and active limited slip differential.

Cross-References

- ▶ Asperities
- ▶ Average Reynolds equations
- ▶ Friction
- ▶ Lubricant Viscosity
- ▶ Lubrication Regimes
- ▶ Lubrication with a Newtonian Fluid
- ▶ Reynolds Equation
- ▶ Stochastic Contact Theories: Theories of Surface Roughness and Applications to Contact Mechanics
- ▶ Surface Characterization and Description
- ▶ Transmission Lubricants

References

E.J. Berger et al., Finite element modeling of engagement of rough and grooved wet clouches. J Tribol. **118**, 137–146 (1996)

J. Choi, M. Nakhaeinejad, M. Bryant, Simulation of wet clutch engagement on electronically-controlled limited-slip differential with

consideration of degraded clutch performance. "in *Proceeding. of* International Joint Tribology Conference ID-41265, (San Francisco, CA, 2010)
J. Deur et al., *Modeling of Wet Clutch Engagement Including a Thorough Experimental Validation*, SAE Int. 2005-01-0877, 2005.
H.S. Hele-Shaw, *Hydraulic Clutch or Coupler*, U.S. Patent 1462600, 1923
Y. Kimura, C. Ohtani, *Observation of Solid Contact with a Contact-Microscope*, Seiken Leaflet, No. 199, The University of Tokyo, 1991 (in Japanese)
S. Natsumeda, T. Miyoshi, Numerical simulation of engagement of paper based wet clutch facing. J Tribol. **116**, 232–237 (1994)
N. Patir, H.S. Cheng, Application of average flow model to lubrication between rough sliding surfaces. J. Lubr. Tech. **101**, 220–230 (1979)
C. Ross et al., *Development of an Electronically-Controlled Limited-Slip Differential (eLSD) for FWD Applications*, SAE Int. 2007-01-0925, 2007

Wetting Ability

► Surface Free Energy

WFDCL (Water-Free Die Casting Lubricants)

► Carrier-Free Die Casting Lubricants

WFL (Water-Free Lubricants) for Carrier-Free Die Casting

► Carrier-Free Die Casting Lubricants

WFRA (Water-Free Release Agents) for Carrier-Free Die Casting

► Carrier-Free Die Casting Lubricants

Wheelburns

► Rolling Contact Fatigue (RCF)

Whipple Thrust Bearing

► Gas Bearings with Narrow Inclined Grooves

White Light Interferometry

Wilfried Bauer[1], Mark Weber[2], Sirichanok Chanbai[2]
[1]Polytec GmbH, Waldbronn, Germany
[2]NanoFocus AG, Oberhausen, Germany

Synonyms

Coherence RADAR; Coherence scanning interferometry; Vertical scanning interferometry

Definition

Vertically scanning white light interferometry is a non-contact optical method for 3D-profiles of rough and smooth surfaces (Dresel and Häusler 1992; Windecker and Haible 1995). Instead of coherent light as it is used in classical interferometers, a polychromatic broadband, incoherent light-source like an LED or a halogen lamp is used. So it is possible to measure the topography not only of smooth but also of rough surfaces without disturbing speckles and, in case of steps, the order of the interference fringes is not lost. The z-resolution can be in the picometer range, independent from the field of view. Changing optics like objectives in a microscope-interferometer does not influence the z-resolution. The field of view depends on the optical setup of the instrument: For telecentric systems, e.g., it can be as large as 50×50 mm^2, for typical microscope-systems its higher lateral resolution depends mainly on the magnification of the objective.

White light interferometry delivers a 3D-point cloud of the complete field of view. From these data, surface- and form-parameters or volumes can be determined by software algorithms. So basically the optical distance of every surface point is measured relative to a reference surface within the instruments. This method is quite old, and in 1895 Michelson published a path difference measurement using the principle of an interferometer (Michelson 1895).

Scientific Fundamentals

Interferometer Setup

The idea of an interferometer is to derive high precision optical path length differences or associated positional

information from the measurement of an optical intensity as a function of a relative displacement of an optical component.

Michelson Interferometer

The general principle of white light interferometers can be explained by the example of the well-known Michelson interferometer, shown in Fig. 1.

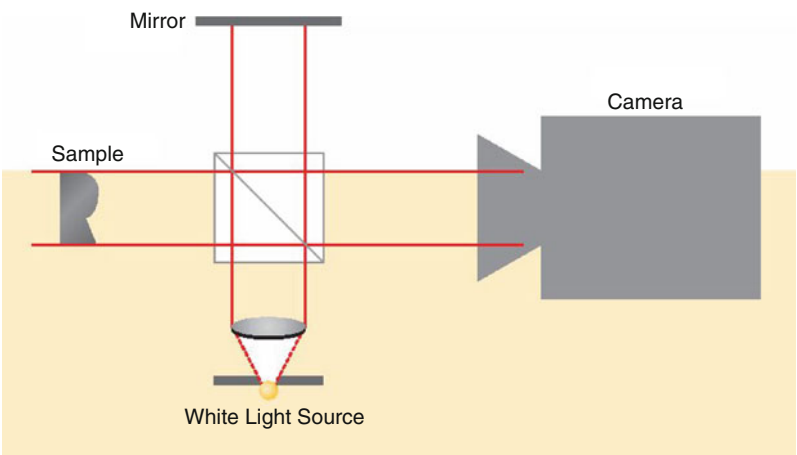
In a Michelson interferometer the light is split into beams, one into the measurement arm to the object and one into a reference arm with a flat mirror. The reflected light from both arms is superimposed. The interference occurs for white light only, when the optical path lengths of the measurement beam and the reference beam are nearly matched and if the optical path lengths of the two arms differs less than half the coherence length of the light source. An object surface having features of different heights leads to a phase pattern that is mixed with the light from the flat reference in the camera image sensor plane. Each pixel of the CCD samples a different spatial position within the image of the object surface. So the complete area within the field of view is detected. A typical white light correlogram (interference signal) is produced when the length of the reference or measurement arm is scanned by a positioning stage through a path length match. By scanning (changing) the measurement beam path length, e.g., by moving the reference surface, a correlogram is generated at each pixel of the detector camera when the path-length condition is met and moved through. The width of the resulting correlogram is the coherence length and depends strongly on the spectral width of the light source. The interference signal of a pixel has maximum modulation when the optical path length of light hitting the pixel is exactly the same for the reference and the object beams. Therefore, the z-value of a surface point imaged by this pixel corresponds to the z-value of the positioning stage when the modulation of the correlogram is greatest. The results of all pixels together lead to the topography of the complete area. The lateral positions of the height values depend on the corresponding object point that is imaged by the pixel matrix.

These lateral coordinates, together with the corresponding vertical coordinates, describe the surface topography of the object.

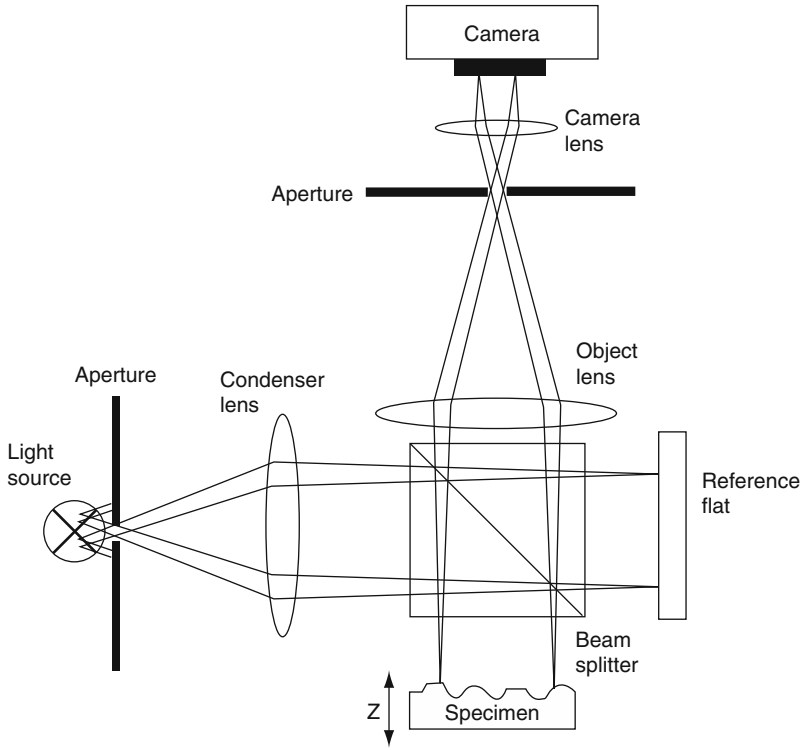
The Twyman–Green interferometer (Fig. 2) is a special configuration of the well-known Michelson interferometer, considering the light source as point source. A coaxial telecentric setup is best for the 3D-profiles of larger areas and also allows measurement of or within deep cuts or drillings without shadowing. Theoretically the z-scan-range is only limited by the scan-stage, which is typically either a high-accuracy stepping motor for larger scan-ranges or a piezo-stage for smaller displacements. Due to the small aperture of a telecentric setup, however, the optical lateral resolution is poor against microscopic setups (Fig. 3) according to Eq. 1 (Abbe or Rayleigh criterion)

$$d = \frac{\lambda}{2n \sin \alpha} \tag{1}$$

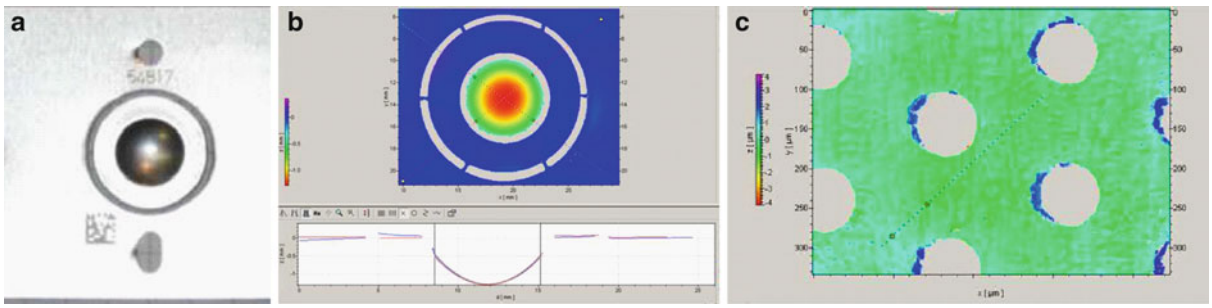
The difference in lateral resolution is clearly shown in the example (Fig. 3) of a nebulizer for medical applications: With the large field of view the shape and form can be measured with a high z-resolution, but the small holes of the nebulizer cannot be obtained; with a microscopic system, however, you can see the holes within the same sample with its material throw-off.



White Light Interferometry, Fig. 1 Setup of a Michelson interferometer

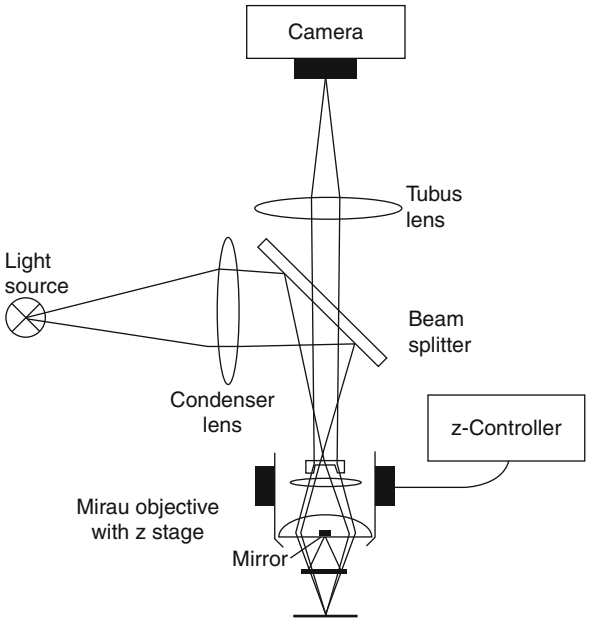


White Light Interferometry, Fig. 2 Optical setup of a Twyman–Green interferometer with a CCD image sensor



White Light Interferometry, Fig. 3 Profiles of a nebulizer (picture from the sample) (**a**), 3D-profile with a white light interferometer with a large field of view (**b**), and with a microscopic setup (50× objective). (**c**) Interferometer with Microscopic Setups

In order to visualize microscopic structures, it is necessary to combine an interferometer with the optical setup of a microscope. Such an arrangement is shown in Fig. 3. This setup is similar to a standard optical microscope. The only difference is an interferometric objective lens. The optical magnification of the image on the CCD does not depend on the distance between tube lens and objective lens if the microscope images the object at infinity. The interference objective is the most important part of such a microscope. Different types of objectives are available. With a Mirau objective, as shown in Fig. 4, the reference beam is reflected back in the direction of the objective front lens by a beam splitter. On the front lens there is a miniaturized mirror of the same size as the illuminated surface on the object. Therefore, for high magnifications, the mirror is so small that its shadowing effect can be neglected. For lower magnifications a Michelson interferometer is integrated into the objective. Moving the



White Light Interferometry, Fig. 4 Schematic layout of an interference microscope with a Mirau objective

interference objective modifies the length of the measurement arm. The evaluation of the correlogram finally leads to the topography. In general, white light interferometers are optical instruments that superimpose reference light with light reflected from the sample surface to obtain the sample's surface topography from the resulting correlogram.

Measurement

To determine surface heights the camera frame output is processed pixel by pixel. No phase-related pixel to pixel information is required. For the measurement the surface is scanned over the interesting z-range and for every step the intensity of the superimposed light from both paths is recorded from every camera pixel with a certain frame rate (given by the camera type) and exposure time (given by the user).

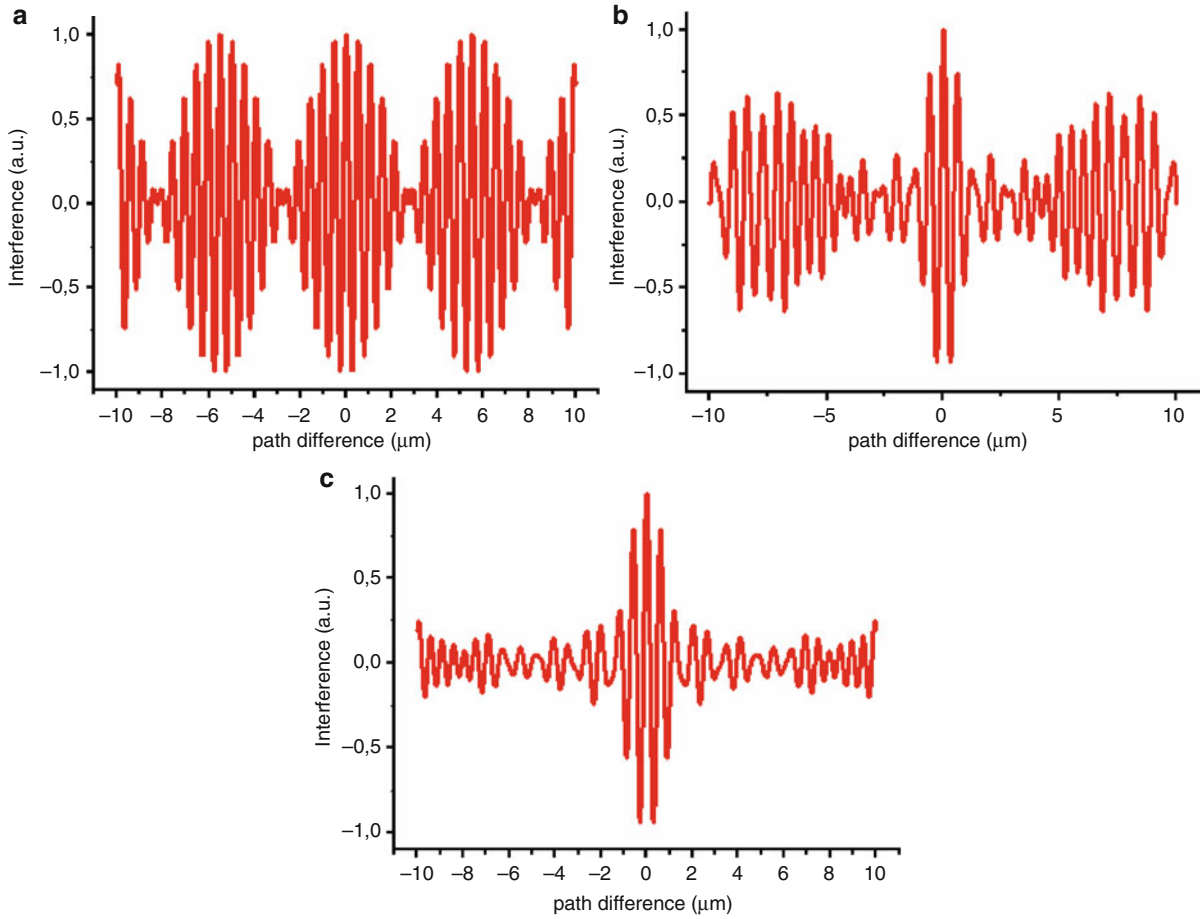
The user has in addition the choice of the step-width: A larger step-width of the stage means shorter measurement time and a small step-width more accuracy, but also more disturbing environmental influences caused by the longer measurement time. The best results are obtained with a short coherence length, an optimized sampling step size, good vibration isolation, and high contrast. The best contrast values are achieved when the intensity of the light from both beam paths is the same. Some white light interferometer devices allow the adjustment of the intensity of at least one path. In general, smooth objects show the most cooperative surfaces. Industrial suppliers of white light interferometers specify the resolution of their instrument by measuring the 3D-profile twice (or even more) under the best environmental conditions and subtracting the first from the second. The result is more or less the noise-value, which is typically in the subnanometer-range. The traceability is ensured by height standards from the National Bureaus of Standards. The analysis described in the VDI guideline 2655-1-1 (VDI guideline 2655 part 1.1 March 2008) gave an uncertainty for step measurements of 5.5 nm. The advantage of white light interferometers against classical interferometers, however, is their ability to characterize both rough and structured surfaces. Such surfaces very often show variant reflectivity from different areas of the surface, which require different exposure times of the camera. Here very often a technique similar to HDR (high dynamic range) techniques of digital image processing systems is used. The smart surface scanning technique, which scans the surfaces two or more times with different camera exposure times and then chooses the result with the best contrast.

Relation Between Spectral Width and Coherence Length and Widths of the Correlogram

The quality and shape of the correlogram have a major influence on the system's resolution and accuracy. The most important parameters of the light source are its wavelength and coherence length. The coherence length defines the width of the correlogram, which again depends on the spectral width of the light source.

To understand the difference between the classical and a white light interferometer the following consideration might be helpful: The superposition of two waves with slightly different wavelengths results in a beat signal, and superposition of five or ten waves with slightly different wavelengths leads to a superposition with a pronounced maximum (see Fig. 5). A broadband light source emits light of many different wavelengths, thus generating a correlogram without side-lobes. In other words, in a white-light interferometer only in the small range of the coherence length all wavelengths have almost the same phase conditions as at the light source and therefore all wavelengths interfere, whereas outside the range the phases and the interferences are randomly distributed and no interference can be observed.

For a more detailed calculation of the relationship between correlogram width, coherence length and spectral width a Gaussian spectrum of the emitted light is assumed, which is a good approximation for a light



White Light Interferometry, Fig. 5 Interference of waves with 2 (**a**), 5 (**b**), and 10 (**c**) different wavelengths

emitting diode (LED). In Fig. 6a the spectral density function (Eq. 2) for a Gaussian spectrum is shown. The normalized spectral-density function is defined according to Eq. 2, where $\Delta\nu$ is the effective 1/e-bandwidth and ν_0 is the mean frequency.

$$S(\nu) = \frac{1}{\sqrt{\pi}\Delta\nu}\exp\left[-\left(\frac{\nu-\nu_0}{\Delta\nu}\right)^2\right] \tag{2}$$

In Fig. 6b the corresponding intensity modulation is shown to be substantial only in the neighborhood of position z_0, where the reference and object beams have the same length and superpose coherently.

The z-range of the positioning stage in which the envelope of intensity modulation is higher than 1/e of the maximum value determines the correlogram width. This corresponds to the coherence length because the difference of the optical path length is twice the length difference of the reference and measurement arms of the interferometer.

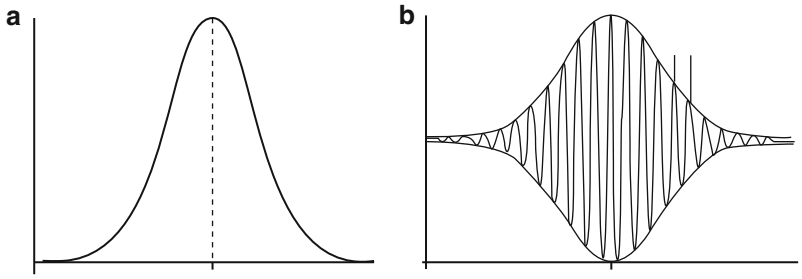
The relationship between correlogram width, coherence length, and spectral width is calculated for the case of a Gaussian spectrum, which is a good approximation for spectral distribution of an LED.

According to the generalized Wiener–Khintchine theorem, the autocorrelation function of the light field is given by the Fourier transformation of the spectral density (Eq. 3), which is measured by interfering the light field of reference and object beams.

$$\begin{aligned} k(\tau) &= \int_{-\infty}^{\infty} S(\nu)\exp(-i2\pi\nu\tau)d\nu \\ &= \exp(-\pi^2\tau^2\Delta\nu^2)\exp(-i2\pi\nu_0\tau) \end{aligned} \tag{3}$$

In the case that the intensities in both interferometer arms are the same, the intensity observed on the screen results in the relation given in Eq. 4

$$I(z) = I_0 \mathrm{Re}\{1 + k(\tau)\} \tag{4}$$



White Light Interferometry, Fig. 6 (**a**) Spectral density function of the light source. (**b**) Light intensity as a function of z- path difference

Here $I_0 = I_{obj} + I_{ref}$ where I_{obj} and I_{ref} are the light intensities from the measurement arm and the reference arm, respectively. The mean frequency $\nu_0 = c/\lambda_0$ can be expressed by the central wavelength, and the effective bandwidth by means of the coherence length L_c, $L_c = c/\pi\Delta\nu$. From Eqs. 3 and 4 the intensity on the screen can be derived (Eq. 5), taking into account that $\tau = 2\cdot(z-z_0)/c$, with c being the speed of light.

$$I(z) = I_0\left(1 + \exp\left[-4\left(\frac{z - z_0}{I_c}\right)^2\right]\cos\left(4\pi\frac{z - z_0}{\lambda_0}\varphi_0\right)\right) \tag{5}$$

Accordingly, Eq. 5 describes the correlogram as shown in Fig. 3b. One can see that the distribution of the intensity is formed by a Gaussian envelope and a periodic modulation with the period $\lambda/2$.

In practice, however, phase shifts at the object surface, inaccuracies of the positioning stage, dispersion differences between the arms of the interferometer, reflections from surfaces other than the object surface, and noise in the CCD can lead to a distorted correlogram. While a real correlogram may differ from the result in Eq. 5, the result clarifies the strong dependence of the correlogram on two parameters: the central wavelength and the coherence length of the light source.

Computation of the Envelope

The envelope function (Eq. 6) is described by the exponential term of Eq. 5.

$$E(z) = \exp\left[-4\left(\frac{z - z_0}{I_c}\right)^2\right] \tag{6}$$

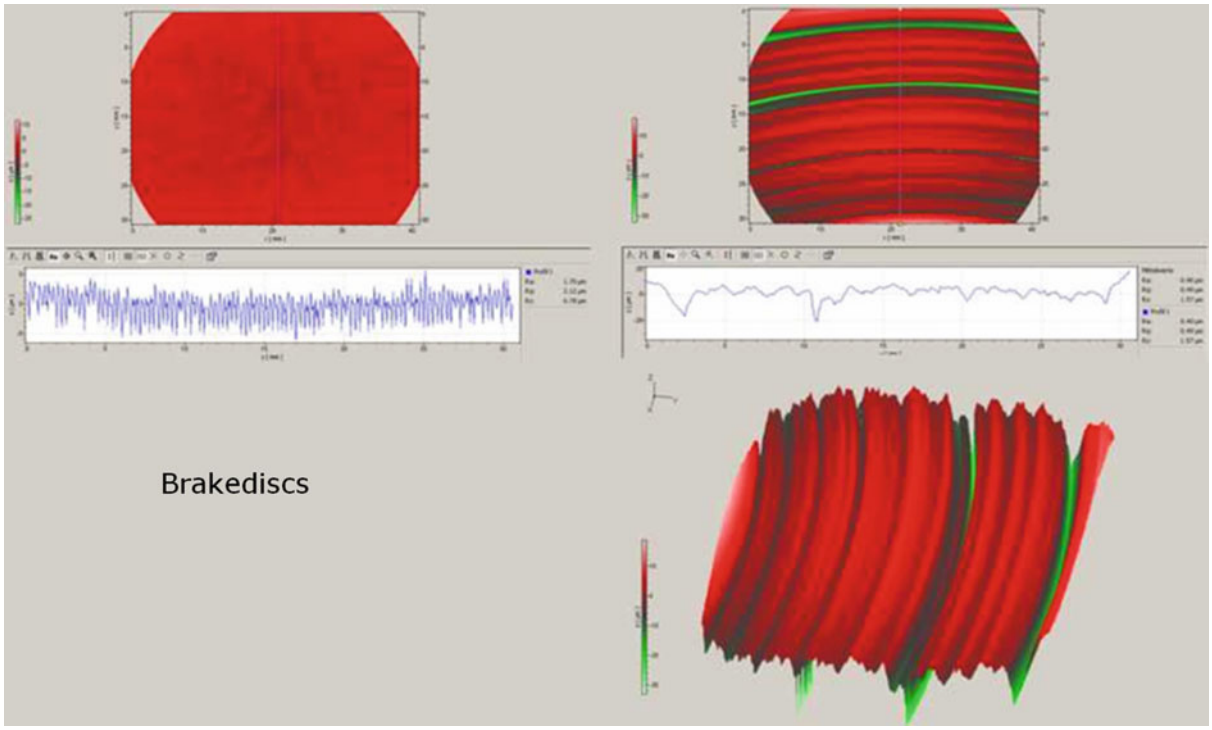
The software calculates the envelope from the correlogram data. One commonly used principle of the envelope calculation is to remove the cosine term of Eq. 5. With the help of a Hilbert transform the cosine term is changed into a sine term. The envelope is obtained by summing the squares of the cosine and sine-modulated correlograms (Eq. 7).

$$E(z) = \sqrt{\left(\exp\left[-4\left(\frac{z - z_0}{I_c}\right)^2\right]\cos\left(4\pi\frac{z - z_0}{\lambda_0}\right)\right)^2 + \left(\exp\left[-4\left(\frac{z - z_0}{I_c}\right)^2\right]\sin\left(4\pi\frac{z - z_0}{\lambda_0}\right)\right)^2} \tag{7}$$

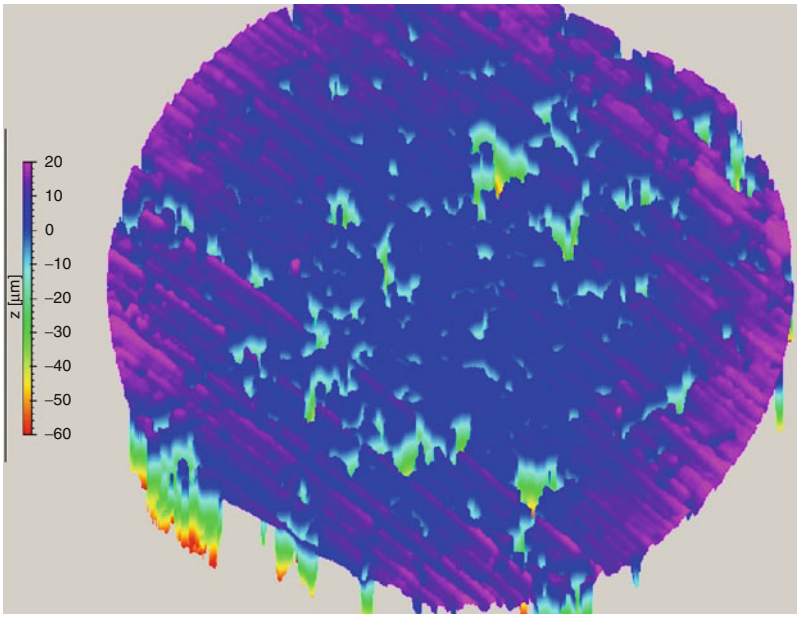
Knowing the center frequency of the light, another approach can be used by fitting a cosine function with the given central wavelengths of the white light to the measured intensity of a pixel. By this, the maximum can also be determined. In addition, for smooth surfaces it is more accurate to look at the zero-crossing of the cosine curve next to the maximum of the envelope. By a phase unwrapping algorithm known from the classical phase shift interferometery you can trace the order for connected surfaces. This tracing is necessary to avoid influences from the optical parts like, for example, dispersion in the beam splitter. For non-connected surfaces or surfaces with steps, a combination of both is done by the software. These are only examples for evaluation. Industrial suppliers of white light interferometers very often use patented algorithms, which might differ. In general, the uncertainty of the calculation of the envelope maximum depends on the following:

- The coherence length.
- The sampling step size of the correlogram.
- Deviations of the z-values from desired values (e.g., due to environmental conditions like vibrations).
- The interference contrast and the roughness of the surface.

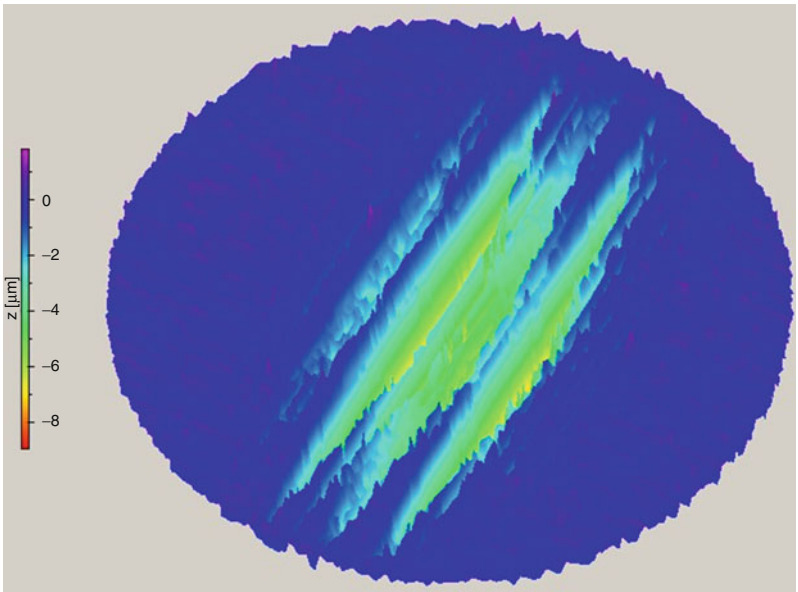
The former part describes a common evaluation algorithm as an example. Other algorithms, such as the more noise-resistant algorithm for the estimation



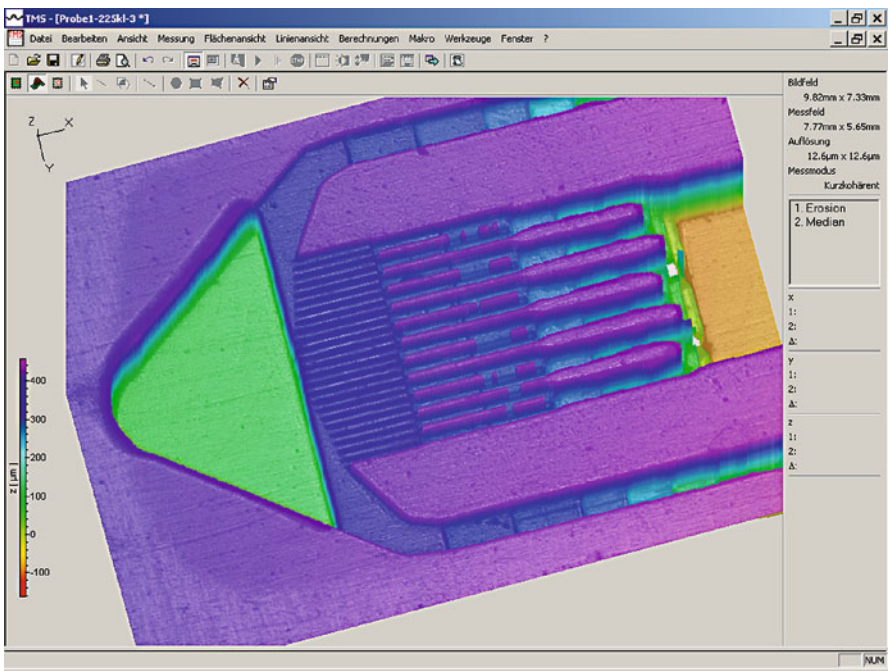
White Light Interferometry, Fig. 7 Profile of brake discs before (*left*) and after treatment (*right*)



White Light Interferometry, Fig. 8 Abrasion of a friction pad



White Light Interferometry, Fig. 9 Example surface for volume determination



White Light Interferometry, Fig. 10 3D-profile of a Bio-MEMS

of the overall signal position using the centroid positions, are described in the ISO 25178–604. An important and more precise evaluation algorithm only for smooth surfaces combines the envelope evaluation with the phase estimation. In short, the envelope determines the fringe order and the evaluation and the height determination is done by interference phase estimation as in classical interferometers. This results in a better resolution, however, the roughness must be smaller, for example, $\lambda/10$.

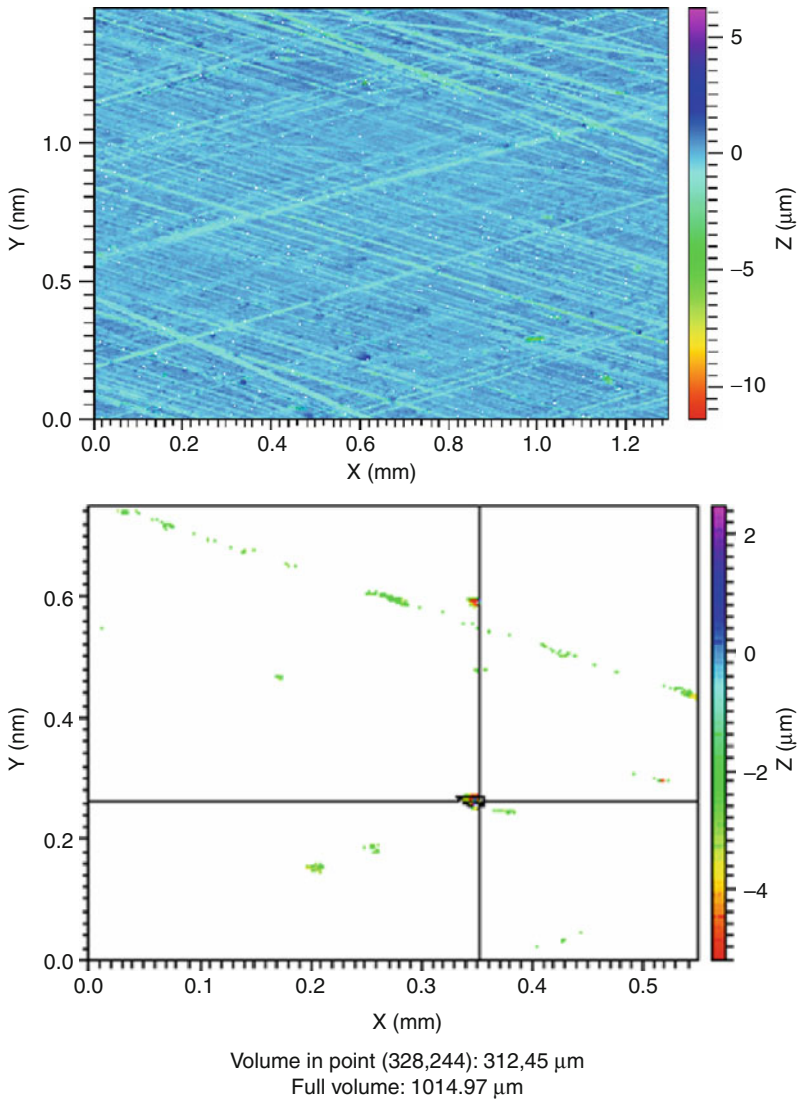
For a good signal to noise ratio of the measured intensity values, it is important to have a high interference contrast. To achieve this, the filters in the reference and object path must be adapted to the reflectivity of the surface and the camera exposure time is chosen to make best use of the dynamic range of the camera.

It should be noted that the determination of the measured z-point is more accurate than the chosen step widths of the stage. Typically a step width of around 100 nm, which gives a reasonable number of measurement points within the correlogram, leading to a resolution of about 1 nm.

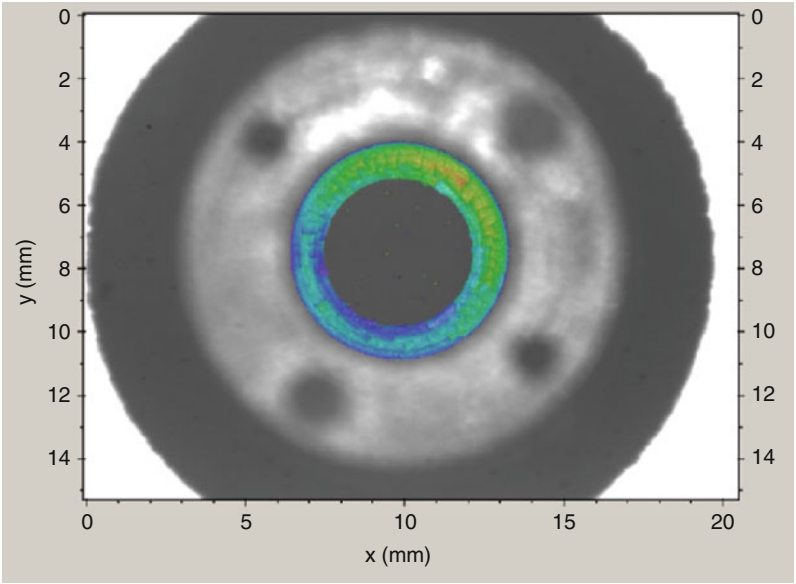
Measurement Examples and Applications

In tribology one of the most important tasks is the "before-and-after" determination. The advantage of white light interferometry is the characterization of fine structures before and after treatments. Examples of measurement data for a brake disc are shown in Fig. 7 and the abrasion of a friction pad in Fig. 8.

Very often, the exact determination of volumes is required. An example surface from a ball on disc test is shown in Fig. 9. A repeatability and reproducibility measurement shows a volume of the groove-like depression of



White Light Interferometry, Fig. 11 Honed surface and pore distribution



White Light Interferometry, Fig. 12 Flatness measurement on a surface at a depth of >40 mm below the top

2.971×10^{-3} mm^3. Other parameters like material distribution, height analysis, etc., can also be determined in an easy way and are more precise than a volume determination with a microscope, and a 3D-profile includes more detailed information on the structure.

Quantitative defect analysis, illustrated in Fig. 10 with the example of a Bio-MEMS surface, is common.

Figure 11 shows a honed surface. The 3D-profile shows the surface structures and a pore analysis of this surface.

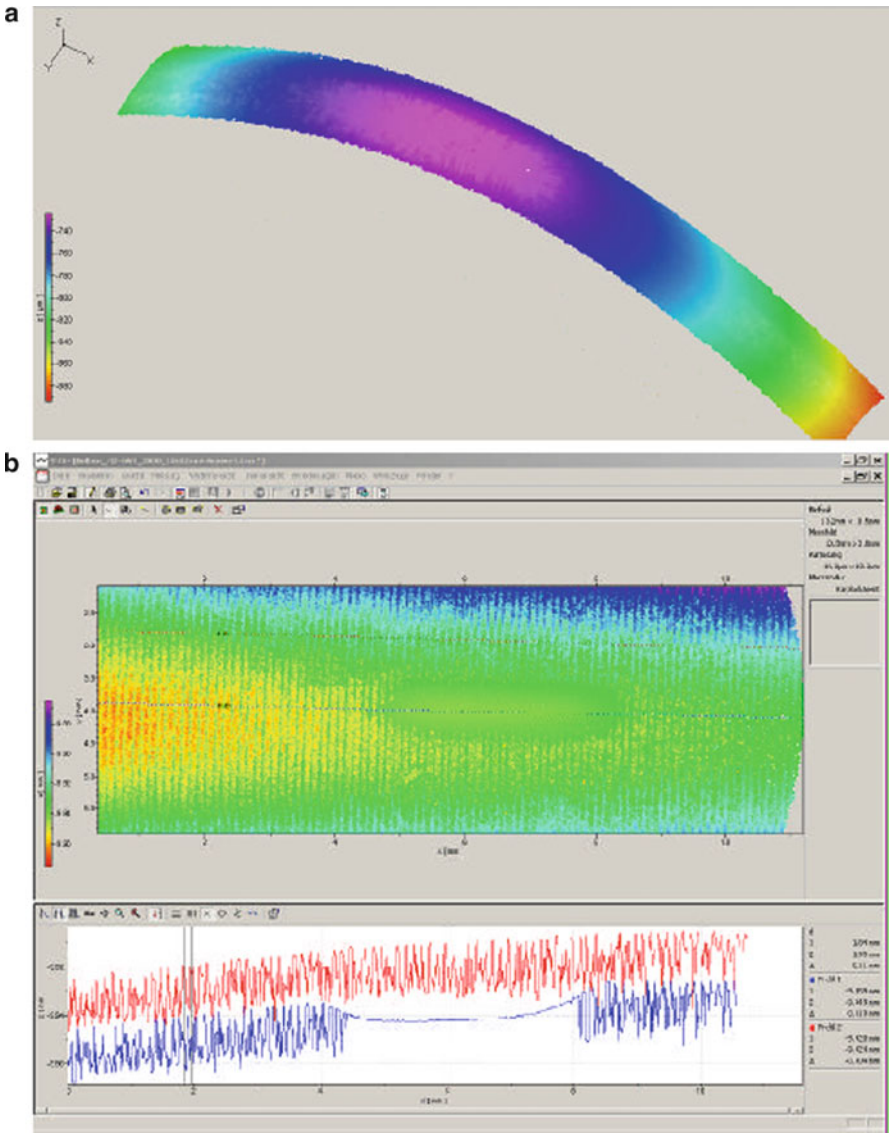
Of course the determination of the Abbott curve and the material distribution is also possible. Because the vertical precision for measurements with white light interferometers is independent from the field of view, the flatness, parallelism, or step height measurements can be done accurately for larger areas. In addition, white light interferometers with a telecentric beam-path can also characterize deep-lying areas, as is shown in Fig. 12. Residual grooves from the machining can also be seen.

Software algorithms allow the exact determination of the zero-level or the subtraction of the form. An example with a curved surface is shown in Fig. 13, where the curvature is subtracted. The reduction of the roughness on the treated part of the surface is clearly obtained.

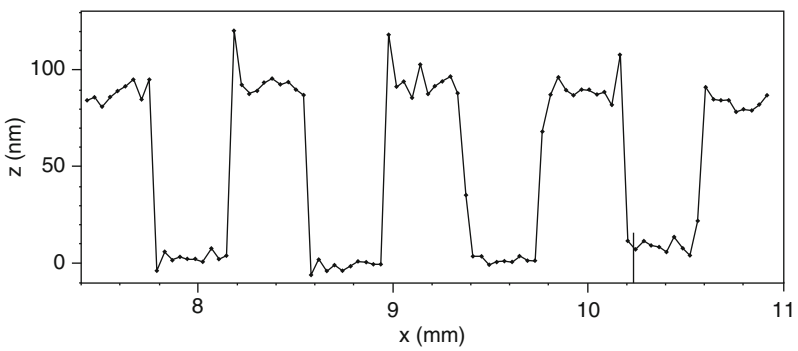
In general, the determination of roughness parameters is quite important. Dynamic sealing surfaces, e.g., require a certain roughness; smooth surfaces lead to unwanted adhesion effects, rough surfaces to higher friction. Unfortunately results obtained by optical methods don't agree completely with the results of standardized tactile measurements due to the different lateral resolution. This, however, is only half of the truth, because artifacts like spiking or the batwing effect, so called because of the shape of the error, also influence the result. The batwing effect is shown in Fig. 14, measured on a grating with a grating constant of 200 µm and a step height of 80 nm, is usually explained as additional interference between different heights of the surface within the lateral resolution of the white light interferometer. Therefore, this typically occurs on edges (for more detailed see Leach and Brown 2008). White light interferometry does not give the correct surface height at the positions close to the step even if the step height is significantly greater than the coherence length.

This height enhancement effect also occurs on rough surfaces with their random height distributions, but it is not so clearly visible as on steps within smooth surfaces. A quantification of this effect can be seen in Fig. 15, a measurement on a grating with a grating constant of 80 µm and a step height of 80 µm, measured by a large area white light interferometer with a optical resolution of about 40 µm. Depending on which location of the grating is imaged to the pixel, an enhancement can be obtained or even no steps can be resolved.

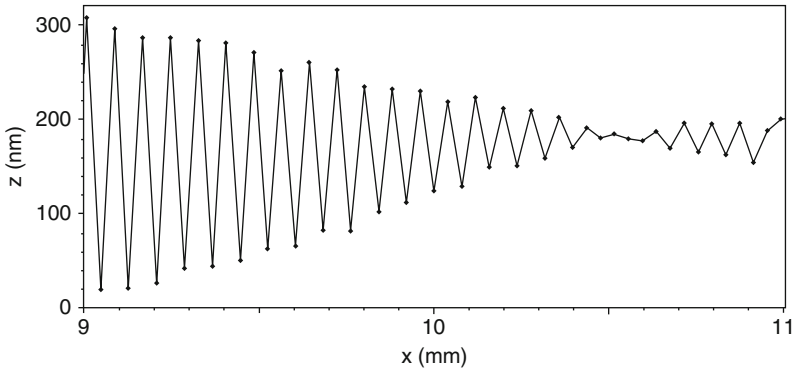
The core idea to achieve comparability is to determine a short cut-off wavelength down to which the spatial frequency components of the surface topography are measured in a correct way (de Groot and Colonna de Lega 2006). The determination of the transfer function is



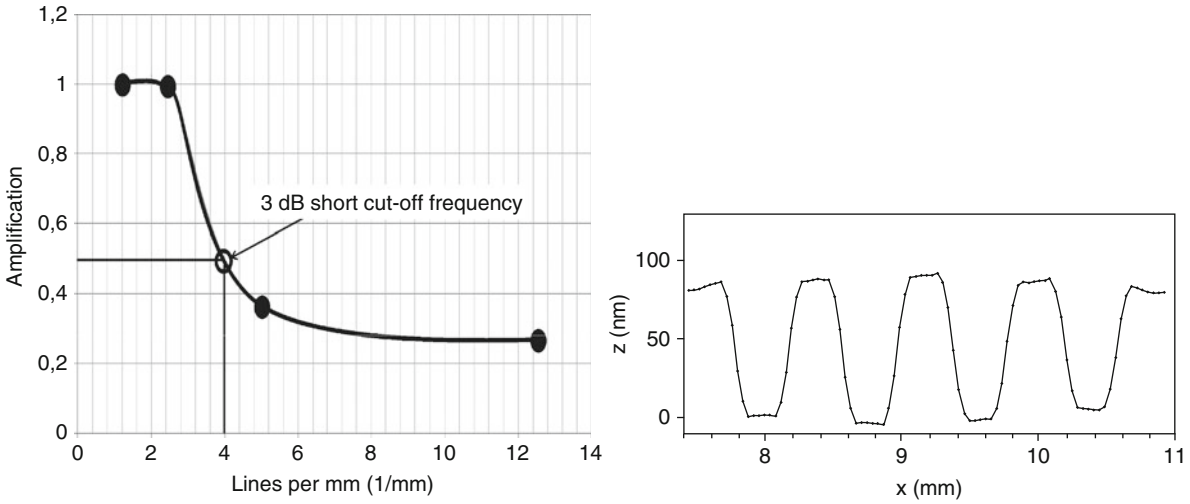
White Light Interferometry, Fig. 13 (**a**, **b**) Abrasion of a *curved* surface



White Light Interferometry, Fig. 14 Topography of a grating showing the so-called "batwings"



White Light Interferometry, Fig. 15 Measurement of a grating with a grating constant of 80 μm with a 40 μm optical resolution



White Light Interferometry, Fig. 16 Determination of the transfer function and calculation results

demonstrated in Fig. 16 with a cut-off wavelength to be about 200 μm for the present instrument with the lateral resolution of 40 μm. The result using this transfer function is shown in Fig. 16.

For microscopic systems the cut-off wavelengths is smaller and on the order of the radius of a tip of tactile measurements or shorter. For comparability, the device-specific transfer function with the longest cut-off wavelength has to be applied on both. This procedure will be described in VDI guideline 2655-1-3 or in ISO 25178–604 (2012) in more detail.

Summary

White light interferometry is a fast and versatile tool for the characterization of surfaces with high z-resolution. It is independent from material, composition, and nature of the surface. It delivers a complete 3D-point cloud of the surface for further evaluations.

All images courtesy Polytec GmbH, Waldbronn, Germany; www.polytec.com

Cross-References

- ▶ Disk Roughness and Defect Monitoring
- ▶ Optical Interferometry
- ▶ Surface Characterization and Description
- ▶ Surface Roughness
- ▶ Surface Statistics and Probability Density Function
- ▶ Topography of Engineering Surfaces
- ▶ Traceability and Calibration of Surface Texture Measurement Instrumentation
- ▶ Wear Quantification by Comparison of Surface Topography Data

References

P. de Groot, X. Colonna de Lega, Interpreting interferometric height measurements using the instrument transfer function, in *Proc. FRINGE 2005*, ed. by W. Osten (Springer, Berlin/Heidelberg, 2006), pp. 30–37

T. Dresel, G. Häusler, Three dimensional sensing of rough surfaces by coherence radar. Appl. Opt. **31**, 919–925 (1992)

R. Leach, L. Brown, *Measurement Good Practise Guide No 108: Guide to the Measurment of Smooth Surface Topography using Coherence Scanning Interferometry* (National Physical Laboratory, Teddington, 2008)

A. Michelson, Determination experimentale de la valeur du metre en longuers dóndes lumineuses. Trav. Mem. Bur. Int. Poids Mes. **11**, 1–42 (1895)

VDI Guideline 2655 Part 1.1: Optical Measurements and Microtopography; Calibration of Interference Microscopes and depth measurement standards for roughness measurrment (March 2008)

R. Windecker, P. Haible, Fast coherence scanning interferometry for smooth, rough and spherical surfaces. J. Mod. Opt. **42**, 2059–2069 (1995)

J. Petzing, J. Coupland, RK. Leach, Measurement Good Practice Guide No. 116, December 2010. "The measurement of rough surface topography using coherence scanning interferometry" (National Physical Laboratory, Teddington, 2010)

White-Etching Layer

► Self-Mating Metal Articulations in the Hip Joint

Wöhler Curve

► Stress-Life Theories

Work of Adhesion and Work of Cohesion

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

Definition

Work of adhesion and work of cohesion are concepts introduced by wetting phenomena of liquids on solid surfaces. In order that these two concepts can be demonstrated clearly, wetting phenomena are introduced briefly.

Wetting Phenomena

There are many wetting phenomena of liquids on solid surfaces in everyday life. Water on a lotus leaf takes the shape of a water bead, and the water bead may roll off due to the force of gravity. If a person immerses his hands in water, a layer of water will dampen them and they will believe that their hands are wet. These are two of the most familiar examples of wetting and nonwetting. When a piece of waxed paper is removed from water, the water film immediately formed at the surface of the waxed paper will rupture into water beads. On the other hand, if an identical piece of waxed paper is removed from oil, the oil film will spread and become very thin without oil beads forming. Wetting is one of the most conventional phenomena we encounter in life and in industrial production. In many applications, such as washing, flotation of minerals, dyeing, the production and use of paint, water-resistant coatings, and bonding-resistant coatings, the wetting of liquids at the surfaces of solids plays an important role.

At the macro-scale, wetting is the process of one fluid replacing another fluid on the surface of a solid. From the angle of micro-scale, the fluid, which dampens a solid, contacts with the molecules at the surface of the solid after replacing the other fluid originally existing at the surface of the solid. There are no molecules of the replaced fluid between the solid and the replacing fluid. The most common wetting phenomenon is one liquid replacing air from the surface of a solid, for example, a drop of water spreads, replacing air at the surface a pane of glass. In 1930, Osterhof and Bartell divided wetting phenomena into three types: adhesion, immersion, and spreading.

Adhesion

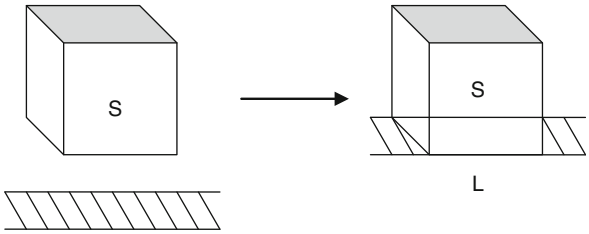
If one liquid phase (L) and one solid phase (S) combine as in Fig. 1, this process is called *adhesion*. The result is that the solid-vapor and liquid-vapor interfaces disappear, producing one new interface of solid-liquid.

Immersion

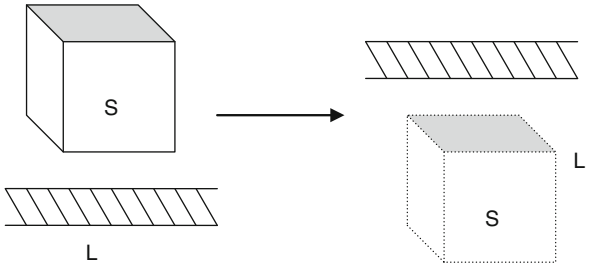
As shown in Fig. 2, a cubic solid (S) is immersed into a liquid (L). If all the gas at the surface of the solid is replaced by a liquid, this process is called *immersion*.

Spreading

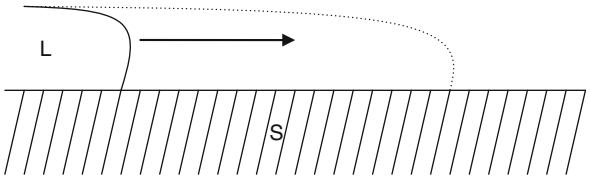
Place a drop of liquid at the surface of a solid (shown in Fig. 3). If the liquid drop automatically spreads into a liquid film under constant pressure and temperature, this process is called *spreading*.



Work of Adhesion and Work of Cohesion, Fig. 1 Adhesion process



Work of Adhesion and Work of Cohesion, Fig. 2 Immersion process



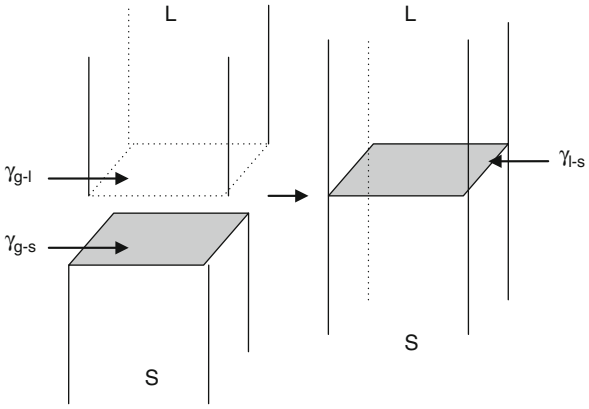
Work of Adhesion and Work of Cohesion, Fig. 3 Spreading process

Fundamental Definition

When a liquid contacts with a solid, the wetting of the liquid at the surface of the solid changes depending on the properties of the interface between the liquid and the solid and those of their respective surfaces. Some liquids drop into a solid and become small, round beads (e.g., a mercury drop on a desk surface). There are also some liquids that can spread into a layer of liquid film (e.g., dropping water on a piece of purified glass panel).

Under constant temperature and pressure, the interfaces of gas-liquid and gas-solid are transformed into a liquid-solid interface, which is shown in Fig. 4. Given that each interface has a unit area, the change of Gibbs' free energy for this course can be expressed as

$$\Delta G = \gamma_{l-s} - \gamma_{g-s} - \gamma_{g-l} \tag{1}$$



Work of Adhesion and Work of Cohesion, Fig. 4 Wetting process of liquid on the surface of solid

$$W_a = -\Delta G = \gamma_{g-s} + \gamma_{g-l} - \gamma_{l-s} \tag{2}$$

where γ_{g-s}, γ_{g-l}, and γ_{l-s} represent the Gibbs' free energy of gas-solid, gas-liquid, and liquid-solid interfaces, respectively. W_a is called *work of adhesion*, which is the maximum value of work done to the outside by the system in the course of liquid-solid adhesion. The larger the value of W_a, the more easily the liquid wets the solid and the stronger the bonding at the interface of the liquid and the solid.

When two liquid surfaces are transformed into a liquid pillar, the change of Gibbs' free energy can be expressed as

$$\Delta G = 0 - 2\gamma_{g-l} \tag{3}$$

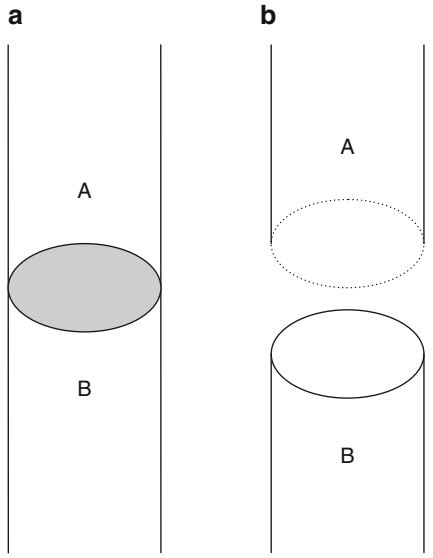
$$W_c = -\Delta G = 2\gamma_{g-l} \tag{4}$$

where W_c is called *work of cohesion*, which is the measure of the solid degree of the liquid combining itself.

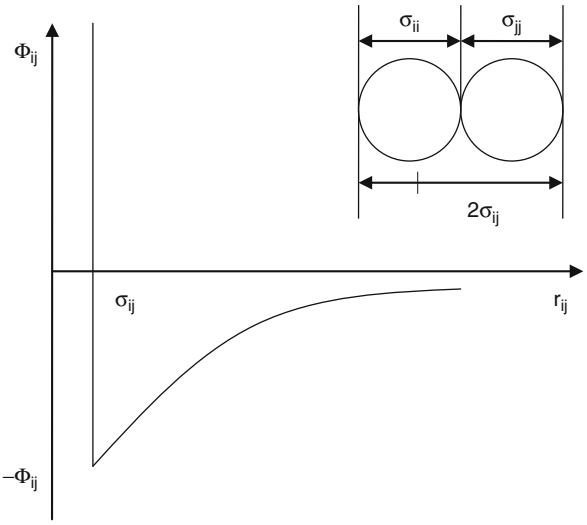
Scientific Fundamentals

Relative Theoretical Correlation

If A, B represents a liquid (L), a solid (S) separately (shown in Fig. 5), and $W_{AA} > W_{AB}$, it can be deduced that $\gamma_L > (\gamma_S - \gamma_{SL})$ from formula $W_{AB} = \gamma_B + \gamma_A - \gamma_{AB}$ and $W_{AA} = 2\gamma_A$, where W_{AA},W_{AB},γ_L,γ_S,γ_{SL},γ_B,γ_A,γ_{AB} represent the work of cohesion of A, work of adhesion of A and B; surface tension of L, S; interface tension of S-L; surface tension of B, A; and interface tension of A-B respectively. Only if the wetting angle is not equal to zero can this condition be satisfied. Therefore, one may conclude that if a kind of binder cannot be spread automatically, it cannot bond (crack of adhesion) before it cracks (crack of cohesion).



Work of Adhesion and Work of Cohesion, Fig. 5 A mode for analysis (**a**) before and (**b**) after the separation of A and B



Work of Adhesion and Work of Cohesion, Fig. 6 Mode of molecular potential energy

Return to the definitions of work of adhesion and work of cohesion. One wonderful feature is that the work making the divided molecules move is connected with the interfacial tension. It is worth noting that the energy that depends on the distances of molecules is the potential energy of the molecule. Postulating that the work driving the cylinder material apart is just the change of the potential energy of molecules, the method can be obtained from which the correlation between interfacial tension and effect of molecule can be established. The potential energy between molecule i and j can be expressed as:

$$\left.\begin{array}{c}\text{when } |r_{ij}| > \sigma_{ij},\ \phi'_{ij} = \varepsilon_{ij}\left\{\frac{\sigma_{ij}}{r_{ij}}\right\}^6 \\ |r_{ij}| \le \sigma_{ij},\ \phi'_{ij} = +\infty\end{array}\right\} \quad (5)$$

where ε_{ij} is the characteristic energy, $|r_{ij}|$ is the distance between the mass centers of molecule i and j, and σ_{ij} is the characteristic length, which is the distance between the centers of molecule i and j when they contact with each other.

The first part of the potential energy is the negative value and is inversely proportional to the hexa-power of the distance of the molecules, which is called the dispersive potential energy of van der Waals or London force. It represents the absorption force between two molecules. The second part of the potential is positive, which represents the repulsion force between two molecules. In fact, this suggests that molecules should be hard balls and should not pierce each other. Figure 6 is the mode of the molecular potential, which is also the plot of formula (5).

When $i = j$, the nearest distance between molecules is σ_{ij} (i.e., the diameter of one molecule). When i ≠ j, generally $\sigma_{ij} = 1/2(\sigma_{ii} + \sigma_{jj})$. Sometimes, it is more appropriate to take ${\sigma_{ij}}^3 = 1/2({\sigma_{ii}}^3 + {\sigma_{jj}}^3)$. It should be pointed out that these two forms both are approximate values. The simple blending rule for dispersive energy is $\varepsilon_{ij} = (\varepsilon_{ii}\varepsilon_{jj})^{1/2}$.

Formula (5) is applicable to nonpolar molecules. It can be written in a more appropriate form as follows:

$$\left.\begin{array}{c}\text{when } |r_{ij}| > \sigma_{ij},\ \phi_{ij} = -\frac{n_i n_j \beta_{ij}}{|r_{ij}|^6} \\ \text{when } |r_{ij}| < \sigma_{ij},\ \phi_{ij} = +\infty\end{array}\right\} \quad (6)$$

where n_i and n_j are the mole number of matter i and j in unit volume respectively, β_{ij} is a new constant, and the unit of ϕ_{ij} is energy/(volume of i)/(volume of j).

The absorption force between A and B, for unit area at a plane interface, is $(-\pi n_A n_B \beta_{AB})/(12{\sigma_{AB}}^2)$, which is determined by employing the formula for interactional energy by Hamauer. In this circumstance, the potential energy can be represented as $(-\pi n_A n_B \beta_{AB})/(12h^2)$, where the least value of h is σ_{AB} (i.e., the beeline between molecule A and B). At this time, A and B is considered as hard ball.

If the molecules of A and B are uniform, the work that causes the system to be divided into two parts along the plane A′B′ can be represented as:

${W_{AA}}^d = 2{\gamma_A}^b$ = (the interactional energy for two molecules with infinite distance) – (the interactional energy for two molecules near to each other)

There is no interaction when the distance between two molecules is infinite (h = ∞). Thus,

$$W_{AA}{}^{d} = 2\gamma_{A}{}^{b} = \frac{\pi n_A n_A \beta_{AA}}{12\sigma_{AA}} \tag{7}$$

where the superscript d represents the contribution of interface tension and work of adhesion to the dispersive potential energy. If the molecules are not polar, $W_{AA} = W_{AA}{}^{d}$ and $\gamma_A = \gamma_A{}^{d}$. Paddy examined the above relationship, whose correction improved upon the potential energy of molecules, liquid theory, and so on. He found that the theoretical value of interface tension for liquid-vapor is consistent with the corresponding experimental value.

Similarly, work of adhesion can be expressed as:

$$W_{AB}{}^{d} = \frac{\pi n_A n_B \beta_{AB}}{12\sigma^2{}_{AB}}. \tag{8}$$

One conclusion for the above equation is that when $\sigma_{AB} \approx \sigma$ and $\sigma_{AA} \approx \sigma_{BB} = \sigma$ (i.e., the sizes of molecules are comparable),

$$W_{AA}{}^{d} \cong 2(\gamma_A{}^{d}\gamma_B{}^{d})^{1/2}.$$

If molecules A and B interact only by dispersion effect,

$$W_{AB} = W_{AB}{}^{d} \cong 2(\gamma_A{}^{d}\gamma_B)^{1/2}.$$

Taking this into the formula $W_{AB} = \gamma_A + \gamma_B - \gamma_{AB}$, then

$$\gamma_{AB} = \gamma_A + \gamma_B - 2(\gamma_A{}^{d}\gamma_B{}^{d})^{1/2}. \tag{9}$$

This is Fowkes' equation, which is a special circumstance of the more general Girifalco–Good equation. However, it is worth noting that (9) is derived when A and B interact only by dispersion effect. If A and B are both nonpolar (i.e., $\gamma_A = \gamma_A{}^{d}$ and $\gamma_B = \gamma_B{}^{d}$), (9) changes as follows:

$$\gamma_{AB} = \gamma_A + \gamma_B - 2(\gamma_A\gamma_B)^{1/2}. \tag{10}$$

Furthermore, when either A or B is nonpolar (i.e., their interaction is nearly only by dispersion effect), (9) is very close. In fact, if A and B are both polar and A is a solid, (9) is also correct. In this case, the main polar interaction energy is zero and the dispersion interaction is overwhelming.

Good–Girifalco Relationship

If two surfaces contact with each other and there is no interaction of molecules at the interface, the interface tension can be expressed as

$$\gamma_{12} = \gamma_1 + \gamma_2.$$

In fact, it is impossible that there is no interaction of molecules at the interface region. Thus, work of adhesion can be defined as the reduction of Helmholtz free energy caused by interaction of molecules while two surfaces contact with each other:

$$W_a = \gamma_1 + \gamma_2 - \gamma_{12}. \tag{11}$$

So,

$$\gamma_{12} = \gamma_1 + \gamma_2 - W_a. \tag{12}$$

This formula, proposed by Good, suggests that the interface tension should be related to the work of adhesion. Work of adhesion can be considered as the reversible work required when one interface is separated into two surfaces.

Similarly, work of cohesion is the reduction of Helmholtz free energy when two similar surfaces adhere together to obtain one uniform object:

$$(W_c)_1 = 2\gamma_1, \text{and } (W_c)_2 = 2\gamma_2. \tag{13}$$

Work of cohesion is also the reversible work required when one uniform object is cut so as to reveal two new surfaces.

Researchers, including Good and Girifalco, proposed the ratio of work of adhesion to the geometric mean for work of cohesion of every bulk phase, which is signified as Φ:

$$\Phi = \frac{(W_a)_{12}}{\left[(W_c)_{11}(W_c)_{22}\right]^{1/2}}. \tag{14}$$

Taking this formula into formula (12), Good and Girifalco's equation can be obtained:

$$\gamma_{12} = \gamma_1 + \gamma_2 - 2\Phi(\gamma_1\gamma_2)^{1/2}, \tag{15}$$

and $\Phi = \Phi_V\Phi_A$, $\Phi_V = \frac{4V_1{}^{1/3}V_2{}^{1/3}}{(V_1{}^{1/3}+V_2{}^{1/3})^2} = \frac{r_{11}r_{22}}{(r_{12})^2}$

$$\Phi = \frac{\frac{3}{4}\alpha_1\alpha_2\left(\frac{2I_1I_2}{I_1+I_2}\right) + \alpha_1{\mu_2}^2 + \alpha_2{\mu_1}^2 + \frac{2}{3}\left(\frac{{\mu_1}^2{\mu_2}^2}{kT}\right)}{\left[\left(\frac{3}{4}\alpha_1{}^2I_1 + 2\alpha_1{\mu_1}^2 + \frac{2}{3}\frac{{\mu_1}^4}{kT}\right)\left(\frac{3}{4}\alpha_2{}^2I_2 + 2\alpha_2{\mu_2}^2 + \frac{2}{3}\frac{{\mu_2}^4}{kT}\right)\right]^{1/2}} = \lambda_{12}/(\lambda_{11}\lambda_{22})^{1/2} \tag{16}$$

where V is the mole volume after interaction, r is the distance between two molecules, I is ionization energy, α is polarizability, μ is permanent dipole moment, and λ is a constant for molecular attractive force.

It can be seen from formula (16) that Φ_A is closely related to the polarity of material. Φ_A obtains its maximum value when the materials of two phases have the same polarity (i.e., Φ_A is equal to 1). If the polarity of the two phases differs more greatly, the value of Φ_A becomes smaller. If both the two phases have no polarity, $\mu_1 = \mu_2 = 0$. So,

$$\Phi_A = \frac{2(I_1 I_2)^{1/2}}{I_1 + I_2}. \quad (17)$$

When $I_1 = I_2$, $\Phi_A = 1$.

The smaller the distances between molecules, the more closely Φ_V approaches to 1.

From the Good–Girifalco relationship, it can be seen that when the parameter Φ of the interaction between two phases increases, the interface tension reduces. The closer the polarity of the two phase materials, the lower the interface tension. Therefore, interface tension is mainly controlled by the polarity of the two phases.

Theoretical Relationship of Relative Polarity

Because the interaction forces between various molecules can be added together linearly, the whole attractive interaction constant of one matter is the result of the contribution from the attractive interaction constant for each type, which is expressed as

$$\lambda_{12} = \lambda_{12}{}^{d} + \lambda_{12}{}^{p} + \cdots\cdots, \quad (18)$$

where the superscripts d and p represent dispersive force and permanent dipole moment, respectively.

Fowkes believed that the work of adhesion across one interface between two phases is the simple sum of the work of adhesion produced by every specific type of interaction between molecules, which is expressed as

$$W_a = W_a{}^{d} + W_a{}^{h} + W_a{}^{p} + W_a{}^{i} + W_a{}^{\pi} + W_a{}^{da} + W_a{}^{e} + \cdots\cdots. \quad (19)$$

In the formula, $W_a{}^{d}$, $W_a{}^{h}$, $W_a{}^{p}$, $W_a{}^{i}$, $W_a{}^{\pi}$, $W_a{}^{da}$, and $W_a{}^{e}$ represent the work of adhesion produced by London dispersive force, hydrogen bond, dipole-dipole, dipole-induced dipole, π bond, donor-acceptor bond, and electrostatic interaction, respectively.

Work of cohesion can be transformed into a formula similar to (19) and $(W_C)_1 = 2\gamma_1$, so surface tension can be expressed as

$$\gamma_1 = \gamma_1{}^{d} + \gamma_1{}^{h} + \gamma_1{}^{p} + \cdots\cdots. \quad (20)$$

The influential interaction force, happening between two different phases across the interface, should be one of those types of interaction forces that are commonly possessed by the two phases. For example, dispersive force is common to various materials. Thus $W_a{}^{d}$ is one of the dominant items on the right side of (19). If one of the two phases only has the interaction of dispersive force, W_a mainly consists of $W_a{}^{d}$.

Generally, surface tension can be expressed as the whole contribution of two parts (i.e., dispersive force and dipole moment):

$$\gamma = \gamma^{d} + \gamma^{p}. \quad (21)$$

In the same way, work of adhesion can be also expressed in a similar form as

$$W_a = W_a{}^{d} + W_a{}^{p}. \quad (22)$$

So, formula (12) can be written as

$$\gamma_{12} = \gamma_1 + \gamma_2 - W_a{}^{d} - W_a{}^{p}. \quad (23)$$

Geometric Mean Equation

From the formula

$$\gamma_1 = \pi n_1{}^{2} \lambda_{11}/32r^2$$

work of cohesion can be also expressed as follows:

$$(W_C)_1 = \pi n_1{}^{2} \lambda_{11}/16r^2. \quad (24)$$

Strictly in accordance with the same method, employing the equation

$$\gamma = \pi n^2 \lambda/32r^2,$$

the work of adhesion for two phases can also be calculated. The result is similar to (24):

$$(W_a)_{12} = \pi n_1 n_2 \lambda_{12}/16 r_{12}{}^{2}. \quad (25)$$

Good, Girifalco, and others proposed formula (14). After ten years, they reexamined some demonstrations relevant to the relation between the attractive interaction constant λ_{11}, λ_{22}, and λ_{12} in this formula and deduced that if there is only the dispersive force interaction between molecules, their relation can be expressed as

$$\lambda_{12} = (\lambda_{11}\lambda_{22})^{1/2}. \quad (26)$$

This formula is the geometric mean relationship. Merging this formula into formula (14),

$$\Phi = \frac{r_{11} r_{22}}{(r_{12})^2}. \quad (27)$$

Combining formulas (12), (13), and (14),

$$\Phi = \frac{\gamma_1 + \gamma_2 - \gamma_{12}}{2(\gamma_1 \gamma_2)^{1/2}}. \quad (28)$$

So,

$$W_a = 2\Phi(\gamma_1 \gamma_2)^{1/2}, \quad (29)$$

$$W_a{}^{d} = 2\Phi(\gamma_1{}^{d} \gamma_2{}^{d})^{1/2}, \quad (30)$$

$$W_a{}^{p} = 2\Phi(\gamma_1{}^{p} \gamma_2{}^{p})^{1/2}. \quad (31)$$

From this it can be known that work of adhesion and work of cohesion can be also represented by the surface tension of each phase and the interface tension. When the

system interacts via the dispersive force, the constants relevant to the attractive interaction can keep the geometric mean relation, and the value of Φ obtained is in the range of 1 ± 0.2. Thus, formula (23) transforms as follows:

$$\gamma_{12} = \gamma_1 + \gamma_2 - 2(\gamma_1{}^d\gamma_2{}^d)^{1/2} - 2(\gamma_1{}^p\gamma_2{}^p)^{1/2}. \quad (32)$$

This formula is the geometric mean equation. It is satisfactory to speculate the value of the interface tension between low and high surface energy materials from the equation.

According to the above finding, when the attractive interaction of dipole moment takes place at the interface of the involved systems with the equal interaction of dipole moment occurring at the interface of two phases, and the dispersive force is nearly equal, the value of Φ in (28) tends to 1. When the interaction of the aforementioned molecules across the interface is controlled by the interaction of dispersive force, the value of Φ is 1. If there is polar force at the interface, and the interactions of both dipole and dispersive force between the two phases are not equal, the value of Φ is less than 1. The closer the polarity, the higher the value of Φ.

Harmonic Average Equation

Wu believed that the harmonic average relation is equivalent to the geometric mean relation. Therefore, the constant of the attractive interaction at interface can be also defined as $\lambda_{12}{}^h = (\frac{1}{\lambda_{11}} + \frac{1}{\lambda_{22}})$. When phases 1 and 2 possess similar polarity, the work of adhesion can be expressed as the harmonic average of the work of cohesion as follows:

$$W_a{}^d = \frac{2(W_c{}^d)_1 \times (W_c{}^d)_2}{(W_c{}^d)_1 + (W_c{}^d)_2} = \frac{4\gamma_1{}^d\gamma_2{}^d}{\gamma_1{}^d + \gamma_2{}^d}, \quad (33)$$

$$W_a{}^p = \frac{2(W_c{}^p)_1 \times (W_c{}^p)_2}{(W_c{}^p)_1 + (W_c{}^p)_2} = \frac{4\gamma_1{}^p\gamma_2{}^p}{\gamma_1{}^p + \gamma_2{}^p}. \quad (34)$$

So,

$$\gamma_{12} = \gamma_1 + \gamma_2 - \frac{4\gamma_1{}^d\gamma_2{}^d}{\gamma_1{}^d + \gamma_2{}^d} - \frac{4\gamma_1{}^p\gamma_2{}^p}{\gamma_1{}^p + \gamma_2{}^p}. \quad (35)$$

This formula is the harmonic average equation. It is relatively satisfactory to speculate the interface tension of the materials of low surface energy from the formula.

Key Applications

The conclusion that the harmonic average relation and the geometric mean relation are equivalent, is derived from the fact that there is no obvious difference between the results obtained by employing the constants of the attractive interaction of these two relationships separately. It is more accurate to calculate the interface tension of polymers by the harmonic average equation than by the geometric mean equation for a polar-polar polymer system. But it is simple to calculate by employing the geometric mean equation.

Applying either the geometric mean equation or the harmonic average equation, one can only detect the contact angle of the tested liquid on the surface of a standard solid, which possesses the composition of dispersive force (e.g., refined solid paraffin), or test the contact angle of a saturated hydrocarbon homologue droplet on the surface of a solid possessing polar composition (e.g., cellulose diacetate film). Then, solving relative simultaneous equations respectively, one can determine the values of γ^d and γ^p of each solid or liquid and calculate the correlative values of interface tension according to these values.

Cross-References

- ▶ Surface Free Energy
- ▶ Surface Tension

References

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

T. Cosgrove, *Colloid Science: Principles, Methods and Applications* (Blackwell, Ames, 2005)

E.A. Irene, *Surfaces, Interfaces, and Thin Films for Microelectronics* (Wiley-Interscience, Hoboken, 2008)

Worm Gear Lubrication and Efficiency

- ▶ Gear Efficiency

Wormgear Lubrication

- ▶ Gear Lubrication