

Tribology

Tribology

Friction and Wear of Engineering Materials

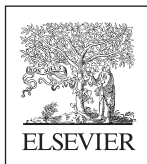
Second Edition

Ian Hutchings

University of Cambridge, UK

Philip Shipway

University of Nottingham, UK



Butterworth-Heinemann
An imprint of Elsevier

Butterworth-Heinemann is an imprint of Elsevier
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States

© 2017 Ian Hutchings and Philip Shipway. Published by Elsevier Ltd. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

ISBN: 978-0-08-100910-9

For information on all Butterworth-Heinemann publications
visit our website at <https://www.elsevier.com/books-and-journals>



Working together
to grow libraries in
developing countries

www.elsevier.com • www.bookaid.org

Publisher: Matthew Deans

Acquisition Editor: Christina Gifford

Editorial Project Manager: Lindsay Lawrence

Production Project Manager: Mohana Natarajan

Cover Image: courtesy of Stephen R. Pearson, University of Nottingham

Typeset by SPi Global, India

Preface to the second edition

Over the 25 years since the first edition was published it has been continually in print, and has attracted an international readership to become a widely used and widely cited text. In collaborating to update and extend this book, we have maintained its core emphasis on friction and wear of materials, but have also strengthened its coverage of the more traditional tribological topics of contact mechanics and lubrication. We have drawn heavily on our experience in teaching courses on tribology at undergraduate and postgraduate level at the Universities of Cambridge and Nottingham, in the annual Cambridge Tribology Course, and in Tutorial sessions at the biennial International Conferences on Wear of Materials, and also as Editors-in-chief of the journal *Wear*.

Readers familiar with the first edition will, we hope, find the level of treatment to be unchanged, and should find some of the new material useful and informative. While the fundamental principles underlying tribological phenomena remain the same, the world in which engineers and scientists operate has changed significantly, with an increased awareness of the environmental implications of the design and use of engineering systems. Tribologists have an important part to play in reducing the global use of energy, of fossil carbon and of strategic and hazardous materials, and also in contributing to human health and wellbeing. In refreshing the book's content we have revised and updated all the chapters and also included several new case studies on topics ranging from magnetic data storage to artificial hip joints that illustrate, we hope, some of the modern engineering applications in which tribological principles play vital roles. In response to demand we have also included questions for each chapter, suitable for classroom use or individual study.

Ian Hutchings
Philip Shipway
October 2016

Preface to the first edition

Tribology, the study of friction, wear and lubrication, is an interdisciplinary subject that draws on the expertise of the physicist, the chemist and the mechanical engineer, as well as the materials scientist or metallurgist. It can therefore be approached from several different viewpoints. Most previous textbooks have focused on topics in tribology that can be modelled with mathematical precision: contact mechanics, fluid film lubrication and bearing design, for example. These are undoubtedly important. In the present book, however, the emphasis is more on an equally important subject that is less amenable to precise quantitative analysis: the behaviour of materials in the context of tribology, and in particular friction and wear.

The book is intended for final-year undergraduate students of the physical sciences and technology, especially students of mechanical engineering, materials science and metallurgy. It should also be valuable for those taking postgraduate or post-experience courses, and since it provides numerous references to the research literature, I hope it will prove a useful source of initial information on tribology for scientists and engineers working in industry.

I am well aware that in offering a textbook on the tribological behaviour of materials I have undertaken a daunting task, particularly in tackling the subject of wear. Professor Duncan Dowson has remarked that whereas the scientific study of friction dates back some 300 years, and that of lubrication more than a century, wear has received similar attention for only 40 years. One can go even further, and suggest that our understanding of wear mechanisms has developed most rapidly only with the widespread use of electron microscopy and instrumental methods of microanalysis over the past 20 years. In a subject so young and so complex, it is inevitable that there will still be competition between theories, confusion over nomenclature and definitions, and inconsistencies between experimental observations. Nevertheless, the foundations of the subject now seem to be well established, and I have tried to present them, as well as more recent developments, in such a way that the reader will be able to appreciate the future advances that will certainly occur.

I have listed recommendations for further reading at the end of each chapter, but the sources of the numerous illustrations, which I have tried to cite accurately in the captions, should also provide valuable points of entry into the research literature. I am very grateful to numerous authors and publishers for their permission to reproduce copyright material.

Ian Hutchings
Cambridge, 1992

Introduction

1

The movement of one solid surface over another is fundamentally important to the functioning of many kinds of mechanisms, both artificial and natural. The subject of this book, *tribology*, is defined as ‘the branch of science and technology concerned with interacting surfaces in relative motion and with associated matters’, and includes the study of friction, wear, lubrication and the design of bearings.

Mechanical systems in which surfaces do not slide or roll against each other are rare, and tribology is therefore a key enabling technology with very wide application. Progress in understanding the surface interactions that lie at the heart of tribology has demanded the skills of mechanical engineers, materials scientists, physicists and chemists; in the context of bio-tribology which concerns living organisms, the discipline becomes even wider and includes the medical and biological sciences. Advances in tribology have underpinned much of the world’s engineering progress. Its reach extends to all aspects of modern technology, in the broadest sense of the word, in which surfaces move against each other, from transport and power generation to medical engineering, food science and cosmetics.

Although the subject of tribology is much wider than a study of friction, friction does play a central role in the performance of many mechanical systems. In some cases, low friction is desirable and even essential. The satisfactory operation of joints, for example, whether hinges on doors, human hip joints or bridge supports, demands a low friction force. Work done in overcoming friction in bearings and gears in machines is dissipated as heat, and by reducing friction we can achieve an increase in overall efficiency. But low friction is not necessarily beneficial in all cases. In brakes and clutches, adequate and controlled friction is essential to dissipate kinetic energy and transfer torque; high friction is similarly desirable between a vehicle tyre and the road surface, just as it is between the human foot and the ground for walking.

Whenever surfaces move over each other, wear will occur: damage to one or both surfaces, generally involving progressive loss of material. Sometimes the wear is imperceptibly slight, but it can also be extremely rapid. In most cases, wear is detrimental, perhaps causing increased clearances between the moving components, unwanted freedom of movement and loss of precision. It often leads to vibration, to increased mechanical loading and yet more rapid wear, and sometimes to fatigue failure. The loss by wear of relatively small amounts of material can be enough to cause complete failure of large and complex machines, and there are unfortunately cases where the root causes of major engineering disasters can be traced back to wear. High wear rates, however, are sometimes desirable. Grinding and polishing, for example, are widely used manufacturing processes that employ wear to remove material in a controlled manner, and an initial small amount of wear is often anticipated and even welcomed during the ‘running-in’ process in some kinds of machinery.

A key method of reducing friction, and often also wear, is to lubricate the system in some way, and the study of lubrication is very closely related to that of friction and

Table 1.1 Methods by which financial savings could be made through improved tribological practice in UK industry. The proportions represent proportions of the total annual saving (which was estimated at £515 million at 1965 prices) (information from HMSO, 1966. Lubrication (Tribology) Education and Research. UK Department of Education and Science)

Savings in maintenance and replacement costs	45%
Savings in losses resulting from breakdowns	22%
Savings in investment through increased life of plant	20%
Reduction in energy consumption through lower friction	5%
Savings in investment through greater availability and higher efficiency	4%
Reduction in manpower	2%
Savings in lubricant costs	2%

wear. Even when an artificial lubricant is not added to a system, components of the surrounding atmosphere (especially the oxygen and water vapour in the air) often play a similar role and must be considered in any study of the interaction of the surfaces.

The word ‘tribology’ (from the Greek: τριβος= rubbing or attrition) was first coined in 1966 by a UK government committee, although friction, lubrication and wear had been studied for many years before then and have a long and fascinating history. That committee also made an estimate of the savings that could be made by UK industry if known tribological principles were widely applied. Similar exercises have subsequently been carried out in several other countries. Their conclusions are all in broad agreement: that at least 1% of the gross domestic product might be saved with minimal further investment in research, and that the potential for even larger savings might exist with further research. The savings arise from several sources. The original survey identified the savings listed in [Table 1.1](#). The reduction in energy consumption through reduced friction was, in money terms, rather a small component of the total; savings in maintenance and replacement costs, in losses consequential upon breakdowns, and in investments through the increased life of machinery were at that time much more important. A subsequent survey in the UK, taking account of the relative increase in the cost of energy since 1966, focused on the savings in energy that can be made by improved tribological design, and identified the kinds of savings shown in [Table 1.2](#).

Substantial advances in tribological understanding and in the application of that knowledge have occurred since the word ‘tribology’ was first defined. We have only to look at the advances in road and air transport, power generation, information technology, manufacturing processes and medical engineering to see important examples. But further development will undoubtedly occur, and it is more important than ever that an understanding of the principles of tribology forms part of the education of every engineer and applied scientist. There is now an increasing, and overdue, awareness of the imperative to conserve the resources of our planet, including fossil fuels and materials. In reducing the environmental impact of human activity, we need not only to reduce the atmospheric release of fossil carbon and generate energy in sustainable ways, but also to use

Table 1.2 Methods by which savings of energy could be made through improved tribological practice in UK industry (information from Jost, H.P., Schofield, J., 1981. Proc. Inst. Mech. Engrs. 195, 151–195)

<i>Direct savings of energy</i>	
Primary:	saving of energy dissipated by friction
Secondary:	saving of energy needed to fabricate replacement parts
Tertiary:	saving of energy content of materials for replacement parts
<i>Indirect savings of energy</i>	
Savings consequential on direct savings, e.g., in plant needed to compensate for frictional losses	

materials, whether in capital equipment or in consumables such as lubricants, in a sustainable and environmentally-friendly manner. Not only should the correct application of tribological principles lead to longer-lasting, more energy- and resource-efficient products, but there are also situations where their neglect can lead to unacceptable performance, to costly product recalls with significant reputational damage for the supplier, and in extreme cases to failures that lead to human injury or even loss of life.

In the following chapters we shall explore various aspects of tribology. We start by examining the topography of surfaces and the way in which they interact when placed in contact ([Chapter 2](#)). The origins of friction, and the frictional responses of metals and non-metals are discussed next ([Chapter 3](#)). Lubricants and lubrication form the subject of [Chapter 4](#), and wear is the topic of [Chapters 5](#) and [6](#). In these two chapters the distinction is made between wear that occurs when two relatively smooth surfaces slide over each other, either lubricated or unlubricated (i.e., sliding wear, covered in [Chapter 5](#)), and wear involving hard particles (i.e., abrasive or erosive wear, covered in [Chapter 6](#)). The terms ‘sliding wear’, ‘abrasion’ and ‘erosion’ are not intended to describe mechanisms of wear, but form broader classifications. As we shall see later, many different mechanisms can be involved in wear and it would be simplistic to attempt to list them here. The important topic of surface engineering is reviewed in [Chapter 7](#), and ways in which tribological principles can be taken into account in the design process are discussed in [Chapter 8](#). The book concludes with a selection of case studies in [Chapter 9](#) that demonstrate the application of the principles developed in the earlier chapters: bearings, automotive tribology, tribology in manufacturing, natural and artificial hip joints, and magnetic data storage. Background material on indentation hardness, corrosion and the definition of wear rate is covered in three appendices.

SOURCES AND FURTHER READING

ASTM, 2015. Standard Terminology Relating to Wear and Erosion, Standard G40.
[Dowson, D., 1998](#). History of Tribology, second ed. Wiley.
[HMSO, 1966](#). Lubrication (Tribology) Education and Research. UK Department of Education and Science.

-
- Hutchings, I.M., 2016. A lot more than lubrication. *Ingenia* 66, (<http://www.ingenia.org.uk/Ingenia/Articles/1009>).
- Jost, H.P., 1989. Tribology—origin and future. *Wear* 136, 1–18.
- Nosonovsky, M., Bhushan, B., 2010. Green tribology: principles, research areas and challenges. *Phil. Trans. Roy. Soc. A* 368, 4677–4694.

QUESTIONS FOR CHAPTER 1

(These general questions can be attempted at any point, but as students' understanding and knowledge of tribology deepens, they should produce fuller answers.)

Question 1.1

Use any sources available to you, including the Internet, to research and write about the following:

- (a) examples of advances in engineering over the past 50 years that can be attributed to better understanding and use of tribological principles;
- (b) examples of failures in engineering systems (which can include household products, power generation systems, air or land transport, and medical devices) due to tribological causes.

In the case of (b), explain how better understanding or application of tribological knowledge would have avoided the failures or mitigated their consequences.

Question 1.2

Use any sources available to you, including the Internet, to research and write about the following:

- (a) the role of tribology in energy conservation;
- (b) the role of tribology in materials conservation;
- (c) tribology and safety.

Surface topography and surfaces in contact

2

2.1 INTRODUCTION

When studied on a sufficiently fine scale, all solid surfaces are found to be uneven. Ultimately, the surface irregularities of the smoothest possible surfaces will be on the scale of individual atoms or molecules; for example, carefully cleaved specimens of the mineral mica can be truly smooth on an atomic scale over areas of several square centimetres. However, the surfaces of even the most highly polished engineering components show irregularities appreciably larger than atomic dimensions, and many different methods have been employed to study their topography. Some involve examination of the surface by electron or light microscopy or by other optical methods, while others use the contact of a fine stylus, or electrical or thermal measurements, or rely on the leakage of a fluid between the surface and an opposing plane. Perhaps the highest resolution can be achieved by the technique of atomic force microscopy (AFM), which can resolve individual atoms, but for most engineering surfaces less sensitive methods are adequate to study their topography.

In this chapter we shall see how surface roughness is measured, and then examine what happens when two surfaces are placed in contact and a load is applied.

2.2 MEASUREMENT OF SURFACE TOPOGRAPHY

The topography of a surface can be represented by a dataset that contains the co-ordinates of points which lie on the surface. However, to resolve the features properly that make up the topography of the surface, the distance between neighbouring measurement points needs to be significantly smaller than the size of those features. Methods are required that allow the co-ordinates of points on the surface to be appropriately measured and recorded.

Stylus profilometry and atomic force microscopy are both techniques in which the co-ordinates of points on a surface are measured by interactions with a probe; the co-ordinates of individual points are measured sequentially, which limits the speed at which measurements can be recorded. The surface topography is described by a dataset describing those co-ordinates, either along a line or across an area. In contact profilometry, a fine stylus is dragged smoothly and steadily across the surface under examination. The position of the stylus in the plane of the surface is recorded (usually via the movement actuation system) and the vertical position of the stylus is monitored via a transducer. The co-ordinates of points regularly spaced along a line are recorded. To record a dataset which takes data from an area of the surface (rather than just a line across the surface), individual profiles can simply be recorded along a series of

parallel lines. AFM uses a much finer stylus on a flexible cantilever, with a piezo actuator commonly being employed to maintain a constant force between the surface and the stylus tip itself (an attractive force is maintained in non-contact mode and a repulsive force in contact mode). With AFM, surface co-ordinates can be measured to a much greater accuracy, but the principles of measuring line profiles or surface topography are the same as for stylus profilometry.

With contact profilometry, an unavoidable limitation results from the shape of the stylus. For reasons of strength, the diamond styli used for standard contact profilometry have pyramidal or conical ends, with minimum included angles of about 60° and tip radii of $1\text{--}2.5\text{ }\mu\text{m}$. The combination of finite tip radius and included angle prevents these styli from penetrating fully into deep narrow features of the surface. In some applications this can lead to significant error. Special styli, with chisel edges and minimum tip radius as small as $0.1\text{ }\mu\text{m}$, can be used to examine very fine surface details where a conventional stylus would be too blunt, but all stylus methods inevitably produce some ‘smoothing’ of the profile due to the finite dimensions of the stylus tip, as illustrated in Fig. 2.1.

With stylus profilometry, the lateral resolution is limited by the tip geometry, with a resolution of 200 nm being typical. The vertical resolution is around 10 nm . AFM can achieve a vertical resolution of 0.1 nm , but the lateral resolution is typically no more than 10 nm .

An issue exists in the examination of very compliant or delicate surfaces by stylus methods because the load on the stylus, although small, may nevertheless cause sufficient local stress to distort or damage the surface. Non-contacting, optical methods of surface measurement are attractive for such applications.

The other main class of methods for measurement of surface co-ordinates is based upon optical measurements, with the two main techniques using either interferometry or optical focus to define the surface position. Both these techniques are microscope-based and use information from the whole field of view to determine surface heights on a pixel-by-pixel basis. They both use computer-based image processing algorithms to determine the surface position for each pixel (using maximum fringe contrast in the interferometry techniques and maximum image sharpness in the focus techniques).

The height resolution of microscope-based techniques is typically $\sim 3\text{ nm}$. The lateral resolution depends on the magnification and the objective lens employed, but is

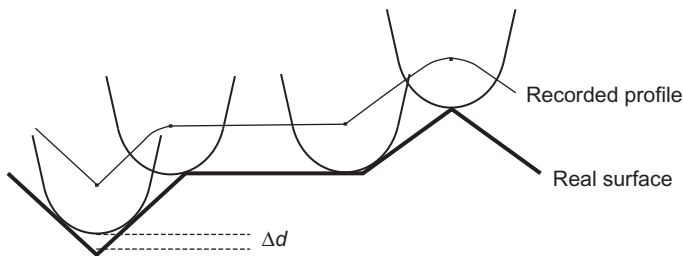


Fig. 2.1 Illustration of the effects of a finite stylus shape on the recorded profile of a surface; the recorded depth of valleys is reduced by Δd , and peaks are rounded off

limited to $\sim 1\text{ }\mu\text{m}$. Lateral resolution can be increased by the use of lenses with higher magnification, but this reduces the measurement area. In most systems, large areas can be analyzed at high resolution by stitching together data from separate measurement areas. Care must be taken with surfaces that are transparent or translucent, and these may need to be sputter coated with a metal to make them fully reflective; a thin sputtered coating (typically 2–20 nm thick) in practice has a negligible effect on the surface profile measurement. In addition, it is commonly observed that the surface position cannot be resolved for every pixel, and the processing software will generally estimate the position of the surface in these pixels based upon measurements from surrounding pixels. The user of such methods needs to be assured that the fraction of pixels where the height has been estimated in this way remains acceptable, and should be aware of this when reporting the results.

Once the topographic data has been collected, it is often represented either graphically, or in the form of one or more quantitative measures as discussed in [Section 2.3](#). Single line profiles can be derived from area-based profile measurements by an interpolation procedure. The graphical representation of the surface profile differs from the shape of a genuine cross-section through the surface for several reasons. The major difference is due to the different magnifications used to represent the dataset in the horizontal and vertical directions. The vertical extent of surface features is nearly always much less than their horizontal scale. It is therefore convenient to compress the graphical record of the surface profile by using a magnification in the direction normal to the surface that is greater than that in the plane of the surface. The ratio of the magnifications will depend on the roughness of the surface, but typically lies between 10 and 5000. In [Fig. 2.2](#), the effect of magnification is illustrated via profilometer recordings of the same surface at different magnification ratios of 50:1, 5:1 and 1:1.

The surface slopes therefore appear much steeper on a typical profilometer record than they really are, and the distorting effect of this horizontal compression, present in nearly all profilometer traces, must be recognized in interpreting them. Thus, although surface slopes may appear very steep on a profilometer trace, in reality they are very rarely steeper than 10° ; surface features on opposing surfaces that appear sharp enough to interlock with each other are rarely so sharp.

2.3 QUANTIFYING SURFACE ROUGHNESS

2.3.1 Quantitative measures of surface topography

It is convenient to differentiate between *roughness*, meaning the small scale irregularities of a surface, and *form error*, which is a measure of the deviation of the shape of the surface from its intended ideal shape (e.g., plane, cylindrical or spherical). The distinction between roughness and form error is arbitrary, although clearly it involves the horizontal scale of the irregularity. The picture may be further complicated by the presence of *waviness*, a periodic surface undulation intermediate in scale between roughness and form error.

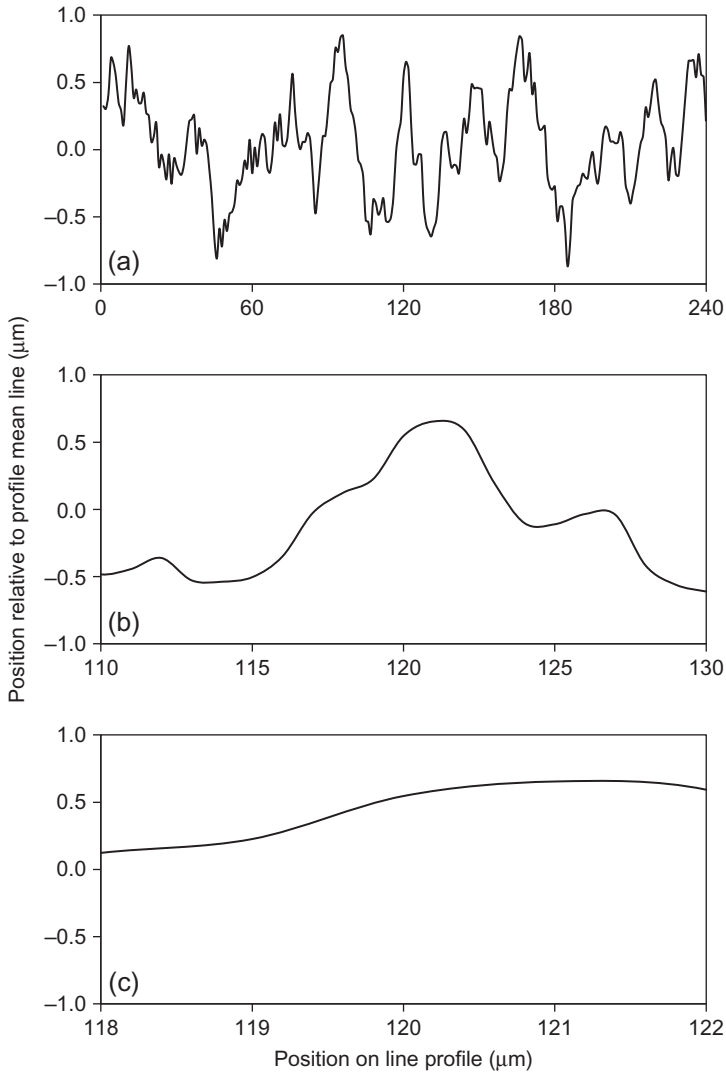


Fig. 2.2 A line surface profile shown with magnification ratios of (a) 50:1, (b) 5:1 and (c) 1:1; the data employed at the lower magnifications are the same as that at the higher magnifications over the line positions as indicated

By various methods, a surface profile can be analyzed to yield the various measurements of interest, namely the roughness (short wavelength irregularities), waviness or form error. Filtering methods may be used; for example, if the roughness of a surface needs to be separated from longer wavelength irregularities (roughness or form), then a sampling length can be defined (as shown in Fig. 2.3), with any measurements associated with those irregularities being made only with reference to the data

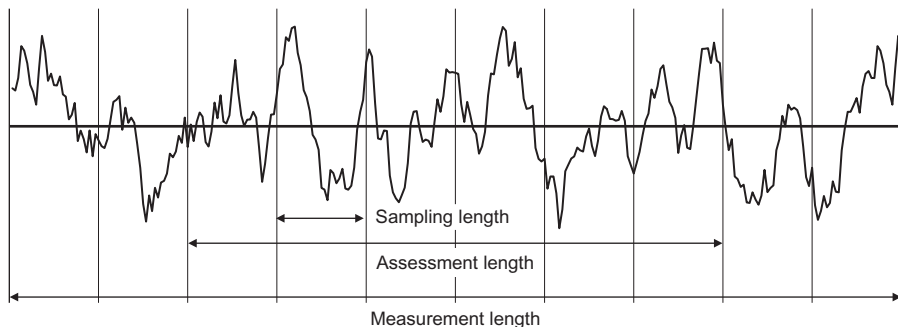


Fig. 2.3 Illustration of terms used in analyzing profilometric data

contained within the sampling length. These measurements can be made over a number of sampling lengths (which together make up the assessment length—see Fig. 2.3) and the values from each of these sampling lengths combined appropriately to yield a representative value.

A dataset that represents the surface co-ordinates over an area contains all the information required to describe the topography of the surface. These datasets are often shown graphically in a 3-D representation as illustrated in Fig. 2.4; in such a representation, the spatial distribution of surface heights can be seen, allowing features such as flaws and directional features such as grinding marks to be visualized. A dataset that has been measured along a single line is essentially a section through that surface; if the surface has no directionality, as in Fig. 2.4(a), then a line profile may represent it adequately, but if the surface has a dominant directionality as in Fig. 2.4(b), care must be exercised in presenting a line profile as representative.

While useful, graphical representations of surfaces (whether area or line profiles) do not provide a sufficiently simple and readily comparable means of describing surface topography. For this purpose, there are many quantitative measures that can be derived from the dataset representing the surface profile. Some quantitative measures were originally derived from line profiles and are still the most commonly quoted parameters; however, as area-based measurement of surface topology has become

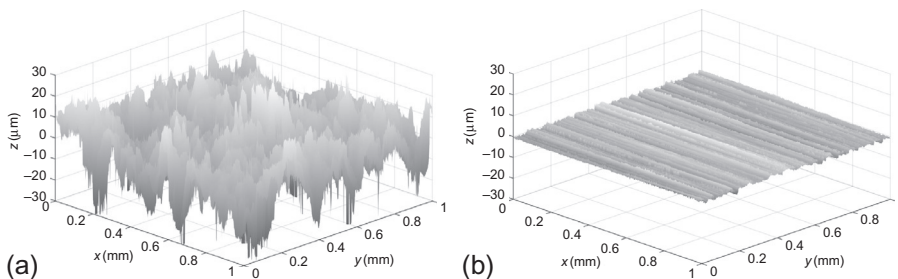


Fig. 2.4 Three dimensional plots of (a) a grit blasted steel surface; (b) a ground steel surface. Both plots are shown on the same scale

commonplace, equivalent measures for area-based datasets have also been derived and standardized, and are becoming more widely used.

The most commonly quoted measure of surface roughness is the *average roughness*, Ra . The average roughness is defined as the arithmetic mean deviation of the surface height from the mean line through the profile. The mean line is defined as the line corresponding to the long wave profile components of the surface; in practice, this is the line of best fit with equal areas of the profile lying above and below it, as shown in Fig. 2.5.

Formally, the average roughness Ra is defined by

$$Ra = \frac{1}{L} \int_0^L |z(x)| dx \quad (2.1)$$

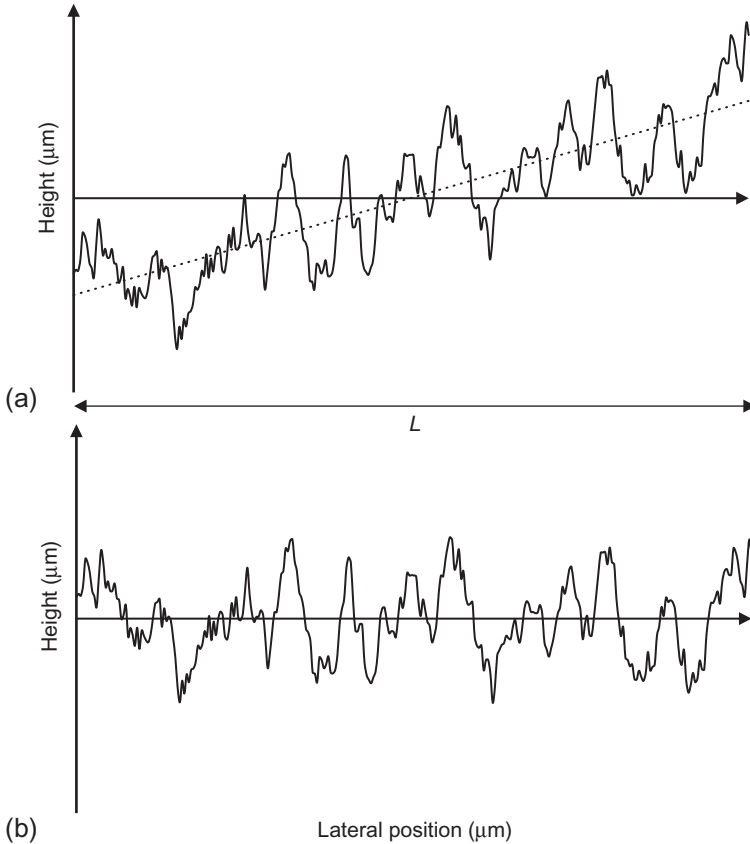


Fig. 2.5 Profile recorded over a sampling length L ; (a) raw data with a line of best fit (mean line) indicated as a dashed line; (b) profile relative to the mean line, from which the line profile-based roughness parameters can be calculated

where z is the height of the surface above the mean line at a distance x from the origin, and L is the length of the profile being considered (i.e., the sampling length as shown in Fig. 2.5). Normally, discrete points equally spaced along a profile are recorded (as opposed to a continuous function), leading to the following definition for average roughness:

$$Ra \cong \frac{1}{N} \sum_{i=1}^N |Z_i| \quad (2.2)$$

where Z_i is the height of the i th point and N is the number of points measured along the sampling length.

The *r.m.s. roughness* (Rq) is defined as the root mean square deviation of the profile from the mean line:

$$Rq = \sqrt{\frac{1}{L} \int_0^L (z(x))^2 dx} \quad (2.3)$$

or for discrete points:

$$Rq \cong \sqrt{\frac{1}{N} \sum_{i=1}^N Z_i^2} \quad (2.4)$$

For many surfaces, the values of Ra and Rq are similar; for a Gaussian distribution of surface heights, $Rq = 1.25 Ra$.

It is inevitable that in attempting to describe a profile by a single number, some important information about the surface topography will be lost. Ra and Rq , for example, give no information on the shapes or spacing of the surface irregularities, and convey no indication of the probability of finding surface heights within certain limits. For a fuller description of the topography of the surface, information is needed about the *probability distribution* of surface heights and the *spatial distribution* of peaks and valleys across the surface.

The need for a method of describing the distribution of surface heights is met by defining an *amplitude density function*, $p(z)$, the value of which, for any height z , is proportional to the probability of finding a point on the surface at height z above the mean line. The quantity $p(z)\Delta z$ is then the fraction of the surface profile that lies at heights between z and $z + \Delta z$ above the mean line, as shown in Fig. 2.6. A symmetrical profile, such as a sine curve, leads to an amplitude density curve that is symmetrical about the position of the mean line. Asymmetry of the surface profile leads to skewing of the amplitude density function, which therefore contains some information about the shapes of surface irregularities as well as their vertical extent. The r.m.s. roughness, Rq , is the standard deviation of the amplitude density function and in this context is sometimes given the symbol σ .

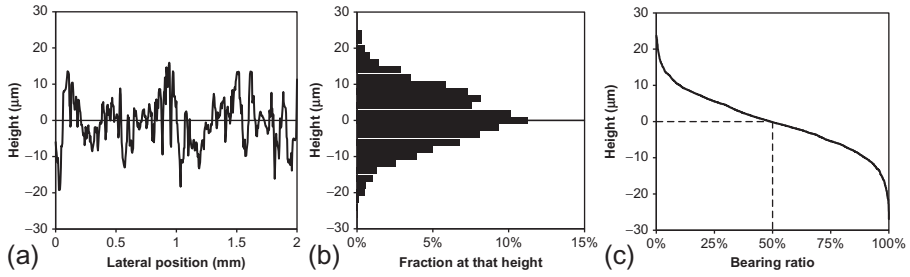


Fig. 2.6 Illustration of the derivation of (b) the amplitude density function and (c) the bearing ratio curve from profile data in (a) (the profile shows a subset of the whole dataset employed in computing the density functions)

The shape of the amplitude density curve can be described by its *skewness*, which provides a measure of its asymmetry, and *kurtosis* (from the Greek word for a hump), a measure of the sharpness of the peak of the distribution curve. Skewness (Rsk) is defined by

$$Rsk = \frac{1}{Rq^3} \int_{-\infty}^{\infty} z^3 p(z) dz \equiv \frac{1}{Rq^3} \frac{1}{L} \int_0^L (z(x))^3 dx \quad (2.5)$$

or

$$Rsk \cong \frac{1}{Rq^3} \left[\frac{1}{N} \sum_{i=1}^N Z_i^3 \right] \quad (2.6)$$

The value of Rsk is dominated by the heights furthest from the mean line, and can be either negative or positive. A symmetrical distribution of heights around the mean line will result in a skewness of zero. High peaks and comparatively shallow valleys lead to a positive skewness, although the majority of the surface heights lie below the mean line. The surface of a porous bronze bush used as a marginally-lubricated bearing (as discussed in [Section 9.2.3](#)) will contain deep pores to retain the oil and would exhibit a negative value of Rsk , as shown in the example in [Fig. 2.7](#).

The kurtosis (Rku) is defined by

$$Rku = \frac{1}{Rq^4} \int_{-\infty}^{\infty} z^4 p(z) dz \equiv \frac{1}{Rq^4} \frac{1}{L} \int_0^L (z(x))^4 dx \quad (2.7)$$

or

$$Rku \cong \frac{1}{Rq^4} \left[\frac{1}{N} \sum_{i=1}^N Z_i^4 \right] \quad (2.8)$$

Unlike skewness, the value of kurtosis does not distinguish between a peak and a valley in the profile. Rku is always positive since the height information is raised

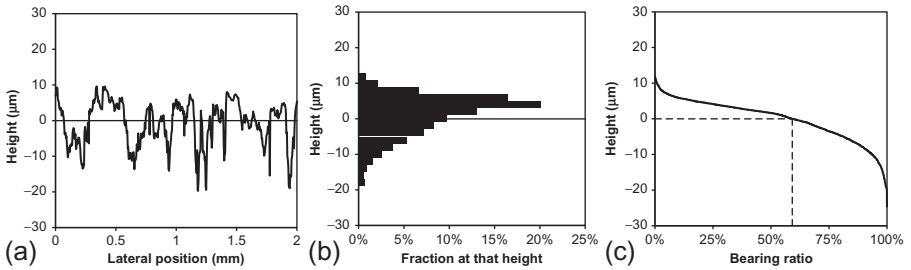


Fig. 2.7 Illustration of (a) a profile with an asymmetric distribution of heights above and below the mean line, together with its corresponding (b) amplitude density function and (c) bearing ratio curve. (The profile shows a subset of the whole dataset employed in the density functions). In this example, Rsk is negative

to an even (fourth) power. Both skewness and kurtosis can be readily calculated from the surface profile. A Gaussian (normal) probability distribution has a skewness value of zero and a kurtosis of 3.0. A kurtosis value greater than 3.0 indicates a broad, flat distribution curve (which could be described as relating to a ‘spiky’ surface), while a lower kurtosis corresponds to a more narrowly-peaked distribution (which might be described as relating to a ‘bumpy’ surface). Many surfaces, especially those resulting from wear processes rather than from deliberate machining, have height distributions that are close to Gaussian. Some combinations of machining process and material, however, can lead to appreciably non-Gaussian distributions.

Another function sometimes used in tribological studies and closely related to the amplitude density function is the *bearing ratio* curve (or *Abbot-Firestone* curve), which represents the fraction of points on a surface above a specified height relative to the mean line. This function can be understood by imagining a straight line, representing the profile of an ideally smooth plane surface, being brought slowly down towards the profile of the surface under investigation. When the plane first touches the surface at a point, the bearing ratio (defined as the ratio of the contact length to the total length of the profile) is zero. As the line is moved further downwards, the length over which it intersects the surface profile increases, and the bearing ratio therefore increases. Finally, as the line reaches the bottom of the deepest valley in the surface profile, the bearing ratio reaches 100%. The bearing ratio curve is a plot of bearing ratio against surface height, as shown in Figs. 2.6 and 2.7, and is the integral of the amplitude density function.

Neither the amplitude distribution nor the bearing ratio curve describes the way in which the hills and valleys are spatially distributed across a surface. Two methods may be used to extract this information from a profile: *autocorrelation* and *spectral analysis*. The *autocorrelation function* $C(\beta)$ of a profile graph is defined by:

$$C(\beta) = \frac{1}{L} \int_0^L z(x)z(x+\beta)dx \quad (2.9)$$

or

$$C(\beta) \cong \frac{1}{N} \sum_{i=1}^N Z_i Z_{i + \beta/s} \quad (2.10)$$

where s is the spacing between individual measurement points along the profile.

The value of the autocorrelation function for some displacement β along the surface is therefore derived by shifting the profile a distance β along the surface, multiplying the shifted profile function by the corresponding unshifted value, and calculating the area beneath the resultant product curve. When the displacement β is zero, the value of the autocorrelation function is a maximum and is simply the mean square roughness (Rq^2).

The autocorrelation function provides a measure of the correlation between the heights of the surface at positions separated by a distance β along the surface. The shape of the curve summarizes statistical information on the characteristic spacings (if any) of the surface features. Any regular undulation of the surface will show up as an oscillation of the same wavelength in the value of the autocorrelation function. For many real surfaces, the autocorrelation function decays steadily to zero as β increases, and may be approximated by an exponential function.

The *power spectral density* function, $P(\omega)$, which conveys direct information about the spatial frequencies (i.e., wavelengths) present in the surface profile, is the Fourier transform of the autocorrelation function:

$$P(\omega) = \frac{2}{\pi} \int_0^\infty C(\beta) \cos(\omega\beta) d\beta \quad (2.11)$$

Power spectral density is a particularly suitable function with which to describe machined surfaces, since it clearly depicts and separates any strong surface periodicities that may result from the machining process.

2.3.2 Areal measures

Contact between real surfaces occurs over areas (rather than along lines), and accordingly, areal profiles can provide understanding that is not available with line profiles. For example, the peaks on a particular line profile will not (in general) represent the highest points on the whole surface, but it will be at the highest points on the surface that contact is first made with a mating smooth surface. An areal profile will also contain information about directionality of surface texture, and about dispersed defects etc., that is generally not available in a line profile. As the instruments by which areal profiles can be measured have become more readily available, there has been a need to standardize both the vocabulary associated with those profiles and the definitions of the quantitative parameters which can be derived from the datasets. Many of the terms have been adapted from physical geography where they are used to describe features of the landscape; for example, terms such as contour line, ridge, dale and saddle point are all part of the standardized vocabulary for areal profiles.

The parameters that have been traditionally used to describe line profiles have generally been extended for use with areal profiles. For example, the areal average roughness (Sa) is defined by:

$$Sa = \frac{1}{A} \iint_A |z(x, y)| dx dy \quad (2.12)$$

As for line profiles, it can also be defined for a dataset of discrete measurement points, as follows:

$$Sa \cong \frac{1}{N} \sum_{i=1}^N |Z_i| \quad (2.13)$$

Equivalent functions for r.m.s. roughness, skewness, kurtosis, autocorrelation length etc. have all been defined for areal profiles. In addition, there are further parameters that can be derived only from areal profile information (such as the texture aspect ratio). Similarly, certain features of the areal surface profile can be characterized, which first involves identification of the features of interest. The features are generally classified as areal features (such as a hill), line features (such as a ridge) or point features (such as a peak). Once the features have been identified, then quantification can take place (such as the density of peaks, or the mean hill area).

2.4 THE TOPOGRAPHY OF ENGINEERING SURFACES

As a result of the manufacturing processes employed (for example, turning, milling or grinding), many surface finishes on engineering components exhibit directionality. This directionality is termed the *lay*, defined as the direction of the dominant surface pattern. If surfaces with a clear lay are characterized via a line profile, then standard methods of measurement generally require the line profile to run across the lay of the surface (i.e., perpendicular to the lay). The profile, and the measures of roughness derived from it, thus correspond to the maximum roughness of the surface, but give no information about the distribution of surface features in a direction along the lay, parallel to the machining marks. [Figure 2.4](#) illustrates the distinct lay of a ground surface, and the contrasting lack of directionality of a shot-blasted surface.

Height data gathered by profilometry may also be used to generate contour maps of surfaces, to study the statistics of feature distributions across two-dimensional areas, and to examine the area of contact between surfaces by numerically modelling their approach.

The topography of a freshly-machined surface depends on the machining process used to generate it as well as on the nature of the material. Likewise, the topography of a worn surface depends on the conditions under which the wear has occurred.

A great deal of information may be conveyed by the use of skewness and kurtosis values, but typically the roughnesses of engineering surfaces are frequently specified

Table 2.1 Typical average roughness values of engineering surfaces finished by different processes (data from Stout, K.J., 1981. *Mater. Eng.* 2, 287–295)

	<i>Ra</i> (μm)
Planning, shaping	1–25
Milling	1–6
Drawing, extrusion	1–3
Turning, boring	0.4–6
Grinding	0.1–2
Honing	0.1–1
Polishing	0.1–0.4
Lapping	0.05–0.4

only by average roughness (*Ra*) values. [Table 2.1](#) lists typical ranges of *Ra* value for engineering surfaces finished by various processes.

2.5 CONTACT BETWEEN SURFACES

When two nominally plane and parallel surfaces are brought gently together, contact will initially occur at only a few points. As the normal load is increased, the surfaces move closer together and a larger number of the higher areas or *asperities* on the two surfaces come into contact.

Since these asperities provide the only points at which the surfaces touch, they are responsible for supporting the normal load on the surface and for generating any frictional forces that act between them. If an electrical current passes from one surface to the other, then it must travel through the points of contact. An understanding of the way in which the asperities of two surfaces interact under varying loads is therefore essential to any study of friction, wear or electrical contact resistance; it is the concept of asperity contact that provides a physical explanation for many common tribological observations (such as the fact that for many materials, the coefficient of friction is independent of the nominal area of contact, as discussed in [Chapter 3](#)).

The macroscopic elastic stress distributions that result when surfaces are loaded against each other are also useful in considering subsurface deformation and damage by mechanisms such as plastic deformation, fracture or fatigue. These macroscopic stresses can be determined on the assumption that the surfaces in contact are smooth at the microscopic level. Many real contacts are non-conforming (such as that between a ball or roller and its raceway in a rolling element bearing, as discussed in [Section 9.2.1](#)), which means that contact takes place over a very small nominal area: even with very moderate loads, the resulting contact pressure can be very high. If two ideal surfaces (i.e., surfaces with no microscopic roughness) are loaded against each

other they will deform, initially elastically. The analysis of the elastic stress fields for such non-conforming contacts was first addressed by Heinrich Hertz in 1881, and is accordingly referred to as *Hertzian contact mechanics*.

2.5.1 Elastic stress fields in normally loaded contacts

Elastic point contact

When a sphere of an elastic material is pressed against another elastic sphere under a normal load W , as shown in Fig. 2.8, contact will occur between the two over a circular area of radius a , given by:

$$a = \left(\frac{3WR}{4E^*} \right)^{1/3} \quad (2.14)$$

Such a contact is generally referred to as a *point contact*. R , the relative radius of curvature of the contacting bodies, is defined in terms of the radii of the two bodies R_1 and R_2 as follows:

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} \quad (2.15)$$

The radii of the individual bodies are positive for convex surfaces and negative for concave. The radius of a plane surface is infinite, and R for a sphere in contact with a plane is then the same as the radius of the sphere.

The relevant elastic modulus is known as the *reduced modulus* (E^*) which depends on the Young's moduli of the two bodies, E_1 and E_2 , and on their respective Poisson's ratios, ν_1 and ν_2 , in the following way:

$$\frac{1}{E^*} = \frac{(1 - \nu_1^2)}{E_1} + \frac{(1 - \nu_2^2)}{E_2} \quad (2.16)$$

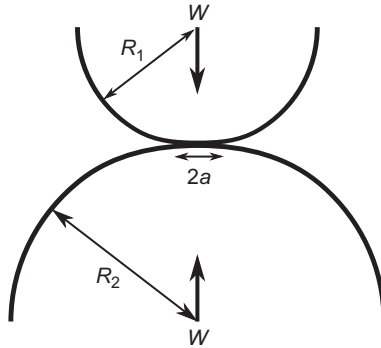


Fig. 2.8 Elastic deformation of a Hertzian point contact between two spherical surfaces under normal load W to form a contact circle with radius a

The pressure $p(r)$ in the contact varies with distance r from the centre of the contact as:

$$p(r) = p_o \sqrt{1 - \frac{r^2}{a^2}} \quad r \leq a \quad (2.17)$$

where p_o is the maximum pressure that occurs at the centre of the contact ($r = 0$). The load supported by the contact is simply the integral of the pressure over its area:

$$W = \int_0^a 2\pi r p(r) dr = \frac{2}{3} p_o \pi a^2 \quad (2.18)$$

The mean pressure acting over the contact area, p_m , is also equal to the applied load divided by the area:

$$W = p_m \pi a^2 \quad (2.19)$$

and therefore

$$p_o = \frac{3}{2} p_m = \frac{3W}{2\pi a^2} \quad (2.20)$$

Combining this with equation 2.14, the contact radius a can be eliminated to yield:

$$p_o = \left(\frac{6WE^{*2}}{\pi^3 R^2} \right)^{1/3} \quad (2.21)$$

There are stresses resulting from this contact which are relevant when considering damage of materials; of particular interest in this context are the radial tensile stress on the surface (which may result in cracking of a brittle material or fatigue if the stresses vary due to repeated movement of the contact) and the shear stress below the surface on the axis of loading (which may result in plastic deformation). The stress distributions are best described in cylindrical polar co-ordinates (r, θ, z) , with the surface being defined as the plane where $z = 0$ and the axis of loading being defined by $r = 0$.

For a material with Poisson's ratio ν , the radial stress in the surface outside the contact area is given by:

$$\frac{\sigma_r}{p_o} = \left(\frac{1 - 2\nu}{3} \right) \left(\frac{a}{r} \right)^2 \quad r \geq a \quad (2.22)$$

and its maximum value occurs at the edge of the contact ($r = a$).

On the axis of the contact ($r = 0$),

$$\frac{\sigma_r}{p_o} = \frac{\sigma_\theta}{p_o} = -(1+\nu) \left\{ 1 - \left(\frac{z}{a} \right) \tan^{-1} \left(\frac{a}{z} \right) \right\} + \frac{1}{2} \left\{ 1 + \left(\frac{z}{a} \right)^2 \right\}^{-1} \quad (2.23)$$

and

$$\frac{\sigma_z}{p_o} = - \left\{ 1 + \left(\frac{z}{a} \right)^2 \right\}^{-1} \quad (2.24)$$

The maximum shear stress at any depth is equal to half the greatest difference in the principal stresses (i.e., $\tau_1 = \frac{1}{2} |\sigma_z - \sigma_r|$). It can be seen that on the surface (i.e., when $z = 0$), $\frac{\tau_1}{p_o} = \left(\frac{1}{4} - \frac{\nu}{2} \right)$; for $\nu = 0.3$, a typical value for Poisson's ratio, the maximum shear stress is one tenth of the maximum contact pressure. However, the maximum shear stress increases with depth until (for $\nu = 0.3$) it reaches a maximum of $0.31p_o$ at a depth of $0.48a$ whereupon it decreases as z continues to increase. The value and position of the maximum shear stress are not particularly sensitive to the value of Poisson's ratio; across the range of values of ν from 0.2 to 0.35 (which covers a large number of metallic alloys, ceramics and some polymers), the maximum shear stress changes from $0.33p_o$ to $0.30p_o$ while its depth varies from $0.45a$ to $0.50a$.

Elastic line contact

A second geometry for which the details of the elastic contact mechanics are particularly useful is that of the *line contact*, applicable to two cylindrical surfaces in contact with their axes parallel to each other, or to a cylinder loaded against a plane and making contact along a line. In this case, the load is described in terms of a force per unit length $\left(\frac{W}{L} \right)$ and produces a rectangular contact patch of length L and width $2b$. The value of b is given by:

$$b = \left(\frac{W}{L} \frac{4R}{\pi E^*} \right)^{1/2} \quad (2.25)$$

The contact pressure $p(x)$ where x is the distance from the centre of the contact patch (along a line perpendicular to the cylinder axes) is given by:

$$p(x) = p_o \sqrt{1 - \frac{x^2}{b^2}} \quad (2.26)$$

The load supported by the contact is again the integral of the pressure over the area of the contact:

$$W = L p_o \int_{-b}^{+b} \sqrt{1 - \frac{x^2}{b^2}} dx = \frac{\pi}{2} L p_o b \quad (2.27)$$

The mean pressure acting across the contact area is also equal to the applied load divided by the area:

$$W = 2 p_m b L \quad (2.28)$$

and hence

$$p_o = \frac{4}{\pi} p_m = \frac{W}{L} \frac{2}{\pi b} \quad (2.29)$$

Combining this with equation 2.25, the contact semi-width b can be eliminated to yield:

$$p_o = \left(\frac{W}{L} \frac{E^*}{\pi R} \right)^{1/2} \quad (2.30)$$

As noted previously for the case of point contact, the stresses resulting from this contact are relevant when considering damage of materials. For the line contact, the maximum shear stress on the axis of loading is $0.30 p_o$ at a depth of $0.78b$. For a line contact, both the value of the maximum stress and its depth below the surface are independent of the value of Poisson's ratio.

2.5.2 Plastic deformation in normally loaded contacts

The onset of plastic deformation

As the normal load between the contacting bodies is increased, one or other of the components may start to deform plastically. Plastic flow first occurs at the point where the yield criterion is first satisfied; if the Tresca criterion is applicable (a reasonable approximation for many ductile metals which also affords arithmetical simplicity over the von Mises yield criterion), this will occur when a maximum shear stress of $Y/2$ is reached, where Y is the uniaxial yield stress of the material. As noted in Section 2.5.1, for a line contact, the maximum elastic shear stress occurs on the axis of loading at a depth $0.78b$ beneath the surface, and has a value of $0.30 p_o$. Plastic deformation therefore initiates when the maximum contact pressure is equal to $1.67Y$. By substituting this into equation 2.30 we can identify the load per unit length W/L at the onset of yield:

$$\frac{W}{L} = 2.78 Y^2 \frac{\pi R}{E^*} \quad (2.31)$$

For spherical point contact (on a material with $\nu=0.3$), the maximum shear stress occurs on the axis of loading at a depth $0.48a$ beneath the surface, and has a value of $0.31p_o$. Plastic deformation therefore initiates when the maximum contact pressure is equal to $1.61Y$. The load for the onset of yield is given by:

$$W = \frac{0.7\pi^3 Y^3 R^2}{E^{*2}} \quad (2.32)$$

Development of plasticity following initial yield

Initially, the region in which plastic flow occurs is small and is surrounded by material that is still elastically deformed, and therefore the plastic strains are limited by the elastic strains of the surrounding material. As the normal load is further increased, the zone of plastic deformation extends until it eventually reaches the free surface. This point may be shown (by finite element analysis) to occur in a metal at a load some 50–100 times that at which plastic flow first initiates. The mean pressure over the contact area has risen by this stage to about $3Y$ and remains at nearly the same value for subsequent increases in load. The independence of the mean contact pressure from the load once full plasticity has been reached, and the constant of proportionality between that pressure and the yield stress of the indented material, provide the basis for *indentation hardness testing* as discussed in [Appendix A](#).

Similar results are found for indentation by strong indenters of all shapes, in particular for spheres, cones and pyramids of different angles, and for flat-ended punches. The constant of proportionality between indentation pressure and yield stress depends on the geometry of the indenter, but is never far from a value of about three. The mean indentation pressure when the deformation is fully plastic is used as a measure of the indentation hardness of the material. A commonly-used geometry is that of the standardized Vickers indenter, a square-based pyramid of diamond with an included angle between its sides of 136° ; the *Vickers hardness* test is discussed further in [Appendix A](#). Vickers hardness is usually quoted in units of kilograms force per square millimetre ($1 \text{ kgf mm}^{-2} \cong 9.8 \text{ MPa}$), but with the symbol for the units omitted and with the number followed instead by the symbol HV (e.g., 100 HV).

For the case of a soft sphere or blunt cone pressed against a rigid plane, in which yield occurs in the sphere or cone rather than in the plane surface, similar results are obtained. Provided the extent of the deformation is not too large, plastic constraint again raises the mean contact pressure to about three times the uniaxial yield stress of the sphere or cone.

When we come to consider the plastic deformation of asperity contacts in [Section 2.5.5](#), we may therefore reasonably assume that when an asperity, of whatever shape, is pressed on to an opposing surface, it does not much matter which component yields; the mean pressure over the contact area will always be of the order of three times the uniaxial yield stress of the softer material. And importantly, the area of contact should be directly proportional to the load.

2.5.3 Elastic stress fields in contacts with additional tractional loading

Many real contacts experience not only normal loading, but also a component of tractional loading parallel to the surface. The final step of our analysis of the macroscopic elastic and plastic deformation of contacts addresses this. If the tractional load is high enough to exceed the frictional force, then sliding of the contact will occur. The analysis of the stress fields is complex (and outside the scope of this book), but a general appreciation of some concepts will assist in understanding material damage in such contacts. Since the contact is sliding, for any point in the contact it is assumed that the local tractional stress, $q(x)$, is given by the following:

$$q(x) = \mu p(x) \quad (2.33)$$

where μ is a constant coefficient of friction.

On the surface, tensile and compressive stresses exist in the direction of sliding, with compressive stresses in front of the contact, and tensile stresses behind. We need to consider the influence of these tensile stresses on surface damage due to either fracture or fatigue. The maximum tensile stress associated with the frictional traction occurs at the trailing edge of the contact. For a Hertzian point contact, the maximum tensile stress σ_x was first described by Hamilton as follows:

$$\sigma_x = p_o \left(\frac{1 - 2\nu}{3} + \frac{4 + \nu}{8} \pi \mu \right) \quad (2.34)$$

For a Hertzian line contact, the maximum tensile stress is given by:

$$\sigma_x = 2\mu p_o \quad (2.35)$$

Below the surface, tractional loads result in an increase in the value of the maximum shear stress, and a movement of the position of the maximum shear stress off the axis of normal loading and closer to the surface as shown in Fig. 2.9 for a line contact, which compares the stress distributions (a) with normal loading only and (b) with additional traction. For a Hertzian line contact, the point of first yield (according to the Tresca yield criterion) occurs on the surface itself if $\mu > 0.25$. In this case, if the shear stress is large enough to cause yield, the constraint provided by the surrounding elastic material (as observed when the maximum shear stress is sub-surface) no longer exists and plastic deformation of much larger magnitude may be expected, resulting in significant damage to the material.

2.5.4 Deformation of a single asperity

Having considered the macroscopic behaviour of ideally smooth contacts, we now consider that of real surfaces coming into contact. As already noted, real surfaces will contact each other only at a number of discrete points (high points on the surfaces), termed the *asperities*. Asperities are revealed by the study of surface profiles to be blunt; as we

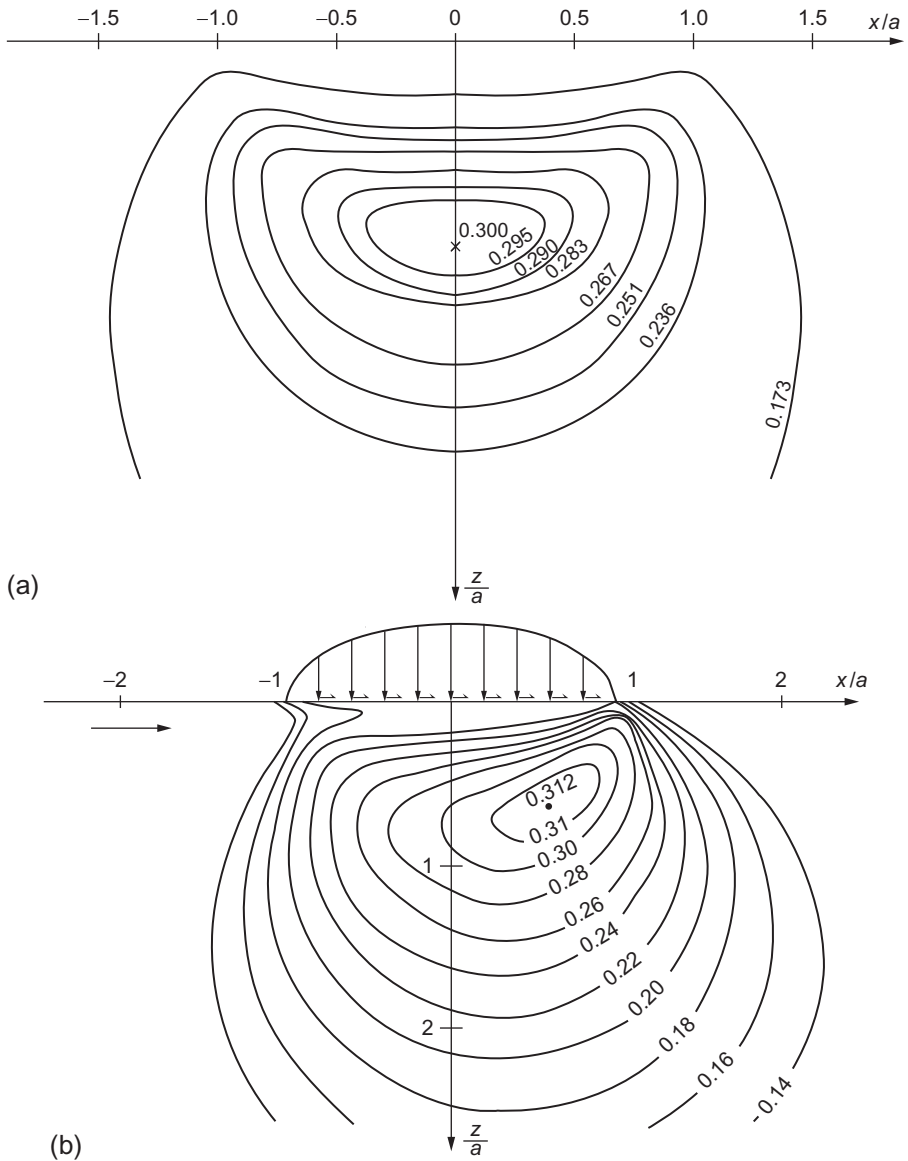


Fig. 2.9 Contours of principal shear stress in Hertzian line contacts: (a) under normal loading alone; and (b) in sliding with a coefficient of friction of 0.2. (after Johnson, K.L., 1985. *Contact Mechanics*. Cambridge University Press)

have seen, surface slopes are very seldom steeper than about ten degrees, and are usually much shallower. In addressing the behaviour of single asperities, it is convenient to model them as perfectly smooth protuberances of spherical shape. Our analysis of the elastic and plastic point contact of macroscopic surfaces can be applied to each asperity individually, noting that the load applied is the load on each asperity, w .

2.5.5 Simple theory of multiple asperity contact

We can use the concept of the asperity contact in an elementary treatment of the contact between two rough surfaces. Although the statistical approach discussed in the next section describes the real case more closely, a simpler model can nevertheless provide a valuable physical picture of the dependence of contact area on normal load.

From consideration of the plane of symmetry between two identical rough surfaces brought into contact, it is clear that the problem may be treated as one of contact between a single rough surface and a rigid frictionless plane. If we make the simplifying assumption that the rough surface consists of an array of spherical asperities of constant radius and height, and that each asperity deforms independently of all the others, then we can immediately apply the results of the previous section.

Each asperity will bear the same fraction of the total normal load, and each will contribute the same area to the total area of contact. By summing the contributions from all the asperities over the whole area of contact, we may show that the total real area of contact A will be related to the total load W in exactly the same way as the individual area of contact for each asperity (πa^2) is related to the load borne by each asperity (w). For the case of purely elastic contact,

$$A \propto W^{2/3} \quad (2.36)$$

and for perfectly plastic behaviour of the asperities,

$$A \propto W \quad (2.37)$$

Real surfaces, however, are not composed of uniform asperities of a single radius and height; both the radii and heights of the surface irregularities are statistically distributed. As the load on a real surface is increased, not only will the area of contact for each individual asperity increase, but more asperities will come into contact and start to carry some load. Under these circumstances, if the *average* area of contact for the contacting asperities remains constant, and the increase in load is borne by a correspondingly increased number of asperities in contact, then even for purely elastic contact, the total area will be directly proportional to the load. To examine whether this is a valid picture of the behaviour of real surfaces, we must look in more detail at the behaviour of a statistically distributed set of asperities.

2.5.6 Statistical theories of multiple asperity contact

One of the first statistical theories for the contact of rough surfaces was presented by Greenwood and Williamson in 1966, and is still widely cited. Although later theories

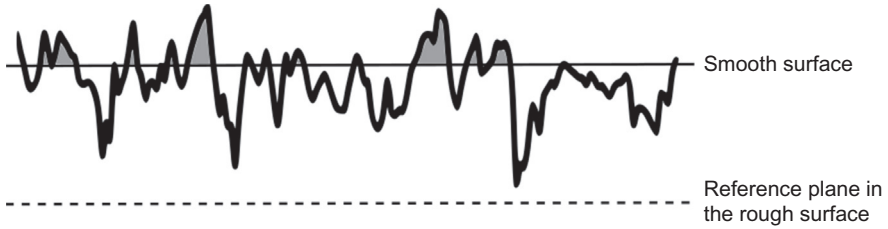


Fig. 2.10 Model for contact between a rough surface and a smooth rigid plane illustrating the assumptions made by Greenwood and Williamson

have progressively removed some of the simplifications made by Greenwood and Williamson, they support the broad conclusions of this model which we shall discuss here in some detail. Subsequent developments of the theory of contact of rough surfaces are discussed in some of the references listed at the end of this chapter, to which the reader is referred for further details.

In the Greenwood and Williamson model, it is assumed that all the contacting asperities have spherical surfaces of the same radius r , and that they will deform elastically under load according to Hertz's equations. Figure 2.10 illustrates the contact between the rough surface and a rigid smooth plane surface assumed in the model.

The height of an individual asperity above the reference plane is z . If the separation between the reference plane and the flat surface, d , is less than z , then the asperity will be elastically compressed and will support a load w which can be predicted from Hertz's theory:

$$w = \frac{4}{3} E^* r^{1/2} (z - d)^{3/2} \quad (2.38)$$

The heights of the asperities will be statistically distributed. The probability that a particular asperity has a height between z and $z + dz$ will be $\phi(z)dz$ where $\phi(z)$ is a probability density function describing the distribution of asperity heights.

It is important to distinguish between the distribution of surface heights, which we have met in Section 2.3.1 as the amplitude density function $p(z)$, and the distribution of asperity or peak heights, $\phi(z)$, the definition of which depends upon a criterion for identifying individual asperities. The method of identification employed by Greenwood and Williamson in their experimental work involved the comparison of each value of surface height, taken from a digitized profilometer trace, with the two neighbouring values, and therefore required the definition of a sampling length.

The probability that an asperity makes contact with the opposing plane surface is the probability that its height is greater than the plane separation, d :

$$\text{prob}(z > d) = \int_d^\infty \phi(z) dz \quad (2.39)$$

If there are a total of N asperities on the surface, then the expected number of contacts, n , will be given by:

$$n = N \int_d^{\infty} \phi(z) dz \quad (2.40)$$

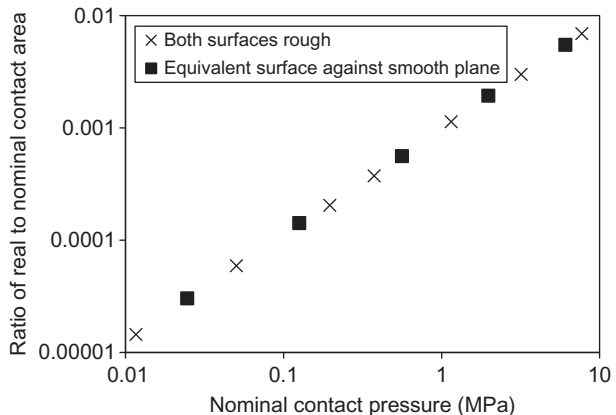
and therefore the total load carried by all the asperities, W , will be

$$W = \frac{4}{3} N E^* r^{1/2} \int_d^{\infty} (z-d)^{3/2} \phi(z) dz \quad (2.41)$$

Greenwood and Williamson studied the behaviour of their model for two different distributions of asperity heights. If the asperity heights follow an exponential distribution, then equation 2.41 may be integrated analytically. The load W is found to be linearly proportional to the total real area of contact, i.e., despite the *nominal* contact pressure increasing linearly with the applied load, the *mean* pressure on the *real* area of contact does not change with the applied load. As the load is increased, the size of each individual contact spot increases, but more asperities come into contact so that the mean size of the asperity contacts remains constant.

Although the exponential distribution provides a fair description of perhaps the highest tenth of all asperities on many real surfaces, the Gaussian (normal) distribution is found experimentally to provide a better model for the height distribution. For a normal distribution of asperity heights, equation 2.41 must be integrated numerically. If physically reasonable quantities are employed, then the results are not greatly different from those obtained with the exponential distribution. Figure 2.11 shows the ratio of the real contact area to the nominal contact area, as a function of the nominal pressure, for model surfaces with properties typical of a metal with hardness ~ 1 GPa. There is an almost linear relationship between the true area of contact and the nominal pressure over three orders of magnitude. Moreover, the results from the simple model of a rough plane

Fig. 2.11 Relationship between the nominal contact pressure and the ratio of the true and nominal contact areas (from Greenwood, J.A., Tripp, J.H., 1970. Proc. Inst. Mech. Eng. 185, 625–633)



surface loaded against a smooth plane are almost indistinguishable from those of a more complex model where two rough plane surfaces are loaded against each other.

The theory of Greenwood and Williamson was derived for purely elastic contact, but it does allow the onset of plastic flow at asperities to be predicted. It is found that the proportion of asperity contacts at which plastic flow has occurred depends on the value of a *plasticity index*, ψ , given by

$$\psi = \frac{E^*}{H} \left(\frac{\sigma^*}{r} \right)^{1/2} \quad (2.42)$$

where H is the indentation hardness of the softer surface (a measure of the plastic flow stress of the asperities) and σ^* is the standard deviation of the distribution of asperity heights. The quantity $(\sigma^*/r)^{1/2}$ is approximately equal to the average slope of the asperities, which can be estimated from profilometric data.

In principle, the proportion of asperity contacts that are plastic is determined by both the value of the plasticity index and the nominal pressure, but in practice the plasticity index dominates the behaviour. For values of ψ less than about 0.6, plastic flow at the asperities would be caused only by extremely high nominal pressures, while for ψ greater than about 1 most asperities will deform plastically under even the lightest of loads. In practice therefore, in using the plasticity index to predict the extent of plastic flow, the applied load on the contact does not need to be considered, since the behaviour is determined to the greatest extent by the surface profile and the mechanical properties of the surface.

Estimates for values of the plasticity index, ψ , can be derived from a knowledge of the mechanical properties of the contact and the surface profile data. Using the basic mechanical properties for three exemplar materials (a metal, a ceramic and a polymer) listed in Table 2.2, the values of (E^*/H) for the various possible contact combinations are presented in Table 2.3, along with the equivalent values of $(\sigma^*/r)^{1/2}$ for which $\psi = 0.6$, and for which the asperity contacts would therefore be elastic.

Smooth engineering surfaces typically have values of $(\sigma^*/r)^{1/2}$ between 0.005 and 0.1, and rough surfaces have values between 0.2 and 0.5. To put these in context, typical values of $(\sigma^*/r)^{1/2}$ are presented in Table 2.4. From the data in Tables 2.3 and 2.4, it can be seen that only the most finely polished surfaces will remain fully elastic in contact ($\psi < 0.6$), but that the use of ceramics and polymers can extend the range of values of $(\sigma^*/r)^{1/2}$ over which an elastic contact may result.

Table 2.2 Mechanical properties of materials covering the three broad classes

	E (GPa)	ν	H (MPa)
Steel	210	0.3	1500
Nylon	3	0.4	120
Silicon nitride	300	0.25	14500

Table 2.3 Values of (E^*/H) (left hand column) and $(\sigma^*/r)^{1/2}$ (right hand column) required to give a plasticity index of 0.6 for the six contact combinations possible for the materials listed in [Table 2.2](#)

	Nylon		Steel		Silicon nitride	
	$\left(\frac{E^*}{H}\right)$	$\left(\frac{\sigma^*}{r}\right)^{1/2}$	$\left(\frac{E^*}{H}\right)$	$\left(\frac{\sigma^*}{r}\right)^{1/2}$	$\left(\frac{E^*}{H}\right)$	$\left(\frac{\sigma^*}{r}\right)^{1/2}$
Nylon	14.9	0.040	29.3	0.020	29.4	0.020
Steel			76.9	0.008	89.4	0.007
Silicon nitride					11.0	0.054

Table 2.4 Typical values of σ^* , r and $(\sigma^*/r)^{1/2}$ (from T.H.C. Childs, Dry and boundary lubricated sliding friction and wear for engine component materials, in Engine Tribology (Ed. C.M. Taylor), Elsevier (1993) pages 51–74)

Material and condition	$\sigma^* / \mu\text{m}$	$r / \mu\text{m}$	$\left(\frac{\sigma^*}{r}\right)^{1/2}$
Bead blasted aluminium	1.37	13	0.32
Ground mild steel	0.6	15	0.20
Ground, lightly polished steel	0.5	85	0.077
Well-polished steel	0.1	500	0.014
Worn, run-in steel	0.065	250	0.016
Hard steel roller bearing	0.024	150	0.013

Theories for the contact between rough surfaces, and also experimental observations, thus suggest that in most practical cases of contact between metals, the majority of asperity contacts will be plastic. The load supported by each asperity is directly proportional to its contact area, and provided that the asperities deform independently, the total real area of contact for the whole surface will be proportional to the normal load and independent of the detailed statistical distribution of asperity heights. For ceramics and polymers, which lead to lower values of the ratio E^*/H , contact is more likely to be elastic. It may also be predominantly elastic for metals if their surfaces are very smooth. But even for elastic contact, as we have seen above, for realistic distributions of surface heights, the total real area of contact will remain roughly proportional to the normal load, since as the load is increased, the number of asperity contacts grows but the mean area of each remains effectively constant.

2.5.7 Contact of coated surfaces

In our discussion up to this point, we have considered contacts between homogeneous bodies. However, many components are surface-engineered by coating or surface modification to enhance their performance. Often, coatings are applied to create a surface with a high hardness to resist wear or surface damage, or to confer a low coefficient of friction in the contact. [Chapter 7](#) is devoted to the important topic of surface engineering.

Solutions for the elastic stress fields in a contact where one (or both) of the bodies has a surface coating are complex, generally requiring numerical methods (such as finite element modelling). However, if the coating is much thinner than the depth at which the maximum shear stress occurs (which depends upon the contact geometry and the loading conditions), then it is reasonable to neglect the presence of the surface coating and to use the Hertzian contact equations based upon the substrate material alone.

Coatings with high hardness are often also brittle, and therefore may fail by fracture. In sliding contacts, the high tensile stress at the trailing edge of the contact may cause brittle failure in a coating; failure by brittle fracture is often also associated with plastic deformation in the underlying substrate material. Another mechanism of failure in such systems is that of coating delamination from the substrate. While the maximum shear stress under normal indentation often occurs at a depth much greater than that of the coating-substrate interface, it must be remembered that under tractional loading (as observed in sliding contacts), the point of maximum shear stress moves towards the surface, and reaches it in many systems when the ratio of tractional to normal load is greater than about 0.25.

2.5.8 Surface energy contributions in elastic contact

Hertzian contact theory describes the elastic contact of many surfaces well, but for contacts where one of the materials has a low stiffness (such as a rubber), there are significant discrepancies between the predictions of this theory and experimental observations. In particular, the contact patch is always larger than that predicted (especially at low loads); as the contact is separated, a finite area of contact remains at zero load, and even persists under a tensile normal force. Several theories have been developed to account for this, of which one will be introduced here. The *Johnson-Kendall-Roberts (JKR) theory* adds an adhesive term between the surfaces, driven by their surface energies. This adhesive term increases the size of the contact at all loads, and accounts for the need for a negative force (the *pull-off force*) to separate the surfaces following contact. For a point contact (e.g., a sphere on a plane), the relationship between contact radius and applied load is now modified from that of equation 2.14 by the existence of the surface energies, as follows:

$$a^3 = \frac{3R}{4E^*} \left(W + 3 \Delta \gamma \pi R + \sqrt{6 \Delta \gamma \pi R W + (3 \Delta \gamma \pi R)^2} \right) \quad (2.43)$$

$\Delta\gamma$ is given by the following equation:

$$\Delta\gamma = \gamma_1 + \gamma_2 - \gamma_{12} \quad (2.44)$$

where γ_1 and γ_2 are the intrinsic surface energies (energy per unit area) of the two materials making up the contact and γ_{12} is the surface energy of the interface between the two materials.

The pull-off load, W_B required to separate the surfaces following contact is then given by

$$W_B = -\frac{3}{2}\pi\Delta\gamma R \quad (2.45)$$

Despite metals and ceramics having much greater surface energies than polymers and rubbers, JKR theory is only realistically applicable to contacts where at least one of the bodies is a low-modulus polymer or rubber. This is because real surfaces are rough and contact therefore occurs between asperities; using Greenwood and Williamson's concept of a distribution of asperity heights, it can be shown that there is competition between the tensile adhesive forces on the lower asperities and the elastic compressive forces associated with contact of the higher asperities. An adhesion parameter can be defined that compares the force needed to deform an asperity by an amount equal to the standard deviation of the asperity heights to the adhesive force for that asperity; if the adhesion parameter is less than unity then the adhesive force is significant. For metals (even with high surface energies), the Young's moduli are so large that a standard deviation of surface heights no greater than a few nanometres is required to bring the adhesion parameter into that range. However, for polymers and rubbers (with low Young's modulus), the adhesive effects predicted by the JKR model are readily observable.

It should be noted that in surface probe techniques such as AFM, the tip on the cantilever is so small that it is commonly in contact with only a single asperity at a time. Here, the multiple asperity situation (which makes the effects of adhesive force only observable only for materials with low modulus) no longer exists, and the effects of adhesion due to surface energy can be observed for all material types.

SOURCES AND FURTHER READING

- Bull, S.J., Berasetegui, E.G., Page, T.F., 2004. Modelling of the indentation response of coatings and surface treatments. *Wear* 256 (9-10), 857–866.
- Greenwood, J.A., Williamson, J.B.P., 1966. Contact of nominally flat rough surfaces. *Proc. Roy. Soc. Lond. A* 295, 300–319.
- Hamilton, G.M., 1983. Explicit equations for the stresses beneath a sliding spherical contact. *Proc. Inst. Mech. Eng. C J. Mech. Eng. Sci.* 197 (1), 53–59.
- Johnson, K.L., 1985. *Contact Mechanics*. Cambridge University Press.

-
- Lardner, T.J., Giovinazzo, R.J., Ritter, J.E., 1992. The effect of high modulus single and double layer coatings on contact stresses. *Philos. Mag. A.* 66 (3), 437–455.
- Leach, R.K., 2014. *Fundamental Principles of Engineering Nanometrology*, second ed. Elsevier.
- McCool, J.I., 1986. Comparison of models for the contact of rough surfaces. *Wear* 107, 37–60.
- Thomas, T.R., 1982. *Rough Surfaces*. Longman.
- Williams, J.A., 2005. *Engineering Tribology*. Cambridge University Press.

QUESTIONS FOR CHAPTER 2

Question 2.1

Figure Q2.1 shows a surface profile across a wear scar in stainless steel formed by a fretting contact. Estimate (i) the slope of the sides of the main trench and (ii) the aspect ratio of the main trench.

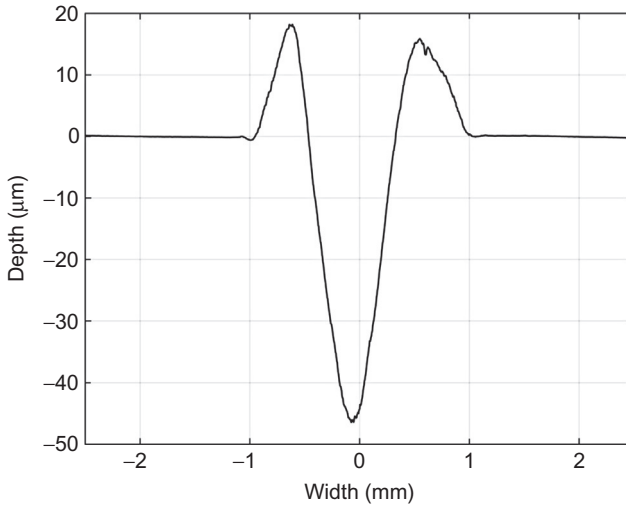


Fig. Q2.1 Profile across a wear scar in stainless steel formed by a fretting contact

Question 2.2

This exercise will enable you to see the effect of different scales of surface profile on measured values of Ra .

- (a) First construct a dataset which we will use as the profile (with a programme such as Matlab or by using a spreadsheet). The profile is made up of the superposition of two sinusoidal waves according to the following equation:

$$y(x) = A_1 \sin\left(\frac{2\pi x}{L_1}\right) + A_2 \sin\left(\frac{2\pi x}{L_2}\right)$$

where $y(x)$ is the surface height at a distance x from the origin, A_1 and A_2 are the amplitudes, and L_1 and L_2 are the wavelengths of the two waves.

Plot a graph of the surface profile from $x=0$ to 10 mm (with a spacing Δx between the data points of 0.001 mm) with the following parameters: $A_1 = 100 \mu\text{m}$; $A_2 = 5 \mu\text{m}$; $L_1 = 20 \text{ mm}$; $L_2 = 0.1 \text{ mm}$.

- (b) To calculate the roughness, you need to remove the form. One way of doing this is to select height data over a sampling length which is shorter than the wavelength of the form. Take the data for the first 1 mm of the profile (the ‘cutoff’), and fit a straight line through the data points (this represents the local equation of the

- form). Subtract this from the original data, leaving a dataset for which the mean value is zero. Using that dataset, calculate the sum of the height deviations from zero (remembering to use the absolute values), and hence calculate the Ra value.
- (c) Leaving the cutoff at 1 mm with form with an amplitude of 100 μm , calculate the Ra value when $L_1 = 3$ mm, when $L_1 = 6$ mm, and when $L_1 = 10$ mm. Comment on the sensitivity of the Ra value to the wavelength of the form.

Question 2.3

- (a) A hardened steel sphere of 6 mm radius lies on a thick plate of the same material and is loaded against it. The uniaxial yield strength of both materials is 1.5 GPa. Stating any assumptions that you make, calculate the applied load and the mean contact pressure at the onset of subsurface yield in the plate; also calculate the depth below the surface at which yield initiates (for both surfaces, $E = 210$ GPa and $\nu = 0.3$).
- (b) To avoid the possibility of yield, a decision is made to reduce the load so that the maximum stress is only half of the yield stress. Calculate the maximum load that can now be carried.

Question 2.4

Reliability issues with a cam and follower have resulted in a decision to attempt to reduce the maximum contact pressure to 80% of its current value by changing the geometry of the contact (all other parameters remain the same). The contact is currently made up of two cylinders with parallel axes (i.e., a line contact); one cylinder has a radius of 80 mm while the other has a radius of 20 mm. To what value would the radius of the *smaller* cylinder need to be changed to achieve the target contact pressure?

Question 2.5

An idealized surface profile has the form of a symmetrical square wave with an amplitude of 1 μm and a wavelength of 10 μm . With the same wavelength, calculate the amplitude of a triangular wave which has the same Ra value. Also, calculate the amplitude of a triangular wave which has the same Ra value but with a wavelength of 100 μm . Discuss the implications of these observations in terms of a surface profile often being described by just a single number.

For the triangular profile, use an Excel spreadsheet or other method to calculate the Rq and Rku values.

Question 2.6

A silicon nitride ceramic plate has a tensile strength (stress to fracture) of 800 MPa, a Young's modulus of 310 GPa, and a Poisson's ratio of 0.27. A 20 mm diameter silicon nitride ball, with the same mechanical properties, slides over the surface of the plate without rotation. Calculate the maximum normal load that could be applied to the ball without risk of tensile fracture at the trailing edge of the contact if the coefficient of friction was (a) 0.5; (b) 0.1; and (c) 0. Comment on the benefits of lubricating sliding contacts between ceramic components.

3.1 INTRODUCTION

We have seen in [Chapter 2](#) that when two solid surfaces are placed together, contact will generally occur only over isolated parts of the nominal contact area. It is through these localized regions of contact that forces are exerted between the two bodies, and it is these forces that are responsible for friction. In this chapter we shall examine the origins of the frictional force, and try to understand the magnitude of the frictional interactions between metals, polymers, ceramics and other materials. We shall also see how to estimate the temperature rise at a sliding contact that results from frictional energy dissipation.

3.2 DEFINITION OF FRICTION

The force known as friction may be defined as the resistance encountered by one body in moving over another. This broad definition embraces two important classes of relative motion: sliding and rolling. The distinction between sliding and rolling friction is useful, but the two are not mutually exclusive, and even apparently ‘pure’ rolling nearly always involves some sliding (see [Section 9.2.1](#)).

In both ideal rolling and sliding, as illustrated in [Fig. 3.1](#), a tangential force F is needed to move the upper body over the stationary counterface. The ratio between this frictional force and the normal load W is known as the *coefficient of friction*, and is usually denoted by the symbol μ :

$$\mu = F/W \quad (3.1)$$

The magnitude of the frictional force is conveniently described by the value of the coefficient of friction, which can vary over a wide range, from about 0.001 in a lightly loaded rolling bearing to greater than 10 for two identical clean metal surfaces sliding in vacuum. For most common materials sliding in air in the absence of a lubricant, however, the value of μ lies in the range from about 0.1 to 1.

3.3 THE LAWS OF FRICTION

Under some conditions of sliding, μ for a given pair of materials and fixed conditions (or absence) of lubrication may remain almost constant as the normal load and apparent area of contact are varied. This observation led to the formulation of two empirical *Laws of Sliding Friction*, often called after Guillaume Amontons who stated them in

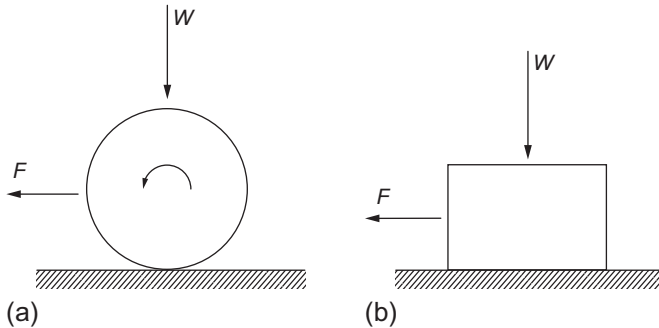


Fig. 3.1 A force, F , is needed to overcome friction and cause motion by (a) rolling or (b) sliding

1699. Leonardo da Vinci, however, had been the first to study and describe them some 200 years earlier; his first statement of the laws dates from 1493.

The Laws of Friction may be stated as follows:

1. The friction force is proportional to the normal load and
2. The friction force is independent of the apparent area of contact

To these is sometimes added a third law, often attributed to Coulomb (1785):

3. The friction force is independent of the sliding velocity

These three Laws of Friction are of varying reliability, but except in some important cases they do provide useful summaries of empirical observations.

The First Law, which may be expressed as

$$F = \mu \cdot W \quad (3.2)$$

amounts to the statement that the coefficient of friction, μ , is independent of the normal load. For many materials under conditions of lubricated or unlubricated sliding this is a fair approximation to what is observed, at least for macroscopic contact areas. Amontons, in the experiments from which his Laws were deduced, used several metals and wood all lubricated with pork fat; his materials therefore experienced *boundary lubrication* (see [Section 4.6](#)). The First Law is also often obeyed for unlubricated sliding. [Figure 3.2](#) shows typical results for the unlubricated sliding, in air, of steel on polished aluminium. The coefficient of friction remained effectively constant although the load was varied by a factor of nearly 10^6 .

Although most metals and many other materials obey the First Law well, polymers and other materials with a very low elastic modulus often do not. The reasons for this difference in behaviour are discussed in [Section 3.8](#).

The Second Law of Friction has not been so widely explored as the First, but it is nevertheless well attested for most materials, with the exception again of polymers. [Figure 3.3](#) shows the coefficient of friction for wooden sliders on an unlubricated steel surface. The normal load was held constant, while the apparent area of contact was varied by a factor of about 250; the value of μ varied very little.

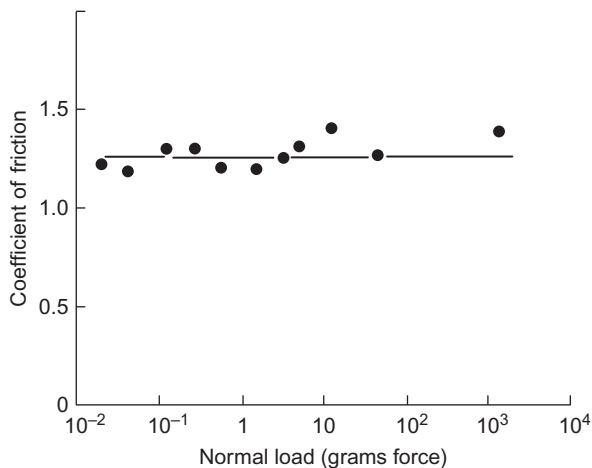


Fig. 3.2 The variation of the coefficient of friction, μ , with normal load, W , for the unlubricated sliding of steel on aluminium in air (from Bowden, F.P. Tabor, D., 1950. *The Friction and Lubrication of Solids*, Clarendon Press, Oxford, by permission of Oxford University Press)

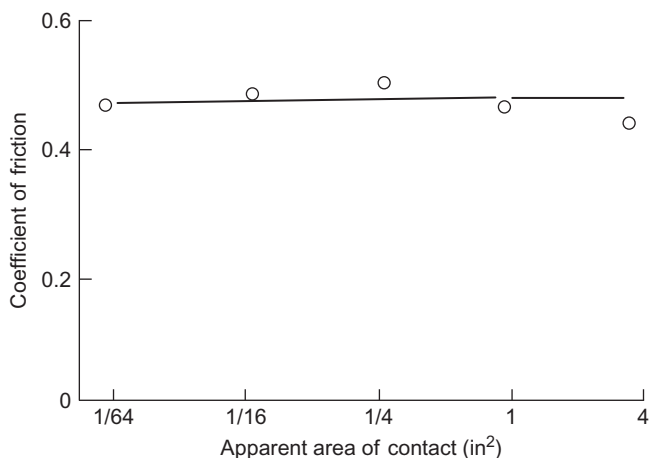


Fig. 3.3 The variation of the coefficient of friction, μ , with apparent area of contact for wooden sliders on an unlubricated steel surface (from Rabinowicz, E., 1995. *Friction and Wear of Materials*. John Wiley)

The Third Law of Friction is rather less well founded than the first two. It is a matter of common observation that the frictional force needed to initiate sliding is usually greater than that necessary to maintain it, and hence that the *coefficient of static friction* (μ_s) is greater than the *coefficient of dynamic (or kinetic) friction* (μ_d). But once sliding is established, μ_d is found for many systems to be nearly independent of sliding velocity over quite a wide range, although at high sliding speeds, of the order of tens or hundreds of metres per second for metals, μ_d falls with increasing velocity (see [Section 3.5.5](#)).

3.4 ORIGINS OF FRICTION

3.4.1 Introduction

Many early investigators, among them Amontons and Coulomb, envisaged that the major contribution to the frictional force arose from mechanical interaction between rigid or elastically deforming asperities.

Figure 3.4 illustrates a simple version of this model, often called the *Coulomb model*, in which the action of rigid wedge-shaped asperities causes the two surfaces to move apart as they slide from position A to position B. It may then be readily shown, by equating the work done by the frictional force to that done against the normal load, that μ is equal to $\tan \theta$. It is in considering the next phase of the motion, from B to C, however, that a fundamental defect of this model becomes apparent. Now the normal load does work on the system, and all the potential energy stored in the first phase of the motion (from A to B) is recovered. No net energy dissipation occurs in the complete cycle, and we must therefore conclude that no frictional force should be observable on a macroscopic scale if the interaction between real surfaces followed the Coulomb model exactly.

Some mechanism of energy dissipation is clearly essential in any satisfactory model for friction, which is a *non-conservative* force. As we shall see below, that mechanism may involve plastic deformation, sometimes in interfacial films rather than in the bulk material, distortion of interatomic bonds in the surface or interfacial material leading to the generation of lattice vibrations (phonons), or even the motion of electrons, while in polymers it often involves the molecular motions that are responsible for viscoelastic behaviour. We shall next consider a simple model for sliding friction. Although this was developed primarily for metals and is of limited applicability even for them, the central concepts, with some modification, are more widely applicable. The behaviour of ceramics, lamellar solids and polymers is discussed in particular detail in later sections.

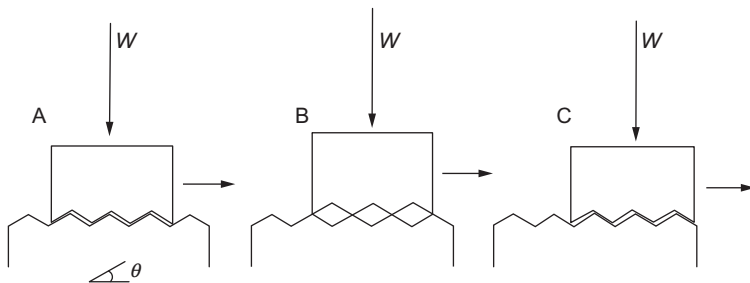


Fig. 3.4 A schematic diagram illustrating the principles behind the Coulomb model for sliding friction. The surface roughness is assumed to have a sawtooth geometry. As sliding occurs from position A to B work is done against the normal load W . The normal load then does an equal amount of work as the surfaces move from B to C

3.4.2 Classical model for sliding friction

The classical model for sliding friction of metals stems from the important work of Bowden and Tabor carried out, mainly in Cambridge, between the 1930s and the 1970s. The Bowden and Tabor model for sliding friction, in its simplest form, assumes that the frictional force arises from two sources: an *adhesion force* developed at the areas of real contact between the surfaces (the asperity junctions), and a *deformation force* needed to plough the asperities of the harder surface through the softer. Although in later developments of the theory it became clear that these two contributions cannot be treated as strictly independent, it is convenient and illuminating to consider them separately; the resultant frictional force F is then taken to be the sum of the two contributing terms, F_{adh} due to adhesion and F_{def} due to deformation.

The adhesion term arises from attractive forces that are assumed to operate at the asperity contacts. At first sight, this assumption may appear implausible; it is, after all, a matter of common experience that when two metal surfaces are pressed against each other they do not generally adhere. However, if the surfaces are truly clean, free from oxide and other surface films and from adsorbed gases, then significant adhesion is observed between metals; such conditions can be achieved under ultra-high vacuum (UHV) conditions (at gas pressures typically $<10^{-8}$ Pa). Strong adhesion, with adhesive forces sometimes greater than the load used to press the surfaces together, is seen under UHV for ductile metals such as copper and gold. In less ductile materials, for example metals with a hexagonal crystal structure and a small number of operating slip systems, or ceramics, the adhesion is found to be weaker.

In very soft and ductile metals such as lead and indium adhesion can readily be demonstrated under ordinary laboratory conditions (Fig. 3.5). If the rounded end of a brass or steel rod is degreased and abraded to remove some of the surface contamination, and then pressed on to the freshly-scraped surface of a block of indium, strong adhesion will occur. Furthermore, when the rod is detached fragments of indium adhere to the rod, showing that the adhesive forces at the junctions are stronger than the cohesive strength of the indium itself. Similar effects are seen in UHV experiments: if a clean iron surface is lightly pressed against a copper surface and then removed, examination of the iron surface reveals traces of copper.

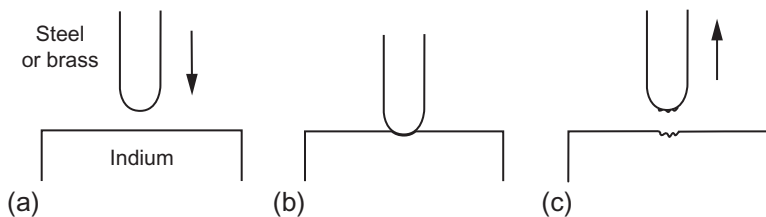


Fig. 3.5 An experiment to illustrate adhesion between metals. A clean steel or brass rod is pressed with a slight twisting motion on to the freshly-scraped surface of an indium block. Appreciable force is needed to detach the rod from the block, and fragments of indium adhere to the steel surface

Simulations of the interaction between dissimilar clean metals in contact, employing molecular dynamics (MD) models for the relevant interatomic forces, also demonstrate significant adhesion and transfer of material. Figure 3.6 shows the result of a molecular dynamics simulation of the contact between a nickel indenter and an initially plane gold surface. As the indenter approaches the surface, the force of adhesion is sufficient to cause distortion of the gold, so that intimate contact occurs when the nickel tip is still some distance above the original level of the surface. The sequence of images starts when the indenter is at the point of maximum indentation into the gold. When the indenter is raised from the surface the junction breaks within the gold and a significant number of gold atoms transfer to the nickel tip.

All this evidence points to the establishment of strong interfacial bonds across asperity junctions, at least for clean surfaces of ductile metals. The reasons why significant adhesion is not observed when two metal surfaces are placed together under ordinary circumstances are twofold: in the first place the surfaces are covered with oxide and adsorbed films that will weaken the adhesion, and second the elastic strains around the asperities when under load generate enough stress to break the asperity junctions during the unloading process, unless the metal is particularly ductile. Thus only in soft ductile metals, and where oxide films are at least partially removed, can appreciable adhesion be demonstrated under ordinary conditions.

The evidence of adhesion experiments and of theoretical modelling further suggests that when two dissimilar metals slide against each other the asperity junctions

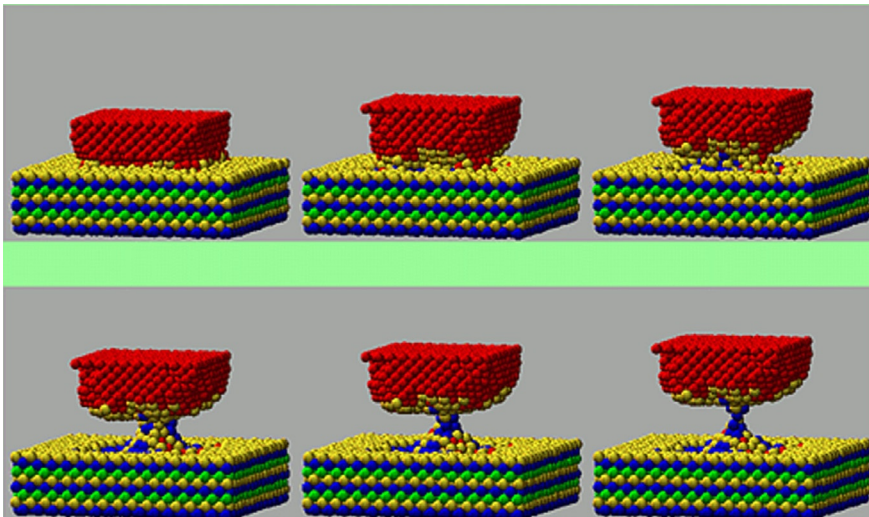


Fig. 3.6 Atomic configurations predicted by theoretical modelling of the contact between a nickel indenter (upper body) and an initially plane gold surface. No consideration is given in the model to adsorbed or reacted oxygen atoms, so that the system modelled corresponds to the contact of atomically clean metals in a perfect vacuum. The images from top left to bottom right show adhesion of gold atoms to the indenter, and progressive formation of a 'neck' of gold atoms drawn out as the indenter is raised from the surface (from Landman, U., Luedtke, W.D., Ringer, E.M., 1992. *Wear* 153, 3–30)

formed will in fact be stronger than the weaker of the two metals, leading to the plucking out and transfer of fragments of the softer metal on to the harder. This is indeed observed, and gives rise to severe wear of the softer metal (see [Chapter 5](#)).

If we denote the true area of contact, the sum of the cross-sectional areas of all the asperity junctions, by A , and assume that all junctions have the same shear strength s , then the friction force due to adhesion will be given by

$$F_{\text{adh}} = As \quad (3.3)$$

We saw in [Section 2.5.6](#) that to a first approximation, whether the contact between asperities is predominantly elastic or predominantly plastic, the real area of contact is almost linearly proportional to the applied normal load. For contact between metal surfaces finished by most conventional engineering processes, the initial asperity contacts will be effectively plastic (see [Section 2.5.6](#)), and we can write

$$W \approx AH \quad (3.4)$$

since the normal stress which the asperities are capable of supporting, if they deform plastically, will be close to the indentation hardness H of the softer material (see [Section 2.5.2](#) and [Appendix A](#)).

The contribution to the coefficient of friction from the adhesive forces is therefore

$$\mu_{\text{adh}} = F_{\text{adh}}/W \approx s/H \quad (3.5)$$

Since the asperity junctions fail by rupture within the weaker of the two materials, we can take s , to a first approximation, to be the shear strength of that material; H is the indentation hardness of the same material. For metals, the indentation hardness is about three times the uniaxial yield stress, as discussed in [Appendix A](#):

$$H \approx 3Y \quad (3.6)$$

The yield stress Y will be about 1.7 or 2 times the yield stress in pure shear, s , the precise factor depending on the yield criterion chosen. We therefore expect that

$$H \approx 5s \quad (3.7)$$

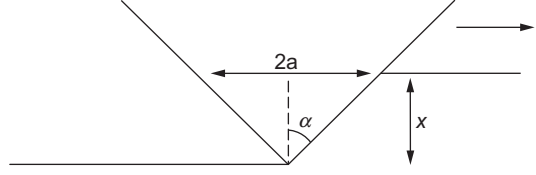
and thus that

$$\mu_{\text{adh}} \approx s/H \approx 0.2 \quad (3.8)$$

The frictional force due to the ploughing of harder asperities through the surface of a softer material, the *deformation* term, may be estimated by considering a simple asperity of idealized shape.

If a rigid conical asperity of semi-angle α (see [Fig. 3.7](#)) slides over a plane surface, the tangential force needed to displace it will be a flow pressure, which we may take as

Fig. 3.7 Model for the deformation component of friction, in which a conical asperity of semi-angle α indents and slides through the surface of a plastically deforming material



the indentation hardness H of the surface material, multiplied by the cross-sectional area of the groove:

$$F_{\text{def}} = Hax = Hx^2 \tan \alpha \quad (3.9)$$

The normal load supported by the asperity is given by

$$W = H\pi a^2/2 = H\pi x^2 \tan^2 \alpha \quad (3.10)$$

The coefficient of friction due to the ploughing term will therefore be

$$\mu_{\text{def}} = F_{\text{def}}/W = (2/\pi) \cot \alpha \quad (3.11)$$

A plane strain model, where the asperity is taken to be a wedge of semi-angle α , leads in a similar way to

$$\mu_{\text{def}} = \cot \alpha \quad (3.12)$$

These relationships are supported by the evidence of experiments in which macroscopic model asperities are dragged across softer metal surfaces.

The slopes of real surfaces are nearly always less than 10° (i.e., $\alpha > 80^\circ$), and therefore from equations 3.11 and 3.12, we would expect μ_{def} to be less than about 0.1.

We therefore conclude from our simple model that, even for a hard metal sliding on a softer one, the total coefficient of friction, representing the contributions from both ploughing and adhesion terms, should not exceed 0.3 or so. For a metal sliding against a counterface of the same material, where the contribution from ploughing should be negligible, μ should be slightly lower, of the order of 0.2.

If we look at the experimental measurements of μ for the unlubricated sliding of metals listed below in Table 3.1, we find that the measured values are in fact typically several times these estimates. The discrepancy is so large that other effects must be playing a role. Two effects dominate: *work-hardening* and *junction growth*.

In the simple model developed above, the material is assumed to have a constant flow stress. However, nearly all metals strain-harden to some extent, and although the normal load is supported by plastic flow some distance from the immediate vicinity of the asperity junctions, the junctions themselves will work-harden significantly, which will tend to raise the relative value of s in comparison with that of H . Work-hardening will therefore tend to increase μ_{adh} , although the phenomenon is difficult to quantify. Probably more important is the effect of junction growth, which will be considered in the next section.

Table 3.1 Typical values of the coefficient of static friction, μ_s , for combinations of metals in air and without lubrication. It must be appreciated that the value of μ_s depends on the precise conditions of the test, and these values should be taken as only broadly representative. Where two values are quoted, they refer to conditions under which the oxide films are intact, and in which they are penetrated (data from Bowden, F.P., Tabor, D., 1964. *The Friction and Lubrication of Solids, Part II*. Clarendon Press, Oxford; Tabor, D., 1973. In: Neale, M.J. (Ed.), *Tribology Handbook*. Butterworths)

	μ_s
<i>(a) Self-mated metals in air</i>	
Gold	2
Silver	0.8–1
Tin	1
Aluminium	0.8–1.2
Copper	0.7–1.4
Indium	2
Magnesium	0.5
Lead	1.5
Cadmium	0.5
Chromium	0.4
<i>(b) Pure metals and alloys sliding on steel (0.13% C) in air</i>	
Silver	0.5
Aluminium	0.5
Cadmium	0.4
Copper	0.8
Chromium	0.5
Indium	2
Lead	1.2
Copper–20% lead	0.2
Whitemetal (Sn based)	0.8
Whitemetal (Pb based)	0.5
α -Brass (Cu–30% Zn)	0.5
Leaded α/β brass (Cu–40% Zn)	0.2
Grey cast iron	0.4
Mild steel (0.13% C)	0.8

3.4.3 Junction growth

In the model just described, we have assumed that the true area of contact is determined solely by the normal load, and that it is unaffected by tangential forces. This is a considerable oversimplification. Whether the metal flows plastically or not is determined by a yield criterion, which takes account of both the normal and the shear stresses acting; a simple illustration will indicate that incorporation of a yield criterion into the model markedly affects its predictions.

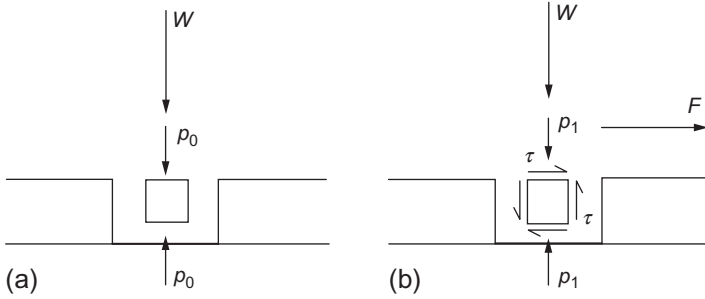


Fig. 3.8 Stresses acting on an element within an idealized asperity pressed against a counterface (a) with no tangential force, and (b) with a tangential force applied. The stress acting in the direction normal to the plane of the diagram is assumed to be zero

Figure 3.8 shows a region of material loaded against a rigid plane surface, representing in very idealized form an asperity contact. The element of material just inside the asperity at (a) is subjected to uniaxial compression by a normal stress p_0 , and we can assume it to be on the point of yielding, since we know that nearly all asperity contacts between metals are plastic. When a tangential stress is then applied to the asperity junction, as at (b), the element of material experiences an additional shear stress τ . For the material to remain at the point of yielding, the normal stress on the element must be reduced to a value p_1 . If the normal load remains constant, then the area of contact must grow: the phenomenon is therefore known as *junction growth*.

The relationship between p_0 , p_1 and τ is determined by the yield criterion. For Tresca's criterion, in which plastic flow occurs at a critical value of the maximum shear stress,

$$p_1^2 + 4\tau^2 = p_0^2 \quad (3.13)$$

For the von Mises yield criterion,

$$p_1^2 + 3\tau^2 = p_0^2 \quad (3.14)$$

Whether we use equation 3.13 or 3.14 does not much matter; both lead to the same qualitative conclusions. Let us examine these for equation 3.13. The normal and shear stresses are given by

$$p_1 = W/A \quad (3.15)$$

and

$$\tau = F/A \quad (3.16)$$

where A is the true area of contact. Note that F here denotes the tangential force, and does not necessarily imply that sliding is actually occurring. We can now substitute in equation 3.13 and obtain

$$W^2 + 4F^2 = A^2 p_o^2 \quad (3.17)$$

In a typical sliding experiment under dead-weight loading, W is constant, while p_o is a property of the material (its yield stress in compression). The real area of contact A will therefore increase with increasing tangential force, and the ratio F/W , the instantaneous value of μ , will also increase steadily.

There is nothing in this model to limit the growth process; in theory it could continue until the whole area of the specimen was actually in contact, and the coefficient of friction would reach a very high value. Under certain conditions junction growth in metals can indeed proceed this far (see Section 3.5.2), but in most practical cases it is limited by the ductility of the material, and by the presence of weak interfacial films. We can model the effect of a weak interface if we assume that it will fail at some shear stress τ_i , less than the shear strength of the bulk material. The maximum possible tangential force is then given by

$$F_{\max} = \tau_i A_{\max} \quad (3.18)$$

and the coefficient of friction is

$$\mu = F_{\max}/W \quad (3.19)$$

If the shear yield stress of the bulk asperity material is τ_o , then from the Tresca yield criterion

$$p_o = 2\tau_o \quad (3.20)$$

An expression for μ can now be derived:

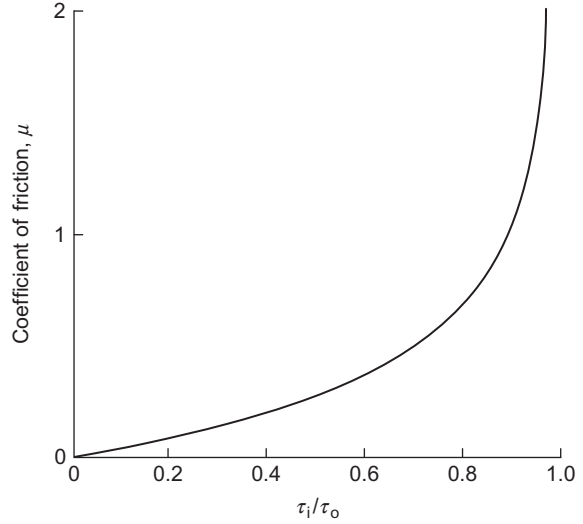
$$\mu = \frac{F_{\max}}{W} = \frac{1}{2 \left((\tau_o/\tau_i)^2 - 1 \right)^{1/2}} \quad (3.21)$$

When the interface has the same shear strength as the bulk material ($\tau_o/\tau_i = 1$), equation 3.21 shows that μ becomes infinite, since junction growth is unlimited. For weaker interfaces μ is finite, and drops rapidly as τ_i/τ_o falls. The dependence of μ on interfacial shear strength is shown in Fig. 3.9.

An interface only 10% weaker than the bulk is sufficient to reduce μ to about 1, while for an interface with half the bulk strength, μ falls to about 0.3. If the interface is very weak, then μ can have an extremely low value: for an interfacial strength one tenth of the bulk value, $\mu = 0.05$.

Weak interfacial films lead to very limited junction growth, and we would then expect essentially the same value of μ predicted from equation 3.21 as from a model

Fig. 3.9 The variation of coefficient of friction, μ , with the ratio between the shear strength of the interface and that of the bulk material (τ_i/τ_o). The function plotted is that given by equation 3.21



which takes no account of junction growth at all. Such a theory can be readily developed: if we assume the idealized asperity in Fig. 3.8 to be separated from the substrate by a weak film of shear strength τ_i , then the frictional force is determined by the shear strength of the film:

$$F = A\tau_i \quad (3.22)$$

and the normal load is supported by the plastic flow stress of the bulk of the asperity:

$$W = Ap_o \quad (3.23)$$

Hence we predict that

$$\mu = F/W = \tau_i/p_o \quad (3.24)$$

Equation 3.21 reduces to the same simple form if $\tau_i \ll \tau_o$ and equation 3.20 is used to relate p_o and τ_o .

Equation 3.24 is important, since it suggests a key way of reducing friction; if a film of material of low shear strength can be interposed between two surfaces, then the coefficient of friction can be lowered. This principle underlies the action of lubricants, discussed in Chapter 4, and is also exploited in the design of some journal bearing materials (Section 9.2.3).

3.4.4 Current views of friction

Some of the assumptions of the classical model discussed above cannot be universally valid, since as we shall see in discussing sliding wear in Chapter 5, significant plastic deformation of asperities and transfer of metal from one surface to another occurs only

under certain conditions and would give rise to an intolerably high rate of wear for many applications. Under many practical conditions, surfaces can slide against each other with much lower wear and yet with significant friction: in these cases plastic deformation of asperities cannot account for the frictional energy dissipation that is observed. While the concept of a real area of contact that is proportional to normal load is still valid, we must look for mechanisms that can provide dissipative processes without involving dislocation motion at the points of contact between asperities.

If two idealized crystalline solids slide over each other as shown in Fig. 3.10(a), the interaction energy between the contacting planes of atoms will be periodic in sliding distance, and if the atoms were rigidly bonded to each other within each body then we would expect there to be no frictional dissipation, for exactly the same reasons we discussed earlier in Section 3.4.1 in dismissing Coulomb's model for friction. In fact, however, the atoms are not rigidly bonded and those close to the interface can vibrate about their equilibrium positions during sliding, giving rise to elastic waves that propagate outwards and carry energy away from the contact. The generation of these lattice disturbances or *phonons* provides one mechanism to account for frictional forces. However, the contact shown in Fig. 3.10(a) is *commensurate*, meaning that the spacings of the atoms in one surface in the direction of sliding is the same as that in the other. Even a small rotation of a three-dimensional crystal, or choice of a different crystal plane, in one of the sliding bodies will result in an *incommensurate* interface as shown in Fig. 3.10(b). Here the variation in energy and interaction force with sliding distance is very much less, and it becomes more difficult to invoke a significant phononic contribution to friction. Static friction between atomically clean surfaces is also difficult to explain by a phononic mechanism. But real surfaces are not clean, and the presence of even small numbers of mobile atoms or molecules such as water or short-chain hydrocarbons at the interface can considerably change the picture, as shown schematically in Fig. 3.10(c). The existence of surface contamination bonded to the surfaces can account not only for dynamic friction that is independent of the commensurability of the interface, but also for static friction.

The origins of friction at the microscopic scale remain the subject of active research, based on molecular dynamics simulation of increasingly complex structures

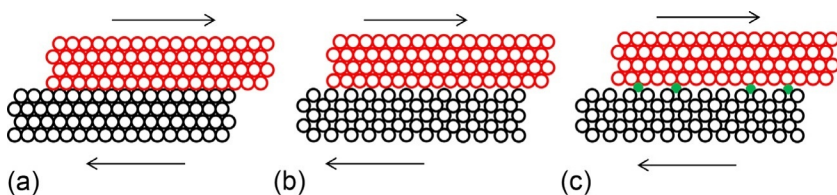


Fig. 3.10 Schematic diagrams showing (a) two crystalline solids sliding over each other with perfect lattice matching (a *commensurate* contact); (b) two crystalline solids with different spacing of surface atoms (an *incommensurate* contact); and (c) an incommensurate contact with contaminant atoms at the interface

as well as experimental measurements of tangential forces in very small-scale contacts. However, quantitative predictive modelling of the macroscopic value of friction force under unlubricated conditions remains elusive.

3.5 FRICTION OF METALS

3.5.1 Introduction

[Table 3.1](#) lists values of the coefficient of static friction μ_s taken from the literature for several combinations of metals in air and without lubrication. As we noted in [Section 3.3](#), the coefficient of dynamic friction μ_d , applicable once sliding is established, would be expected to be somewhat lower.

It must be borne in mind in all quantitative discussion of friction that the precise value of μ depends critically on the experimental conditions under which it is measured; the usefulness of the figures in [Table 3.1](#) lies more in their relative magnitudes than in their absolute values. In some cases, as shown by the examples below, to quote a single value for μ is very misleading, since the coefficient depends so strongly on the conditions. We shall discuss some of the underlying reasons for the values of μ listed in [Table 3.1](#) in the following sections.

3.5.2 Clean metals in high vacuum

If metal surfaces are cleaned in high vacuum and then placed into contact, strong adhesion is usually observed. The surfaces do not need to be atomically clean for them to adhere: merely heating them in vacuum to drive off adsorbed gases is often sufficient. The coefficient of friction under these conditions has a very high value, typically 2 to 10 or even more, and gross seizure frequently follows when sliding is attempted. Strong metallic bonds form across the interface, and when the surfaces are pulled apart metal is transferred from one body to the other. With little or no interfacial contamination present, the extent of junction growth is limited only by the ductility of the asperity material. The coefficient of friction is therefore very high. Sliding under ultra-high vacuum conditions can occur in space engineering, and special measures must be taken in designing sliding components to operate in this demanding environment. Solid lubricants and thin soft metallic films (see [Section 4.7](#)) can provide valuable protection.

3.5.3 Self-mated metals in air

In many practical applications, metals slide against one another in air. Coefficients of friction are then much lower than in vacuum, and typically, for unlubricated sliding, lie in the range from 0.5 to 1.5. [Table 3.1\(a\)](#) lists some representative values of μ , although it must be remembered that frictional behaviour depends, sometimes markedly, on the composition and microstructure of the materials and on the conditions of measurement.

Gold has a particularly high value of μ among the metals listed. It forms no oxide film although adsorbed gases will be present in air, and the asperity junctions therefore tend to be strong. Gold is also very ductile, so that considerable junction growth can occur. Both these factors contribute to a high frictional force, although not so high as that found in high vacuum.

All other metals oxidize in air to some extent, forming oxide films typically between 1 and 10 nm thick within a few minutes exposure of an atomically clean surface. These films play a critical role in determining the sliding behaviour, since friction between oxide surfaces, or between oxide and bare metal, is almost always less than between surfaces of bare metal.

Figure 3.11 illustrates the effect of oxygen on the sliding friction of pure iron. In high vacuum, strong adhesion and seizure occur. The admission of only a small pressure of oxygen allows sliding, although with a high coefficient of friction. As more oxygen is admitted to the system the value of μ drops, until it eventually reaches the value normally measured in air at ambient pressure.

The effect of an oxide film in mitigating friction can be destroyed to some extent, however, if the film is penetrated during sliding. Figure 3.12 shows a common form of behaviour, here seen in copper. At low normal loads, the oxide films effectively separate the two metal surfaces and there is little or no true metallic contact. The electrical resistance of the interface is high, and the track formed by the slider appears smooth and polished. The coefficient of friction is low, perhaps because the oxide acts as a low shear strength film, but more probably because its low ductility limits junction

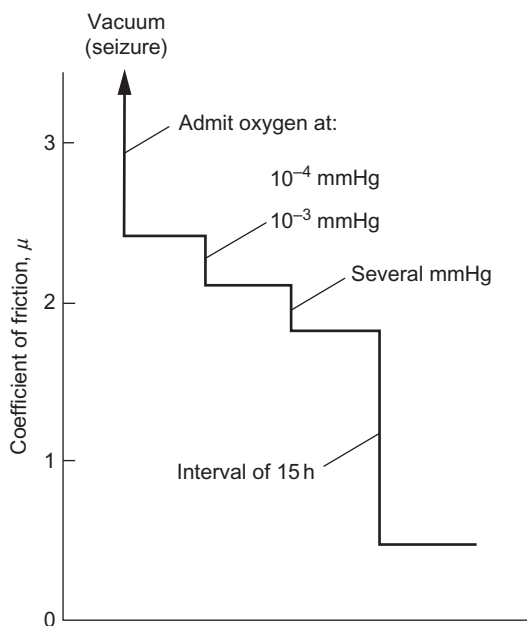
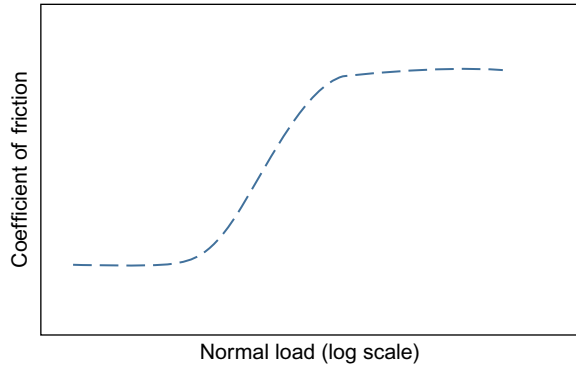


Fig. 3.11 The effect of oxygen on the friction of pure iron. In vacuum, the value of μ is very high; as the oxygen concentration is increased, so the coefficient of friction falls (from Buckley, D.H., 1981. *Surface Effects in Adhesion, Friction, Wear and Lubrication*. Elsevier)

Fig. 3.12 Schematic diagram showing how the coefficient of friction varies with applied normal load for copper sliding against copper in air, unlubricated



growth. As the normal load is increased, a transition occurs to a higher value of μ : in the case of copper in these experiments from less than 0.5 to greater than 1.5. The track of the slider now exhibits considerable surface damage with evidence of metallic transfer, the electrical resistance of the contact falls to a low value, and we deduce that metallic contact is occurring between the copper asperities.

Transitions of this kind are common in other metals also, although the change in μ may not be so great as in copper. In aluminium, for example, as listed in [Table 3.1\(a\)](#), μ increases from 0.8 to 1.2 as the oxide film breaks down. Cadmium shows almost no change in μ , although electrical resistance measurements confirm that an insulating film is present at low loads which is broken down at higher loads. In all these cases, when the surfaces are separated by protective oxide films the friction is essentially being measured between oxide surfaces. At higher loads the surface films may deform and fracture allowing true metallic contact to occur, and the friction often (but not always) rises. As might be expected, the wear rates associated with these two regimes are markedly different, and are further discussed in [Chapter 5](#).

For some very soft and ductile metals, notably tin and indium, metallic contact between the surfaces occurs even at the lightest loads, and the coefficient of friction is consequently high and does not change markedly with load. Here the oxide films, although formed, are penetrated easily since the soft substrate offers little mechanical support. At the other extreme of behaviour, chromium forms a thin but very strong oxide film, and exhibits over a wide range of load no metallic contact (as determined by electrical resistance measurements) and a constant low coefficient of friction.

In summary, the friction of pure metals sliding against themselves in air, unlubricated, is often controlled by the presence of surface oxides. If the oxide film is not broken during sliding, surface damage is slight and the oxide itself determines the coefficient of friction. The value of μ is often (but not always) lower than at higher loads where the oxide film is penetrated and metallic contact between asperities occurs. Under these conditions considerable surface damage and rapid wear occur. In certain metals no transition occurs because the protective oxide coating is retained over a wide range of load, while in others where the oxide is penetrated even at very light pressure the contact is always metallic.

3.5.4 Dissimilar metals and alloys

Oxide films are also important in the friction of dissimilar metals and alloys in air. In general, the coefficient of friction for an alloy tends to be rather less than that for its components as pure metals.

The sliding friction of steels has been widely studied; μ varies with both composition and microstructure, and also often depends on load. Figure 3.13 shows typical behaviour.

The solid curve, for a 0.4% carbon steel, shows a transition from a relatively high value of μ at low loads to a lower value at high loads. It has been suggested that the transition results from the stratified structure of the oxides present on the steel. The uppermost layer is Fe_2O_3 , while beneath this lie layers of Fe_3O_4 and finally FeO next to the metal itself. According to this model, the transition then results from penetration of the Fe_2O_3 layer at the higher loads. Some other steels do not show such a pronounced transition: the broken line in Fig. 3.13 (for a steel of lower carbon content) illustrates commonly observed behaviour.

The effects of minor constituents on the frictional behaviour of alloys can be marked, as a result of surface segregation. Aluminium in steels, for example, segregates to the surface and in vacuum tends to increase the coefficient of friction. Oxidation of the surface, however, produces a layer of aluminium oxide which promotes a reduction in friction. The effects of surface segregation are complex, but can provide a way of changing the frictional properties of an alloy through relatively small modifications to its composition.

Table 3.1(b) lists representative values of μ for various alloys sliding against a low-carbon steel. Some alloys, notably leaded α/β brass (copper–zinc), grey cast iron and copper–lead, stand out as showing particularly low coefficients of friction. The reason in all three cases is the same: all contain phases which provide thin interfacial

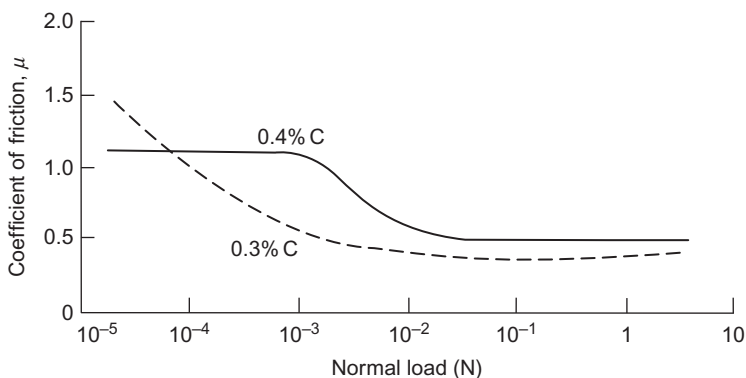


Fig. 3.13 The variation of coefficient of friction with normal load for steels sliding against themselves in air, unlubricated. Results are shown for two different plain carbon steels, with carbon contents of 0.4% (solid line) and 0.3% (broken line) (from Bowden, F.P., Tabor, D., 1964. *The Friction and Lubrication of Solids*, Part II. Clarendon Press, Oxford, by permission of Oxford University Press)

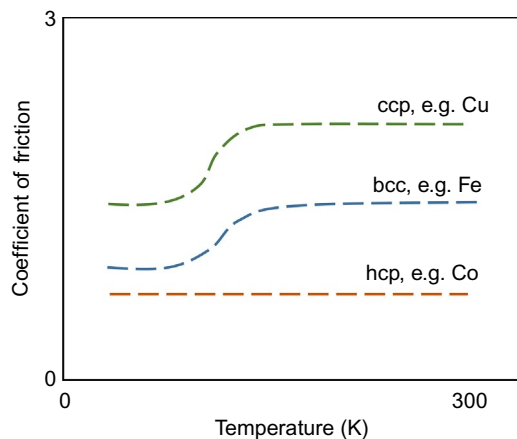
films of low shear strength. Lead brass and copper–lead both contain dispersions of metallic lead, which has very little solubility in the matrix phases. On sliding, a thin weak film of lead is formed on the surface and results in a low value of μ by the mechanism expressed in equation 3.24. In grey cast iron the low shear strength film is provided by the graphite constituent (see Section 3.7). These alloys with intrinsically low coefficients of friction in dry sliding against steel, which do not depend on the formation of a protective oxide layer, have wide engineering use: copper–lead and other copper-based alloys containing lead or bismuth are used in journal bearings (see Section 9.2.2), while the good tribological properties of grey cast iron add to its attractions of low cost and high damping capacity for applications such as heavy machine tool slideways. It is worth noting that the traditional bearing alloys, the tin- and lead-based whitemetals, do not have intrinsically low coefficients of friction against steel (Table 3.1(b)). Other factors, as discussed in Section 9.2.2, determine their suitability for this purpose.

3.5.5 Effect of temperature

When the temperature of a sliding metal is increased, several effects will occur: its mechanical properties will change, the rate of oxidation and other chemical reactions at the surface will increase, and phase transformations may take place. All these will influence its frictional behaviour.

Figure 3.14 shows schematically the variation of μ in ultra-high vacuum for metals of the three common crystal structures sliding against themselves, and illustrates how the influence of temperature on their plastic deformation behaviour changes the coefficient of friction. For both the cubic close-packed (ccp) and body-centred cubic (bcc) metals, transitions in friction occur. In the ccp metals these are associated with a marked change in the work-hardening rate (which is higher at low temperatures), while in the bcc metals the transition coincides with the ductile-brittle transition.

Fig. 3.14 Schematic diagram showing how the coefficient of friction varies with temperature, for pure metals with three different crystal structures sliding against themselves in ultra-high vacuum (Sketch showing behaviour reported by Bowden, F.P., Childs, T.H.C., 1966. Proc. Roy. Soc. Lond. A312, 451–466)



The hexagonal close-packed (hcp) metals show no change in friction with temperature over this range, since their mechanical properties do not alter significantly, but there is a clear correlation between the ductilities of the individual metals and their values of μ . Titanium and zirconium, being fairly ductile, resemble the ccp metals in their frictional behaviour, whereas beryllium and cobalt, with their ductility limited by the small number of slip systems operating at these low temperatures, exhibit lower values of μ . In these experiments, where the influence of interfacial films was deliberately excluded by performing them in vacuum, it is clear that μ was largely determined by the ductility of the metal at the points of contact. In general, as the ductility of a metal increases, so does the value of μ .

When metals are heated in air, their rate of oxidation will increase and so the thickness, and possibly also the nature, of the oxide films will change. This provides a second mechanism through which friction can vary with temperature.

Figure 3.15 illustrates this effect, for the sliding of an austenitic stainless steel against pure nickel. As the temperature is increased, the rising ductility of the metals leads to a steady rise in μ , until suddenly at 750°C a thick film of nickel oxide forms on the nickel and the friction drops sharply to a lower value. On cooling, the oxide continues to separate the asperities, and the low friction is maintained down to much lower temperatures. Similar behaviour is commonly observed in steels. A surface layer of Fe_3O_4 provides a lower coefficient of friction than Fe_2O_3 ; the predominant oxide which forms is determined by the temperature and by the compositions of the alloy and of the gas phase. Other surface films formed by gas-solid reactions can change the frictional behaviour in suitable systems: chlorine, for example, reacts strongly with many metals and the resultant chloride films can reduce friction.

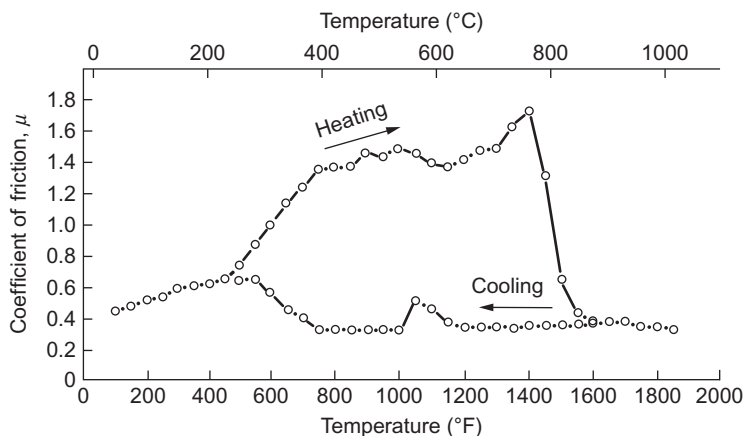


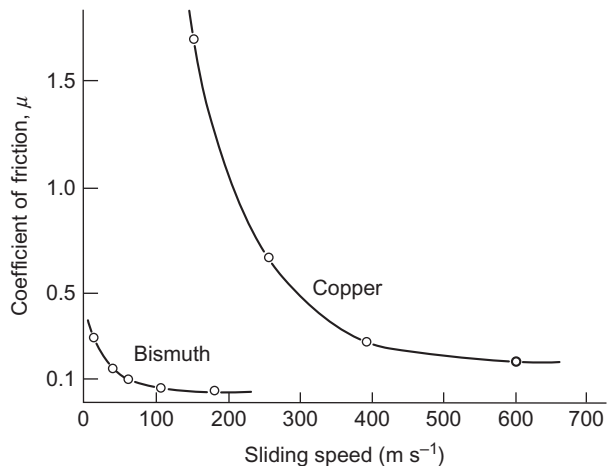
Fig. 3.15 The dependence of coefficient of friction on temperature for an austenitic stainless steel sliding against pure nickel in air. The two curves relate to data gathered while the specimens were being heated (upper curve) and subsequently cooled (lower curve) (from Peterson, M.B., Florek, J.J., Lee, R.E., 1960. Trans. Am. Soc. Lubr. Eng. 3, 101–115)

Phase transformations, through their influence on the mechanical properties of the materials, can result in large changes in friction. The most drastic effect is that due to melting; as a metal approaches its melting point its strength drops rapidly, and thermal diffusion and creep phenomena become more important. The resulting increased adhesion and ductility at the points of contact lead to a pronounced increase in friction. When one of the sliding surfaces actually becomes molten, however, and therefore loses its shear strength, the friction force drops to a low value determined by viscous forces in the liquid layer. This occurs in the sliding of metals at very high speeds (typically $> 100 \text{ m s}^{-1}$: see Fig. 3.16) and also in the sliding of a ski over ice or snow. In both these cases, the dissipation of frictional work generates enough local heat to raise the temperature at the interface to the melting point, and sliding then takes place under conditions of effectively hydrodynamic lubrication (see Section 4.4). Models for frictional heating are discussed in Section 3.9. At low sliding speeds, or at sufficiently low ambient temperatures, a molten film does not form and the friction is controlled by the interaction of the solid surfaces.

Solid state phase transformations also influence friction. Figure 3.17 shows the variation of the friction of cobalt with temperature. Cobalt exhibits a rather sluggish transformation (with an equilibrium temperature of 417°C) from its low-temperature hexagonal close-packed structure, with limited slip ductility, to a cubic close-packed structure which is fully ductile. The coefficient of friction associated with the ccp structure is high, whereas that of the hcp is low. The change in frictional behaviour is first observed at an ambient temperature below the equilibrium temperature for the transformation because frictional heating raises the local temperature at the contact.

Tin is another metal exhibiting a polymorphic transformation which leads to a change in friction, as seen in Fig. 3.18. Below the transformation temperature of 13°C grey tin (cubic, with the diamond structure) is the stable allotrope, while above, white tin (body-centred tetragonal) is stable. The change in μ shown in Fig. 3.18 is fully reversible with temperature. The addition of a small amount of bismuth to tin

Fig. 3.16 The variation of coefficient of friction with sliding speed for pure bismuth and copper sliding against themselves (from Bowden, F.P., Tabor, D., 1964. *The Friction and Lubrication of Solids, Part II*. Clarendon Press, Oxford, by permission of Oxford University Press)



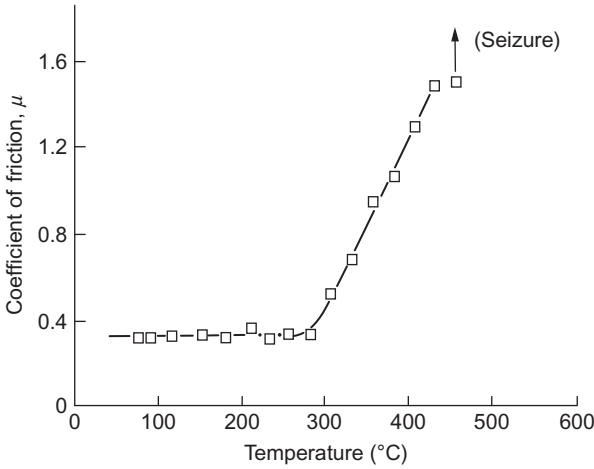


Fig. 3.17 The variation of coefficient of friction with temperature for cobalt sliding against itself in vacuum (from Buckley, D.H., 1981. *Surface Effects in Adhesion, Friction, Wear and Lubrication*, Elsevier)

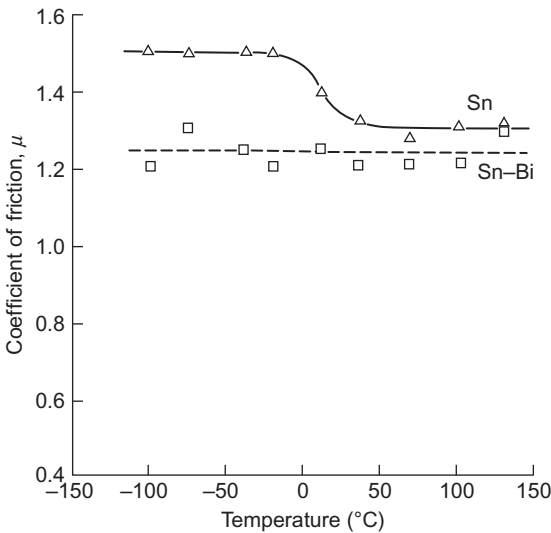


Fig. 3.18 The variation of coefficient of friction with temperature for tin sliding against itself (solid curve). The addition of a small amount of bismuth suppresses the allotropic transformation to grey tin (broken curve) (from Buckley, D.H., 1981. *Surface Effects in Adhesion, Friction, Wear and Lubrication*, Elsevier)

suppresses the transformation from the white to the grey allotope, and the coefficient of friction for the tin–bismuth alloy therefore does not change with temperature.

Another class of phase transformation affecting friction in metals is the order–disorder transformation, which occurs in many alloy systems. The copper–gold alloy of composition Cu_3Au has been investigated (Fig. 3.19). At low temperatures Cu_3Au has a long-range ordered cubic structure but an order–disorder transformation occurs at 390°C . The mechanical properties change markedly with the transformation: both the elastic modulus and the hardness increase on ordering, and as can be seen from Fig. 3.19, there is also a pronounced decrease in friction.

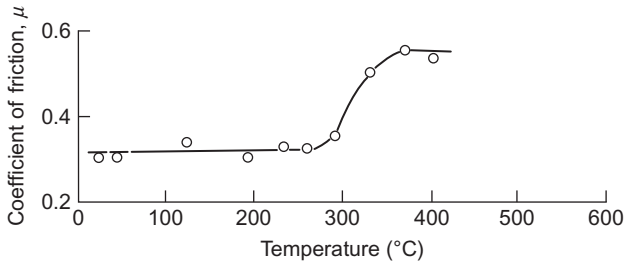


Fig. 3.19 The variation of coefficient of friction with temperature (in vacuum) for the alloy Cu_3Au which exhibits an order-disorder transformation at 390°C (from Buckley, D.H., 1981. *Surface Effects in Adhesion, Friction, Wear and Lubrication*. Elsevier)

3.6 FRICTION OF CERAMIC MATERIALS

The ceramic materials of most interest to the tribologist are the so-called *engineering ceramics*, which combine low density with excellent mechanical properties (e.g., high strength, hardness and stiffness) up to high temperatures. Typical examples of bulk engineering ceramics are silicon nitride (Si_3N_4), silicon carbide (SiC), alumina (Al_2O_3) and zirconia (cubic and/or tetragonal ZrO_2). Although these materials may be nominally pure, they usually contain additives, which may be present either in small amounts to assist the fabrication process (e.g., as sintering aids), or in more substantial proportions as alloying additions. Ceramic materials are also widely applied as thick or thin coatings on substrates of different materials by various surface engineering processes, as discussed in [Chapter 7](#); ceramics used for thin coatings (typically a few micrometres thick) include titanium nitride (TiN) and other transition metal nitrides and carbides, as well as carbon in various forms, including diamond and diamond-like carbon (DLC).

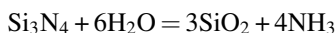
The major differences in mechanical behaviour between the engineering ceramics and metals arise from the different nature of the interatomic forces: ionic or covalent in ceramics, rather than metallic bonding. Ionic bonding, in ceramics such as MgO or Al_2O_3 , leads to crystal structures with only a small number of independent slip systems available for dislocations, fewer than the five necessary to accommodate a general plastic strain. Covalent bonding, as in SiC , TiC or diamond, leads to very narrow dislocations which move only under high stress, even though five independent slip systems may exist.

Ceramic materials of either bond type thus show only limited plastic flow at room temperature, and correspondingly, much less ductility than metals. Also, as we saw in [Section 2.5.6](#), asperity contact between ceramics is more likely to be elastic than in metals. The large plastic strains associated with asperity contact in metals do not therefore occur in ceramics except at high temperature, and although adhesive forces (of covalent, ionic or van der Waals origin) are present between ceramic materials in contact, the coefficient of friction never reaches the very high values seen in clean

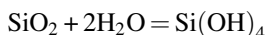
metals sliding in the absence of oxygen (see [Section 3.5.2](#)). The value of μ for ceramic-ceramic contacts lies typically in the range from 0.25 to 0.8, although some lower values will be discussed below. These are similar to the values seen for metallic couples sliding in air in the presence of intact oxide films, and indeed, there is much similarity between the contact of oxidized metal surfaces and that of bulk oxide ceramics.

Wide variability is typically seen in reported values of μ for engineering ceramics, and it would therefore be misleading to state even representative coefficients of friction for particular materials. Environmental factors are responsible for a good part of this variation. Despite their reputation for chemical inertness, the surfaces of most ceramics are susceptible to tribochemical reactions which lead to the formation of surface films and thus modify their frictional behaviour. These reactions occur much more rapidly at a sliding contact than on a free surface at the same bulk temperature. Local high ‘flash’ temperatures at the asperity contacts (discussed further in [Section 3.9](#)), the exposure of atomically clean surfaces by the wear process, and direct mechanical stimulation of the reaction (e.g., in the highly strained region at an asperity junction or crack tip) are all mechanisms that can accelerate surface reactions at sliding contacts.

Non-oxide ceramics in air commonly form oxide films on their sliding surfaces. For example, silicon nitride, silicon carbide, titanium nitride and titanium carbide have all been found after sliding to have reacted significantly with oxygen, derived from oxygen in the air or from water vapour. Oxide ceramics will react with water, whether it is present as liquid or as vapour: alumina and zirconia, for example, both form hydrated surface layers on sliding in humid air. In the case of non-oxide ceramics, oxidation can be followed by hydration, so that the nature of the surface films formed on silicon nitride in humid air or water is controlled by both the reactions



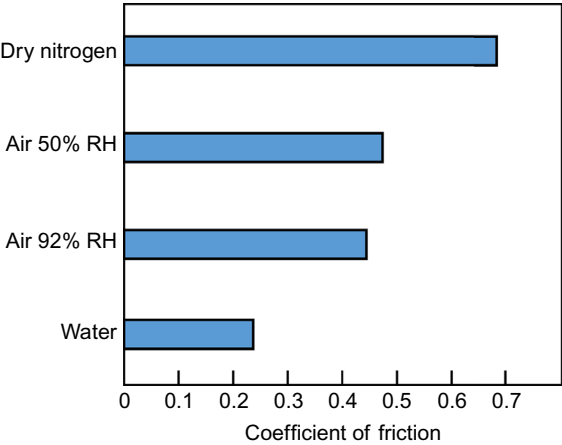
and, if hydration occurs fully,



These tribochemical effects are responsible for the influence of atmospheric composition on friction which is commonly observed in ceramics. [Figure 3.20](#), for example, shows how the value of μ falls with the increasing availability of water for silicon nitride sliding against itself in dry nitrogen gas, in air of two different humidity levels, and in liquid water. The reactions outlined above lead to the formation and hydration of a silica film at the interface, lowering its shear strength and thereby reducing the coefficient of friction.

In some ceramics the effects of surface films can be very marked: diamond and titanium nitride provide good examples. Both show high friction when sliding against themselves in vacuum: with diamond, μ reaches ~ 1 after cleaning or repeated sliding in vacuum. Yet in air, much lower friction is measured: $\mu = 0.05$ to 0.15 for diamond, and typically 0.1 to 0.2 for titanium nitride. Significant surface oxidation has been reported in titanium nitride and appears to be responsible for its low friction in air,

Fig. 3.20 Illustration of the effect of environment on the friction of hot-pressed silicon nitride. The values of μ are derived from pin-on-disc tests at a sliding speed of 150 mm s^{-1} , carried out in the environments indicated (data from Ishigaki, H., et al. 1985. In: Ludema, K.C. (Ed.), *Wear of Materials*. ASME, pp. 13–32)



but in diamond the surface modification is due to adsorption of a gaseous species, rather than the formation of a reaction product. The particularly low friction of diamond on diamond in air is due to the very low adhesive force between the surfaces in the presence of the adsorbed contamination, together with a small contribution from other dissipative processes.

Diamond-like carbon (DLC) coatings sliding against themselves can show very low friction under certain conditions. As discussed in [Section 7.4.3](#), these carbon-based materials are amorphous, with various proportions of sp^2 (graphitic) and sp^3 (diamond-like) bonding, as well as significant amounts of hydrogen and possibly other species present in some variants. When smooth DLC surfaces are passivated with hydrogen, derived from the environment or from the material itself, then the friction is low, but if hydrogen or some other passivating species is absent then adhesion between the opposing carbon surfaces can be high, leading to relatively high friction. [Table 3.2](#) illustrates these effects. The non-hydrogenated DLCs (a-C and ta-C) show high friction in vacuum or dry air, but much lower friction in humid atmospheres (and in water), since the surfaces become passivated with hydrogen derived from the water. For the hydrogenated DLC (a-C:H), the friction is low even in vacuum and dry air. [Figure 3.21](#) provides an additional illustration of the behaviour of

Table 3.2 Typical ranges of coefficient of friction for non-hydrogenated (a-C and ta-C) and hydrogenated (a-C:H) DLC coatings sliding against themselves in dry air and in humid air (data from Sutton, D.C., et al., 2013. *Friction* 1, 210–221)

Relative humidity (%)	<5	5–100
a-C or ta-C	0.3–0.8	0.08–0.2
a-C:H	0.003–0.3	0.02–0.5

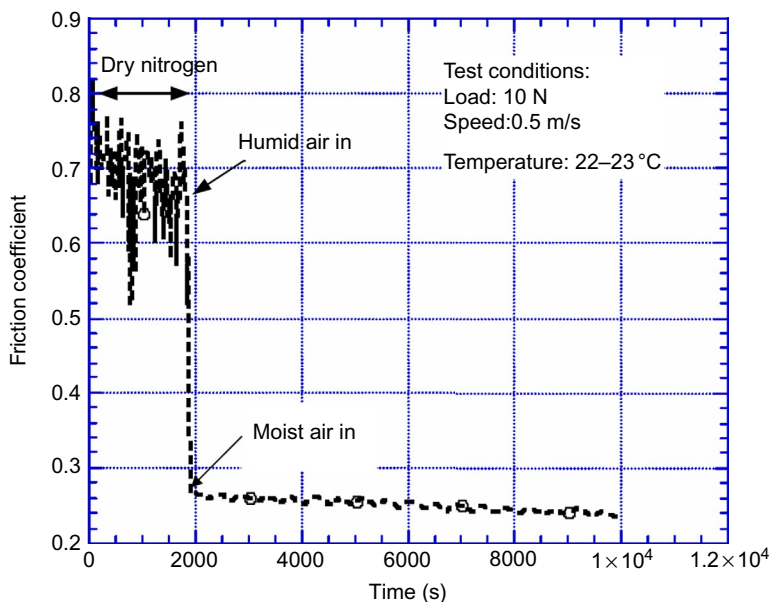


Fig. 3.21 Effect of moist air on the frictional behaviour of a ta-C DLC coating sliding against itself. The friction is very high in dry nitrogen but falls rapidly as soon as the surfaces are exposed to humid air (from Erdemir, A., Donnet, C., 2006. *J. Phys. D Appl. Phys.* 39, R311-R327, © IOP Publishing. Reproduced with permission. All rights reserved)

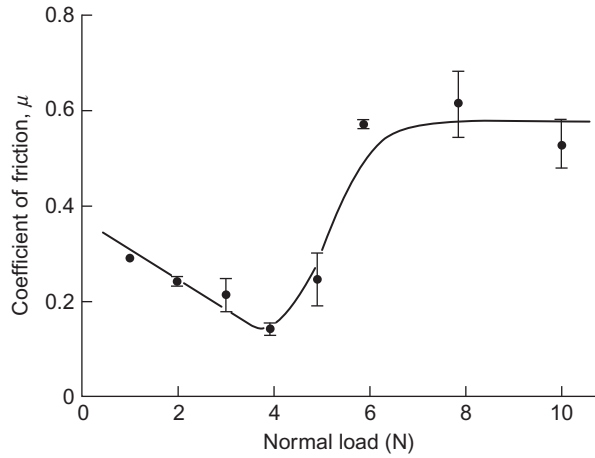
non-hydrogenated DLCs: the friction between ta-C films in dry nitrogen is high, but falls abruptly when the surfaces are exposed to moist air. The strong adhesive forces between the surfaces in dry nitrogen are suppressed as they react with oxygen and water.

As well as surface chemical reactions, a second factor which can be important in the friction of ceramics is the extent of fracture on the sliding surfaces. As we shall see in considering the wear of these materials (in [Section 5.10](#)), under some conditions widespread brittle fracture can occur in the contact zone: often intergranular in polycrystalline ceramics, but under more severe conditions, transgranular. The occurrence of fracture leads to increased friction, since it provides an additional mechanism for the dissipation of energy at the sliding contact.

Fracture is readily produced in concentrated contacts, such as a hard sharp pin or stylus sliding against a flat. [Figure 3.22](#), for example, shows results from experiments in which a diamond cone was slid over a single crystal of silicon carbide. At low loads, plastic grooving of the silicon carbide occurred, with no fracture; the value of μ was relatively low. As the load was increased brittle fracture occurred around the sliding track, similar to that shown in [Fig. 6.15](#), leading to a higher coefficient of friction.

The effects of atmospheric composition, temperature, load, sliding speed and time of sliding on the friction of ceramics can usually be interpreted in terms of changes in

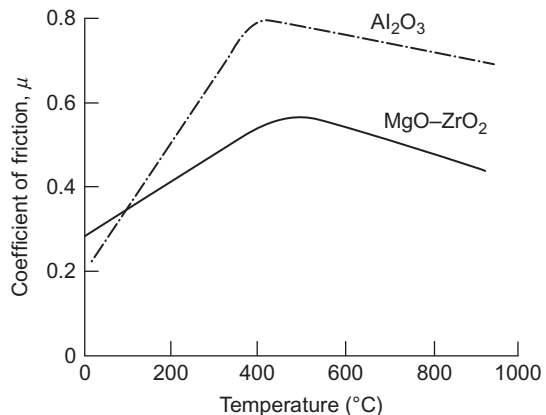
Fig. 3.22 Variation of coefficient of friction with normal load for a 60° diamond cone sliding over the (0001) face of a silicon carbide single crystal. The increase for loads above ~4 N is associated with fracture along the sliding path (from Adewoye, O.O., Page, T.F., 1981. *Wear* 73, 247–260)



the tribochemical surface films and the extent of fracture in the contact region. Both load and sliding speed will also affect the rate of frictional energy dissipation and hence the temperature at the interface; at sufficiently high temperatures increased plasticity will occur in most ceramics and will also affect the frictional behaviour. The influence of these factors can be large. Figure 3.23, for example, shows the variation of coefficient of friction with temperature for alumina and zirconia, sliding in self-mated couples in air. The initial marked rise in friction with temperature has been attributed to the removal of adsorbed water from the interface.

The effects of interfacial temperature on tribochemical processes are also often assumed to be responsible for the changes in friction with sliding speed commonly observed in many ceramic systems: Fig. 3.24 shows representative results for silicon nitride and silicon carbide in self-mated sliding, and similar behaviour has also been reported in alumina and zirconia. Fracture may also play an important role in some cases, where above a critical sliding velocity the onset of fracture coincides with a sharp increase in both friction and sliding wear rate.

Fig. 3.23 Variation of coefficient of friction with temperature for magnesia-partially stabilized zirconia and alumina sliding against themselves (from Hannink, R.H.J., 1984. *Wear* 100, 355–366)



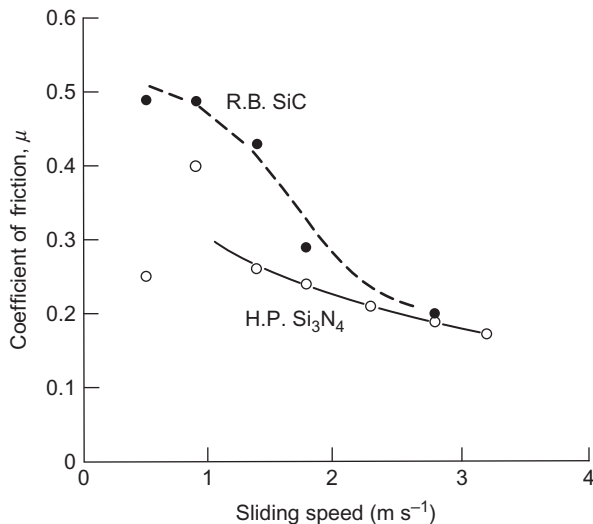


Fig. 3.24 Variation of coefficient of friction with sliding speed for reaction-bonded silicon carbide and hot-pressed silicon nitride samples sliding in self-mated couples in air (data from Cranmer, D.C., 1985. *J. Mater. Sci.* 20, 2029–2037)

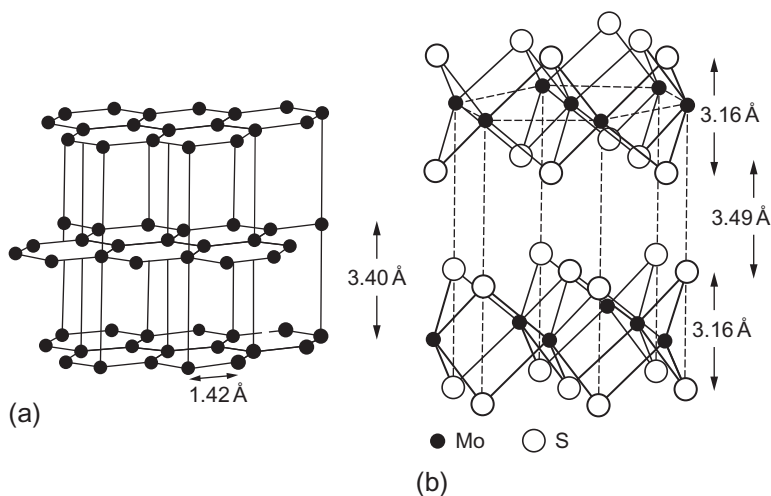


Fig. 3.25 The structures of (a) graphite and (b) molybdenum disulfide (MoS₂)

3.7 FRICTION OF LAMELLAR SOLIDS

Several materials with lamellar structures exhibit low values of friction under certain conditions, and are therefore of interest as solid lubricants; this application is discussed in detail in [Section 4.7](#). Foremost among these materials are graphite (an allotrope of carbon) and molybdenum disulfide (MoS₂), the structures of which are shown in [Fig. 3.25](#).

In each case the bonding between the atoms within the layers of the structure is covalent and strong, while that between the layers is considerably weaker. In graphite the interplanar bonding is primarily from van der Waals forces, with a weak covalent contribution resulting from interaction between the π -electron orbitals of the carbon atoms. The interplanar bond energy is about one tenth to one hundredth of that between atoms within the layers. In molybdenum disulfide the bonding between the layers of sulfur atoms is due only to van der Waals forces, and is rather weaker than in graphite. Both materials are strongly anisotropic in their mechanical and physical properties: in particular, they are much less resistant to shear deformation in the basal plane (i.e., parallel to the atomic planes) than in other directions.

The low friction of both graphite and molybdenum disulfide is associated with their lamellar structures and weak interplanar bonding, but by no means all compounds with similar structures show low friction, and the low friction values cannot therefore be ascribed to these factors alone.

The sliding friction of graphite against itself or other materials in air is low; typically, $\mu \approx 0.1$. If the surface of graphite is examined by electron diffraction after sliding, it is found that the basal planes have become oriented nearly parallel to the plane of the interface, with a misalignment of the order of 5° . The friction of graphite depends strongly on the nature of the ambient atmosphere. In vacuum or in dry nitrogen, μ is typically 10 times greater than in air, and graphite under these conditions wears very rapidly. Controlled addition of gases and vapours reveals that the low friction and wear of graphite depends on the presence of oxygen, water vapour, or other condensable vapours.

Figure 3.26 illustrates the effect on the sliding wear rate of graphite of varying the partial pressure of several gases. Oxygen shows an effect similar to that of water vapour, but at pressures about 100 times greater. Figure 3.27 (upper curve) shows how the coefficient of friction of graphite increases dramatically with decreasing air pressure. Although some features of the behaviour of graphite are still not well understood, it is clear that the adsorption of certain molecules is necessary for a

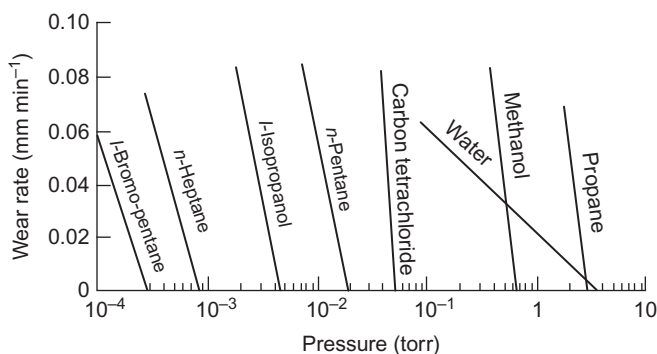


Fig. 3.26 The wear rate of graphite in sliding experiments under different partial pressures of various gases (from Savage, R.H. Schaefer, D.L., 1956. *J. Appl. Phys.* 27, 136–138, with the permission of AIP Publishing)

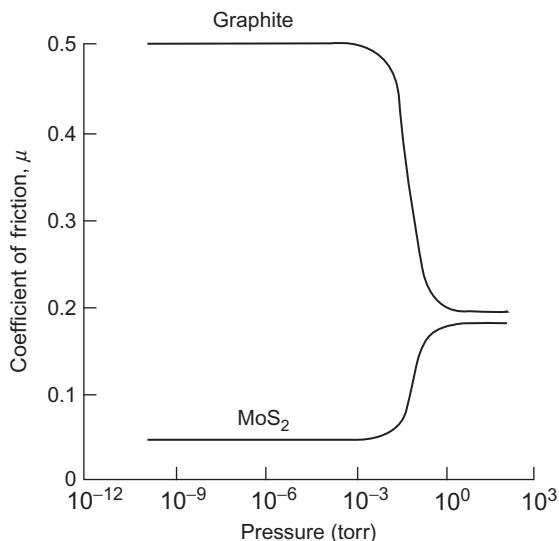


Fig. 3.27 The variation of the coefficient of friction for graphite and molybdenum disulfide with oxygen concentration (air pressure) (from Buckley, D.H., 1981. *Surface Effects in Adhesion, Friction, Wear and Lubrication*. Elsevier)

low value of μ . The friction between graphite lamellae in their plane appears always to be low; they are low-energy surfaces and show little adhesion. But the edges of the lamellae contain high-energy ‘dangling bonds’ which will bind strongly to other edge sites or to basal planes. In sliding friction some edge sites will always be exposed, and so the friction of graphite in vacuum is high. Condensable vapours lower the friction by adsorbing selectively to the high-energy edge sites, saturating the bonds, reducing the adhesion, and thus lowering the friction. Only a small concentration of adsorbed molecules is needed to produce this effect.

Molybdenum disulfide, unlike graphite, exhibits an intrinsically low coefficient of friction. [Figure 3.27](#) (lower curve) shows how the friction depends on air pressure; MoS_2 shows a low value of μ in air, which in contrast to the behaviour of graphite falls even lower in vacuum. For MoS_2 the addition of condensable vapours raises the coefficient of friction. Like graphite, MoS_2 forms an oriented film on a sliding surface, with the basal planes tending to be aligned parallel to the surface; bonding of the crystallites to the surface is probably aided by the internal polarization of the lamellae which results from their ‘sandwich’ structure. A coefficient of friction of ~ 0.1 is typically found for sliding between basal planes; for edge-oriented crystallites sliding against basal planes, μ is two or three times higher. The low coefficient of friction of MoS_2 observed in practice can therefore be ascribed to the strong orientation of the films of MoS_2 produced by sliding, and the intrinsically low adhesion and shear strength between the basal planes of the MoS_2 structure.

Many other compounds with lamellar structures exist. Some, like boric acid, talc, tungsten disulfide, graphite fluoride, cadmium chloride and lead iodide, do show low friction and are potentially useful as solid lubricants; these are discussed further, together with practical applications of graphite and molybdenum disulfide as lubricants, in [Section 4.7](#). Others, such as mica and calcium hydroxide, do not show low values of μ .

3.8 FRICTION OF POLYMERS

3.8.1 Introduction

Contact between polymers, or between a polymer and a metal, is often predominantly elastic. In this important respect the friction of polymers differs fundamentally from that of metals. As we saw in [Section 2.5.6](#), the ratio E^*/H , where E^* is the reduced modulus and H is the hardness of the softer material, determines the extent of plasticity in the contact region through equation 2.42; the surface topography is also important. For contact between metals, the value of E^*/H is typically of the order of 100 or greater, while for contact against many of the softer (i.e., low modulus) polymers E^*/H is of the order of 10. The plasticity index ψ (see [Section 2.5.6](#)) for contact on a soft polymer thus has only about one tenth of the value for a metal, and the contact is therefore almost completely elastic, except against very rough surfaces. Common polymers for which this is true include polyethylene (high and low density—HDPE and LDPE), polytetrafluoroethylene (PTFE), nylons, polymethylmethacrylate (PMMA) and some epoxy resins. Rubbers, too, make contact elastically.

A second factor that plays an important role in the friction of polymers is the strong time-dependence of their mechanical properties: most polymers are viscoelastic, and also show a marked increase of flow stress with strain rate.

Coefficients of friction between polymers sliding against themselves, or against metals or ceramics, commonly lie in the range from 0.1 to 0.5, although values outside this range are also seen. In these materials, for which Amontons' Laws are not even broadly applicable, μ varies so much with normal load, sliding speed and temperature that a list of coefficients of friction for specific materials would be of little value.

The friction of polymers, like that of metals, can be attributed to two sources: a *deformation* term, involving the dissipation of energy in quite a large volume around the local area of contact, and an *adhesion* term originating from the interface between the slider and the counterface. The regions where these two sources of friction originate are illustrated in [Fig. 3.28](#), for the sliding of a hard asperity over a polymer surface. As for metals, the distinction between the deformation and adhesion components of friction is somewhat artificial, and under many circumstances no clear demarcation can be made. However, in some experiments, and in some practical applications, one term dominates and can then be examined in isolation.

In the next two sections we shall discuss the contributions to polymer friction from these sources. Since in most practical applications polymers slide against more rigid counterfaces, we shall concentrate on this situation; much of our discussion is also applicable to the case of polymers sliding against other polymers.

3.8.2 Friction due to deformation

We can isolate the *deformation* component of friction by eliminating adhesion; this can effectively be achieved by rolling a hard sphere or cylinder over a polymer surface treated with lubricant to reduce the interfacial adhesion, or by dragging a slider across it under well-lubricated conditions. [Figure 3.29](#) illustrates the case of rolling.

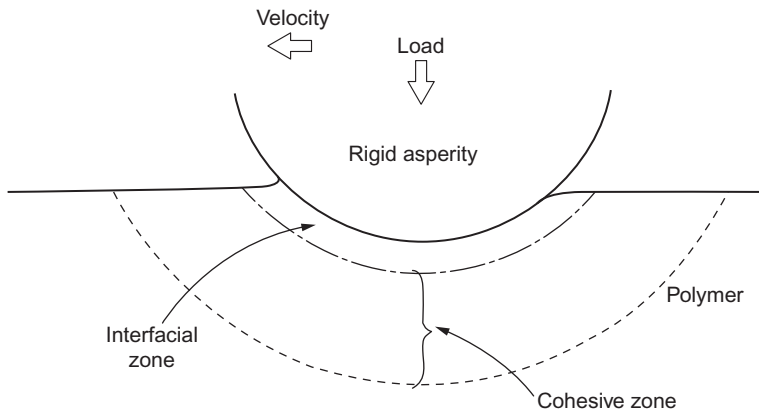


Fig. 3.28 The origins of the friction associated with the sliding of a hard smooth asperity over a polymer surface (from Briscoe, B.J., 1981. *Tribol. Int.* 14, 231–243)

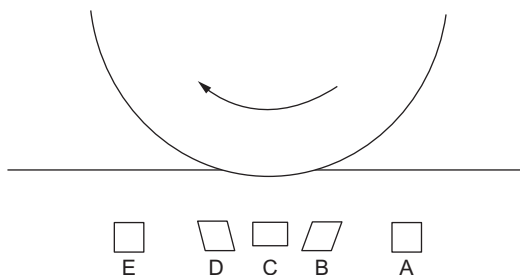


Fig. 3.29 Deformation of elements in a polymer under a rolling sphere (from Briscoe, B.J., 1981. In: Friedrich, K. (Ed.), *Friction and Wear of Polymer Composites*. Elsevier, pp. 25–59)

The resistance to rolling arises from energy dissipation in the bulk of the polymer beneath the sphere, and is a direct result of its viscoelastic response. As the sphere in Fig. 3.29 rolls towards the right, elements of the polymer in its path become progressively deformed, while those behind it recover their undeformed shape. An individual element will experience the deformation cycle illustrated by the sequence ABCDE. In a viscoelastic material, energy will be dissipated during this cycle as heat; the frictional force will be equal to the energy dissipated per unit distance moved by the force.

It can be shown that if a fraction β of the total energy input is dissipated (i.e., the fraction $(1 - \beta)$ is recovered mechanically), then for a sphere of radius R rolling under a normal load W the frictional force F_{def} will be given by

$$F_{\text{def}} = 0.17\beta W^{4/3} R^{-2/3} (1 - v^2)^{1/3} E^{-1/3} \quad (3.26)$$

Here ν is Poisson's ratio for the polymer, and E is the real part of Young's modulus (i.e., for a viscoelastic material, the elastic modulus derived from the component of strain in phase with a cyclically varying applied stress). For a well-lubricated conical slider of semi-angle α , rather than a rolling sphere, the friction force is given by

$$F_{\text{def}} = \left(\frac{\beta W}{\pi} \right) \cot \alpha \quad (3.27)$$

For deformation by a sphere or a cone, the fraction β is found to be about two to three times the fraction of the energy lost in a single cycle of pure shear deformation (the 'loss factor' α), reflecting the more complex cycle of deformation experienced by the polymer under rolling or sliding contact. The values of α and E vary with loading frequency and temperature in a polymer, and can be measured independently from F_{def} in separate experiments.

Good agreement is generally found between the predictions of equation 3.26 and experimental measurements of the rolling friction force for polymers. Figure 3.30 shows results for the rolling of a steel sphere on PTFE over a range of temperature. The correlation between F_{def} and the quantity $\alpha E^{-1/3}$, as measured in independent experiments, is good; the peaks in the curves are associated with specific energy-absorbing motions of the polymer molecules. Similar dependence is also found on the rate of loading (i.e., on the speed of rolling or sliding).

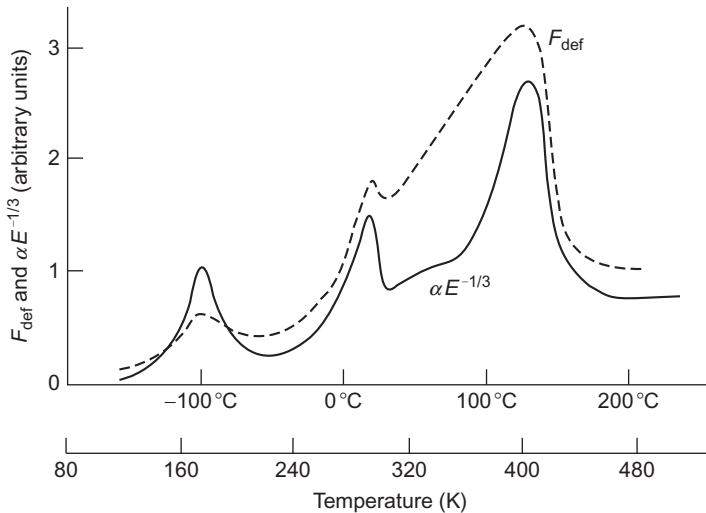


Fig. 3.30 Dependence of the deformation component of the frictional force (F_{def}) and the quantity $\alpha E^{-1/3}$ (see text) on temperature, for a steel sphere rolling over PTFE (48% crystalline). The units employed for the ordinate of the graphs are arbitrary (from Ludema, K.C., Tabor, D., 1966. *Wear* 9, 329–348)

3.8.3 Friction due to adhesion

If a smooth polymer surface slides against a relatively smooth rigid counterface so that the contribution of deformation to friction is negligible, then the frictional force originates primarily in *adhesion* between the two surfaces, and can be analyzed in the same way as the adhesion component of friction for metals (Section 3.4). The analysis differs, however, in that the contact of a polymer against a typical engineering surface will be predominantly elastic rather than plastic.

At low loads and for moderately rough surfaces, where the true area of contact is a small fraction of the apparent area, the contact area remains effectively proportional to the normal load (Section 2.5.6) and the coefficient of friction is therefore constant and independent of the normal load. Figure 3.31 illustrates this case, for the friction of crossed cylinders at low loads, when their surfaces have been lathe-turned to produce moderate surface roughness. When the load is increased however, or if the surfaces are smooth and polished, the coefficient of friction is found to fall with increasing load. Under these conditions, although the frictional force remains proportional to the real area of contact, that area is no longer proportional to the normal load.

At high loads, or for very smooth surfaces, the elastic deformation at the points of contact between the cylinders is so great that individual asperities on the contacting surfaces are ‘ironed out’, and the situation approximates to the contact of a single giant asperity. We saw in Section 2.5.1 that for a single spherical asperity deforming elastically, the area of contact will be proportional to the load raised to the power 2/3, and the same exponent applies for the contact of elastic cylinders.

We would therefore expect that under these conditions

$$\mu \propto A/W \propto W^{2/3}/W \propto W^{-1/3} \quad (3.28)$$

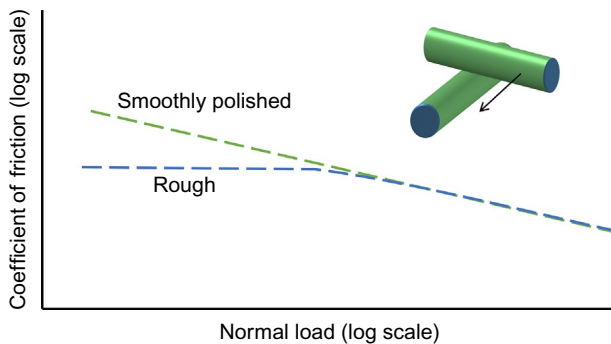


Fig. 3.31 Schematic diagram showing how the coefficient of friction varies with normal load for sliding of crossed cylinders of a glassy polymer with different surface roughness (Sketch showing behaviour reported for PMMA by Archard, J.F., 1957. Proc. Roy. Soc. Lond. A243, 190)

In practice, fair agreement is found with this relationship, although the exponent of normal load may not be exactly $-1/3$. Similar dependence of the coefficient of friction on load is found for the heavily loaded contact of a rounded polymer body against a hard smooth substrate.

The adhesion responsible for this component of polymer friction results primarily from the weak bonding forces (e.g., hydrogen bonding and van der Waals forces) which are also responsible for the cohesion between the polymer chains themselves in the bulk of the material. The relative magnitude of the forces may be described in terms of the surface energies of the polymer and substrate. Surface energies of non-polar polymers (e.g., polyethylene) are typically 20 to 30 mJ m^{-2} , while those of polar polymers (e.g., nylons) have up to twice these values. Polar polymers therefore tend to show stronger adhesion. The junctions formed by adhesion show many similarities to those responsible for the friction of clean metal surfaces. Failure of the junction tends to occur within the bulk of the polymer, rather than at the interface; however, only very limited junction growth occurs, and the high values of μ associated with extensive junction growth in metals are never observed in polymers.

Many polymers sliding against hard counterfaces (e.g., metals) transfer detectable films of polymer on to the counterface. The formation and behaviour of the transfer films are important factors in the friction and wear of these polymers. Once a transfer film has formed, subsequent interaction occurs between the polymer and a layer of similar material, irrespective of the composition of the substrate. On further sliding the polymers may continue to wear by adding material to the transfer film, since the interfacial bond to the counterface is often stronger than that within the bulk of the polymer itself. The sliding wear of polymers is discussed further in [Section 5.11](#).

The properties of thin films of polymers may be investigated by placing the film between a rigid slider and a hard substrate. If sliding occurs under a sufficiently high load, the real area of contact will be equal to the apparent area and a direct comparison can then be made between the measured coefficient of friction and the value predicted from independent measurements of the shear yield strength of bulk samples of the polymer. Agreement is generally good. The coefficient of friction depends on the load, since the yield stress of a polymer, unlike that of a metal, depends strongly on the hydrostatic component of stress. The shear yield stress τ varies with the applied hydrostatic pressure P in the following way:

$$\tau = \tau_0 + \alpha P \quad (3.29)$$

where τ_0 and α are constants for a particular polymer. For the case of a thin polymer film sandwiched between rigid surfaces, $P = W/A$ and $F = \tau A$. It therefore follows that

$$\mu = \frac{F}{W} = \frac{\tau}{P} = \frac{\tau_0}{P} + \alpha \quad (3.30)$$

Good agreement is generally found between measured values of μ for polymers which form transfer films and those calculated from this equation, when the values of τ_0 and α are measured in conventional mechanical tests.

From equation 3.30, it can be seen that the coefficient of friction is expected to fall with increasing hydrostatic pressure (and hence with increasing normal load), and this is indeed observed in practice for polymers which form transfer films. For very high pressures, the value of μ tends to α . Equation 3.30 is also found to predict the friction of bulk samples of many polymers if P is taken to be the bulk hardness of the polymer.

Two common polymers, linear (i.e., unbranched: high density) polyethylene (HDPE) and polytetrafluoroethylene (PTFE) exhibit appreciably lower coefficients of friction than most others and have important applications as solid lubricants and bearing materials. Their molecular structures are characterized by linear, unbranched chains without bulky or polar side groups, giving weak intermolecular bonding and a high degree of crystallinity. Although these polymers form transfer films on a hard counterface, subsequent sliding in the same direction tends to occur at the interface between the bulk polymer and the film, leading to low wear rates. The coefficient of friction for initial sliding on a clean hard substrate is not particularly low (typically $\mu = 0.2$ to 0.3) and the transfer film is of the order of micrometres thick. As sliding progresses, the coefficient of friction drops to a much lower value (perhaps as low as 0.05 for PTFE); the transfer film becomes much thinner and contains molecular chains strongly oriented parallel to the sliding direction. For subsequent sliding with the same orientation of the slider on the same track, the friction remains low. But if the slider is rotated, destroying the molecular alignment between the slider and the transfer film, the friction and wear rate rise substantially.

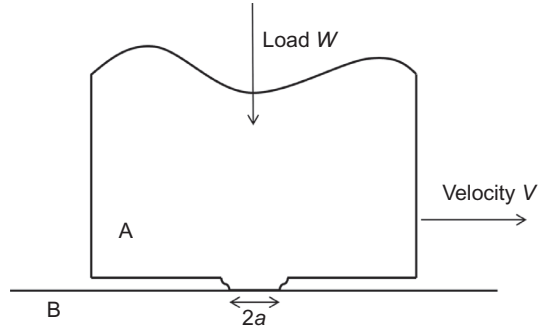
The behaviour of HDPE and PTFE has been attributed to the smoothness of the molecular chains in these materials, although it seems that a low surface energy, associated with weak intermolecular forces, is also needed to allow the very thin oriented films to be drawn out from the bulk.

3.9 FRICTIONAL HEATING

When two surfaces slide against each other and work is done against friction, heat is generated at or very close to the sliding interface. This heat energy represents nearly all of the frictional work, since only a small proportion of it (up to a maximum of a few percent) is stored in dislocations or other structural defects, or is used in the creation of new surface in debris particles or cracks. It is therefore easy to see that there will be a temperature rise at and near the contact, but not so straightforward to predict its value. Under some conditions the rise in temperature will be unimportant, but in others, as we shall see later, it may have a critical influence on the behaviour of the tribological system. Frictional heating may cause a local reduction in the strength of a material or even result in melting; it may cause enhanced oxidation or other chemical reaction with the surrounding atmosphere or lubricant additives; or it may cause tensile stresses due to thermal expansion mismatch or phase transformations that result in fracture in some brittle materials.

Figure 3.32 illustrates a simple example of a sliding contact that we will use to discuss the challenges involved in estimating frictional temperature rises. A cylindrical pin carrying a protuberance of radius a slides over a plane surface at velocity V . If the normal load is W and the coefficient of friction is μ , then the frictional power

Fig. 3.32 Schematic diagram showing sliding contact between a body A with a protuberance of circular cross-section, and a semi-infinite body (B). Frictional heat generated at the contact patch is conducted into both bodies



dissipated is μWV . We shall assume that all of this power represents heat that is generated at the interface between the moving body A and the stationary body B, and that it is conducted into the two bodies; we ignore other mechanisms of heat transfer such as convection and radiation. The small region of contact between the two bodies may represent a single asperity, or perhaps a larger area such as the end of a rounded pin in a pin-on-disc test, as discussed in Section 5.2. The materials of the two bodies, which may be different, are characterized by their thermal conductivities K_A and K_B , and thermal diffusivities κ_A and κ_B : thermal diffusivity is defined by $\kappa = K/\rho C$ where ρ is density and C is specific heat per unit mass. We assume that the cylinder A has a finite length. Body B is semi-infinite, and initially at the same temperature as A.

The rate of heat generation per unit area at the contact, q , will be given by

$$q = \frac{\mu WV}{\pi a^2} \quad (3.31)$$

This total heat flux will be split between the two bodies, with a flux q_A passing by conduction into body A and q_B into body B, where $q = q_A + q_B$.

Body A experiences a steady flow of heat from the contact, which will cause the temperature to rise locally at the contact area and also more generally within the cylinder. If there is a uniform heat flux over the circular contact, then the maximum temperature rise at the contact, described as the *flash temperature* ΔT_{fA} , is given by

$$\Delta T_{fA} = \frac{q_A a}{K_A} \quad (3.32)$$

The more general bulk temperature rise at the bottom face of body A will be much less than the flash temperature, since the heat input from the contact then becomes spread over the whole cross-sectional area of body A; to make an estimate of the contact temperature in the early stages of sliding we shall ignore it.

The contact patch moves across the surface of body B, and to estimate its effect there we therefore need to solve a problem involving a moving heat source. The temperature distribution beneath the surface of body B at any location will vary with time. As the heat source passes over it the temperature will first rise and then fall again as the heat diffuses into the bulk. If the heat source moves very slowly then the problem can

be treated as that of a stationary source and the maximum flash temperature at the surface ΔT_{fB} will be given by the same expression as that used for body A:

$$\Delta T_{fB} = \frac{q_B a}{K_B} \quad (3.33)$$

For a rapidly moving heat source, the heat has little time to diffuse downwards so that the heating effect will be concentrated close to the surface; and the time for which any point on the surface of body B is being heated will be short. The net result is that the maximum flash temperature rise at the surface will be smaller than in the case of a slowly moving contact (for the same value of q). Precise calculation of the flash temperature is complicated and several different models have been used that make different assumptions, for example about the shape of the contact area and the distribution of the heat flux over it. For a constant heat flux over a circular contact area (which might be appropriate for a plastically deforming asperity), a simple algebraic expression for the peak flash temperature that is valid for a wide range of source velocities V is as follows:

$$\Delta T_{fB} = \frac{2q_B a}{K_B \sqrt{4 + \pi J}} \quad (3.34)$$

where $J = Va/2\kappa_B$. For $V=0$, equation 3.34 becomes the same as equation 3.33.

The quantity J is a dimensionless group referred to as the *Péclet number*, and provides a measure of how far the heat from the moving source diffuses into the surface. The characteristic distance x travelled by heat from a source in time t can be estimated from

$$x \approx \sqrt{\kappa t} \quad (3.35)$$

(The conduction of heat leads to a distribution of temperature rise beneath the surface. This expression is exact for the r.m.s. distance travelled from a point source in an infinite medium, and approximate for other geometries and boundary conditions.)

So the time t_1 taken for heat to travel to a depth a from the surface can be approximated by

$$t_1 = \frac{a^2}{\kappa} \quad (3.36)$$

and the time taken for the contact itself to travel through its own diameter $2a$ along the surface is

$$t_2 = \frac{2a}{V} \quad (3.37)$$

The Péclet number J can be interpreted as the ratio of these two times: $J = t_1/t_2$. For small values of J ($J \ll 1$) the heating effect becomes similar to that from a static source of heat, while for large J ($J \gg 1$), frictional heating is concentrated close to the surface of body B.

It remains to estimate the proportions of the total heat flux q that are conducted into the two bodies. A widely-used assumption is that the temperatures at the surfaces of the two bodies must be the same. By putting $\Delta T_{tA} = \Delta T_{tB}$ and combining equations 3.32 and 3.34 we obtain an expression for the ratio between the heat fluxes into the two bodies:

$$\frac{q_A}{q_B} = \frac{2K_A}{K_B\sqrt{4 + \pi J}} \quad (3.38)$$

If both bodies have the same thermal conductivity, we see that at low sliding speeds (e.g., $J < 0.1$) the heat is partitioned almost equally between the two, but that as the sliding speed increases, so does the proportion of the heat conducted into body B.

SOURCES AND FURTHER READING

- Blau, P.J., 2001. The significance and use of the friction coefficient. *Tribol. Int.* 34, 585–591.
- Bowden, F.P., Tabor, D., 1950, 1964. *The Friction and Lubrication of Solids, Part I and Part II*. Clarendon Press, Oxford.
- Briscoe, B.J., Tabor, D., 1978. Friction and wear of polymers. In: Clark, D.T., Feast, W.J. (Eds.), *Polymer Surfaces*. John Wiley, pp. 1–23.
- Buckley, D.H., 1981. Surface Effects in Adhesion, Friction, Wear and Lubrication. *Tribology Series No. 5*. Elsevier.
- Buckley, D.H., Miyoshi, K., 1989. Tribological properties of structural ceramics. In: Wachtman, J.B. (Ed.), *Structural Ceramics. Treatise on Materials Science and Technology*, vol. 29. pp. 293–365.
- Clauss, F.J., 1972. *Solid Lubricants and Self-Lubricating Solids*. Academic Press.
- Erdemir, A., Donnet, C., 2006. Tribology of diamond-like carbon films: recent progress and future prospects. *J. Phys. D. Appl. Phys.* 39, R311–R327.
- Hutchings, I.M., 2016. Leonardo da Vinci's studies of friction. *Wear* 360–361, 51–66.
- Kennedy, F.E., 2001. Frictional heating and contact temperatures. In: Bhushan, B. (Ed.), *Modern Tribology Handbook*, vol. 1. CRC Press, pp. 235–272.
- Kennedy, F.E., Lu, Y., Baker, I., 2015. Contact temperatures and their influence on wear during pin-on-disk tribotesting. *Tribol. Int.* 82, 534–542.
- Krim, J., 2002. Resource Letter: FMMLS-1: friction at macroscopic and microscopic length scales. *Am. J. Phys.* 70, 890–897.
- Lancaster, J.K., 1972. Friction and wear. In: Jenkins, A.D. (Ed.), *Polymer Science*. vol. 2. North-Holland, pp. 959–1046.
- Ling, F.F., Pan, C.H.T. (Eds.), 1988. *Approaches to Modeling of Friction and Wear*. Springer-Verlag.
- Persson, B.N.J., 1998. *Sliding Friction*. Springer.
- Rabinowicz, E., 1995. *Friction and Wear of Materials*. John Wiley.
- Rigney, D.A. (Ed.), 1981. *Fundamentals of Friction and Wear of Materials*. ASM.
- Ringlein, J., Robbins, M.O., 2004. Understanding and illustrating the atomic origins of friction. *Am. J. Phys.* 72, 884–891.
- Singer, I.L., Pollock, H.M. (Eds.), 1992. *Fundamentals of Friction*. Kluwer.
- Tabor, D., 1972. Friction, lubrication and wear. In: Matijevic, E. (Ed.), *Surface and Colloid Science*. vol. 5. John Wiley, pp. 245–312.

QUESTIONS FOR CHAPTER 3

Question 3.1

Friction is said to be a ‘non-conservative’ force. Explain what is meant by the term ‘non-conservative’, and give another example of a non-conservative force. Give three examples of conservative forces.

Question 3.2

Give four examples of situations where high friction is desirable, and four in which friction should be low. For each example, explain why the high or low coefficient of friction is important. Try to give examples from as broad a range of fields as possible.

Question 3.3

I place a sheet of paper on a table, and a drinking glass full of water on top of the paper, 100 mm from the far end of the sheet. If I pull the sheet gently towards me, the glass moves with the paper. If I suddenly jerk the paper and move it quickly, I can withdraw it completely leaving the glass standing on the table. Explain these observations. For the second case, estimate how far the glass will have moved relative to the table, if the coefficient of dynamic friction between the glass and the paper is 0.4 and the paper is withdrawn at an average speed of 1 m s^{-1} relative to the table. State any assumptions you make. Is there an upper limit to the coefficient of friction for which it would be possible to withdraw the paper at this speed?

Question 3.4

- Explain what is meant by ‘junction growth’ and discuss its importance in the sliding friction of metals.
- For a metal that obeys the von Mises yield criterion, show that a simple model for junction growth applied to a single asperity, as illustrated by Fig. 3.8, leads to the expression:

$$W^2 + 3F^2 = A^2 p_o^2$$

where W is the normal load on the contact, F is the tangential force, p_o is the yield stress of the material in uniaxial compression and A is the real area of contact. The von Mises criterion can be expressed as

$$(\sigma_1^2 - \sigma_2^2)^2 + (\sigma_1^2 - \sigma_3^2)^2 + (\sigma_2^2 - \sigma_3^2)^2 = 2p_o^2$$

where σ_1 , σ_2 and σ_3 are the principal stresses and p_o is the yield stress in uniaxial tension (Y).

- (c) If the bulk material has a shear yield stress τ_o and the interface fails at a shear stress τ_i , use the model from part (b) to derive an expression for the coefficient of friction μ . Hence estimate the coefficient of friction for a system in which $\tau_i = 0.1 \tau_o$.

Question 3.5

Discuss the following statements critically. Illustrate your answers with appropriate examples.

- (a) All metals in air carry a surface oxide film. The sliding friction between metals is therefore the same as that between their oxides.
- (b) Ceramics are unreactive materials and so their frictional behaviour does not depend on their environment.

Question 3.6

A steel ball rolls on the plane surface of a thermoplastic amorphous polymer which has a glass-transition temperature of -20°C . Discuss how the coefficient of rolling friction might be expected to vary with temperature between -150°C and $+100^{\circ}\text{C}$.

Question 3.7

- (a) In analyzing heat flow in a sliding contact a dimensionless group known as the Péclet number J can be defined by $J = Va/2\kappa_B$ where V is the sliding speed, a is the radius of the circular contact patch and κ_B is the thermal diffusivity of the body over which the contact is moving. How can the value of J be interpreted in physical terms? Discuss qualitatively how the distribution of heat flux between the two bodies and the resulting temperature distributions will vary as the Peclet number is varied between very low ($J \ll 1$) and very high ($J \gg 1$) values.
- (b) A brake in a piece of office machinery consists of a fixed ball sliding against the rim of a rotating disc. Two different designs are proposed: one with a nylon ball and a steel disc, and the other with a steel ball and a nylon disc. In both designs the coefficient of sliding friction between the nylon and the steel can be assumed to be 0.5, and the contact patch to be circular with a diameter of 1 mm. The load pressing the ball against the disc is 10 N and the disc rotates at a maximum surface speed of 1 m/s. The material properties are as follows:

	Thermal conductivity, K $\text{W m}^{-1} \text{K}^{-1}$	Density, ρ kg m^{-3}	Specific heat, C $\text{J kg}^{-1} \text{K}^{-1}$
Steel	50	7800	450
Nylon	0.25	1100	1300

Use the method described in [Section 3.9](#) to calculate, for each design of brake:

- (i) the maximum power dissipated in friction
- (ii) the value of the Péclet number J
- (iii) the maximum flash temperature rise at the sliding interface

Explain in qualitative terms why the interfacial temperature rise is considerably greater for one design than for the other. Discuss the proposed design of brake critically. What further information would you need in order to optimize the design of the brake?

Lubricants and lubrication

4

4.1 INTRODUCTION

We have seen in the previous chapter that the coefficient of sliding friction for dry metals, and for many ceramics and some polymers, is rarely lower than 0.5, and in many cases is rather higher. Such high values of μ in engineering applications would often lead to intolerably high friction forces and frictional energy losses, as well as frictional heating. For most practical uses, therefore, *lubricants* are used to reduce the frictional force between surfaces.

A lubricant functions by introducing between the sliding surfaces a layer of material with a lower shear strength than either of the surfaces themselves, or the interface. In some lubricated systems, the lubricant may not completely prevent asperity contact, although it reduces it and may also reduce the adhesive forces between the surfaces. In other cases, the lubricant completely separates the surfaces. Thus to a greater or lesser extent, the use of a lubricant will always reduce the rate of sliding wear, and this is another substantial benefit of lubrication.

A wide variety of materials (gases, liquids or solids) may be used as lubricants. It is convenient to distinguish between various types of lubrication. In *hydrodynamic* lubrication the surfaces are separated by a fluid film, which is usually thick in comparison with the asperity heights on the bearing surfaces. The hydrostatic pressure in the film causes only small elastic distortion of the surfaces, which to a first approximation can be treated as rigid bodies. *Elastohydrodynamic* describes the case where the local pressures in the lubricant are so high and the lubricant film so thin that elastic deformation of the surfaces can no longer be neglected; indeed, it is a vital feature of this regime of lubrication. In *boundary* lubrication, the surfaces are separated by adsorbed molecular films, usually laid down from an oil or grease containing a suitable boundary lubricant; appreciable asperity contact may nevertheless occur. *Solid lubricants* function by providing a solid interfacial film which either exhibits low shear strength or results in an interface with low shear strength.

We shall discuss these four types of lubrication below, but since in the first three the lubricant is in most cases an oil or grease, we shall first describe the properties and compositions of lubricating oils and greases, starting with a discussion of *viscosity*.

4.2 VISCOSITY

The most important property of an oil for lubricating purposes is its *viscosity*. Viscosity provides a measure of the resistance of a fluid to shearing flow, and may be defined as the shear stress on a plane within the fluid, per unit velocity gradient normal to that plane.

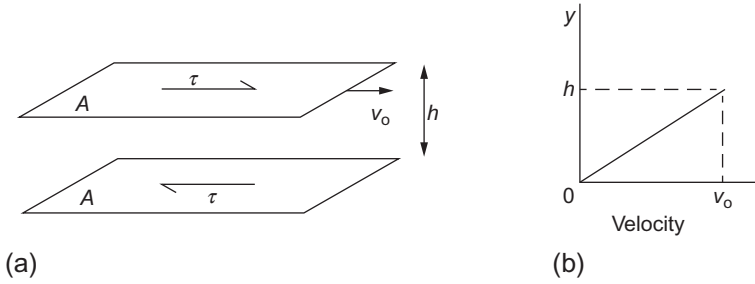


Fig. 4.1 Viscosity is defined by considering the shear stress τ acting on two planes within the fluid, of area A and separated by a distance h , moving with relative velocity v_o

Figure 4.1 illustrates this definition. We consider two planes within the fluid, one of which is moving parallel to the other with a relative velocity v_o (Fig. 4.1a). The velocity gradient dv/dy in the fluid is assumed to be constant and uniform between the two planes (Fig. 4.1b) and is given by

$$\frac{dv}{dy} = \frac{v_o}{h} \quad (4.1)$$

The shear stress acting on the planes is τ . The viscosity η is defined by the relationship

$$\tau = \eta \frac{dv}{dy} \quad (4.2)$$

An exactly equivalent definition, which may be more readily understood by those with a background in solid mechanics, is provided by the equation

$$\tau = \eta \frac{d\gamma}{dt} = \eta \dot{\gamma} \quad (4.3)$$

where γ is the shear strain in the fluid. A fluid, since it has no shear strength, can support no static shear stress, but its viscous nature leads to a dynamic shear stress τ which is proportional to the velocity gradient as described by equation 4.2, or, equivalently, to the rate of change of shear strain $\dot{\gamma}$ (often called the shear rate) as described by equation 4.3. The constant of proportionality is the viscosity of the fluid, η : the shear stress developed per unit shear rate.

Viscosity as defined above is termed *dynamic viscosity*, and has dimensions $\text{mass} \times \text{length}^{-1} \times \text{time}^{-1}$. The SI unit of dynamic viscosity is therefore the pascal second ($1 \text{ Pa s} = 1 \text{ kg m}^{-1} \text{ s}^{-1}$). Dynamic viscosities are rarely quoted in these units, however; more common is the millipascal second ($1 \text{ mPa s} = 10^{-3} \text{ Pa s}$) or an equivalent, older unit, the *centipoise* ($1 \text{ cP} = 1 \text{ mPa s}$). Typical lubricating oils have dynamic viscosities in the range 2–400 mPa s; the viscosity of water at 20°C is almost exactly 1 mPa s.

The ratio of dynamic viscosity to fluid density (η/ρ) sometimes occurs in calculations. This quantity is known as the *kinematic viscosity*; its SI unit is $\text{m}^2 \text{s}^{-1}$, and a commonly used, older unit is the *centistokes* ($1 \text{ cSt} = 1 \text{ mm}^2 \text{s}^{-1} = 10^{-6} \text{ m}^2 \text{s}^{-1}$). Since many lubricating oils have a density of $850\text{--}900 \text{ kg m}^{-3}$, their kinematic viscosities in cSt tend to be typically 10%–15% higher than the corresponding dynamic viscosity value in mPa s (or cP).

The viscosity of an oil can be measured in many different ways. The simplest method is perhaps to measure the time taken for a certain volume of oil to flow through an orifice or capillary tube of fixed geometry under gravity. Under these conditions the ratio of dynamic viscosity to density is clearly relevant, and so the method measures kinematic viscosity directly, although calibration of the method is needed to obtain a value in absolute units. Historically, relative values of viscosity were sometimes all that was required, and so the viscosity was taken to be the time required for the sample to flow through a particular type of viscometer, in units of seconds. Viscosity measurements are now usually made with an instrument (a *shear rheometer*) which measures the torque acting on a rotating body separated from a fixed body by a gap of known geometry, with the fluid under test being placed in the gap. The measurement is commonly made with parallel rotating discs (flat-on-flat or cone-on-flat), or with concentric cylinders. For a plate with a conical surface rotating about the axis of the cone against a flat disc, the shear rate in the fluid is constant over the whole sample volume. Rheometers are generally designed to carry out automated tests in which both shear rate and temperature are varied in order to measure the fluid properties over a wide range.

Several methods of describing viscosity apart from the SI units have been used in the past, and are still in use in the lubricant industry. Most important are the widely-used ISO/ASTM grades and the SAE (US Society of Automotive Engineers) grades for automotive oils. Charts and tables, and on-line converters, are available to allow viscosity scales to be compared, although such comparisons are always approximate and depend on the test conditions, since as we shall see below, different oils show different variation of viscosity with temperature and shear rate. The ISO/ASTM viscosity grade (usually denoted ISO VG and defined by standard ISO 3448) relates directly to the kinematic viscosity in cSt at 40°C , with the ‘standard’ grades varying by a factor of about 1.5 between neighbouring steps: for example, an oil described as ISO VG 100 has a kinematic viscosity of $100 \text{ mm}^2 \text{s}^{-1}$ at 40°C . The SAE values are defined in terms of viscosity limits at different temperatures, with the suffix W denoting a viscosity value at low (‘winter’) temperature. The viscosities of all oils fall rapidly with increasing temperature. Since the standard temperatures at which the SAE viscosities are measured are different from those used in other measurement systems, precise comparison of the viscosity scales can only be made for a particular oil.

Over a wide range of viscosity the variation of viscosity with temperature can be represented by a single equation (the MacCoull-Walther equation):

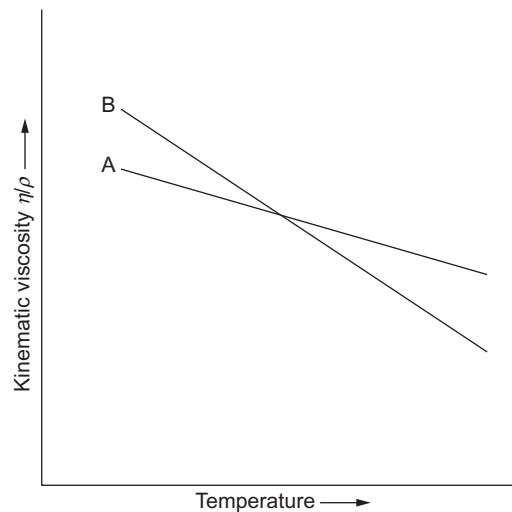
$$\log_{10} \log_{10}(\text{viscosity in cSt} + 0.7) = A - B \log_{10} T \quad (4.4)$$

where T is the absolute temperature (in Kelvin). The quantities A and B are constants for a particular oil. Although this equation is strictly only valid for mineral oils, it has been successfully applied to synthetic oils as well.

The use of equation 4.4 is straightforward and allows the viscosity of an oil to be estimated at any temperature when its values at two other temperatures are known. In some cases, a graphical method is useful and graph paper is available on which the dependence of viscosity on temperature for any oil can be plotted as a straight line. Different oils will show different temperature dependence of viscosity on such plots. For some purposes this temperature dependence must be described, and the concept of *viscosity index* is often used. The behaviour of two different oils is illustrated in Fig. 4.2. Both oils have the same viscosity at a certain temperature, but the viscosity of oil A drops off less rapidly at higher temperatures than B. Viscosity index (VI) is quantified by comparing the behaviour of an oil with that of two reference oils of known VI. The reference oils chosen when the scale was originally established in 1929 were arbitrarily assigned VI values of 0 and 100, but viscosity indexes outside this range are now commonly encountered and the original standard oils have long been unavailable. One simple method of estimating the VI for an oil is to use a nomogram (a diagram showing various scales across which a straight line is drawn to relate values between the different scales). More accurate methods are available which use tables or computer software; details are provided in standards ISO 2909 and ASTM D2270. Most commercial lubricants have a VI of about 100; higher values are needed for oils that will experience a wide range of operating temperature, such as *multigrade* oils for motor vehicle engines which typically have a VI of 150 or higher.

The viscosity of an oil also depends on hydrostatic pressure. Under some conditions of lubrication (elastohydrodynamic lubrication, discussed in Section 4.5) this effect is extremely important. A simple model (the Barus equation) suggests an exponential dependence:

Fig. 4.2 The variation of kinematic viscosity with temperature for two mineral oils. Oil A has a higher viscosity index than oil B



$$\eta = \eta_o \exp(\alpha P) \quad (4.5)$$

where η_o and α are constants for a particular oil and P is the hydrostatic pressure. α has been shown empirically for mineral oils to be related to the viscosity at zero pressure, η_o , by the relationship

$$\alpha \approx (0.6 + 0.965 \log_{10} \eta_o) \times 10^{-8} \quad (4.6)$$

where η_o is in mPa s and the units of α are Pa^{-1} . More accurate models for the pressure dependence of viscosity are also available, but the Barus equation is widely used.

4.3 COMPOSITIONS AND PROPERTIES OF OILS AND GREASES

Most lubricating oils are of mineral origin, and contain several different hydrocarbon species with mean molecular weights generally between about 300 and 600. Some typical molecular structures are shown in Fig. 4.3. Predominant molecules are

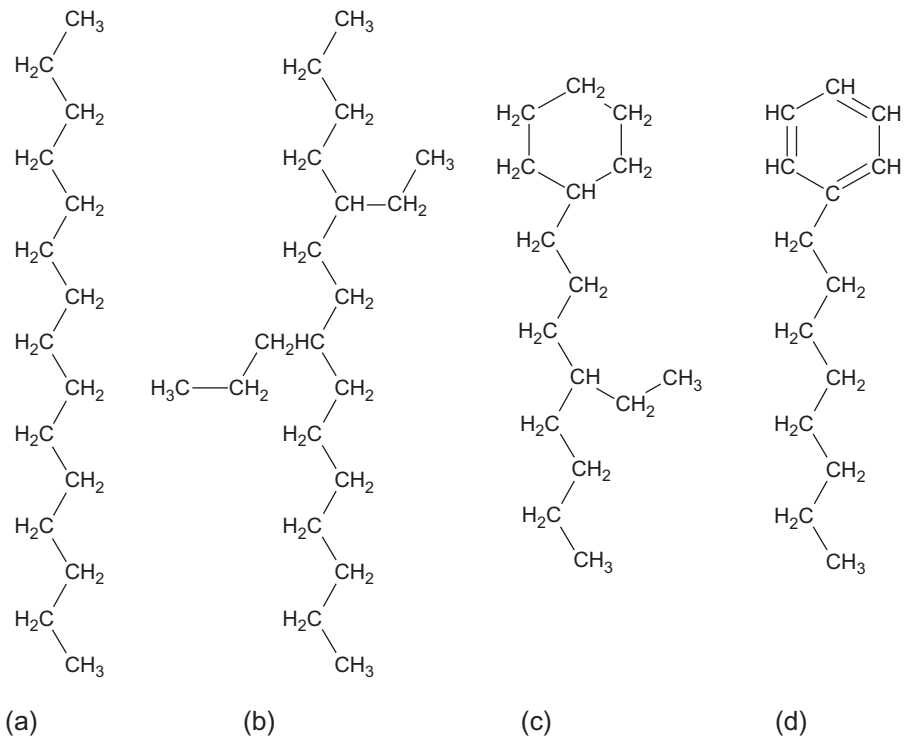


Fig. 4.3 Examples of structures of some hydrocarbon molecules present in oils: (a) paraffin (alkane); (b) branched paraffin (alkane); (c) naphthenic; (d) aromatic

saturated long-chain hydrocarbons (*paraffins*) with straight or branched chains containing 15–30 carbon atoms, and saturated 5- or 6-membered hydrocarbon rings (*naphthenes*) with attached sidechains up to 20 carbon atoms long. Aromatic components are also present in small proportions, consisting of one or more benzene rings with saturated sidechains. Many other constituents may be present in smaller quantities. Mineral oils may be broadly classified as *paraffinic*, in which most of the carbon atoms are present in paraffin chains, or *naphthenic*, in which the sidechains are short and the proportion of carbon atoms in naphthene rings is only slightly less than that in the sidechains, or *mixed*. Synthetic paraffinic hydrocarbons (e.g., *poly-alpha-olefins*, PAOs) are increasingly used in high-performance lubricants as their structures and properties can be better controlled than those of natural mineral oils.

Other synthetic oils are used as lubricants under conditions where mineral oils would be unsuitable, for example at very high or low temperatures, or where low flammability is vital. Some examples are listed in Table 4.1. Among typical synthetic oils are *organic esters* which lubricate over a wide temperature range and are used for example in gas turbine engines, *polyglycols* which have excellent boundary lubricating properties (see Section 4.6) and decompose at high temperature without leaving solid residues, and *silicones* which are chemically very stable, can operate at high and low temperatures, and are electrically insulating. All synthetic oils, however, are more expensive than natural mineral oils. They are therefore used only in specialized and demanding applications.

The viscosity of a mineral oil depends strongly on its composition, with higher viscosity being associated with higher molecular weight. Generally, paraffinic oils have a higher VI than naphthenic oils, but the VI can be modified considerably by the use of suitable polymer additives, such as polybutene or polyacrylics. These increase the viscosity index of an oil and are often used as *viscosity index improvers*. As the temperature rises, these long-chain molecules occupy progressively larger volumes and increase the viscosity of the oil-polymer blend, thus increasing the value of VI. Under conditions of high shear rate, however, certain VI improvers may lose their effectiveness, and the viscosity will then drop sharply. Figure 4.4 illustrates this effect for two

Table 4.1 Examples of some synthetic lubricants, compared with mineral hydrocarbon oil (data from Lansdown, A.R., 1982. Lubrication. Pergamon Press, New York)

	Diester	Phosphate ester	Methyl silicone	Polyphenyl ether	Mineral oil
Maximum <i>T</i> in air (°C)	210	120	180	320	150
Minimum <i>T</i> (°C)	–35	–55	–50	0	0
VI	145	0	200	–60	0–140
Boundary lubrication	Good	V. good	Poor	Fair	Good
Reason for use	High <i>T</i>	Fire resistant	High <i>T</i>	High <i>T</i>	

mineral oils which have the same value of viscosity at low shear rate, but differ in their behaviour at high rate. Prolonged shearing of the additive molecules can also lead to chain scission in some VI improvers, with permanent reduction in their effectiveness. Resistance to shear thinning and degradation varies between different VI improvers.

A variety of other additives may be used to improve the lubricating qualities of a mineral oil or to prolong its life; additives typically make up 10–20 wt% of an automotive engine lubricant, as shown in Table 4.2. Boundary lubricants (such as antiwear

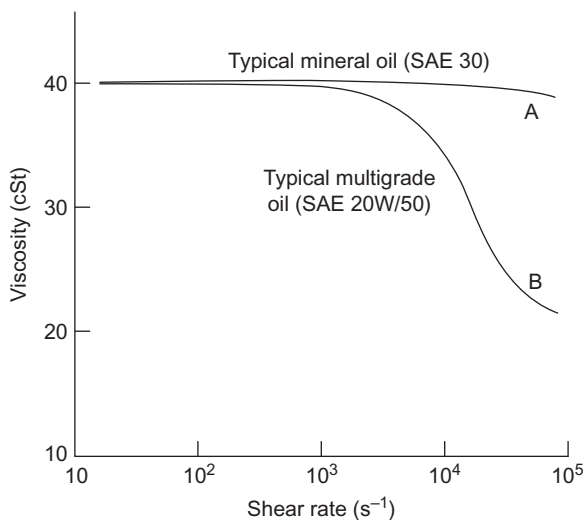


Fig. 4.4 The variation of viscosity with shear rate for two mineral oils with the same viscosity at low shear rates. Oil A is a plain mineral oil, while oil B contains viscosity index improvers which lose their effectiveness at high shear rates. (from Lansdown, A.R., 1982. Lubrication. Pergamon Press, New York)

Table 4.2 Typical ranges of composition of automotive engine lubricants (from Bell, J.C., 1993. In: Taylor, C.M. (Ed.), Engine Tribology. Elsevier, pp. 287–301, © Shell Global Solutions International B.V.)

Function	Component	Concentration (wt%)
Base oil (mineral or synthetic)		75–95
Friction and wear	VI improver	0–6
	Antiwear additive	0.5–2
	Friction modifier	0–2
	Corrosion inhibitor	0–1
Contamination/cleanliness	Antioxidant	0–1
	Dispersant	1–10
	Detergent	2–8
Maintain fluid properties	Pour-point depressant	0–0.5
	Anti-foam additive	0–0.001

additives and friction modifiers) are discussed in [Section 4.6](#) and are used to reduce sliding wear and friction when the lubricating film is so thin that significant asperity contact occurs. Paraffinic oils thicken and become ‘waxy’ at low temperatures; certain polymers can be added as ‘pour-point depressants’ to improve the flow properties when cold. Oils deteriorate in service by oxidation; although naturally occurring oxidation inhibitors are present in mineral oils, antioxidants are also frequently added, together with additives to neutralize the acidic products of oxidation. Detergents and dispersants may also be added to reduce or prevent deposits resulting from oxidation or thermal degradation of the oil.

For some purposes, *grease* is preferable to oil as a lubricant. A grease is a solid or semi-fluid compound consisting of an oil and a thickening agent (5–20 wt%). Most greases incorporate mineral oils, although greases based on synthetic oils are used in specialized applications (for example, silicone greases for use at very high or low temperatures). The thickening agent is often a soap (i.e., a metallic salt of a carboxylic acid) or a clay mineral (e.g., bentonite). Greases based on calcium or lithium soaps are commonly used; calcium greases are cheaper, while lithium greases will tolerate a wider range of temperature. Similar additives to those used in oils can be incorporated; boundary lubricant additives may be used, together with solid lubricants such as graphite and molybdenum disulfide.

With its semi-solid consistency, grease is less easily displaced from bearing surfaces than oil. It will therefore not drain from the bearing under gravity, and also forms a very effective ‘squeeze film’ lubricant (see [Section 4.4](#)). Grease also provides a good seal against contamination by dirt or moisture. However, unlike oil, grease provides a very poor means of transferring frictional heat away from the sliding surfaces, and also usually leads to higher friction than an oil, due to viscous forces.

4.4 HYDRODYNAMIC LUBRICATION

Under conditions of *hydrodynamic lubrication* the sliding surfaces are separated by a relatively thick film of fluid lubricant, and the normal load is supported by the pressure within this film, which is generated hydrodynamically. For hydrodynamic lubrication the opposing surfaces must be *conformal*; that is, they must be so closely matched in dimensions that they are separated by only a small gap over a relatively large area. Examples of conformal surfaces are shown in [Fig. 4.5](#). The simplest geometry is provided by opposed planes ([Fig. 4.5a](#)); a common geometry for rotating shaft bearings is the plain journal bearing illustrated in [Fig. 4.5\(b\)](#) in which the surfaces are concentric cylinders. The gap between the two surfaces is filled with the lubricating fluid: oil or grease, or less commonly water, air, or other liquid or gas.

The pressure which supports the normal load in hydrodynamic lubrication results from viscous forces within the lubricant, which in turn result from the relative motion of the two surfaces. For a lubricant film to generate a hydrodynamic pressure between two sliding surfaces, the gap between them must converge. The geometry must therefore be as shown in [Fig. 4.6](#), although the degree of convergence has been

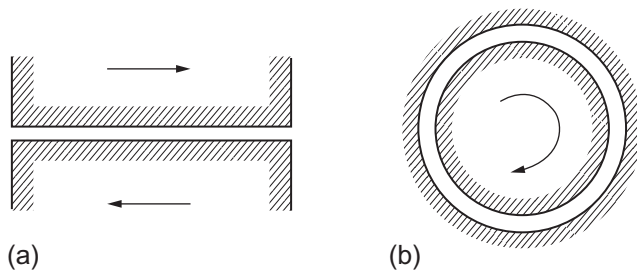


Fig. 4.5 Examples of conformal bearing geometries: (a) opposed planes; (b) plain journal bearing

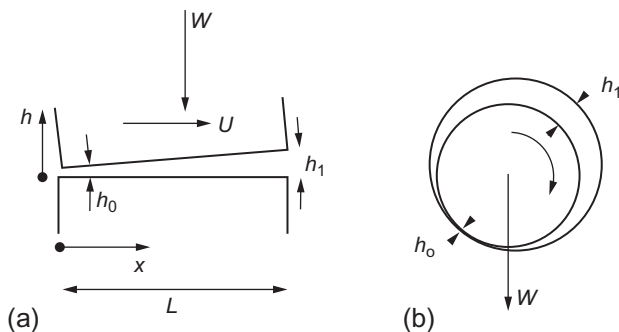


Fig. 4.6 Bearings with the geometries shown in Fig. 4.5, under conditions of hydrodynamic lubrication. The degree of convergence of the surfaces has been exaggerated for clarity

considerably exaggerated here for clarity. The separations of the surfaces, and the angles of convergence, are typically very small: a plain journal bearing lubricated with oil will have a mean lubricant film thickness of the order of one thousandth of the journal diameter, while the maximum and minimum film thicknesses may differ by a factor of four or five.

The distribution of pressure within a hydrodynamic lubricant film is described by the *Reynolds equation*, which is derived from the general (Navier–Stokes) equations for fluid flow by assuming that the flow is laminar, that the fluid film is thin compared with its other dimensions, and that the dominant forces are due to viscosity. Even with these simplifications, the general form of the Reynolds equation is not readily solved analytically, but further simplifications can be made for many problems of interest. For detailed discussion and derivation of the Reynolds equation, together with its application to bearing design, the reader is referred to the specialized texts on lubrication theory listed at the end of the chapter.

The problem is simpler when the fluid is incompressible (e.g., a liquid rather than a gas) and any variations in its viscosity with temperature, pressure or shear rate are neglected. We shall here consider the important case of uniform tangential sliding

of two planes, forming a wedge as shown in Fig. 4.6(a). The variation of pressure p with distance x is given by the simplified Reynolds equation

$$\frac{dp}{dx} = -6\eta U \frac{(h - h^*)}{h^3} \quad (4.7)$$

where h^* is the separation of the surfaces at the point of maximum pressure ($dp/dx=0$), U is the sliding velocity and the other quantities are as defined in the diagram.

This type of sliding bearing, the tilted pad, can be used in linear slideways. Alternatively, by attaching a ring of pads to a collar surrounding a shaft a rotary thrust bearing can be constructed to carry an axial load. The pressure distribution can be found by integrating equation 4.7, and employing the boundary conditions that $p=0$ at $x=0$ and $x=L$. Further integration gives the normal load W carried by the bearing (per unit width) as

$$W = 6\eta KU \left(\frac{L^2}{h_o^2} \right) \quad (4.8)$$

Here

$$K = \frac{\ln(1+n)}{n^2} - \frac{2}{n(2+n)} \quad (4.9)$$

where $n=(h_1/h_o)-1$. The numerical value of K therefore depends on the ratio of the inlet and outlet film thicknesses only, and is in fact rather insensitive to that ratio. The maximum load capacity of a tilted pad bearing occurs when h_1 is about $2.2 h_o$; then $K=0.027$.

A steadily loaded rotating journal bearing, as illustrated in Fig. 4.6(b), can be readily visualized as similar to a tilted pad slider which is ‘wrapped round’ on itself. Solution of the Reynolds equation is more complicated than for the tilted pad bearing, but the forms of the equations for the load capacity are, not surprisingly, similar. The load per unit width carried by a journal bearing is given by

$$W = S\eta U \left(\frac{R^2}{h^2} \right) \quad (4.10)$$

where h is the mean film thickness (radial clearance) in the bearing, R is the journal radius, and U is its peripheral speed. S is a dimensionless number, the Sommerfeld number, which is determined by the width to diameter ratio of the bearing and by the eccentricity of the shaft within its housing. (It should be noted that there are several different definitions of Sommerfeld number in the literature, and care must be taken in comparing values from different sources.)

The eccentricity e of the shaft is defined as the distance between the centre of the shaft and the centre of the bearing bush, and hence $e=(h_1 - h_o)/2$, where h_1 and h_o are

the maximum and minimum radial clearances as shown in Fig. 4.6(b). The *eccentricity ratio* ε is given by $\varepsilon = e/c$ where c is the mean radial clearance: $c = (h_1 + h_o)/2$.

For a short bearing with $w/D < 1/3$ where w is the width of the bearing and D is its diameter, the following analytical expression for S in terms of (w/D) and ε can be obtained by solving the one-dimensional Reynolds equation:

$$S = \left(\frac{w}{D}\right)^2 \frac{\pi\varepsilon}{(1-\varepsilon^2)^2} (0.621\varepsilon^2 + 1)^{0.5} \quad (4.11)$$

For longer bearings, values of S must be obtained by numerical methods, but design charts are readily available in textbooks on lubrication from which values can be read off for any eccentricity and length of bearing. They can also be used to solve the inverse problem of deducing the eccentricity, and hence the minimum film thickness, from a knowledge of the maximum load by calculating the value of S from equation 4.10.

If the shaft is concentric in the housing, then no convergent film is formed and clearly no load can be carried by the bearing; under these conditions $\varepsilon = 0$ and $S = 0$. A typical bearing under load might have $h_1/h_o = 4$ and hence $\varepsilon = 0.6$; then, for a bearing with its width equal to its diameter ($w/D = 1$), S would be about 2.5. For an infinitely wide bearing ($w/D = \infty$) under the same conditions, $S \approx 6.5$. As the ratio h_1/h_o increases, so S rises and the load capacity of the bearing will rise. The maximum value of h_1/h_o which can be allowed depends on the roughness of the journal and housing: too thin a lubricant film would allow asperity contact to occur, the friction would rise, wear would be rapid, and the bearing would no longer be efficient.

The frictional force in a journal bearing can also be predicted. The total tangential force F per unit width acting on the periphery of the journal is given to a fairly good approximation for an infinitely wide bearing by the *Petrov equation*:

$$F = 2\pi\eta UR/h \quad (4.12)$$

The mean coefficient of friction is therefore given approximately by:

$$\mu = \frac{F}{W} = \left(\frac{2\pi}{S}\right) \left(\frac{h}{R}\right) \quad (4.13)$$

This simple expression provides a useful estimate of μ , although it is most accurate when the shaft and housing are nearly concentric; μ becomes higher than the value given by equation 4.13 as the bearing becomes more eccentric. From this equation it is clear that in a well-designed hydrodynamic journal bearing the coefficient of friction can be very low: typically down to about 0.001 for a clearance ratio (h/R) of about the same value.

As the load on the bearing is reduced, or its rotational speed increased, the eccentricity and hence the Sommerfeld number S will decrease, leading to a rise in the coefficient of friction. On the other hand, if the load is increased or the speed is dropped

then S and the eccentricity ratio h_1/h_o will rise, resulting in a decrease in the coefficient of friction. However, as the minimum film thickness diminishes further, asperity contact between the two surfaces becomes more probable and an increasing proportion of the load is carried by direct asperity interaction rather than being supported by the fluid film. This is referred to as *mixed* or *partial lubrication*, and the coefficient of friction therefore rises. At very high loads or very low sliding speeds the support provided by the fluid film is negligible and the system is in the regime of *boundary lubrication*, discussed in Section 4.6. This behaviour is illustrated in Fig. 4.7, which shows the general form of the *Stribeck curve* relating coefficient of friction to the product $\eta U/W$ for a hydrodynamically lubricated bearing.

We have so far considered hydrodynamic effects associated with the wedge or entraining flow generated when two surfaces slide past each other. In some important practical applications, however, the surfaces have appreciable relative motion in the perpendicular direction. This type of motion results in *squeeze-film* lubrication, which occurs in pad or journal bearings subjected to varying normal loads. The lubrication of a big-end bearing on the crankshaft of an internal combustion engine (see Section 9.3.2), which is subjected to large cyclic loading by the varying force on the piston, is very largely controlled by squeeze-film phenomena rather than by the effects of wedge flow.

The simple geometry shown in Fig. 4.8 leads to the following form of Reynolds equation for squeeze-film action:

$$\frac{dp}{dx} = 12\eta V \left(\frac{x - x^*}{h^3} \right) \quad (4.14)$$

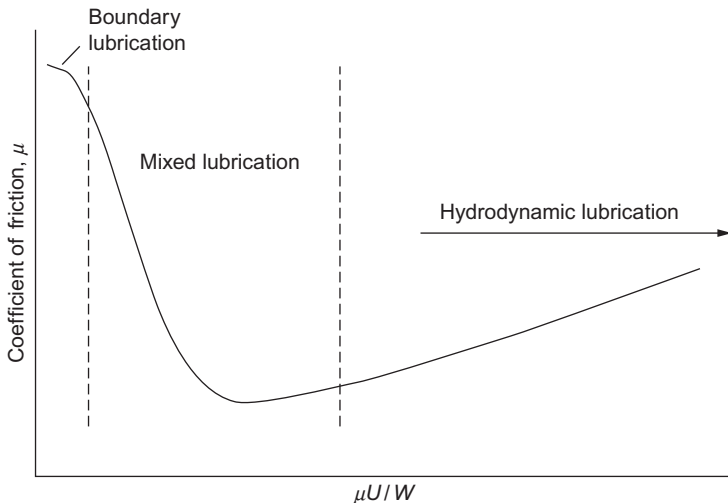


Fig. 4.7 The variation in frictional drag (expressed as the coefficient of friction μ) with the quantity $\eta U/W$ for a lubricated sliding bearing: the Stribeck curve

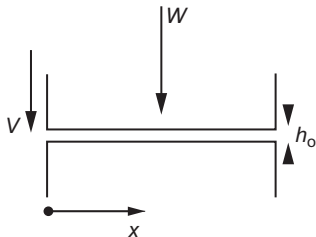


Fig. 4.8 Two opposing surfaces separated by a lubricant film and moving towards each other give rise to squeeze-film lubrication

where V is the velocity of approach of the surfaces and x^* is the position of maximum pressure (i.e., the point at which $dp/dx=0$). This equation can be integrated for suitable boundary conditions to give an expression for the normal load W per unit width carried by infinitely wide plates separated by a squeeze-film:

$$W = \eta V \left(\frac{L^3}{h_0^3} \right) \quad (4.15)$$

Comparison with equation 4.8 reveals that the normal load supported by a squeeze film varies as $1/h_0^3$ whereas that due to sliding of a wedge varies as $1/h_0^2$. The normal velocity of approach of the surfaces V need therefore be no more than about h_0/L times the sliding velocity U for squeeze-film effects to become important in a tilted pad bearing. Similar conclusions apply to the case of a journal bearing. In many cases, however, one of these effects dominates; the general Reynolds equation, which takes account of both wedge and squeeze-film forces, is complicated to solve, and in many practical cases the problem can be simplified by neglecting one effect or the other.

It was mentioned above that gases can be used as lubricants in fluid film bearings. *Gas bearings* have some advantages over oil lubricated bearings at high speeds; they offer very low friction and can operate over an extremely wide temperature range. The viscosity of a gas increases with temperature, which may offer advantages in some applications. Air bearings use a freely-available, environmentally-friendly and non-flammable lubricant. However, since the lubricant film thickness in a gas bearing tends to be less than that for an oil-lubricated bearing, the manufacturing tolerances and surface finish of the shaft and bearing bush must be appreciably more accurate. The form of Reynolds equation used to analyze the behaviour of a gas bearing must take account of the compressibility of the gas; the reader is referred to specialized texts on lubrication and bearing design for further discussion.

Mean bearing pressures (normal load/area) in oil-lubricated hydrodynamic bearings running under steady load are usually less than 2 MPa, although in impulsively loaded systems such as internal combustion engine crankshaft bearings the pressures can reach a transient maximum of ~ 50 MPa, with local maxima in the oil film perhaps five times higher still. Gas bearings operate at much lower pressures, typically up to

100 kPa. The materials from which hydrodynamically lubricated bearings are made must be carefully selected; this aspect is discussed in [Section 9.2.2](#).

4.5 ELASTOHYDRODYNAMIC LUBRICATION

If the contacting surfaces are *counterformal* (i.e., non-conforming), involving nominally line or point contact, then the local pressures in the contact zone will generally be much higher than those encountered in hydrodynamic lubrication. Examples of counterformal geometries are illustrated in [Fig. 4.9](#).

The contact between gear teeth, between a cam and its follower, or between a ball and its track in a ball bearing, all involve *concentrated* contacts of very small area. Here the local pressures between steel components typically range up to several gigapascals. Under these conditions the dependence of the lubricant viscosity on pressure plays an important role, as does the elastic deformation of the bearing surfaces. Lubrication in these circumstances is known as *elastohydrodynamic* (often abbreviated as *EHL*).

The lubricated contact of soft elastic bodies, such as rubbers, also involves a type of EHL, although then the pressures are much lower and have no effect on the viscosity of the lubricant; only the elastic deformation of the contacting surfaces need be considered. This type of lubrication is sometimes known as *soft EHL* to distinguish it from *hard EHL* which occurs when both surfaces have a high elastic modulus.

In hard EHL, the lubricant film is typically a few micrometres or tenths of a micrometre thick. According to the classical hydrodynamic theory discussed in [Section 4.4](#), films much thinner than this should be associated with the prevailing loads, and we would therefore expect asperity contact to occur. The classical hydrodynamic theory needs to be modified in two ways to explain why in practice thicker films are formed, and why in EHL the surfaces can still be separated by a continuous fluid film and damaging asperity contact can be avoided.

The very high pressure in the film locally increases the viscosity of the lubricant, and this results in a film thickness greater than that predicted by the hydrodynamic theory. We have seen that the pressure dependence of viscosity of an oil can be described by equation 4.5. For a typical mineral lubricating oil, the

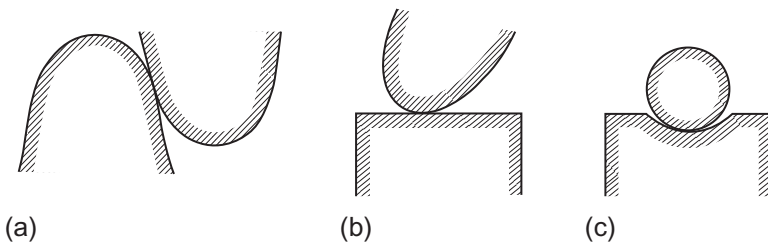


Fig. 4.9 Examples of counterformal contacts: (a) gear teeth, (b) a cam and its follower, (c) a ball in a bearing race

pressure coefficient of viscosity α may be $\sim 2 \times 10^{-8} \text{ Pa}^{-1}$. At the mean pressures occurring in a hydrodynamically lubricated bearing, the increase in viscosity will be only a few percent, but in EHL it can be very high indeed. At a pressure of 500 MPa, for example, the viscosity of the oil will be over 20,000 times that at atmospheric pressure, and it will behave within the contact zone much more like a solid than a liquid.

The second effect, which is also the only important effect in soft EHL, is that due to elastic distortion of the surfaces. When an elastic sphere is pressed against a rigid plane, initial contact occurs at a point (Fig. 4.10a). As the normal load is increased (Fig. 4.10b), the contact region expands and the contact area can be calculated from Hertz's theory (see Section 2.5.1). If a lubricant film is present, and the sphere slides over the flat, the pressure distribution and surface deformations predicted by Hertz's

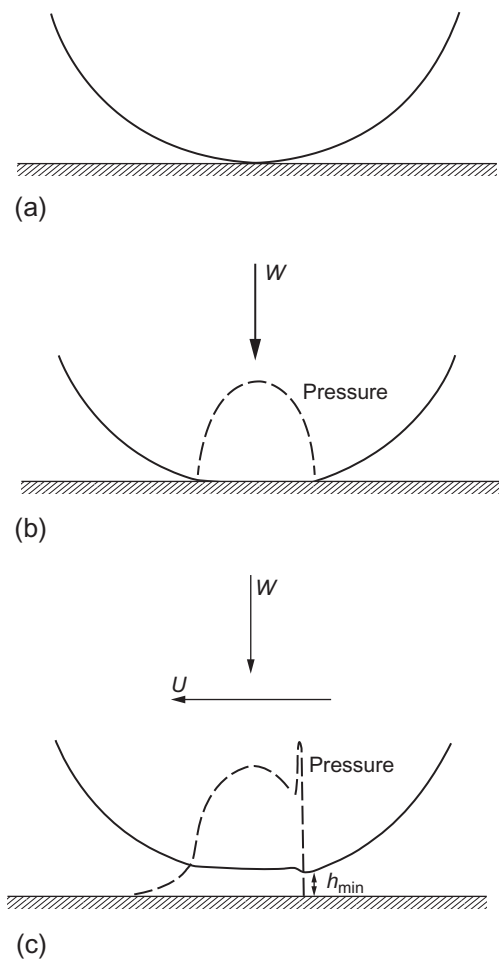


Fig. 4.10 Contact between a sphere and a plane under conditions of elastohydrodynamic lubrication: (a) point contact under zero normal load and with no sliding motion; (b) area of contact due to the normal load only; (c) elastic distortion of the sphere (shown exaggerated for clarity) under combined normal load and sliding motion, when the surfaces are separated by a lubricant film.

equations must be modified. Theoretical investigation of hard EHL involves solving Reynolds equation while taking account of the variation of lubricant viscosity with pressure (the *piezoviscous* effect), and also allowing for the elastic distortion of the bounding surfaces caused by the hydrodynamically generated pressure distribution. The film predicted by this theory has a thickness profile in the direction of sliding which is shown schematically in Fig. 4.10(c). It is nearly parallel for most of its length, and then develops a sharp constriction in the exit region, in which its thickness is reduced typically by about one fifth. Associated with this constriction is a local sharp spike in the pressure distribution.

The diagrams in Fig. 4.10 show a cross-section in the direction of sliding. In fact, for a sphere on a plane the minimum EHL film thickness, an important quantity because it determines the extent of asperity interaction, occurs towards the sides of the contact patch. It is rather smaller than the value suggested by Fig. 4.10(c). Unless major simplifications are introduced in the theory, numerical methods are necessary to solve the equations. A useful empirical power-law dependence for the minimum EHL film thickness can be derived from the computed numerical results. Several different expressions have been published, which differ only slightly. One version (due to Hamrock and Dowson) for the contact between a sphere of radius R and a plane surface is written in terms of commonly-used dimensionless groups as follows:

$$\mathcal{H}_{\min} = 1.79 \mathcal{U}^{0.68} \mathcal{G}^{0.49} \mathcal{W}^{-0.073} \quad (4.16)$$

The groups are defined by:

$$\mathcal{H}_{\min} = \frac{h_{\min}}{R}, \quad \mathcal{U} = \frac{\eta_0 u}{E' R}, \quad \mathcal{G} = \alpha E', \quad \mathcal{W} = \frac{W}{E' R^2}$$

where $h_{\min}$ is the minimum film thickness, E' is the equivalent elastic modulus of the surfaces (which is twice the reduced modulus E^* given by equation 2.16), η_0 and α are as defined for equation 4.5, and W is the normal load. u is the entraining velocity, which is the mean velocity of the two surfaces: for the case of a sphere sliding over a stationary plane with velocity U , $u = U/2$. Similar expressions to equation 4.16 are available for other contact geometries.

It is instructive to compare equation 4.16 with the relationship predicted for classical hydrodynamic conditions (i.e., for a constant fluid viscosity, and no deformation of the surfaces: a *rigid isoviscous* model). For hydrodynamic lubrication, the minimum film thickness, again for a sphere sliding on a plane, is given by:

$$h_{\min} = 140 \frac{u^2 \eta^2 R^3}{W^2} \quad (4.17)$$

which can be written in terms of the dimensionless groups as:

$$\mathcal{H}_{\min} = 140 \mathcal{U}^2 \mathcal{W}^{-2} \quad (4.18)$$

For the piezoviscous elastic EHL model (equation 4.16) the dependence of film thickness on normal load (i.e., on the dimensionless group $\mathcal{W}$) is very slight, whereas for the hydrodynamic case the dependence is much stronger ($h \propto W^{-2}$). The exponents for sliding speed and viscosity are also much lower in EHL (via the group $\mathcal{U}$) than for hydrodynamic lubrication (0.68 compared with 2).

A power law relationship can be similarly used to indicate how film thickness varies in soft EHL. Here the pressure dependence of lubricant viscosity is not important; we use an *isoviscous elastic* model. An empirical equation, again for a sphere on a plane, is:

$$\mathcal{H}_{\min} = 2.80 \mathcal{U}^{0.65} \mathcal{W}^{-0.21} \quad (4.19)$$

The dependence of film thickness on load is stronger in this case than for hard EHL, and the elastic modulus also has a greater effect. The dependence on viscosity and sliding speed is much the same.

The minimum lubricant film thickness, together with the surface roughness, determines when full fluid film lubrication begins to break down. It is useful to consider the *lambda ratio*

$$\lambda = h_{\min} / \sigma^* \quad (4.20)$$

where σ^* is the r.m.s. roughness of the two surfaces, defined by

$$\sigma^{*2} = Rq_1^2 + Rq_2^2 \quad (4.21)$$

Rq_1 and Rq_2 are the r.m.s. roughness values for each surface (see equation 2.3).

The value of λ provides a measure of how likely, and how severe, asperity interactions will be in lubricated sliding. For $\lambda > 3$, a full fluid film separates the two surfaces, asperity contact is negligible and both the friction and wear should be low. However, many concentrated contacts in machinery operate with $\lambda < 3$. The regime $1 < \lambda < 3$ corresponds to *partial lubrication* or *mixed lubrication*; under these conditions some contact between asperities occurs.

The breakdown of the lubricant film with decreasing film thickness is associated with a sharp rise in friction, as illustrated in Fig. 4.7 for low values of $\eta U/W$. Thermal effects, such as the reduction of viscosity, activation of lubricant additives and even thermal degradation of the lubricant, undoubtedly play an important role under these conditions: a typical rate of viscous energy dissipation in a thin EHL oil film has been estimated at 100 TW m^{-3} , equivalent to dissipating the entire electrical power output of the USA in a volume of 5 litres.

At extremely high loads or low sliding speeds, when the film thickness falls further and λ drops below 1, increasingly severe surface damage (*scuffing*, discussed in Section 5.9) occurs on sliding, and the behaviour of the system depends critically on the properties of *boundary lubricants* and *antiwear additives*, if present. These will be discussed in the next section.

4.6 BOUNDARY LUBRICATION

Under very high contact pressures, or at very low entrainment velocities, hydrodynamic forces are insufficient to maintain a continuous lubricant film between sliding surfaces, and direct contact will occur between the asperities. Even in systems that normally operate with a relatively thick oil film, there will be shut-downs when the surfaces are stationary and the liquid lubricant is squeezed out from the contact. High friction and wear rates will then occur on start-up unless the surfaces are protected in some way. There are also situations in operating machinery, as in the contact between a cam and follower, where the entrainment velocity becomes zero for a short time and the hydrodynamically-generated film collapses. Various additives known as *friction modifiers*, *antiwear additives* and *EP additives* are usually incorporated into oils and greases to protect the surfaces in this regime which is termed *boundary lubrication*. All have different functions and modes of operation.

One group of boundary lubricants, *organic friction modifiers* (OFMs), acts by forming adsorbed molecular films on the surfaces; repulsive forces between the films then carry much of the load, and intimate contact between unprotected metallic asperities with associated adhesion is prevented or limited, and the friction is reduced. [Figure 4.11](#) illustrates the mechanism of operation of a classic OFM, a long-chain carboxylic acid, on metal surfaces. The lubricant molecules are adsorbed with the polar end-groups adhering strongly to the oxide layer present on the metal. The molecular chains tend to align perpendicularly to the surface, stabilized by their mutual repulsion, and form dense layers of hydrophobic chains typically 2–3 nm long. When the two layers come together, most of the normal load is carried by the interaction of the hydrocarbon chains, and there are only small areas of naked asperity contact. The frictional force is lower than for unlubricated sliding, and although some wear does occur, it is substantially less severe than if the surfaces were unprotected. Many different OFMs are available, all characterized by a polar head group that bonds to the metal surface, and an oil-soluble hydrocarbon tail.

The lowering of friction produced by an OFM generally increases with its molecular weight, and hence with the length of the hydrocarbon chain. [Figure 4.12](#) shows how the coefficient of friction for very slow sliding of steel on steel varies with the size of the OFM molecule (in these experiments, straight-chain alcohols and carboxylic acids). For these molecules, appreciable chemical affinity exists between the acid

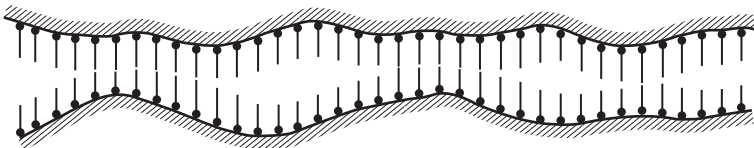


Fig. 4.11 The mechanism of operation of one type of friction modifier. Polar end-groups on hydrocarbon chains bond to the surfaces, providing layers of hydrophobic lubricant molecules that reduce direct contact between the asperities

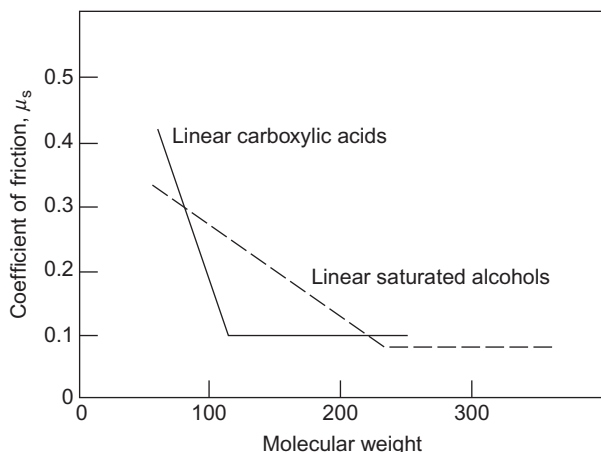


Fig. 4.12 Variation in coefficient of friction with molecular chain length for steel surfaces lubricated with carboxylic acids and alcohols. (data from Bowden, F.P., Tabor, D., 1950. *The Friction and Lubrication of Solids*. Clarendon Press, Oxford)

or hydroxyl group and the metal oxide, and the adsorption process involves a degree of chemical reaction (*chemisorption*).

Some oils naturally contain some molecular species with friction modifying properties; vegetable oils, such as castor oil and rapeseed oil, contain more naturally-occurring OFMs than mineral oils. The presence of these friction modifiers is associated with the qualitative terms *oiliness* and *lubricity*. In practical applications, additional OFMs are often added to the oil. These might be long-chain carboxylic acids (fatty acids) such as stearic acid (=octadecanoic acid, $C_{17}H_{35}COOH$), carboxylic acid amides (e.g., oleic acid amide, $C_8H_{17}CHCHC_7H_{14}CONH_2$), esters such as butyl stearate ($C_{17}H_{35}COOC_4H_9$) or glyceryl monooleate ($CH_2OHCHOHCH_2OCOC_7H_{14}CHCHC_8H_{17}$), or alcohols (e.g., cetyl alcohol = hexadecanol, $C_{16}H_{33}OH$). The structures of stearic acid and glyceryl monooleate are shown in Fig. 4.13. Complete coverage of a surface by a layer of OFM only one molecule thick is, in principle, sufficient for protection. These boundary lubricants are therefore effective in oils at very low overall concentrations. Typically up to 2 wt% is added, since for continuing protection by these and other surface-active additives, enough must be present to ensure that the film is rapidly renewed as soon as it becomes damaged or worn away. The reduction in friction does depend on the lubricant molecule being bonded to the sliding surface; if the temperature is raised thermal activation will cause the additive to desorb from the surface, losing its effectiveness in reducing friction.

Another class of molecules with boundary lubricating properties consists of the so-called *extreme pressure* (EP) additives. These are substances that react with the sliding surfaces under the severe conditions in the contact zone to give compounds with low shear strength, thus forming a lubricating film at precisely the location where it is needed. Originally the high pressures in the contact region were considered to play a role in the action of these additives, hence the name 'extreme pressure'. It is now clear, however, that the transient high temperatures associated with sliding in a very thin EHL film, or with the rapid mechanical interaction of asperities (see Section 3.9), are more important in promoting the surface reaction. Additives of this type usually contain sulfur, phosphorus or, less commonly, chlorine, and function by adsorbing on

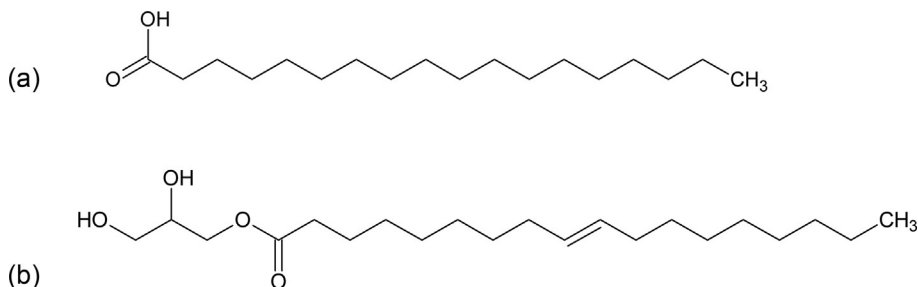


Fig. 4.13 Structures of two organic friction modifiers: (a) stearic acid (octadecanoic acid); (b) glyceryl monooleate

the metal surface and then reacting in the contact zone to form local sulfide, phosphide or chloride films which are easily sheared. EP additives are used where contact conditions are severe: for example, in metalworking processes such as forming, cutting and wire-drawing where contact pressures are high enough to cause plastic flow, flash temperatures are high and fresh, highly reactive metal surfaces are being continually created. They may also be needed to lubricate some gear tooth contacts. Since these compounds function by reacting with the metal surfaces, and some removal of surface material will therefore be inevitable in sliding, the balance is often fine between mechanical damage to the surfaces if insufficient EP protection is provided, and excessive chemical attack by the additives themselves if they are too aggressive. In contrast to the OFMs discussed in the previous paragraph which desorb from the surface as the temperature is raised, EP additives require a high temperature to react and provide protection; lubricant formulations are therefore designed to ensure that surface protection is provided by a package of additives over a continuous range of temperature.

Reduction of wear under boundary lubrication conditions is particularly important in automotive applications, where long lives are needed for engine and transmission components. One particular antiwear additive, zinc dialkyldithiophosphate (ZDDP), is very widely used; apart from its excellent antiwear performance it also acts as an anti-oxidant. The structure of ZDDP is shown in Fig. 4.14(a). Steel surfaces sliding in an oil containing ZDDP form patchy areas of a protective film with a complex structure that provides a low shear strength interface between opposing asperities. The structure of the film and the mechanisms by which it forms have been extensively studied. Tribochemical processes cause decomposition and oxidation of ZDDP at the asperity contacts, leading to a tribofilm with graded composition that consists predominantly of iron oxide and sulfide close to the steel surface, covered with an amorphous region (with zinc cations as glass network-modifiers) of short-chain polyphosphates and thiophosphates, with longer chain polyphosphates in the outer region, as shown schematically in Fig. 4.15. The organic content increases progressively towards the bulk of the oil. The tribofilm thickness is typically about 100 nm.

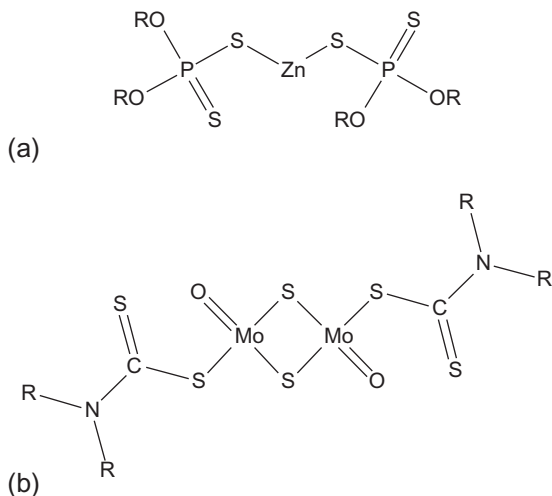


Fig. 4.14 Structures of two common boundary lubricant additives: (a) zinc dialkyldithiophosphate (ZDDP), an antiwear additive; (b) molybdenum dialkyldithiocarbamate (MoDTC), a friction modifier. In both cases, the symbol R represents a short hydrocarbon chain, e.g., C_4H_9 or C_5H_{11} .

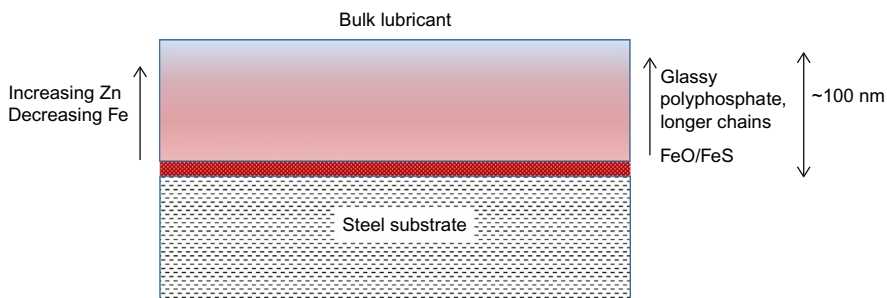


Fig. 4.15 Schematic diagram showing the structure of the layered tribofilm formed under boundary lubrication conditions on a sliding steel surface in the presence of ZDDP

The gradation of composition is matched by a gradation of properties, with the regions closer to the metal being stiffer and having a higher shear strength than the outer material. Growth of the tribofilm from ZDDP is enhanced by both temperature and pressure, so that conditions at sliding asperity contacts favour local film growth and regeneration of the film if it is damaged.

Although antiwear additives such as ZDDP are very effective at preventing direct metallic contact between asperities and thus mitigating wear, they do not reduce the friction and may even increase it. For this reason, they tend to be used in conjunction with friction modifiers. The OFMs discussed above are most useful at low to moderate temperatures, and for automotive engine lubricants an additional approach is needed. Oil-soluble molybdenum compounds are commonly used, such as molybdenum dialkyldithiophosphate (MoDDP) which is the molybdenum analogue of ZDDP,

and molybdenum dialkyldithiocarbamate (MoDTC), with the structure shown in Fig. 4.14(b). The action of these organo-molybdenum friction modifiers results from their decomposition to form MoS_2 at the sliding interface, which as we have seen in Section 3.7 has a lamellar structure with low shear strength. The use of ZDDP and MoDTC together appears to be particularly beneficial in reducing both friction and wear, and provides an example of the complex interactions that are common between lubricant additives. Additives with no obvious tribological function, such as anti-oxidants, dispersants and detergents, modify and in some cases may usefully enhance the action of antiwear additives and friction modifiers.

Most boundary lubricants have been developed to function well with the ferrous surfaces (steel or cast iron) that are common in many engineering applications. It should not be assumed that they will be so effective for non-ferrous materials because the chemical reactions by which they function have been optimized for ferrous materials. For example, ZDDP is not so effective as an antiwear agent for aluminium alloys as for steels, and the lubrication of ceramics and coated materials poses a separate set of challenges.

4.7 SOLID LUBRICATION

Solid materials which exhibit low coefficients of friction may be used as lubricants in preference to liquid or gas films for several reasons. One component of a bearing may be composed of or coated with the solid lubricant, or made from a composite containing it, in order to produce a ‘self-lubricating’ system (i.e., a system with a low interfacial shear strength) which needs no external source of lubricant during its lifetime. Such bearings offer obvious savings in maintenance and lubrication costs. Under certain conditions solid lubrication may be the only feasible system, for example in food-processing machinery or healthcare equipment where contamination by a liquid lubricant must be avoided, or in space (or other high vacuum environments) where a liquid lubricant would evaporate. At high temperatures, liquid lubricants decompose or oxidize; suitable solid lubricants can extend the operating temperatures of sliding systems beyond 1000°C , while maintaining relatively low coefficients of friction.

We saw in Sections 3.7 and 3.8 that graphite, molybdenum disulfide (MoS_2) and polytetrafluoroethylene (PTFE) under appropriate conditions all exhibit low coefficients of friction; these three materials are used in the great majority of applications requiring solid lubricants. They can be applied as thin coatings, incorporated into composite bearing materials (see Section 9.2.3), or, in the case of graphite and PTFE, used as bulk bearing materials. Graphite and MoS_2 can also be applied as replenishable lubricants in the form of dry powders. They are also sometimes incorporated into lubricating oils and greases as suspensions of fine particles; they come into play under severe contact conditions and provide boundary lubrication.

All three of these materials, however, suffer from limitations. Their use at high temperature is in each case limited by decomposition or oxidation in air: graphite to $\sim 500^\circ\text{C}$, MoS_2 to $\sim 300^\circ\text{C}$, and PTFE to $\sim 250^\circ\text{C}$. In addition, the lubricating properties of graphite depend on the presence of condensable vapours (see [Section 3.7](#)), while high atmospheric humidity adversely affects the friction and wear rate of MoS_2 .

Several other materials have been proposed as solid lubricants, and are used in certain applications. They may be broadly classified as organic polymers, lamellar solids, other inorganic compounds, and soft metallic films.

The only polymer used at all widely as a lubricant, in thin films on harder substrates, is PTFE. Several other polymers, notably nylon, polyimides, acetals, ultra-high molecular weight polyethylene (UHMWPE) and polyetheretherketone (PEEK), show relatively low friction and good wear resistance when sliding against hard counterfaces (usually metallic). These are generally used, as PTFE is also used, as components of composites or in bulk form, as bearing materials rather than as lubricants.

Many other materials apart from graphite and MoS_2 have lamellar structures, and although this provides no guarantee that the compound will act as a solid lubricant, many do. Transition metal dichalcogenides with structures analogous to MoS_2 have been investigated: NbS_2 , TaS_2 , WS_2 and the diselenides of the same metals all act as solid lubricants. None is used as widely as MoS_2 , however, although some offer lower friction and have other advantages. For example, WS_2 (tungsten disulfide) resists oxidation better than MoS_2 , and can lubricate at significantly higher temperatures.

Other inorganic compounds with layer structures vary in their lubricating action. Some, such as CdCl_2 , CdI_2 and CoCl_2 , typically show $\mu < 0.1$ and are potentially valuable lubricants, while others show high friction and are of no interest for lubrication. Hexagonal boron nitride (BN), which is isoelectronic with carbon and has a structure similar to graphite, has good oxidation resistance, but shows rather high friction in air ($\mu = 0.2\text{--}0.4$). Boric acid (H_3BO_3) also has a lamellar structure, and can form by the reaction of borided steel or boron carbide with oxygen and water vapour to create a self-lubricating system; it can also be used as a dry powder. Like graphite, both boron nitride and boric acid are sensitive to the presence of certain vapours which react with sites on the edges of their molecular planes (see [Section 3.7](#)): a relatively low vapour pressure of heptane, for example, is sufficient to halve the coefficient of friction for boron nitride, while water vapour has a similar beneficial effect for boric acid.

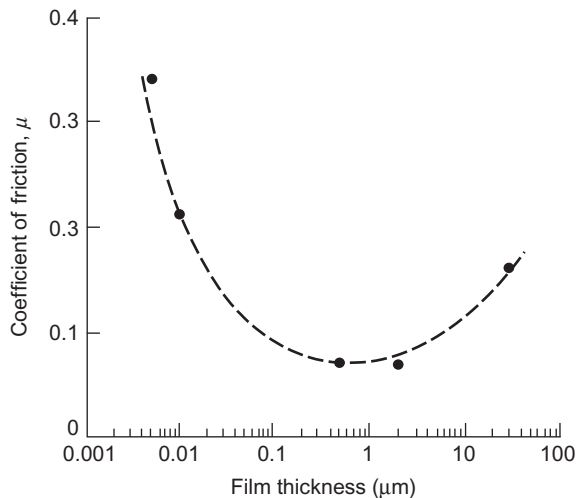
Certain inorganic solids without lamellar structures also possess low shear strengths, and some of these are useful as lubricants, particularly at high temperatures. A wide range of compounds has been investigated: calcium and barium fluorides (CaF_2 and BaF_2), lead oxide and sulfide (PbO and PbS) and boric oxide (B_2O_3) can be used as high temperature lubricants either in bulk composite form or incorporated into coatings. Ceramic-bonded composite films containing CaF_2 can give a coefficient of friction as low as 0.1 at temperatures up to 1000°C . It is not easy, however, to find a lubricant which will perform well at low temperatures as well as at high.

Thin films of soft metals constitute another class of solid lubricants. As was shown in [Section 3.4.3](#), the coefficient of friction for sliding over a thin soft metal layer on a harder substrate may be simply estimated as

$$\mu = \tau_i / p_o \quad (4.22)$$

where τ_i is the shear strength of the film and p_o is the indentation pressure of the substrate material. The friction force is determined by the shear strength of the film material, whereas the normal load is carried by deformation of the substrate; hence the two quantities τ_i and p_o can be varied independently, leading to a low value of μ . Lead, tin, indium, silver and gold have all been widely and successfully used as solid lubricant films, usually deposited by electroplating or a vapour deposition process (see [Chapter 7](#)). Very thin coatings (as thin as 0.1 μm) will suffice to lubricate suitably smooth hard substrates, and metallic films of this thickness can be used to lubricate rolling element bearings for use in space or ultra-high vacuum, or for high temperature operation. Coefficients of friction can be very low (for example, 0.06 for In, 0.09 for Sn and 0.20 for Ag films on a steel substrate with a steel slider) although the sliding friction for bulk specimens of these soft metals would in contrast be extremely high. As the thickness of the metal film is varied, so the coefficient of friction changes. [Figure 4.16](#) shows the behaviour of a curved steel slider moving over indium films deposited on hard tool steel. There is an optimum film thickness for minimum friction. Thinner films cannot prevent asperity contacts between the base metals, while for thicker films the frictional behaviour becomes more similar to sliding on a bulk specimen of the soft metal.

Fig. 4.16 The coefficient of friction for a steel sphere sliding against a tool steel surface coated with a thin film of indium metal, plotted against the thickness of the indium film. (data from Bowden, F.P., Tabor, D., 1950. *The Friction and Lubrication of Solids*. Clarendon Press, Oxford, by permission of Oxford University Press)



SOURCES AND FURTHER READING

- Arnell, R.D., Davies, P.B., Halling, J., Whomes, T.L., 1991. Tribology: Principles and Design Applications. Macmillan.
- ASM, 1992. ASM Handbook, Vol. 18, Friction, Lubrication and Wear Technology. ASM International.
- Bhushan, B. (Ed.), 2001. Modern Tribology Handbook. CRC Press.
- Booser, E.R. (Ed.), 1983. CRC Handbook of Lubrication, vols. 1 and 2. CRC Press.
- Cameron, A., 1981. Basic Lubrication Theory, third ed. Ellis Horwood.
- Hamrock, B.J., 1994. Fundamentals of Fluid Film Lubrication. McGraw Hill.
- Jones, M.H., Scott, D., 1983. Industrial Tribology, Tribology Series No. 8. Elsevier.
- Lansdown, A.R., 1982. Lubrication: A Practical Guide to Lubricant Selection. Pergamon Press.
- Mortier, R.M., Fox, M.F., Orszulik, S.T. (Eds.), 2010. Chemistry and Technology of Lubricants. Springer.
- Neale, M.J. (Ed.), 1993. Lubrication: A Tribology Handbook. Butterworth Heinemann.
- Spikes, H., 2015. Friction modifier additives. Tribol. Lett. 60, 5.
- Stachowiak, G.W. (Ed.), 2005. Wear—Materials, Mechanisms and Practice. John Wiley.
- Stachowiak, G.W., Batchelor, A.W., 2014. Engineering Tribology, fourth ed. Butterworth-Heinemann.
- Williams, J.A., 2005. Engineering Tribology. Cambridge University Press.

QUESTIONS FOR CHAPTER 4

Question 4.1

- (a) Explain how the assumptions used to predict film thickness in ‘hard elastohydrodynamic’ lubrication (hard EHL) differ from those for ‘hydrodynamic’ lubrication.
- (b) A steel ball, 20 mm in diameter, slides over a flat polished steel plate at a speed of 2 m s^{-1} under a normal load of 10 N. The contact region is flooded with mineral oil with a viscosity of 10 mPa s. Estimate the minimum thickness of the oil film beneath the ball if the conditions are assumed to be (a) hydrodynamic lubrication; and (b) hard EHL. The ball has a roughness Rq of $0.02 \text{ }\mu\text{m}$ and the plate has a roughness Rq of $0.03 \text{ }\mu\text{m}$. Comment on the extent of asperity contact between the ball and the plate. What effect on the frictional force would result from the addition of a friction modifier to the oil?

[For the steel, Young’s modulus $E = 210 \text{ GPa}$ and Poisson’s ratio $\nu = 0.3$. For the oil, the pressure coefficient of viscosity $\alpha = 2 \times 10^{-8} \text{ Pa}^{-1}$]

Question 4.2

Explain the term ‘viscosity index’. Give two examples of applications where a lubricating oil with a high viscosity index would be desirable. Under what conditions might a low viscosity index be acceptable?

Question 4.3

Water has a dynamic viscosity of 1 mPa s at 20°C and shows almost no variation of viscosity with hydrostatic pressure (i.e., the pressure coefficient of viscosity $\alpha \approx 0$). Suggest and discuss two applications where water might be an effective lubricant, instead of a hydrocarbon oil. In your answer, consider other properties of the lubricating fluid as well as its viscosity.

Question 4.4

A hydrodynamically lubricated journal bearing, with width equal to its diameter, has the following dimensions: journal diameter 50 mm, mean radial clearance $25 \text{ }\mu\text{m}$. The bearing is lubricated with an oil with a dynamic viscosity of 10 mPa s. If the shaft is rotating at 1500 revolutions per minute and the minimum acceptable oil film thickness is $10 \text{ }\mu\text{m}$, what is the maximum radial load that can be carried by the bearing? For these conditions, estimate the effective coefficient of friction in the bearing and the power dissipated in friction. What happens to this energy?

Question 4.5

Five samples of a hydrocarbon oil were formulated with the following additives:

Oil A: no additive

Oil B: lauric acid (dodecanoic acid, $C_{11}H_{23}COOH$)

Oil C: sulfur-containing EP additive

Oil D: ZDDP

Oil E: ZDDP+MoDTC

Each oil was used to lubricate a reciprocating sliding contact between a steel ball and a steel plate, and the friction and wear rate (of the ball) were measured in tests at two different temperatures, T . The results were as follows.

	$T = 40^{\circ}\text{C}$	$T = 200^{\circ}\text{C}$
Test series I	High friction+low wear	High friction+low wear
Test series II	Low friction+low wear	High friction+high wear
Test series III	High friction+high wear	High friction+high wear
Test series IV	Low friction+low wear	Low friction+low wear
Test series V	High friction+high wear	Low friction+low wear

Which oil was used in which test series? In each case explain the reasoning behind your answer.

Question 4.6

In order to increase the efficiency of petrol (gasoline) or diesel-powered road vehicles there is a trend to reduce the viscosity of the engine oil. Explain how this is beneficial. Discuss the implications of lower oil viscosity for the lubrication of the following components: (a) crankshaft main bearings; (b) cam/follower contacts in the valve train; (c) piston ring/cylinder bore contacts at the top of the stroke. (You may need to do some research to understand the function of these components, and the conditions they experience. [Section 9.3.2](#) will also be helpful.)

Question 4.7

Identify three examples of applications where a solid lubricant would be preferable to a liquid lubricant. For each example, explain the advantages of using a solid lubricant and suggest one or more suitable candidates. Give reasons for your suggestions.

5.1 INTRODUCTION AND TERMINOLOGY

This chapter is concerned with the wear that occurs when two solid surfaces slide over each other. In most practical applications sliding surfaces are lubricated in some way, and the wear that occurs is then termed *lubricated sliding wear*. In some engineering applications, however, and in many laboratory investigations, the surfaces interact in air without a lubricant. The resulting wear is then often called *dry sliding wear*, although it usually takes place in ambient air of appreciable humidity. We shall consider both dry and lubricated wear in this chapter, but we shall deliberately exclude the wear associated with the presence of hard particles, either as a separate component between the sliding surfaces or as an element of the structure of one or both surfaces. This type of wear is termed *abrasive wear*, and is discussed in [Chapter 6](#), together with other kinds of wear caused by hard particles. The distinction between abrasive wear and sliding wear employed here is somewhat artificial, and certainly not respected by nature, since under some conditions sliding wear can generate debris which then causes further wear by abrasion. The boundary between different types of wear is not a distinct one.

The term *adhesive wear* is sometimes used to describe sliding wear, but can be misleading. As we shall see, adhesion plays an important role in sliding wear, but it is only one of the several physical and chemical processes that may be involved. We shall therefore use ‘sliding wear’ as a general term in preference. The words *scuffing*, *scoring* and *galling* are often associated with severe sliding wear. *Scuffing* occurs in inadequately lubricated metallic contacts (as discussed in [Section 5.9](#)), and is associated with macroscopically observable changes in the surface texture, with features related to the direction of relative motion. *Scoring* involves the formation of grooves and scratches in the direction of sliding and can also imply scratching by abrasive particles (described in [Section 6.3.1](#)). *Galling* represents a more severe form of scuffing, due to local welding, and is associated with gross surface damage. The word often refers to damage resulting from unlubricated sliding at low speeds, characterized by severely roughened surfaces and transfer or displacement of large fragments of material. Galling may occur in nominally lubricated systems when the lubricant film breaks down, and can be followed by seizure of the surfaces and consequent gross failure of the sliding system.

The complexity in terminology reflects the complexity of the processes that underpin wear in sliding. As two surfaces slide against each other, there are three main issues to be considered:

Stresses and damage: as surfaces slide against each other, stresses (both normal and tractional) result in mechanical damage to the materials, typically by plastic deformation or by fracture; this damage may occur at the asperity level, or at a much more macroscopic level.

Thermal effects: as discussed in detail in [Section 3.9](#), when surfaces slide against each other, frictional work is dissipated as heat, resulting in temperature increases in the bodies, both at the asperity scale and macroscopically. Macroscopic changes in temperature may influence the properties of the subsurface material (such as yield strength), but the temperature increases of individual asperities at the sliding surfaces are expected to be very much higher, and will strongly influence both material properties and the nature and rates of chemical reaction with the surrounding environment (be that a gaseous environment or a formulated lubricating oil).

Chemical reactions and interactions of the surfaces: the surfaces of all materials generally interact with the environment in which they are placed, by the adsorption of atoms or molecules or by chemical reaction (see [Appendix B](#)). For metals in air, this reaction typically involves the formation of a thin layer of metal oxide on the surface (with the presence of the oxide itself restricting further growth in thickness of the oxide). In sliding contacts, these reactions may take place at a much higher rate due to the high asperity flash temperatures generated in sliding (see [Section 3.9](#)) and by the fact that the reaction layers (such as oxides) that are formed (and which would normally restrict further chemical reaction of the material with the environment) can be mechanically removed by the sliding action, resulting in the exposure of clean underlying material which will then rapidly react with the environment. In the case of metals, this enhanced rate of oxidation (by the continual formation and removal of oxide) can result in the formation of thick beds of oxide debris in sliding contacts; these beds often exert significant influence on subsequent behaviour and rates of wear. As already indicated, the action of sliding of two surfaces against each other often results in the disturbance or removal of existing surface layers (adsorbed species or oxides), resulting in the nature of the contact changing from contact between surfaces covered with adsorbed or reacted material, to contact between the underlying materials themselves. Depending on the nature of these two materials, this may result in chemical reactions or even in mechanical bonding.

The nature and rate of wear will thus be controlled by the stresses and mechanical damage, thermal effects, and chemical reactions and interactions of the surfaces. These are all influenced by both the physical and chemical responses of the surfaces. Moreover, these influences cannot be treated independently, since the response of the contact depends on the complex interactions between them which can change with time as the contact develops. Indeed, it is this very complexity that has made the development of an all-encompassing model for sliding wear so elusive. In seeking to understand damage by sliding, it is critically important to understand both the mechanical and chemical nature of the contact (and the way that each is influenced by the other), and to base hypotheses and models upon an understanding of these processes built on careful microstructural observation of the nature of the degradation.

Laboratory tests that aim to evaluate material performance relevant to a particular service environment must use conditions appropriate to that environment, so that the complex balance of physical and chemical interactions is representative of the service conditions. This is a complex process, but by recognizing this complexity and being aware of the issues, the tribologist can be aware of the limitations of laboratory test programmes and can design them to ameliorate the risks of deploying new materials

into tribologically aggressive environments. This chapter outlines a framework by which the basic underlying physical processes which govern wear in sliding can be described, and thus provides a foundation from which rates and mechanisms of wear can be understood and influenced.

5.2 TESTING METHODS

Many different experimental methods have been used to study sliding wear. Laboratory investigations of wear are usually carried out either to examine the mechanisms by which wear occurs, or to simulate practical applications and provide useful design data on wear rates and coefficients of friction. For both purposes, selection, control and measurement of all the variables that may influence wear are very important. Wear rate and friction are often critically dependent on the sliding conditions; even apparently minor changes in conditions can lead to radical changes in the dominant mechanism and associated rate of wear. Careful choice of conditions, and then their close control and monitoring, is essential if the results of a test are to be useful either as a simulation of a practical application, or for wider scientific purposes.

In a sliding wear test, the main aspects that must be considered are:

- the materials of which each of the two bodies in the couple are made, and in some cases the method of surface preparation;
- the test geometry including both the shape and dimensions of the samples;
- the applied load and contact pressure;
- the sliding speed; and
- the test environment (the nature of the environment surrounding the contact, including its temperature).

In designing a test, each of these needs to be considered and justified. It is important to remember that behaviour of a tribological contact is related to behaviour of the whole *system* and not just of a *material*, and that one can never define the tribological behaviour of a material without a full description of the system. It is for this reason that an understanding of the underlying physical mechanisms of damage and wear is essential, since it is this which may allow the conclusions to be extrapolated reliably beyond the strict confines of the test conditions.

Figure 5.1 shows the geometrical arrangements employed in several common types of sliding wear testing apparatus. The word *tribometer*, first used in 1774 for an instrument intended to measure friction, is sometimes used for such apparatus; more recently the term *tribotester* and its associated verb have been coined.

The most common test rigs employ a pin or block pressed against a disc, either on the flat face (Fig. 5.1a) or on the rim (b), a pin loaded against a plane sample with reciprocating motion (Fig. 5.1c), or rotating discs, either rim-loaded or face-loaded (which produce relative sliding in all cases where $\omega_1 \neq \omega_2$) (Fig. 5.1d and e). In most of the geometries used for sliding wear testing, the contact is asymmetrical, i.e., the nature of the contact conditions experienced by one body in the couple is not the same as that experienced by the other body. The key difference is normally that a point on the surface of one body is always in contact with the other surface throughout the test,

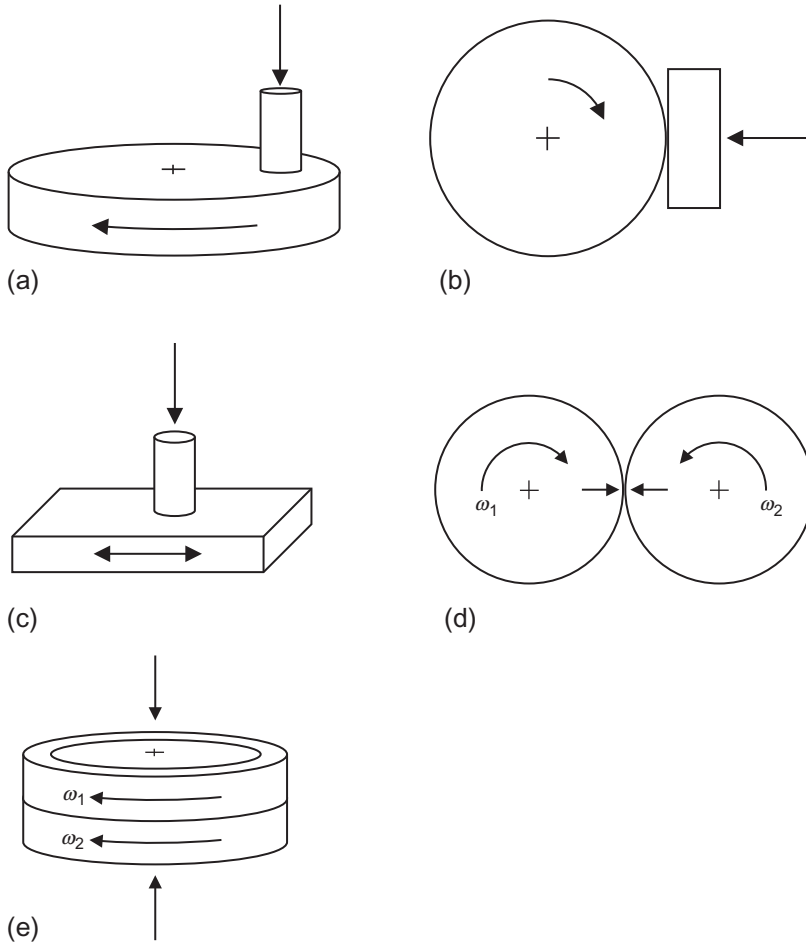


Fig. 5.1 Geometries commonly employed in sliding wear testing; (a) pin-on-disc; (b) block-on-ring; (c) pin-on-reciprocating plate; (d) twin disc; (e) ring-on-ring

whereas for the other body, a point on the surface is only in intermittent contact with the other body. This asymmetry causes many differences, such as in the number of interactions per asperity on the two surfaces, the thermal profile, and the ways that chemically-formed species are created by interactions with the environment and removed as debris by mechanical interactions. Except for the case illustrated in [Fig. 5.1\(e\)](#) (and the specific case of [Fig. 5.1\(d\)](#) when $\omega_1 = -\omega_2$), the methods shown in [Fig. 5.1](#) are all asymmetric, and therefore the two bodies, even of the same material, will almost certainly experience different rates of wear. [Figure 5.1\(e\)](#) illustrates a ring-on-ring geometry where all points on the contact surfaces of both bodies remain in contact throughout the test; as with the pin-on-disc test ([Fig. 5.1a](#)), the sliding velocity varies across the contacting surface with distance from the axis of rotation, but the

effect of this can be minimized by making the width of the ring surface small compared with the radius.

In the pin-on-disc test (Fig. 5.1a), the disc can be large enough for several wear tests to be conducted on the same disc at different radii. However, the conditions in these tests will not be the same, even if the rotational velocity and the number of revolutions are adjusted between tests to ensure that the sliding speed and the total sliding distance remain constant. For a point on the disc (which is in intermittent contact), the time between passages of the pin will be different, which will change the thermal history and the time for any other changes that occur during this period. The number of times that it interacts with the pin during the test will also be different.

In asymmetric arrangements, one component of the mating pair, commonly the pin or block, is usually treated as the *specimen* and is the component for which the wear rate is measured, while the other, often the disc, flat or ring, is called the *counterface*. However, as we have noted earlier, wear is a system response, and in wear testing, the wear of the counterface is often just as important as the wear of the specimen itself. For example, in seeking to reduce wear by selection of different specimen materials, little is gained if the overall wear of the system (specimen and counterface together) is increased.

Additionally, contacts may be described as *conformal* or *counterformal* (sometimes called *concentrated*). *Conformal* contact initially occurs over an extended nominal contact area (e.g., with a flat-ended pin-on-disc (Fig. 5.2a), or a conforming block-on-ring (Fig. 5.2b)), whilst a counterformal or concentrated contact typically occurs initially only at a point or line (e.g., for a spherically-ended pin-on-disc (Fig. 5.2c), or a plane block-on-ring (Fig. 5.2d)). In counterformal contacts, much higher initial nominal contact pressures are achieved than is generally the case with conformal contacts; however,

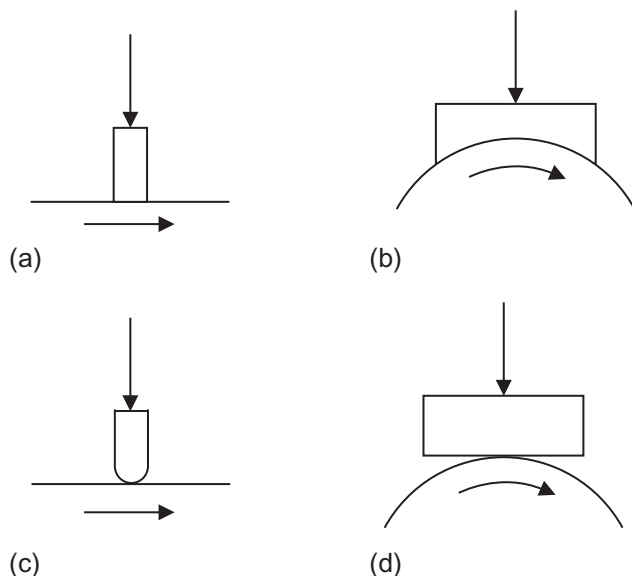


Fig. 5.2 Geometries of (a, b) conformal contacts and (c, d) counterformal contacts

wear will change the macroscopic contact geometry, and thus lead to changes in the contact pressure that may influence the subsequent rates of wear. Changes in nominal contact pressure of more than two orders of magnitude are not uncommon for a spherically-ended pin-on-flat contact during a wear test, although the average asperity contact pressure is broadly independent of such changes (see [Section 2.5.6](#)). Whilst such undesirable changes are avoided by the use of conformal contacts, there are issues associated with alignment of the specimen faces in conformal contacts which are absent (or much less severe) if a counterformal contact geometry is utilized.

[Figure 5.3](#) illustrates the case of a pin-on-flat geometry with a small angular misalignment, $\Delta\theta$. This results in a poorly-defined initial contact pressure which will be much higher than the intended contact pressure; moreover, it may vary between nominally identical tests, and potentially lead to low levels of repeatability. As the contact wears, the desired conforming geometry will eventually be reached. If the amount of wear required to achieve the conforming geometry is only a small fraction of the wear in the whole test, then these issues may be insignificant; however, in tests with thin surface-engineered layers (such as PVD coatings discussed in [Section 7.4.3](#)) or for materials with a very low wear rate (such as some ceramics), this may not be the case. The use of a defined counterformal contact (such as a spherically-ended pin-on-flat) may then be the best compromise.

The dimensions of specimens in laboratory wear testing usually lie in the range from millimetres to tens of millimetres; in asymmetric tests, the pins and blocks are typically smaller than 25 mm in size, while the counterface rings and discs are typically several tens of millimetres in diameter. Loads over a wide range are employed, from fractions of a newton to several kilonewtons. Sliding speeds similarly cover a broad span, from fractions of a millimetre per second to tens or even hundreds of metres per second in specialized apparatus. The variables that affect friction and wear, and which must therefore be controlled in wear testing, range from the most obvious such as speed and load to the less obvious, such as specimen size and orientation.

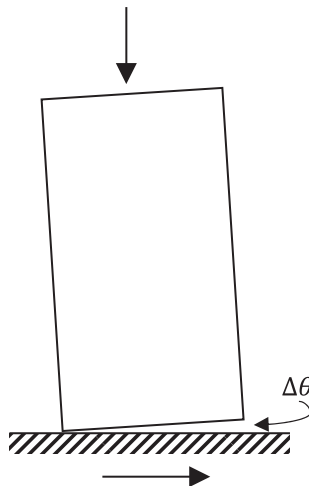


Fig. 5.3 Illustration of misalignment in the setup of a conformal contact

Several sliding wear test methods are the subject of national and international standards. Examples are the tests for sliding wear with a block-on-ring geometry (ASTM G77), wear with a pin-on-disc geometry (ASTM G99), galling with a rotating pin-on-flat geometry (ASTM G98) and fretting with a ball-on-flat geometry (ASTM G204). Some standards are specific to particular specimen types (such as CEN EN 1071-12 which addresses reciprocating wear testing of ceramic coatings and ASTM G176 which addresses sliding wear testing of polymeric materials).

As we shall see in the next section, a simple model for wear suggests that the volume of material removed from a sliding body should be proportional to the distance slid and to the normal load carried by the contact (which is equivalent to saying that the wear depth is proportional to the contact pressure over the contact region). Whilst the proportionality with distance slid is often observed in the steady state, initial transients often exist, and these need to be dealt with appropriately in any calculations of the wear rate (see [Appendix C](#)). Wear is usually measured either by removing the specimen from the test apparatus and determining the mass or volume change, or by continuously measuring its position during the test with a displacement transducer and deducing the wear from its change in dimensions. In the case of a counterformal contact, wear causes the nominal area of contact to increase, and the wear volume can therefore be quantified by measuring the size of the wear scar on the specimen; this method of determining wear is commonly used, for example, in the spherically-ended pin-on-disc ([Fig. 5.2c](#)). Obviously, in any test where a specimen is removed for determination of wear and then replaced in the test apparatus with a view to continuing the test, care must be taken to replace the specimen in exactly the same position and orientation as previously; this is a non-trivial requirement, and interrupted tests should be validated by comparison with tests that have been run in the same way and for the same duration, but without the interruptions.

It is generally advisable to measure and record the friction force continuously during a wear test. This is usually done by measuring the tangential force or the torque on one of the bodies. A continuous friction record not only provides numerical values for μ , but also allows changes in sliding behaviour to be monitored (for example, by a rise or fall in μ , or a change from a smooth friction trace to an irregular one). These changes often herald changes in surface nature or topography, or in the mechanism of wear. Running-in, for example, can be explored in this way, as can the breakdown of protective oxide or lubricant films. Knowledge of such changes is sometimes more valuable than measurement of absolute wear rate. For some purposes, continuous measurement of electrical contact resistance between the specimen and counterface is informative, since it provides a means of assessing the thickness of oxide or EHL films. Electrical capacitance measurements can provide another source of such information.

Wear under sliding conditions depends on the distance slid, but also to some extent on both the sliding speed and the duration of the test, independently. The sliding speed affects the rate of frictional energy dissipation, and hence the temperature at the interface as discussed in [Section 3.9](#). It certainly cannot be assumed that one wear test will produce the same results as another of half the duration at twice the sliding speed, since abrupt transitions in wear mechanism and rate may occur as the sliding speed is changed, as discussed in [Appendix C](#).

Wear rate also depends on the nominal contact pressure between the sliding surfaces, and, as is shown in the following sections, transitions are commonly induced

by changes in contact pressure. The linear dimensions of the specimen are also important, independently of the contact pressure, since wear debris formed near the leading edge of a long specimen will have more influence during its passage through the contact zone than it would have with a shorter specimen.

Apart from the major variables of normal load, contact area, sliding speed and testing time, several other factors must also be considered and monitored in wear testing. The ambient testing temperature is important through its influence on the mechanical properties of the materials and on thermally activated chemical processes, although these may often be dominated by frictionally-generated temperature rises. In lubricated systems, temperature will also be important through its effect on lubricant viscosity. Atmospheric composition is extremely important; reactive components such as water vapour and oxygen strongly influence wear rates and mechanisms in all classes of materials. The orientation of the apparatus can also, perhaps surprisingly, affect the results of testing: different behaviour may be seen if wear debris falls readily away from the contact area under gravity, rather than being retained on the counterface.

All the types of test discussed above can be applied either unlubricated ('dry') or lubricated, and the lubricant can be introduced in various ways. In a lubricated system the supply of lubricant, and the details of the pressure distribution within the lubricant film and its resulting thickness, are important factors.

The comprehensive list of influences given above may suggest that it is never possible to produce a valid laboratory simulation of a practical application, and that the only valid wear test is a service trial. This is emphatically not the case, but the tribologist must always be aware of the possible consequences of the test being an entirely inappropriate representation of the actual conditions of service. Contact stresses, thermal conditions, sliding speeds and chemical environment are all vital factors in any wear test, and it is critical to ensure, by measurement of friction and by close examination of the worn surfaces and wear debris after the test, that the mechanism of wear is the same in the test as in the service application; this should be the basis of the justification of the selected test conditions. Only then can the results of a laboratory test be applied with confidence to a practical problem.

5.3 SIMPLE THEORY OF SLIDING WEAR: THE ARCHARD WEAR EQUATION

When two surfaces in contact slide over each other, one or both of the surfaces will suffer wear. A simple theoretical analysis of this type of wear, due originally to Holm and Archard, will be given here. By virtue of its simplicity, it highlights the main variables that influence sliding wear; it also yields a method of describing the severity of wear by means of the wear coefficient, K , which is valuable and widely used. Although the model was originally developed for metals, it can provide some insight into the wear of other materials as well.

The starting point for the model is the assumption, discussed in detail in [Chapter 2](#), that contact between the two surfaces will occur where asperities touch, and that the

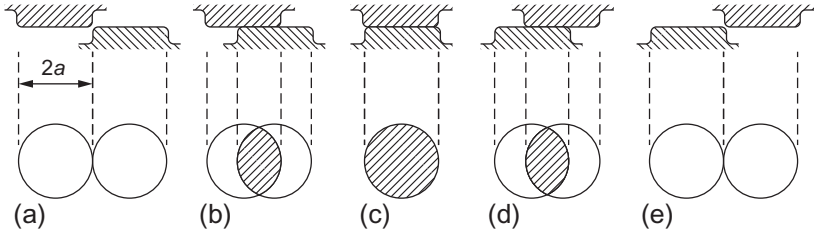


Fig. 5.4 Schematic diagram showing the evolution of a single contact patch as two asperities move over each other (after Archard, J.F., 1953. *J. Appl. Phys.* 24, 981–988)

true area of contact will be equal to the sum of the individual asperity contact areas. This area will be closely proportional to the normal load, and it can be assumed that under most conditions, for metals at least, the local deformation of the asperities will be plastic (see [Section 2.5.6](#)).

[Figure 5.4](#) shows a single asperity contact, which we assume to be circular in plan view and of radius a , at various stages of evolution during sliding. In [Fig. 5.4\(c\)](#), the contact has reached maximum size, and the normal load supported by it, δW , will be given by

$$\delta W = P\pi a^2 \quad (5.1)$$

where P is the yield pressure for the plastically deforming asperity (which will be close to its indentation hardness H , see [Section 2.5.2](#) and [Appendix A](#)).

As sliding proceeds, the two surfaces become displaced as shown in [Fig. 5.4\(d\)](#) and [\(e\)](#), and the load originally borne by the asperity is progressively transferred to other asperity junctions that are in the process of forming elsewhere in the contact. Continuous sliding leads to the continuous formation and destruction of individual asperity contacts. Wear is associated with the detachment of fragments of material from the asperities, and the volume of each wear fragment will depend on the size of the asperity junction from which it originated. It is assumed that the volume of material removed by wear, δV , will be proportional to the cube of the contact dimension a , which implies that the shape of the wear particle should be independent of its size.

For argument's sake, the volume can be taken to be that of a hemisphere of radius a (although this should not be taken to imply that the worn particles are necessarily hemispherical in shape), giving

$$\delta V = \frac{2}{3}\pi a^3 \quad (5.2)$$

Not all asperity contacts give rise to wear particles. Let us suppose that only a proportion κ does so. The average volume of material δQ worn away per unit sliding distance due to sliding of the one pair of asperities through a distance $2a$ is therefore given by

$$\delta Q = \frac{\kappa \delta V}{2a} = \frac{\kappa \pi a^2}{3} \quad (5.3)$$

and the overall wear rate Q arising from all the asperity contacts is the sum of the contributions over the whole real area of contact:

$$Q = \sum \delta Q = \frac{\kappa}{3} \sum \pi a^2 \quad (5.4)$$

The total normal load W is given by

$$W = \sum \delta W = P \sum \pi a^2 \quad (5.5)$$

and hence

$$Q = \frac{\kappa W}{3P} \quad (5.6)$$

It is convenient to combine the factor of $1/3$ into the constant of proportionality, by putting $K = \kappa/3$, and to assume that $P = H$, the indentation hardness. We can then rewrite the equation in the form:

$$Q = \frac{KW}{H} \quad (5.7)$$

This equation, relating the volume worn per unit sliding distance, Q , to the macroscopic quantities W , the normal load, and H , the hardness of the softer surface, is often called the *Archard wear equation*. The dimensionless constant K is usually known as the *wear coefficient*.

The wear coefficient K is of fundamental importance, and provides a valuable means of comparing the severity of wear processes in different systems. However, for engineering applications the quantity K/H is often more useful. This is given the symbol k and is normally termed the *specific wear rate*; it represents the volume of material removed by wear per unit distance slid per unit normal load on the contact, and is often quoted in units of $\text{mm}^3 \text{m}^{-1} \text{N}^{-1}$. The measure of wear provided by k is particularly helpful for comparing wear rates in different classes of material. In some materials, elastomers for example, there are basic problems with the use of the dimensionless coefficient K since the plastic indentation hardness H cannot be readily defined (see [Appendix A](#)).

Although we have so far defined K in terms of the probability of each asperity interaction resulting in the production of a wear particle, this is not the only possible interpretation. It may also be taken to reflect the number of cycles of deformation required by each asperity before a fragment of material is removed by a fatigue process, or alternatively it can be correlated with the size of the wear particle produced by every asperity contact.

From a fundamental viewpoint, equation 5.7 follows inevitably from the assumption that the wear rate (volume of material removed per unit sliding distance) depends only on the normal load and on the hardness or yield strength of the softer surface; it is

the only dimensionally correct relationship that is possible between Q , W , and H . The volume removed per unit sliding distance, Q , has the dimensions of area, and the quantity W/H also represents an area of basic importance in contact processes: the real area of contact for fully plastic asperities. K can therefore be interpreted as the ratio between these two areas.

Equation 5.7 implies that if K is a constant for a given sliding system, then the volume (or mass) of material lost by wear should be proportional to the distance slid (i.e., Q should be constant), and if the normal load W is varied then the wear rate should vary in proportion.

It is found by experiment that for many systems, the loss of material by wear is indeed proportional to the sliding distance (and so, for sliding at constant velocity, to the time). Transient behaviour is sometimes observed at the start of sliding, until equilibrium surface conditions have become established (see [Appendix C](#)); the wear rate during this initial *running-in* period may be either higher or lower than the corresponding steady-state wear rate, depending on the nature of the running-in process. Given the existence of such transients, care must be exercised in defining the wear rate from measurements of wear as outlined in [Appendix C](#).

Strict proportionality between the wear rate and the normal load is less often found. Although for many systems the wear rate varies directly with load over limited ranges, abrupt transitions from low to high wear rate, and sometimes back again, are often seen with increasing load. [Figure 5.5](#) illustrates this behaviour. In this example, the brass specimen exhibits a rapid increase in wear rate above a critical load, although below and above that load its behaviour is described well by the Archard equation. The reasons for such transitions will be explored in detail in the following sections; they typically correspond to substantial changes, by factors of 100 or more, in the wear coefficient K with normal load.

No mention has so far been made of the influence on wear rate of the apparent area of contact, or of the sliding velocity. According to equation 5.7, the wear rate Q should be independent of these factors. For many purposes this assertion is true, although in some systems, as discussed below, sharp transitions in wear rate are seen with increasing sliding velocity or load.

Some typical values of K , for dry sliding wear in pin-on-ring tests for a range of materials are shown in [Fig. 5.6](#). It is instructive to contrast these values with the coefficients of friction for sliding between the same materials listed in [Table 3.1](#). The values of wear coefficient illustrated here vary over nearly four orders of magnitude (from ca. 10^{-3} to ca. 10^{-7}), much more than the variation in the coefficient of friction, and no correlation is apparent between the two sets of data.

The Archard wear equation provides a valuable means of describing the severity of wear via the wear coefficient K , but its validity cannot be used to support or reject any particular mechanism of material removal. As we have seen, the wear coefficient is open to several interpretations, and it will be shown in [Chapter 6](#) that an equation of identical form can also be derived from completely different starting assumptions for material removal by abrasive wear. The mechanisms actually responsible for material removal in sliding wear, and for the wide range of values of K observed in practice, will be explored in the following sections.

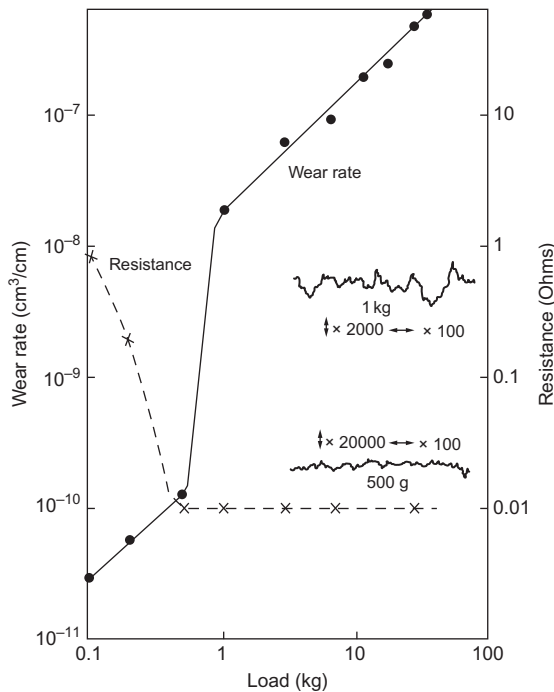
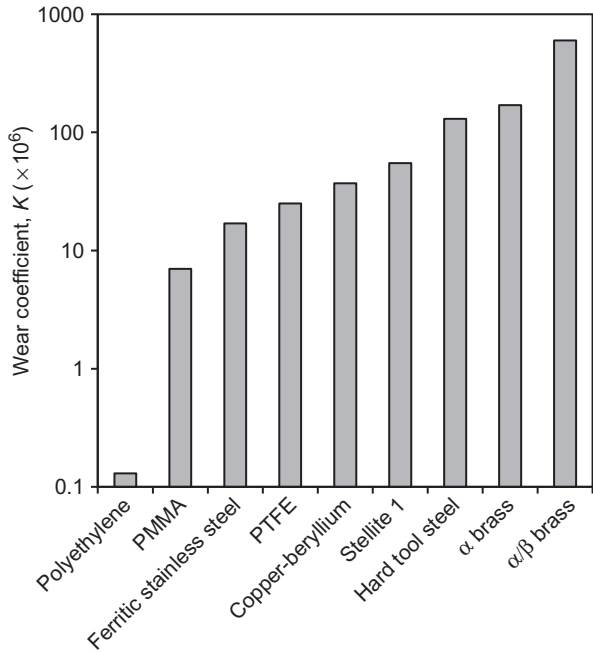


Fig. 5.5 Wear rate and electrical contact resistance for a leaded α/β brass pin sliding against a hard Stellite ring, as a function of normal load. Note the sharp transition in wear rate, associated with a change in contact resistance (from Hirst, W., Lancaster, J.K., 1956. *J. Appl. Phys.* 27, 1057–1065)

Fig. 5.6 Dimensionless wear coefficients for various materials sliding against tool steel in unlubricated pin-on-ring tests in air. It should be noted that these data relate to specific test conditions, and take no account of the mechanism of wear (data from Archard, J.F., Hirst, W., 1956. *Proc. Roy. Soc. Lond.* 236A, 397–410)



5.4 OBSERVATIONS OF DAMAGE DURING THE SLIDING WEAR OF METALS

During the sliding wear of metals, damage is evident in a number of ways, as follows;

- (i) there are microstructural changes in the material below the wearing surface associated both with mechanical deformation and with the thermal history associated with the frictional power dissipation;
- (ii) a layer of material that differs chemically from the bulk of the specimen is often formed on the wearing surface;
- (iii) particulate material (often termed *wear debris*) is expelled from the contact;

The subsurface microstructures, worn surface and debris are commonly examined by optical microscopy, SEM and TEM. A wide range of analytical techniques, such as nanoindentation, X-ray diffractometry, electron backscatter diffraction (EBSD) in the SEM, energy- and wavelength-dispersive X-ray microanalysis and Auger electron spectroscopy can also provide compositional and structural information.

Critical examination of both the worn surfaces and the wear debris can provide useful information about the mechanisms of wear, although it is never easy to discriminate unequivocally between the possible mechanisms that may be operating in a particular case. The worn surfaces will presumably exhibit fragments of debris at all stages of their life cycle, and thus it is not always easy to identify a clear progression in the wear process. Similarly, it is important to question whether the debris collected from a wear test is in the same state as that in which it first became detached from the surface, or whether its appearance (and even nature) may have been changed subsequently. The rolling and deformation of debris particles between sliding surfaces, and the oxidation of metallic debris after detachment, are examples of possible changes.

5.4.1 Microstructural changes below the wearing surface

Below the surface layers, significant changes to the microstructure are often observed in what is sometimes known as the *tribologically transformed zone* (TTZ). In some cases, severe deformation occurs at some considerable distance below the surface, whereas in other cases it is restricted to a much smaller depth. [Figure 5.7](#) illustrates schematically the various regions of microstructural modification using a common nomenclature that has been developed by several investigators. In Zone 1, bulk material remains undeformed. Zone 2 contains plastically deformed material, with plastic shear strains increasing towards the surface. Grain reorientation and refinement may be found towards the boundary between Zones 2 and 3.

Material near the top of Zone 2 often shows a lamellar structure resulting from alignment of dislocation cell walls and sub-grain boundaries parallel to the sliding surface. The microstructure is similar to that seen in metals subjected to large plastic shear strains in other ways, for example in torsion testing. At large strains, there is

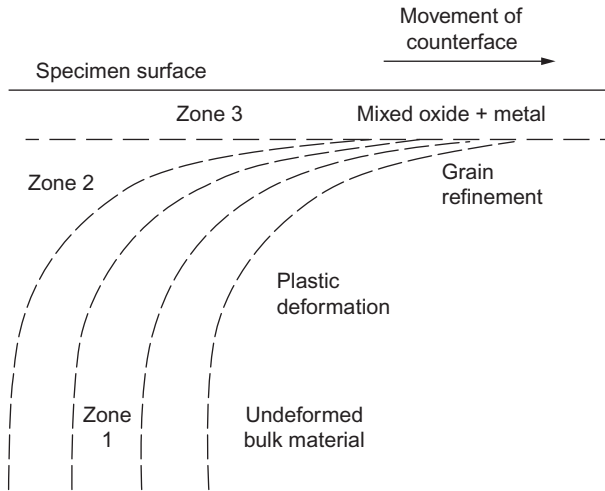


Fig. 5.7 Schematic diagram showing the regions of the TTZ (adapted from Rice, S.L., Nowotny, H., Wayne, S.F., 1981. In: Rhee, S.K., Ruff, A.W., Ludema, K.C. (Eds.), *Wear of Materials*. ASME, pp. 47–52)

a tendency for shear to localize, and fracture originating from this concentrated deformation may be responsible for the formation of the initial wear particles.

The subsurface strains commonly become very high close to the surface, and result in significant work-hardening unless the structure undergoes annealing promoted by the increases in temperature resulting from the frictional power dissipation. If annealing occurs, it is often accompanied by recrystallization of the microstructure. In areas where the deformed grain structure is preserved (i.e., no recrystallization has occurred), the equivalent strain, ε can be estimated from the following relationship:

$$\varepsilon \cong \frac{\tan \theta}{\sqrt{3}} \quad (5.8)$$

where θ is the angle that the microstructural features make with the normal to the worn surface.

Figure 5.8 shows an example of the hardening and equivalent strain below the surface of a silver-copper alloy following sliding against a stainless steel; strains up to 700% are observed with significant increases in hardness.

5.4.2 The nature of the surface layers

Two distinct types of surface layer (Zone 3) are observed on bodies that have been exposed to sliding wear. In the first type, the Zone 3 material is predominantly made up of metal oxides, with these oxides typically containing contributions from both partners of the couple (the specimen and the counterface). The oxide layer may be

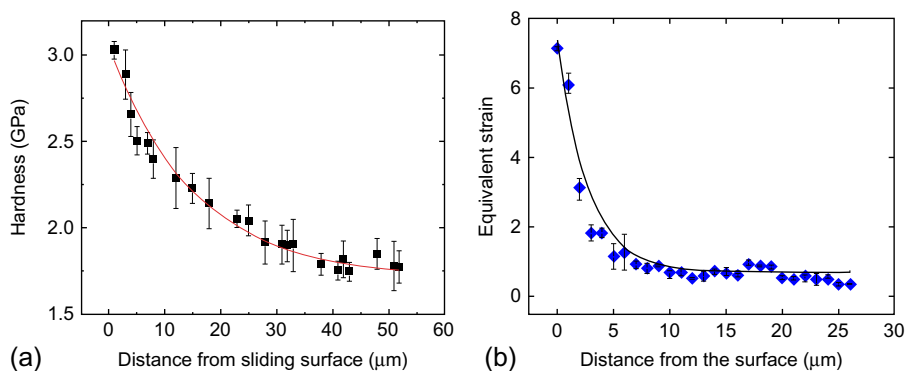


Fig. 5.8 (a) Nanoindentation hardness profile and (b) strain profile of eutectic Ag–Cu alloy after sliding against a stainless steel counterface (from Cai, W., Bellon, P., 2013. *Wear* 303(1–2), 602–610)

made up of loosely agglomerated particles, or form a dense coherent layer on the wearing surface which is very smooth; such a coherent oxide layer is generally termed a *glaze*.

The second type of Zone 3 material is distinct from the first in that it is mainly metallic in nature; it is typically very fine-grained (with sub-micrometre grains, sometimes as small as 10 nm) and contains components not present in the original bulk material, such as metal transferred from the counterface and oxides. This material of mixed composition, which contains intimately mixed particles of oxide and counterface material, shows many similarities to material produced by the process of mechanical alloying.

5.4.3 The nature of the debris

As with the materials found in Zone 3, the debris can generally be of two distinct forms. In the first, the debris is made up of fine particles (typically 0.01–1 μm in size) which are predominantly metal oxides, although these may bond together to form larger agglomerates as part of the wear process. In the second, the debris is composed of much larger metallic particles, perhaps 20–200 μm in size, which may even be individually visible to the naked eye; these fragments always tend to be plate-like, small in thickness compared with their other dimensions.

An example of the changes in the nature of the debris for sliding of an aluminium alloy against a steel counterface is presented in Fig. 5.9, with the change being observed over a relatively small range of sliding speeds. At the lowest speed, the debris is fine and consists of oxides associated with both the aluminium and steel. In contrast, at the highest sliding speed, the debris is present as much larger metallic flakes, formed solely from the aluminium alloy.

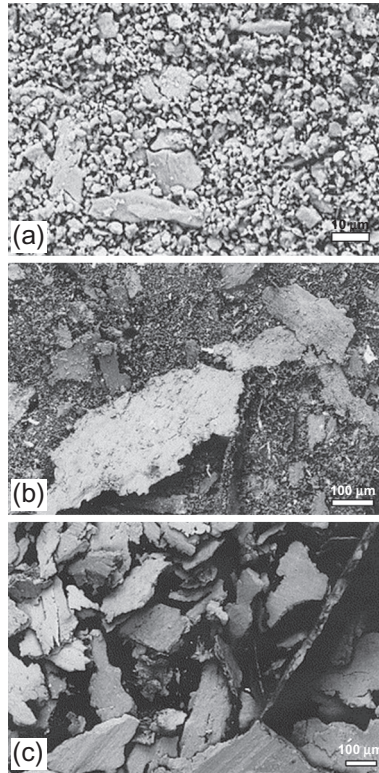


Fig. 5.9 The morphology of wear debris resulting from the sliding of an aluminium alloy against a bearing steel at different sliding velocities: (a) 0.8 m s^{-1} ; (b) 1.2 m s^{-1} ; (c) 1.4 m s^{-1} (from Wilson, S., Alpas, A.T., 1999. *Wear* 225–229(Part 1), 440–449)

5.5 MECHANISMS OF THE SLIDING WEAR OF METALS

5.5.1 *Plasticity-dominated wear*

Many different mechanisms have been proposed for plasticity-dominated wear of metals, but differ in the detailed processes by which material becomes removed.

Early theories of sliding wear (as discussed in [Section 5.3](#)) suggested that material is removed as lumps or fragments from asperity peaks by adhesive processes, and some debris particles certainly have irregular blocky shapes of dimensions appropriate to such processes. Several mechanisms have been proposed by which asperity contacts can lead directly to the formation of wear debris. [Figure 5.10](#) shows a representative example, in which plastic flow at an asperity tip is followed by detachment of a wear particle.

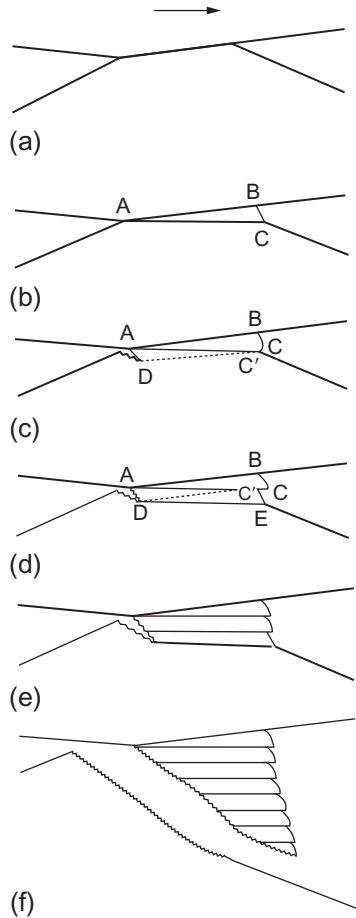


Fig. 5.10 Diagram illustrating one way by which detachment of a fragment of material might result from plastic deformation of an asperity tip (from Kayaba, T., Kato, K., 1979. In: Ludema, K.C., Glaeser, W.A., Rhee, S.K. (Eds.), *Wear of Materials*. ASME, pp. 45–56)

In this case, plastic shearing of successive layers (for example along the plane AC) occurs in conjunction with the propagation of a shear crack (AD), along which the particle detaches. Adhesive forces are not necessary for material to be removed by such a mechanism, which depends only on the mechanical interaction between two asperities, but they must be invoked to explain any subsequent attachment of debris to one of the sliding surfaces.

Another model for wear supposes that the formation of a debris fragment by asperity rupture is followed immediately by its adhesive transfer to the counterface, to form a new asperity on that surface (Fig. 5.11). Further sliding causes more fragments to be

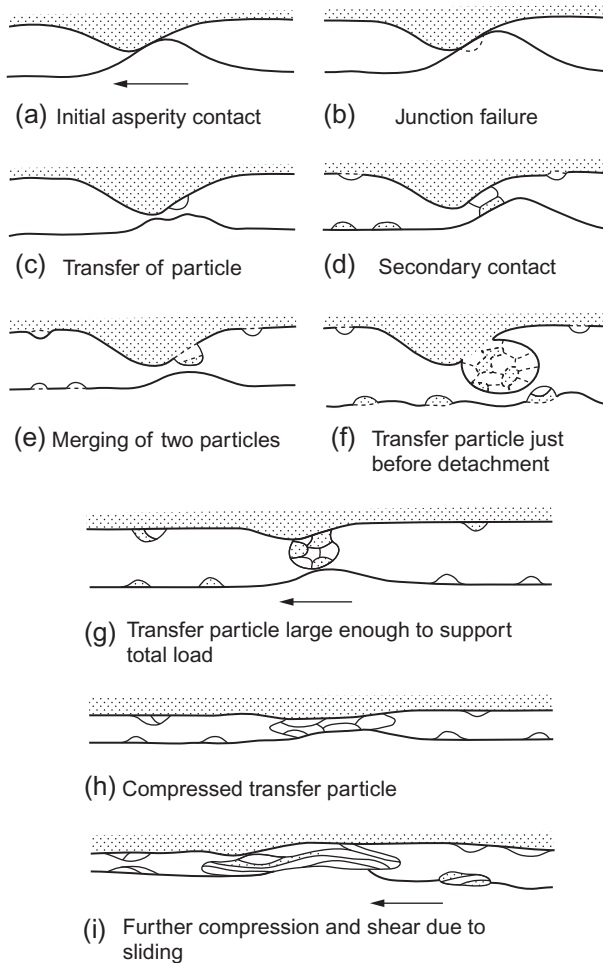


Fig. 5.11 Formation of transfer particle by asperity rupture and aggregation (after Sasada, T., 1984. *Wear research in Japan: trends and future directions*. *Wear* 100(1), 561–577)

formed which adhere to the original fragment, until a much larger conglomerate wear particle eventually becomes detached. This particle of debris may be roughly equiaxed, as shown in Fig. 5.11(f), or it may be flattened and elongated in the direction of sliding by further plastic deformation, forming a plate-like composite wear particle (as illustrated in Fig. 5.11i).

Plate-like metallic debris particles are often found in sliding wear (as shown for example in Fig. 5.9c), with ratios of thickness to width or length usually in the range from 0.1 to 0.5. Several other mechanisms have been proposed to explain their

formation. For example, shear fracture of the asperity summit in Fig. 5.10 along the plane AC would generate a thin flat particle (ABC). Another mechanism has become known as the *delamination theory*. This involves the nucleation of subsurface cracks and their propagation parallel to the surface. The cracks originate within the plastically deformed material beneath the surface (Zone 2 in Fig. 5.7), perhaps (but not necessarily) from voids caused by the shearing of inclusions, and after growing and linking together, eventually extend up to the free surface. A plate-like wear particle thus becomes detached.

Although several attempts have been made to model the sliding wear of metals as a fatigue process, the validity of this approach is not well established. Models have been proposed for the propagation of subsurface fatigue cracks under cyclic (elastic) shear forces, while others have assumed that asperities are removed as a result of low-cycle (macroscopically plastic) fatigue. There is experimental evidence for the existence of subsurface fatigue cracks under some conditions of sliding in some materials, but there is disagreement about whether the rate-controlling step in wear by a high-cycle fatigue mechanism might be the initiation of cracks or their propagation. It is certainly true that no simple correlation is found between the wear rates of metals and their fatigue properties. This may be because the cracks formed in wear are much shorter than those used to establish conventional fatigue data, or because they are stressed in shear rather than in tension, or because the surface material in which they propagate is very different from the bulk material used in fatigue tests.

The complexity of plasticity-dominated sliding wear is well illustrated by the diversity of the mechanisms discussed above. Experimental evidence has been published for each of them, yet it would be unrealistic to claim that any one represents the ‘true’ mechanism for sliding wear under these circumstances. Some will produce debris in a single sliding pass, while others require many passes. Transfer layers are necessary for some mechanisms but ignored in other models. Several different mechanisms may well operate in sequence or simultaneously during the lifetime of a sliding system, and details of the surface topography, composition and microstructure of the near-surface material, and the nature of the surface films and adsorbed species will all be very important.

5.5.2 Oxidative wear

The majority of metals have a negative free energy of oxidation, and their surfaces are therefore generally covered with a thin layer of oxide. Appendix B provides some background on oxide growth. The rate of oxidation can be increased significantly by continual removal of the oxide from the surface by asperity-asperity interactions (thus exposing fresh metal surface to react with oxygen), by mechanical disturbance of the oxide, or by increases in temperature that promote diffusion of species through the oxide layers and thus allow the oxide to grow. Flash temperatures of several hundred kelvin can readily be generated in sliding and the magnitude of these temperatures

depends strongly on velocity, as discussed in [Section 3.9](#). Only moderate sliding speeds are needed with most metals before these hot-spot temperatures become high enough to cause significant surface oxidation. For example, 700°C is reached at around 1 m s⁻¹ for steels. Growth of a substantial oxide film suppresses the plasticity-dominated mechanisms discussed in the previous section, by (i) reducing the shear strength of the interface and hence the development of the large subsurface shear strains necessary for these mechanisms to operate, and (ii) limiting the number of metal-metal asperity contacts. In oxidative wear, debris is removed from the oxide layer; Quinn has suggested that this occurs by delamination of the oxide film following its growth to a critical thickness of between 1 and 3 µm. Under vacuum or in an inert atmosphere, of course, oxidative wear cannot occur, and plasticity-dominated processes then occur over a much wider range of conditions.

The role of temperature in oxidative wear is clear, and various mechanisms have been suggested by which thermal oxidation can lead to wear. At low ambient temperatures significant oxidation occurs only at asperity contacts, and is associated with the transient flash temperatures. The fact that some FeO is found in the wear debris from steels sliding at low bulk temperature, for example, suggests that it forms at the hot-spots rather than at the lower mean temperature of the surface, since FeO grows under static conditions only at temperatures above 570°C.

At higher ambient temperatures, however, general oxidation of the surface outside the contact zone becomes more important, and may even dominate. In the oxidative wear of steels, the predominant oxide present in the debris changes with sliding speed and ambient temperature: at low speeds and temperatures α -Fe₂O₃ dominates, with FeO forming at high speeds and temperatures and Fe₃O₄ under intermediate conditions.

Both the measurement and estimation of temperature rises due to sliding are difficult. Determining the temperature dependence of oxidation rates due to sliding therefore poses problems. It is nevertheless clear that oxide growth during sliding, like thermal oxidation under static conditions, obeys an Arrhenius equation:

$$k_p = Ae^{(-Q/RT)} \quad (5.9)$$

Here, k_p is the parabolic rate constant for growth of the oxide film (see [Appendix B](#)), T is the absolute temperature and R is the gas constant. Q , the *activation energy*, determines the temperature dependence of the growth rate, while A , the *pre-exponential factor* or *Arrhenius constant*, acts as a multiplying factor.

The data in [Table 5.1](#) indicate clearly that oxidation under sliding conditions is much more rapid than would be expected from static oxidation models, especially at lower temperatures, and this conclusion is confirmed by many studies. Oxide thicknesses that would take a year to grow under static conditions can develop in hours or even minutes during sliding wear. This strong enhancement of oxidation by sliding may result from increased diffusion rates of ions through a growing oxide layer that has a high content of defects (e.g., voids, dislocations and vacancies) due to mechanical perturbation.

Table 5.1 Typical values of Arrhenius constant A and activation energy Q for oxidation of steels under static conditions and in sliding contacts (from Sullivan, J.L., Quinn, T.F.J., Rowson, D.M., 1980. Developments in the oxidational theory of mild wear. Tribol. Int. 13(4), 153–158)

Temperature range:	<450°C	450–600°C	>600°C
Arrhenius constant, A ($\text{kg}^2 \text{m}^{-4} \text{s}^{-1}$):			
Static oxidation	1.5×10^6	3.2×10^{-2}	1.1×10^5
Under sliding conditions	10^{16}	10^3	10^8
Activation energy, Q (kJ mol^{-1}):	208	96	210

Many detailed mechanisms have been proposed by which oxide particles become detached as wear debris. Growth of oxide on an asperity tip immediately after it has been scraped clean by contact with the counterface may be followed by removal of that oxide in the next contact event. This process has been termed ‘oxidation-scrape-reoxidation’. The fine particles of oxide removed in this way might agglomerate into larger transfer particles before finally being released as debris.

Experimental evidence suggests that under many circumstances an appreciable thickness of oxide (perhaps several micrometres) builds up on the metal surface before it becomes detached as debris particles, and mechanisms have been suggested in which ‘islands’ of oxide grow until they reach a critical thickness, when they spall off. These islands may be homogeneous oxide layers growing by tribochemically enhanced oxidation of the surface. Alternatively, the oxide islands may be aggregates of fine particles generated at asperity contacts by oxidation-scrape-reoxidation, or perhaps aggregates of fine metallic particles that have oxidized after detachment from the asperity peaks. One such mechanism, illustrated schematically in Fig. 5.12, involves an additional stage that occurs in some cases, namely the formation of a glaze by the agglomeration and sintering of oxide particles to form a smooth and more durable surface layer. The possibilities of transfer and back-transfer of oxide particles, and of abrasive wear caused by hard oxide particles, add further complexities to the range of mechanisms that have been proposed. As with plasticity-dominated wear, there is undoubtedly no single mechanism that applies to all cases, and the conditions under which the various possible mechanisms become important are still only poorly understood.

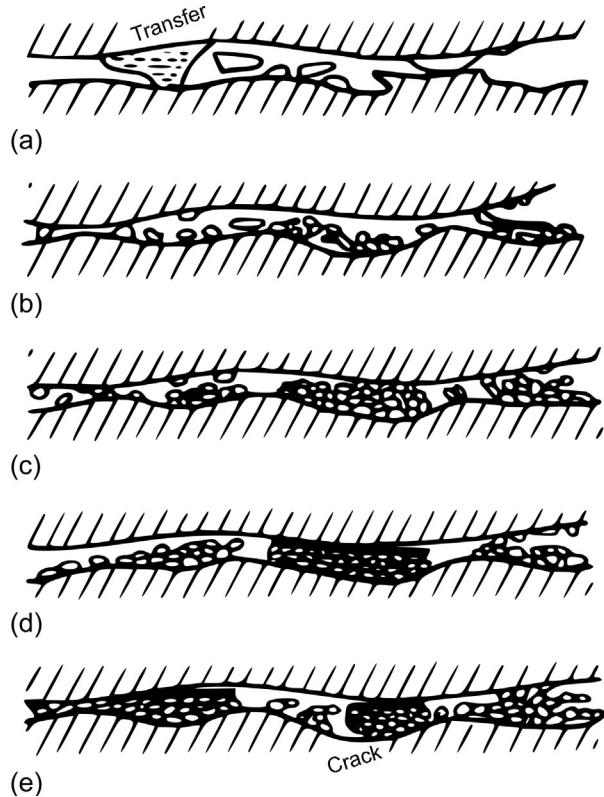
5.6 MILD AND SEVERE WEAR OF METALS

Figure 5.5 shows a transition in wear rate above a critical load. This transition is associated with a change in the mechanism of material removal, and in this section, links are drawn between the rates and mechanisms of wear and their dependence upon the tribological conditions to which they are exposed.

The regime of wear at low loads, below a transition of the type seen in Fig. 5.5, is often called *mild wear*, while that leading to the higher rate above the transition is

Fig. 5.12 Schematic representation of a model for the development of wear-protective oxide layers from wear debris particles.

(a) Generation of metallic wear debris. (b) Comminution, oxidation and agglomeration of debris particles. (c) Compaction of agglomerated debris particles. (d) Development of 'glaze' layer over compacted particle layer under some conditions. (e) Breakdown of wear-protective layers and development of new protective layers in other areas (from Jiang, J., Stott, F.H., Stack, M.M., 1994. *Wear* 176, 185–194)



termed *severe wear*. These terms, although apparently vague and subjective, are sufficiently well-defined in the context of the sliding wear of metals to be useful, and they are widely employed.

The major differences between the states of mild and severe wear are listed in [Table 5.2](#). Wear rates due to severe wear are nearly always so high as to be completely unacceptable in engineering applications, with wear coefficients, K , of up to 10^{-2} – 10^{-3} ; even rates of unlubricated mild wear, with K values typically below 10^{-4} – 10^{-5} , may be intolerable under many circumstances.

The transition between mild and severe wear results from a change in the nature of the sliding contact. In severe wear, there is extensive metallic contact over the whole real area of contact, as shown by the low electrical contact resistance. The transition from severe to mild wear occurs when two competing processes balance: the rate of exposure of fresh metal surface by plasticity-dominated processes (as discussed in [Section 5.5.1](#)), and the rate of oxidation of that surface by the surrounding atmosphere. In mild wear, the sliding surfaces are separated by oxide films, with only occasional direct metallic contact; this explains the relatively high contact resistance. In many sliding systems, the presence of the oxide films result in a lower coefficient of friction

Table 5.2 Features of mild and severe wear

	Mild wear	Severe wear
Rates of wear	Low	High (typically 100–1000 times greater than those observed in mild wear)
Surface morphology	Smooth	Rough
Zone 3 material	Oxides (typically from both surfaces)	Heavily deformed metallic material (typically from both surfaces) with incorporated oxide particles
Debris	Fine oxide particles	Coarse metallic flakes
Electrical contact resistance	High	Low

than that observed in severe wear (see [Section 3.5.3](#)). In the unlubricated sliding of brass on steel, for example, the coefficient of friction is ca. 0.15 in the mild wear regime, compared with 0.25–0.3 when severe wear occurs. The debris formed by mild wear consists of mixed oxides of copper, zinc and iron, originating from both the brass slider and the steel ring, and the wear rates of both are comparable despite their different hardnesses.

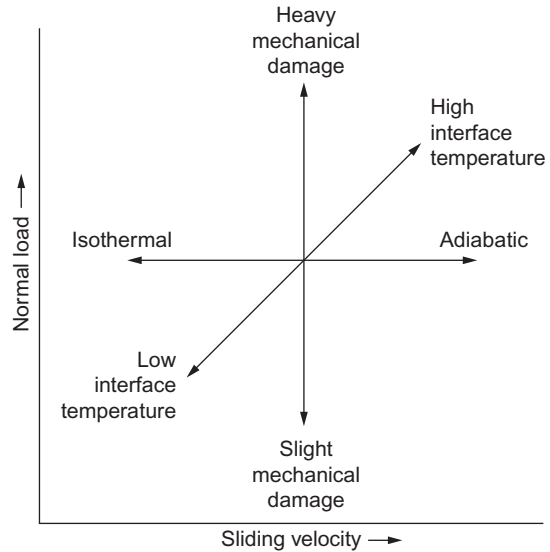
The mechanisms of removal of the oxide to form wear debris depend on the precise conditions of load, speed and temperature and can vary between different regions of the contact area. Flakes of oxide may detach as a result of cyclic loading, but abrasion by hard oxide particles can also be important. Abrasion (discussed in [Chapter 6](#)) can be a significant contributor to wear in sliding contacts.

Since oxide formation plays a crucial role in the process of mild wear, any factor that changes the rate of oxidation of the sliding surfaces will influence the transition between mild and severe wear. Temperature is the most obvious variable affecting oxidation rate: the temperature at the sliding interface will depend on both the ambient temperature and on the frictional power dissipation, which in turn depends on the sliding speed and load. The effects of both are illustrated schematically in [Fig. 5.13](#).

In the sliding wear of metals, no unique wear mechanism operates over a wide range of conditions. Rather, there are several mechanisms that change in relative importance as the sliding conditions are varied. This conclusion is also true for other sliding systems: transitions in the dominant wear mechanism, and in the associated rate of wear, commonly occur with variation of normal load and sliding speed, and also in some cases with sliding time (or distance). The main factors controlling the importance of the underlying mechanisms are *mechanical stresses*, *temperature* and *oxidation phenomena*. Consideration of all three factors is essential in understanding the sliding wear of metals; it must also be appreciated that the conditions at the interface may be very different, especially in terms of temperature, from those in the surroundings.

The complexity of sliding wear arises from the fact that all three controlling factors are interrelated, and may be influenced by both load and sliding velocity. [Figure 5.13](#)

Fig. 5.13 Diagram illustrating the combined influences of load and sliding speed on the sliding wear process in metals



illustrates schematically how the extent of mechanical damage (due to surface stresses), and also the interface temperature, will depend on the load and sliding velocity. Increasing the load leads directly to higher stresses, and these will result in more severe mechanical damage. Both the load and the sliding velocity influence the interface temperature. Together, they control the power dissipated at the interface (since that is the product of the sliding speed and the frictional force).

Additionally, the sliding speed influences the process of heat conduction from the interface, as discussed in [Section 3.9](#). At low sliding speed, the interface temperature will remain low; in the limit, sliding would be isothermal. At high speed, the heat diffuses only a short distance during each asperity interaction and the interface temperature is therefore high. A high interface temperature leads to high chemical reactivity of the surfaces, causing, for example, rapid growth of oxide films in air. It will also reduce the mechanical strength of asperities and near-surface material, and may even in extreme cases cause melting. We shall discuss the influence of mechanical stresses, temperature and oxidation further in the following paragraphs.

The important mechanical stresses are the normal stress at the surface of each sliding body, and the shear stresses at and below the surface. As we saw in [Chapter 2](#), the normal stress at the asperity contacts will, under conditions of plastic contact, be close to the indentation hardness of the softer body. But if the surfaces are very smooth, closely conforming or lightly loaded then the contacts may be elastic (see [Section 2.5.6](#)). The normal load applied to the system therefore controls the extent of plastic flow at the asperity contacts; if the load is low enough, or the surfaces conform well enough, wear will proceed only very slowly, perhaps by an elastic (e.g., high-cycle fatigue) process.

The magnitude and position of the maximum shear stress depend on the coefficient of friction (see Section 2.5.3). For $\mu < 0.25$, the maximum shear stress and associated plastic flow will lie beneath the surface, and the plastic strain accumulated by each sliding pass is small. This condition is typical for a lubricated system, or for one carrying a protective oxide layer. For $\mu > 0.25$, however, the maximum shear stress lies at the surface and large shear strains can be accumulated. Several wear mechanisms dominated by plastic flow have been proposed, involving asperity adhesion and shear (as originally envisaged by Archard), or nucleation and growth of subsurface cracks leading to the formation of lamellar wear particles ('delamination wear'), and other models involving fatigue crack propagation. At relatively low sliding speeds and high loads, these 'plasticity-dominated' wear mechanisms prevail and lead to severe wear.

As already discussed in Section 5.5.2, oxidation phenomena are also important in sliding wear, since nearly all metals form oxide films in air. The rate of film growth depends, as in static oxidation, strongly on temperature; however, the local temperature at the sliding interface may be substantially higher than that of the surroundings, and may also be enhanced at the asperity contacts by transient local flash temperatures. Perhaps the simplest method to estimate the highest possible flash temperature is to suppose that for a brief moment the whole load applied to the system, both normal and tangential, is carried by a single asperity, and to determine the interface temperature resulting from this instantaneous rate of local power dissipation. The temperature rise varies with sliding speed because the power dissipation depends on velocity and also because the motion of the heat source over the counterface must be considered, as discussed in Section 3.9. Figure 5.14 shows the predictions of this model, for steels of different hardness.

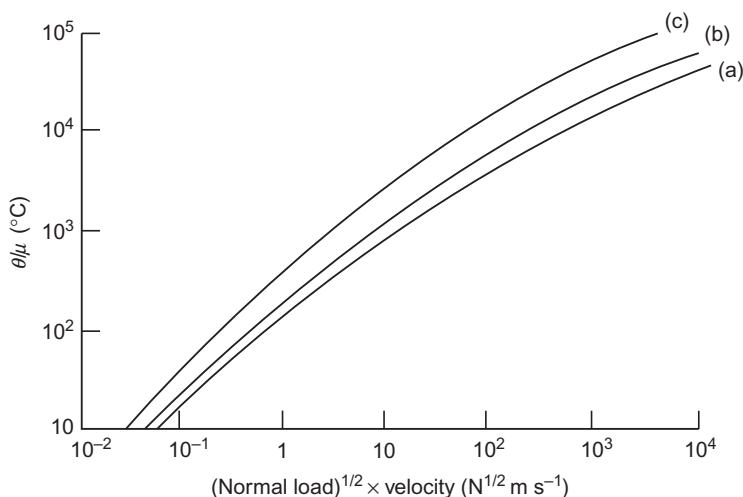


Fig. 5.14 Maximum attainable flash temperature for sliding steel contacts. Three different values are assumed for the hardness of the steel: (a) 150 HV; (b) 250 HV; (c) 850 HV (from Archard, J.F., 1958. *Wear* 2, 438–455)

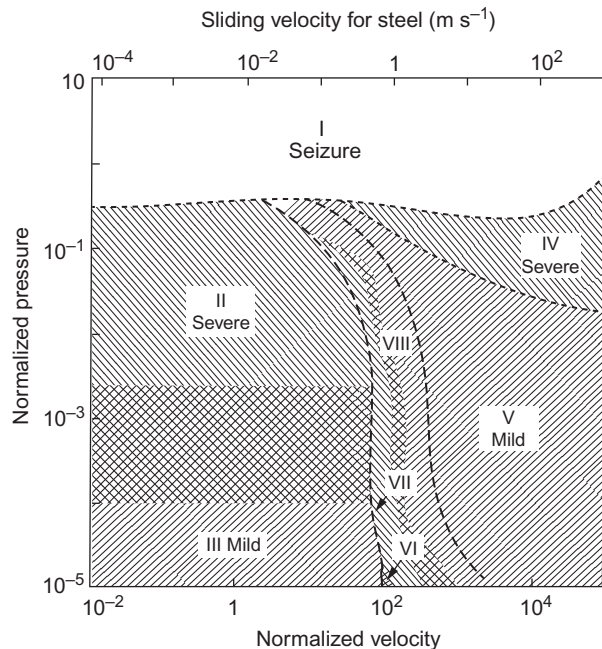
Flash temperature rises predicted in this way may exceed 1000 K, and steady interfacial temperature rises due to the mean frictional power dissipation can add a further hundred kelvin or so. Temperatures of this order have been confirmed experimentally by thermoelectric and infra-red measurements. Such high temperature transients can lead to local phase transformations: for example, to the formation of austenite, which then quenches to martensite, in carbon and low-alloy steels. They are also responsible for rapid surface oxidation.

5.7 WEAR-REGIME MAPS FOR METALS

A useful method of depicting the various regimes in which different modes of wear dominate is provided by a *wear-regime map* or *wear-mode map*. Figure 5.15 shows an example for steels sliding in air in the common pin-on-disc geometry.

Although the details of the map are specific for steels, the general form will be similar for the unlubricated sliding of most metals in air. The map is plotted with the same coordinates as the diagram in Fig. 5.13. Dimensionless variables are employed: a normalized contact pressure (defined as the normal load on the specimen pin, divided by the nominal contact area times the indentation hardness of the softer material), and a normalized velocity (which can be thought of as the sliding speed divided by the speed of heat flow). Representative values of sliding speed are also given along the upper abscissa, for typical steels.

Fig. 5.15 Wear mode map for the unlubricated sliding of steel in the pin-on-disc configuration (from Lim, S.C., Ashby, M.F., Brunton, J.H., 1987. *Acta Metall.* 35, 1343–1348)



The dominant modes of wear, and the resulting regions on the diagram, are determined in two ways: by plotting empirical data from experiments performed under widely differing conditions, and by the use of simple analytical models for wear rates due to the various mechanisms. Although wear-regime maps can provide only approximate predictions of wear rates, they nevertheless give a valuable illustration of the regimes over which the different wear mechanisms are important.

Eight distinct regimes can be identified in [Fig. 5.15](#). In regime I, associated with very high contact pressure, gross seizure of the surfaces occurs. This regime corresponds to catastrophic growth of asperity junctions so that the real area of contact becomes equal to the apparent area. The boundary of this regime can be calculated from the junction growth model discussed in [Section 3.4.3](#).

In regime II, corresponding to high loads and relatively low sliding speeds, the thin surface oxide film on the steel is penetrated at asperity contacts, high surface tractions occur, and metallic wear debris is formed; this is a regime of severe wear. At lower loads, in regime III, the oxide is not penetrated, and a low rate of mild wear results due to removal of particles from the oxide layer itself.

In regimes II and III, thermal effects are negligible, whereas in regimes IV and V they become very important. At high loads and sliding speeds (regime IV) the frictional power dissipation is so high, and thermal conduction so ineffective at removing heat from the interface, that melting occurs. Although this leads to a low coefficient of friction (see [Section 3.5.5](#)), viscous forces in the molten layer continue to dissipate energy and wear is rapid and severe, with metal being removed in the form of molten droplets. This extreme form of wear occurs, for example, as the driving band of an artillery shell slides along a gun barrel. It is also put to use in the process of *friction cutting*, in which materials are cut by a rapidly rotating disc of steel with a smooth periphery.

At low contact pressure, but high sliding speed, we enter regime V. Here the interface temperature is still high, but below the melting point of the metal, and the surface oxidizes rapidly. An extreme type of oxidational wear occurs, in which a thick, hot oxide layer deforms plastically and may even melt, and where the wear debris consists of oxide from this layer, rather than metal. Following our earlier definition (see [Section 5.6](#)), this is a form of mild wear.

Regimes VI, VII and VIII occur over narrow ranges of sliding velocity, and represent various types of transitional behaviour between the extremes of low speed, isothermal conditions, and high speed, adiabatic conditions. At low loads, all three regimes can be experienced as the sliding speed is increased. In regime VI, thermal effects begin to play a role, not by raising the interface temperature significantly, but through the occurrence of hot-spots at asperity contacts. These local flash temperatures lead to patchy oxide growth, and the main source of debris in this regime of mild wear is the detachment of oxide particles by spalling, typically when they reach a thickness of about 10 μm . At higher loads, metallic contact between asperities occurs, despite the enhanced oxidation at hot-spots, and the severe wear regime VII, with metallic debris, is entered.

Regime VIII, associated with higher flash temperatures, occurs when martensite is formed at the interface. This results from local heating of the asperities, followed by

quenching by conduction of heat into the bulk. It is probable that the local temperature must rise above the α - γ allotropic transformation temperature for iron (910°C), rather than simply above the eutectoid temperature (723°C), since only then could austenitization occur on the extremely short timescale of a hot-spot transient. Dissolution of cementite in ferrite, which would be necessary below 910°C , is rather sluggish, whereas dissolution of cementite in austenite is extremely rapid. There is also some evidence that the transformation is facilitated by shear stresses at the contact points. The high strength of the martensite provides local mechanical support for the surface oxide film, and mild wear proceeds by the removal of this oxide, in a regime similar in many ways to regime VI.

The patterns of behaviour illustrated in Fig. 5.15 are, necessarily, simplified. No account is taken, for example, of possible transitions in wear behaviour with sliding distance or time. Changes in composition and microstructure of the steel, or in ambient temperature, will obviously shift the boundaries between the various regimes, which in any case are not rigidly defined. In the map shown, regimes of mild and severe wear overlap to a marked extent. These areas of overlap represent uncertainty about the precise location of the transition, not only because of variation in behaviour between different steels and experimental conditions, but also because in some cases the transitions themselves are poorly defined. Different areas of the rubbing surface may simultaneously exhibit different modes of wear.

Wear maps nevertheless show clearly how relatively small changes in sliding speed or contact pressure can lead to transitions between mild and severe wear, and how more than one transition might be generated by changing only one variable.

Transitions between regimes of mild and severe wear, in which mild wear is caused by oxidation phenomena and severe wear involves metallic contact, are commonly observed in many metals. As we have seen, they may result from changes in normal load or sliding speed, but can also occur with time or sliding distance. They are then associated with *running-in*. If a transition occurs during running-in, it is usually from severe to mild wear, and is associated with a smoothing of surface topography, or the formation of an interfacial layer of oxide or a hard phase such as martensite.

5.8 FRETTING WEAR OF METALS

The term *fretting* denotes small oscillatory movement between two solid surfaces in contact. The direction of the motion is usually, but not necessarily, tangential to the surfaces. When the amplitude of the motion lies in a range typically from 1 to $300\text{ }\mu\text{m}$, surface degradation occurs which is called *fretting damage* or *fretting wear*.

Although fretting wear can be regarded formally as reciprocating sliding wear with very small displacement, there are enough differences in both wear rates and mechanisms to merit the use of a distinct term. The key difference between fretting wear and sliding wear is that in the former, the debris is more prone to remain in the contact area (due to the small displacement amplitude) and this influences the progression of wear.

Whereas sliding wear usually results from deliberate movement of the surfaces, fretting often arises between surfaces that are intended to be fixed in relation to each other, but which nevertheless experience a small oscillatory relative movement. These small displacements often originate from vibration. Typical examples of locations where fretting may occur are in hubs and discs press-fitted to rotating shafts, in splined couplings, in riveted or bolted joints, between the strands of wire ropes, and between the rolling elements and their tracks in nominally stationary ball and roller races. It may also occur between items packed inadequately for transport, where vibration can lead to fretting damage at the points of contact between the items. Fretting wear can lead to loosening of joints, resulting in increased vibration and a consequently accelerated rate of further wear. Since the debris formed by fretting wear is predominantly oxide, which in most metals occupies a larger volume than the metal from which it originates, fretting wear can also lead to seizure of parts that are designed to slide or rotate with a small clearance. Whether fretting leads to increased clearance or to seizure depends on the ease with which the wear debris can escape from the contact region.

A further important phenomenon associated with fretting damage is the development of fatigue cracks in the damaged region, leading to large reductions in the fatigue strength of a cyclically-loaded component. The greatest reduction in fatigue strength is seen when the fretting process and cyclic stressing occur together; the cyclic displacement responsible for fretting damage is often directly linked to the cyclic load which causes the fatigue cracks to grow. Fatigue failure originating from fretting damage is given the name *fretting fatigue*.

Fretting wear is usually studied in the laboratory in systems of simple geometry: typically sphere-on-flat, cylinder-on-flat or crossed cylinders. For elastic loading, the conditions in the contact zone can be analyzed theoretically. We saw in [Section 2.5.1](#) that for elastic contact of a sphere or cylinder on a plane surface under normal load, the contact pressure reaches a maximum at the centre of the contact circle, and falls to zero at the edges. [Figure 5.16](#) illustrates this pressure distribution for a circular contact, and also shows a plan view of the area of contact.

If a small cyclic tangential force is superimposed on the normal force, some displacement may occur between the surfaces around the edges of the contact zone, where the normal pressure is lowest and the frictional stress opposing movement is therefore least ([Fig. 5.16c](#)). The contact zone can then be divided into two regions: a central area where there is no relative tangential movement, and a surrounding annular zone in which *microslip* occurs. Similar behaviour is seen if a cyclic torque (about a vertical axis) is applied to the sphere, rather than a tangential force. Fretting damage occurs in the microslip region. As the amplitude of the cyclic tangential force is increased, the central area within which no slip occurs shrinks, until eventually slip occurs over the whole contact area ([Fig. 5.16d](#)).

In laboratory studies of fretting, it is common to apply a macroscopic displacement Δ and to measure the force Q which resists that displacement; from these measurements, the concept of a fretting loop can be defined (see [Fig. 5.17](#)). Due to the compliance of the system, the applied displacement amplitude (Δ^*) will not be the same as the amplitude of slip in the contact (δ); moreover, there is often a variation in force

Fig. 5.16 (a) The distribution of elastic normal stress beneath a sphere pressed against a flat. (b)–(d) show plan views of contact with increasing values of applied cyclic tangential force. The shaded regions represent areas over which local slip occurs between the surfaces

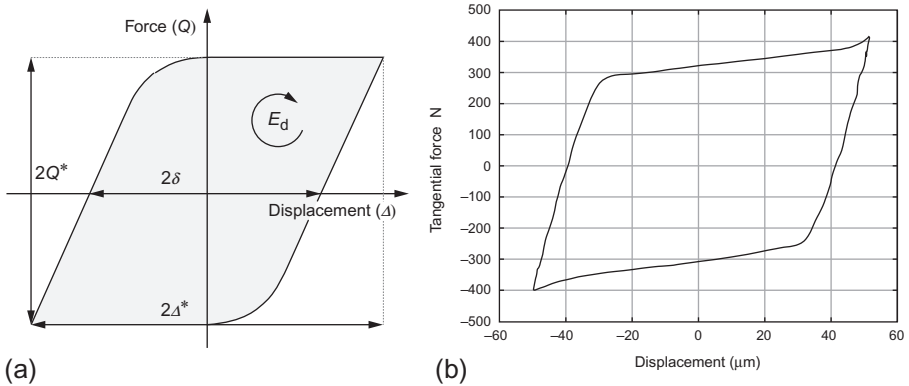
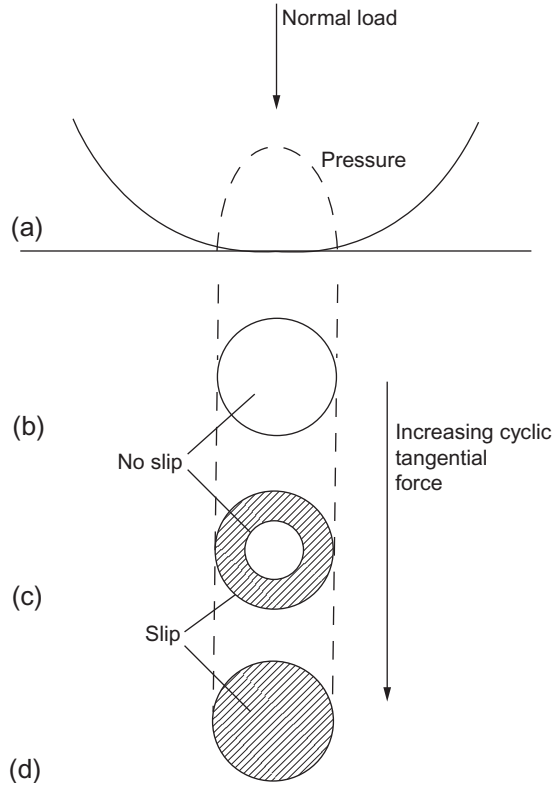


Fig. 5.17 (a) Schematic fretting loop with terms as defined in the text; (b) experimental fretting loop (from Jin, X., Sun, W., Shipway, P.H., 2016. Tribol. Int. 102, 392–406)

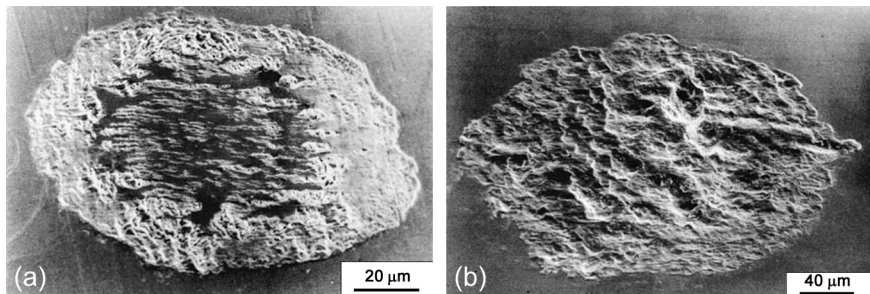


Fig. 5.18 Surface fretting damage caused by small-scale cyclic displacement of a stainless steel sphere against a flat of the same material: (a) partial sticking leads to less damage in the centre of the contact area, while (b) gross slip causes damage over the whole region. The elliptical shape of the contact areas results from the tilt employed in the SEM imaging (from Bryggman, U., Söderberg, S., 1986. *Wear* 110, 1–17)

during sliding, which is commonly associated with geometrical development of the wear scar. The energy dissipated per cycle, E_d , is given by the area enclosed by the loop, and the total frictional energy dissipated in a test is the sum of these areas over all cycles. Figure 5.18 shows examples of the fretting damage developed on a steel flat in laboratory sphere-on-flat experiments, where (a) microslip has occurred over only part of the contact circle, and (b) slip and damage have taken place over the whole contact area.

The regimes of normal load and displacement amplitude corresponding to complete sticking, mixed sliding and microslip, and gross slip in such experiments are plotted in Fig. 5.19. At a fixed normal load, increasing the amplitude of the cyclic tangential displacement leads to an increase in the extent of slip; a similar effect results from *reducing* the normal load at a fixed amplitude.

Fretting wear occurs most severely in the area of the contact zone undergoing slip, and is usually quantified in terms of the mass or volume loss from the surface. The wear rate is found to be approximately proportional to the normal load, and it is useful then to express it as a *specific wear rate*, k , as defined in Section 5.3 for continuous sliding wear. The specific wear rate is the volume removed by wear per unit sliding distance per unit normal load; the sliding distance is simply related to the duration of the test, the frequency of the vibration and the cyclic slip amplitude (remembering that the slip amplitude is not the same as the applied displacement amplitude). The value of k provides a measure of the severity of wear, and allows the results of tests at different normal loads and amplitudes to be compared. Alternatively, the wear volume can be plotted as a function of the total energy dissipated (i.e., the sum of E_d (see Fig. 5.17a) over the test duration), and the specific wear rate is then quoted in terms of volume lost per unit energy dissipated (which is dimensionally identical to the normal definition of k). In the example of fretting of titanium alloys shown in Fig. 5.20, the ‘Archard work’ (the product of the total distance slipped and the applied load) and the energy dissipated are directly proportional to each other, with the constant of proportionality being equal to the coefficient of friction.

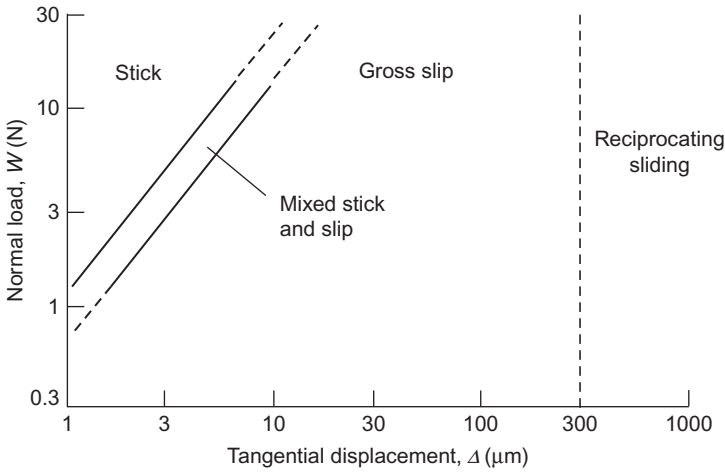


Fig. 5.19 Schematic illustration of fretting regimes for counterformal contact between stainless steel surfaces (sphere-on-flat geometry) (from Vingsbo, O., Söderberg, S., 1988. On fretting maps. *Wear* 126, 131–147)

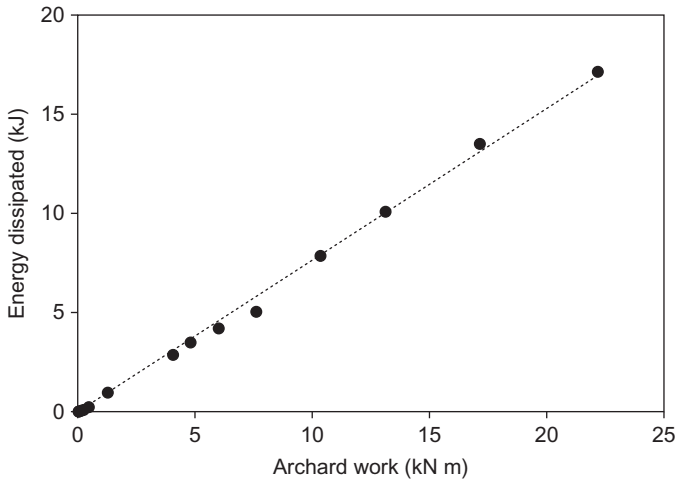


Fig. 5.20 Correlation between ‘Archard work’ and frictional energy dissipated in fretting of titanium alloys; the gradient of 0.77 represents the coefficient of friction (data from Heredia, S., Fouvry, S., 2010. *Wear* 269(7–8), 515–524)

Figure 5.21 shows schematically how the fretting wear rates of steels, defined in this way, depend on the applied displacement amplitude. At very small amplitudes ($<1\ \mu\text{m}$), the wear rate is negligibly small. Under these conditions, no significant microslip occurs in the contact and there is consequently very little damage. As the amplitude is increased, microslip develops and the wear rate rises slowly. Once

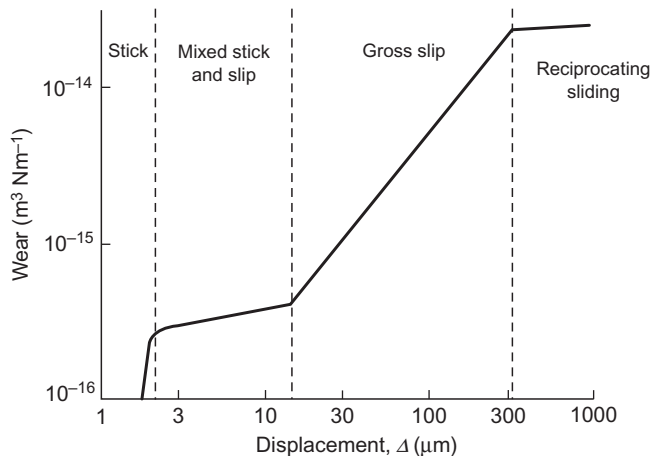


Fig. 5.21 The variation of fretting wear rate with amplitude of displacement Δ for a wide range of steels (after Vingsbo, O., Söderberg, S., 1988. On fretting maps. *Wear* 126, 131–147)

gross slip occurs (at amplitudes $> \sim 10 \mu\text{m}$) a rapid rise in k occurs, which eventually levels off to the constant value expected for reciprocating sliding wear at amplitudes $> \sim 300 \mu\text{m}$. The fretting amplitudes at which, on the one hand, damage becomes negligible, and on the other the process becomes indistinguishable from reciprocating sliding, are the subject of debate. They certainly depend on the mechanical properties of the material. Various values have been reported; the threshold for damage in some systems is $< 0.5 \mu\text{m}$, perhaps as low as $0.1 \mu\text{m}$, while the transition to sliding wear (identified by a value of k that is independent of amplitude) has been observed at amplitudes as small as $50 \mu\text{m}$.

The fatigue life of a cyclically-loaded specimen also varies with the amplitude of fretting displacement at its surface. A minimum in the fatigue life occurs at the transition from mixed microslip and sticking in the contact zone to gross slip. It can be shown that in the mixed stick–slip regime (as shown in Fig. 5.16c), the highest alternating stresses in the surface occur at the boundary between the two regions, and it is therefore at that point that fatigue cracks would be expected to initiate. It is important to note that the conditions giving the shortest fatigue life do not correspond to those giving the highest wear rate; in practice, failure by fretting fatigue is most likely at loads and fretting amplitudes that are responsible for only low rates of wear since here, high stresses are not reduced by wear and cracks can initiate and grow in regions of material before that material is removed by the wear process.

The mean sliding velocity during fretting is very low; a displacement amplitude of $20 \mu\text{m}$ at a frequency of 50 Hz provides a mean velocity of only 2 mm s^{-1} . Reference to Fig. 5.15 suggests that unidirectional sliding wear in steels at such low speeds and at moderately high normal pressure should result in severe wear, with metallic wear debris. This indeed occurs during the early stages of fretting wear, and gives rise to low electrical contact resistance and to initial wear rates that are comparable with those in unidirectional sliding.

However, fretting differs from continuous sliding, even at an equivalent velocity, in that the escape of the wear debris from the contact zone is more difficult. Further fretting in an oxidizing environment causes plastic smearing of the metallic debris fragments formed during the earliest stages, combined with their oxidation. In the steady-state fretting wear of steels, the wear debris consists of very fine ($0.01\text{--}0.1\text{ }\mu\text{m}$) plate-like particles, mainly of $\alpha\text{-Fe}_2\text{O}_3$ with some metallic iron, of a characteristic red-brown colour, redder than the normal corrosion product. This is sometimes colloquially known as ‘cocoa’ or ‘red rust’. The wear debris appears to form by delamination from a very thin composite surface layer of mixed oxide and metal, and the associated wear rate is substantially lower than that in the initial period of severe wear. The wear rate falls as loss of debris from the contact is made more difficult; for example, a moderate increase in the environmental temperature of fretting can cause the debris to sinter in the contact (forming a glaze), resulting in a substantial reduction in the rate of wear.

Other metals wear in a similar way, although the duration of the initial stage depends on their mechanical properties and rate of oxidation. Titanium, for example, forms oxide debris almost immediately from the start of fretting wear. In some materials the fretting debris may be hard enough once formed to cause further wear by abrasion (see [Section 6.3](#)). The debris formed by aluminium, for example, is mainly fine $\gamma\text{-Al}_2\text{O}_3$, black in colour and very hard compared with the aluminium itself.

The importance of oxidation in fretting wear, and the fact that the debris after the initial stages is predominantly oxide, has led to the use of the term *fretting corrosion* as a synonym for fretting wear, although as we have seen, the earliest stages of fretting wear do not involve appreciable chemical attack. It is preferable to use the more general term ‘fretting wear’ to denote all types of wear due to fretting motion, and to restrict the term ‘fretting corrosion’ to cases, as in the steady-state fretting wear of most metals in air, where the debris is predominantly the product of a chemical reaction.

Removal of oxygen or other chemically reactive species from the atmosphere leads to a substantial change in the fretting behaviour of metals (and often to a change in the wear rate too). Fretting in an inert atmosphere leads to the transfer of metallic debris from one surface to the other, but little escapes from the contact zone. The volume changes associated with the formation of oxide are absent, but substantial damage does accrue which may lead to crack initiation and fretting fatigue.

Water vapour also influences the rate of fretting wear in air, although there is still debate about the mechanism; clearly, the water vapour either changes the nature of the debris, or changes the way in which the debris is retained in (or lost from) the contact. Whatever the mechanism, it is clear that monitoring (and preferably control) of humidity in experimental studies of fretting is of critical importance.

5.9 WEAR OF METALS IN LUBRICATED CONTACTS

In a fluid-lubricated system, the ratio λ between the thickness of the lubricant film and the root-mean-square asperity height (see equation 4.20) determines the regime of lubrication. As we saw in [Section 4.5](#), increasing the normal load on the contact,

decreasing the sliding speed or decreasing the lubricant viscosity all tend to reduce the film thickness. Under full-film hydrodynamic or elastohydrodynamic lubrication (EHL) conditions, where the asperities on the opposing surfaces do not come into contact, then the wear rate will be very low.

If λ falls below about 3, however, some asperity contact occurs, and we enter the regime of partial (or mixed) EHL. The wear rate will then inevitably be higher than under full-film conditions. For $\lambda < \sim 1$, only the presence of boundary or solid lubricants can prevent the wear rate from rising to the level found in an unlubricated system; such lubricants are also invaluable in mitigating the wear rate in the mixed lubrication regime.

Figure 5.22 illustrates these regimes of lubrication and wear schematically. It also shows the variation of coefficient of friction μ , which follows the Stribeck curve familiar from Fig. 4.7. Typical values of the specific wear rate (k) for metals are indicated. Although the absolute values of k vary considerably between different systems, the relative changes between the various regimes represent widely observed behaviour.

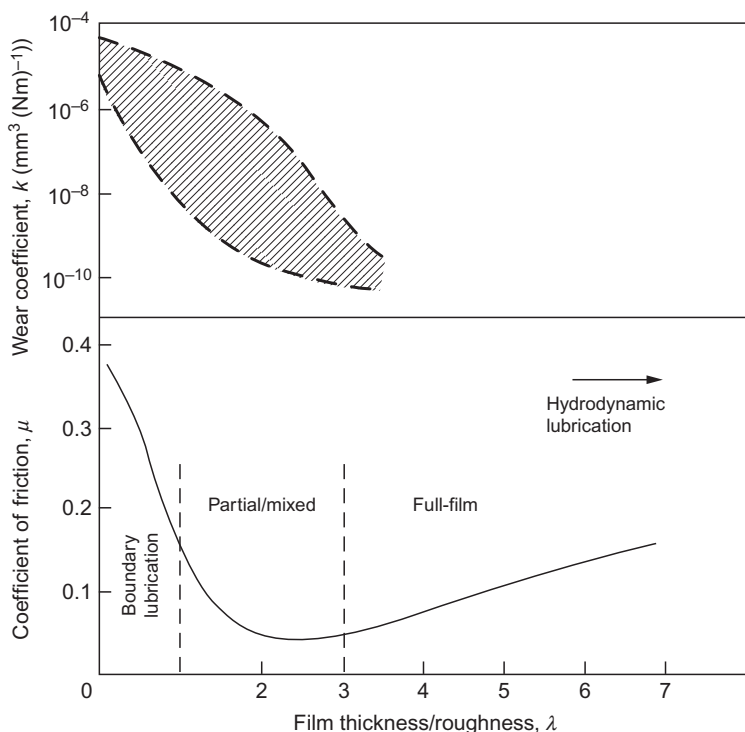


Fig. 5.22 Regimes of lubrication and wear in the lubricated sliding of metals, as a function of the ratio of film thickness to surface roughness, λ (from Czichos, H., Habig, K.-H., 1985. In: Dowson, D., et al. (Eds.), *Mixed Lubrication and Lubricated Wear*, Proc. 11th Leeds-Lyon Symposium on Tribology, Butterworths, pp. 135–147)

An alternative method of representing data on lubricated wear is provided by the *IRG transition diagram*, developed by the International Research Group on Wear of Engineering Materials sponsored by the OECD. The IRG diagram displays regimes of lubricated wear on a map with coordinates representing normal load and sliding speed. The boundaries between the regimes are determined from experiments with concentrated (counterformal) contact geometries, such as sphere on flat or crossed cylinder methods (see Section 5.2).

Figure 5.23 shows a typical IRG transition diagram for steel sliding on steel, completely submerged in an oil bath at constant temperature. Boundaries separate three regions, conventionally labelled I, II and III.

Regime I corresponds to partial EHL. The coefficient of friction is typically 0.02–0.1, and remains steady at this value for long periods of running after a very brief, transient higher value. The wear rate during this initial period may be quite high, but drops rapidly and continues to fall with further sliding, eventually becoming very small (with specific wear rate $k < \sim 10^{-9} \text{ mm}^3 \text{ N}^{-1} \text{ m}^{-1}$). This behaviour may be explained in the following way. On initial sliding, frictional interactions lead to local heating and oxidation of asperities, and material is removed by oxidative wear. The regime may be described as one of mild wear. As high spots are removed from the surfaces by wear, the effective value of λ increases, and the conditions change from partial to full-film EHL. Friction and wear thus both fall to low and stable levels.

In regime II, the initial coefficient of friction is higher than in regime I, typically 0.3–0.4 for steels. The wear rate is also higher, although still tolerable for some engineering applications ($k \approx 10^{-6}$ – $10^{-8} \text{ mm}^3 \text{ N}^{-1} \text{ m}^{-1}$). In this regime, boundary

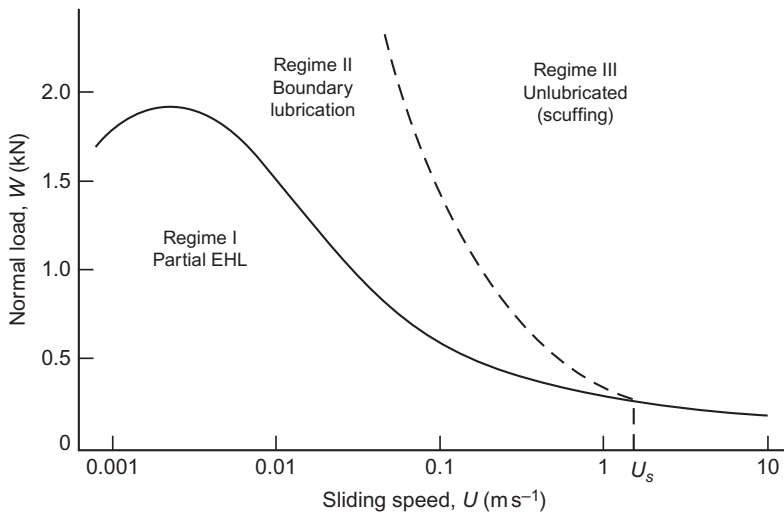


Fig. 5.23 Example of an IRG transition diagram for counterformal contact between steel components, fully submerged in an oil bath (from de Gee, A.W.J., Begelinger, A., Salamon, G., 1985. In: Dowson, D., et al. (Eds.), *Mixed Lubrication and Lubricated Wear*, Proc. 11th Leeds-Lyon Symposium on Tribology, Butterworths, pp. 105–116)

lubrication dominates; fluid film effects are negligible because of the high contact pressures and consequent low values of λ . After a longer running-in period than in regime I, μ falls (to about 0.1); this drop is associated primarily with general oxidation of the sliding surfaces (despite the access of oxygen being more limited by the presence of the lubricant). As wear continues and the surfaces become smoother, λ increases, fluid film lubrication can become important and stable sliding under partial or even full-film EHL conditions can be maintained. Wear under steady-state conditions in regime II is oxidative in character (despite the lubricant itself restricting access of oxygen into the contact), although the initial higher wear rate during running-in is associated with adhesive mechanisms involving unprotected asperity contacts, and can be viewed as incipient scuffing (see below).

In regime III, conditions are so severe that no lubricant film (boundary or partial EHL) can prevent metallic contact. Rapid wear occurs between the effectively unlubricated surfaces. The coefficient of friction is typically 0.4–0.5, with $k > \sim 10^{-5} \text{ mm}^3 \text{ N}^{-1} \text{ m}^{-1}$. The term *scuffing* is often used to describe this regime of wear in concentrated contacts, which is associated with the breakdown of lubrication.

The breakdown of partial EHL, leading either to boundary lubrication (i.e., a transition between regimes I and II) or at higher sliding speeds directly to unlubricated sliding and scuffing (I to III) is indicated by the lower transition curve (solid line) in Fig. 5.23. This transition can be understood by reference to equation 4.16, which shows that the film thickness in EHL depends, among other factors, on the viscosity of the lubricant and on the sliding velocity. Increasing either of these leads to a thicker film, and hence raises the load-carrying capacity of the contact. At very low sliding speeds (below about 1 mm s^{-1} in Fig. 5.23), the transition load does increase with sliding speed, since the viscosity of the oil remains effectively constant. At higher speeds, however, thermal effects become important, and largely determine the shape and location of the lower transition curve. As the sliding speed is increased, frictional heating in the contact zone leads to a drop in the local viscosity of the lubricant; this effect outweighs that of the increased sliding velocity in equation 4.16, and the film thickness drops. The load-carrying capacity of the contact therefore falls with increasing speed. To a good approximation the falling part of the lower transition curve fits the equation:

$$WU^n = \text{constant} \quad (5.10)$$

The value of the exponent n varies with the temperature-dependence of viscosity of the lubricant, and lies in the range 0.3–0.8 for a wide range of mineral and synthetic oils.

The lower transition is largely associated with mechanical effects, being controlled primarily by the roughness of the surfaces, the lubricant viscosity and its dependence on temperature and pressure. A subsidiary role is also played by chemical effects, as indicated by the fact that transition loads in the absence of oxygen may be significantly lower than in air; the composition of the lubricant, particularly the presence of boundary lubricating or EP agents (see Section 4.6) is also important.

The upper transition (the broken curve in Fig. 5.23), from boundary lubrication (regime II) to unlubricated sliding and scuffing wear (regime III) depends, in contrast, mainly on chemical effects. It is independent of the viscosity of the lubricant and of the initial roughness of the surfaces, but strongly influenced by the lubricant chemistry and by the microstructure and composition of the surfaces.

The phenomenon of scuffing in lubricated systems is of great practical importance, since it leads to unacceptably high friction and rates of wear. There is evidence that scuffing first occurs when the surface temperature reaches a critical value, possibly because this causes some desorption of the boundary lubricant, resulting in exposure of a critical fraction of bare, unprotected surface. Scuffing can therefore be avoided by ensuring a low rate of frictional energy dissipation, through low friction, low sliding speed or low load. As can be seen from the transition diagram (Fig. 5.23), at low sliding speeds some warning of impending scuffing may be given, since the system must pass through the boundary lubricated regime (II) before full scuffing occurs. At higher speeds, however, above the bifurcation speed U_s , the system will immediately move with increasing load from the desirable state of partial EHL (regime I) to catastrophic scuffing (regime III).

A tenacious boundary lubricant film, together with suitable EP or antiwear additives, is very valuable in increasing the scuffing load. Oxidation of the surface can also be valuable, both in aiding adhesion of the boundary lubricant, and through the lubricating action of the oxide itself. Fe_3O_4 , for example, provides a ductile film of low shear strength on steel asperities during lubricated sliding in air. Fe_2O_3 , in contrast, forms on steels during sliding in the presence of water, and is abrasive; it enhances rather than reduces the propensity for scuffing.

The process of running-in under lubricated conditions usually involves the formation of beneficial oxide films, typically tens of nanometres thick, by local thermal oxidation of high spots, as well as the smoothing of the surfaces by wear of the higher asperities. Solid lubricant films, and other surface coatings can also raise the upper transition curve very substantially.

The other important factor that controls the onset of scuffing is the nature of the sliding surfaces. Dissimilar metals show less tendency for adhesion at asperity contacts than similar metals (see Sections 3.5 and 8.6.2), and should therefore show higher scuffing loads. However, in many engineering applications the use of steels is dictated by reason of strength and economy. Alloying elements in steels have been shown to influence the scuffing transition, both by changing the chemical composition of the surface and therefore the strength of adhesion of boundary lubricants, and also by altering the microstructure of the steel; for example, an increase in the proportion of retained austenite in a steel tends to produce a downward shift in the upper transition curve.

Similar effects are observed in fretting under lubricated conditions. Fretting damage may be reduced by the presence of the lubricant, but can instead be enhanced due to the reduction in the rate of oxide formation associated with the presence of the lubricant. As with sliding wear, the lubricant viscosity will be affected by sliding velocity (dependent upon stroke and frequency of oscillation in fretting) and the applied load. Lubrication of a fretting contact will change the mechanism and rate of degradation;

this may be beneficial, but care needs to be exercised that fatigue-inducing plasticity-dominated damage does not ensue. Significant damage has been observed in grease-lubricated fretting contacts under small fretting strokes, where the lubricant entrainment into the contact is poor and the grease rather effectively limits the oxidation of the wearing surfaces (and the reduced wear rates that commonly follow).

5.10 SLIDING WEAR OF CERAMICS

5.10.1 Introduction

We have seen in [Section 3.6](#) that ceramic materials differ from metals in the nature of their interatomic bonding, and that this leads to very limited capacity for plastic flow at room temperature. In comparison with metals, they are therefore much more inclined to respond to stress by brittle fracture rather than by gross plastic deformation.

In some ceramics, notably oxides, crack growth is sensitive to environmental influences. In alumina, for example, the presence of water increases the crack growth rate dramatically. Environmental factors can also influence plastic flow in many ceramics, by affecting the mobility of near-surface dislocations. This example of a *chemo-mechanical effect* is also known as the *Rehbinder effect*. Surface chemical reactions, such as oxidation in susceptible materials, occur in ceramics just as in metals, and the rates of these reactions may be greatly enhanced by sliding.

We therefore see similarities in sliding wear between ceramics and metals, in that some local plastic flow can occur in both, and tribochemical reactions can modify the sliding interface and also lead to the formation of transfer films. However, gross plastic flow is much less important in ceramics, and tribochemical effects are correspondingly more significant than in metals. Brittle fracture, essentially absent in metals, plays a major role in the sliding wear of ceramics. Environmental factors, through their influence on both flow and fracture, are very important.

5.10.2 Brittle fracture

As we saw in [Section 3.6](#), the coefficient of friction for the unlubricated sliding of ceramics in air lies typically in the range 0.25–0.8. At low applied loads, the tangential forces will be too low to cause fracture, and sliding under these conditions will result in relatively low rates of wear, with mechanisms of wear that do not involve fracture (see [Fig. 5.24a](#)). Sliding under higher loads results in significant tangential forces, which may lead to fracture rather than plastic flow. Typically, such fracture takes place along grain boundaries, leading to the removal of individual grains and a rough surface as seen in [Fig. 5.24\(b\)](#).

The relative ease of plastic flow in ceramic materials is usually strongly anisotropic, depending on crystallographic orientation. For example, sliding on single crystals of Al_2O_3 parallel to prismatic planes in the c -axis direction leads to extensive fracture, whereas in other directions on these planes, or on the basal plane, it is accompanied by plastic flow. It is therefore quite possible for the surface of a polycrystalline ceramic to

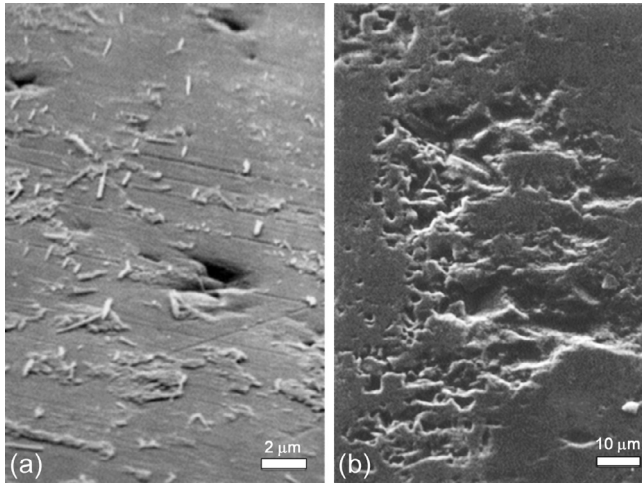


Fig. 5.24 Scanning electron micrographs of wear tracks formed on an Al_2O_3 ball against a flat Al_2O_3 specimen tested in air at 11.5 mm s^{-1} sliding speed under two loads: (a) 9.8 N where cylindrical wear debris (rolls) and some grooves were observed on the wear track; (b) 62.7 N where the loss of entire grains by intergranular fracture is observed (from Fischer, T.E., et al., 2000. *Wear* 245, 53–60)

exhibit regions of intergranular fracture alongside areas of predominantly plastic flow, associated with differences in the crystallographic orientation of the grains.

Concentrated (counterformal) contact under higher loads may result in fracture on a much larger scale, and can be understood in terms of the Hertz elastic stress distribution. (For contact by a sharp hard indenter the stress field and the resulting fracture are different, and discussed in [Section 6.3.2](#)) For a sphere pressing against a plane surface (with no traction), we saw in [Section 2.5.1](#) that the radius of the contact circle, a , is proportional to the cube root of the normal load, w (equation 2.14). The radial tensile stress σ_r in the plane surface outside the contact area reaches a maximum value $\sigma_{\max}$ just at the edge of the circle of contact, given by

$$\sigma_{\max} = p_o \left(\frac{1 - 2\nu}{3} \right) \quad (5.11)$$

Here p_o is the maximum normal stress exerted on the contact area, and ν is Poisson's ratio. It is this tensile stress component $\sigma_{\max}$ that gives rise to fracture. When the normal load reaches a critical value, a crack initiates just outside the contact circle, and rapidly propagates to form a Hertzian cone crack, as shown in [Fig. 5.25](#).

If a tangential force is applied to the contact (see [section 2.5.3](#)), as in sliding, then the stress distribution is modified and the normal load necessary to initiate fracture is greatly reduced. For example, in experiments with a TiC sphere on a flat of the same material, the critical normal load for fracture was reduced by a factor of 10 under sliding conditions in air (with $\mu \approx 0.2$), compared with its value with no tangential force.

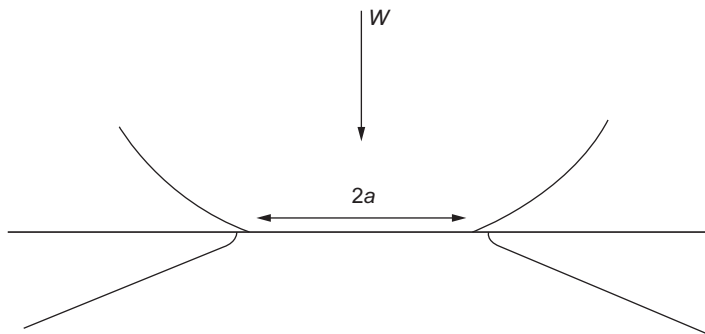


Fig. 5.25 Sectional view showing the geometry of a Hertzian cone crack formed by a sphere loaded normally on to the plane surface of a brittle material

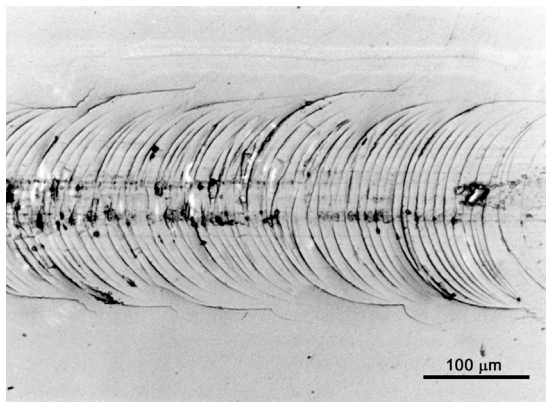


Fig. 5.26 Series of arc-shaped fractures caused by sliding of a sphere over a brittle solid under normal load. This example shows the damage due to sliding of a tungsten carbide sphere on a soda-lime glass plane. The sphere slid from left to right

Sliding in vacuum resulted in a greater tangential traction ($\mu \approx 0.9$) and a corresponding further reduction, by another factor of 50, in the normal force needed for cracking. Under sliding conditions, the cracks which form no longer intersect the surface in complete circles, but in a series of arcs as shown in Fig. 5.26.

Ceramics with low fracture toughness form such cracks more readily than tougher materials, and cracking is also favoured by high contact stresses (e.g., high loads and concentrated contact) and by large surface tractions (i.e., a high coefficient of friction).

5.10.3 Wear transitions: mild and severe wear

Sliding wear mechanisms in ceramics can be dominated by fracture, tribochemical effects or plastic flow. Transitions between regimes characterized by each of these commonly lead to sharp changes in wear rate with load, sliding speed or environmental conditions (e.g., humidity or oxygen content), as illustrated by Fig. 5.24. The terms *mild wear* and *severe wear* are often used, as for metals, to describe conditions on either side of a transition.

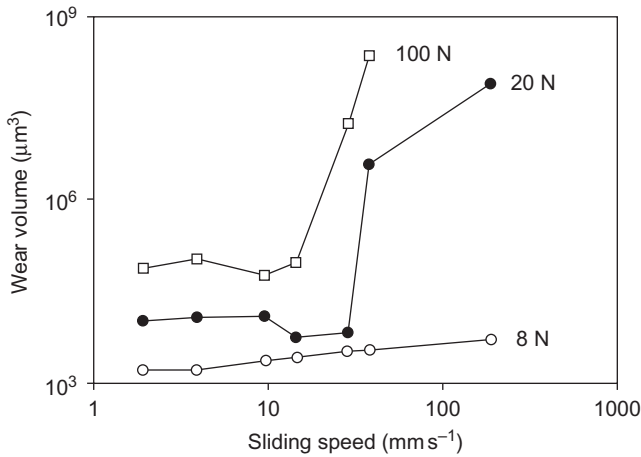


Fig. 5.27 Wear volumes for polycrystalline alumina as a function of sliding speed for three different loads, derived from experiments in a ball-on-three-flats sliding geometry (after Hsu, S.M., Shen, M.C., 1996. *Wear* 200(1–2), 154–175)

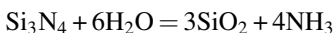
Mild wear in ceramics is associated with a low wear rate, smooth surfaces, a steady friction trace and mechanisms of wear dominated by plastic flow or tribochemical reactions (as shown in Fig. 5.24a). The wear debris is often finely divided, and may be chemically different from the bulk sliding material, for example through oxidation or hydration.

Severe wear, in contrast, causes a higher wear rate together with a rougher surface (see Fig. 5.24b), a fluctuating friction trace and mechanisms of wear dominated by brittle fracture. The wear debris is often angular and not chemically different from the substrate. Ceramic materials generally obey the Archard equation (equation 5.7) well enough, by showing linear dependence of wear on sliding distance and load, for the wear coefficient (K) or specific wear rate (k) to be useful quantities. Typical values of K for engineering ceramics undergoing severe wear are 10^{-4} – 10^{-2} , with K for mild wear being lower by factors of 10–100 or even more.

Figure 5.27 shows data for the sliding wear of polycrystalline alumina specimens, derived from experiments with a sphere-on-flat geometry under dry conditions. At low loads, the system exhibits mild wear across the range of sliding speeds examined. Under higher loads, transitions to severe wear, involving intergranular fracture, can be induced by increases in the sliding velocity, with the transition taking place at lower velocities as the load is raised. Such data can be visualized in the form of a wear map.

5.10.4 Chemical effects

Sliding of Si_3N_4 in air containing water vapour results in the formation of a surface layer of hydrated silicon oxide by the reactions outlined in Section 3.6, of which the initial step is the formation of SiO_2 :



The formation of silica, followed often by some degree of hydration, provides the source of the wear debris. Having a lower shear strength than the bulk ceramic, this silica layer also lowers the coefficient of friction. It is possible that adsorbed water on the surface also increases the ease of plastic flow by the Rehbinder effect (see [Section 5.10.1](#)). As the sliding speed is raised, increasing the interface temperature, the tribochemically reacted layer ceases to provide protection and the coefficient of friction rises. The increased surface shear stress causes cracking, and a transition ensues from mild wear (largely tribochemical in origin) to severe wear (with extensive brittle fracture). The resulting increase in surface roughness is responsible for the rise in the coefficient of friction. Self-mated silicon nitride sliding submerged in water can exhibit a coefficient of friction as low as 0.002. A mechanism has been proposed in which, when sliding in water, the silica layer builds to a critical thickness before delaminating (Fig. 5.28).

The effect of water (either as a vapour or as a liquid) on the wear of Si_3N_4 provides an example of the environmental sensitivity commonly observed in the sliding wear of ceramics. [Figure 5.29](#) shows how the same transition in Si_3N_4 can be induced in a different way: by changing the humidity of the surrounding air. A similar tribochemical reaction occurs in SiC , also leading to the formation of silicon oxide, but in this case the presence of water leads to an increase in wear rate.

Alumina and zirconia also show strong sensitivity to water, which in both ceramics also causes the wear rate to increase, typically tenfold in comparison with dry sliding. This phenomenon may reflect the increased surface plasticity of these ceramics in the presence of water (the Rehbinder effect) which can also be detected from microhardness

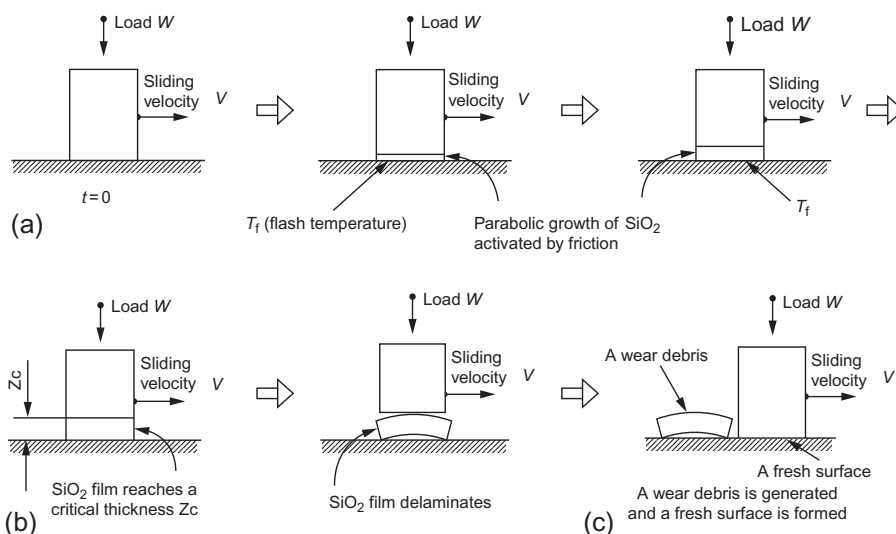


Fig. 5.28 A conceptual model for the tribochemical wear of silicon nitride sliding in water: (a) SiO_2 film grows; (b) SiO_2 film reaches a critical thickness and delaminates; (c) wear debris is generated and a fresh surface is formed (from Xu, J., Kato, K., 2000. *Wear* 245(1–2), 61–75)

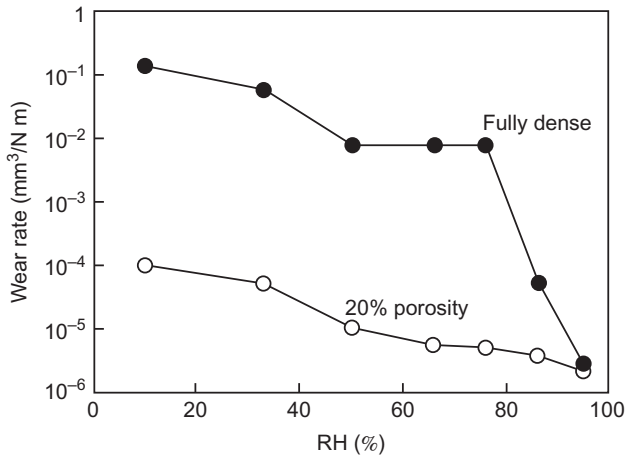


Fig. 5.29 Variation of wear rate of silicon nitride pair as a function of relative humidity for a dense silicon nitride sphere sliding against silicon nitride discs, both fully dense and with 20% porosity (from Takadoun, J., Houmid-Bennani, H., Mairey, D., 1998. *J. Eur. Ceram. Soc.* 18(5), 553–556)

measurements; on the other hand, it may be associated with enhanced crack growth (stress-corrosion cracking) which is common in oxide ceramics exposed to water.

Chemical effects can also be important in sliding at high speeds, where surface heating leads to increased rates of reaction. Oxidation, for example, is the major mechanism of wear in carbon/carbon composite materials used in high performance (e.g., aircraft) brakes. Many other non-oxide ceramics oxidize at high interfacial temperatures. Phase transformations may also result from the high temperatures; for example α - Si_3N_4 is formed on the sliding surface of β - Si_3N_4 in this way, and the rapid wear of diamond if it is used to machine or grind steel is caused by local transformation of the diamond to graphite.

5.10.5 Lubricated wear

Ceramic materials respond to conventional lubricants in a similar way to metals. Effective lubrication lowers the incidence of asperity contact, reduces the surface shear forces, and decreases the wear rate. However, even under lubricated conditions chemical effects can be very important. Surface oxidation of Si_3N_4 and SiC occurs even in mineral oils, and the mechanism of lubricated wear of these ceramics is probably tribochemical. Boundary lubricants such as octadecanoic acid (stearic acid) function on ionic materials to which their polar end groups can bond; for example, Al_2O_3 is lubricated effectively by stearic acid, whereas SiC (with covalent bonding) is not. However on some ionic ceramics, notably Al_2O_3 and ZrO_2 , such lubricants lead to sharply enhanced wear rates despite the lowered coefficient of friction. This effect appears to be due to stress-corrosion cracking which leads to intergranular fracture.

A phenomenon that can occur in the lubricated sliding of materials with low thermal conductivity and high thermal expansion coefficients is *thermoelastic instability*. Zirconia is particularly susceptible to this. Contact between asperities in thin-film lubricated sliding causes intense local power dissipation and consequent heating. In a material that conducts heat poorly and shows high thermal expansion, the local temperature rise can significantly distort the topography of the surface and cause further asperity contact to be concentrated in the same area. The instability results in very high flash temperatures and rapid wear, and is responsible for poor performance of zirconia under lubricated sliding conditions.

5.11 SLIDING WEAR OF POLYMERS

5.11.1 Introduction: interfacial and cohesive wear

In contrast to metals and ceramics, polymers exhibit lower coefficients of friction, with values typically between 0.1 and 0.5, whether self-mated or sliding against other materials (see [Section 3.8](#)). They are therefore often used unlubricated in tribological applications such as bearings and gears, usually sliding against harder counterfaces.

Polymers are much more compliant than metals or ceramics, with values of elastic modulus lower by factors of 10–100. Their strengths are also much lower, and it is therefore reasonable to consider metallic or ceramic counterfaces when sliding against polymers to act as rigid bodies. Nearly all the deformation due to contact or sliding takes place within the polymer, and the surface finish of the hard counterface has a strong influence on the mechanism of the resulting wear.

If the counterface is smooth, then wear may result from adhesion between the surfaces, and involve deformation only in the surface layers of the polymer. On the other hand, if the counterface is rough then its asperities will cause deformation in the polymer to a significant depth; wear then results either from abrasion associated with plastic deformation of the polymer, or from fatigue crack growth in the deformed region. These two classes of wear mechanism, involving surface and subsurface deformation respectively, have been termed *interfacial* and *cohesive* wear processes.

As with metals and ceramics, chemical effects can also play important roles in the wear of polymers, for example through environmental effects on fatigue crack growth, or through surface degradation which constitutes a further example of an interfacial wear process. We shall examine the various cohesive and interfacial wear processes in the following sections.

The level of counterface roughness at which the transition from interfacial to cohesive wear mechanisms occurs depends on the nature of the polymer, but corresponds typically to R_a values between 0.01 and 1 μm . In general, susceptible polymers sliding on highly polished hard counterfaces will experience adhesive wear, while turned or ground surfaces (see [Table 2.1](#)) promote cohesive mechanisms. The transition between the two regimes can lead to a minimum in wear rate at a certain surface roughness, as illustrated in [Fig. 5.30](#). The roughness at which the transition occurs can also be influenced by environmental factors and especially by the presence of species that reduce adhesion at the interface.

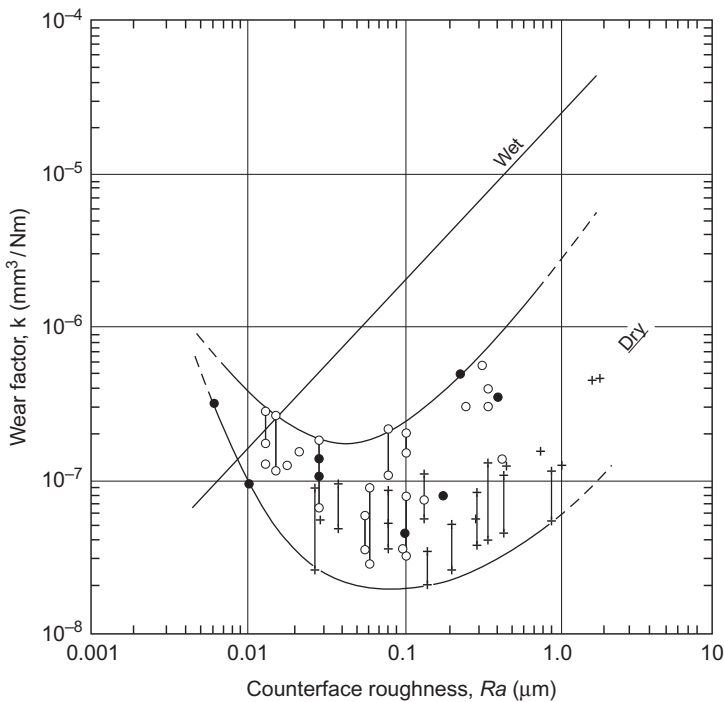


Fig. 5.30 Wear rate of ultra-high molecular weight polyethylene (UHMWPE) sliding against a steel counterface, as a function of roughness of the steel surface (from Dowson, D., et al. 1985. Influence of counterface topography on the wear of ultra-high molecular weight polyethylene under wet and dry conditions. *Polym. Wear Control* 12, 171–187)

5.11.2 Cohesive wear mechanisms

Cohesive wear results from the deformation of surface and subsurface material, caused by the passage of a protuberance on the counterface over the polymer surface. The protuberance may be an asperity on a hard surface arising from its topography, or a particle of harder material partially embedded into a softer counterface, or possibly a lump of polymeric debris transferred to the counterface. The resulting deformation in the polymer may be either plastic (permanent) or elastic (recoverable). In the first case the mechanism of wear can be termed abrasion, while in the second it is associated with fatigue.

The distinction between these two types of mechanism in the wear of polymers is not sharp. It is fairer to envisage a progressive change, rather than an abrupt transition from one mechanism to the other as the mechanical properties of the polymer are changed. [Figure 5.31](#) shows schematically how the relative proportions of abrasion and fatigue depend on the surface roughness and the nature of the polymer. In elastomers, for example, with low elastic modulus, the contact deformation will be almost

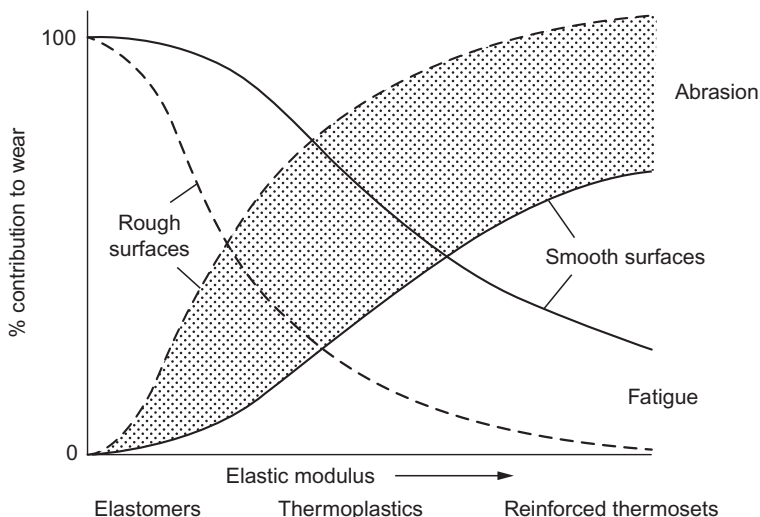


Fig. 5.31 Schematic diagram showing the variation in importance of the abrasion and fatigue mechanisms in polymers, as defined in the text, with the elastic modulus of the polymer and the roughness of the counterface (from Evans, D.C., Lancaster, J.K., 1979. In: Scott, D. (Ed.), *Wear, Treatise on Materials Science and Technology*, vol. 13. Academic Press, pp. 85–139)

totally elastic and fatigue mechanisms will dominate. High modulus polymers such as thermosets, on the other hand, show appreciable plastic deformation due to asperity contact and suffer wear by abrasion. We shall now examine the processes of wear by abrasion and fatigue.

Abrasive wear is discussed in detail in [Section 6.3](#). The mechanisms of interest here result from plastic displacement (ploughing or cutting) of the polymer by the rigid protuberances on the counterface. For this to occur, the contact must be plastic, and the value of the plasticity index ψ discussed in [Section 2.5.6](#) can be used to establish whether this will be so. The term E^*/H in the expression for ψ (see equation 2.42) is much smaller (one fifth or less) for contacts involving polymers than it is for contacts between metals. The other term in the plasticity index expression depends on the surface roughness; the roughness at which the contact becomes predominantly plastic is therefore much greater for polymers than for metals. For polymers the plastic regime typically starts at a counterface roughness (Ra value) greater than a few micrometres.

The simple theory of abrasive wear by plastic deformation presented in [Section 6.3.1](#) predicts that the wear rate should be proportional to the normal load. This behaviour is indeed seen in polymers for wear by this mechanism. Equation 6.4 also predicts that the wear rate should be inversely proportional to $\tan \alpha$, where $(\pi/2 - \alpha)$ is the mean surface slope and depends on the mean roughness Ra . Abrasive wear rates of polymers do increase with counterface roughness, but considerably more rapidly than equation 6.4 suggests.

The theory also predicts that the wear rate should be inversely proportional to the hardness, H . In practice, H correlates only with the general trend of behaviour for polymers in this regime of wear, as can be seen from Fig. 6.19. The correlation of abrasive wear rate with hardness is much worse than for metals. The reason for this disparity is twofold. As discussed in Appendix A, hardness defined as the load per unit area of the residual impression in an indentation test provides a measure of rather different properties for metals and for polymers. In metals it accurately reflects the ease of plastic flow and correlates closely with the yield stress, whereas in polymers, much of the material displaced around the indentation is accommodated elastically and the measured hardness incorporates contributions from both elastic and plastic sources. The second reason for the poor correlation is that even in the harder polymers sliding on rough counterfaces, appreciable elastic deformation occurs; this causes damage by fatigue processes that are not controlled by plastic properties.

A much better correlation is found between abrasive wear rates and values of $1/\sigma_u \epsilon_u$, where σ_u and ϵ_u are the ultimate tensile stress and elongation for the polymer, measured in a conventional tensile test. Although the conditions under which σ_u and ϵ_u are measured differ considerably, especially in terms of strain rate, from those at a sliding contact, the validity of this correlation, sometimes called the *Ratner-Lancaster correlation*, is widely recognized. Figure 5.32 shows the good agreement found for several thermoplastics.

The product $\sigma_u \epsilon_u$ is roughly proportional to the area under the stress-strain curve for the polymer to the point of tensile failure, and thus provides a measure of the work done in producing tensile rupture.

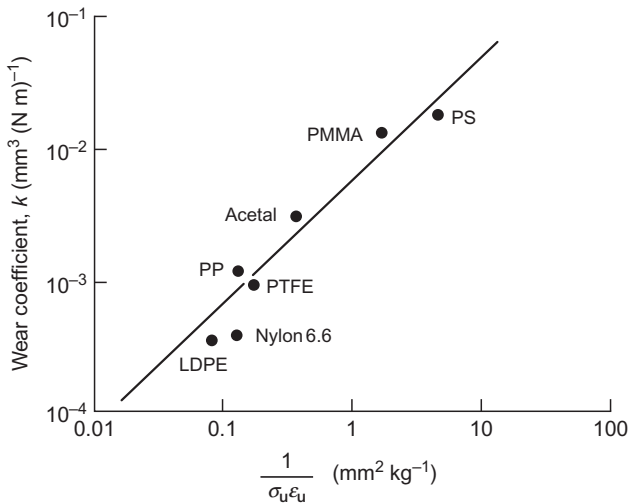


Fig. 5.32 The Ratner-Lancaster correlation between the wear rates of polymers under predominantly abrasive conditions and the reciprocal of the product of the stress and strain at tensile rupture (from Briscoe, B.J., 1981. Tribol. Int. 14, 231–243)

Figure 5.31 shows that for low modulus polymers, such as the softer thermoplastics and elastomers, fatigue processes occur rather than abrasion. Fatigue is also important in harder polymers sliding against smooth counterfaces. Wear due to fatigue results from the formation of cracks associated with predominantly elastic deformation. Damage is cumulative, and develops over a number of contact cycles. Particles of wear debris become removed by the growth and intersection of cracks.

In a simple model of fatigue wear, the wear rate would be expected to correlate with the rate of fatigue crack growth. This usually follows the Paris equation:

$$\frac{da}{dN} = A(\Delta K)^n \quad (5.12)$$

where da/dN is the increase in crack length a per stress cycle, ΔK is the range of stress intensity to which the growing crack is exposed during each cycle, and A and n are empirical constants. The use and validity of this equation for fatigue crack growth in many materials are discussed in basic texts on materials science and engineering. The value of the exponent n lies typically in the range from 1.5 to 3.5 for elastomers and 3–10 for rigid thermoplastics and thermosets.

ΔK is proportional to the range of stress $\Delta\sigma$ experienced by elements of the polymer during sliding. If it is assumed that the crack length at which material becomes detached is much greater than the size of the defect from which it grows, then integration of equation 5.12 shows that the number of cycles needed to remove a wear fragment, N_f , is proportional to $1/\Delta\sigma^n$. The wear rate will be inversely proportional to N_f . $\Delta\sigma$ can be assumed for simplicity to be the mean elastic contact stress due to a single spherical asperity of radius r under a load w , and depends on these quantities as follows (see Section 2.5.1):

$$\Delta\sigma \propto w^{1/3} r^{2/3} \quad (5.13)$$

The radius a of the contact area is given by:

$$a \propto w^{1/3} r^{1/3} \quad (5.14)$$

If the breadth and width of the wear particle removed after N_f cycles are both proportional to a , and its depth is independent of a (being just the crack length at which detachment occurs), then the volume of material removed per unit sliding distance, q , will be given by:

$$q \propto \frac{a^2}{N_f} \quad (5.15)$$

Hence, combining the preceding equations,

$$q \propto r^{2(1-n)/3} w^{(2+n)/3} \quad (5.16)$$

We would therefore expect the overall wear rate due to fatigue to depend strongly on the surface roughness (since this controls r and also the number of asperity contacts per unit area). This is indeed observed: a tenfold increase in the roughness of the counterface (e.g., from 0.1 to 1 $\mu\text{m } Ra$) can lead to an increase in wear rate of a polymer by as much as one hundred times.

Strong dependence of wear rate on normal load is also observed, often following the power law suggested by equation 5.16. The exponent of load depends on both the nature of the polymer and the topography of the counterface.

Despite the success of fatigue models in explaining the dependence of wear on load and roughness, direct microscopic evidence for the presence of fatigue cracks during dry sliding wear is sparse. This may be because they are obscured by other changes in the surface layers of chemical or thermal origin. However, strong evidence for a link between wear and fatigue is seen from experiments carried out in liquid environments. Figure 5.33 shows how the wear rate of polyethersulfone (PES) sliding against a stainless steel counterface depends on the nature of the liquid present. Under these sliding conditions fluid film lubrication did not occur, nor was abrasion an important mechanism.

Also plotted in the figure are the values of A and n from the Paris equation (equation 5.12), derived from conventional fatigue experiments carried out in the same liquids. On the horizontal axis are plotted values of δ_s , the solubility parameter for the

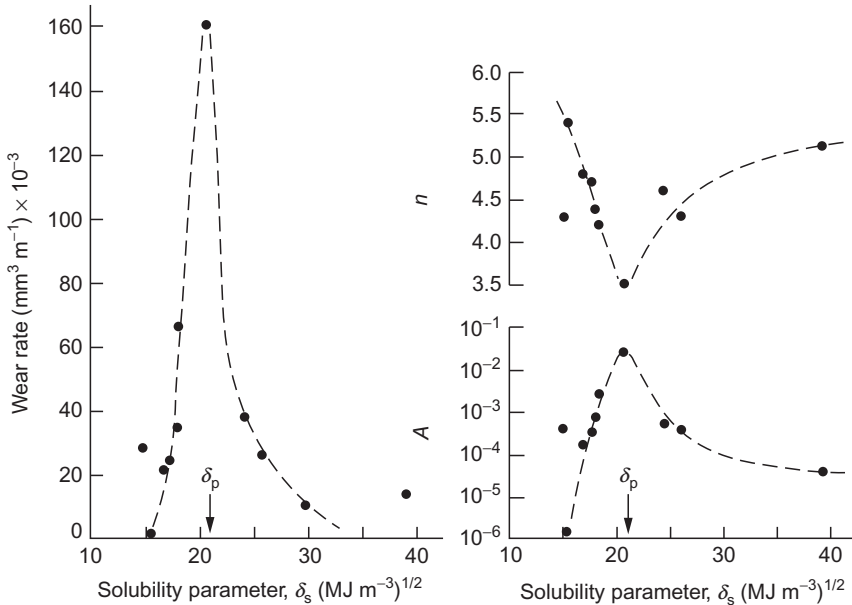


Fig. 5.33 The wear rate of polyethersulfone (PES) sliding against a smooth stainless steel counterface in a variety of solvents under boundary lubrication conditions. The liquids are described by the values of the solvent solubility parameter, δ_s . Also shown are the values of A and n from the Paris equation 5.12 (from Atkins, A.G., Omar, M.K., Lancaster, J.K., 1984. *J. Mater. Sci. Lett.* 3, 779–782)

solvent. The value of δ_s provides a measure of the chemical affinity between the liquid and the polymer; when δ_s is equal to δ_p , the solubility parameter for the polymer, this affinity is a maximum. Two important points arise from Fig. 5.33. First, the wear rate correlates strongly with the mutual affinity of the solvent and the polymer; similar results are found for many, though by no means all, other polymers. Second, the wear rate correlates even more strikingly with the crack growth rates measured in conventional fatigue tests.

Elastomers are distinguished from other engineering polymers by their low elastic modulus and large values of ultimate tensile elongation. True abrasive wear by a cutting mechanism occurs only on sliding against counterfaces with extremely sharp, needle-like asperities, or in lubricated conditions (such as in water—see Fig. 5.34a). More usually, wear occurs by progressive crack growth, with cracks propagating from the surface under cyclic loading. A rough counterface is not essential for the development of these cracks, since they can arise directly from surface tractions associated with sliding on a smooth counterface with a high coefficient of friction. This crack growth process in elastomers often leads to the formation of a characteristic ‘abrasion pattern’ on the surface.

The pattern, as shown in Fig. 5.34(b) consists of regular parallel ridges lying perpendicular to the direction of sliding. In cross-section the ridges have a saw-tooth profile; each ridge is associated with the progressive growth of a single crack into the rubber. Excellent agreement is found between wear rates measured for this type of wear and the predictions of a model based on incremental crack growth. The wear rate depends strongly on the frictional traction at the surface. With a high coefficient of friction, and for elastomers of low strength, the crack may propagate far enough in one contact cycle to generate a wear particle; this debris can be deformed by rolling between the sliding surfaces into a characteristic elongated shape, leading to this wear process being sometimes termed ‘roll formation’.

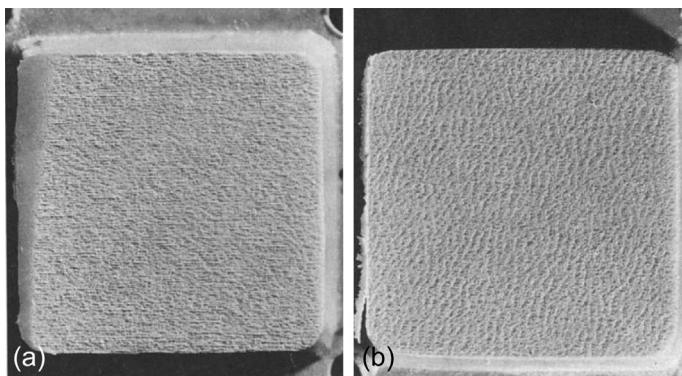


Fig. 5.34 Patterns formed during abrasion of unfilled nitrile rubber on P60 silicon carbide paper (rubber sliding to the right). The test pieces were 25 mm square. (a) Wet with score lines; (b) dry with classical abrasion pattern (from Muhr, A.H., Roberts, A.D., 1992. Rubber abrasion and wear. *Wear* 158(1) 213–228)

5.11.3 Interfacial wear mechanisms

In interfacial wear, material is removed by processes occurring close to or in the surface of the polymer. The most important such process is wear by adhesion, which is directly related to the adhesive component of sliding friction discussed in Section 3.8.3.

Adhesive wear occurs only when the counterface is smooth, and involves the transfer of polymer to the harder counterface and its subsequent removal as wear debris. In this respect the process is similar to that seen in metals. A running-in period often occurs before steady-state conditions are reached; during steady-state wear, the wear rate is often directly proportional to the normal load over quite a large range (see Fig. 5.35), in conformity with the Archard equation (equation 5.7). Corresponding values of the specific wear rate k lie typically in the range 10^{-6} – 10^{-3} mm³ N⁻¹ m⁻¹.

Not all polymers show adhesive wear of this type. Thermosets, for example, do not form transfer films, but wear instead by fatigue processes or abrasion, or at high interfacial temperatures by thermal degradation (pyrolysis) of the surface—another form of interfacial wear. In amorphous polymers such as polystyrene, polyvinylchloride (PVC) and polymethylmethacrylate (PMMA) below their glass transition temperatures (T_g), the interface between polymer and counterface is weaker than the bulk polymer, and wear occurs mainly by fatigue or abrasion as discussed in the previous section. Once the interface temperature rises above T_g , however, amorphous thermoplastics wear by adhesive transfer.

Wear results from the adhesion of polymer material to the counterface by electrostatic forces (including van der Waals forces). The junction between the polymer and the counterface is stronger than the polymer itself, and failure occurs within its bulk, leaving a transferred fragment. Repeated sliding over the counterface leads to the progressive build-up of a transferred layer which eventually becomes detached. Under

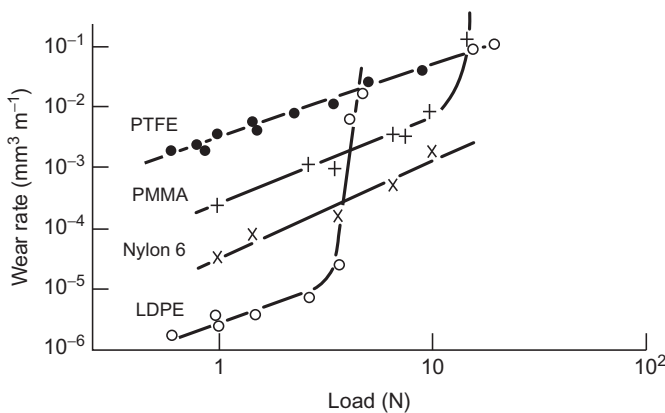


Fig. 5.35 The variation of steady-state wear rate with normal load for polymers sliding against a smooth mild steel counterface ($R_a = 0.15$ μm) under unlubricated conditions (from Evans, D.C., Lancaster, J.K., 1979. In: Scott, D. (Ed.), *Treatise on Materials Science and Technology*, vol. 13, pp. 85–139)

these conditions, sometimes known as ‘normal’ transfer, polymer is transferred to the counterface without significant chain scission or chemical degradation, in irregular lumps or patches typically 0.1–10 μm thick. The polymer chains may show some orientation in the direction of sliding. Back-transfer to the polymer may also occur, as may incorporation of material from the counterface into the transfer layer. The rate of wear appears to be dictated by the rate of removal of the transfer film from the counterface, rather than by the rate of polymer transfer into the film. There is evidence that polymers are reluctant to transfer further on to their own transfer films, and if the transfer film adheres strongly to the counterface, then the wear rate is low.

A small number of polymers, notably polytetrafluoroethylene (PTFE), high density polyethylene (HDPE) and ultra-high molecular weight polyethylene (UHMWPE), form transfer layers of a rather different type under certain conditions, which are associated with low friction and wear rates. These were discussed in [Section 3.8.3](#). At low sliding speeds (typically $<0.01 \text{ m s}^{-1}$), a very thin transfer layer is laid down, perhaps 5–10 nm thick. In this layer the polymer chains are highly oriented in the direction of sliding. The layer adheres well to the counterface, and further sliding occurs between the surface of the bulk polymer, which contains similarly oriented molecules, and the transfer film on the counterface.

The conditions under which this regime of low friction and wear is maintained are precarious; a return to ‘normal’ transfer occurs with rougher counterfaces or higher sliding velocities. However, the thin oriented transfer layer can be stabilized by measures which improve its adhesion to the counterface. Suitable fillers in the polymer will do this and lower the wear rate by a factor of 10^2 – 10^3 , without significant change in the coefficient of friction. The incorporation of lead and copper oxides into HDPE, for example, provides an effective enhancement of its wear resistance against steel, although details are unclear of the chemical mechanisms by which the adhesion of the transfer film is improved. Metallic fillers (e.g., bronze) also appear to contribute to a chemical enhancement of film adhesion. Other fillers, for example carbon particles, are thought to reduce wear rates by abrading the counterface slightly; this process will reduce the surface roughness and also provide a cleaner surface, both tending to increase the adhesion of the transfer layer. Fillers of these types are widely used in composite polymeric bearing materials (see [Section 9.2.3](#)), which may also contain solid or dispersed liquid lubricants, as well as solid filler particles that carry a proportion of the normal load.

Thermoplastic polymers generally have low thermal conductivities and low softening or melting temperatures compared with metals; both factors lead to marked thermal softening under much milder sliding conditions than would be needed for metals. The phenomenon is, however, exactly the same as the transition from effectively isothermal to almost adiabatic sliding conditions discussed in [Section 5.6](#). The same method of analysis used to generate the curves in [Fig. 5.14](#) indicates that the rapid increases in wear rate of low density polyethylene (LDPE) and polymethylmethacrylate (PMMA) at the higher loads in [Fig. 5.35](#) arise from thermal softening. These transitions occur when the flash temperature at the interface reaches the softening point of the polymer. Above this thermal transition the rate of wear is high, with material being removed from a molten layer at the surface of the polymer.

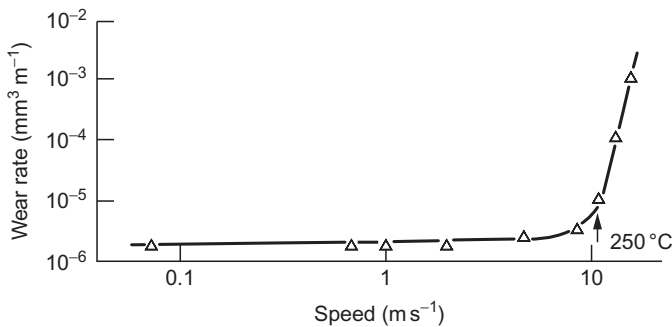


Fig. 5.36 The variations of steady state wear rate with sliding speed for Nylon 6.6 sliding against a smooth mild steel counterface ($Ra = 0.15 \mu\text{m}$) under unlubricated conditions (from Evans, D.C., Lancaster, J.K., 1979. In: Scott, D. (Ed.), *Treatise on Materials Science and Technology*, vol. 13, pp. 85–139)

The transition has been described as one from ‘mild’ to ‘severe’ wear, although these terms have different meanings from those used for metals or ceramics.

The same effect is also responsible for a sharp increase in wear rate for thermoplastics above a certain sliding speed, as shown for Nylon 6.6 in Fig. 5.36; the arrow indicates the speed at which, under these experimental conditions, the calculated flash temperature reaches 250°C , the softening point of the polymer.

The nature of the counterface also influences the conditions at which thermal softening occurs. For a metallic counterface, a good conductor of heat, the critical speed may be some hundred times higher than for a polymeric counterface with much lower conductivity. Part of the beneficial effect of metallic fillers in reducing the wear rates of polymer composites in bearing applications (see Section 9.2.3) is due to the increase in thermal conductivity caused by the filler.

5.11.4 Lubricated wear of polymers

Hydrodynamic lubricant films or even EHL films (see Chapter 4) will separate the sliding surfaces and reduce the contact stresses, lowering the wear rates due to all the mechanisms discussed above. Polymers, with their low values of elastic modulus, remain in the EHL regime at much lower sliding speeds and higher normal loads than metals. In the regime of boundary lubrication, however, when the load is carried predominantly by solid contact, the mechanical effects of lubrication are often of secondary importance compared with the chemical and physical effects of the lubricant on the surface of the polymer. Although lubricants always reduce the coefficient of friction, they may either increase or decrease the rate of wear.

Boundary lubricants, reviewed in Section 4.6, act by forming adsorbed layers on the sliding surfaces. They bond poorly to the low energy surfaces of polymers, and so have little effect on the friction and wear of self-mated polymers, although they may still play a useful role when they can chemisorb on to a metallic counterface.

Liquid lubricants interact with polymers in several ways. In some cases, in glassy thermoplastics for example, they can cause increased plasticity of the surface which leads to reduced wear. However, the same polymers can also experience stress-crazing or cracking in other liquids, resulting in greatly enhanced wear. We have already seen in Fig. 5.33 how the relative affinity of the polymer and the liquid can influence the rate of fatigue crack growth, and hence the rate of wear by fatigue mechanisms.

Polymers which show low wear rates because they form thin oriented transfer films tend to suffer greater wear when lubricated, because the lubricant interferes with the adhesion of the transfer film to the counterface. Water has a particularly marked influence in this respect; for example, the wear rates of PTFE composites containing glass fibres or bronze particles, sliding against metal counterfaces, may be 100–1000 times greater when wet than when dry.

SOURCES AND FURTHER READING

- Briscoe, B.J., 1981. Wear of polymers: an essay on fundamental aspects. *Tribol. Int.* 14, 231–243.
- Briscoe, B.J., Tabor, D., 1978. Friction and wear of polymers. In: Clark, D.T., Feast, W.J. (Eds.), *Polymer Surfaces*. John Wiley, pp. 1–23.
- Buckley, D.H., Miyoshi, K., 1989. Tribological properties of structural ceramics. In: Wachtman, J.B. (Ed.), *Structural Ceramics, Treatise on Materials Science and Technology*. vol. 29, pp. 293–365.
- Fischer, T.E., 1988. Tribochemistry. *Annu. Rev. Mater. Sci.* 18, 303–323.
- Friedrich, K. (Ed.), 1986. *Friction and wear of polymer composites. Composite Materials Series*, vol. 1. Elsevier.
- Jahanmir, S., 2002. Wear transitions and tribochemical reactions in ceramics. *Proc. Inst. Mech. Eng. J. J. Eng. Tribol.* 216 (6), 371–385.
- Lim, S.C., Ashby, M.F., 1987. Wear-mechanism maps. *Acta Metall.* 35, 1–24.
- Ling, F.F., Pan, C.H.T. (Eds.), 1988. *Approaches to Modeling of Friction and Wear*. Springer-Verlag.
- Merchant, H.D., Bhansali, K.J. (Eds.), 1987. *Metal Transfer and Galling in Metallic Systems*. TMS.
- Peterson, M.B., Winer, W.O. (Eds.), 1980. *Wear Control Handbook*. ASME.
- Quinn, T.F.J., 1983. Review of oxidative wear. Part I: the origins of oxidative wear. *Tribol. Int.* 16 (5), 257–271.
- Rigney, D.A. (Ed.), 1981. *Fundamentals of Friction and Wear of Materials*. ASM.
- Rigney, D.A., 1988. Sliding wear of metals. *Annu. Rev. Mater. Sci.* 18, 141–163.
- Rigney, D.A., et al., 1990. Viewpoint set on materials aspects of wear. *Scripta Met.* 24, 799–844.
- Scott, D. (Ed.), 1979. *Wear, Treatise on Materials Science and Technology*. Vol. 13. Academic Press.
- Sullivan, J.L., 1987. The role of oxides in the protection of tribological surfaces. *Inst. Mech. Engrs. Conf. Series* 5, vol. 1, 283–302.
- Waterhouse, R.B., 1972. *Fretting Corrosion*. Pergamon.
- Xu, J., Kato, K., 2000. Formation of tribochemical layer of ceramics sliding in water and its role for low friction. *Wear* 245 (1–2), 61–75.

QUESTIONS FOR CHAPTER 5

Question 5.1

A steel ball slides against a ceramic disc in a pin-on-disc sliding wear test. The ball has a radius, R , of 12.7 mm and the wear track diameter on the disc is 30 mm. The normal load is 10 N and the disc rotates at 1 revolution per second for a test duration of 2000 revolutions. The coefficient of friction is measured as 0.7.

The wear rate, k , of the ball is $1 \times 10^{-5} \text{ mm}^3 \text{ m}^{-1} \text{ N}^{-1}$.

Assume that the disc exhibits no wear, and that the wear scar on the ball takes the form of a spherical cap of base radius a . The volume of a spherical cap (when the ratio of cap base radius to ball radius is less than 0.1) is given by the following equation:

$$V_{\text{cap}} \approx \frac{\pi a^4}{4R}$$

- Ignoring any elastic deformation, plot a graph of the nominal contact pressure with the number of disc revolutions.
- Calculate the power dissipated in the contact, and hence calculate the power density (i.e., the frictional power per unit contact area) at the beginning and the end of the test.
- In order to make efficient use of the specimen material, another test is performed on the same disc. This time, the track diameter is reduced to 15 mm. Both the rotational speed and the number of revolutions are increased by a factor of two to achieve the same sliding speed and sliding distance. Discuss how these changes will affect the wear of the disc.

Question 5.2

In a fretting test, a steel cylinder of radius 6 mm and length 10 mm is loaded on to a steel plane to form a line contact (i.e., the axis of the cylinder is parallel to the plane), and the cylinder is displaced perpendicularly to its axis with an applied displacement amplitude, Δ^* of 25 μm under an applied load of 250 N. The contact is submerged in a lubricant bath. The Young's modulus and Poisson's ratio of the steel are 207 GPa and 0.28, respectively.

- The rig and contact together have a stiffness of 40 MN m^{-1} . At the beginning of the test, the coefficient of friction is observed to be 0.2. Calculate the contact slip amplitude, δ .
- Assuming an elastic contact (see [Chapter 2](#)), calculate the semi-width of the contact, b .
- In fretting, a portion of the contact will always remain in contact throughout the oscillation, termed the covered width (see [Fig. Q5.1](#)). The covered width, b_c , is defined by the following equation:

$$b_c = \begin{cases} 0, & b < \delta \\ (b - \delta), & b \geq \delta \end{cases}$$

Calculate the covered width at the beginning of the test

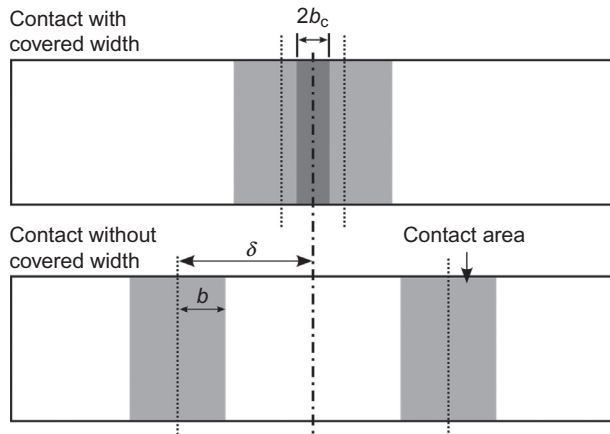


Fig. Q5.1 Diagram illustrating the concept of the covered width in fretting of a line contact

- (d) A second test is performed, identical in all aspects to the first except that a larger displacement amplitude ($\Delta^* = 100 \mu\text{m}$) is employed. Here, a reduced coefficient of friction is observed. Calculate the covered width at the start of the test, and thus explain the reduction in coefficient of friction.

Question 5.3

Explain the following observations:

- During sliding of a cylindrical steel pin against a steel disc in a pin-on-disc test in air, the coefficient of friction is initially around 0.7, but then falls to 0.5 as the test proceeds. At the end of the test, the wear track on the disc is rough and black in colour, with a black powder on the disc surface outside the wear track; chemical analysis of the wear track itself in an electron microscope shows a high oxygen level.
- A test identical to that described in part (a) is conducted, but this time the test is conducted in vacuum. The coefficient of friction remains high through the test, the wear volumes are much greater, and the track on the disc surface following the test is very rough and silvery in colour.
- A test identical to that described in part (a) is conducted, but this time the test is conducted in air at 300°C . At the end of the test, the wear track on the disc is very smooth (shiny) and black in colour, with no debris on the disc surface outside the wear track; chemical analysis of the wear track itself shows a high oxygen level.
- An alumina ball is slid against an alumina disc in air under an applied load. At low loads, the wear rate is low and the wear track is smooth; at slightly higher loads, the wear rate increases by a factor of 100, with generation of large amounts of powdered material on the disc following the test.

- (e) A silicon nitride ball slides against a silicon nitride disc under an applied load in the laboratory and exhibits low friction and low rates of wear. The pair is then deployed in an actuator on a satellite in space, and fails rapidly.
- (f) A steel ball slides against a polymer disc under an applied load. The rates of wear are low, and the tribologist running the test programme realizes that to reduce errors in measurement of the wear volumes, the test needs to be accelerated. Higher sliding velocity and increased normal load are both investigated, but in both cases, the coefficient of friction is very different from that observed in the initial tests: the specific wear rates are orders of magnitude greater, and the wear tracks have very different appearance.

Question 5.4

A twin disc rig (Fig. 5.1d) runs with two 200 mm diameter discs loaded against each other with a force of 200 N. They rotate in opposite senses, one with a peripheral speed of 2 m s^{-1} and other one with a peripheral speed of 2.2 m s^{-1} . For a coefficient of friction of 0.8, calculate the frictional power dissipated. Also, calculate the interval between contacts for a point on the periphery of each of the discs; in general terms, comment on how this is likely to affect the thermal history of points on the periphery of each disc.

Wear by hard particles

6

6.1 INTRODUCTION AND TERMINOLOGY

In this chapter we shall concentrate on *abrasive wear* and *erosion*. In abrasive wear, material is removed or displaced from a surface by hard particles, or sometimes by hard protuberances on a counterface, forced against and moving along the surface. Several qualifying terms can be used in describing abrasion. A distinction is often made between *two-body abrasive wear*, as illustrated in Fig. 6.1(a), and *three-body abrasive wear*, shown in Fig. 6.1(b). Two-body abrasion is caused by hard protuberances on the counterface or hard particles attached to it, while in three-body abrasion the hard particles are free to roll and slide between two, perhaps dissimilar, sliding surfaces.

In fact, the distinction between fixed and loose particles implied by the terms ‘two-body’ and ‘three-body’ is not sufficient to describe the wear process. While particles firmly attached to one surface will necessarily slide over the counter-surface, if they are unattached and form ‘third bodies’ they may either slide or roll in the gap between the surfaces. In some situations, particles may even slide in one region of a contact, but roll in another. It is therefore often more helpful to describe abrasion as either *sliding abrasion* or *rolling abrasion*, depending on the motion of the particles. A drill bit cutting rock will experience two-body, sliding abrasion; grit particles entrained between sliding surfaces, perhaps present as a contaminant in a lubricating oil, may experience either sliding or rolling motion although both situations would correctly be described as three-body abrasion. Wear rates due to rolling abrasion are generally lower than those due to sliding abrasion, although the various mechanisms of material removal in the two cases may differ only in relative importance rather than in nature.

Other terms sometimes used to describe abrasive wear are *high-stress* and *low-stress* abrasion. In high-stress abrasion, the crushing strength of the abrasive particles is exceeded and they are broken up during the wear process, while in low-stress abrasion the particles remain unbroken. The term *gouging abrasion* is often used to describe high-stress abrasion by large lumps of hard abrasive material, for example in rock crushing machinery; gouging results in the removal of material from the worn surface in large fragments. The scale of surface deformation above which the term ‘gouging’ is appropriate is, however, poorly defined.

Sometimes wear is caused by hard particles striking the surface, either carried by a gas stream or entrained in a flowing liquid (Fig. 6.1c). This type of wear is called *erosion*, often qualified as *solid particle erosion* or *solid impingement erosion* to distinguish it from the damage caused by the impact of liquid jets or drops. Particle velocities in erosive wear are commonly between 5 and 500 m s⁻¹, although the term can still be applied outside this range. If the hard particles are carried by a liquid, the wear may be termed *slurry erosion*.

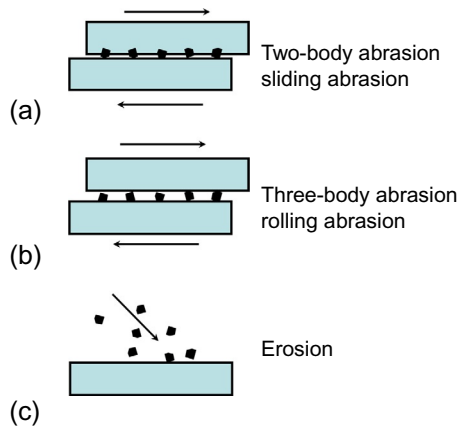


Fig. 6.1 Illustration of the differences between (a) two-body abrasion, (b) three-body abrasion and (c) erosion

The particles responsible for erosion or abrasion may be intrinsic to a given application; for example, tools used to cultivate soil or excavate sand must inevitably be exposed to hard particles, as must pipelines carrying ore/water slurries in mining operations. In other situations, the particles may be contaminants that are difficult to avoid, such as fine airborne grit particles which find their way into lubricating oil. Hard particles may also be generated locally by oxidation or wear from components of the tribological system; Fe_2O_3 wear debris, for example, may be produced in the sliding wear of steels, and can then cause further damage by abrasion. Sliding abrasion may result simply from differences in hardness and roughness between two sliding surfaces, incompletely lubricated: a relatively rough alumina ceramic sliding against a steel counterface would cause wear of this kind.

In some cases significant corrosion may occur at the same time as abrasive or erosive wear, and the resulting rate of material removal can then be higher, sometimes much higher, than the rate of corrosion or wear alone. These are examples of *tribocorrosion*, in which *synergy* between the mechanisms of wear and corrosion is a common feature. Fundamental aspects of corrosion and tribocorrosion are reviewed in [Appendix B](#).

Abrasive and erosive wear is not always deleterious. These processes can also be employed usefully, in grinding and polishing processes and in methods of cutting and shaping materials. Chemical mechanical polishing, discussed in [Section 9.4.2](#), is an important process in the semiconductor industry.

Although in practical cases the sources and nature of the abrasive and erosive particles may differ, the resulting wear processes have much in common which we shall explore in the rest of this chapter.

We have so far assumed that the particles responsible for abrasion or erosion are ‘hard’. In the next section we shall define more closely what is implied by this term, and discuss the properties of common abrasive and erosive particles that are important in wear. We shall then discuss in detail the phenomena of abrasive and erosive wear for all types of material, and examine relevant methods of wear testing. The definition and calculation of wear rate are discussed further in [Appendix C](#).

6.2 PARTICLE PROPERTIES: HARDNESS, SHAPE AND SIZE

6.2.1 Particle hardness

The hardness of the particles involved in abrasion or erosion influences the rate of wear: particles with lower hardness than that of the surface cause much less wear than harder particles. For particles significantly harder than the surface, then the exact value of their hardness matters much less. This behaviour is illustrated in Fig. 6.2, which shows the relative wear rates in two-body (sliding) abrasion of a wide range of metals and ceramics, abraded by various types of grit particle. The wear rate becomes much more sensitive to the ratio of abrasive hardness H_a to the surface hardness H_s when H_a/H_s is less than ~ 1 .

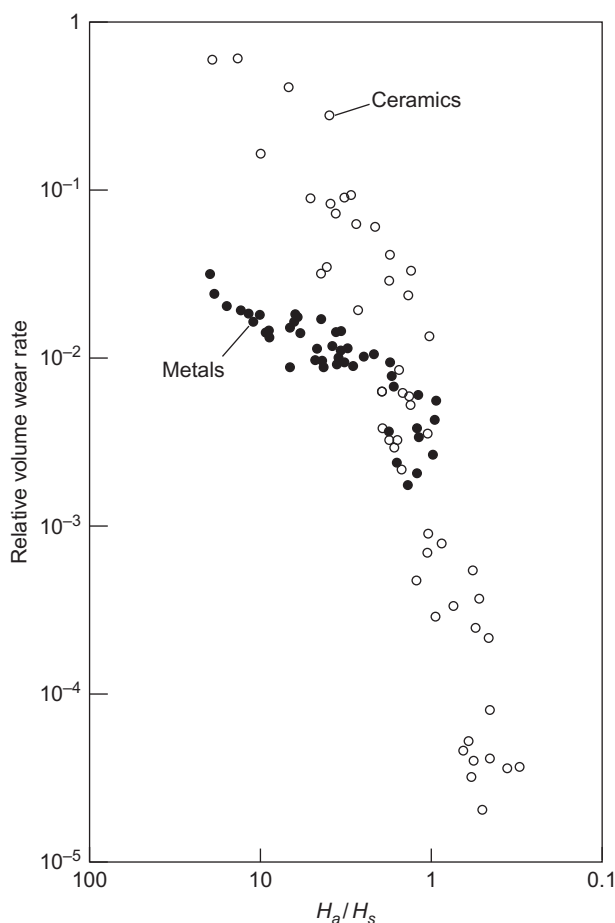


Fig. 6.2 Relative volume wear rate plotted against the ratio of the hardness of the abrasive to that of the surface (H_a/H_s) for a range of metallic and ceramic materials and abrasive particles, for two-body abrasion (data from Moore, M.A., 1978. Mater. Eng. Appl. 1, 97–111)

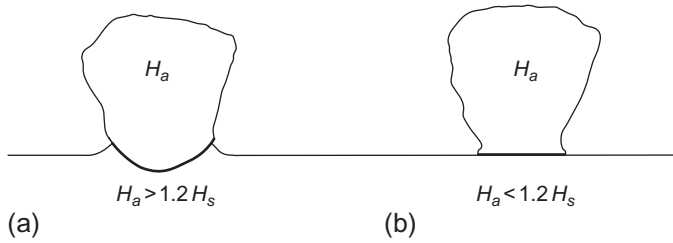


Fig. 6.3 Illustration of contact between a grit particle under normal load and a plane surface. (a) If $H_a > \sim 1.2 H_s$, the particle will indent the surface. (b) If $H_a < \sim 1.2 H_s$, plastic flow will occur in the particle, which will be blunted

The reason for this behaviour can be understood by examining the mechanics of contact between a discrete grit particle and a plane surface. If the surface material flows plastically once its yield point is exceeded, significant plastic flow will occur in the surface when the mean contact pressure reaches about three times its uniaxial yield stress Y , as discussed in [Section 2.5.2](#) and [Appendix A](#). This contact pressure is the indentation hardness of the surface, and depends little on the detailed shape of the particle. Plastic indentation of the surface will occur as the normal load on the particle is increased only if the particle can sustain this contact pressure without deforming ([Fig. 6.3a](#)). If the particle fails by flow or fracture before the pressure on the surface reaches $\sim 3Y$, then insignificant plastic deformation will occur in the surface ([Fig. 6.3b](#)).

For a spherical particle pressed against a flat surface, the maximum contact pressure is about 0.8 times the indentation hardness of the particle material. We would therefore expect a blunt (i.e., near-spherical) abrasive particle of hardness H_a to cause plastic indentation in a surface of hardness H_s if H_s is less than $\sim 0.8 H_a$; that is, if $H_a/H_s > \sim 1.2$. A similar result can be derived from slip-line field theory for other contact geometries, and it is observed experimentally that abrasive grit particles of any shape will cause plastic scratching only if H_a/H_s is greater than ~ 1.2 . Abrasion under conditions where $H_a/H_s < 1.2$ is sometimes termed *soft abrasion*, in contrast to *hard abrasion* when $H_a/H_s > 1.2$.

The observation that a certain minimum ratio of hardness is needed for one material to be able to scratch another provides the physical basis for the scale of hardness devised by the German mineralogist Friedrich Mohs in 1822. Mohs assigned integer hardness numbers to a sequence of 10 minerals, each of which would scratch all those, but only those, below it in the scale. As [Fig. 6.4](#) shows, the ratio of indentation hardness between neighbouring standard minerals (except diamond) is nearly constant. For the Mohs scale it is about 1.6, rather higher than the minimum necessary to cause scratching, but allowing a wide range of hardness to be spanned with only 10 standards.

Knowing that erosion or abrasion will lead to much greater wear rates when the particles have more than about 1.2 times the hardness of the surface, we can now examine the hardnesses of common abrasive particles and structural materials.

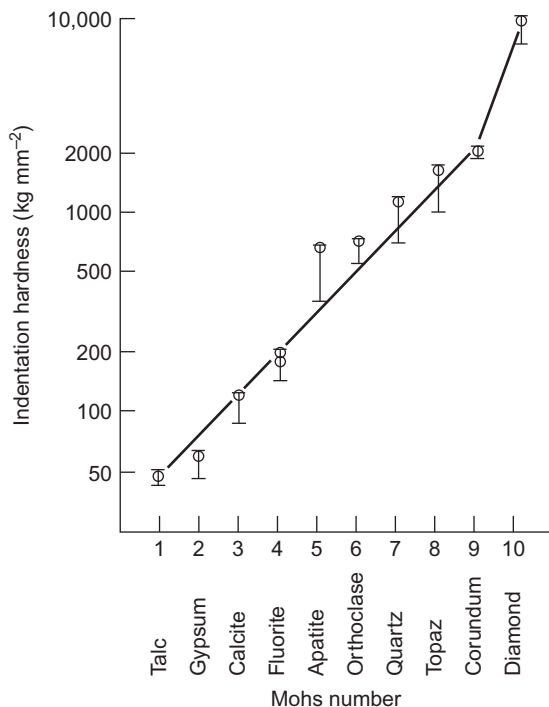


Fig. 6.4 Comparison of indentation hardness (measured by the Vickers or Knoop method) with Mohs hardness number for 10 standard minerals (from Tabor, D., 1954. *Proc. Phys. Soc. B* 67, 249–257, © IOP Publishing. Reproduced with permission. All rights reserved)

Table 6.1 lists values of Vickers hardness (see [Appendix A](#)) for a range of materials. Silica (quartz) is the most commonly occurring natural abrasive contaminant, constituting about 60% of the Earth's crust, and has a hardness of about 800 HV (i.e., $800 \text{ kgf mm}^{-2} \approx 7.8 \text{ GPa}$). The hardest martensitic steels have a hardness about 1.2 times this, and it is therefore clear that steels and non-ferrous metals will be especially vulnerable to abrasive wear and erosion by quartz particles. Materials containing harder phases, such as cemented carbides and 'wear-resistant' alloys, will usually also have softer constituents. Ceramic materials may be sufficiently hard and homogeneous to suffer little plastic deformation from common abrasive or erosive particles, and as we shall see below and in [Chapter 7](#) they can confer valuable wear resistance when used either as bulk materials or as coatings.

6.2.2 Particle shape

Most particles responsible for abrasive or erosive wear are roughly equiaxed, but there can be considerable variation in their angularity depending on their origins. Wear rates depend strongly on the shapes of the particles, with angular particles causing greater wear than rounded particles. The reasons for this are discussed below, in [Sections 6.3](#) and [6.4](#).

Table 6.1 Hardness values for common abrasive particles, bulk structural materials, hard coating materials and alloy phases (It should be noted that the precise value of hardness will depend on the composition and microstructure of the material: for example on grain size, level of cold work and state of heat treatment. In many materials hardness also depends significantly on the indenter load used to measure the hardness. The values given here, from a wide variety of sources, must therefore be taken only as a rough guide to relative values of hardness)

Material	Hardness (HV)
<i>1. Typical abrasive materials</i>	
Diamond	6000–10,000
Cubic boron nitride (CBN)	~4500
Boron carbide, B ₄ C	2700–3700
Silicon carbide	2100–2600
Alumina (corundum)	1800–2000
Quartz (silica)	750–1200
Garnet	600–1000
Magnetite, Fe ₃ O ₄	370–600
Soda-lime glass	~500
Fluorite, CaF ₂	180–190
<i>2. Phases or constituents of steels and cast irons</i>	
Ferrite, α -Fe	70–200
Pearlite (plain C)	250–320
Pearlite (alloyed)	300–460
Austenite (12% Mn)	170–230
Austenite (low alloy)	250–350
Austenite (high Cr)	300–600
Martensite	500–1000
Cementite, Fe ₃ C	840–1100
Chromium/iron carbide, (Fe,Cr) ₇ C ₃	1200–1600
<i>3. Non-ferrous metals (pure unless indicated)</i>	
Aluminium (commercial purity)	25–45
Aluminium alloy (age-hardened)	100–170
Brass (α or β)	80–140
Chromium (cast)	100–170
Chromium (electroplated)	500–1250
Copper (commercial purity)	40–130
Copper-beryllium	150–400
Gold	30–70
Indium	1
Iron	70–200
Lead	4
Molybdenum	160–180
Nickel	70–230
Nickel (electroplated)	200–400
Rhodium (electroplated)	800

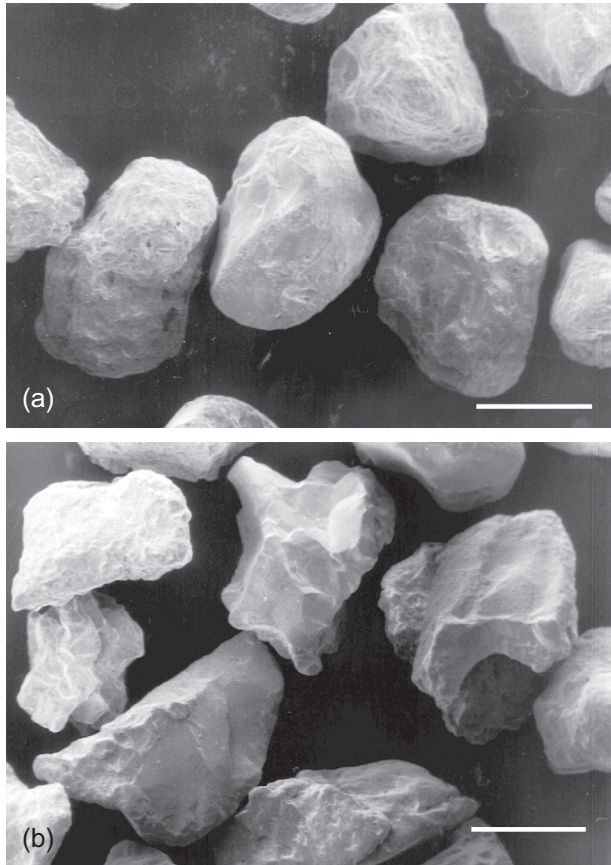
Table 6.1 Continued

Material	Hardness (HV)
Silver	25–80
Tin	5–6
Tungsten	260–1000
Zinc	30–35
<i>4. Ceramic materials (bulk or as coatings) (see also Section 1 above)</i>	
<i>Carbides:</i>	
Chromium carbide, Cr_7C_3	1600
Chromium carbide, Cr_3C_2	1300
Hafnium carbide	2270–2650
Molybdenum carbide, Mo_2C	1500
Niobium carbide	2400–2850
Tantalum carbide	1800–2450
Titanium carbide	2000–3200
Tungsten carbide, WC	2000–2400
Vanadium carbide	2460–3150
Zirconium carbide	2360–2600
<i>Nitrides:</i>	
Chromium nitride	2200
Hafnium nitride	1640
Niobium nitride, NbN	1400
Niobium nitride, Nb_2N	1720
Tantalum nitride	1220
Titanium nitride	1200–2000
Vanadium nitride	1520–1900
Zirconium nitride	1150
<i>Borides:</i>	
Chromium diboride	1800
Hafnium diboride	2250–2900
Molybdenum diboride	2350
Tantalum diboride	2450–2910
Titanium diboride	2200–3500
Tungsten diboride	2400–2660
Vanadium diboride	2070–2800
Zirconium diboride	2250–2600

Angularity is not straightforward to define. [Figure 6.5](#) shows two shapes of quartz particle: rounded and angular. Differences in particle shape of this magnitude can result in differences in wear rate by a factor of 10 or more, yet the angularity of abrasive particles is seldom measured quantitatively. This is largely because of the difficulty of identifying and quantifying the features of a complex three-dimensional shape that are responsible for its abrasivity.

One of the simplest descriptions of shape is based on measurements of the perimeter and area of a two-dimensional projection of the particle, usually generated by

Fig. 6.5 SEM micrographs of silica particles: (a) rounded and (b) angular (Courtesy of A. J. Sparks. Scale bars 100 μm)



optical microscopy. The *roundness factor* or *circularity* F can then be defined as the ratio between the actual area A of the projection, and the area of a circle with the same perimeter P as the projection. In terms of these quantities,

$$F = \frac{4\pi A}{P^2} \quad (6.1)$$

If $F = 1$ the projection is a circle; the more the outline of the particle departs from circular, the smaller the value of F . By averaging the values of F derived from the two-dimensional outlines of many particles, oriented randomly, an indication can be gained of the departure from sphericity of three-dimensional particles.

Some success has been achieved in correlating abrasive wear rates with values of the roundness factor F , although it provides only a crude measure of the deviation of the particle from a perfect sphere. Many other, more sensitive measures of angularity have been proposed, but the simple circularity F has the advantage that it is widely used in other contexts and is readily computed by standard image-analysis software.

6.2.3 Particle size

The sizes of abrasive particles cover a wide range. Those responsible for most abrasive and erosive wear are between 5 and 500 μm in size, although polishing may employ submicrometre particles, while gouging wear may involve hard objects some tens or even hundreds of millimetres across.

A consistent pattern of behaviour is found in laboratory studies of both abrasion and erosion of metals, as shown in Fig. 6.6. The units of wear plotted there are explained further in Sections 6.3 and 6.4; for the present it is sufficient to note that if particles of different sizes were equally effective at removing material, the wear rates expressed in these units would be constant. In fact, as Fig. 6.6 shows, wear rates for particles smaller than about 100 μm drop markedly with decreasing particle size. Similar behaviour is seen for quite different particle materials and different metals.

Several explanations have been proposed for this effect of particle size on the abrasion and erosion of metals, and it is likely that no single mechanism is responsible for the whole effect. Some contribution may be due to a true size effect in the strength of the metal itself, also seen in indentation and scratching experiments, associated with the difficulty of nucleating or moving dislocations in a small volume. This increase in local flow stress as the scale of the deformation is diminished leads to a reduction in wear rates by plastic processes (further discussed in Sections 6.3.1 and 6.4.1), and hence to a particle size effect. It is also likely that features of the shapes of real abrasive particles play a role. The extreme edges and corners of apparently angular particles are to some extent rounded rather than being perfectly sharp, so that

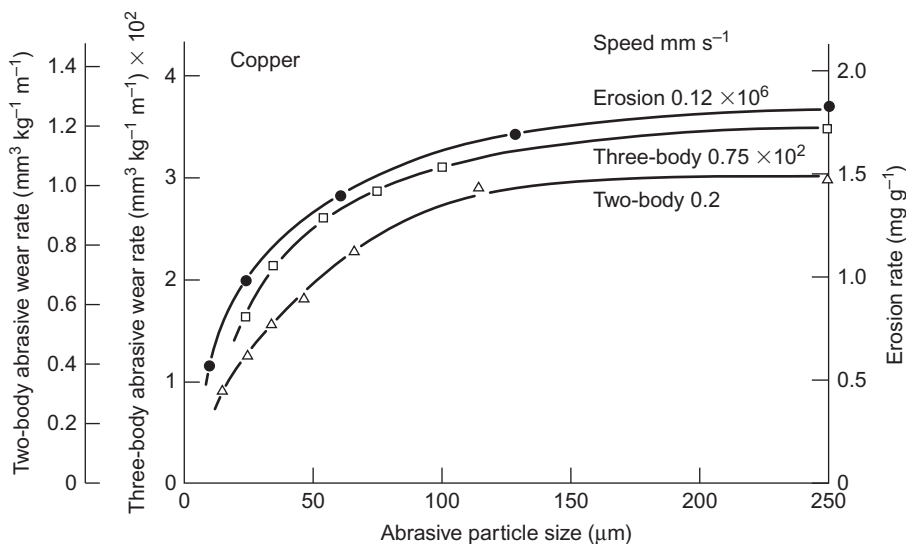


Fig. 6.6 Wear rates of copper under conditions of two-body and three-body abrasion and erosion, due to silicon carbide particles of different sizes (from Misra, A., Finnie, I., 1981. *Wear* 65, 359–373)

the pattern of deformation caused by the particle will depend on the depth to which it indents the surface. For small angular particles the rounding of the edges and corners will have a proportionately greater effect than for large particles, and so their effectiveness in removing material by wear will be more similar to that of macroscopically rounded particles than to that of angular particles with geometrically sharp corners.

Behaviour like that illustrated in Fig. 6.6 is seen in metals and other materials where wear involves plastic flow. As we shall see below, wear of brittle materials may involve fracture, and they then exhibit an even stronger dependence of wear rate on particle size. Whatever the dominant mechanism, the fact that small particles cause proportionately less wear than larger ones is fortunate. It means that because methods of removing contaminant particles from a system, such as filtration or centrifugal separation, eliminate large particles more readily than smaller ones, they can reduce wear rates by abrasion and erosion very effectively (see Section 8.4).

6.3 ABRASIVE WEAR

Mechanisms of abrasive wear can involve both plastic flow and brittle fracture. Under some circumstances plastic flow may occur alone, but both often occur together, even in materials conventionally thought of as ideally brittle. Models for abrasive wear by each type of mechanism in isolation have been developed which usually ignore the possibility of the other. In order to understand abrasive wear in simple terms, we shall also separate the two groups of mechanisms, and examine in the following sections models for abrasive wear by plastic deformation and by brittle fracture. We shall then discuss the behaviour of engineering materials in the light of these somewhat idealized models.

6.3.1 *Abrasive wear by plastic deformation*

A simple model for abrasive wear involves the removal of material by plastic deformation. It is closely related to the model for the ploughing (deformation) component of friction examined in Section 3.4.2.

Fig. 6.7 shows an abrasive particle, idealized as a cone of semi-angle α , being dragged across the surface of a ductile material which flows under an indentation pressure P . It forms a groove in the material, and wear is assumed to occur by the removal of some proportion of the material that is displaced by the particle from the groove. As we saw in Section 3.4.2, the normal load w carried by the particle is supported by plastic flow beneath the particle, which causes a pressure P to act over the area of contact between the particle and the surface. Since the cone is moving and therefore in contact only over its front surface,

$$w = P \frac{\pi a^2}{2} = \frac{1}{2} P \pi x^2 \tan^2 \alpha \quad (6.2)$$

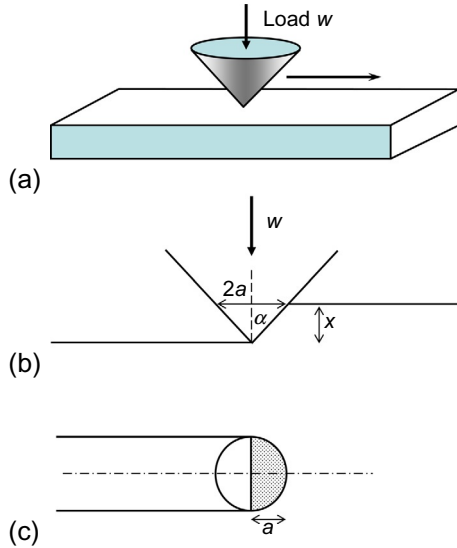


Fig. 6.7 Geometry of contact between an idealized conical abrasive particle and a surface: (a) perspective view; (b) section in elevation; (c) section in plan view

The volume of material displaced from the groove by the cone in sliding a distance l along the surface is $l a x$, or $l x^2 \tan \alpha$. The quantities a and x are defined in Fig. 6.7. So if a fraction η of the material displaced from the groove is actually removed as wear debris, then the volume of wear debris produced by this one particle per unit sliding distance, q , will be given by

$$q = \eta x^2 \tan^2 \alpha \quad (6.3)$$

We can substitute for x^2 from equation 6.2, to find

$$q = \frac{2\eta w}{\pi P \tan \alpha} \quad (6.4)$$

Summing over many abrasive particles, and assuming that $P \approx H$, the indentation hardness of the material, we can then show that the total volume removed per unit sliding distance Q is given by

$$Q = \frac{KW}{H} \quad (6.5)$$

where W is the total applied normal load and the constant K depends on the fraction η of displaced material actually removed, and on the geometry of the abrasive particles (i.e., on α).

Equation 6.5 is exactly the same as the Archard equation for sliding wear (equation 5.7), although derived from completely different initial assumptions. K , the dimensionless wear coefficient, can be used as a measure of the severity of wear. Typical values of K in the two-body abrasive wear (sliding abrasion) of metals lie between $\sim 5 \times 10^{-3}$ and $\sim 50 \times 10^{-3}$. For three-body abrasion K is lower, typically between 0.5×10^{-3} and 5×10^{-3} . As in the case of sliding wear, it is sometimes more useful to consider the specific wear rate k ($=K/H$), with the usual units $\text{mm}^3 \text{N}^{-1} \text{m}^{-1}$.

Another quantity sometimes used to express the severity of abrasive wear, which allows particularly revealing comparisons between different abrasion conditions and with erosive wear, is the specific energy for material removal, U . This is defined as the frictional work expended per unit volume of material removed. In terms of the quantities discussed above,

$$U = \frac{\mu W}{Q} \quad (6.6)$$

where μ is the coefficient of friction (ratio of tangential to normal force) between the sliding bodies. U can also be expressed in terms of μ and k :

$$U = \frac{\mu}{k} \quad (6.7)$$

For two-body (sliding) abrasion of metals in air, μ lies typically between 0.4 and 1, while for three-body (rolling) conditions it is often lower: 0.2–0.5. A rough estimate of U can therefore readily be made from equation 6.7 and the value of the specific wear rate k .

The simple model for wear by plastic deformation which leads to equation 6.5 predicts that the volume of material removed by sliding abrasion should be directly proportional to the sliding distance, and also to the normal load. This behaviour is usually observed in practice, as illustrated in Fig. 6.8 for three ductile metals subjected to sliding abrasion by silicon carbide particles.

Equation 6.5 also suggests that the wear rate should vary inversely with the hardness of the material, H . Many pure metals do behave in this way, although alloys often exhibit more complex behaviour. Figure 6.9 shows the results of sliding abrasion tests on a range of annealed pure metals (open points) and heat-treated and work-hardened steels (solid points). The relative wear resistance, which is proportional to the reciprocal of the volume wear rate Q , is plotted against the indentation hardness of the bulk metals. The experimental points for the pure metals lie close to the straight line through the origin expected from equation 6.5. However, the results for the steels (solid points) do not lie on this line, but on other lines with different slopes. In order to understand the reasons for this, we must explore the factors which influence the value of the quantity K in equation 6.5.

In plotting wear resistance against the bulk hardness of the metals in Fig. 6.9, we ignored the fact that the material at the worn surface will have been strain-hardened by plastic flow, and that its hardness will therefore generally be greater than that of the

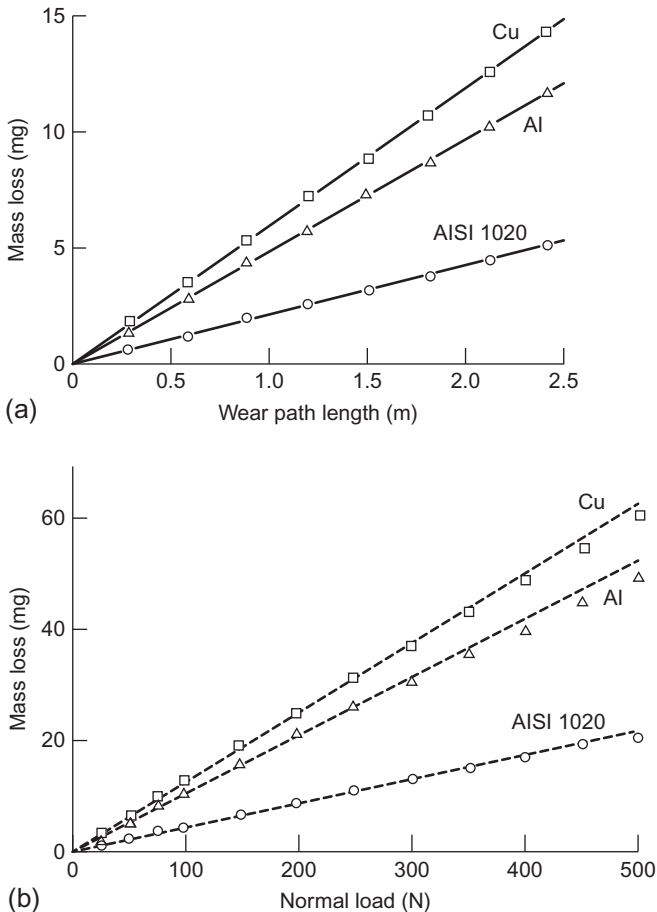


Fig. 6.8 Wear, measured as mass loss, for samples of copper, aluminium and 0.2% carbon steel (AISI 1020) subjected to two-body abrasion by 115 μm silicon carbide particles: (a) variation with sliding distance; (b) variation with normal load (from Misra, A., Finnie, I., 1981. *Wear* 68, 41–56)

bulk. Abrasion introduces very high shear strains into the surface material; in copper and brass, for example, true strains of up to 8 have been recorded. The shear strain decreases with depth into the bulk as shown in Fig. 6.10; the depth of deformation and the strain at a given depth are both proportional to the depth of indentation of the abrasive particles. The energy of frictional work expended in abrasion is largely accounted for by the plastic work done in this deformed region. As a result of strain hardening, the flow stress of the deformed surface material may be two or even three times that of the bulk, although its precise value depends on the response of the particular metal to the very high shear strains, and possible high temperatures and strain rates, to which it is subjected during abrasion. Better correlation is therefore often

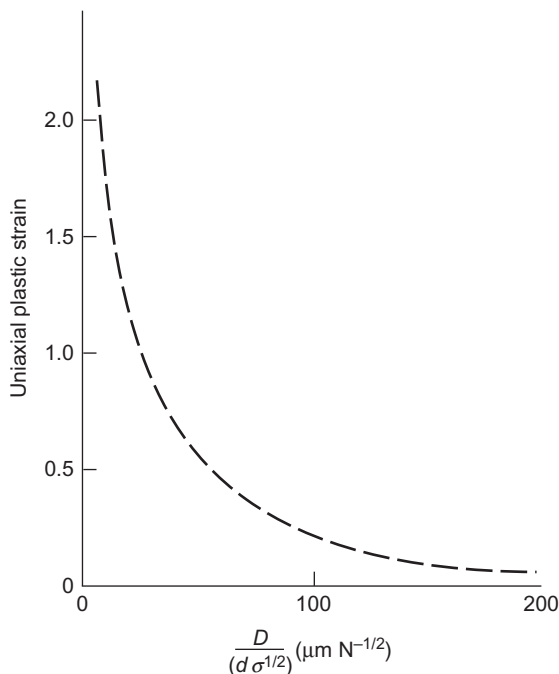
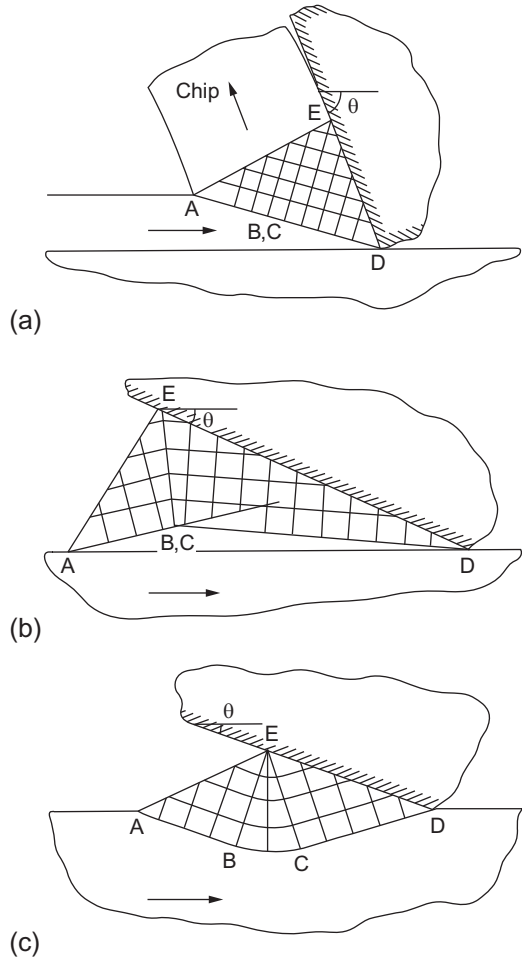


Fig. 6.10 Distribution of plastic strain with depth beneath abraded metal surfaces (here, a copper/silver laminate). In this graph the depth D is scaled by dividing by the product of the particle size d and the square root of the contact pressure σ (the normal load per unit area) (graph based on data from Moore, M.A., Douthwaite, R.M., 1976. *Met. Trans.* 7A, 1833–1839, with permission of Springer)

three distinct modes of deformation due to a rigid two-dimensional wedge (an idealized abrasive particle) sliding over a rigid-plastic material from right to left. **Figure 6.11(a)** shows the *cutting* mode, in which material is deflected through a shear zone and flows up the front face of the particle to form a chip. This mode of deformation is exactly the same as that caused by a single-point tool in an orthogonal machining process (see [Section 9.4.1](#)). In this mode, all the material displaced by the particle is removed in the chip. **Figure 6.11(c)** depicts the other extreme case, the *ploughing* mode, in which a ridge of deformed material is pushed along ahead of the particle, in much the same way that a wrinkle can be pushed along a piece of cloth lying on a table. In the ploughing mode, no material is removed from the surface; material from the raised ridge flows beneath the particle. A simple distinction between these two modes is that in the case of cutting, material flows up the front face of the particle, whereas in ploughing it flows down.

A third mode of deformation, representing intermediate behaviour, is illustrated in **Fig. 6.11(b)**. Limited slip, or even complete adhesion, occurs between the front face of the particle and a raised ‘prow’ of material. It is not possible to construct a steady-state slip-line field for this mode, and **Fig. 6.11(b)** represents one stage in the development of the prow. The deformation consists of the growth and eventual detachment (by shear along the line AD) of the raised prow, a sequence which is repeated continuously. This mode has been termed *wedge formation* and leads, like the cutting mode, to removal of material from the surface.

Fig. 6.11 Slip line fields for the deformation of a perfectly plastic material caused by the sliding of a rigid two-dimensional wedge from right to left. Three distinct modes can be identified, as discussed in the text: (a) cutting; (b) wedge formation; (c) ploughing. The angle θ is termed the 'attack angle' (from Challen, J.M., Oxley, P.L.B., 1979. *Wear* 53, 229–243)



The modes of deformation shown in Fig. 6.11(b) and (c) have already been discussed in Section 3.4.2 in the context of sliding friction; but in that case, since asperities on surfaces have very shallow slopes, the possibility of the cutting mode of deformation which will occur only at high angles of attack, was ignored.

Analysis of the forces involved in each of the three modes of deformation shown in Fig. 6.11 allows the operative mode (that which requires the lowest tangential force) to be identified for any sliding conditions. The controlling factors are θ , the *attack angle* of the particle, and the shear strength of the interface between the particle and the surface. This can be expressed as the ratio f between the shear stress at the interface and the shear yield stress of the plastically deforming material. Perfect lubrication would imply $f=0$, while complete adhesion would lead to $f=1$. For $f<0.5$, only two modes of deformation are possible: cutting and ploughing. Low values of θ favour ploughing, while values greater than a critical value θ_c lead to cutting. For $f>0.5$, all three modes

can operate, with transitions occurring from ploughing to wedge formation and from wedge formation to cutting as the attack angle is increased.

The idealized modes of deformation shown in Fig. 6.11 are two-dimensional. In practice, abrasive grit particles will cause three-dimensional flow patterns that are more difficult to analyze, but which are nevertheless analogous to the three discussed above. In the ploughing mode, material is displaced both below the particle and to the sides of the groove, and little or none is directly removed. Cutting and wedge formation both lead to wear. The transition from ploughing to the other, debris-forming, modes occurs at a critical attack angle θ_c . Figure 6.12 illustrates these regimes of deformation for the case of a hard spherical indenter sliding over a rigid-plastic material. The depth of penetration of the sphere determines the effective attack angle, which is also plotted in the figure. Although in the rigid-ideal plastic model θ_c depends only on f , for a real material it depends also on the work-hardening rate and on its elastic properties: specifically, on the ratio E/H between the Young's modulus and the surface hardness. θ_c lies typically between 30° and 90° for most metals; a high value of E/H leads to a high value of θ_c .

The distribution of attack angles of the abrasive particles in two-body wear depends on the geometry of the abrasive counterface; an example, for silicon carbide-coated abrasive paper, is shown in Fig. 6.13. The figure shows how the value of θ_c determines the proportion of the abrasive particles which deform the surface by cutting or wedge formation rather than ploughing, and hence influences the value of K .

We can readily explain the effect of lubrication on abrasive wear by plastic deformation (discussed further in Section 6.3.4), since by lowering the friction between the particle and the surface, lubrication leads to more particles cutting, and hence to a higher rate of wear. The effect of particle shape discussed in Section 6.2.2 can also be understood, since the angularity of the particles controls the distribution of attack angle θ . Angular particles will tend to present higher attack angles, leading to a greater proportion of particles which cut, and thus to a higher wear rate, than more rounded particles. The lower wear rate associated with three-body (rolling) abrasion can also be explained, since if the abrasive particles are free to roll between the surfaces then

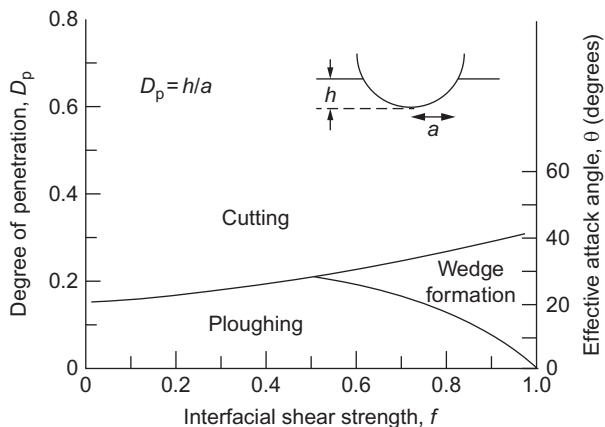


Fig. 6.12 Modes of deformation observed in the sliding of a hard spherical indenter on α -brass, a plain carbon steel (0.45% C) and an austenitic stainless steel (AISI 304). The depth of penetration of the sphere D_p determines the effective attack angle θ which is also plotted (from Hokenigawa, K., Kato, K., 1988. Tribol. Int. 21, 51–57)

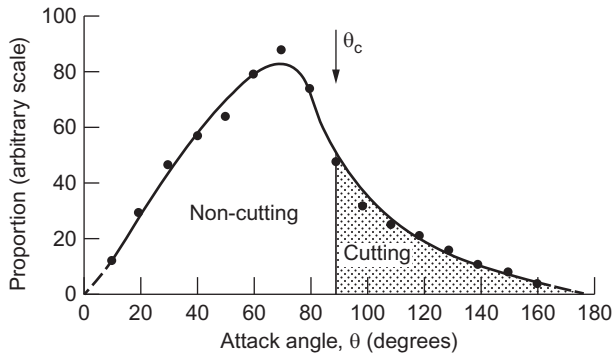


Fig. 6.13 Frequency distribution of attack angle of the contacting abrasive particles in unused 220 mesh size (70 μm) silicon carbide abrasive paper. From curves like this, the proportion of particles for which $\theta > \theta_c$ can be calculated. In this case, $\theta_c = 90^\circ$ has been assumed (from Mulhearn, T.O., Samuels, L.E., 1962. *Wear* 5, 478–498)

indentation and ploughing will occur more often, and cutting less often, than if the particles are fixed to the counterface as in two-body (sliding) abrasion.

When cutting or wedge formation occurs, by no means all the material displaced by the particle is removed as wear debris. Some material will be displaced without being removed, either into side ridges or accommodated by elastic deformation in the bulk away from the groove. The proportion actually removed during cutting or wedge formation varies between different metals. It tends to increase with the hardness of the metal, and so falls with an increase in the ratio E/H defined above. It also decreases with increasing ductility of the metal (defined as the tensile fracture strain or alternatively as the strain at which cracking occurs in rolling experiments).

We have now discussed the influence on K of three factors: the extent of the strain hardening of surface material during abrasion (which causes the surface hardness to become higher than that of the bulk), the fraction of the particles which deform the material by cutting or wedge formation rather than ploughing, and the proportion of the groove volume which is actually removed in cutting or wedge formation. All depend to some extent on the properties of the particles as well as of the metal. A further factor that we incorporated into K in deriving equation 6.5 depends solely on the geometry of the particle, and is equal to $2/(\pi \tan \alpha)$ for a cone. We have seen that experimentally determined values of K for two-body abrasion of metals lie typically between 0.05 and 0.005. Values of this order can be readily understood in terms of the factors we have just enumerated. For angular abrasive particles, α may be around 60° , while for more rounded particles it will be greater. The factor $2/(\pi \tan \alpha)$ will therefore be less than 0.4. Experimental studies suggest that the proportion of particles which remove material by cutting or wedge formation may be between 0.1 and 0.6, and that for those the proportion of the groove volume actually removed may be from 0.2 to 0.9. The combination of these three factors yields values of K from ~ 0.2 to less than $\sim 8 \times 10^{-3}$, in good agreement with the range actually observed in the two-body abrasion of metals, when work-hardening is taken into account.

6.3.2 Abrasive wear by brittle fracture

The second idealized picture of abrasive wear that we shall examine is one in which material removal takes place by brittle fracture, with a negligible contribution from the mechanisms associated with plastic flow and discussed in the previous section.

As we saw in Section 5.10.2, if a brittle material is indented at a sufficiently high load by a blunt (e.g., spherical) body, and the contact stresses remain elastic, then a Hertzian cone crack will form as illustrated in Fig. 5.25. If the indenter slides over the surface a series of incomplete conical cracks forms, intersecting the surface in a row of circular arcs (Fig. 5.26). Neither a single complete conical crack, nor a row of partial cone cracks, leads readily to material removal, although repeated passages of the indenter will lead to the cracks growing and linking up, and thus eventually to wear. If the slider is a hard angular abrasive particle, however, local plastic deformation can occur at the point of contact and cracks of a different geometry form which can immediately result in wear.

Figure 6.14 shows how cracks form in a brittle solid subjected to a point load. Cracks like these grow beneath a sharp rigid indenter which generates an elastic–plastic stress field, such as a cone, pyramid or hard irregular grit particle. They are entirely different from the Hertzian cracks that form in the elastic stress field under a blunt indenter.

At the point of initial contact, very high stresses occur. Indeed, if the tip of the indenter were perfectly sharp (i.e., with zero radius of curvature) there would be a stress singularity at this point. These intense stresses (shear and hydrostatic compression) are relieved by local plastic flow or densification around the tip of the indenter; the zone of deformed material is indicated by the letter D in Fig. 6.14. When the load on the indenter increases to a critical value, tensile stresses across the vertical mid-plane initiate a *median vent crack*, indicated by M in Fig. 6.14. Further increase in load is accompanied by progressive extension of the median crack. On reducing the load, the median crack closes (Fig. 6.14d). Further unloading (Fig. 6.14e and f) is accompanied by the formation and growth of *lateral vent cracks* (labeled L). The formation of these lateral cracks is driven by residual elastic stresses, caused by the relaxation of the deformed material around the region of contact. As unloading is completed (Fig. 6.14f), the lateral cracks curve upwards, terminating at the free surface.

The median cracks, like the Hertzian cone cracks due to a blunt indenter, propagate down into the bulk of the solid with increasing load on the indenter, and do not grow further on unloading. They are not associated in the first instance with the removal of material. Lateral cracks, in contrast, can lead directly to wear.

Lateral cracks form only when the normal load on the indenter has exceeded a critical value, w^* . The value of w^* depends on the fracture toughness of the material, K_c , and on its hardness, H . According to one theory,

$$w^* \propto \left(\frac{K_c}{H} \right)^3 K_c \quad (6.8)$$

K_c is usually taken to be the mode I plane strain fracture toughness K_{Ic} , as measured in conventional notched tensile or bending tests, although the value of toughness

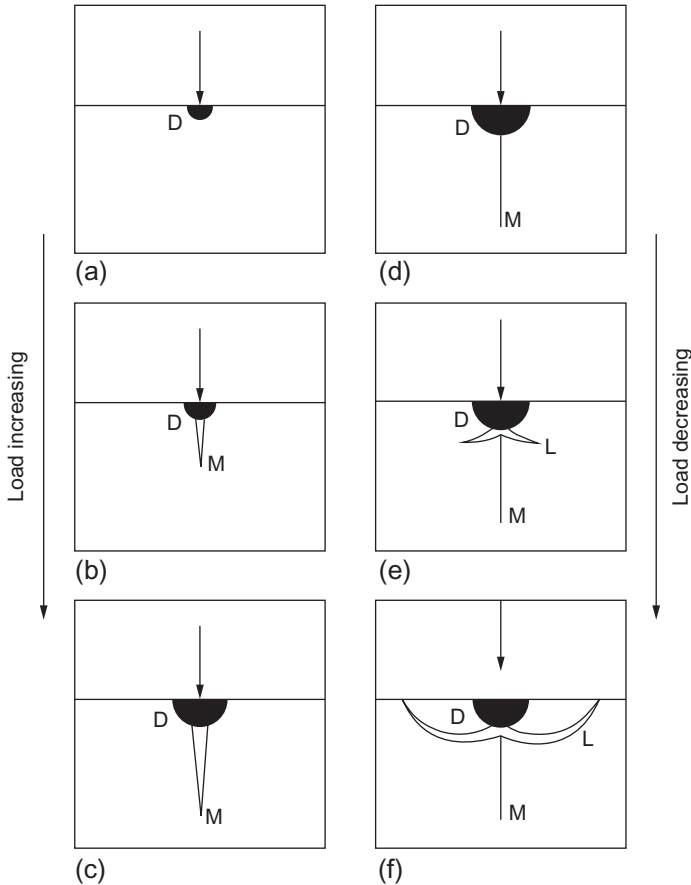


Fig. 6.14 Diagram showing crack formation in a brittle material due to point indentation. The normal load increases from (a) to (c), and is then progressively reduced from (d) to (f) (from Lawn, B.R., Swain, M.V., 1975. *J. Mater. Sci.* 10, 113–122, with permission of Springer)

applicable to indentation fracture, which occurs under a different stress distribution and at a different size scale, is likely to be significantly different. The ratio H/K_c for a material provides a useful measure of its brittleness: a low value of this *brittleness index* corresponds to a high value of w^* , and therefore indicates a material which is reluctant to fracture on indentation. The quantity $(K_c/H)^2$ has dimensions of length, and can be related to a critical scaling dimension above which fracture will occur during contact by a sharp indenter.

One model for the abrasive wear of brittle materials is based on the removal of material by lateral cracking. As a sharp particle slides over the surface forming a plastic groove, lateral cracks grow upwards to the free surface from the base of the sub-surface deformed region (Fig. 6.15), driven by the residual stresses associated with the deformed material. It is assumed that material is removed as chips from the region

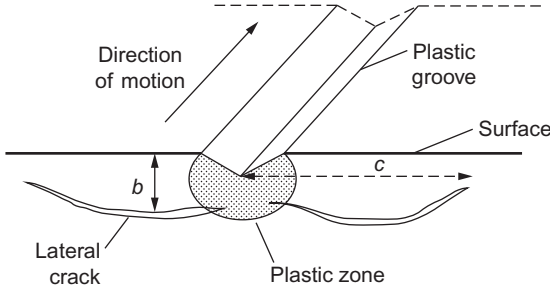


Fig. 6.15 Schematic illustration of material removal in a brittle material by the extension of lateral cracks from beneath a plastic groove (redrawn from Evans, A.G., 1979. In: Hockey, B.J., Rice, R.W. (Eds.), *The Science of Ceramic Machining and Surface Finishing II*. National Bureau of Standards Sp. Pub. 562, US Govt. Printing Office, pp. 1–14)

bounded by the lateral cracks and the free surface, and the volume wear rate is then estimated from the volume of this region. The sideways spread of the cracks, c , is given by

$$c = \alpha_1 \frac{(E/H)^{3/5} w^{5/8}}{K_c^{1/2} H^{1/8}} \quad (6.9)$$

where w is the normal load on the particle, E and H are the Young's modulus and indentation hardness of the material respectively, and K_c is its fracture toughness. α_1 is a constant which depends only on the particle shape.

The depth of the lateral cracks b is assumed to be proportional to the radius of the plastic zone, and is given by

$$b = \alpha_2 \left(\frac{E}{H} \right)^{2/5} \left(\frac{w}{H} \right)^{1/2} \quad (6.10)$$

where α_2 is another geometrical constant. An upper limit for the volume removed by lateral fracture, per particle, per unit sliding distance is $2bc$. If N particles are in contact with the surface, each carrying the same load w , then the volume wear rate per unit sliding distance due to all the particles, Q , will be

$$Q = \alpha_3 N \frac{(E/H) w^{9/8}}{K_c^{1/2} H^{5/8}} \quad (6.11)$$

where α_3 is a material-independent constant.

Other analyses of the same model lead to slightly different results, since they use alternative methods of calculating c and b . One such approach leads to:

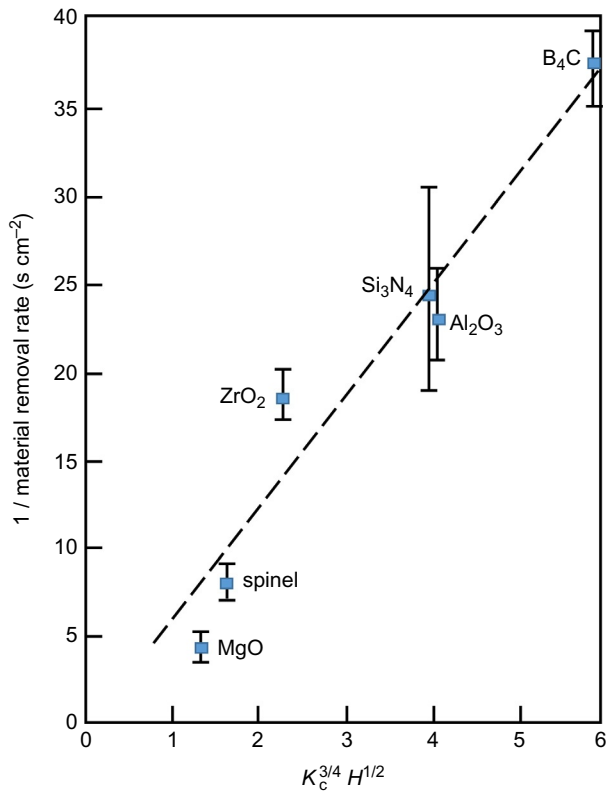
$$Q = \alpha_4 N \frac{w^{5/4}}{K_c^{3/4} H^{1/2}} \quad (6.12)$$

where α_4 is a constant. In this case, Q does not depend on E .

The practical consequences of equations 6.11 and 6.12 are similar, since the ratio E/H does not vary greatly between different hard brittle solids. Both equations predict wear rates that are inversely proportional to both hardness and toughness, raised to powers of about 1/2. Experimental measurements of the abrasive wear rates of some ceramic materials show fair agreement with the predictions of the lateral fracture models. Figure 6.16 shows the correlation between the material removal rate by abrasive machining (i.e., two-body abrasive wear involving brittle fracture) and $K_c^{-3/4}H^{-1/2}$ for several ceramic materials; a very similar correlation is seen with $K_c^{-1/2}H^{-5/8}$.

An important feature of equations 6.11 and 6.12 is that in each case the exponent of w , the normal load on each particle, is greater than 1. This means that the wear rate by lateral fracture is not directly proportional to the normal load, marking an important difference between this mechanism and those involving plastic deformation discussed in the previous section. This dependence on load leads to an apparent influence of particle size on wear rate that is quite distinct from the effects discussed in Section 6.2.3. If we assume there to be N abrasive particles spread over an area A of the surface, each carrying the same normal load w , then the total

Fig. 6.16 Correlation between the reciprocal of material removal rate in abrasive machining (i.e., two-body abrasion) and the quantity $H^{1/2}K_c^{3/4}$, for several ceramic materials (from Evans, A.G., Wilshaw, T.R., 1976. *Acta Met.* 24, 939–956)



applied load W is Nw . If the particles are each of linear dimension d then N will be proportional to Ad^{-2} . (This supposes that the pattern in which the particles are distributed over the surface scales linearly with d —a reasonable supposition for many cases of abrasive wear.) Taking equation 6.12 to represent the wear rate, it follows that the volume removed per unit sliding distance by all the particles will be given by

$$Q = \alpha_5 \frac{W^{5/4} d^{1/2}}{A^{1/4} K_c^{3/4} H^{1/2}} \quad (6.13)$$

For a fixed load W and apparent contact area A , we would therefore expect the wear rate to increase with the square root of the particle size.

Several other models for abrasive wear involving fracture have been proposed, based on various assumptions about the geometry and origin of the cracks. Some are applicable only to particular classes of material. For example, Fig. 6.17 shows how wear can occur in grey cast iron, which contains flakes of brittle graphite. The frictional traction imposed on the surface as an abrasive particle slides across it causes cracks to open on the planes of the graphite lamellae lying normal (or nearly

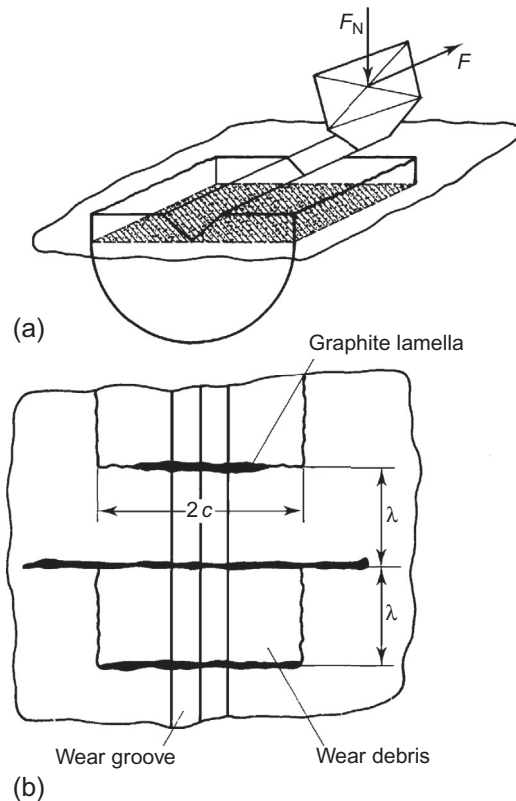


Fig. 6.17 (a) Perspective and (b) plan views showing crack formation modified by the presence of weak graphite lamellae during the abrasive wear of grey cast iron (from Zum Gahr, K.-H., 1987. *Microstructure and Wear of Materials*. Elsevier)

normal) to the sliding direction. Material is then detached by the propagation of shear cracks parallel to the surface, at a depth equal to the depth of penetration of the abrasive particle. The size of the debris fragment is therefore defined by this depth, by the width of the cracks associated with the graphite flakes ($2c$ in Fig. 6.17b), and by the mean separation of the flakes λ . This model leads to a wear rate due to N particles given by

$$Q = \alpha_6 N \frac{w^{3/2} H^{1/2}}{K_c^2} \quad (6.14)$$

where α_6 is a factor which incorporates the coefficient of friction for the system, the geometry of the particles and the spatial distribution of the graphite lamellae. Exactly the same functional dependence of Q on w , H and K_c is found for a similar analysis of wear by fracture in a material containing microcracks or brittle grain boundaries. These analyses are of course only valid when the load on each particle is sufficient to cause fracture in the cast iron; this condition is determined by a relationship similar to equation 6.8.

The models for wear by brittle fracture described here have several important features in common. They predict wear rates considerably higher than those due to plastic mechanisms; in Figs. 6.15 and 6.17, for example, the volume encompassed by the cracks is much greater than that displaced from the plastic groove. They suggest that the wear rate should increase more rapidly than linearly with the applied load. In this respect the models depart significantly from the linear dependence assumed in the plastic models and in the Archard equation for sliding wear (see Section 5.3). As we have seen above, this leads to an increase in wear rate with abrasive particle size which is quite separate in origin from the size effect associated with plastic deformation. The models all predict an inverse correlation between wear rate and some power of the material's fracture toughness: the dependence on toughness is often stronger than the dependence on hardness. Finally, they all suggest that wear by fracture will occur only when a critical load on each abrasive particle is exceeded.

This last point is important, for it means that if the abrasive particles are sufficiently small and numerous in the contact region, the load carried by each may be below the threshold w^* needed to cause cracking (equation 6.8). This is indeed observed. In soda-lime glass, for example, a load of some 0.1–1 N on a single sharp particle is needed to cause indentation fracture. Below this threshold a hard abrasive particle will cause only plastic deformation, and wear will occur by the plastic processes discussed in Section 6.3.1. Plastic scratches up to about 1 μm wide can be formed in soda-lime glass, conventionally thought of as extremely brittle, without fracture. We therefore see a transition in wear mechanism in brittle materials with increasing load or particle size, which originates from this effect. At low loads or with small particles, fracture may be suppressed and abrasive wear may occur by plastic processes. For higher loads, or larger particles, brittle fracture occurs, leading to a sharply increased wear rate.

6.3.3 Abrasive wear of engineering materials

We have seen how abrasive wear may occur by mechanisms dominated either by plastic deformation (Section 6.3.1) or by brittle fracture (Section 6.3.2). In the first case the hardness of the counterface is an important factor in determining its wear resistance, whereas in the second the fracture toughness is more important, although hardness still plays a role. Even in materials with low fracture toughness, fracture will not always occur; it is favoured by severe contact conditions, for example by large, hard, angular abrasive particles under high normal load. Under conditions where fracture is not important, a fair correlation is found between the resistance of a material to abrasive wear and its indentation hardness. Figure 6.9 showed this behaviour for pure metals and some steels, and Fig. 6.18 illustrates the performance of a much wider range of materials. It must be stressed that the trends indicated in Fig. 6.18 are found only when the abrasive particles are ‘hard’ compared with the material being abraded; as we saw in Section 6.2.1, in practice this means that their indentation hardness must be at least about 1.2 times that of the surface.

It is evident from Fig. 6.18 that while there are correlations between wear resistance and hardness within each generic group of materials, the same correlation is not found in different groups. For example, a grey cast iron, a white cast iron and a ceramic may have the same bulk hardness but exhibit very substantial differences in wear resistance. We shall briefly examine the reasons for these differences.

We saw in Section 6.3.1 that if the wear mechanism is dominated by plastic deformation, an important factor in determining the wear rate (via the factor K in equation 6.5) is the fraction of the plastic groove volume removed as wear debris. This

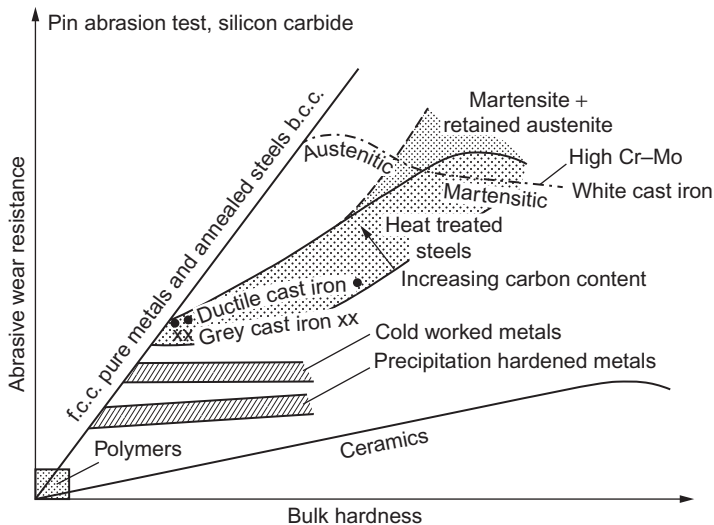


Fig. 6.18 Two-body abrasion resistance (1/volume wear rate) of various materials plotted against bulk hardness (from Zum Gahr, K.-H., 1987. *Microstructure and Wear of Materials*. Elsevier)

fraction varies between different classes of material. Low values of the ratio E/H (the ratio between Young's modulus and hardness; see Section 6.3.1) favour deformation by cutting rather than ploughing, which leads to correspondingly large fractions of the groove volume removed as wear debris. We can therefore understand why ceramic materials, for which E/H is lower than for metals, show higher rates of wear at the same hardness. Polymers also have low values of E/H , and thus also have lower wear resistance than metals of the same hardness. As we noted in Section 5.11.2, there is generally a poorer correlation between abrasive wear rate and hardness for polymers, as shown in Fig. 6.19, than with the product of ultimate tensile strength and elongation (as illustrated in Fig. 5.32).

Pure metals of the same crystallographic structure generally show close proportionality between E and H , and so have effectively constant values of E/H . Thus under given abrasion conditions the value of K varies little for these pure metals, and the wear resistance ($1/Q$) is directly proportional to H as predicted by equation 6.5.

In heat-treated steels, where the elastic modulus E varies insignificantly with composition and microstructure, an increase in hardness leads to a decrease in E/H , and hence to a higher value of K . The benefits gained from hardening steels are therefore not so high as would be expected if K remained constant. A more detailed

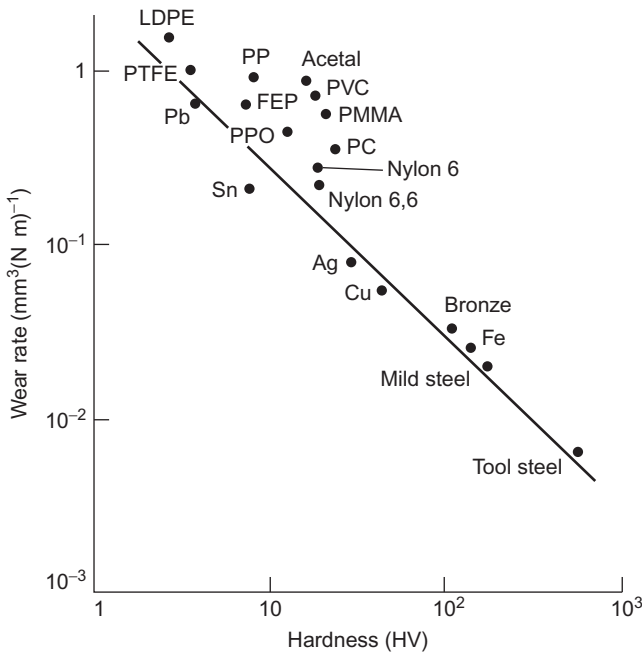


Fig. 6.19 Correlation between two-body abrasive wear rate (with silicon carbide abrasive) and bulk hardness for several polymers and metals (from Evans, D.C., Lancaster, J.K., 1979. In: Scott, D. (Ed.), *Treatise on Materials Science and Technology*, vol. 13. Academic Press, pp. 85–139)

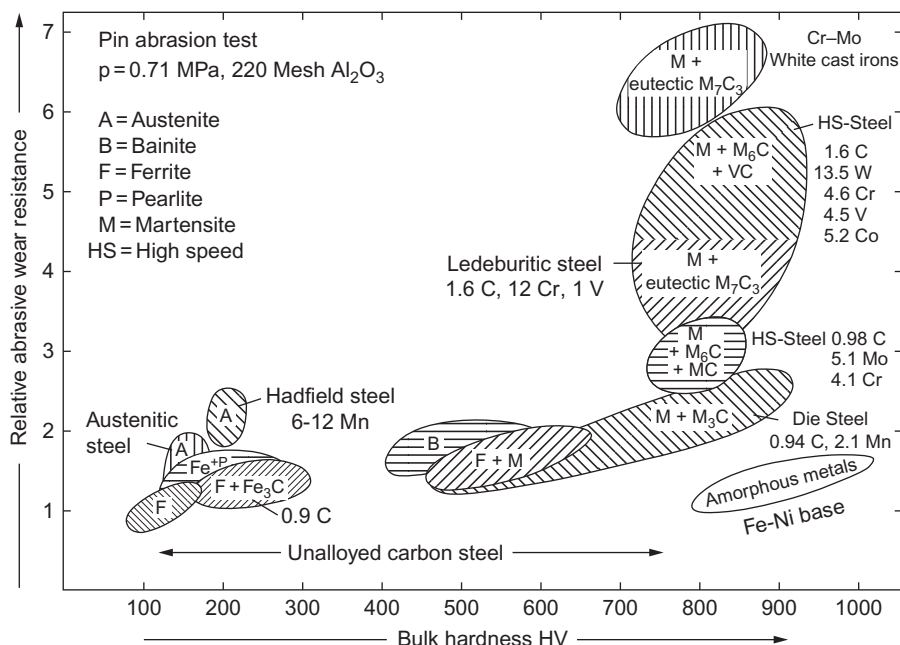


Fig. 6.20 Effect of microstructure and composition on the relative two-body abrasive wear resistance of steels and cast irons against $70 \mu\text{m}$ alumina particles (from Zum Gahr, K.-H., 1987. *Microstructure and Wear of Materials*. Elsevier)

representation of the behaviour of steels is shown in Fig. 6.20. As explained in Section 6.3.1, hardening by cold work has no effect on abrasion resistance, because the wear process itself introduces even more severe surface strains. In contrast, an increase in carbon content, even if the hardness remains the same, leads to higher abrasion resistance. Different trends in the dependence of wear resistance on hardness occur for steels of different microstructure. Austenitic steels show greater wear resistance at the same hardness than pearlitic or bainitic steels, whereas martensitic steels exhibit lower resistance. These observations can be understood in terms of the ductility and strain-hardening rates of the various microstructures: the lower ductility of martensite compared with austenite leads to removal of a greater fraction of the groove volume as wear debris, and hence to a higher value of K . Pearlitic and bainitic steels show intermediate behaviour. Retained austenite, for example in a martensitic or bainitic structure, is beneficial for resistance to abrasion by hard particles since it strain hardens during abrasion and exhibits high ductility.

Some materials in Fig. 6.18, notably the white cast irons and very hard ceramics, show a decrease in wear resistance with increasing bulk hardness. We can understand this by recalling that under severe contact conditions in brittle materials wear by fracture may occur (see Section 6.3.2). In these circumstances maximum wear resistance will be achieved in a material with intermediate values of hardness and toughness,

since in general materials with high hardness have low fracture toughness, and vice versa. Figure 6.21, which although schematic is supported by experimental data for a wide range of materials, illustrates this point. It shows the relationship between fracture toughness and both wear resistance ($1/Q$) and hardness, under severe abrasion conditions. Materials with low fracture toughness, such as ceramics, tend to be hard, and their resistance to abrasive wear by brittle fracture increases with their toughness. Materials of high toughness, such as metals, tend in contrast to be softer, and suffer abrasive wear by plastic deformation rather than brittle fracture. For these materials, wear resistance increases with increasing hardness, and so falls with increasing toughness. At intermediate values of hardness and toughness, in materials such as hard tool steels and white cast irons, both hardness and toughness play important roles. The simplified picture represented by Fig. 6.21 also shows how increasing the severity of the contact conditions, for example by increasing the size of the abrasive particles, their angularity or the normal load, leads to an increase in the extent of the fracture-dominated regime, and to a consequent shift in the peak in wear resistance towards higher values of fracture toughness.

Many materials exposed to abrasive wear contain hard phases within a softer matrix. Such microstructures may be developed during casting or forging, or by heat treatment, or the materials may be composites synthesized from two or more distinct components. For example, hard transition metal carbides (e.g., VC, NbC, M_3C , M_6C , M_7C_3 , etc.) form important constituents of hard steels, white cast irons (e.g., Ni-Cr, high Cr or high Cr-Mo irons), cobalt-based alloys (e.g., ‘Stellites’) and cemented carbides (hardmetals, e.g., WC-Co). All these materials are often used as wear resistant materials. Composite materials, with polymer, ceramic or metal matrices and harder reinforcing constituents, are also sometimes used in applications where they are exposed to abrasive wear.

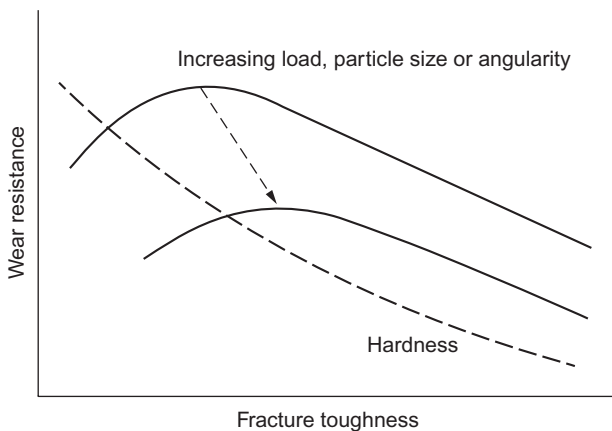


Fig. 6.21 Relation between fracture toughness and resistance to ‘hard’ abrasive wear for metallic and ceramic materials (from Zum Gahr, K.-H., 1987. *Microstructure and Wear of Materials*. Elsevier)

The response of all these materials depends on the size of the hard phase regions compared with the scale of the deformation caused by individual abrasive particles. The scale of this deformation can be described either by the width or the depth of the indentation formed by each particle (i.e., a or x in Fig. 6.7). If this dimension is substantially greater than the size of the hard particles (or fibres, in a fibre-reinforced composite) and their separation, then the material will behave very much like a homogeneous solid. This situation is illustrated schematically in Fig. 6.22(a). A finely dispersed hard second phase causes an increase in the flow stress of the matrix (e.g., by dispersion or precipitation hardening) which leads in general to increased wear resistance. Thus there is a direct correlation between carbide volume fraction in quenched and tempered steels and their resistance to abrasive wear. Furthermore, for a given carbide volume fraction the wear resistance is improved by a finer carbide distribution, leading to a shorter inter-particle spacing. Coherent or semi-coherent hard phases (formed for example by precipitation hardening) have a greater effect on abrasion resistance than incoherent particles. The benefits of hard carbide phases are also clear from Fig. 6.20.

This general rule that a high volume fraction of hard second phase particles is desirable for abrasion resistance is not, however, universally true. The matrix must also possess adequate toughness. In some materials where the matrix itself is brittle and the particle-matrix interface is weak, hard particles can act as internal stress-concentrators; cracks initiate at the interface and then propagate through the matrix. A high volume fraction of the second phase in such a material then leads to enhanced wear by fracture mechanisms.

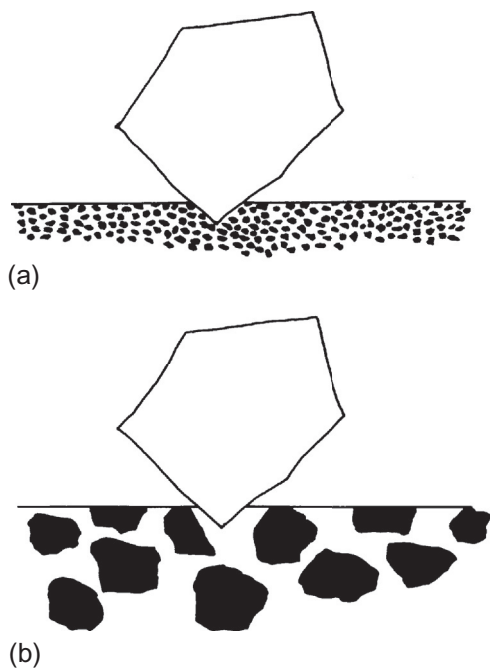


Fig. 6.22 Illustration of the importance of the relative sizes of the particle contact zone and the hard phase regions in the abrasive wear of a two-phase (or composite) material. In (a) the material responds in a homogeneous manner, whereas in (b) the response is heterogeneous (from Hutchings, I.M., 1994. *Mater. Sci. Technol.* 10, 513–517, by permission of the publisher, Taylor & Francis Ltd.)

If the hard phase particles are comparable in size with the scale of the abrasion damage, or larger, then the material will respond heterogeneously (see Fig. 6.22b). The action of an abrasive particle on a region of the hard phase can lead to plastic flow or to fracture, depending on the load carried by each abrasive particle and on its geometry, as well as on the sizes and the mechanical properties of the abrasive and hard phase particles and of the matrix. If the matrix mean-free path (a measure representing the distance between adjacent particles) is large compared to the scale of the abrasion damage, then preferential matrix removal may occur leading to a lack of support for the hard-phase which is subsequently removed by pull-out. Finally, it should be noted that the wear rate will also depend on the strength of the interface between the hard phase and the matrix.

An example of the complexities possible in these circumstances is provided by the behaviour of white cast irons. Figure 6.23 shows how the abrasive wear resistance of a range of Cr–Mo white cast irons varies with carbide content. The series of alloys contained varying volume fractions of $(\text{Fe}_3\text{Cr}_4)\text{C}_3$ carbides in an austenitic matrix. The bulk indentation hardness of the alloys increased almost linearly with carbide volume fraction (Fig. 6.23a). Pin-on-disc abrasion tests (see Section 6.3.5) with garnet abrasive particles showed fair correlation between wear resistance and bulk hardness (Fig. 6.23b): an increase in hardness led to lower wear rate as predicted by equation 6.5. With silicon carbide abrasive, however, (Fig. 6.23c) an inverse correlation was found: the wear rate *increased* with carbide content and with bulk hardness. Under these conditions of two-body sliding abrasion the hard silicon carbide particles penetrated the brittle carbide phase, causing it to crack, whereas the softer garnet particles did not cause fracture. Such behaviour is not simply a function of the hardness of the abrasive material; it also depends on the conditions of the test. In a three-body rubber-wheel abrasion test (see Section 6.3.5) the wear rates of exactly the same white irons due to alumina abrasive fell steadily with increasing carbide content (Fig. 6.23d), but with softer quartz abrasive (Fig. 6.23e) the wear rate reached a minimum at about 30% carbide volume fraction and then rose again. This rise in wear rate was again associated with fracture of the carbides. In this case it was probably caused by preferential wear of the softer matrix by the quartz, which exposed the massive carbide particles and rendered them more vulnerable to fracture.

Cemented carbides are often used to resist abrasive wear and form another class of composite materials containing both hard and soft phases: in this case a high volume fraction (typically 60–95%) of very hard carbide particles in a softer and more ductile metallic binder. The most commonly used carbide is tungsten carbide, with a hardness of 1900–2100 HV, in a cobalt matrix. The bulk (average) hardness of cemented carbides ranges from 800 to 2000 HV. High hardness is associated with a low binder content and small carbide grains. With such high bulk hardnesses, the abrasion of cemented carbides can often be considered as ‘soft’ abrasion (see Section 6.2.1). Wear by quartz, for example, with a hardness of about 800 HV, can fall into this category. Appreciably harder abrasive particles, such as silicon carbide or diamond, will cause ‘hard’ abrasion. Both the mechanisms of wear and the resulting wear rates depend strongly on the ratio of hardness between the abrasive particles and the cemented carbide, and transitions between hard and soft abrasion with different abrasive particles are commonly encountered. Under

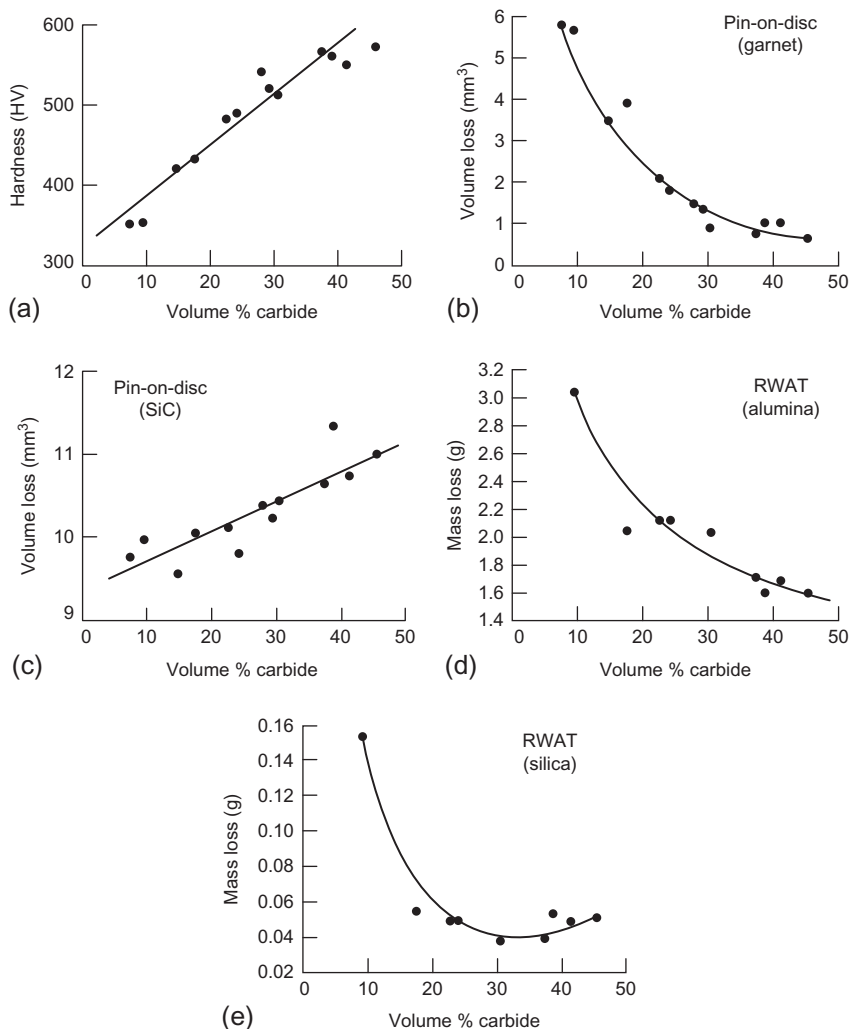
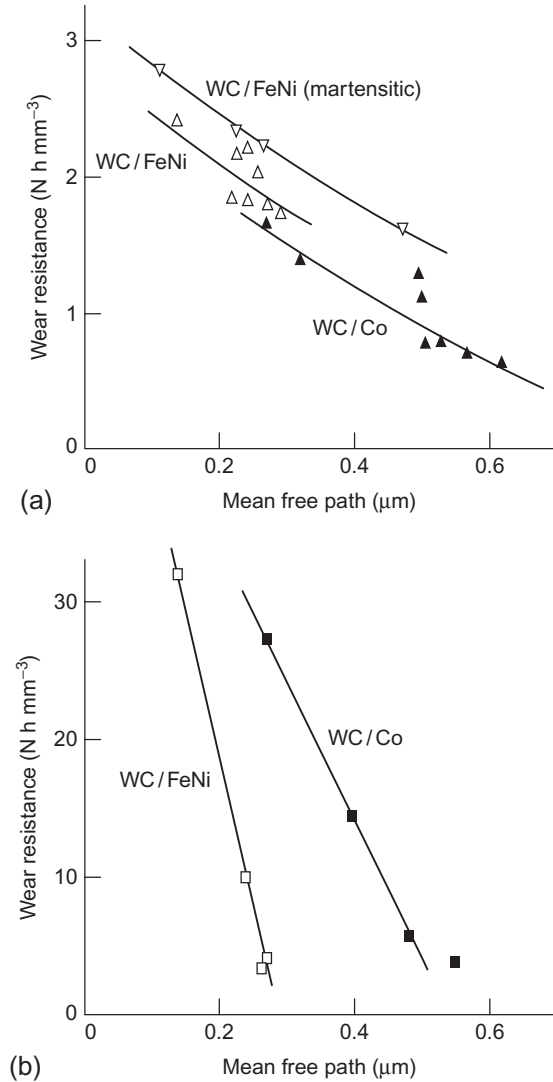


Fig. 6.23 Abrasive wear rates of a range of high Cr–Mo white cast irons, containing different volume fractions of carbides: (a) bulk microhardness of the alloy; (b) pin on disc (two body) tests with 100 μm garnet particles; (c) pin on disc (two body) tests with 90 μm silicon carbide particles; (d) rubber wheel (effectively three body) tests with 250 μm alumina particles; (e) rubber wheel (effectively three body) tests with 250 μm quartz particles (data from Zum Gahr, K.-H., Eldis, G.T., 1980. *Wear* 64, 175–194, and Fulcher, J.K., Kosel, T.H., Fiore, N., 1983. *Wear* 84, 313–325)

hard abrasion conditions, abrasive particles cause plastic deformation in both the carbide particles and the binder phase. The contact area for individual abrasive particles is usually much larger than the carbide particle size (typically 0.3–3 μm), and wear occurs predominantly by plastic ploughing and cutting, with some local associated fracture in the more brittle composites. A correlation is

Fig. 6.24 Resistance of tungsten carbide cemented carbides with cobalt and iron-nickel binders to two-body abrasive wear: (a) by silicon carbide particles ('hard' abrasion); (b) by quartz particles ('soft' abrasion) (from Larsen-Basse, J., 1983. In: Viswanadham, R.K., Rowcliffe, D.J., Gurland, J. (Eds.), *Science of Hard Materials*. Plenum Press, pp. 797–813, with permission of Springer)



found between the abrasion resistance and the mean free path in the binder phase between the carbide grains, as shown in Fig. 6.24(a). A short mean free path, due to a high volume fraction of fine carbide particles, gives the highest abrasion resistance. It also leads to high bulk hardness. Figure 6.24(a) also illustrates the effect of the strength of the metallic binder phase. An iron-nickel alloy, with higher strength than cobalt, gives greater wear resistance for the same mean free path. Increasing the strength of the binder further by heat treatment (which forms martensite) causes still further increase in wear resistance.

In the 'soft' abrasion regime, when the abrasive particles are softer than the cemented carbide, the wear rate is lower and the mechanisms of wear are different. [Figure 6.24\(b\)](#) shows that the wear rate of cemented tungsten carbides by quartz is about one tenth of that due to silicon carbide in the hard abrasion regime. The softer abrasive particles cannot produce plastic grooves in the carbide particles. Instead, the cyclic normal and tangential forces applied to the carbide grains cause them to move slightly, and these repeated small displacements cause gradual extrusion of the metallic binder from between the grains. The extruded binder can then be removed by the abrasive particles by plastic deformation. Removal of the binder leads either to detachment of the carbide particles or to their cracking, perhaps by a fatigue mechanism. As in the case of hard abrasion, there is a strong inverse correlation between wear resistance and the mean free path in the binder, presumably because this influences the ease of extrusion of the binder material.

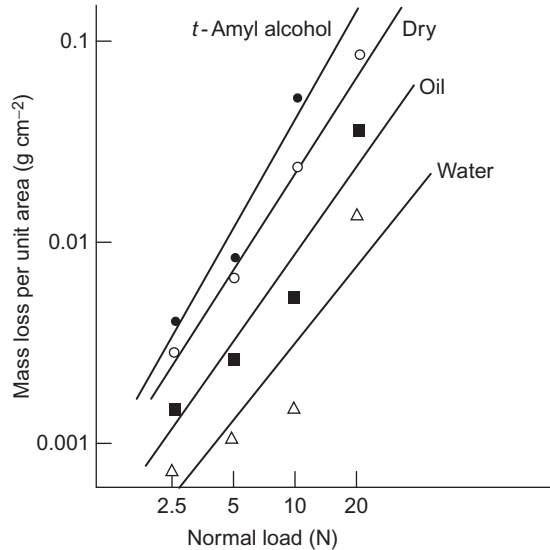
The complex behaviour of these composite materials illustrates clearly the general rule that 'wear resistance' can never be regarded as an intrinsic material property. The wear rate, and the mechanisms of wear, depend not only on the composition and microstructure of the material, but also to a very important extent on the conditions to which it is exposed.

6.3.4 Effects of lubrication and environment, and abrasion-corrosion

Lubrication and the nature of the environment (e.g., atmospheric oxygen content or humidity) can both act as powerful influences on wear by abrasion. In an environment where significant corrosion occurs in conjunction with abrasion, additional mechanisms come into play and the overall process becomes one of *tribocorrosion*.

The effect of lubrication on abrasive wear contrasts strongly with that on sliding wear. As we saw in [Chapter 5](#), lubrication in sliding wear tends to reduce wear by lowering the tangential stresses on the surface and by decreasing the incidence and severity of asperity contact. Between relatively smooth surfaces, protective lubricant films are readily formed. Abrasive particles, however, will often be larger than the hydrodynamic or elastohydrodynamic film thickness in a lubricated system, and the lubricant cannot then prevent contact between the particle and the counterface. In abrasive wear, therefore, lubrication will not cause the large reduction in wear rate seen in the absence of hard particles, and often instead results in an increase. For example, the abrasive wear rates of metals under two-body conditions may be increased by factors of up to three when lubricating oil is applied. The main reason for this effect has already been outlined in [Section 6.3.1](#): the lubricant reduces the friction between an abrasive particle and the metal surface, thereby widening the range of attack angles for which the cutting mode of deformation will occur, leading to increased efficiency of material removal. Lubrication may also have the subsidiary effect of inhibiting adhesion of wear debris to the abrasive particles, so slowing the degradation of abrasivity under multiple-pass conditions by capping or clogging (see [Section 6.3.5](#)). For these reasons, enhancement of wear rate by lubrication would be expected whenever plastic abrasion mechanisms are dominant.

Fig. 6.25 Comparison of two-body abrasive wear rates due to silicon carbide particles of sintered glass-bonded alumina in different environments (from Hines, J.E., Bradt, R.C., Biggers, J.V., 1977. In: Glaeser, W.A., Ludema, K.C., Rhee, S.K. (Eds.), *Wear of Materials*. ASME, pp. 462–467, by permission of the publisher)



The effect of lubrication in brittle materials is less consistent. Although reducing the tangential force by lubrication may reduce the tendency to cracking, lubricants may also, through local chemical effects at the crack tip, cause enhanced crack growth rates in the material being abraded. Environmental factors can also provoke a transition under otherwise constant abrasion conditions from fracture-dominated wear to plastic processes. A series of tests on alumina (Fig. 6.25) illustrates the effects of lubricants and other liquids on abrasive wear in a ceramic. The wear rate varied by a factor of up to 10 between different liquid environments; the wear rates in water-soluble oil and in water were lower than in air, whereas that in tertiary amyl alcohol was significantly higher.

For plastics materials and elastomers, the influence of lubrication depends on the dominant wear mechanism. A hard thermoplastic, suffering abrasion by plastic processes, will respond in a similar way to a ductile metal, and lubrication may well increase the wear rate. If the material is an elastomer, and the wear mechanism is one of progressive crack growth driven by surface tractions (see Section 5.11.2), then lubricants will decrease the rate of wear by reducing the frictional forces. For this reason, the abrasive wear rates of rubber vehicle tyres in wet conditions are much lower (typically one tenth the value or less) than when dry.

Atmospheric composition can also influence abrasive wear rates, with oxygen content and humidity being potent factors. The effects of the gaseous environment are often complex; even in ductile metals the effects can vary in magnitude and even in direction. In general, oxygen provides a lubricating action in the abrasion of metals. Rapid oxidation of the freshly exposed metal surface at the tip of the abrasive particle produces a thin oxide film with low shear strength and reduces adhesion between the particle and the metal. A similar effect is observed in conventional metal cutting, where aggressive small molecules such as oxygen, chlorine or tetrachloromethane can react rapidly with clean metal both on the workpiece and on the cutting tool itself

and reduce the frictional force on the tool (see [Section 9.4.1](#)). The abrasive wear of metals in the absence of oxygen is usually less severe than with it present; not only is cutting less efficient, but in two-body sliding abrasion increased adhesion between metallic debris and the abrasive particles also leads to a progressive reduction in wear rate, due to capping and clogging (as discussed in [Section 6.3.5](#)).

There is good evidence for significant dependence of abrasion rates in air on the level of atmospheric humidity. An extreme example is seen in the abrasion of magnetic recording heads by tape coated with iron oxide ($\gamma\text{-Fe}_2\text{O}_3$) particles, where an increase in wear rate by a factor of up to 10 has been noted with an increase in the relative humidity of the surrounding air from 1% to 50%. Much smaller effects occur in laboratory two-body abrasion tests, which are probably related to the influence of humidity on the mechanical properties of the abrasive particles, as well as to a lubricating action similar to that of oxygen. As we saw in [Section 5.10.4](#), environmental factors play an important role in the sliding wear of ceramic materials, and similar effects would be expected to occur just as readily when the ceramic constitutes the abrasive particles as when it forms the abraded surface.

If the environment favours significant corrosion as well as abrasion, then there may be *synergy* between the two processes and the overall rate of material removal can be greater (in some cases much greater) than the sum of the two processes acting separately. The fundamental concepts of corrosion and tribocorrosion are presented in [Appendix B](#); here we consider the mechanisms by which the processes of corrosion and abrasion may interact. The degradation is called *abrasion–corrosion* if the ratio between W , the mechanical contribution from abrasion, is greater than C , the chemical contribution from corrosion (i.e., $C/W < 1$). It is *corrosion–abrasion* if $C/W > 1$. As outlined in [Appendix B](#), the most common source of the enhanced corrosion rate caused by abrasion (ΔC_w) is the mechanical removal of the passive oxide film from the metal surface; the film then regrows rapidly and the overall rate of corrosion may be much higher than that in the absence of mechanical action. This mechanism is important, for example, in the abrasion–corrosion of many metallic alloys at high temperatures in air, or in corrosive and abrasive aqueous slurries. However, other synergistic mechanisms are possible. For example, in the abrasion of WC–Co cemented carbides by an acidic aqueous slurry, corrosion of the cobalt binder phase leads to its removal and the consequent loosening and easier detachment of the discontinuous tungsten carbide phase, and thus to a large and positive value for ΔW_c . [Figure 6.26](#) shows experimental data for the abrasive–corrosive wear of several different bulk metals, cemented carbides and coatings in aqueous media, plotted to highlight the relative importance of the mechanical (W) and corrosion (C) components of the total degradation rate (T). For only a few cases is $C/W > 1$; for most examples, the mechanical contribution dominates, although the corrosion component is still significant.

6.3.5 Testing methods for abrasive wear

The most commonly used laboratory tests for abrasive wear employ either a pin-shaped specimen sliding against fixed abrasive, or a rotating wheel or ball sliding against a plane specimen with loose abrasive particles being continuously fed between the two.

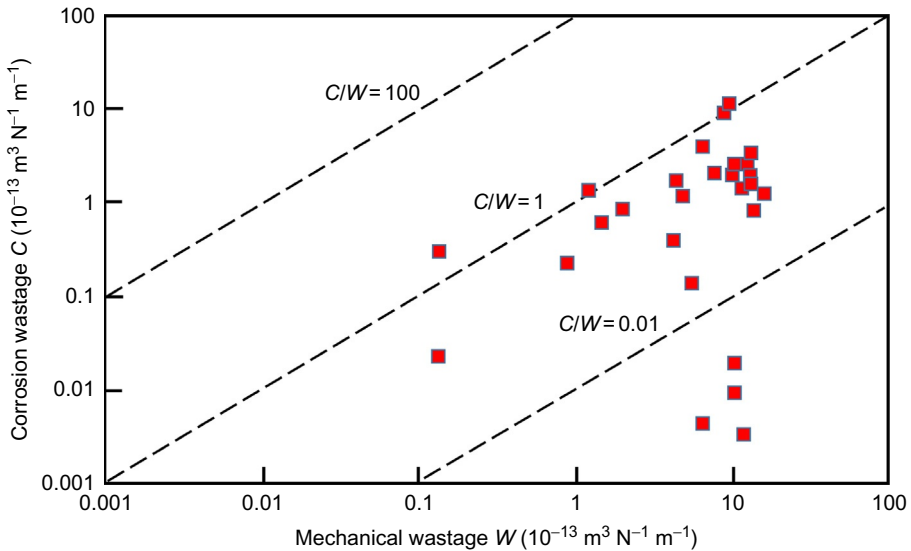


Fig. 6.26 Example of a tribo-corrosion regime map, for abrasion–corrosion of various materials including sintered cemented carbides, stainless and duplex stainless steels and thermally sprayed WC coatings. (from Wood, R.J.K., 2007. *J. Phys. D: Appl. Phys.* 40, 5502–5521, © IOP Publishing. Reproduced with permission. All rights reserved)

Figure 6.27(a)–(c) shows three common variants of the method in which a specimen pin slides against fixed abrasive particles, causing two-body (sliding) abrasive wear. Commercial bonded-abrasive paper or cloth is usually used for the counterface, carrying evenly distributed grit particles of a narrow size distribution, bonded to the substrate by a strong resin. The wear rate due to such fixed abrasive particles decreases with repeated passes of the specimen over the same track. Several mechanisms are responsible for this progressive reduction in abrasivity: fracture of particles, leading to a decrease in the number of cutting points; removal of whole particles from the binder resin (‘shelling’); rounding of the contacting areas of particles by chemical, mechanical or thermal mechanisms (‘attrition’); adhesion of wear debris to the tips of particles (‘capping’); and accumulation of wear debris in the spaces between particles so that it carries part of the applied load (‘clogging’). The rates and relative importance of these mechanisms vary with specimen material, and also with load, sliding speed, atmosphere and other factors. In order to avoid the problems introduced by degradation of the abrasive, it is often ensured that the specimen always slides against fresh abrasive. In the pin-on-disc geometry (Fig. 6.27a), this can be achieved by moving the pin radially on the disc during the test, so that it describes a spiral track. Alternative geometries involve linear sliding over a rectangular sheet of abrasive paper (Fig. 6.27b), or moving the pin parallel to the axis of a rotating cylinder covered with abrasive particles (Fig. 6.27c). Apparatus of this type is specified in a widely-used international standard procedure for abrasion testing of rubber (ISO 4649; see also ASTM D5963 and DIN

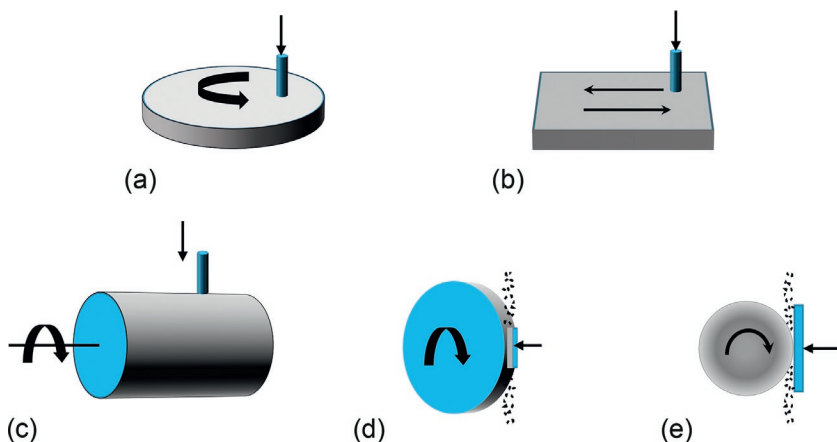


Fig. 6.27 Schematic illustration of five common methods used to measure abrasive wear rates of materials: (a) pin on abrasive disc; (b) pin on abrasive plate; (c) pin on abrasive drum; (d) rubber wheel abrasion test; (e) micro-scale abrasion test

53516). In each of these cases, a constant load is applied to the pin, often by a dead weight. The wear rate is usually measured by weighing the pin before and after the test, although in some designs an electrical transducer is used to monitor the position of the pin continuously, and hence measure its instantaneous length. In this way a continuous record of wear can be obtained during the test.

Figure 6.27(d) and (e) illustrates a second class of abrasive wear test, in which loose abrasive particles are supplied either as dry powder or mixed with a liquid to form a slurry. The specimen is in the form of a plate or block, pressed under constant load against the surface of a rotating wheel or ball. The method shown in Fig. 6.27(d), the dry sand/rubber wheel abrasion test, has been adopted as a standard (ASTM G65) and the wheel consists of a rubber rim of defined hardness (i.e., elastic modulus) moulded on to the surface of a steel disc. In the standard specification the radial thickness of the rubber rim is 12.7 mm, and the wheel has a width of 12.7 mm and an overall diameter of 228.6 mm; it rotates at 200 r.p.m. The force pressing the specimen against the wheel is also specified. Dry silica (quartz) particles of a narrow size distribution and from a specified source are fed at a constant rate into the contact region. Wear is measured by weighing the specimen; the standard procedure prescribes the duration of the test, as well as other test variables such as the load, sliding speed, hardness of the rubber and particle feed rate.

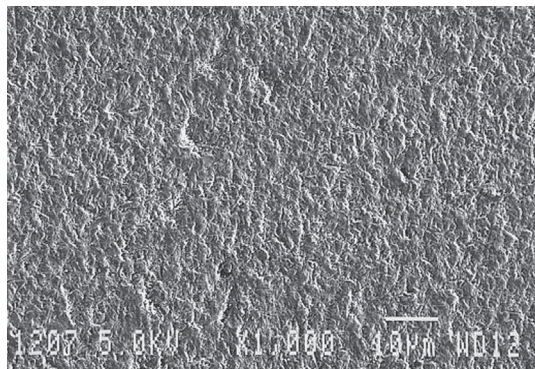
The test shown in Fig. 6.27(e) is usually called ‘micro-scale abrasion’ or ‘ball-cratering’. The ball is typically a hard steel bearing ball, 25 mm in diameter, and the abrasive particles (about 4 μm in diameter) are fed to the contact zone as a concentrated slurry, suspended in water. The wear scar produced in the initially plane specimen surface by this test is typically 1–2 mm in diameter, and has the geometry of a spherical cap with the same radius as the ball; the wear volume can therefore be accurately calculated from a measurement of the diameter of the wear scar in the

surface of the specimen. Because the wear scar occupies such a small area of the specimen, it is possible to carry out several tests on a relatively small sample. The depth of the wear scar is also small ($\sim 10\text{--}20\text{ }\mu\text{m}$ for wear scar diameters of 1–1.5 mm), and the method is therefore well suited to measuring the wear rates of thin coatings. Methods exist to extract the wear rates of both coating and substrate from a suitable series of tests, even when the wear scar has penetrated through the coating so that the wear volume contains contributions from both the coating and the substrate. The derivation of a value of wear rate from experimental measurements is discussed further in [Appendix C](#).

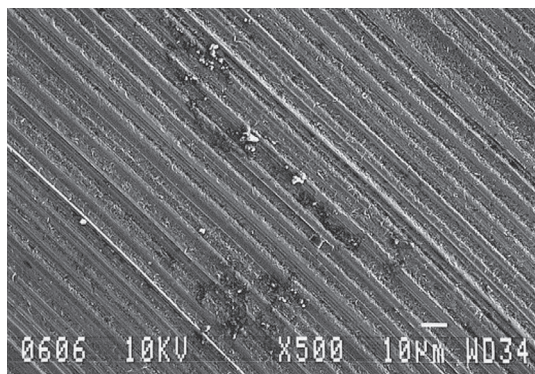
The methods shown in [Fig. 6.27\(d\) and \(e\)](#) have also been modified to study abrasion–corrosion, by suspending the abrasive particles in a corrosive aqueous medium. Examples have included acidic or alkaline solutions, real or synthetic seawater or chloride solutions, and saliva. Measurements carried out with and without the presence of the corrosive agent can be used to explore the synergy between corrosion and wear, as discussed in [Appendix B](#). The micro-scale abrasion test in particular has also been adapted so that electrochemical measurements can be made during wear tests, for example by using the standard 3-electrode potentiodynamic method (described in textbooks on corrosion), or so that corrosion of the sample can be inhibited by applying a sufficient cathodic potential. For these tests an insulating material, typically a ceramic, is used for the rotating ball.

The two groups of test shown in [Fig. 6.27](#) subject the specimens to rather different conditions. The dry sand rubber wheel and micro-scale abrasion tests involve loose abrasive particles that indent the wheel or ball to a greater or lesser extent during abrasion. Examination of the wear scars shows that in some cases the particles are dragged through the contact zone in a sliding motion, while in others they roll; both rolling and sliding can occur in the same test, for example in different regions of the contact area, or at different times during the test. With a compliant rubber wheel, even if particles do not roll during contact, they can certainly rotate significantly. In doing so they will tend to move away from ‘cutting’ orientations in order to minimize the frictional energy dissipation, so that conditions in this test tend to be closer to rolling abrasion than to sliding. Measured wear rates ($K \approx 5 \times 10^{-4}$ for metals) certainly compare more closely with values associated with three-body rolling abrasion than with sliding abrasion. Particle motion in the micro-scale abrasion test depends on the concentration of the particles in the slurry, the applied load and the relative hardness of the ball and sample. Both rolling and sliding motion can be achieved by changing the test conditions, as illustrated in [Fig. 6.28](#).

The tests that involve pin specimens sliding on fixed abrasives ([Fig. 6.27a–c](#)), in contrast, produce true two-body sliding conditions. In many cases, probably because they are readily available, commercial alumina or silicon carbide abrasive papers are used. As we have seen (in [Section 6.2.1](#)), these abrasive materials are much harder than naturally occurring abrasive particles, and care must be taken in applying the results of wear tests with such hard particles to practical conditions in which the abrasive particles are usually much softer. The hardness, shape and size of the abrading particles will all affect the wear rate, for the reasons outlined in the previous sections: they will control the relative importance of the various mechanisms of wear, and the apparent wear resistance of materials may change markedly when the results from different types of abrasion test, employing different abrasive particles and different conditions, are compared.



(a)



(b)

Fig. 6.28 SEM images of the worn surfaces of quenched and tempered tool steel produced by micro-scale abrasion at a normal load of 0.25 N with (a) high concentration of abrasives in the slurry; showing rolling abrasion, and (b) low concentration of abrasives in the slurry, showing grooving abrasion (from Trezona, R.I., Allsopp, D.N., Hutchings, I.M., 1999. *Wear* 225–229 (Part 1), 205–214)

A further effect that contributes to a difference between the results of different types of abrasion test concerns the compliance of the support provided to the abrasive particles. In the rubber wheel test (Fig. 6.29a), abrasive particles in contact with harder phases in a two-phase material or composite can penetrate more deeply into the rubber and thus carry less of the total load than if the support is more rigid, as occurs with a steel ball in the micro-scale abrasion test and is commonly the case in two-body abrasion tests. There will therefore be a greater tendency for hard brittle phases to fracture with a rigid abrasive support (Fig. 6.29b) than with a more compliant backing.

Quite apart from effects due to the mechanical design of the test method, there are strong effects of the abrasive material alone, as shown earlier in Fig. 6.23 and further illustrated in Fig. 6.30. This shows the wear rates for two martensitic ferrous alloys, abraded in a pin-on-plate test (Fig. 6.27b). The two materials were a plain carbon steel (0.9% carbon) and a 27% chromium white cast iron, with nearly the same bulk hardness. Against silicon carbide abrasive their wear rates were similar, but they showed very different performance when abraded against garnet; alumina gave intermediate behaviour. The main reason for these differences lies in the response of the two materials, and in particular of the hard chromium carbide phase in the cast iron, to abrasives of different hardness. Extensive fracture of this brittle carbide occurs with silicon

Fig. 6.29 Illustration of the effect of the compliance of the abrasive support in abrasive wear. If the support is compliant (a) the particles in contact with harder regions of the counterface can deflect the support and thus carry less of the total load than if the support is rigid, as at (b). (from I. M. Hutchings, in *Materials Science and Technology* 10, 513–517, 1994, by permission of the publisher, Taylor & Francis Ltd.)

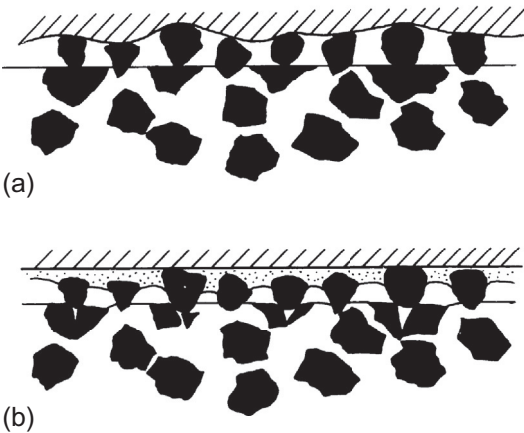
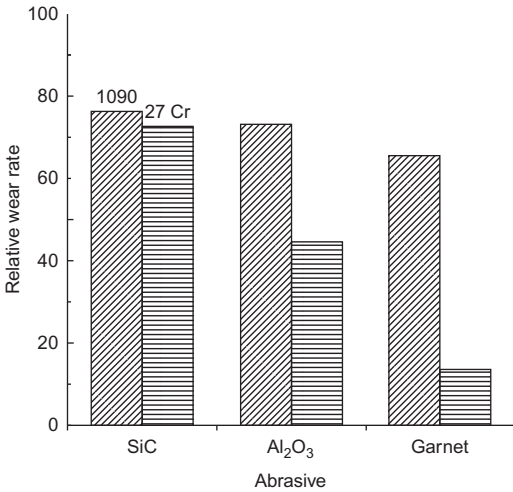


Fig. 6.30 Relative wear rates of two martensitic ferrous alloys (AISI 1090 steel with 0.9% carbon, and 27% Cr 2% Mo white cast iron) with similar values of bulk hardness (735 and 765 HV respectively), measured in pin-on-plane two-body abrasion tests with silicon carbide, alumina and garnet abrasives (data from Gundlach, R.B., Parks, J.L., 1977. In: Glaeser, W.A., Ludema, K.C., Rhee, S.K. (Eds.), *Wear of Materials*. ASME, pp. 211–216)



carbide abrasive (which is harder than the carbide—see [Table 6.1](#)), whereas garnet, being softer, does not cause fracture.

As in the case of sliding wear discussed in [Chapter 5](#), care must be taken in applying the results of laboratory tests to practical applications. At the very least, it must be established that the mechanisms of wear under the two sets of conditions are the same. In order to provide closer simulation of particular applications, many other abrasive wear tests have been devised. One example is provided by a laboratory-scale jaw-crusher test, used to simulate the wear by gouging abrasion that occurs in larger-scale mineral processing plant in the mining industry, and adopted as a standard (ASTM G81). In the jaw-crusher, irregular lumps of rock are crushed as they fall between two flat plates (the jaws) that move relative to each other. [Figure 6.31](#) shows a simple example in which one plate moves in an arc relative to a fixed plate; more complex motions are used in

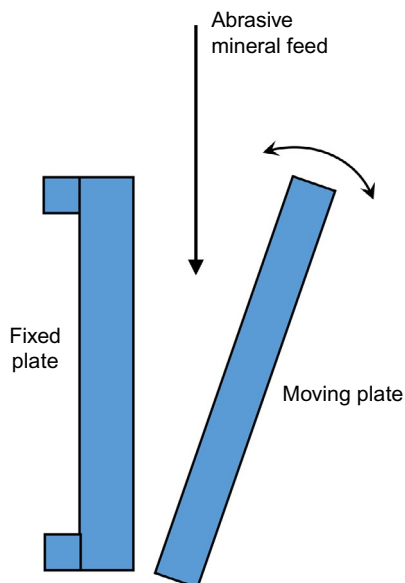


Fig. 6.31 Side view of the moving and stationary plates in one design of jaw crusher wear test apparatus.

some designs. The wear rate of a plate-shaped specimen mounted on one of the jaws is determined by weighing it after a fixed amount of standard rock (typically up to 50 mm in size) has been crushed. The dimensions of the apparatus and the operating conditions must of course be standardized. In order to compensate for the inevitable slight variability in test conditions, the other plate (or the neighbouring half of the same jaw in some implementations) is made from a standard alloy, and its wear rate is also measured. In this way, the wear rates of the test specimen and the standard material can be determined simultaneously and then expressed as a ratio.

The results of all these types of abrasion test are influenced by many experimental variables, and for test results to be reproducible between different laboratories careful attention must be paid to measuring and standardizing all the conditions. In the dry-sand/rubber wheel abrasion test, for example, which is closely defined, a series of wear rate measurements under the standard conditions on the same bulk material with the same apparatus will typically show a distribution of wear rates with a standard deviation of some 3% of the mean value. Between tests carried out under nominally identical conditions in different laboratories, the standard deviation may be 5% or more. At least part of this variability can be ascribed to the abrasive particles. Particle shape in particular has a strong influence on wear rate (see [Section 6.2.2](#)), and some variation in abrasivity will inevitably be found between different batches of abrasive even from the same source. In many cases it may be helpful to use a standard reference material to calibrate the test conditions, as outlined above for the jaw-crusher test.

6.4 EROSION BY SOLID PARTICLE IMPACT

The wear process known as *solid particle erosion* occurs when discrete solid particles strike a surface. It differs from three-body abrasion, which also involves loose particles, primarily in the origin of the forces between the particles and the wearing surface. In abrasion the particles are pressed against the surface and move along it, usually because they are trapped between two sliding surfaces (Fig. 6.1b). In erosion (Fig. 6.1c), several forces of different origins may act on a particle in contact with a solid surface. These are shown in Fig. 6.32. Neighbouring particles may exert contact forces, and a flowing fluid if present will cause drag. Under some conditions, gravity may be important. However, the dominant force on an erosive particle, which is mainly responsible for decelerating it from its initial impact velocity, is usually the contact force exerted by the surface. In abrasive wear, we saw that the amount of material removed depends on the normal load pressing the particles against the surface and on the distance slid (Sections 6.3.1 and 6.3.2). In erosion, the extent of wear depends instead on the number and mass of individual particles striking the surface, and on their impact velocity.

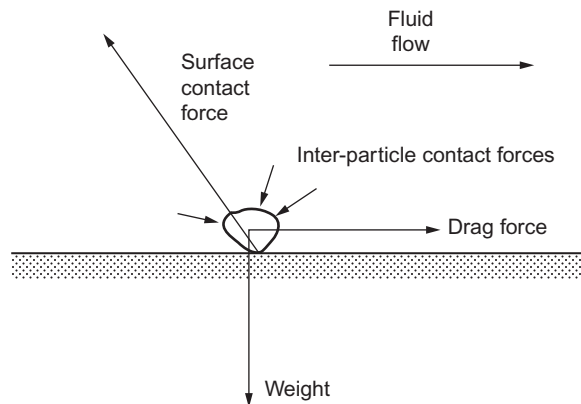
As in the case of abrasion, mechanisms of erosive wear can involve both plastic deformation and brittle fracture. We shall examine these in the following sections. Erosion of metals usually involves plastic flow, whereas more brittle materials may wear predominantly either by flow or by fracture depending on the impact conditions, as discussed below.

6.4.1 Erosive wear by plastic deformation

Before looking in detail at the mechanisms of erosion involving plastic deformation, it is helpful to examine the behaviour of a single hard particle striking a softer surface at normal incidence (Fig. 6.33).

We shall make the simplest possible assumptions: that the particle does not deform and that the problem can be analyzed quasi-statically (i.e., by ignoring dynamic effects

Fig. 6.32 Diagram showing the forces which can act on a particle in contact with a solid surface (from Hutchings, I.M., 1987. Chem. Eng. Sci. 42, 869–878)



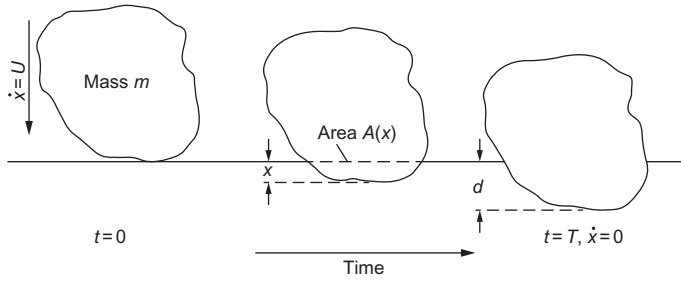


Fig. 6.33 The process of penetration of a rigid particle into the plane surface of a plastically deforming material, at normal incidence. Initial contact occurs at time $t=0$, and the particle comes to rest at time $t=T$

such as wave propagation and strain-rate sensitivity). The only force assumed to be acting is the contact force exerted by the surface. We shall further assume that the deformation of the surface is perfectly plastic, with a constant indentation pressure (hardness) H . At time t after initial contact, the particle of mass m will have indented the surface to a depth x ; the cross-sectional area of the indentation at the surface will be $A(x)$, where $A(x)$ is determined by the shape of the particle. The upward force decelerating the particle will be that due to the plastic flow pressure acting over the area $A(x)$, and the equation of motion of the particle can therefore be written as

$$m \frac{d^2x}{dt^2} = -HA(x) \quad (6.15)$$

For simple particle shapes this equation can readily be solved analytically, but for our purposes we wish to know only the final volume of the indentation, when the particle comes to rest at a depth d (Fig. 6.33). At this point the work done by the retarding force will be equal to the initial kinetic energy of the particle, which is assumed to have an initial velocity U :

$$\int_0^d HA(x) dx = \frac{1}{2}mU^2 \quad (6.16)$$

The final volume of the indentation V is given by

$$V = \int_0^d A(x) dx \quad (6.17)$$

and therefore since H is assumed to be constant

$$V = \frac{mU^2}{2H} \quad (6.18)$$

The material displaced from the indentation can suffer several possible fates: it may be accommodated by elastic deformation of material away from the indentation, it may form a rim of plastically deformed material around the indentation, or it may be removed in some way as wear debris. We shall discuss below possible mechanisms by which material can be removed, but for the present shall assume only that some fraction K of the material displaced from the indentation is removed as wear debris. We can therefore write

$$\text{mass of material removed} = K\rho \frac{m U^2}{2H} \quad (6.19)$$

where ρ is the density of the material being eroded and K is a dimensionless factor.

Summation of equation 6.19 over many impacts suggests that the total mass of material removed from the surface should be proportional to the total mass of the erosive particles that have hit it. Figure 6.34 shows how the mass lost from a surface varies with the total mass of erosive particles that have struck it. For some materials, particles may become embedded in the surface and cause an initial mass gain, as

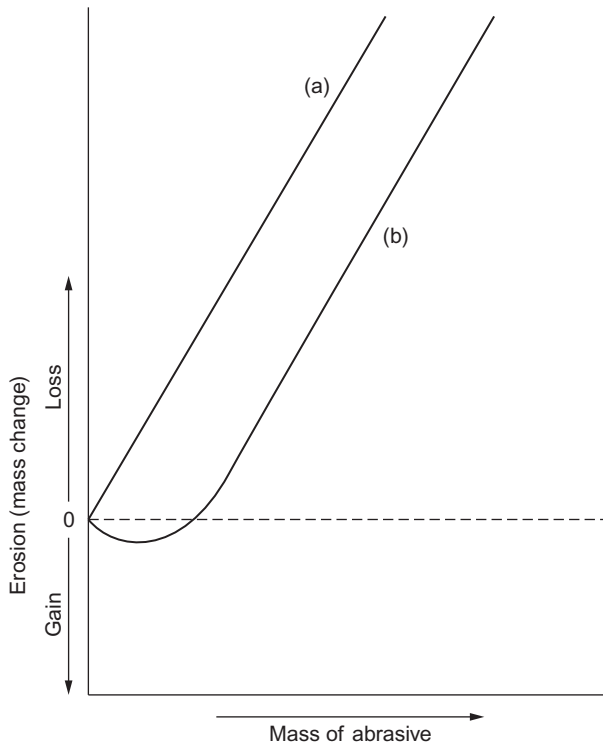


Fig. 6.34 Typical dependence of mass lost from the surface on the total mass of erodent particles which have struck it. Line (a) corresponds to linear erosion behaviour, with no incubation. Under some conditions, however, incubation behaviour (as in curve b) is observed

shown by curve (b). After this incubation period, which is observed mainly with soft target materials and tends to be more pronounced at high angles of incidence, the erosion proceeds linearly with the mass of particles striking the surface. For most ductile target materials and most types of erodent particles, however, any incubation period is negligible and the mass lost from the surface is closely proportional to the total mass of erodent particles that have struck the surface: line (a) in Fig. 6.34 is followed. The linear relationship observed in steady-state erosion allows a simple definition of erosion, E , to be used:

$$E = \frac{\text{mass of material removed}}{\text{mass of erosive particles striking the surface}} \quad (6.20)$$

E is dimensionless, and equation 6.19 becomes

$$E = \frac{K\rho U^2}{2H} \quad (6.21)$$

A comparison of equation 6.21 with equation 6.5 for abrasive wear is instructive. Both predict wear rates that are inversely proportional to H , the hardness of the surface. The role of the applied normal load in abrasive wear (W in equation 6.5) is played in erosion by the quantity $\rho U^2/2$. In both cases, the severity of wear is determined by a dimensionless *wear coefficient* K , which as we saw in Section 6.3.1 is analogous to the Archard coefficient for sliding wear. K can be thought of as a measure of the efficiency of the material removal process; if all the material displaced by the erosive particle were removed, K would be unity. For the erosion of metals, K lies typically in the range 5×10^{-3} to 10^{-1} , very similar to the values observed in two-body abrasion.

Equation 6.21 provides only a crude estimate of the factors controlling erosive wear; it ignores, for example, any variation of erosion rate with impact angle. To improve our understanding, we must examine in more detail the interaction between a hard particle and the surface of a ductile material.

The geometry of the deformation due to the impact of a hard particle depends on the impact velocity, on the shape and orientation of the particle and on the impact angle. Impact angles in erosion are usually defined relative to the plane of the surface, as shown in Fig. 6.35. For normal impact, $\theta = 90^\circ$, while at glancing incidence θ tends to zero. The erosion of ductile materials (e.g., most metals) depends strongly on impact angle, as illustrated in Fig. 6.35 (curve a), typically showing a maximum at $20\text{--}30^\circ$ and falling to one half to one third of the peak wear rate at normal incidence.

Studies of the impact of single particles on to metals at 30° impact angle show three basic types of impact damage, illustrated in Fig. 6.36. Rounded particles deform the surface by *ploughing*, displacing material to the side and in front of the particle (Fig. 6.36a). Further impacts on neighbouring areas lead to the detachment of heavily-strained material from the rim of the crater or from the lip at its end. This type of deformation embraces both the ploughing and wedge-forming modes of abrasion

Fig. 6.35 Typical dependence of erosion (as defined by equation 6.20) on impact angle θ (defined as the angle between the impact direction and the surface). Ductile metals commonly show peak erosion at a shallow impact angle (curve a), while brittle materials often show maximum wear for normal incidence (curve b)

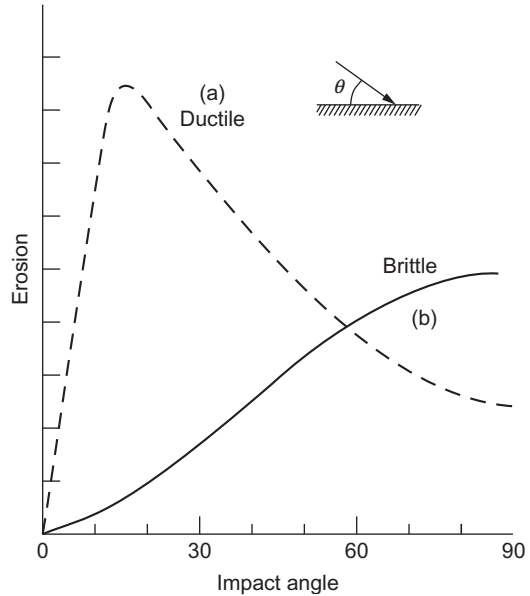
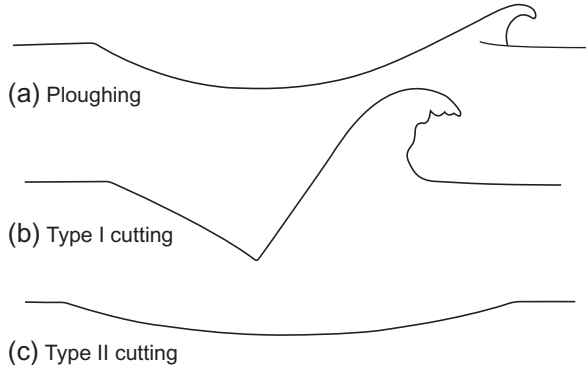
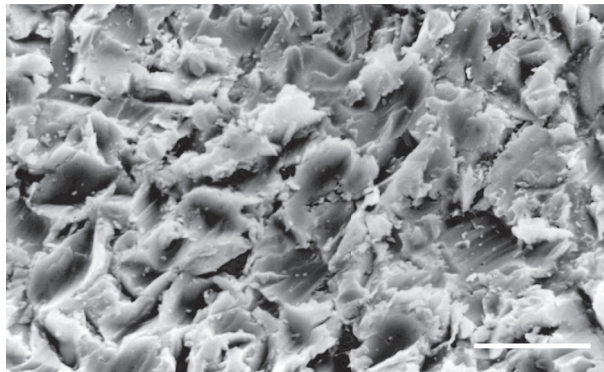


Fig. 6.36 Sections through impact sites formed by hard particles on a ductile metal, showing typical shapes. The impact direction was from left to right. (a) Ploughing deformation by a sphere; (b) type I cutting by an angular particle, rotating forwards during impact; (c) type II cutting by an angular particle, rotating backwards during impact

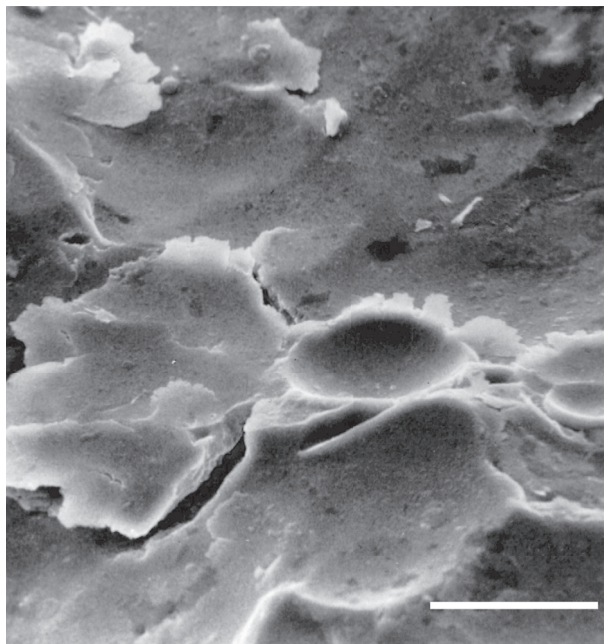


shown in Fig. 6.11(b) and (c). The deformation caused by an angular particle depends on the orientation of the particle as it strikes the surface, and on whether the particle rolls forwards or backwards during contact. In the mode which has been termed *type I cutting* (Fig. 6.36b), the particle rolls forwards, indenting the surface and raising material into a prominent lip, which is vulnerable to removal by subsequent nearby impacts. If the particle rolls backwards (Fig. 6.36c), a true machining action can occur, in which the sharp corner of the abrasive grain cuts a chip from the surface. This is *type II cutting* and occurs over only a narrow range of particle geometries and impact orientations.

While clear distinctions can be drawn between these three types of deformation when single particles strike a plane surface, classification is not so simple when randomly oriented particles of irregular shape impinge on a previously eroded and thereby roughened surface. Nevertheless, features similar to those caused by single particle impacts can be distinguished on surfaces eroded by the impact of many particles. [Figure 6.37\(a\)](#) shows, for example, a steel surface eroded by silicon carbide particles. Erosive wear is associated with the detachment by plastic rupture of metal displaced from the impact sites into raised crater rims and lips. Although each



(a)



(b)

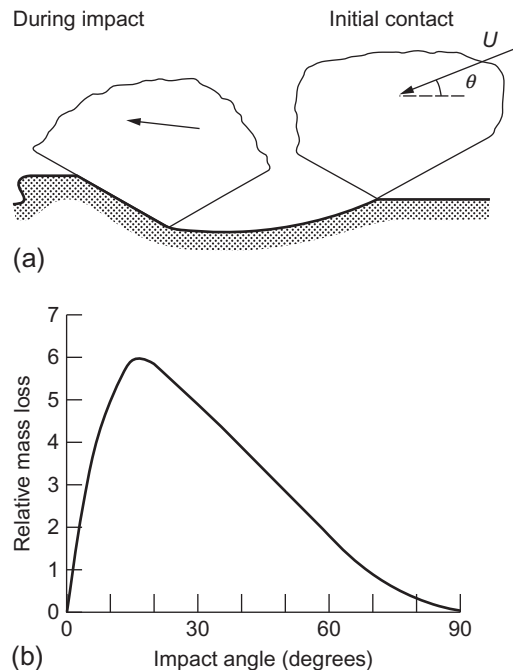
Fig. 6.37 Metal surfaces after erosion by hard particles: (a) mild steel eroded at 30° impact angle and 55 m s^{-1} by angular silicon carbide particles; (b) aluminium eroded at 90° by spherical glass beads at 60 m s^{-1} . Scale bars: (a) $10 \mu\text{m}$; (b) $50 \mu\text{m}$ (from Cousens, A.K., 1984. Ph.D. Dissertation, University of Cambridge, with permission, and Cousens, A.K., Hutchings, I.M., 1983. *Wear* 88, 335–348)

impact displaces material from the indentation formed, it will often not become detached as wear debris until it has experienced several cycles of plastic deformation and become severely work-hardened.

The shape of the abrasive particles influences the pattern of plastic deformation around each indentation (as in abrasive wear—[Section 6.3.1](#)) and the proportion of the material displaced from each indentation which forms a rim or lip. More rounded particles lead to less localized deformation, and more impacts are required to remove each fragment of debris. An increased impact angle has a similar effect. In the extreme case of spherical particles at normal incidence, material is removed only after neighbouring impacts have imposed many cycles of plastic deformation, and the surface looks very different from one eroded by angular particles ([Fig. 6.37b](#)).

Despite the evidence that most of the wear debris formed in erosion has been deformed by several impacts, many theoretical models advanced for erosion are basically models for the impact of a single particle striking a plane surface, with empirical extrapolation of the results to the practical case of multiple impact. One such approach extends the simple model for normal impact described above to oblique impact, by solving the equations of motion for a rigid angular particle striking a perfectly plastic material at a shallow angle, and assuming that material is removed only by a cutting action. [Figure 6.38](#) shows an idealized two-dimensional model of a rigid particle cutting into the plane surface. The volume of material removed is taken to be that swept out by the motion of the particle tip (as distinct from the total volume displaced by the indentation), and the model therefore

Fig. 6.38 Contact geometry assumed in a theoretical model for the erosion of ductile materials by a cutting mechanism: (a) shows the particle at initial contact and also during impact; (b) shows the angular dependence $f(\theta)$ predicted by the simplest form of the theory (from Finnie, I., McFadden, D.H., 1978. *Wear* 48, 181–190, and Finnie, I., 1995. *Wear* 186–187, 1–10)



predicts zero erosion for normal incidence. Various simplifying assumptions are possible: one method of analysis, in which the forces on the particle are assumed to act at its extreme tip, leads to the expression

$$E = \frac{K\rho U^2}{H} f(\theta) \quad (6.22)$$

The value of K depends on the geometry of the particle and on the fraction of particles actually cutting in an idealized manner (or, alternatively, on the fraction of the volume swept out by the particle tip which is actually removed as wear debris). The function $f(\theta)$ predicted by the theory is shown in Fig. 6.38, and is similar to the experimentally observed curve (a) in Fig. 6.35, although it falls to zero for normal incidence.

A similar but more realistic model, in which the point of action of the forces on the particle is allowed to move during the impact, leads to a more complex expression for E which can be approximated by

$$E = \frac{K_1 \rho U^n}{H} f_1(\theta) \quad (6.23)$$

where the velocity exponent n lies typically between 2.0 and 2.5, and is itself a function of the impact angle θ .

The models which lead to equations 6.22 and 6.23 assume that material is removed by individual particles in a cutting action (similar to type II cutting in Fig. 6.36c), and are valid only for shallow angles of incidence. At high impact angles debris becomes detached only after repeated deformation, and models which take account of this fact are more applicable. The extreme case of erosion at normal incidence by spherical particles, in which cutting can play no role, may be modelled in two ways: by assuming that surface material becomes detached when the accumulated plastic strain reaches a critical value, or by treating the problem as one of low-cycle fatigue caused by the cyclic plastic deformation associated with successive particle impacts. Both approaches, with suitable assumptions about the mechanics of impact, lead to similar conclusions despite the differences in their initial assumptions. The erosion rate should follow the equation

$$E = \frac{K_2 \rho \sigma^{1/2} U^3}{\epsilon_c^2 H^{3/2}} \quad (6.24)$$

where σ is the density of the spherical erosive particles, and ϵ_c is the critical plastic strain at which detachment of wear debris occurs. Equation 6.24 differs from equations 6.22 and 6.23 notably in the higher exponent of the velocity U , and in the fact that two distinct properties of the surface material determine its erosion resistance: not only its hardness H , but also a failure strain ϵ_c , which can be thought of as a measure of the material's ductility under erosion conditions.

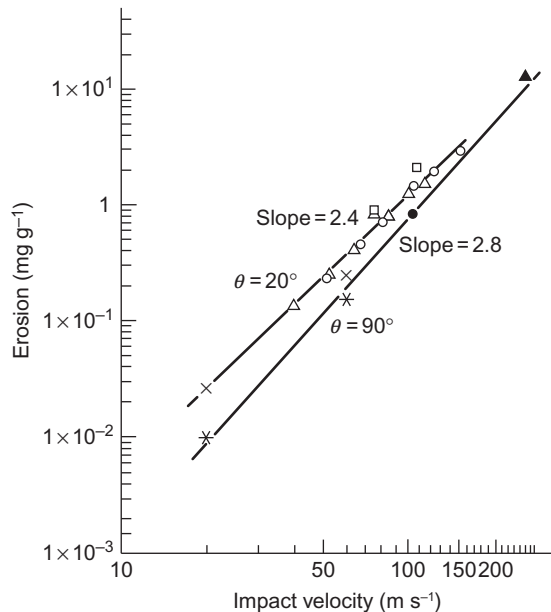
In practice, the erosion of metals does show strong sensitivity to particle impact velocity. The dependence is often expressed in the form

$$E \propto U^n \quad (6.25)$$

and values of n between about 2.3 and 3.0 are commonly reported. The velocity exponent n is nearly always greater than the value of 2.0 predicted by the simpler models, and often lies around 2.4 for ductile metals at impact angles close to that of maximum erosion. There is some suggestion that higher values of n are associated with steeper angles of impact, as shown in Fig. 6.39 for the erosion of copper. This increase may be associated with an increase in the number of particle impacts needed to remove each fragment of wear debris, and to a consequent change in the mechanism from one dominated by single impact events to one better described as fatigue or accumulation of plastic strain.

All the theoretical models outlined above predict that erosion rates due to mechanisms involving plastic deformation should be inversely proportional to the hardness of the material, raised to a power of either 1 (equations 6.21, 6.22 or 6.23) or 3/2 (equation 6.24). Equation 6.24 also predicts a dependence on the ductility of the material. Since these equations also predict that the mass removed per unit mass of particles, E , should be proportional to the density of the material ρ , it is helpful in investigating the dependence on hardness alone to plot the quantity E/ρ against hardness. E/ρ represents the volume of material removed by unit mass of particles, and is a particularly useful measure of erosive wear in the context of design (see Chapter 8). Figure 6.40 shows

Fig. 6.39 Measurements of erosion rate of copper, from a range of different investigations, plotted against impact velocity (log scales) (from Ives, L.K., Ruff, A.W., 1979. In: Adler, W.F. (Ed.), *Erosion: Prevention and Useful Applications*, ASTM Sp. Tech. Pub. 664, pp. 5–35)



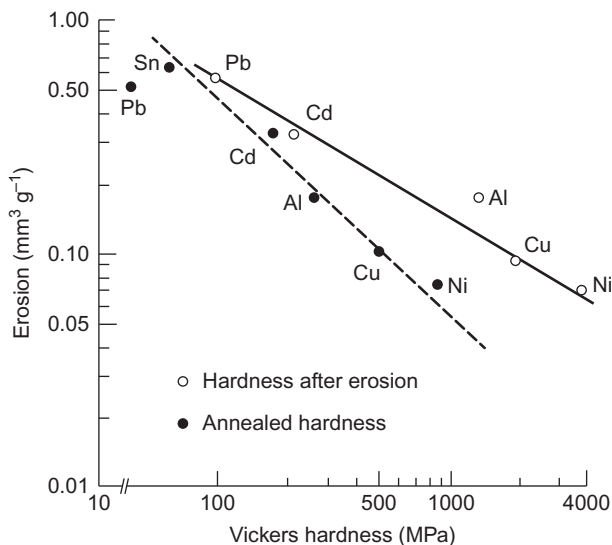


Fig. 6.40 Volume erosion (volume removed per unit mass of erodent particles) for a range of metals, plotted against two different measures of hardness: the hardness of the annealed metal, and the hardness of the surface material after erosion. The erosion tests were performed with 60 mesh silicon carbide particles at 75 m s^{-1} , at an impact angle of 20° (from Sheldon, G.L., 1977. *Trans. ASME: J. Eng. Mater. Technol.* April, 133–137, by permission of the publisher)

the volume erosion, defined in this way, for several pure metals, plotted against their hardness in the annealed state as well as against the hardness of the eroded surfaces. The surface material becomes heavily work-hardened by particle impacts, and this is reflected in the difference between the two sets of hardness values. Although fair correlation is found between volume erosion and $(\text{hardness of the annealed metal})^{-1}$, better correlation is seen with $(\text{hardness of the eroded surface})^{-0.6}$. It is the hardness of the eroded surface, of course, that would be expected to determine the steady-state erosion rate according to equations 6.21–6.24.

Similar results are shown in Fig. 6.41, in which ρ/E , the reciprocal of the volume wear rate and thus a measure of erosion resistance, is plotted against indentation hardness. Once again, the linear behaviour predicted by equations 6.21 to 6.23 is found only for some annealed metals. The dependence for work-hardened metals is not linear, and strikingly, the hardening of steels produces no corresponding increase in wear resistance. The results in this figure can be compared with those for abrasive wear shown in Fig. 6.9. Pure metals generally show quite good correlation between hardness and resistance to wear by both abrasion and erosion, although there are exceptions (molybdenum and tungsten, for example, in Fig. 6.41). Alloys, including steels, show a weaker dependence of abrasion resistance on hardness; against erosive wear they show almost no increase, and in some cases even a decrease, in resistance with increasing hardness.

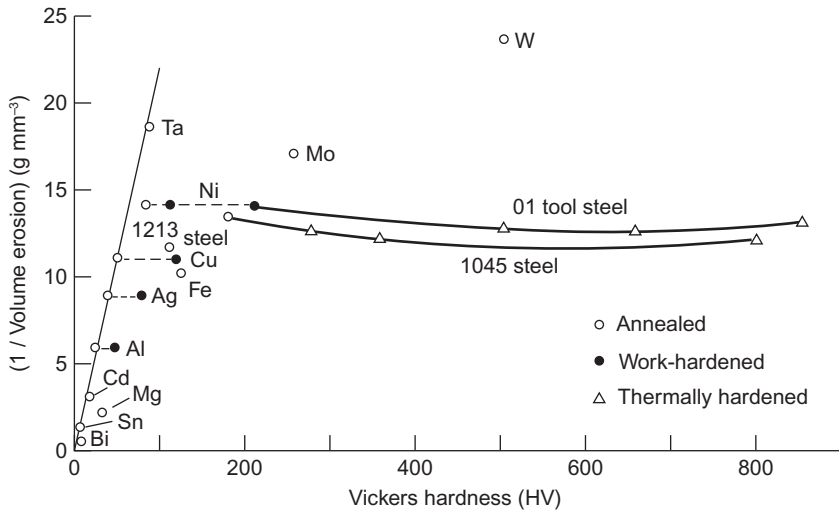


Fig. 6.41 Dependence of erosion resistance ($1/\text{volume erosion in mm}^3 \text{ g}^{-1}$) on Vickers hardness for several pure metals and steels, for various states of work hardening and heat treatment (data from Finnie, I., Wolak, J., Kabil, Y., 1967. *J. Mater.* 2, 682–700)

There are several reasons why bulk hardness is a poor predictor of the erosion resistance of metals, some of which also apply to wear by abrasion and have already been discussed in [Section 6.3.1](#). Better correlation is found between erosion resistance and the hardness of a surface after work-hardening by erosion, as might be expected, but the dependence is weaker than that suggested by theory. The degree to which plastic flow is localized around each particle impact site, which will influence the susceptibility of displaced material to removal, is probably important. This in turn is related to the work-hardening rate of the metal, and also to the ratio between elastic modulus and hardness, as discussed earlier for abrasion. In the case of alloys, the effect of an increase in hardness may well be offset by a decreased strain-hardening rate, which leads to removal of a larger proportion of the metal displaced by each impact.

It is important to remember that the impact of an erosive particle occurs not only on a surface which is heavily strain-hardened, but also over a very short time scale, leading to extremely high strain rates in the deformed material. Some idea of the mean strain rates can be derived from the simple model for normal impact discussed at the beginning of this section. For the normal impact of a sphere on to a rigid-plastic surface, it can be shown that

$$\dot{\epsilon} \approx \frac{2^{2/3} U^{1/2}}{5\pi r} \left(\frac{3H}{2\sigma} \right)^{1/4} \quad (6.26)$$

where $\dot{\epsilon}$ is the mean plastic strain rate, σ is the density of the sphere and r is its radius. [Figure 6.42](#) shows strain rates derived from equation 6.26 for the impact of silica

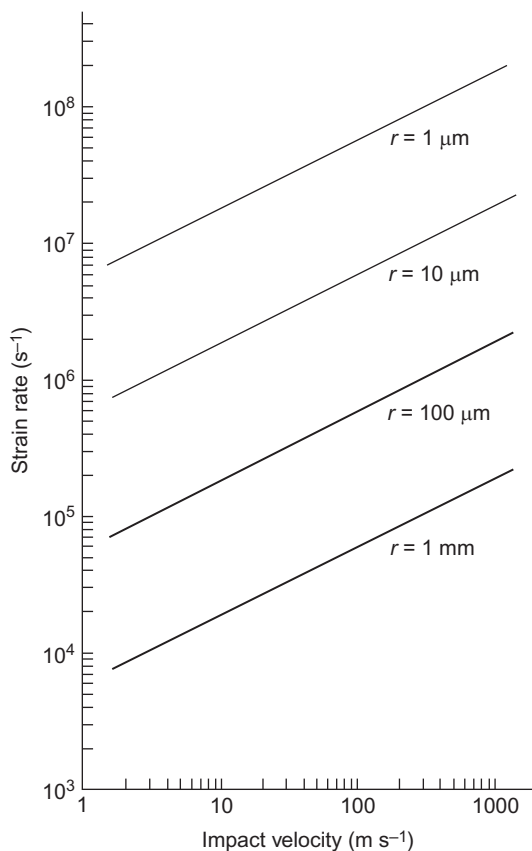


Fig. 6.42 Mean plastic strain rates predicted by equation 6.26 for the impact of silica spheres of different radius on to a mild steel surface

spheres on to mild steel. The deformation caused by erosive particles will typically occur at strain rates of about 10^6 s^{-1} . At such high strain rates, conventional metallurgical strengthening mechanisms such as precipitation, solid solution and grain boundary hardening have much less influence than at lower rates. It is likely that the negligible differences in erosion rate between members of the same alloy family, despite widely different quasi-static hardness, are at least partly due to the very high strain rates involved in erosion.

6.4.2 Erosive wear by brittle fracture

When the impact of an erosive particle causes brittle fracture, material is removed from the surface by the formation and intersection of cracks. As we saw in the case of abrasive wear in Section 6.3.2, even if the mechanism responsible for the detachment of the wear debris is brittle fracture, there will often also be some plastic flow in the material around the point of contact of an angular particle. Although crack patterns

caused by impact differ in detail from those produced by quasi-static indentation, the general morphology shown in Fig. 6.14 is nevertheless observed when the material is homogeneous and brittle and the particle is sufficiently hard and angular. More rounded or softer particles tend to cause purely elastic deformation and conical Hertzian fractures of the form shown in Fig. 5.25.

The extent of cracking due to particle impact is most severe when the impact direction is normal to the surface, and erosion under these conditions is then most rapid. In Fig. 6.35, curve (b) shows the dependence of erosion on impact angle for a typical case where wear occurs by brittle fracture. Erosion falls monotonically with angle away from 90°, in contrast to the behaviour when wear occurs by plastic deformation (curve a). Models for erosion by brittle fracture have therefore been developed mainly for impact at normal incidence. Most of these models assume crack patterns of one of the two types discussed above, and lead to expressions of the form:

$$E \propto r^m U^n \quad (6.27)$$

where r is the radius of the particle.

Models for material removal by the formation and intersection of Hertzian cone cracks suggest that the values of the exponents m and n should be related to the Weibull constants for the material, which describe the statistical distribution of fracture stress among samples or areas of the surface. Such models lead to values of m typically around 1, and a velocity exponent n in the range from 2.6 to 3.0.

For the angular particles which are more commonly encountered in practice, models for erosion by elastic-plastic fracture of the types illustrated in Fig. 6.14 are more applicable. Material is assumed to be removed by the intersection of lateral cracks with each other and with the surface. One model estimates the contact force exerted on the particle during impact by assuming that the pressure resisting penetration is the quasi-static hardness of the surface (as in the model leading to equation 6.21). The contact force is then used in a semi-empirical analysis to predict the extent and depth of the lateral cracks formed, and hence the volume of material removed by a single impact (a cylinder bounded by the surface and the lateral cracks). This predicts a volume erosion (volume removed per unit mass of erodent particles, E/ρ) given by

$$\frac{E}{\rho} \propto r^{0.7} U^{2.4} \frac{\sigma^{0.2} H^{0.1}}{K_c^{1.3}} \quad (6.28)$$

The most important material property determining erosion resistance is seen to be fracture toughness (K_c), with hardness (H) being much less significant; similar predictions are made by other models, with the role of fracture toughness being dominant. In contrast to the models for erosion by plastic deformation, there is a significant dependence of erosion E on particle size, r .

In experimental studies of the erosion of brittle materials the mass or volume loss from the surface depends linearly on the total mass of erodent particles, as shown in Fig. 6.34 and discussed in Section 6.4.1. Some incubation behaviour may be observed

before the steady-state regime. Velocity exponents are commonly in the range from 2 to 4, consistent with the predictions of the simple theories for material removal by cone or lateral cracking. Substantial dependence of erosion on particle size is also seen, with exponents for r commonly lying between 0.7 and 1. There is considerable variation of erosion between different materials, and attempts have been made to find a correlation with values of the fracture toughness K_c and hardness H measured in conventional mechanical tests. The results generally do not conform well with the predictions of any one theory. Figure 6.43, for example, shows experimental data for the erosion of a range of brittle materials by angular silicon carbide particles, plotted against the predictions of equation 6.28. Although there is a fair correlation between the measured erosion and the relevant combination of K_c and H , the slope of the line is greater than the value of 1.0 predicted theoretically. In this case better correlation is found with the following expression generated by multiple regression analysis, which involves a rather greater dependence on both hardness and toughness as well as a slightly higher velocity exponent:

$$\frac{E}{\rho} = r^{0.9} U^{2.8} \frac{\sigma^{0.4} H^{0.5}}{K_c^{1.9}} \quad (6.29)$$

The models so far discussed assume that the material being eroded is homogeneous and isotropic, but some important materials which erode by brittle mechanisms are heterogeneous on the scale of the damage caused by the erosive particles. Examples are cemented carbides, other composite materials with ceramic or some polymeric matrices, two-phase ceramics such as reaction-bonded silicon carbide, and ceramic materials containing significant numbers of microcracks, pores or inclusions, or weak grain boundaries. So far, no satisfactory predictive models for the erosion of these materials have been developed, and even the mechanisms of material removal are in many cases

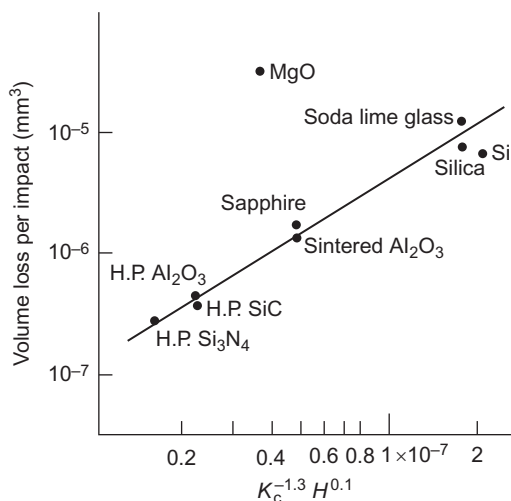


Fig. 6.43 Comparison of the measured values of erosion rate (expressed in terms of volume loss per impact) for a range of ceramic materials eroded at normal incidence by 150 μm silicon carbide particles at 63 $m s^{-1}$ with predictions from quasi-static theory (equation 6.28) (from Wiederhorn, S.M., Hockey, B.J., 1983. *J. Mater. Sci.* 18, 766–780, with permission of Springer)

unclear. The models based on lateral cracking fail to provide accurate predictions of erosion rate. In Fig. 6.43 for example, the inconsistency between the erosion rate of polycrystalline MgO and that of the other materials is associated with the nature of the impact damage in the MgO samples, which cracked readily along the grain boundaries and did not form lateral cracks.

6.4.3 Erosion of engineering materials

We have seen that the dependence of erosion, defined by equation 6.20, on impact angle varies with the mechanism by which wear occurs. If plastic deformation dominates, then maximum wear occurs at a shallow angle (Fig. 6.35 curve a), whereas erosion by brittle fracture is most rapid for normal incidence (Fig. 6.35 curve b). These two types of behaviour are so characteristic that they are often described as *ductile* and *brittle* erosion behaviour. It is important to note, however, that the angular dependence of erosion is not a characteristic of the material alone, but depends also on the conditions of erosion. For example, although most metals eroded by hard angular particles do show typical 'ductile' behaviour (Fig. 6.35a), erosion by spherical particles, even of a ductile metal such as mild steel, can lead to apparently 'brittle' angular dependence, although wear still occurs by purely plastic processes. Alloys of high hardness and low ductility may also show their maximum erosion rate at normal incidence; an example, for a low alloy bearing steel after different heat treatments, is shown in Fig. 6.44. At comparatively low hardness and high ductility, the steel shows characteristic 'ductile' behaviour, while with high hardness it shows apparently 'brittle' behaviour, although microscopic examination shows no sign of brittle fracture. The occurrence of brittle fracture cannot safely be deduced from the angle-dependence of erosion alone.

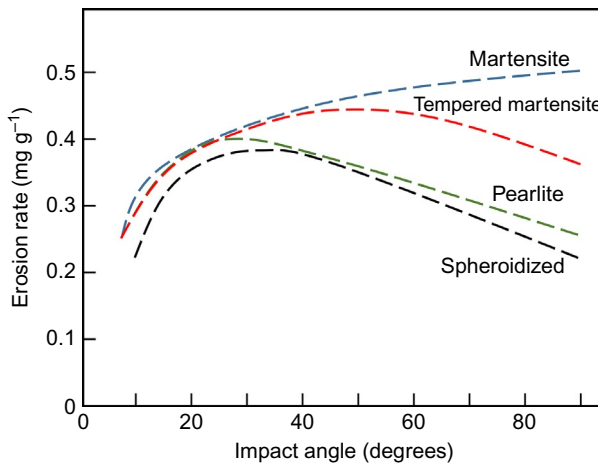


Fig. 6.44 Erosion rate of AISI 1078 steel samples (0.8% C) with different microstructures, as a function of impact angle. Alumina particles were used, mean size $36\ \mu\text{m}$ at $99\ \text{m s}^{-1}$ (from Zum Gahr, K.-H., 1987. *Microstructure and Wear of Materials*. Elsevier; data from McCabe, L.P., Sargent, G.A., Conrad, H., 1985. *Wear* 105, 257–277)

Materials conventionally thought of as brittle, such as ceramics, glasses and some polymers, usually suffer peak erosion at normal incidence. However, on erosion by very small particles, nominally brittle materials can show truly ductile behaviour, with material being removed only by plastic deformation and maximum erosion occurring at a shallow angle. This transition from truly brittle to ductile mechanisms is illustrated in Fig. 6.45 for the erosion of soda-lime glass by silicon carbide particles. The reason for this behaviour is exactly the same as that discussed in Section 6.3.2 in the context of abrasive wear. Fracture occurs only when the indentation size produced by each particle exceeds a certain threshold. The indentation size is determined by the particle size and the impact conditions; on reducing the particle size or the impact velocity, the impact events remain plastic and the material then erodes in the same way as a ductile metal. As we saw in Section 6.3.2, the quantity H/K_c provides a measure of the brittleness of a material, and this ratio has dimensions $(\text{length})^{-1/2}$. The dimensionless group K_c^2/rH^2 , where r is the radius of the erodent particle, provides a useful guide to the nature of the dominant erosion mechanism. For low values of K_c^2/rH^2 the material response will be dominated by fracture, while for high values extensive plastic flow will occur.

Metallic alloys show remarkably little difference in erosion rate between members of the same alloy system, despite substantial differences in hardness. Data for two steels were shown in Fig. 6.41; these observations have been confirmed in numerous other studies. Figure 6.46, for example, shows erosion rates under the same conditions for ranges of ferritic, martensitic and austenitic steels, aluminium alloys, nickel alloys and copper alloys. The lack of variation within each alloy system is striking, as is the small range of variation between the different alloy systems. The general conclusion that alloy selection can be of little benefit in reducing erosive wear is true provided the hardness of the alloy is less than that of the erodent particles themselves. If an alloy

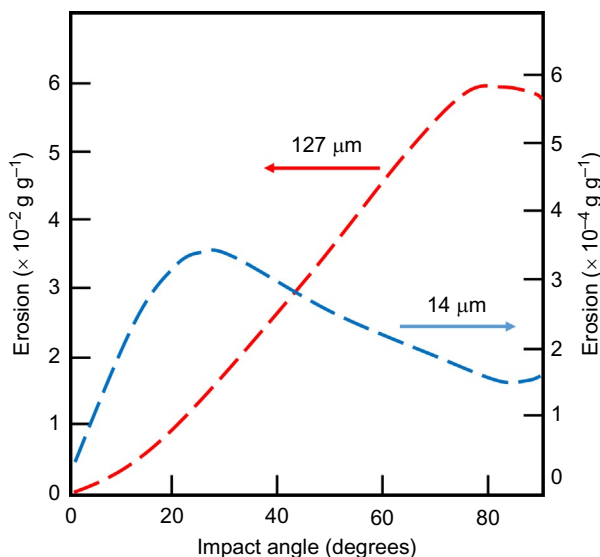


Fig. 6.45 Dependence of erosion rate on impact angle for soda lime glass eroded by 14 μm and 127 μm silicon carbide particles at 136 m s^{-1} (data from Sheldon, G.L., 1970. Trans. ASME D: J. Basic Eng. 92(3), 619–626)

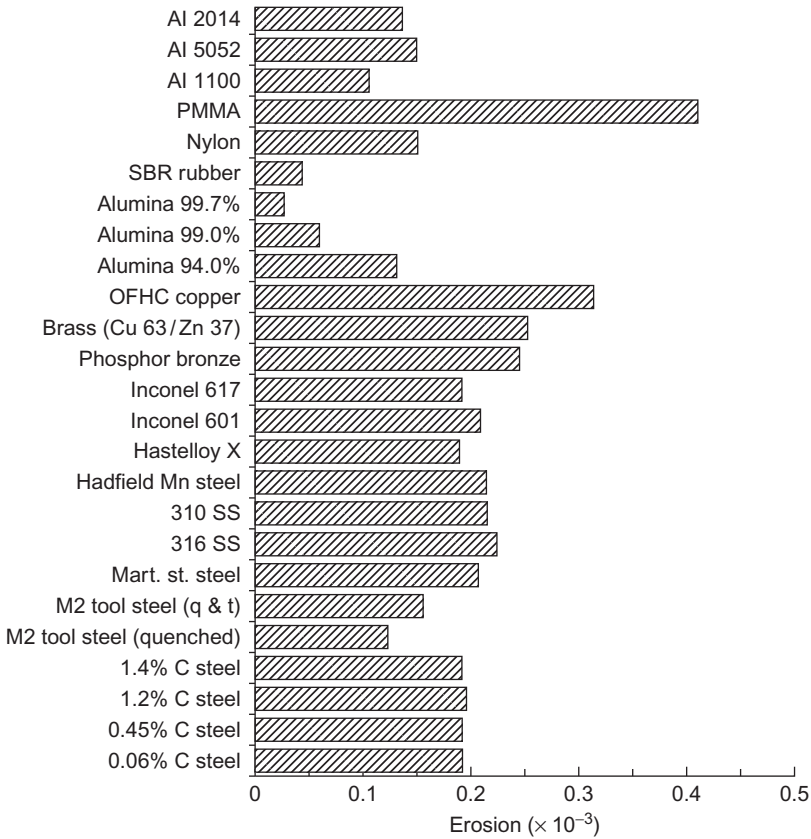


Fig. 6.46 Erosion rates (mass loss per unit mass of erodent) for a range of materials exposed to olivine sand particles (Mohs hardness 6.5–7), 350–500 μm in size, at 45° impact angle and 66 m s^{-1} (data from Söderberg, S., et al. 1981. Tribol. Int. December, 333–343)

harder than the erodent can be found, erosion rates may be reduced by the effect discussed in Section 6.2.1 and illustrated in Fig. 6.2. This effect probably explains the relatively low erosion rate of the quenched tool steel shown in Fig. 6.46.

Ceramic materials show considerably greater variation in erosion resistance than metals. We have seen that erosion rates predicted theoretically from measured values of hardness and fracture toughness may be grossly in error if the mechanism of erosion is different from that postulated in the model. Most ceramics are harder than quartz, but nevertheless show significant erosion rates when eroded by this material. Data for the three different polycrystalline sintered aluminas in Fig. 6.46 typify the range of behaviour for these materials. The worst sample showed a rather lower mass loss than the steels, but the corresponding volume loss (because of its lower density) was actually higher. The best alumina ceramic gave a volume loss less than 20% of that of the carbon steels. In these tests the erosion resistance of glass-bonded sintered aluminas

correlated with alumina content, but this rule does not hold generally; in other work a material with 88% alumina content has been found to give higher erosion resistance than one with 97.5%. More important, probably, is the grain size, the nature and strength of the grain boundary phase and the proportion and distribution of porosity.

Binder phase content has been found to be important in the erosion resistance of other ceramics and cemented carbides. For example, reaction-bonded silicon carbide containing some 10% free silicon is eroded by alumina particles some eight or nine times as fast as hot-pressed silicon carbide of >99% theoretical density. Similar behaviour is seen in silicon nitride: reaction-bonded silicon nitride containing residual porosity may erode at 10 times the rate of pressureless sintered and hot-pressed specimens. Although broad trends can be discerned it is not possible to predict the erosion resistance of ceramic materials with any precision. Materials of nominally the same composition and fabrication method produced by different manufacturers via the same route may show completely different behaviour in an erosion test.

Cemented carbides have been widely studied. As for the case of abrasive wear (Section 6.3.3), the dominant mechanism of erosion depends on the scale of individual particle contacts. If the erosive particles are small enough, wear can occur by preferential erosion of the metallic binder phase, leading to undercutting and eventual removal of intact carbide grains. If the binder regions are too small to allow this type of localized attack, then overall erosion of the carbide and binder occurs, usually by ductile mechanisms if the erodent particles are small. The combination of these mechanisms leads to a strong and monotonic dependence of erosion rate on binder content, as illustrated in Fig. 6.47 for the erosion of WC–Co cemented carbides by a fine silica slurry (with a particle size down to 1 μm). With larger erodent particles, fracture of the brittle carbide grains can become important, and can lead to a peak in the erosion rate at an intermediate value of binder content. Figure 6.48 illustrates this behaviour for WC–Co cemented carbides eroded by 100 μm silicon carbide particles. At shallow

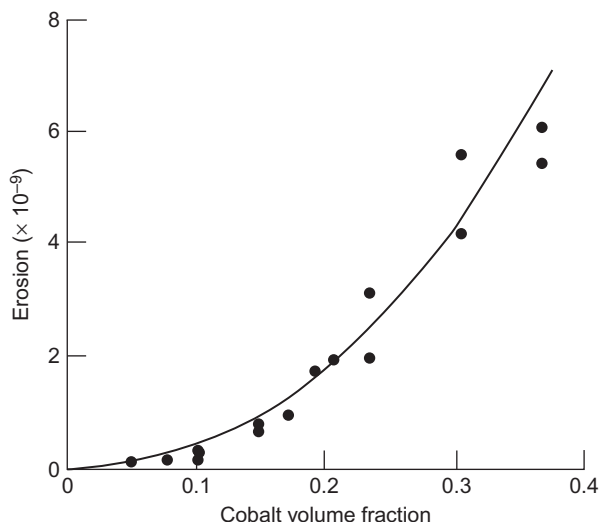


Fig. 6.47 Dependence of erosion rate on volume fraction of cobalt binder phase for a series of tungsten carbide/cobalt cermets eroded by a slurry of 10 μm silica particles in oil, at 133 m s^{-1} and a nominal impingement angle of 90° (from Shetty, D.K., Wright, I.G., Clauer, A.H., 1987. *Wear* 114, 1–18)

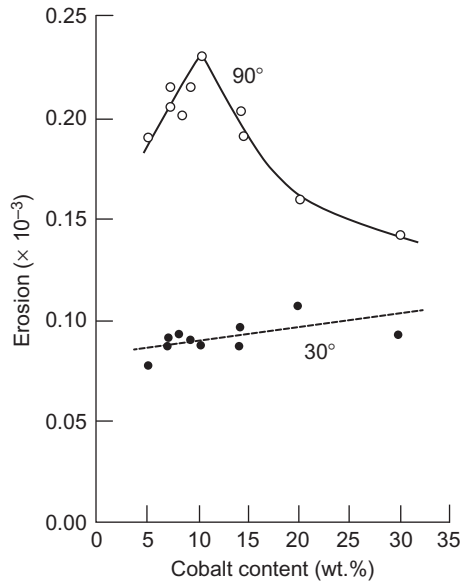


Fig. 6.48 Dependence of erosion rate on volume fraction of cobalt binder phase for a series of tungsten carbide/cobalt cermets eroded by airborne $100\text{ }\mu\text{m}$ silicon carbide particles at 40 m s^{-1} and at impact angles of 30° and 90° (from Pennefather, R.C., et al. 1988. *Mater. Sci. Eng. A* 105/106, 389–394)

angles of incidence, a general trend of increasing erosion rate with binder content is seen, although not so strong as that shown in Fig. 6.47. For normal incidence, however, a sharp peak in erosion rate is seen at a cobalt binder content of $\sim 10\text{ wt.}\%$. Under these conditions the dominant mechanism of erosion is fragmentation of the carbide grains, which is most prevalent at this level of binder content.

Polymers, including plastics materials and elastomers, show very wide variation in erosion resistance. Three polymers are included in Fig. 6.46. Their erosion rates, in terms of mass loss under the conditions of the test, range from more than twice that of the steels (in the case of PMMA) to only one quarter (styrene-butadiene rubber). The mechanisms and angular dependence of erosion vary correspondingly widely. Ductile thermoplastics, such as nylon, acetal, polycarbonate, polypropylene and polyvinylchloride erode by mechanisms involving plastic deformation, as for metals, and experience maximum wear at shallow impact angles. More brittle thermoplastics and thermosets such as polystyrene, PMMA and some epoxy resins erode by brittle fracture, and show maximum erosion rates at high angles. As for other materials eroding by brittle fracture, a transition may occur with sufficiently small erosive particles to wear by plastic mechanisms.

Some elastomers show excellent resistance to erosion at normal incidence and moderate velocities, but they show greater sensitivity to both impact angle and velocity than either conventionally ductile or brittle materials. Figure 6.49 illustrates the

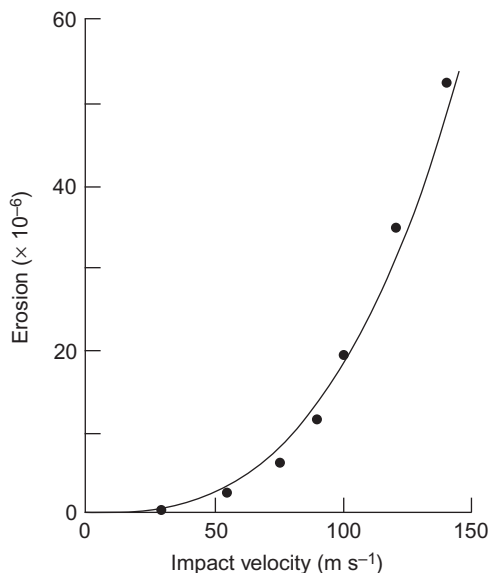


Fig. 6.49 The dependence of erosion (mass removed per unit mass of erodent) on impact velocity for natural rubber, eroded by 120 μm silica sand particles at 30° impact angle (from Arnold, J.C., Hutchings, I.M., 1990. *Wear* 138, 33–46)

dependence of erosion on impact velocity for a natural rubber compound, which shows a very steep rise in erosion above about 50 m s^{-1} . For this material the erosion at 90° impact angle is around one quarter that at 30°. The mechanisms of erosion in rubbers are similar to those of abrasive wear (see [Section 5.11.2](#)), with tearing, crack propagation and fatigue all playing a role. Attempts to correlate erosion rate with tensile strength are generally unsuccessful, nor does the Ratner–Lancaster correlation appear to be useful. High rebound resilience, however, (defined as the relative height of rebound of a hard sphere striking the rubber surface) has been found to be associated with good erosion resistance in a range of unfilled rubber compounds, and in some systems a low elastic modulus also leads to a low erosion rate.

6.4.4 Erosion–corrosion

In several important practical applications, materials are exposed to particle impact while also being subject to high-temperature oxidation or aqueous corrosion. These conditions then lead to material loss by processes known as *erosion–oxidation* or *erosion–corrosion*. Examples include the wear of boiler tubes and combustor components in particle-laden flue gases, or of turbine blades in a gas turbine engine, where erosion is combined with high-temperature oxidation, and the wear of the impellers and casings of pumps handling corrosive slurries, for example in the mining industries. Other instances occur in marine engineering, the oil and gas industry, and thermal power generation. As we have also seen in the context of both sliding wear ([Section 5.5.2](#)) and abrasive wear ([Section 6.3.4](#)), in erosion–corrosion the overall material loss rate can be enhanced by the mechanical removal of a surface protective film which leads to higher corrosion, giving a significant synergistic effect (see [Appendix B](#)).

The mechanisms involved in these types of tribo-corrosion can, however, also be more complex, as shown by the following examples.

Figure 6.50 shows how the wastage rate of a low alloy steel varied with temperature under conditions of combined erosion and high temperature oxidation, for samples moved through a heated bed of angular alumina particles fluidized by an air flow. These experiments were carried out to simulate the conditions in a fluidized bed coal combustor. The wastage in each case shows a characteristic peak at an intermediate temperature. At lower temperatures the dominant mechanism is the removal of oxide by particle impact. The oxidation rate, and hence the wastage rate, increases with temperature until at 400–500°C, depending on the severity of the erosive action, the oxide becomes thick enough to become mechanically protective and the wastage rate then decreases with further increase in temperature. An increase in the ductility of the oxide with temperature probably also plays a role.

In the example of the erosion–corrosion of aluminium by a slurry of silica sand particles in aqueous sodium chloride solution shown in Fig. 6.51, the total wastage rate T at all angles of slurry jet impingement was considerably greater (by up to a factor of two times) than the sum of the rates of pure erosion W_o (measured with the slurry with no chloride addition) and of pure corrosion C_o (measured with a jet of chloride solution containing no particles). Electrochemical measurements showed that the increase in the corrosion rate due to erosion was much less than the total magnitude of the synergistic effect, indicating that the effect mainly arises from the effect of corrosion on erosion: i.e., in terms of the variables defined in Appendix B, $\Delta W_c > \Delta C_w$. Detailed examination of the mechanism of material removal, which in this case involved the detachment of heavily-deformed flakes of metal from the

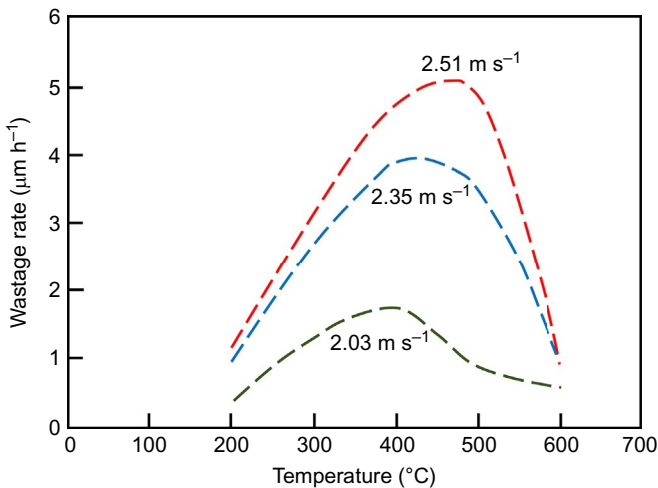


Fig. 6.50 Variation of wastage rate with temperature for a 2.25 Cr–1 Mo steel exposed to erosion–oxidation by angular alumina particles in air in a fluidized bed. The velocities represent the average speed of relative motion between the steel surface and the particles (from Rogers, P.M., et al., 1995. *Wear* 186–187, 306–315)

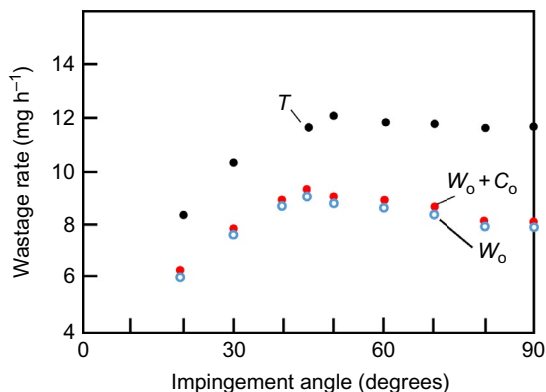


Fig. 6.51 Erosion–corrosion of commercially pure aluminium in a slurry of silica particles in aqueous sodium chloride solution in a jet impingement rig. Material removal rate is plotted against impingement angle. T = total wastage rate; W_0 = erosion rate in the absence of chloride ions; C_0 = corrosion rate with no particles (from Li, Y., Burstein, G.T., Hutchings, I.M., 1995. *Wear* 186–187, 515–522)

surface, suggested that the higher erosion rate in the presence of corrosion probably resulted from stress-corrosion cracking that enhanced the detachment of the flakes. Other mechanisms that have been proposed to explain an enhancement of erosion by corrosion include the removal of a work-hardened layer by corrosion, or a corrosion-induced increase in surface roughness, but both of these mechanisms could be discounted in this example.

6.4.5 Testing methods for erosive wear

Laboratory-scale erosion testing is performed for several reasons: to provide data on absolute and relative wear rates under specific conditions, to validate theoretical models, and to study the mechanisms of wear. The first of these objectives gives information of direct value to the design engineer, while the others are of more value in improving our understanding of erosive wear. For the results of a test to be useful and reproducible, the impact conditions (particle velocity, flux and impact angle) must be accurately defined, and the particles and material being tested should be well characterized.

To establish a value for erosion rate it is essential to measure the mass loss from the specimen (usually determined by accurate weighing) after exposure to several different amounts of erodent, as discussed in [Appendix C](#). Incubation behaviour (see [Section 6.4.1](#)) can lead to misleading results in erosion testing if it is not recognized and allowed for. It is dangerous, for example, to test a material for a single time period, or with a single quantity of erodent, and consider the resultant mass or volume loss to represent a trustworthy measure of erosion. Test procedures should be used that will identify any incubation period, and provide an accurate measurement of the rate of wear in the steady-state regime after incubation.

The erodent particles used in a laboratory test should be selected carefully if the results are to be useful in a practical application. Not only should the particle size be the same as that in the practical case, but the particle material and angularity should also be the same. As we saw for abrasive wear ([Section 6.3](#)), the material of the particles can influence the wear mechanisms as well as the rate of wear. In an extreme

case even the ranking of materials in terms of relative erosion rate can be changed by the use of different erodent particles.

Methods commonly used for laboratory erosion testing can be divided into those in which the particles are accelerated in a gas or liquid stream, and those where circular motion is used to achieve the impact velocity. Figure 6.52 shows schematic diagrams of four types of testing method. In the method shown at (a), particles are accelerated in a fluid stream along a nozzle which may be either parallel-sided or of more complex shape (e.g., a converging-diverging nozzle). They then strike the target material which is held some distance from the end of the nozzle at a fixed angle. This test, often called the *jet impingement* or *gas-blast* method, can be used with gas-borne particles (usually with air as the carrier gas) or with a liquid slurry, often with oil or water as the fluid medium. The few standard methods for erosion testing by airborne particles (e.g., ASTM G76) employ gas-blast methods. Nozzles of widely different dimensions can be used for different applications: practical tests have covered the range from about 1 to 50 mm diameter. A small nozzle results in a correspondingly small area of wear on the specimen, but requires only low feed rates of fluid and particles. A larger nozzle is more suitable for testing materials with heterogeneous microstructures of large scale, such as some coarse-grained ceramic materials and composites, but demands a correspondingly greater supply of fluid and erodent. Impact velocities in the range from 10 to 100 m s⁻¹ are readily achievable, although with long acceleration distances in the nozzle higher speeds are possible for gas-borne particles. In most tests of this type, particles are used only once, obviating any problems of particle degradation by impact. With very large nozzles, however, the quantity of erodent required may make recirculation of the particles essential, and care must then be taken to ensure that the

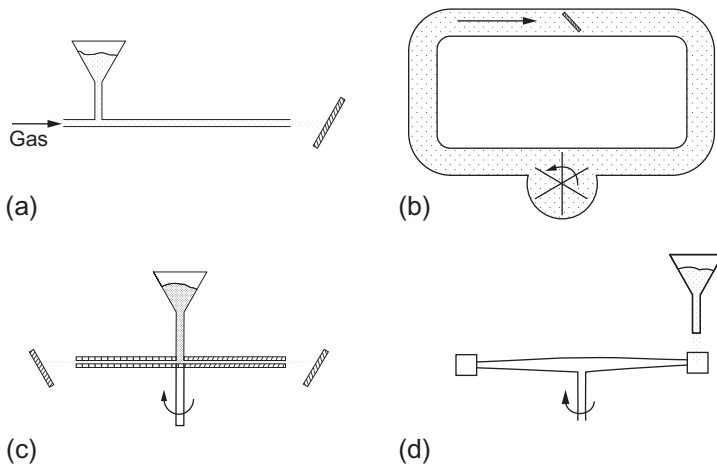


Fig. 6.52 Schematic illustration of four different methods of exposing specimens to erosive wear in laboratory tests: (a) jet impingement method; (b) recirculating loop; (c) centrifugal accelerator; (d) whirling arm rig

erosion rate does not change during the test due to fragmentation or attrition of the erodent particles which changes their size or shape.

Problems of particle degradation also occur in the *pumped loop* or *recirculating loop* test shown in Fig. 6.52(b). Here a two-phase flow of particles and fluid (gas or liquid) is driven around a loop of pipework. The method is valuable for establishing the wear rates of pipework components such as bends or valves in pneumatic or hydraulic conveying systems. It can also be used to examine the behaviour of materials by completely immersing specimens in the flow. Although the schematic diagram suggests that the flow passes through the pump, in some practical designs wear in the pump is avoided by separating the particles from the fluid and then recombining them after the fluid has been pumped. In some cases, the issue of particle size is avoided by passing them only once through the system, but this may involve the use of a large amount of erodent material.

The method shown in Fig. 6.52(c), sometimes called a *centrifugal accelerator*, uses circular motion to generate a continuous stream of particles; it is usually used in air or vacuum. The erodent particles are fed into the centre of the rotor, and move outwards along radial tubes or channels, leaving the rotor at a speed governed by the peripheral speed of the rotor. Their motion is not tangential to the rotor because by the time they reach the rim they have acquired some radial velocity, although in some designs steps are taken to reduce the radial velocity to negligible proportions. Stationary specimens are arranged around the rim of the rotor, and the method can be used to compare the erosion behaviour of up to twenty or so different specimens simultaneously.

In the apparatus shown in Fig. 6.52(d), two specimens on the ends of a balanced rotor move at high speed through a slowly falling stream of particles, striking them at the peripheral speed of the rotor and at an angle determined by the orientation of the specimens. This *whirling arm* test is usually carried out in vacuum to eliminate aerodynamic effects on the particles and reduce the power needed to turn the rotor. A similar principle is used in the *slurry-pot* test in which a rotor carrying specimens is immersed in a tank containing a slurry of liquid and particles; in this type of test, however, the impact angle and velocity can be difficult to define, and it is more suited to comparison of materials than to absolute measurement of erosion rates.

As we have seen in earlier sections, the particle impact velocity is the most important variable that influences erosion rate. In erosion testing it is particularly important that the velocity be held constant, and preferably known accurately. Uncertainty in impact velocity is the source of the greatest error in erosion tests, which even when repeated under nominally identical conditions with the same materials in different laboratories can give significantly different results.

In slurry erosion tests, it is sometimes reasonable to assume that the particle velocity is the same as that of the liquid, which can often be measured by simple methods. In the jet impingement test, for example, the volume of slurry leaving the nozzle of known dimensions in a fixed time provides a direct measure of the exit velocity of the slurry jet. In most test methods for airborne particles, however, independent measurement of the particle velocity is necessary. Computational models are available to predict particle velocities in gas-blast erosion testing, but in order to make accurate predictions they must take account not only of the size, density and drag coefficient (which will depend

on the shape) of the particles, but also of the nozzle wall friction and the effects of particle-wall and even particle-particle interactions. The plume of particles leaving the end of a cylindrical nozzle contains some particles that diverge quite markedly from the nozzle axis, showing that impacts between particles and the nozzle wall can be significant. Theoretical models for particle velocity are therefore no substitute for accurate measurements under realistic operating conditions.

Multiple flash photography, particle image velocimetry (PIV) and laser Doppler velocimetry (LDV) are all potentially useful methods for velocity measurement. A simple and cheap mechanical method has been used in the past in which the stream of particles passes through a slit in a disc mounted on a rotating shaft, and forms an erosion scar on a second disc, also fixed to the shaft. An estimate of the mean particle velocity can be made by measuring the displacement of the erosion scar from the scar formed when the discs are not rotating, or rotating in the opposite sense. This method is not, however, as accurate as the optical methods, nor does it provide quantitative information on the velocity distribution or the distribution of particle trajectories in the flow.

An important variable that is often ignored in erosion testing is the *flux* of particles striking the surface, measured in terms of the number of particles or their total mass impinging on unit area per unit time. The time interval between successive impacts at any point on the specimen will depend on the flux. In practical cases of erosive wear the particle flux may be low, but a laboratory test may nevertheless be carried out at a higher flux in order to obtain a measurable amount of wear in a reasonable time. If the results of an *accelerated test* like this are to be applied to predict erosion under practical conditions, it must be established that the effect of particle flux on erosion is not significant. For most materials at moderate fluxes this is true. However, at very high fluxes the erosion rate may be affected by interactions between particles striking the surface and those rebounding from it, or possibly by thermal effects as the surface heats up under a high rate of kinetic energy dissipation. The latter effect depends also upon the particle impact velocity, since this together with the mass flux controls the impinging kinetic power density as follows:

$$P_D = \frac{1}{2}JU^2 \quad (6.30)$$

where P_D is the power density (power per unit area), J is the mass flux and U is the impact velocity.

The conditions suggested in the ASTM standard test for gas-blast erosion testing (ASTM G76) result in a mass flux of almost $20 \text{ kg m}^{-2} \text{ s}^{-1}$ in the accelerating nozzle but only $2 \text{ kg m}^{-2} \text{ s}^{-1}$ at the specimen surface (because the particle stream expands between the nozzle tip and the specimen). With the impact velocity of 30 m s^{-1} suggested in the standard, this results in a kinetic power density of 900 W m^{-2} . However, much greater mass fluxes and impact velocities are sometimes employed, leading to kinetic power densities in excess of 100 kW m^{-2} . (This is, however, still very small compared with the power densities commonly employed in thermal hardening of steels—see [Section 7.2.1](#)). In ductile materials (such as metals and polymers),

most of this power is dissipated by plastic deformation in a plastic zone of depth x_p . This depth will be approximately the same as the plastic indentation depth, which can be estimated from the simple model for plastic impact of a spherical particle described by equation 6.15 to give:

$$x_p = RU \sqrt{\frac{2\sigma}{3H}} \quad (6.31)$$

where R and σ are the particle radius and density respectively and H is the hardness of the material being eroded. In practice, metals with high thermal conductivities experience no significant bulk temperature rise, but in polymers (whose thermal conductivities are typically between a hundred and a thousand times less than for metals), heating can occur which significantly changes their properties. The gas stream in a gas blast test may provide some convective cooling to the specimen surface which is absent with other methods such as the centrifugal accelerator, adding further possible complexity to the interpretation of the test results.

An additional influence of particle flux on erosion rate may occur when chemical reactions take place on the eroded surface. These reactions may, but need not, be conventional corrosion processes. For example, the erosion of natural rubber depends strongly on particle flux at low fluxes, probably due to local oxidative attack of the material at the particle impact sites; this dependence can be eliminated by the incorporation of an anti-oxidant into the rubber formulation. For all these reasons, high particle fluxes should be avoided in laboratory erosion tests, or at least, their implications should be fully appreciated.

All the erosion test methods illustrated in Fig. 6.52 can in principle also be adapted to study erosion–corrosion. High temperature oxidation can be combined with erosion in all of these methods, although the practicalities of heating the samples and erodent, as well as heating and containing the gas environment, add considerably to the complexity and cost of the apparatus. Erosion–corrosion tests in liquid slurries, usually but not always water-based, can be carried out with apparatus involving either jet impingement or rotating specimens (a ‘slurry pot’ test). In aqueous slurries the separate contributions to material loss from mechanical (erosion) and chemical (corrosion) origins can be deduced from separate experiments carried out with an inert slurry or under conditions where corrosion of the specimen is electrochemically inhibited, as discussed in Section 6.3.5 and Appendix B.

SOURCES AND FURTHER READING

- Adler, W.F. (Ed.), 1979. *Erosion: Prevention and Useful Applications*. American Society for Testing and Materials. ASTM STP 664.
- Briscoe, B.J., Adams, M.J. (Eds.), 1987. *Tribology in Particulate Technology*. Adam Hilger.
- Friedrich, K. (Ed.), 1986. *Friction and Wear of Polymer Composites*, Composite Materials Series, vol. 1. Elsevier.
- Landolt, D., Mischler, S., 2011. In: *Tribocorrosion of Passive Metals and Coatings*. Woodhead Publishing.

- Larsen-Basse, J., 1990. The role of microstructure and mechanical properties in abrasion. *Scripta Met. Mater.* 24, 821–826.
- Miller, J.E., Schmidt, F. (Eds.), 1987. *Slurry Erosion: Uses, Applications and Test Methods*. ASTM. Sp Tech. Pub. 946.
- Misra, A., Finnie, I., 1982. A review of the abrasive wear of metals. *Trans. ASME: J. Eng. Mater Technol.* 104, 94–101.
- Preece, C.M. (Ed.), 1979. *Erosion, Treatise on Materials Science and Technology*, vol. 16. Academic Press.
- Rabinowicz, E., 1995. *Friction and Wear of Materials*, second ed. John Wiley.
- Rigney, D.A. (Ed.), 1981. *Fundamentals of Friction and Wear of Materials*. ASM.
- Wood, R.J.K., 2007. Tribo-corrosion of coatings: a review. *J. Phys. D Appl. Phys.* 40, 5502–5521.
- Zum Gahr, K.-H., 1987. *Microstructure and Wear of Materials*, Tribology Series No. 10. Elsevier.

QUESTIONS FOR CHAPTER 6

Question 6.1

- (a) Assume that Fig. 6.7(b) represents a cross-section through a long prismatic wedge-shaped particle indenting a metal surface and moving to the right. Hence derive an expression analogous to equation 6.4 for q , the volume of material removed by abrasive wear per unit distance moved by a wedge-shaped particle, in terms of the load on the particle w , the indentation hardness of the metal H , the fraction of the material displaced from the groove that is removed by wear η , and the semi-angle of the wedge α .
- (b) Use a similar approach to that used in part (a) to analyze the process of abrasive wear of a ductile material by a spherical particle, and derive an expression for q in terms of w , H , η and the radius of the particle R . Discuss the reasons why angular particles cause higher wear rates for ductile metals in two-body (sliding) abrasion than rounded particles.

Question 6.2

- (a) Explain why the abrasive wear rate of a brittle material by hard angular particles would be expected to be considerably greater than the wear rate predicted by equation 6.4.
- (b) In an abrasive wear test carried out with an alumina ceramic pin sliding against silicon carbide abrasive paper in a pin-on-disc geometry, with the pin tracing a spiral track on the disc, the mass lost from the pin by wear over a period of 1 minute was 0.2 g. The rotational speed of the disc was 60 r.p.m.

Use an appropriate model to estimate the mass which would be lost from a pin made from soda-lime glass when abraded for the same time, under exactly the same test conditions. State your assumptions clearly.

The properties of the alumina and soda-lime glass were as follows:

	Density ρ (kg m^{-3})	Hardness H (HV)	Fracture toughness K_{IC} ($\text{MPa m}^{1/2}$)	Young's modulus E (GPa)
Soda-lime glass	2500	540	0.75	75
Alumina ceramic	3900	1600	4.0	400

- (c) Discuss how the relative wear rates of these two materials might differ: (i) if the pin moved repeatedly on the same circular track instead of on a spiral track; and (ii) if the abrasive particles were garnet rather than silicon carbide.

Question 6.3

In the dry-sand rubber wheel abrasion test illustrated in Fig. 6.27(d), loose abrasive particles are fed into the gap between a fixed plane specimen and a rotating cylindrical

wheel with a compliant rubber rim. In a particular test arrangement the effective length of the wear zone (over which the particles are in contact with both the specimen and the wheel) is 20 mm. The wheel is 230 mm in diameter and 12.5 mm wide and is narrower than the specimen. The silica abrasive particles have a mean diameter of 200 μm . The density of silica is 2700 kg m^{-3} .

- (a) By treating the particles as uniform spheres close-packed in a single layer in the wear zone, estimate the maximum number of particles that can be present in the wear zone at any time. Hence, if the wheel is rotating at 200 r.p.m., estimate the maximum mass flow rate of abrasive $\dot{M}$ (in kg per second) for which all the abrasive particles fed into the apparatus will pass through the wear zone. How will $\dot{M}$ vary with particle diameter?
- (b) The specimen is pressed against the wheel with a force of 130 N. Estimate the load supported by each abrasive particle if the particle feed rate is (i) $\dot{M}$; (ii) $2\dot{M}$; (iii) $0.5\dot{M}$. Assume in all cases that the particles are present as a single layer in the wear zone.
- (c) Explain why in abrasive wear the load supported by each particle can have an important influence on the wear rate. In your answer, consider the behaviour of both the particles and the material being abraded.

Question 6.4

In the micro-scale abrasion test illustrated in Fig. 6.27(e) a steel ball with radius R is rotated against the surface of a flat specimen under a normal load W in the presence of hard abrasive particles. A crater with the same spherical radius as the ball is formed in the specimen by abrasive wear. After the surface of the ball has moved through a distance S relative to the specimen the diameter of the wear crater in the surface of the specimen is b .

- (a) Assuming that wear follows the simple model described by equation 6.5, show that for the micro-scale abrasion test

$$SW \approx \frac{1}{k} \left(\frac{\pi b^4}{64R} \right)$$

for a crater depth much less than R .

[The volume V of a spherical cap with spherical radius R and depth h is given by

$$V = \frac{\pi h^2}{3} (3R - h)]$$

- (b) A micro-scale abrasion test was carried out with a steel ball 25.4 ± 0.01 mm in diameter on a tool steel sample using an aqueous slurry of 4 μm silicon carbide particles. The ball rotated at 60 ± 1 r.p.m. and the force between the ball and the sample was 0.20 ± 0.01 N. After the test had been run for 500 ± 1 s the wear crater diameter was measured as 1.02 ± 0.01 mm. Calculate the specific wear rate k and

estimate the experimental error in this value. How could the experimental method be improved to increase the confidence in this result?

- (c) The test was then repeated on a separate area of the sample under identical conditions, but with the slurry now containing 3.5 wt% sodium chloride in addition to the silicon carbide. The crater diameter was then 1.14 mm. Explain why this crater was larger than that in the previous test, and suggest further experiments to study the process of material removal by abrasion in the presence of chloride ions.

Question 6.5

Assume that the normal impact of a hard erosive particle on a ductile metal is similar to a process of quasi-static plastic indentation, as illustrated in Fig. 6.33 and represented by equation 6.15.

- (a) Show that for a spherical particle of radius R and density σ impacting a metal with indentation hardness H , for a maximum depth of indentation much less than R , the time taken for it to come to rest t_p is independent of the initial impact velocity and is given approximately by:

$$t_p = \frac{\pi}{2} R \sqrt{\frac{2\sigma}{3H}}$$

Hence estimate the impact duration for a silica sand particle 50 μm in diameter, with density 2700 kg m^{-3} , impinging on a steel surface with a hardness of 250 HV.

- (b) By equating the initial kinetic energy of a spherical particle with impact velocity U to the plastic work done in forming the indentation, show that the depth x of the final indentation is given approximately by:

$$x = RU \sqrt{\frac{2\sigma}{3H}}$$

[The volume V of a spherical cap with height x and base diameter $2a$ is given by:

$$V = \frac{\pi x}{6} (3a^2 + x^2)]$$

- (c) For a silica sand particle striking a steel surface at 50 m s^{-1} , compare the depth of the plastic zone beneath the indentation (assumed to be approximately equal to x as calculated in part (b)) with the approximate distance for heat diffusion in the steel over the impact time calculated in part (a). What do these results tell you about thermal effects close to the impact site?

[The thermal diffusivity of the steel (see Section 3.9) is $11 \times 10^{-6} \text{ m}^2 \text{ s}^{-1}$.]

Question 6.6

The table below shows the mass loss (in mg) from specimens of an austenitic stainless steel in a series of gas-blast erosion experiments under three different sets of conditions, measured as a function of the total mass of erodent. The erodent particles were either spherical glass beads or angular olivine sand, all with the same size range

(100–150 μm), and the impact velocity (60 m s^{-1}) was the same in all the experiments. The nominal impact angle was either 90° or 30° , as shown.

Mass of erodent (g)	0	100	200	300	400	500	600
Round glass 90°	0	-0.8^{a}	-0.6^{a}	0.5	2.3	3.7	5.4
Angular olivine 90°	0	4.8	10.3	14.8	20.4	25.3	31.1
Angular olivine 30°	0	13.6	27.8	42.1	55.8	70.3	83.9

^aA negative number represents a mass gain.

- (a) Plot suitable graphs and calculate the steady-state erosion rate E (expressed as mass loss per unit mass striking the sample) for the three sets of conditions. Comment on the relative values of E and explain why they are different, describing the physical processes occurring during erosion.
- (b) To save time, instead of weighing the specimen after each exposure to 100 g of erodent an inexperienced tribologist proposes to carry out a single erosion test under each set of conditions, using 500 g of erodent and weighing the sample before and after the test. Explain why this method might lead to serious errors in the determination of E .
- (c) Approximately what value of E might you expect for erosion of the same steel by angular olivine sand 200–250 μm in diameter at an impact velocity of 100 m s^{-1} and an impact angle of 45° ? Explain and discuss your assumptions critically.

7.1 INTRODUCTION

Surface engineering involves the enhancement of certain properties of the surface of a component independently from those of the underlying substrate material. The enhancements may be in areas as diverse as visual appearance, tactile properties, optical properties, wettability, corrosion resistance or (the focus of this chapter) tribological behaviour. There are two common objectives in the use of surface engineering for tribological applications: to increase the wear resistance or damage resistance of the component, and to modify its frictional behaviour. In some cases, both are achieved together.

The processes by which a component can be surface engineered may be divided into three basic groups, as summarized in [Fig. 7.1](#).

- The first group consists of processes that modify the existing surface in some way *without* a change in composition, such as transformation hardening and surface melting.
- The second group consists of processes that modify the existing surface in some way *with* a change in composition of the surface layer being a critical feature of the process. Processes in this group may modify the existing crystal structure by forming a solid solution or by lattice disruption, or may result in changes to the transformation behaviour (e.g., carburizing of steels which leads to a change in maximum hardness and hardenability). However, they may also directly result in the formation of new phases, distinct from those of the substrate, by reaction between elements from the substrate and those introduced by the process. In some instances, the surface engineered layer is composed of these new phases in the form of precipitates within a matrix (e.g., aluminium or vanadium nitrides formed by nitriding of a suitable steel), while in other cases the new phases form a distinct layer. This reaction layer may well have all the properties that one would normally associate with a coating (a different phase with a distinct boundary between the layer itself and the substrate), but these processes are distinct from coating processes since the new surface phase is not added to the substrate but forms from it. The boundary between the substrate and the reaction layer is normally more diffuse than that between a true coating and a substrate, and the bond strength is normally higher. Anodizing (in which an alumina layer is formed on an aluminium alloy) and boronizing (where layers of borides may be formed on iron, titanium or cobalt-based alloys) are examples of such processes.
- The third group consists of processes which apply a material to the surface, and are generally referred to as coating processes. Here, the surface engineered layer is applied to the surface (as opposed to being a modification of the original material) and accordingly, the range of potential methods and materials is vast. Painting, weld hardfacing, electroplating and chemical vapour deposition are examples of processes within this group.

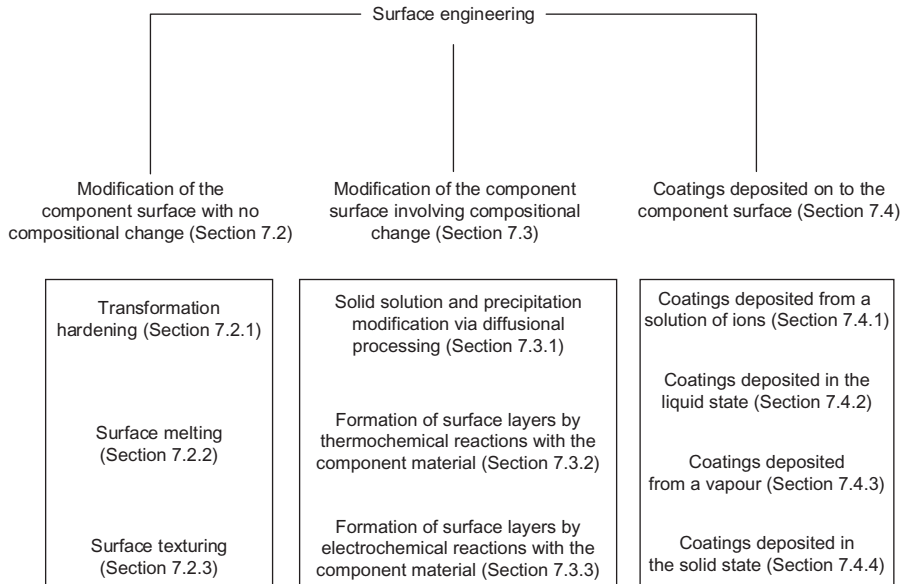


Fig. 7.1 Taxonomy of surface engineering methods

In this chapter we shall examine all the general classes of surface engineering process outlined in Fig. 7.1; novel coatings and processes are constantly being developed and considered for applications, and rather than attempt to cover all of these, a framework will be presented in which such innovations can be understood.

7.1.1 *Delamination of surface engineered layers*

Before describing surface engineering processes, it is important to understand how a surface engineered component differs from a homogeneous component in the context of a tribological contact, since this will inform our understanding of the key differences between types of process.

A homogeneous material can be defined by a range of properties that will govern its tribological response (in terms of friction and wear). Components can be surface engineered to enhance the response of the component, with the surface engineered layer exhibiting a combination of properties that provides a more favourable tribological response. However, the system is not only defined by the properties of the surface engineered layer and the properties of the substrate, but also by the properties of the interface, which can be classified as being either sharp or diffuse (Fig. 7.2).

When a surface engineered component is subjected to a tribological ‘challenge’ (which might, for example, be rolling contact in a bearing, sliding between gear teeth or abrasion by a rough hard counterbody) its response is associated not just with the behaviour of the surface material, but also with the material below the surface. If the conditions are sufficiently severe, they may cause the surface engineered layer to

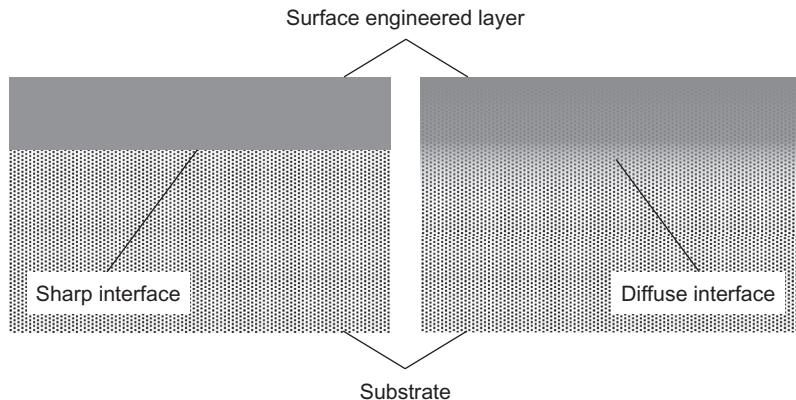


Fig. 7.2 The nature of interfaces in surface engineered components

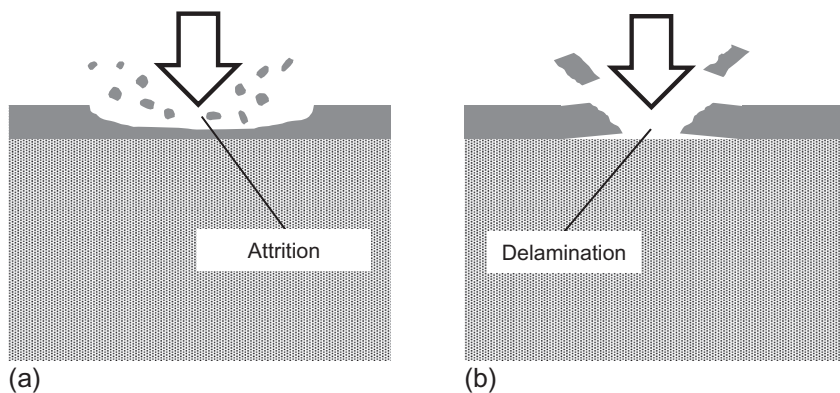


Fig. 7.3 Illustration of surface engineered layer exhibiting (a) progressive wear; (b) delamination

be removed by delamination (i.e., detachment at the interface) rather than by progressive wear, as shown in Fig. 7.3. Whether debonding occurs at the interface will depend on the strength of the substrate-coating bond, the intensity of the tribological conditions and the thickness of the coating layer (which together control the stresses at the interface), and any other driving forces for the coating to delaminate. When the interface is diffuse, in contrast, there is less risk of separation between the surface layer and the underlying material.

For systems with sharp interfaces, it is not straightforward either to describe or to measure the strength of the interface. The interface will typically exhibit very different properties in tension than in shear. The strength of an interface is often described in terms of either a uniaxial strength (σ_i) or in terms of a surface energy (G_i). In most surface engineering processes, considerable effort is devoted to increasing the interfacial strength, for example by cleaning and etching of the surface before

coating (as employed in electroplating and vapour deposition processes), grit blasting (commonly used before thermal spraying) or use of a bond coat—an intermediate layer between the coating and substrate which makes a strong bond to each material (even though the coating and substrate themselves may exhibit a low bond strength to each other).

In the deposition of a coating on to a substrate, residual stresses build up in the system (driven by deposition stresses and thermal stresses associated with cooling from the processing temperature). There are often stress gradients in coatings, but if it is assumed that (i) the coating is thin compared to the size of the component; (ii) the stress in the coating is uniform throughout its thickness and (iii) the coating is in a biaxial stress state, then the strain energy density (energy per unit volume) in the coating, Λ_c , (the integral of elastic stress and strain) can be estimated from:

$$\Lambda_c = \frac{\sigma_c^2}{E_c}(1 - \nu_c) \quad (7.1)$$

where σ_c is the residual stress in the coating, and E_c and ν_c are the Young's modulus and Poisson's ratio of the coating respectively. The strain energy released per unit area on debonding of the coating from the substrate is the product of the strain energy density (Λ_c) and the coating thickness, t , and if this product is greater than the interfacial surface energy, G_i , the debonding will be energetically favourable; once initiated, it will tend to proceed. Provided that the residual stress does not change as the coating thickness increases, the strain energy release rate is proportional to coating thickness but the interfacial surface energy is independent of thickness. Thus thicker coatings will show a greater tendency to delaminate.

Designers are often very concerned about coating delamination, since an estimate of the life of a component based on its rate of wear (or other progressive degradation mechanism) will be over-optimistic if failure is actually by delamination. The potential for early and unexpected catastrophic failure has meant that in some industries, the potential of surface engineering to improve the performance of components has not been exploited. Quantification of these risks to allow designers to specify coatings with confidence remains a challenge in certain applications, although in others, surface engineering is very widely used.

7.2 MODIFICATION OF THE COMPONENT SURFACE WITH NO COMPOSITIONAL CHANGE

7.2.1 Transformation hardening

The simplest method of producing a region of hard material at the surface of a softer carbon steel component is by *transformation hardening*, in which the surface material is rapidly and selectively transformed by heating to austenite, which is then quenched to form martensite and subsequently tempered, as illustrated in Fig. 7.4. The method is restricted in principle to materials that show a suitable phase transformation, and is

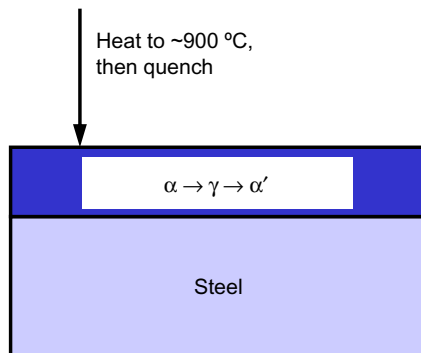


Fig. 7.4 In transformation hardening of steel the surface is rapidly heated to form austenite, which is then quenched to form martensite

widely applied only to ferrous alloys. The maximum hardness that can be achieved by this method is modest as we shall see below, but since it is rapid and relatively simple to implement, transformation hardening is widely used for steel components such as gear teeth, camshafts and crankshafts, cutter blades and various bearing surfaces. No new material is deposited, and thus there are no concerns about interfacial failure.

Bulk transformation hardening of steels is commonplace, and typically involves austenitization in a furnace before quenching to martensite. In both bulk and surface hardening of steels, the heat generally diffuses into the body from the surface (or a region very close to the surface in the case of induction hardening); only in the case of electrical resistance heating is the heat generated relatively uniformly throughout the body. The key difference in surface hardening is that the heat is applied quickly to the surface so that diffusion of the heat away from the surface and into the bulk is restricted. In this way, a steep temperature gradient can be developed, with the surface region being austenitized whilst the bulk of the material remains ferritic (and cannot therefore transform to martensite). This localization of heat close to the surface means that care needs to be exercised to ensure that melting does not occur. On removal of the heat source following surface austenitization, the heat is removed either by an external colder medium or by conduction into the colder bulk material, to achieve a cooling rate that results in martensite formation.

The depth of surface transformation hardening thus depends on the rate of heat input into the surface layer, the time for which that occurs, and the subsequent rate of heat loss, both from the surface and by conduction into the cooler underlying material. High power input for a short time will favour a shallow martensitic layer.

The source of heat used in transformation hardening determines the power density. Common methods use oxy-acetylene or oxy-propane flames (*flame hardening*), high frequency electrical induction heating (*induction hardening*), and lasers or electron beams (*laser hardening* or *e-beam hardening* respectively).

In both flame and induction hardening, an external quenchant is used to cool the workpiece after surface austenitization: commonly, water jets or a water bath. Typically, depths of hardening in the range 0.25–6 mm can be achieved by both methods.

Better control of depth is attained through induction heating, since the high frequency eddy currents responsible for heating are localized in the surface of the workpiece by the 'skin effect'. The depth to which the alternating current penetrates, δ , depends on its frequency f , and also on the temperature, which affects both the magnetic permeability μ and the resistivity ρ of the workpiece material. The skin depth δ is given by

$$\delta = \sqrt{\frac{\rho}{\pi \mu f}} \quad (7.2)$$

The permeability of iron falls sharply to unity at its Curie temperature of 770°C when it ceases to be ferromagnetic; the relationship becomes approximately

$$\delta \sim \frac{C}{\sqrt{f}} \quad (7.3)$$

$C \approx 20 \text{ mm Hz}^{1/2}$ at 20°C, and $500 \text{ mm Hz}^{1/2}$ at 800°C. Typically, frequencies between 3 and 500 kHz are used to produce hardened layers 0.5–5 mm in depth with input power densities (power per unit surface area) of about 20 MW m^{-2} . Figure 7.5 indicates the approximate range of depths and process temperatures used in both flame and induction hardening of steels.

In *laser hardening*, a laser is used as a heat source. A variety of laser types can be used, (CO_2 , Nd-YAG, diode etc), often selected on the basis of wavelength and beam shape (with wavelength significantly influencing the absorbance of the laser by metallic surfaces). In a typical setup, a beam from a high power (0.5–15 kW) laser

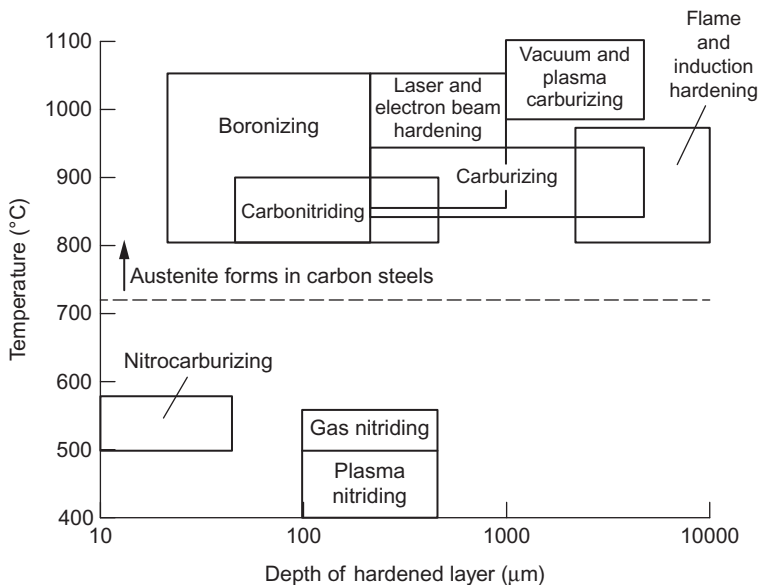


Fig. 7.5 Comparison of the process temperature and depth of hardened material produced by various methods of surface modification of steels

is directed on to the steel surface. The beam is focused to produce a spot 1–2 mm in diameter. By moving the source or workpiece, the spot can be made to traverse the surface in any desired pattern. Large areas are typically treated by scanning the beam in a raster pattern, whilst small areas of a larger component can also be readily treated. As the laser beam passes over a point on the surface, it is first rapidly heated (at a rate up to 10^6 K s^{-1}), then cooled by conduction of heat into the surrounding and underlying material. The cooling rate by conduction is so high (typically $>10^4 \text{ K s}^{-1}$) that no external quenchant is required. Figure 7.6 shows the range of power densities and heating times over which transformation hardening occurs. If the power density is increased (for example by focusing the beam more sharply) or the heating time is prolonged, surface melting will occur; processes in which this is used are discussed briefly in Section 7.2.2. The high power densities and short heating times used in laser transformation hardening result in high temperatures and shallow depths of hardening, as indicated in Fig. 7.5. A significant advantage over flame and induction hardening is that bulk heating and consequent distortion of the component are absent. The higher temperatures in laser hardening also lead to more rapid dissolution of carbides than in flame or induction hardening.

Electron beam hardening is similar in many respects to laser hardening. The energy input is provided by an electron beam focused to a diameter of about 3 mm, giving a similar power density to that used in the laser process ($10\text{--}100 \text{ MW m}^{-2}$ —see Fig. 7.6).

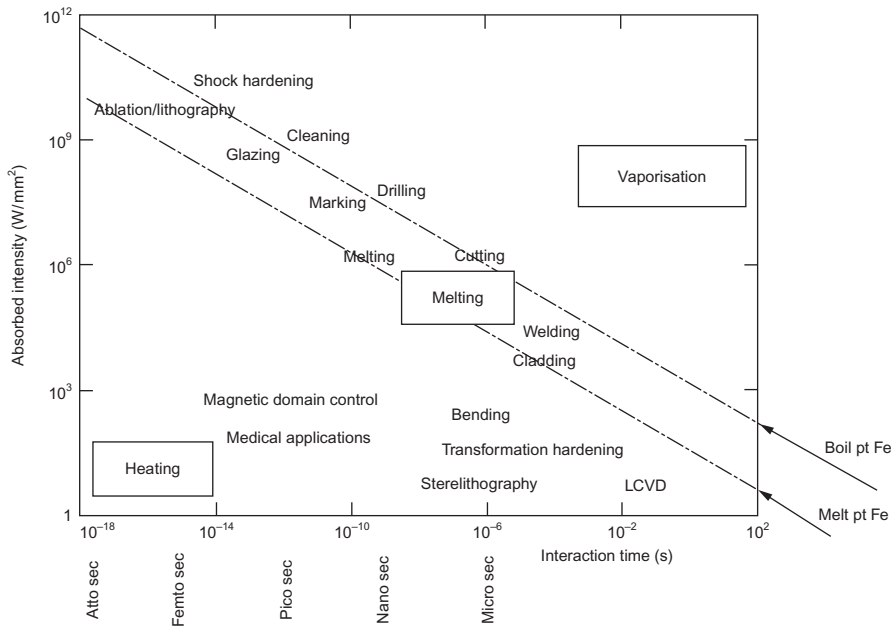


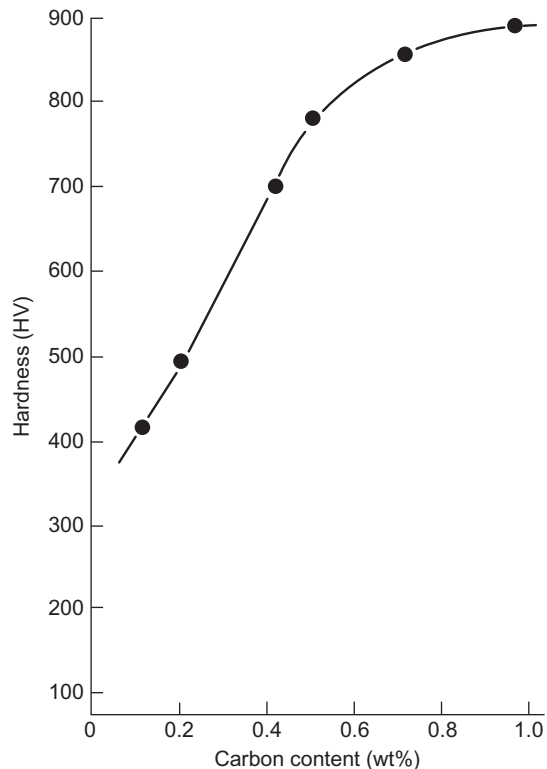
Fig. 7.6 Range of laser processes mapped against power density and interaction time. The diagonal lines represent lines of constant temperature for the boiling and melting point of iron (from Steen, W.M., Mazumder, J., 2010. *Laser Material Processing*, fourth ed. Springer, London)

The beam is moved over the surface by electromagnetic deflection, and the workpiece may also be moved in order to treat the desired area. The electron beam process differs from laser hardening in that it must be performed in a moderate vacuum (at a pressure of 1–10 Pa). Depths of hardening of up to about 2 mm can be achieved (see Fig. 7.5).

Figure 7.7 shows how the hardness of untempered martensite formed in plain carbon steels varies with carbon content. A further modest increase in hardness, typically by up to 100 HV, may result from alloying additions in low alloy steels. A steel of high carbon content would in principle be desirable for maximum hardness after transformation hardening, but in practice other factors limit the composition. Carbon contents above 0.5 wt.% give low toughness in the hardened layer, and lead to increased susceptibility to quench cracking. On the other hand, low carbon contents (<0.3 wt.%) require very rapid cooling rates ($>400 \text{ K s}^{-1}$) in order for martensite to be formed, which cannot be achieved in flame and induction hardening. For these reasons, these two processes are usually applied only to steels within a narrow range of carbon content: typically 0.4 wt.%–0.5 wt.%. Alloy steels may be used as well as plain carbon, as can cast irons with suitable matrix compositions.

The rapid cooling produced by self-quenching in laser and electron-beam hardening means that steels of lower carbon content can be satisfactorily hardened, although

Fig. 7.7 Hardness of untempered martensite as a function of carbon content for plain carbon steels



the hardness of the martensite formed will not be high. A further feature of the rapid cooling rate is retention of some austenite in the hardened layer, which becomes more prevalent in steels of higher carbon content; as we saw in [Section 6.3.3](#), retained austenite can lead to enhanced resistance to abrasive wear.

All the transformation hardening processes are characterized by a short process time, and all are to some extent suited to localized treatment of small areas. All are limited in the range of materials that can be hardened, and in the maximum hardness which can be produced. While flame and induction hardening are well suited to hardening of large areas, laser hardening in particular finds wide application due to its ability to harden steels of low carbon content without distortion and its ability to be applied selectively to specific areas on a component where it is required.

7.2.2 *Surface melting*

A further method of modifying surface properties without change in composition employs the liquid–solid phase transformation, by locally melting material and allowing it to resolidify. If solidification is rapid, the process causes homogenization and refinement of the microstructure. It may also lead to supersaturation and the formation of non-equilibrium phases, and even to glass formation in suitable materials. As noted above for transformation hardening, in order to produce localized surface melting and a rapid cooling rate, a high input power density is needed. This is usually achieved by laser or electron-beam heating, although electric arc welding methods with non-consumable electrodes (e.g., the tungsten inert gas process—TIG) can also be used. As no new material is deposited, there are no concerns about interfacial failure.

In laser and electron beam surface melting, higher power densities are employed than in transformation hardening (see [Fig. 7.6](#)). By adjusting the power and process time, the temperature profile can be controlled, which in turn determines the cooling rate once the power input ceases. Although simple melting results in no change in composition, the refinement of microstructure caused by rapid solidification leads generally to a significant increase in hardness in the treated surface compared with that of the bulk. In contrast to transformation hardening, the method can be applied to both ferrous and non-ferrous alloys and potentially even to non-metals. The effect of microstructural refinement in tribological applications may extend beyond a simple increase in hardness. For example, remelting of thermally sprayed coatings can be employed to eliminate the splat-like microstructural features exhibited by such coatings (see [Section 7.4.2](#)). [Figure 7.8](#) shows micrographs of a thermally sprayed Stellite 6 coating, both in the as-deposited state and following laser remelting; in this case, the laser remelting resulted in a reduction in hardness, but an increase in wear resistance by almost a factor of two was observed when sliding against a martensitic stainless steel.

Low carbon steels are unsuitable for surface melting treatment, since soft δ -ferrite is formed and retained in the quenched surface layer. At carbon contents between 0.4 wt.% and 0.9 wt.%, however, the treated region consists of martensite and retained austenite which can be hardened by further heat treatment, and the process yields significant benefits.

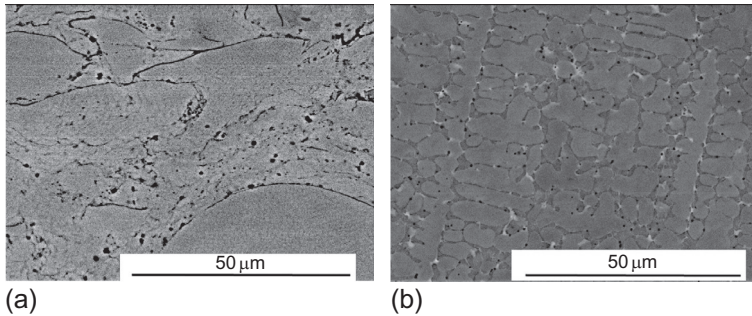


Fig. 7.8 Back-scattered scanning electron micrographs of HVOF sprayed Stellite 6 coating; (a) as sprayed; (b) following laser remelting (from Houdková, Š., et al. 2016. Surface and Coatings Technology. <http://dx.doi.org/10.1016/j.surfcoat.2016.09.012>)

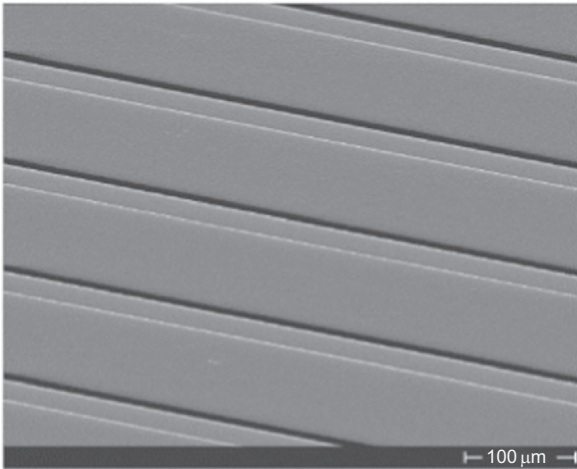
Cast irons respond well to surface melting, developing fine-grained structures similar to those formed in conventional chill-casting. The TIG melting process is widely applied to wearing surfaces on crankshafts, camshafts and cam followers, and allows a very hard surface to be combined with a tough core material.

7.2.3 Surface texturing

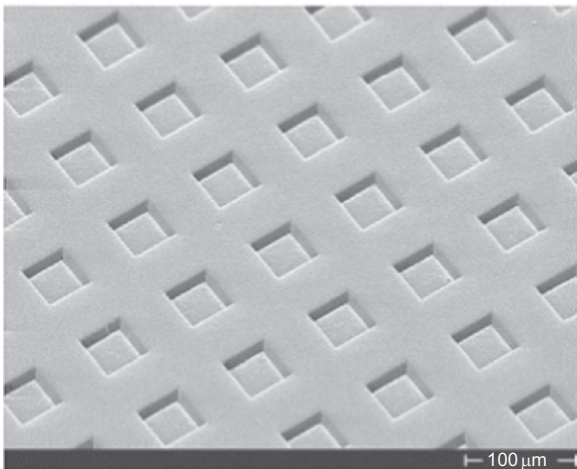
Surfaces texturing involves the development of a topographical pattern on the surface, typically in the form of grooves or dimples, on a scale that is generally much smaller than the size of the tribological contact. Such features may affect the contact in a number of ways, such as by altering the flow of lubricant and the lubricant film thicknesses in lubricated contacts, or by altering the way that debris is retained in a sliding contact or expelled from it.

There is significant interest in the use of textured surfaces in lubricated bearings, with golf-ball-like dimpled textures showing significant promise. However, detailed understanding of the way that texturing affects tribological contacts is far from complete, with reports of texturing being ineffective or even deleterious in certain test programmes. It is clear that the effects of texturing need to be considered alongside a consideration of the more general tribological aspects of the contact. For example, textured surfaces of the type illustrated in Fig. 7.9 were then coated with a low-friction DLC coating and slid against a steel ball. In unlubricated conditions, texturing of the surfaces disrupted the formation of a transfer film on the steel counterbody and resulted in higher friction than when the balls were slid against the equivalent surface with no texture. However, in boundary lubricated sliding of the same system, some (but not all) textures resulted in lower coefficients of friction than with the equivalent surface with no texture.

Texturing of a surface in this way can be achieved by a variety of means, such as mechanical indentation, lithography, electrical discharge machining (EDM) and laser



(a)



(b)

Fig. 7.9 Two surface textures with 25% surface coverage: (a) grooves; (b) square pits (from Pettersson, U., Jacobson, S., 2003. *Tribol. Int.* 36 (11), 857–864)

engraving. The main variables are the size and shape of the topographic features, their area fraction and any orientation of the pattern with respect to the shape and relative motion of the contact.

7.3 MODIFICATION OF THE COMPONENT SURFACE INVOLVING COMPOSITIONAL CHANGE

The range of surface microstructures that can be achieved by transformation hardening and by surface melting (without alloying or addition of second phases) is limited by the composition of the starting material, and the properties that can be achieved are

therefore restricted. Methods in which the composition of the surface is locally altered, on the other hand, can produce microstructures and associated mechanical properties that are completely different from those of the substrate.

Two classes of solid state processes can be identified in which atomic transport by diffusion leads to surface modification, and the majority of these are *thermochemical processes*. In the first class, diffusion of small atoms into the surface leads to the formation of an interstitial solid solution in the substrate material, and sometimes to the formation of compounds as very fine precipitates. *Carburizing* and *carbonitriding* are examples. In processes of the second type, chemical reaction occurs between the diffusing atoms and constituents of the substrate, causing the formation of a distinct layer of a new compound at the surface. Examples of this type of process are provided by *nitriding*, *nitrocarburizing*, *boronizing* and *chromizing*. For all processes that depend upon thermally-enhanced diffusion of species, the diffusion distance (x) (and therefore the depth of modification) depends upon the diffusion coefficient of the species (D) and the time (t) of processing as follows:

$$x \approx \sqrt{Dt} \quad (7.4)$$

This equation is essentially the same as that for the conduction of heat (equation 3.35). As discussed in [Appendix B](#), the diffusion coefficient is very sensitive to temperature, exhibiting Arrhenius behaviour (equation B.6).

In their end result, these processes are similar to the coating processes discussed later in this chapter, but a clear distinction can nevertheless be drawn in that the surface layer, while of different composition from the substrate, is formed by chemical reaction with the substrate material. For this reason the interface between the surface layer and the bulk material is often more diffuse, and of higher strength, than for an externally applied coating.

7.3.1 Solid solution and precipitation modification of surfaces via diffusional processing

Processes of this type depend on diffusion of species into the component (generally at high temperature), giving a surface engineered layer with a diffuse interface with the bulk of the component. There are therefore no concerns about interface strength.

The most widely used thermochemical process for surface hardening of steels is *carburizing*, sometimes called *case hardening*, shown schematically in [Fig. 7.10](#). Carburizing involves the diffusion of atomic carbon into steel from the surface, to produce a surface region or *case* of enhanced carbon concentration, typically up to several millimetres deep (see [Fig. 7.5](#)). The process is applied to plain carbon or low alloy steels of low initial carbon content, typically 0.15 wt.%–0.2 wt.% C, which provide a core that will remain tough even in the as-quenched condition. Carburizing is carried out in the austenite phase region, typically at temperatures of 900°C or higher; at these temperatures the diffusion of carbon in austenite is rapid. The carbon concentration in the surface layer may be enhanced to 0.7 wt.%–0.9 wt.% by carburizing, producing a

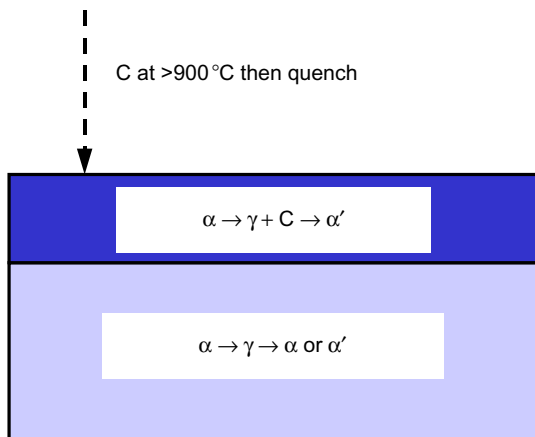


Fig. 7.10 Carburizing involves diffusion of carbon into the surface of the steel while it has an austenitic structure. Quenching results in the formation of martensite near the surface with a higher carbon content than in the bulk; the underlying material may transform to martensite or ferrite, depending on the local hardenability and cooling rate

maximum hardness of up to about 900 HV (see Fig. 7.7). In some methods, the component is quenched immediately after the carburizing process, while in others it is cooled slowly and can then be machined to final dimensions in a soft condition before the final heat treatment to produce the hardened case of lightly tempered martensite.

In all methods of carburizing, the martensitic transformation that provides the mechanism for hardening also causes distortion of the component due to the associated change in lattice volume. Although this distortion can be reduced by careful practice, it can never be entirely eliminated. Dimensional tolerances after carburizing are therefore lower than after methods of surface treatment that are carried out in the ferrite phase region. The process is widely used for rotating shafts and bearing components, cam followers, gears and camshafts. In addition to the high surface hardness and consequent wear resistance which is of tribological interest, the martensitic transformation produces a compressive residual stress in the surface that substantially increases the fatigue life.

The depth of hardening and the hardness achieved depend on the time, temperature and carbon activity at the surface during carburizing, and these in turn depend on the process used. *Gas carburizing* is a common method. Traditionally the components are heated to $\sim 900^\circ\text{C}$ in an atmosphere of carbon monoxide, hydrogen and nitrogen; recent developments use a mixture of methanol and nitrogen. In *pack carburizing*, sometimes used to obtain very thick cases, the components are heated in sealed boxes with a granular medium containing charcoal and an 'energizer' (typically barium carbonate). Carburizing takes place in the same way as in the traditional gas process, through the action of carbon monoxide formed from the carbon and residual oxygen within the box. Since the diffusion of carbon in iron is thermally activated, at higher temperatures it becomes significantly more rapid. Despite the limitations imposed by furnace design and heating times, the temperature for gas carburizing can be raised to 1000°C with substantial shortening of the process time.

Vacuum carburizing and *plasma carburizing* both operate at high temperature, about 1050°C. In vacuum carburizing, the components are heated in moderate vacuum, and methane or propane is then admitted to the furnace at a low pressure. Reaction of the gas at the surface of the hot steel provides the source of carbon. Following saturation of the surface with carbon, a further period under vacuum at high temperature is then allowed for it to diffuse inwards. In plasma carburizing, a glow discharge in methane at low pressure is used to deposit carbon on the surface of the hot substrate which is held at a negative potential. As with the vacuum process, a short carburizing period is then followed by a longer diffusion period. Both vacuum and plasma carburizing are energy-efficient processes, and because of the high process temperature, result in much deeper case hardening than lower temperature methods of the same duration. Figure 7.11 illustrates hardness profiles in a low carbon steel carburized for approximately the same time by the gas method, by the vacuum method with two different gas pressures, and by the plasma method. A much longer carburizing time would be required for the lower temperature gas method to achieve the same depth of hardening as the other methods.

Carbonitriding is a very similar process to carburizing, involving the simultaneous diffusion of both carbon and nitrogen into austenite in a low carbon steel

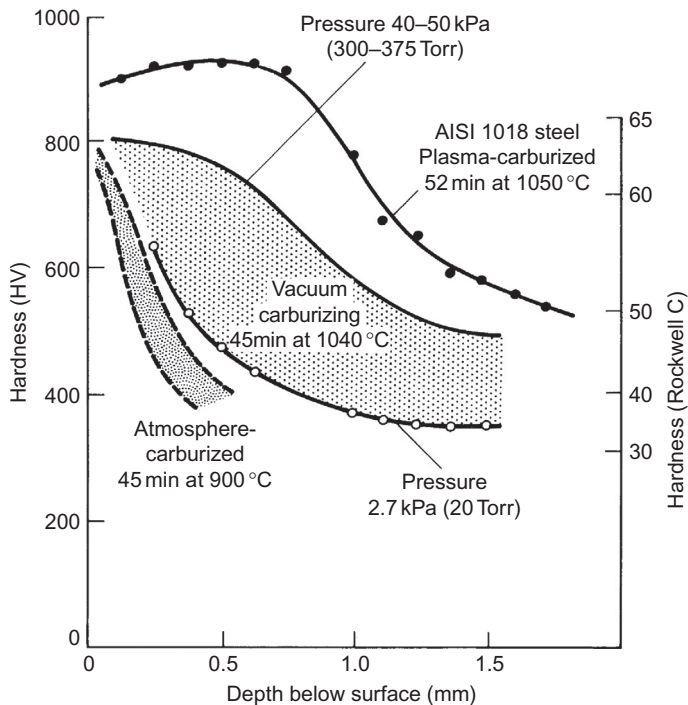


Fig. 7.11 Variation of hardness with depth in a plain 0.18 wt.% carbon steel carburized by different methods (from Grube, W.L., Gay, J.G., 1978. Metall. Trans. 9A, 1421)

(<0.25 wt.% C). The process is typically carried out at 800–900°C to produce case depths from 0.05 to 0.75 mm (Fig. 7.5). The final nitrogen concentration in the hard layer is 0.5 wt.%–0.8 wt.%, with a carbon concentration similar to that achieved in carburizing. The effect of the nitrogen is to increase the hardenability of the surface layer; oil quenching rather than water quenching can therefore be used, reducing the risk of quench-cracking. The nitrogen also increases the resistance to tempering of the martensite formed. In general, carbonitrided steels are more resistant to sliding wear than those carburized to the same hardness.

Two methods are commonly used for carbonitriding: a gas phase method and a molten salt treatment. *Gas carbonitriding* is similar to gas carburizing, and can be achieved simply by the addition of ammonia to the gas mixture used for carburizing. *Salt bath carbonitriding*, sometimes misleadingly called *liquid carburizing*, involves immersion of the components in a molten salt bath containing sodium cyanide (typically 45% NaCN, 40% Na₂CO₃, 15% NaCl) at 880°C. Treatment times are up to 1 h, and the parts may be quenched straight from the salt bath.

One final process to be considered in this group is that of *ion implantation*, and in this case the diffusion of the atoms into the surface is not thermally activated, but is driven by the kinetic energy of ions arriving at the surface. If ions of high energy strike a surface, they can penetrate to a macroscopic distance, and change the properties of the material to this depth. Ions commonly used for implantation in surface engineering include N⁺, N²⁺, C⁺ and B⁺, and also metal ions: e.g., Ti⁺, Al⁺ and Y⁺. Their energies are usually in the range from 50 to 200 keV, giving penetration depths of less than 1 µm. The peak concentration, at an overall dose of 10¹⁷ ions cm⁻², is >10 at.%. Doses of this level are about the minimum that give practically useful changes in tribological properties, and are several orders of magnitude greater than those used to modify the electronic properties of semiconductors. Because of the shallow depth of material affected by ion implantation, the process causes negligible change in the dimensions or surface finish of the substrate. It can be applied to ceramics and cermets as well as to metals.

Implantation of energetic ions alters the structure of the surface material in two ways: through the introduction of the implanted species, which may form a solid solution or compound, and also through radiation damage—the introduction of lattice defects through the displacement of matrix atoms by ion-induced collisions. In many materials, ion implantation leads to a significant decrease in sliding wear rate, for which several reasons have been suggested. Implantation of some metals by certain ions also leads to a reduction in friction, which may be directly responsible for the reduced wear rate by lowering the tangential forces transmitted on sliding. For example, implantation by Ti⁺ ions reduces the coefficient of friction for dry sliding of a hard bearing steel on itself from 0.6 to 0.3, probably by modifying the surface composition, its oxidation characteristics and the nature of the oxide formed. It also leads to a lower wear rate. However, implantation of N⁺ ions also reduces the rate of wear of relatively ductile steels (though not hard martensitic steels), despite causing no reduction in friction. In this case, the main source of the increased wear resistance is probably the changes in near-surface mechanical properties: in particular, hardness and strain-hardening rate. Microhardness measurements at very small loads have revealed

increases in the surface hardness of N⁺-implanted steels by 50% to 100%. Steels containing nitride formers (e.g., aluminium, chromium or vanadium) show a greater effect than others. A further effect of ion implantation that may contribute to tribological benefits is the introduction of residual compressive stress in the surface regions. Typical applications are in polymer injection molding equipment, such as metallic extruder screws, barrels, gates and moulds where ion implantation has been shown to reduce friction, wear and sticking.

7.3.2 Formation of surface layers by thermochemical reactions with the component material

The second group of thermochemical processes consists of those in which a reaction occurs between a constituent of the substrate and an externally-supplied chemical species, either hardening the substrate through precipitation of the reaction product, or forming a hard layer of the reaction product at the surface. In cases where a distinct layer is formed, the interface can often be described as sharp, and in these cases, the properties of the interface itself must be considered when assessing its suitability for use in tribologically demanding environments. We shall examine four processes in this class: *nitriding*, in which nitride compounds are formed as fine precipitates close to the surface; *nitrocarburizing*, which forms a layer of iron carbonitride; *boronizing*, in which iron boride layers are formed; and *chromizing* which forms a chromium carbide layer. The hardnesses resulting from these surface treatments are listed in Table 7.1, together with values for steels hardened by other methods for comparison.

Nitriding, shown schematically in Fig. 7.12, differs from the thermochemical hardening processes discussed in the previous section in that it is carried out at a lower temperature: 500–570°C (see Fig. 7.5). In this range, carbon steels are ferritic. The process is applied to steels that contain the nitride-forming solute elements aluminium, chromium, molybdenum, titanium, tungsten or vanadium. The most effective

Table 7.1 Maximum hardnesses attainable in steels by various thermal and thermochemical methods of surface modification (data from Child, H.C., 1980. Surface Hardening of Steel, Engineering Design Guide No. 37. Oxford University Press)

Process (and compound formed)	Hardness (HV)
Thermal hardening (0.5 wt.% C steel)	700
Carburizing and carbonitriding	850–900
Nitriding:	
Cr-Mo steel	650
Cr-Mo-V steel	900
Cr-Mo-Al steel	1100
Nitrocarburizing (forms Fe ₂ (C,N))	500–650
Boronizing (forms FeB, Fe ₂ B)	1500
Chromizing (forms Fe ₂ Cr ₃)	1500
Toyota diffusion (TD) process	3000–4000

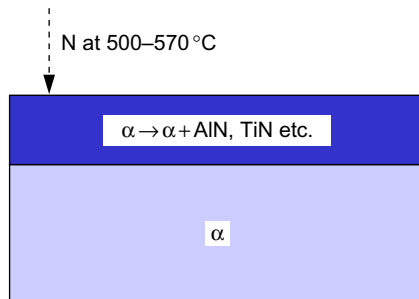


Fig. 7.12 Nitriding involves the diffusion of nitrogen into a steel containing nitride-forming elements while in a ferritic condition, to form nitride precipitates

hardening, giving values up to 1100 HV, is obtained in steels containing aluminium or titanium. In nitriding, atomic nitrogen is formed at the surface of the steel and diffuses inwards, reacting with the solute atoms to form very fine nitride precipitates typically 5–15 nm in size. The lattice strains associated with these precipitates are high enough to nucleate dislocations, which together with the precipitates themselves have a strong hardening effect. The nitriding process is not restricted to ferritic steels, but can also be applied to austenitic stainless steels which contain nitride-formers. The hardening produced by nitriding is retained at temperatures up to 500°C; in contrast, the hardness of the martensitic layers produced by carburizing starts to fall above 200°C.

In addition to the high hardness which gives nitrided surfaces good tribological properties, nitriding also produces a compressive residual stress which enhances the fatigue strength. Because the process is operated in the ferritic phase region for carbon and low alloy steels, heat treatment of the core material must be carried out *before* nitriding. Steels with about 0.4 wt.% C are therefore commonly used to give reasonable core strength after tempering at the nitriding temperature. The relatively low process temperature and the absence of phase transformations mean that there is very little distortion associated with nitriding. However, a disadvantage of the low temperature is the relatively sluggish diffusion of the nitrogen atoms into the surface, necessitating long treatment times and limiting the practical thickness of the hardened layer (see Fig. 7.5).

Two main methods are used for nitriding steels. In *gas nitriding*, the parts are heated to 530°C in a stream of ammonia gas (often mixed with other gases). The ammonia dissociates at the metal surface, forming nitrogen which diffuses inwards. Process times are long, measured in days rather than hours: 4 days may be needed to develop a hard layer 500 μm thick. In *plasma nitriding* or *ion nitriding*, the steel component is placed in a chamber containing nitrogen and hydrogen at a pressure of 10–1000 Pa, and a plasma discharge is established at a potential of 500–1000 V with the workpiece as cathode. The electrical power dissipation heats the steel surface, which is bombarded with nitrogen ions. The process is energy-efficient and about three times as fast as gas nitriding at the same temperature. It can also be operated

at lower substrate temperatures (down to 350°C) and can therefore be applied to high strength steels, such as tool steels, which would be over-tempered at a higher process temperature.

In both gas nitriding and plasma nitriding a thin surface layer of iron nitrides (mixed ϵ -Fe₂N (h.c.p.) and γ' -Fe₄N (c.c.p.)) may form, known as *white layer*. This usually has a deleterious effect, since although hard it cracks easily and can promote fatigue failure in the substrate. Formation of white layer can be avoided by careful control of the gas composition (in gas nitriding) or the process conditions (in plasma nitriding); if formed, it can be removed by grinding or by chemical dissolution.

In the process known as *nitrocarburizing*, the formation of a compound layer very similar to the white layer is deliberately employed to provide a hard and wear-resistant surface on cheap steel substrates. In addition to the relatively thin hard layer, the method also produces significant hardening of the underlying steel through inward diffusion of nitrogen. The composition of the hard layer depends on the details of the process: it is usually described as ϵ -Fe₂(C,N) although it may also contain some oxygen and sulfur. Like nitriding, nitrocarburizing is carried out on carbon steels in the ferritic condition, typically at around 570°C; mild or very low alloy steels are commonly used, although the process can also be used with higher strength alloys. The hardness of the surface layer is lower than that produced by nitriding (typically 600–900 HV), but the process is much faster. Layer thicknesses up to about 20 μ m are used, taking about 2 h to form. Molten salt bath processes, often misleadingly referred to as *salt bath nitriding*, provided the earliest nitrocarburizing methods and were based on sodium cyanide and cyanate. The proprietary names of these processes (e.g., ‘Tufftride’) are sometimes used generically. The early bath compositions are now being replaced with thiocyanate baths of lower toxicity, and gas and plasma processes have also been developed, often using methane (or another hydrocarbon) as the source of carbon, in addition to ammonia. The iron carbonitride layer usually exhibits microporosity which can be exploited by introduction of an organic sealant or a lubricant to further enhance the properties of the component.

Boronizing, also known as *boriding*, involves the diffusion of boron into the surface of a metal to form a hard layer of metal boride. Since it depends on the reaction between the boron and the component being treated, its application is generally limited to steels, titanium-based alloys and cobalt-based hardmetals. The process is predominantly applied to steels and, in this case, is carried out above the austenite transformation temperature, being held at temperatures of between 800°C and 1050°C (see Fig. 7.5) for several hours. In steels, boronizing forms compound layers up to 200 μ m thick, often consisting of two distinct phases: an outer layer of FeB (orthorhombic), and an inner layer next to the substrate of Fe₂B (body-centred tetragonal). The proportions of the two phases depend on the composition of the boronizing environment and the alloy content of the steel (higher alloy content favours FeB formation). The iron boride layers are very hard (>1500 HV); since this is significantly higher than many commonly occurring abrasives, boronizing has been employed in situations requiring abrasive wear resistance. The structure of the two-phase boride layers on a low-carbon steel is shown in Fig. 7.13.

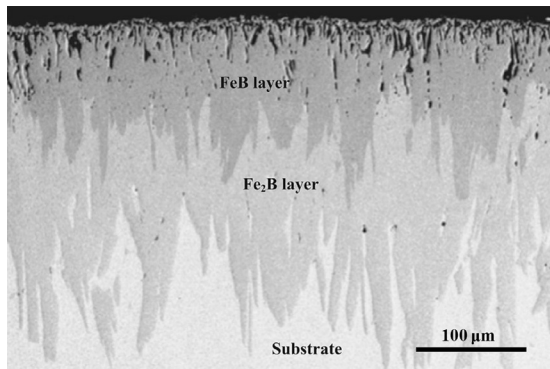


Fig. 7.13 Cross-sectional SEM micrograph of AISI 1018 steel pack-borided at 1000 °C for 6 h (from R. Carrera-Espinoza, et al. 2016, *Wear* 362–363, 1–7).

The two iron boride phases have different coefficients of thermal expansion and therefore develop differential residual stresses on cooling from the processing temperature, with the FeB being under tensile stress. Cracking at the interface between the phases sometimes originates from this source. Formation of FeB can be reduced or prevented by controlling the process conditions, giving a preferable single-phase compound layer. The irregular, deeply indented interface between the Fe₂B layer and the steel substrate in Fig. 7.13 is characteristic of boride layers formed on many steels, and provides a strong mechanical bond to the underlying steel. Alloying elements in the steel not only reduce the rate of boride layer growth (and hence the thickness for fixed treatment conditions), but also change the morphology of the interface. Chromium, nickel and carbon all have similar effects: increasing their concentration leads to a smoother interface. At a chromium content of 12 wt.% in a low carbon steel, the interface is completely smooth.

The most common process is *pack boronizing*, carried out in a similar way to pack carburizing. The granular boronizing medium consists of a mixture of boron carbide (B₄C), an inert diluent such as silicon carbide or alumina, and an activator (e.g., KBF₄). The function of the activator is to transport boron in the gas phase to the steel surface; it is regenerated by reaction with the boron carbide. The boride layer formed on low alloy steels is typically 150 μm thick for a treatment time of 6 h at 900°C. Pack boronizing is an attractive process because it can be carried out with simple plant. Other methods involving either molten salt baths or gas phase processing (with BCl₃ as the boron source) are available, and plasma boronizing is under development. In common with other surface treatment processes carried out in the austenite region, boronizing may be followed by heat treatment in order to develop optimum properties in the core steel.

Boronizing is not restricted to steels. The process can also be applied to tungsten carbide/cobalt cemented carbides, in which cobalt borides (CoB, Co₂B and Co₃B) and tungsten carboborides (W₂(C,B)₅ and W(C,B)₂) are formed, and to titanium and its alloys which form titanium boride layers (TiB and Ti₂B) with hardnesses up to 2500 HV.

Boronizing is one of several processes sometimes collectively known as *metallizing* or *metalliding*, in which hard surface layers are formed by reaction between a constituent of the substrate and one or more metals or metalloids. Chromizing is another such process for the surface hardening of carbon and tool steels, in which chromium reacts with carbon from the steel to form a surface layer of chromium carbide. In this process the rate-controlling step is the diffusion of carbon towards the surface of the steel, and the process may have to be preceded by carburizing to ensure that the initial carbon concentration is sufficiently high. The surface layer formed by chromizing is characterized by very high hardness (>1500 HV) which is retained to high temperature (typically 700°C). Pack, gas-phase and molten salt bath processes are all available, and treatment temperatures from 900°C to 1000°C produce chromium carbide layers up to $40\text{ }\mu\text{m}$ thick.

Other transition metal carbides, as well as chromium carbide, are formed in the salt bath process known as the *Toyota Diffusion (TD) process*. The method was developed primarily to provide extremely hard wear-resistant layers of chromium, niobium, titanium and vanadium carbides on tool and die steels. Layer thicknesses are typically $5\text{--}10\text{ }\mu\text{m}$, formed by immersion for up to 10 h in a salt bath containing molten borax and iron alloys of the appropriate carbide-forming element. The salt bath temperature, between 800°C and 1050°C , is usually chosen so that the steel can be quenched directly from the bath. Hardness values of $3200\text{--}3800$ HV have been reported for vanadium and titanium carbides, and $2500\text{--}3000$ HV for niobium carbide, formed by the TD process. These carbides can also be formed by other thermochemical processes; for example, a pack process conducted in a reactive gas (methane + argon) has been developed for chromium, titanium and vanadium carbides, and other processes based on gas-phase and plasma treatments are available.

A widely-used and simple process applied to sintered iron components for tribological applications (such as bearing surfaces in small refrigeration compressors) is *steam oxidation*. The process involves exposing the sintered part to steam at a temperature of $500\text{--}550^{\circ}\text{C}$ for about 1 h. Initially used to seal the network of interconnected pores and thus make the component impervious to liquids and gases, steam oxidation treatment improves other properties as well. The magnetite (Fe_3O_4) formed on the surface and in the interconnected porosity provides an increase in hardness and in mechanical properties as a whole, and can offer benefits of reduced wear rate and friction and an increase in corrosion resistance in moderately aggressive environments.

Although we have concentrated on processes applied to steels, similar principles can be adopted to form hard surface layers on non-ferrous metals. Processes exist for surface hardening alloys of copper, aluminium and titanium, in which the electrolytic deposition of a metallic alloy on the surface is followed by heat treatment; inter-diffusion during the latter phase leads to the formation of hard compounds, giving a hard surface layer typically $20\text{--}30\text{ }\mu\text{m}$ thick. For example, in the *Delsun* process for brasses and bronzes, an alloy of antimony, cadmium and tin is electrodeposited, followed by a diffusion treatment at 400°C . The surface layers formed by such processes are much softer than the layers formed on steels by the methods previously outlined (e.g., $450\text{--}600$ HV on brasses and bronzes, and $200\text{--}500$ HV on aluminium alloys), but are nevertheless substantially harder than the substrate alloy and give significant tribological benefits.

7.3.3 Formation of surface layers by electrochemical reactions with the component material

The third group in this class consists of processes in which an electrochemical reaction occurs between the substrate and a solution of ions, forming a hard layer of the reaction product at the surface. In these cases, the interface is generally sharp, and accordingly, there is a need to consider the properties of the interface itself when assessing its suitability for use in tribologically demanding environments.

Anodizing refers to the formation of an oxide layer by a process where the component to be coated forms the anode in an electrochemical cell. Anodizing is commonly applied to certain aluminium alloys, although it can also be applied to titanium, magnesium and zinc. In anodizing of aluminium alloys, a layer of alumina is grown by anodic reaction of the surface itself in a sulfuric acid solution. The alumina film contains pores that allow the electrolyte to continue to have access to the metallic surface, and therefore allow the alumina film to continue to grow. An anodized layer can confer a number of benefits such as greater paint adhesion, corrosion resistance or wear resistance.

Aluminium alloys are attractive materials for many applications (often due to their low density) but have poor wear resistance and a tendency to gall. To provide wear resistance, an anodized coating layer between 25 and 150 μm can be formed by *hard anodizing*. A hard anodized surface on an aluminium alloy would normally have a hardness in the range 350–650 HV. The hard anodizing bath employs an aqueous sulfuric acid electrolyte (10–20 wt% acid) at temperatures normally between 0°C and 10°C with current densities between 2 and 4 A dm^{-2} . These conditions tend to minimize the dissolution of the alumina in the acid solution and result in a less porous structure than in conventional anodizing. Hard anodized surfaces tend to be rough, and are usually ground or lapped to the finish required for the application.

Not all aluminium alloys are suitable for hard anodizing: alloys containing more than 10 wt.% silicon or 5 wt.% copper, for example, experience severe local attack in the anodizing bath. Low alloy contents usually lead to harder films. It should also be noted that the process causes some loss of fatigue strength (by up to 50%). Anodized coatings are porous, and use can be made of this porosity by impregnating the coating after it is formed with solid lubricants such as PTFE or MoS_2 . Proprietary coatings made in this way can give very low friction (determined by the solid lubricant) and high hardness (determined by the alumina film).

Titanium alloys are also useful structural materials due to their low density, but in situations where motion between mating surfaces occurs, galling again may result. To alleviate this problem to some extent, titanium can be anodized in either caustic electrolyte or dilute acids (based on phosphoric acid). The anodized layers are only of the order of micrometres thick and thus only benefit components subject to moderate mechanical stresses.

Plasma electrolytic deposition involves the formation of arc plasma discharges in an aqueous solution which ionize gaseous media from the solution and promote reaction with the substrate material. Whilst plasma electrolysis has been employed

to nitride or carburize steels, *plasma electrolytic oxidation* (PEO) (sometime termed *micro-arc oxidation: MAO*) has been employed commercially to form coating layers on materials such as alloys of aluminium, titanium and magnesium. Whilst the technology itself is not new, it has only relatively recently been developed into a commercial process for the production of wear-resistant surfaces. The process is commonly operated with an alkaline bath at potentials above the break-down voltage of the oxide film that is growing (typically in the region of 120–350 V), and growth continues through plasma thermochemical interactions. As in anodizing, the formation of the coating depends on reactions involving the substrate material and thus bond strengths tend to be relatively high. The nature of the coating depends on the alloying elements within the substrate; PEO treatment of aluminium alloys is more tolerant of elements such as copper and silicon than the anodizing process. In aluminium alloys, dense, wear-resistant layers up to 0.5 mm thick have been reported with hardnesses of greater than 2000 HV. Figure 7.14 illustrates a typical PEO layer on an aluminium alloy. To further modify or enhance the properties of the coating, fine particles of materials such as hard carbides or solid lubricants can be introduced into the electrolyte which then become integrated into the coating.

7.4 COATINGS DEPOSITED ON TO THE COMPONENT SURFACE

The processes considered in this section deposit material on to the surface of a component to form a coating. In all cases except that of weld hardfacing, the interface between the coating and the substrate is sharp.

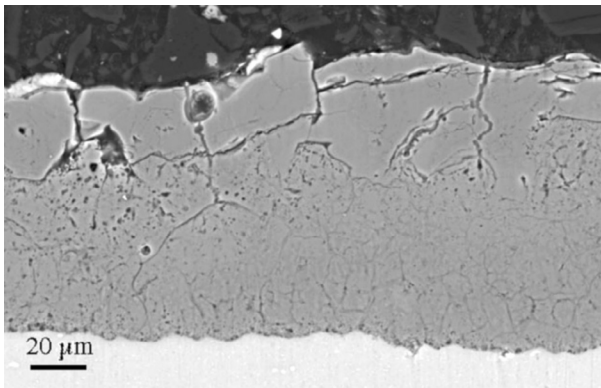
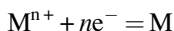


Fig. 7.14 SEM image (back-scattered electron mode) of a polished section from a PEO coating on an aluminium alloy substrate, showing the fine network of channels created by individual discharge events, representing a porosity level of the order of a few percent (from Curran, J.A., Clyne, T.W., 2006. *Acta Materialia* 54 (7), 1985–1993)

7.4.1 Coatings deposited from a solution of ions

One of the earliest processes for depositing a coating of one metal on to the surface of another was *electroplating*, and the method is still relatively widely used to produce coatings with a tribological function.

Electroplating or *electrodeposition* involves the reduction of metallic ions at the surface of the substrate, which is made the cathode in an electrolytic cell, as shown in Fig. 7.15. The relevant cathodic reaction is:



Clearly, the component to be plated must be electrically conducting, and electrodeposition is therefore normally on to metallic components. Components are thoroughly cleaned and may be etched before the deposition process begins.

Many different metals and alloys can be electrodeposited from aqueous solution, although the plating of alloys is more complex than that of pure metals since the rates of deposition of the relevant cations depend upon the bath chemistry and deposition conditions. Additions to the plating bath (complexants) are normally used to limit variability in the final product. The morphology and structure of a typical alloy electrodeposit can be seen in Fig. 7.16. Of the vast array of systems that can be electrodeposited, the coatings of major tribological interest are chromium and nickel. These coatings are often termed *hard chrome* and *hard nickel*, to distinguish them from the much thinner coatings with different properties. Whilst thinner coatings are typically employed for other purposes (e.g., to confer corrosion resistance or for decorative reasons), they may also confer some favourable tribological properties (e.g., a thin chromium deposit on a threaded fastener may confer galling resistance).

The components on to which electrodeposits are applied are often made of steel. Electrodeposition of chromium and nickel from aqueous solution is conducted at low temperature (below 70°C), and so the properties of the steel are not influenced by thermal effects. However, steel substrates can suffer hydrogen embrittlement caused by absorption of atomic hydrogen generated during the plating process (by the reduction of hydrogen ions from water). In this regard, the main concern relates to plating of high strength steel components (tensile strength >1000 MPa); such steels

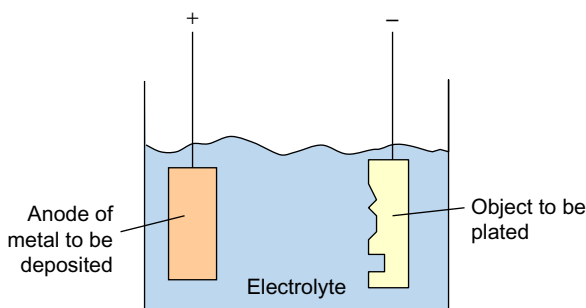


Fig. 7.15 Schematic diagram of the electroplating process

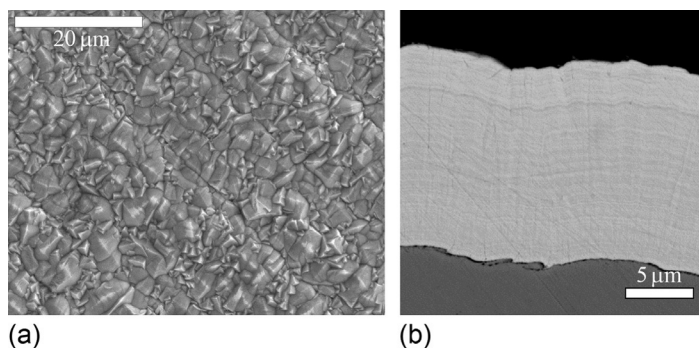


Fig. 7.16 SEM images of cobalt–tungsten alloy electrodeposit: (a) surface topography; (b) cross-section (from Weston, D.P., et al. 2009. *Wear* 267 (5–8), 934–943)

are heat treated following plating (typically between 190°C and 230°C with the time of treatment depending upon the strength of the steel) to remove the hydrogen and reduce the risk of hydrogen embrittlement.

One of the concerns during electroplating is the ability to produce a uniform coating on a component; a lack of uniformity results from an uneven current density across the surface, and this may be affected by the macroscopic shape of the component, or by much smaller features (such as threads on a threaded fastener). The ability to produce a uniform coating on components with complex geometries is known as the *throwing power* of the bath, and this can be controlled to some extent by the plating parameters and chemical additions to the bath. Despite this, in many systems, the problem of non-uniformity in coating thickness remains; for example, maintaining a uniform coating thickness on the root, flank and crest of the threads of a fastener remains a significant challenge in the bulk processing of components.

Chromium electrodeposits for tribological purposes are hard (850–1250 HV as deposited) and have the advantage of being corrosion resistant in many normal environments. They are plated from a chromic acid bath, with thicknesses typically from 10 to 500 µm, much thicker than the chromium plating used for decoration. The microstructure of electroplated hard chromium is extremely fine-grained (~10 nm), containing oxide inclusions and often significant microcracking. In oil lubricated systems (such as hydraulic actuators), this microcracking provides a means of lubricant retention on the component surface, further enhancing performance. The use of electroplated chromium is now being reduced because of concerns about worker and environmental safety and legislative pressures related to hexavalent chromium (Cr(VI)). Users of chromium plating are therefore having to consider alternative methods of surface engineering, as discussed further in [Section 8.6.5](#).

Nickel is also electrodeposited from acidic solution, producing coatings with a hardness up to 400 HV. Layers up to several millimetres thick can be plated, for example, to salvage worn parts.

A process also used to form nickel-based coatings for tribological applications is *autocatalytic* (or *electroless*) plating, in which the deposition of metallic nickel from a solution containing nickel ions and a reducing agent is catalyzed by the substrate surface and no external current is applied. Reducing agents containing phosphorus and boron are used to produce autocatalytic coatings of nickel containing about 10 wt.% phosphorus or 5 wt.% boron respectively. Electroless nickel-phosphorus has a hardness of ~ 500 HV, and nickel-boron ~ 700 HV; coatings can be applied to a wide range of metallic and non-metallic substrates. As for electroplating, autocatalytic plating can cause hydrogen embrittlement in steels and suitable precautions must be taken. Since the local rate of deposition is controlled simply by the chemistry of the bath and the surface, production of a coating of uniform thickness is relatively straightforward. Coating thicknesses depend on the material being plated, but typically range from 5 to 50 μm .

Both electrolytic and autocatalytic processes can be modified to incorporate fine particles (typically 0.5–5 μm in size) into the growing film; the particles are simply added to the bath as a suspension, with their volume fraction in the bath being used to control the concentration in the deposit. Hard particles such as diamond, silicon carbide or alumina, as well as solid lubricants such as PTFE, can be incorporated into the growing metallic matrix to provide coatings with enhanced wear resistance or lower friction. Several proprietary coating processes have been developed in this way.

Phosphating (or phosphate coating) is part of a group of *conversion coatings* and describes a process whereby an insoluble adherent phosphate coating is formed on a surface (normally of iron, steel, galvanized steel or aluminium) by immersion in an acidic solution of iron, zinc or manganese phosphate. These salts are only soluble in acidic solutions, and the reaction of the substrate with the acid locally raises the pH, resulting in precipitation of the phosphate salt on the component surface in the form of a coating. Phosphate coatings are commonly employed to provide wear resistance in situations where sliding motion is present and are normally used in conjunction with oils or greases, but can provide some level of lubrication in their own right.

Heavy manganese phosphate coatings can be used to prevent galling. The work-piece surface is seeded with manganese phosphate by dipping it into an agitated aqueous suspension of fine crystals of that compound; the component is then transferred to the phosphating bath where the seed crystals act as nuclei for the formation of a fine, compact coating. Such a phosphating treatment would be operated typically between 95°C and 100°C for a period between 10 and 15 min and would result in a coating thickness of between 4 and 8 μm .

7.4.2 Coatings deposited in the liquid state

When coatings are applied in the liquid state (or a mixed liquid–solid state), the solid coating is formed either by solidification from a melt, by chemical reaction or by evaporation of a solvent or carrier (normally aqueous or organic). Solvent/carrier-based systems are commonly used for organic coatings, but are also relevant to sol–gel coating which is currently being explored as a means to deposit wear resistant coatings.

However, most wear resistant coatings in this category are applied in the molten or semi-molten state.

There are main two classes of coating process in which the coating material is applied in the molten state, which we can conveniently consider together as *fusion processes*. These are *welding* and *thermal spraying*. In welding, the coating material is melted in contact with the substrate by techniques very similar to those used to join materials; the surface of the substrate is heated during the process to the melting temperature of the coating material, and some melting of the substrate occurs. In thermal spraying, the coating material is melted some distance from the substrate, and projected towards it in the form of fine molten droplets; the substrate remains relatively cool and the coating is formed and built up by solidification of the droplets as they strike the surface. These two types of method differ in the nature of the interface between the coating and substrate, the temperature of the substrate during the coating operation, and also in the nature and thickness of the coating that can practicably be applied.

Welding methods, often known as *hardfacing processes*, are best suited to the application of thick coatings, from 1 to 50 mm or more, of metallic materials. The coating can be formed on the surface of a component in *weld hardfacing* by any fusion welding method (such as manual metal arc (MMA), metal inert gas (MIG), tungsten inert gas (TIG), plasma transferred arc (PTA) etc.). Alternatively, a laser can be employed as the source of heat. The process is then known as *laser cladding* and is used to deposit coatings typically between 0.5 and 3 mm in thickness; the conditions used are shown in Fig. 7.6. In all these processes, the coating material is normally fed into the fusion zone either as a filler rod, consumable electrode, a previously applied paste or (in the case of laser cladding) as a powder.

In hardfacing processes, the surface of the substrate melts, mixes with and to some extent dilutes the coating material. The composition of the coating therefore differs from that of the feedstock material and this dilution can affect the properties of the deposited layer. The degree of dilution of the deposit varies with the welding method but can be as high as 30%. The high power input and short processing times in laser cladding mean that in this process, dilution can be restricted to much lower levels.

Coating delamination is rarely an issue with these processes, as substrate melting ensures that the interface between the substrate and coating is not abrupt and the bond between the two is very strong. As in conventional welding used for joining, the substrate will exhibit a heat affected zone. This may result in thermally-induced changes in the substrate properties (e.g., by the formation of martensite in steels or the uncontrolled ageing of aluminium alloys) which must be taken into account. Thermal stresses (due to thermal gradients and different coefficients of thermal expansion) and the low ductility of some deposits can also cause cracking of the coating in some cases.

Thick hardfacings are most suited to applications where the wear rate of the coating will inevitably be high, or where the applied loads will produce high stresses at a considerable depth beneath the surface, as discussed in Section 8.6. They are therefore widely used to provide resistance to high stress abrasive wear (see Sections 6.1 and 8.6.4): for example, in mining, quarrying and agricultural applications. Welding methods can be used to apply hard facings to localized areas; several techniques are

readily portable (such as manual metal arc, TIG and MIG welding) and can therefore be used to apply or repair coatings on-site. In contrast, weld hardfacing of components requiring higher precision is often performed with automated welding procedures before machining or grinding to a final finish.

The restriction that hardfacing materials must be molten at the temperature of the welding process imposes limitations on the materials that can be used. Ceramics, for example, are in general too refractory. For this reason, welded coatings are usually metallic or contain a fusible metallic binder; these same materials can also be fabricated by casting into bulk components, and are often also used in that form. Hardfacings can be grouped together, in order of increasing hardness, as austenitic manganese steels, hard (martensitic and tool) steels, cast irons with high carbide content (e.g., chromium-containing austenitic and martensitic irons) and cemented carbides (usually tungsten carbide/cobalt). Nickel- or cobalt-based alloys may also be used. The melting points of these materials limit the choice of substrate, since in the welding process its surface is heated to the same temperature as the coating material (typically $\sim 1400^\circ\text{C}$). Considerations of miscibility in the liquid state, and the possible formation of intermetallic compounds by interdiffusion, also constrain the choice, and in practice hardfacing methods are applied most widely to steels.

In the *thermal spraying* processes, the second group we shall consider in this section, many of the restrictions of the welding methods are removed. Here the temperature of the molten droplets of coating material is much higher than that of the substrate. Since the droplets solidify rapidly on striking the surface, to which the rate of heat transfer is much lower than in welding, substrate temperatures remain typically below 200°C . Not only can coatings of refractory metals and ceramics be applied by these methods, the choice of possible substrate materials is also much wider than for hardfacing. Coatings are deposited via a series of individual particles which form splats in the microstructure (Fig. 7.17). These splats often (but not always) have an oxide layer surrounding each one of them, and thus the splat-like microstructure can be a source of coating weakness in tribologically aggressive environments; cracking is often observed to proceed along splat boundaries.

Two generic sources of heat can be used in thermal spraying: combustion of a liquid or gaseous fuel (as in *flame spraying* and *high-velocity oxy-fuel (HVOF) spraying*)

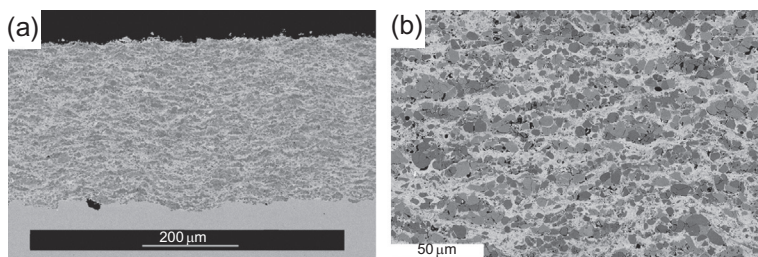


Fig. 7.17 Cross-sectional microstructure of HVOF-sprayed NiCr-Cr₃C₂ coating: (a) overview; (b) detail (from Bolelli, G., et al. 2016. Wear 358–359, 32–50)

or an electrical discharge (as in *electric arc spraying* and *plasma spraying*). The key features which influence the way in which the coating builds up are the particle temperature and particle velocity upon impact. The temperature of the particles is controlled by the temperature of the flame (or arc or plasma—as appropriate) and their residence time within it (which is controlled by the particle velocity and the point in the flame into which the feedstock material is fed), along with the distance between the flame and the surface on to which the coating is being deposited (termed the *standoff distance*). Similarly, the velocity of the particles depends on the speed of expansion of the flame, arc or plasma, the residence time and the standoff distance. Between them, the particle impact velocity and temperature control (amongst other things) the bond strength with the substrate, the coating porosity and the cohesive strength of the coating, all of which are key issues for its use in a tribologically aggressive environment.

In thermal spraying, the substrate does not melt, and the interface with the coating is therefore sharply defined. The surfaces of components are grit blasted before spraying to provide a mechanical key to facilitate bonding, but retention of blasting grit at the interface can act as a source of fatigue failure under certain loading conditions.

Even a strong interfacial bond may still fail if sufficient stress is applied to it, and an important source of such stress is differential thermal contraction between the coating and substrate as they cool together from the deposition temperature. Residual stresses from this source limit the usable thickness of many ceramic coatings to 0.2–0.5 mm; for metallic coatings on metallic substrates, the difference in expansion coefficients is usually less, and thicker coatings (up to several millimetres) are readily deposited by thermal spraying. In order to improve the interfacial adhesion of some sprayed coating materials (particularly ceramics) *bond-coats* are often used. These are intermediate strongly-adherent coatings that provide a strong mechanical key for the final sprayed coating. Nickel–aluminium powder is often used to form a bond-coat in thermal spraying; an exothermic reaction which occurs when the particles are heated in the spraying process raises their temperature still further and ensures good bonding. Molybdenum powder, which bonds readily to most substrates, is also used. In some thermal spraying processes, the composition of the powder can be changed progressively by mixing during spraying, so that coatings of graded composition can be produced, with a steady decrease in the concentration of the bond-coat material into the coating away from the surface of the substrate.

In *flame spraying* the coating material is fed as a wire or fine powder into a flame, usually of oxy-acetylene although propane or hydrogen is sometimes used as a fuel. [Figure 7.18](#) illustrates the principle of operation of a wire-fed spraying torch; in this design, an external supply of air is used to enhance the acceleration of the molten droplets once they are formed. The flame temperature is of the order of 3000°C; the droplets are heated to over 2000°C and strike the substrate at about 100 m s⁻¹. Metallic wires are readily sprayed; rods of ceramic materials and combustible tubes containing powders of ceramics or cemented carbides can also be used. In other designs of flame-spraying gun, the material is fed directly into the flame in powder form.

High velocity oxy-fuel (HVOF) spraying now has the largest share of the thermal spraying market. Fuel (commonly hydrogen, propylene or kerosene) is burnt in an internal combustion chamber, and the hot flame leaves the chamber through a nozzle

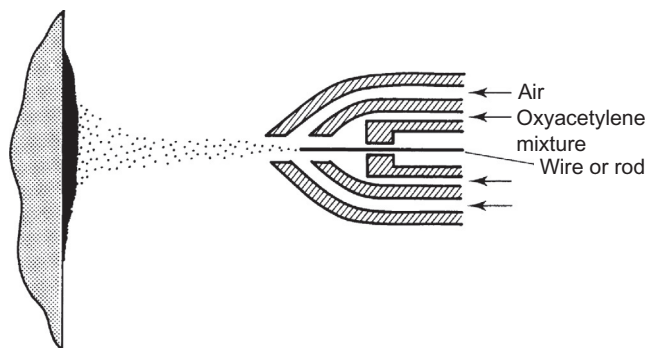


Fig. 7.18 Schematic diagram of a wire-fed flame spraying torch

at high pressure and expands supersonically. Coating material in powder form is injected into the flame. Subtleties in the design of the system have led to a wide variety being available commercially. Differences centre around the fuel type (gaseous or liquid fuel, with the former generally resulting in a lower velocity flame), nozzle design and the position of powder injection into the flame. Injection into the combustion chamber where the gases have low velocity results in high heat transfer to the particles, whereas injection into the nozzle where the gas velocities are higher results in less heat transfer. The gas temperature is similar to that in flame spraying (of the order of 3000°C) but the particle velocities are much higher (typically in the range of $600\text{--}900\text{ m s}^{-1}$). This high velocity results in reduced particle temperatures and increased impact velocities.

Another method of coating in which thermal and kinetic energy is derived from combustion and which gives well-bonded coatings of low porosity is the *detonation-gun* or *D-gun* process. In this proprietary method, powdered coating material is fed into a tube in which pulsed detonations of an oxy-acetylene gas mixture take place. These rapidly repeated explosions (4–8 times per second) give a flame temperature of $\sim 3000^{\circ}\text{C}$ and accelerate the molten droplets to $\sim 800\text{ m s}^{-1}$. At such high velocity, the droplets form a dense well-consolidated coating on the substrate, with porosity levels as low as 0.5%–1%.

The *electric arc spray* process is similar in principle to flame spraying, but uses an electrical heat source rather than combustion; an arc is struck between consumable wire electrodes giving a local temperature of $>4000^{\circ}\text{C}$, and a jet of gas (usually air) is used to project the molten droplets of the wire material at the substrate. In all these thermal spraying processes carried out in air, significant entrainment of air occurs during deposition, which leads to porosity in the final coating. Inevitably, some oxidation of the molten droplets also occurs in the spraying of metals. Porosity levels in the coating may be as high as 20%, reducing the mechanical strength, although for some tribological applications in lubricated sliding, controlled porosity can provide beneficial local reservoirs of lubricant.

Plasma-spraying methods are widely used; a typical design of plasma spray gun is shown schematically in Fig. 7.19. A plasma is formed in an inert gas (usually argon

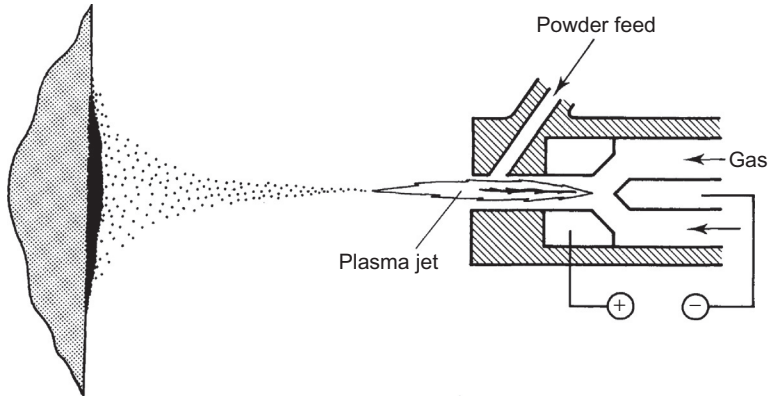


Fig. 7.19 Schematic diagram of a plasma spray gun

with a small proportion of hydrogen or helium) by a high-energy electric arc (typically 40 kW). The coating material is fed into the plasma as a fine powder, where it melts. The very high temperature of the plasma ($\sim 15000^{\circ}\text{C}$ with argon) enables a wide range of coating materials to be sprayed. Rapid expansion of the hot gas accelerates the molten droplets to $250\text{--}500\text{ m s}^{-1}$, and the combination of high melting temperature with high kinetic energy leads to coatings of lower porosity than those produced by flame spraying. Oxidation of metallic coating materials is also less pronounced, since the carrier gas is inert. However, plasma spraying is usually carried out in air (*air plasma spraying*, APS), and entrapment of gas still causes some porosity; levels are typically 1–10%, with the higher levels being associated with high melting point ceramic materials. Lower porosity can be achieved in *vacuum plasma spraying* (VPS), or *low-pressure plasma spraying* (LPPS), in which spraying takes place in a partial vacuum. Very low levels of porosity and correspondingly stronger and stiffer coatings result. Figure 7.20 illustrates the microstructures of alumina coatings produced by plasma spraying in air and in vacuum; the difference in porosity is evident.

The thermal spraying processes can be used to apply coatings of a wide range of materials, including metals and cemented carbides; refractory ceramic materials such as oxides can also be applied by the detonation-gun and plasma spraying processes which involve higher thermal and kinetic energies. In each case, the properties of the coating depend strongly on the process conditions. The two most important considerations are the strength of the coating and the strength of its bond with the substrate. Strong coatings are associated with complete fusion between the molten droplets on striking the surface, and with low porosity. As we have seen, the porosity is determined to some extent by the nature of the coating process and some methods (notably the HVOF, detonation-gun and LPPS/VPS processes) are capable of producing coatings with lower porosity than others, although this depends in all cases upon the operating parameters. The effect of coating microstructure can also be very significant. For example, pure alumina coatings deposited by LPPS have been found to show resistance to wear by solid particle erosion (with silica erodent) up to ten times

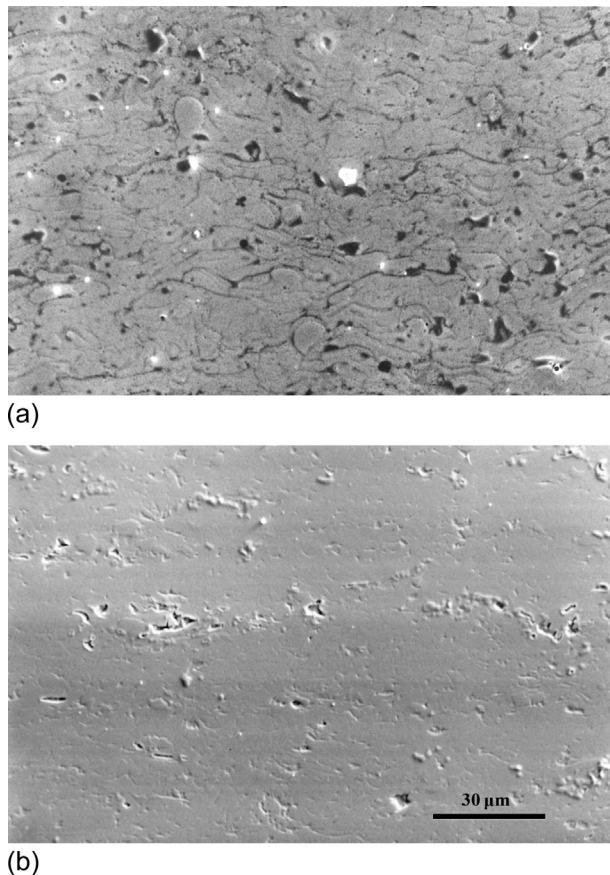


Fig. 7.20 Microstructures of alumina coatings produced by plasma spraying (a) in air (APS), and (b) in low pressure (LPPS). (from Zhang, X.S., Clyne, T.W., Hutchings, I.M., 1997. Relationship between microstructure and erosive wear resistance of plasma sprayed alumina coatings. *Surf. Eng.* 13 (5), 393–401)

that of a coating of the same composition plasma-sprayed in air. This difference can be largely ascribed to the difference in porosity between coatings produced by the two methods. Thermally sprayed coatings can also crack and debond cohesively along the inter-splat boundaries, which are generally weak areas within the microstructures.

The third group of coatings to be considered in this section are *polymer-based coatings* that are deposited as either liquid paints or slurries, or as viscous pastes (putties), and solidify by evaporation of a solvent or by chemical reaction (e.g., cross-linking). The polymers are typically epoxy resins and polyurethanes. Engineering paints can be applied by a variety of techniques such as spraying, brushing, dipping or printing and are filled either with hard ceramic phases (to resist abrasion or erosion) or with materials such as graphite, MoS₂ or PTFE to provide a dry-film lubricant coating. Substrate surfaces are typically prepared by grit-blasting followed by a chemical pre-treatment

(such as phosphating or anodizing), depending upon the metal being coated. Service temperatures of these paints are typically up to 300°C. Dry film lubricant paints and slurries are used in applications as diverse as piston rings in internal combustion engines, dovetail joints in fan blade-disc assemblies in turbofan aero-engines and to prevent squeaking at metal-to-plastic joints in automobile interior assemblies. Pastes, on the other hand, are generally employed in applications requiring abrasion or erosion resistance, commonly for in-situ repair of failed components. They contain high proportions of ceramic fillers and are trowelled on to previously grit-blasted surfaces; the most common applications are in pump impellers, cyclones, chutes and hoppers handling bulk minerals and particle-laden liquid flows.

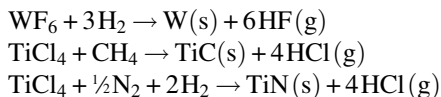
7.4.3 Coatings deposited from a vapour

Methods in which a coating is formed from the vapour phase can be divided into two groups, described as *chemical vapour deposition* (CVD) and *physical vapour deposition* (PVD). We shall examine these two groups in turn.

Chemical vapour deposition

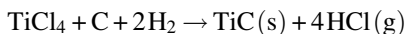
Chemical vapour deposition involves thermally-induced chemical reactions at the surface of a substrate, with reagents supplied in gaseous form. If the substrate is heated, these reactions may involve the substrate material itself, but often they do not.

The simplest CVD process involves the pyrolytic decomposition of a gaseous compound on the substrate to provide a coating of a solid reaction product. Although methods have been developed to deposit many metals as well as alumina in this way, the organic metal compounds used are hazardous and other methods tend to be preferred. Reaction of metal halides, for example, with hydrogen, nitrogen or methane is commonly used to provide coatings of pure metals or their nitrides or carbides. For example, the following reactions are used to produce solid coatings of tungsten metal, titanium carbide and titanium nitride respectively:



With mixtures of metal halides, hydrogen, oxygen, nitrogen, hydrocarbons and boron compounds, coatings of a wide range of metals and their oxides, nitrides, carbides and borides can be formed by CVD.

Although the reactions listed above involve only gaseous reagents, the substrate material in some cases also plays a significant role. For example, the rate-controlling step in the growth of a TiC layer on carbon steels or cemented carbides from TiCl₄ is the reaction



which involves carbon from the substrate. This reaction can lead to decarburization of the substrate immediately beneath the coating, and this must be taken into account in selecting a suitable substrate material.

The decomposition processes that form the coating require an energy source, and many variants on the CVD process exist, all of which result in differences in the resulting coatings. Typical energy sources are the hot surface of the component being coated, a hot filament, electromagnetic radiation (radio frequency or microwave) or a DC plasma. In some cases, combustion of the process gases themselves can be used as the energy source, such as combustion of ethyne (acetylene) in the CVD deposition of diamond-like carbon (DLC). When the substrate itself is heated, temperatures between 600°C and 1100°C are typical and these will generally result in significant (and generally undesirable) microstructural changes; steels, for example, will often be heated into the austenite phase region and the coating process must then be followed by suitable heat treatment to optimize the properties of the substrate. Many process developments have been driven by the desire to reduce the deposition temperature, and in processes such as plasma-enhanced CVD, the temperature of the substrate does not need to be increased above ambient.

Equipment for CVD coating usually consists of a reaction vessel, as shown in Fig. 7.21, and associated gas-handling equipment. The CVD process is controlled by varying (i) the temperature or energy input into the process and (ii) the composition and pressure of the gas mixture. The optimum deposition rate is generally achieved at a gas pressure below atmospheric; the rate depends strongly on the process type and conditions along with the material being deposited, but is typically $1\text{--}20\text{ nm s}^{-1}$. Final thicknesses of up to $10\text{ }\mu\text{m}$ are commonly deposited, but thicker layers may be formed for some applications. Coatings formed by CVD often have a columnar microstructure, although the initial deposit is sometimes equiaxed; the grain size and microstructure depend strongly on the process conditions. In many cases, the most desirable microstructure may not be produced at a high deposition rate, and a compromise

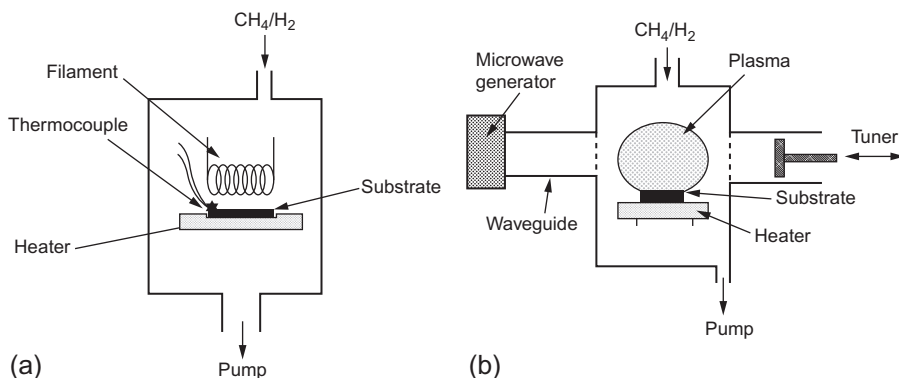


Fig. 7.21 Two types of CVD reactor (a) hot filament CVD reactor; (b) microwave plasma-enhanced CVD reactor (from May, P.W., 1995. *Endeavour* 19, 101–106)

may have to be struck between a high growth rate, desirable for economic reasons, and a fine-grained microstructure with the most desirable tribological properties.

In cases where the component surface is at a high temperature, appreciable inter-diffusion can occur between coating and substrate, and such CVD coatings therefore show strong adhesion. Inter-diffusion can, however, lead to problems in some systems with the formation of brittle intermetallic compounds at the interface which result in poor mechanical properties, and potentially to coating delamination. To avoid this problem, undercoats of other materials which do not form such compounds with either the substrate or the final coating can be used as diffusion barriers. For some applications, a sequence of several CVD coatings can be applied, which not only provide chemical compatibility and reduce the problems of inter-diffusion with the substrate, but also provide a gradation of mechanical, chemical and physical properties. For example, cemented carbide cutting tools coated with a three-layer sequence of titanium carbide, titanium carbonitride and titanium nitride show significantly longer lives than tools with a single layer coating; multilayer coatings with as many as ten components have also been used in tool applications (see [Section 9.4.1](#)).

Physical vapour deposition

In the processes collectively known as physical vapour deposition (PVD), the coating material is transported to the surface in atomic, molecular or ionic form, derived by physical rather than chemical means from a solid, liquid or gaseous source. Chemical reactions may, but need not, occur at the surface of the substrate, which is typically at a temperature between 50°C and 500°C. At low temperatures, the coating process may not significantly change the microstructure or properties of the underlying substrate material.

The simplest PVD process is *evaporation*, which has been used for many years to produce coatings on glass lenses and other optical components. The principle of the method is illustrated in [Fig. 7.22\(a\)](#). The coating is formed by the evaporation of material from a molten source, often heated by an electron beam. The process is carried out in vacuum ($\sim 10^{-3}$ Pa), and is restricted to coating materials of relatively low boiling point that do not dissociate on heating, such as metals. If a suitable reactive gas is present (e.g., methane, oxygen or nitrogen), chemical reactions can be induced with the evaporated species on the substrate surface, and coatings of compounds can be laid down. The process is then known as *reactive evaporation*.

In evaporative processes, the kinetic energy of the atoms striking the substrate results only from their thermal energy. On a cool substrate they lose this energy quickly, and little mixing occurs between the surface atoms of the substrate and those of the coating. Adhesion is therefore weak, and is further reduced by any contaminants (adsorbed gases etc.) on the surface. Some improvement in adhesion can be achieved by heating the substrate during or after coating to allow some inter-diffusion, but the bond strengths of evaporated coatings are often too low for tribological purposes. Another limitation of evaporative methods is that the coating is transported along a line of sight from source to substrate. An even coating is difficult to achieve and the coating of complex shapes is sometimes impossible.

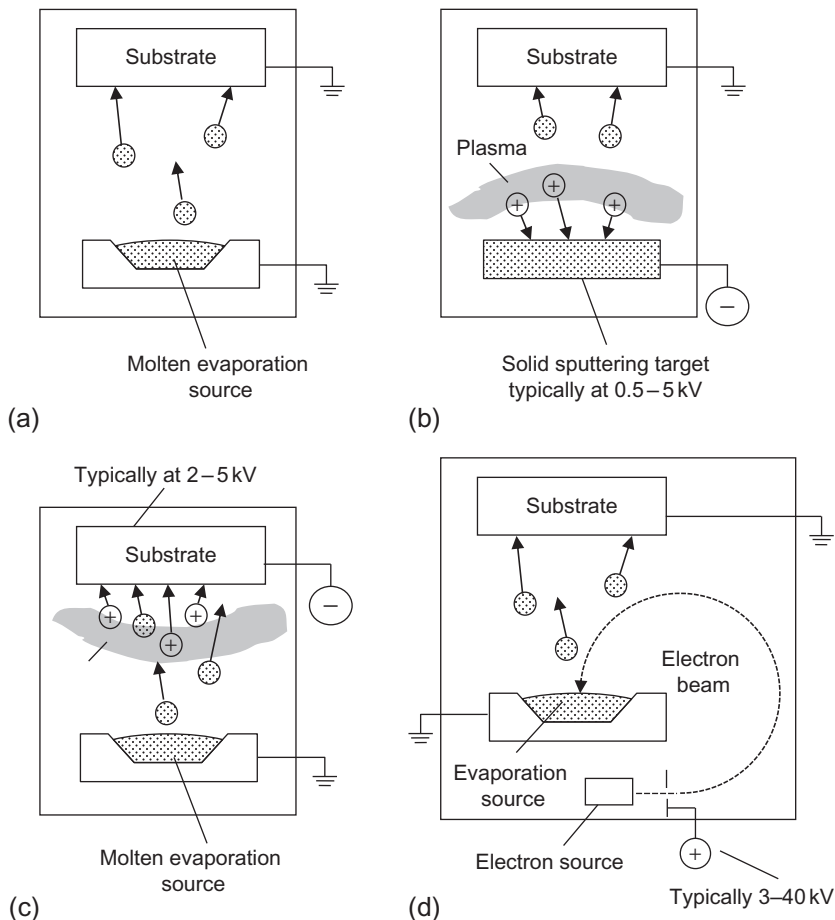


Fig. 7.22 Schematic diagrams of generic physical vapour deposition (PVD) processes: (a) evaporation; (b) sputtering; (c) ion plating; (d) electron-beam evaporation

In *sputtering*, the energy to transport material from the source to the substrate is supplied by energetic heavy gas ions. The process is illustrated in Fig. 7.22(b). Positive ions are formed in a glow discharge in a low pressure gas (usually argon at 0.1–10 Pa) and strike the solid source material, which is negatively biased, with energies up to a few keV. This ion bombardment causes atoms to be sputtered from the target which then strike the substrate a short distance away. With a direct current source, the process is limited to coatings of conducting materials, but the use of an alternating (radio-frequency) source removes this limitation. The diode arrangement shown in Fig. 7.22(b) is the simplest; higher sputtering rates can be gained by increasing the ionization efficiency, and magnetron sources are often used for this purpose. Diode systems give coating rates typically of $1\text{--}100\text{ nm min}^{-1}$, while with a magnetron source the rate may be as high as $2\text{ }\mu\text{m min}^{-1}$.

Sputtering is a versatile method that can be used to form coatings of a wide range of materials. The process may also be modified by the incorporation of a reactive gas (*reactive sputtering*), and this process is widely used to deposit oxides, nitrides and carbides. Because atoms sputtered from the source have much higher energies (typically tens of eV) than the thermal energies of evaporated atoms (fractions of 1 eV), they undergo more inter-diffusion and mixing with the surface atoms of the substrate, resulting in stronger adhesion than exhibited by evaporated coatings.

The third important PVD process is *ion plating*, in which atoms or molecules of the coating material are evaporated from a hot source into a glow discharge, usually in argon at a pressure of 0.1–1 Pa. [Figure 7.22\(c\)](#) illustrates the process. The vapour source may be heated resistively or by an electron beam, or may be provided by an arc struck against a solid source material. Some atoms of the vapour become positively ionized in the plasma and are accelerated towards the substrate, which is held at a negative potential of 2–5 kV. Others move under their thermal velocities, enhanced by collisions with the energetic argon ions, and also strike the substrate, which is also being bombarded continuously by argon ions. The high energies of the atoms at the substrate surface, together with the scattering provided by collisions with the argon ions, provides a uniformly distributed coating with good adhesion through mixing at the interface. Cleaning of the substrate before coating is readily carried out by sputtering in the glow discharge. Deposition rates for metals of several $\mu\text{m min}^{-1}$ for metals can be achieved by ion plating. As with the other PVD processes described previously, incorporation of a reactive gas into the system allows compounds to be formed at the surface. The process is then termed *reactive ion plating*. This process is used extensively to form coatings of titanium nitride, with the titanium metal being evaporated into a glow discharge in a mixture of nitrogen and argon gases surrounding the workpiece, which reaches a temperature of about 400°C. The process can be applied, for example, to tool steels and other hard substrate materials that are microstructurally stable at this temperature, without causing the substrate to become softer. Analogous coatings such as TiAlN and TiCN are deposited by similar methods.

In a further process, *electron-beam PVD* (EBPVD), the coating material is evaporated from a solid target by the action of a high energy electron beam, with an accelerating voltage typically from 3 to 40 kV ([Fig. 7.22d](#)). The method can give high coating rates, and is widely used to produce *thermal barrier coatings* (TBCs) 100–500 μm thick on gas turbine blades for aircraft propulsion and power generation. High thermodynamic efficiency in a gas turbine engine is favoured by a high combustion temperature, and the primary function of a thermal barrier coating is to reduce the rate of heat transfer from the hot combustion gas into the blade, allowing the blade which is internally cooled to operate in a gas stream that is well above the melting temperature of the blade material. EBPVD TBCs are made from ceramics (typically yttria-stabilized zirconia) and have a columnar microstructure, with the bonds between adjacent columns being relatively weak. If the coating is exposed to tension perpendicularly to the columns, for example by thermal expansion mismatch with the substrate, this microstructure can provide a level of strain tolerance that avoids cracking, by reversible separation of the columns. [Figure 7.23](#) shows an example

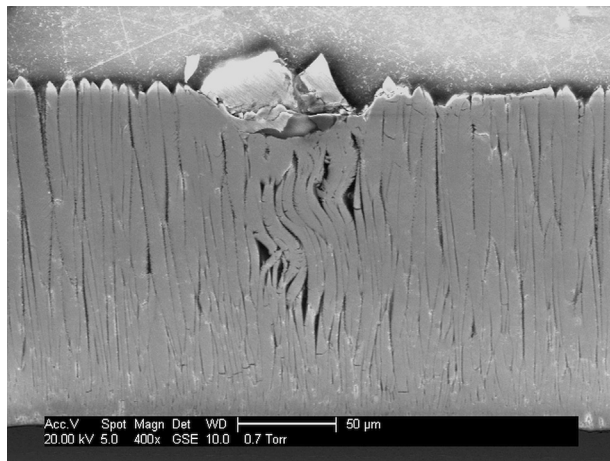


Fig. 7.23 Cross-sectional SEM micrograph illustrating column buckling in an EBPVD thermal barrier coating following large particle impact (from Wellman, R.G., Deakin, M.J., Nicholls, J.R., 2005. *Wear* 258 (1–4), 349–356)

of an EBPVD TBC, illustrating the columnar structure and also showing its response to impact by a single hard particle, as might occur on ingestion of sand into an aero-engine.

All the processes discussed above and grouped together as PVD methods have a good deal in common. Intermediate techniques can be devised that incorporate features of more than one of the simpler processes described here, and it is possible to combine different processes, either simultaneously or sequentially, to treat the substrates in a single vacuum chamber. In this way, the properties of coatings and interfaces can be tailored for particular applications. In order to achieve even coatings, the components are often mounted in jigs that rotate about one or more axes within the deposition chamber. Multilayer coatings can be deposited by arranging for the components to travel repeatedly past two or more different target materials during the deposition treatment. Considerable further development of such processes will undoubtedly occur.

DLC coatings

Diamond-like carbon (DLC) coatings are attractive tribological materials due to their high hardness, low coefficients of friction against themselves and other materials (see [Section 3.6](#)) and low rates of wear. They are now widely used in applications ranging from automotive components (e.g., camshafts and tappets) to biomedical implants, although it should be noted that due to the high solubility of carbon in steels (and its reactions with other metals), DLC coatings are rarely used to coat cutting tools. The low cost of production has even resulted in their application to blades in disposable razors. The term DLC covers a broad family of amorphous carbon coatings that are characterized by their predominant bonding type (either trigonal sp^2 bonding (a-C)

Table 7.2 Characteristics of DLC films and other carbon-based materials (from Robertson, J., 2002. Mater. Sci. Eng. R 37, 129–281)

	sp ³ (%)	H (%)	Hardness (GPa)
Diamond	100	0	100
Graphite	0	0	
Glassy C	0	0	3
Evaporated C	0	0	3
Sputtered C	5	0	
ta-C	80–88	0	80
a-C:H hard	40	30–40	10–20
a-C:H soft	60	40–50	<10
ta-C:H	70	30	50
Polyethylene	100	67	0.01

or tetrahedral sp³ bonding (ta-C)) and hydrogen content, as illustrated in Fig. 7.24. Table 7.2 lists the sp³ content, hydrogen content and hardness of some typical DLC coatings, as well as diamond, graphite and other carbon-based materials for comparison. The hardness of DLC can vary from around 80 GPa (close to the hardness of diamond itself) down to <10 GPa (less than the hardness of many ceramics and similar to a martensitic steel). DLC coatings can be deposited by a wide variety of both physical and chemical vapour deposition techniques including magnetron sputtering and plasma enhanced CVD. They exhibit high bond strengths to carbide-forming metals, and bonding to silicide-forming metals can be enhanced by the deposition of a very thin (2–4 nm) silicon interlayer. As discussed in Section 3.6, the coefficient of friction associated with sliding against a DLC coated surface depends upon the

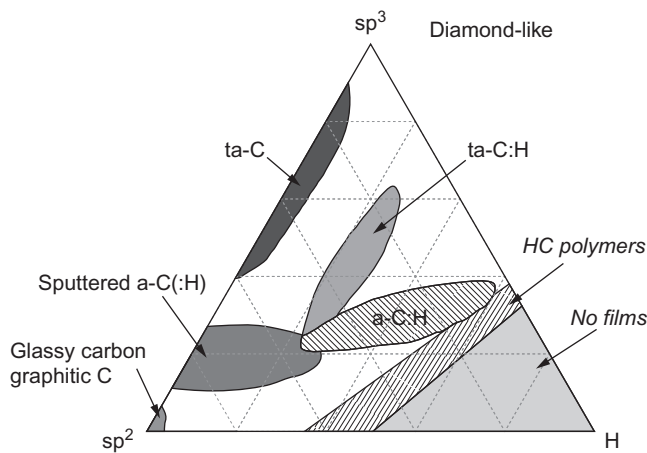


Fig. 7.24 Ternary phase diagram for DLC films as a function of sp², sp³ and hydrogen contents (from Robertson, J., 2002. Mater. Sci. Eng. R 37, 129–281)

local environment (most significantly, the presence of oxygen and water) and upon the nature of the DLC film itself (bond type and hydrogen content) and ranges from 0.005 to 0.5 or more. When low friction is observed, graphitized transfer layers are found on the surfaces, indicating the importance of these transfer layers in this context.

CVD and PVD processes using a hydrocarbon gas (typically methane or ethyne) are often employed to generate softer DLC films with significant hydrogen contents. PVD processes using solid carbon precursors lead to the formation of hydrogen-free films, with the sp^2 to sp^3 ratio being governed by the process type; for example, conventional sputtering produces films dominated by sp^2 bonding whereas unbalanced magnetron sputtering generates harder films dominated by sp^3 bonding. In all process types (both PVD and CVD), considerable variation in film properties and growth rate occurs with process conditions. Numerous variants are possible: DLCs are sometimes doped with other elements (such as nitrogen, sulfur, tungsten or titanium) which play a significant role in determining the properties of these materials. Multi-layer DLC-tungsten carbide coatings (the individual layers being of the order of nanometres thick) are commercially available and show high durability in oil-lubricated conditions.

7.4.4 Coatings deposited in the solid state

For completeness, it should be mentioned that there are a few niche techniques in which coatings are deposited in the solid state. These include *cold-gas dynamic spraying* in which particles are accelerated to very high velocities in a gas stream (with little or no external heating) and bond with a substrate upon impact; and *friction surfacing* where a rod is rotated under pressure against a substrate and traversed across it, whereupon material from the rod is deposited as a coating. Both processes depend upon the disruption of oxides on the surfaces of both the coating and substrate materials by heavy local plastic deformation, resulting in the formation of a metallic bond.

Cold-gas dynamic spraying (otherwise known as *cold spraying* or *kinetic metallization*) involves the deposition of coatings from a particulate feedstock. Particles are accelerated to very high velocities (of the order of 600 m s^{-1} depending upon the powder type) in a supersonic gas stream (originally helium, but heated nitrogen is now preferred for economic reasons) that is created by the use of a de Laval (convergent–divergent) nozzle. The particles deform heavily on impact with the substrate and a coating can be built up. The technology is still in its infancy and while it has been adopted in certain applications, it is not currently being widely applied commercially to deposit wear-resistant coatings. However, the ability to coat lightweight structural materials such as aluminium or titanium with more wear resistant surfaces, and the ability to conduct repairs without thermal damage to the substrate may be a major driver for the expansion of this technology.

In friction surfacing, which is similar to *friction stir welding*, the resulting coating is relatively flat and commonly deposited to a thickness between 0.5 and 3 mm depending upon operating parameters and materials. The coverage of the coating

depends upon the relative movement of the substrate and the depositing rod. Friction surfacing results in significant heating of the materials. The rod reaches high temperatures whilst being deformed and there are some similarities to a hot forging process. Friction surfacing has been employed to produce wear-resistant coatings for applications such as machine knives.

7.5 TRIBOLOGICAL BEHAVIOUR AND ITS EVALUATION

The tribological response of a surface engineered component may depend simply on the properties of the surface engineered layer, or it may be controlled by properties of the interface or even of the substrate, depending on the severity of the tribological conditions, the thickness of the surface engineered layer and the properties of the interface. For example, if the substrate yields plastically due to the applied load, a brittle surface engineered layer will tend to fracture. Under given loading conditions, the tendency for the substrate to yield will decrease with increasing substrate hardness and with increasing thickness of the surface engineered layer. Interface debonding can result from a combination of low bond strength and severe tribological conditions.

Detailed modelling of the stress field in the coating and substrate, together with knowledge of the properties of the coating, the substrate and the interface (including any residual stresses present) can be used to predict the mechanical response of the system. But to evaluate and compare the tribological behaviour of different coatings and for quality assurance purposes, empirical investigation of the behaviour of a coated system is still invaluable. Many laboratory tests used for bulk materials as described in [Sections 5.2, 6.3.5 and 6.4.5](#) can be applied directly to surface engineered components, especially in the case of thick surface engineered layers where, for example, measurable mass losses occur in a wear test before the coating is penetrated. However for thin coatings, this is not always the case, and more suitable tests are available. One of the most widely used is the *scratch test*, in which a Rockwell C diamond cone indenter (see [Appendix A](#)) is dragged across the coating surface under a load that is increased linearly with distance. Coating damage can be assessed in terms of cracking and delamination, and samples compared in terms of their resistance to damage. [Figure 7.25](#) illustrates the test method and typical damage that is seen in CVD and PVD films. The coating is characterized by the critical loads at which certain types of damage are first seen: in this case cracking, followed by edge-chipping, followed by cross-track cracking.

Some wear tests are particularly suitable for thin coatings. One example is the micro-scale abrasion test shown in [Fig. 6.27\(e\)](#), in which a fine abrasive slurry (typically containing 4 μm diameter particles) is fed into the gap between a rotating ball and the sample under test. Provided a sufficiently low load is used, the abrasive particles become entrained into the ball-surface interface and generate a wear scar that has the same spherical radius as the ball. The wear rates (and hence specific wear rates) of both the coating and the substrate can be determined from simple geometrical

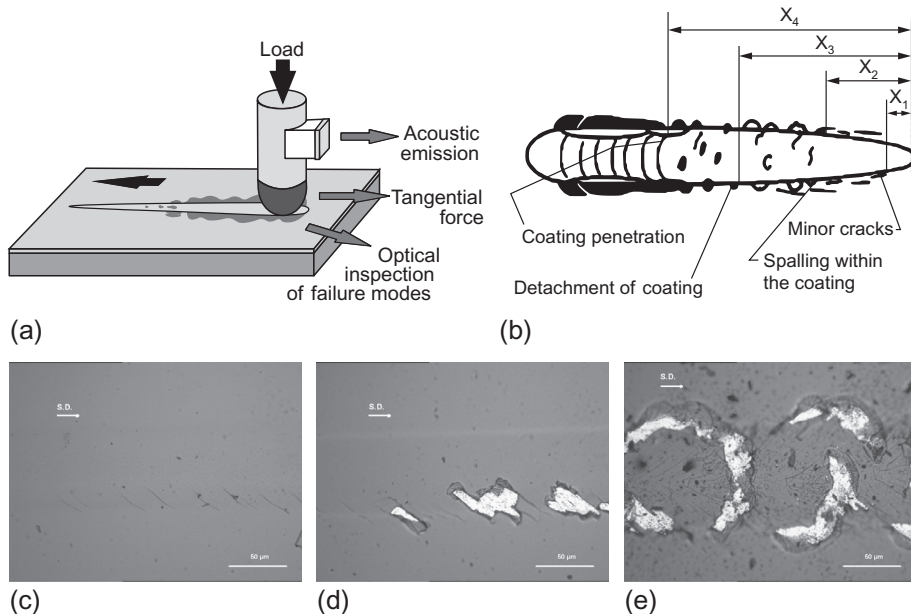


Fig. 7.25 Scratch testing of coatings. (a) schematic diagram of the test method; (b) failure events that can occur; (c)–(e) examples of events on carbon film showing (c) cracking at 15 N load, (d) edge chipping at 20 N load, and (e) cross-track cracking at 30 N load (from Jennett, N.M., Gee, M.G., 2006. Mechanical testing of coatings. In: Mellor, B.G. (Ed.), Surface Coatings for Protection Against Wear. Woodhead Publishing)

measurements of the outer diameter of the wear scar and the diameter of the interface between the coating and the substrate within the wear scar. The same test method, sometimes known as *ball-cratering* or *dimple grinding*, can also be used as a simple method to determine the coating thickness. The test, while not non-destructive, results in a comparatively small wear scar (typically ~ 1 mm in diameter) that can often be located in a non-critical area of the component.

SOURCES AND FURTHER READING

- ASM, 1991. Metals Handbook, 10th ed. Friction, Lubrication and Wear Technology, vol. 4. ASM International.
- Budinski, K.J., 1988. Surface Engineering for Wear Resistance. Prentice Hall.
- Bull, S.J., Rickerby, D.S., 1991. Evaluation of coatings. In: Rickerby, D.S., Matthews, A. (Eds.), Advanced Surface Coatings: A Handbook of Surface Engineering. Springer.
- Child, H.C., 1980. Surface Hardening of Steel, Engineering Design Guide No. 37. Oxford University Press.
- Dearnley, P.A., 2017. Introduction to Surface Engineering. Cambridge University Press.

- Erdemir, A., Donnet, C., 2006. Tribology of diamond-like carbon films: recent progress and future prospects. *J. Phys. D Appl. Phys.* 39, R311–R327.
- Gabel, M.K., Donovan, D.M., 1980. Wear resistant coatings and treatments. In: Peterson, M.B., Winer, W.O. (Eds.), *Wear Control Handbook*. ASME, New York, pp. 343–371.
- Hainsworth, S.V., Uhure, N.J., 2000. Diamond like carbon coatings for tribology: production techniques, characterisation methods and applications. *Int. Mater. Rev.* 52 (3), 153–174.
- HMSO, 1986. *Wear Resistant Surfaces in Engineering*. UK Department of Trade and Industry.
- Peterson, M.B., Ramalingam, S., 1981. Coatings for tribological applications. In: Rigney, D.A. (Ed.), *Fundamentals of Friction and Wear of Materials*. ASM, Metals Park, pp. 331–372.
- Ramalingam, S., 1980. New coating technologies for tribological applications. In: Peterson, M.B., Winer, W.O. (Eds.), *Wear Control Handbook*. ASME, New York, pp. 385–411.
- Towler, B., 1978. *Flame Deposition, Engineering Design Guide No. 25*. Oxford University Press.

QUESTIONS FOR CHAPTER 7

Question 7.1

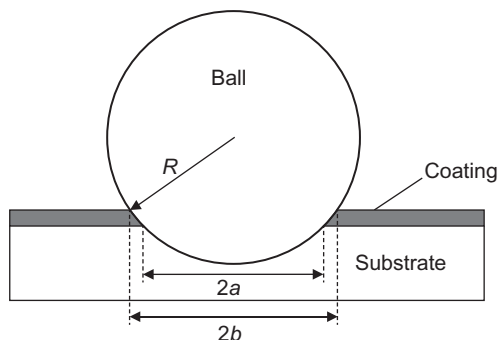
During the coating or surface treatment of components, certain problems can arise that affect either the coating or the component being coated. For each of the following examples, describe the problem and its origin. Then suggest methods that might be used to eliminate the problem.

- (a) Hydrogen embrittlement of high strength steels in electroplating.
- (b) Variation of coating thickness over a component following electrodeposition.
- (c) High levels of porosity in a thermally sprayed coating.
- (d) Debonding of a thermally sprayed coating following spraying.

Question 7.2

The micro-scale abrasion test described in [Section 6.3.5](#) and shown in [Fig. 6.27\(e\)](#) can be used to form a crater or ‘dimple’ in the surface of a coated component from which the thickness of the coating can be determined. The method involves measuring the chordal diameter of the overall crater ($2b$) and of the boundary between the coating and the substrate ($2a$), by optical microscopy. The geometry is shown in the diagram below.

- (a) Derive an expression for the thickness of the coating t in terms of the chordal radii b and a and the radius R of the ball. Make the assumption that the depth of the crater is much less than R .
- (b) This method was used with a 25 mm diameter ball to measure the thickness of a PVD coating on a steel substrate. The outer chordal diameter of the wear crater was 1.20 mm and the chordal diameter of the coating-substrate boundary was 1.07 mm. Calculate the thickness of the coating.
- (c) Discuss the limitations of this method for measuring coating thickness.



Question 7.3

- (a) Explain why, while the surface of a steel component can be hardened by laser treatment in a matter of seconds, hardening by carburizing can take several hours.
- (b) Estimate the depths of the hardened layers formed in a steel component by the following two processes. In each case, state your assumptions clearly.
 - (i) Laser-hardening by a laser beam 1 mm in diameter that passes over the surface at a speed of 0.1 m s^{-1} .
 - (ii) Carburizing for a period of 4 h at a temperature of 900°C .

[Assume that the steel has a thermal conductivity of $50 \text{ W m}^{-1} \text{ K}^{-1}$, a density of 7800 kg m^{-3} , and a specific heat capacity of $470 \text{ J kg}^{-1} \text{ K}^{-1}$. The diffusion coefficient D for carbon in austenite at temperature T is given by $D = D_0 \exp(-Q/RT)$, where $D_0 = 10^{-5} \text{ m}^2 \text{ s}^{-1}$, $Q = 135 \text{ kJ mol}^{-1}$ and $R = 8.31 \text{ J mol}^{-1} \text{ K}^{-1}$.]

Question 7.4

Surface engineering methods are proposed for the following applications to enhance the tribological performance. For each case,

- (i) Suggest a suitable surface engineering process (and material where appropriate) that might be used.
 - (ii) Describe and justify an appropriate laboratory test method to aid in coating selection.
 - (iii) Identify the benefits and shortcomings of the test method selected.
 - (iv) Identify further information that you would need to refine your test methodology.
-
- (A) A layer approximately 50 mm thick to provide wear resistance for a steel paddle in a concrete processing plant.
 - (B) Treatment of a steel die in a plastic injection moulding operation to reduce problems from adhesion between the die and the product
 - (C) Protection of helicopter engine turbine blades from erosion damage by ingested sand.
 - (D) Prevention of wear between the faces of a bolted joint in an application where vibration is present.
 - (E) Provision of scratch resistance to plastic lenses for eye-glasses (spectacles).

Question 7.5

The erosion durability of a thin PVD coating for use in an aircraft engine is required. An erosion test is carried out with a gas-blast erosion test with silica sand as the erodent and an impact velocity of 70 m s^{-1} . The radius of the circular scar from which the coating has been removed from the substrate is measured as a function of the mass of erodent delivered. The results are as follows:

Mass of erodent/g	Scar radius/mm
10	2.5
20	3.0
60	4.0
100	4.5

The scar radius, r , is related to the total mass of erodent, m , as follows:

$$r = \alpha \ln m - \alpha \ln (2\pi\alpha^2 Q_c)$$

where α is a constant and Q_c is the critical dose for coating removal under those conditions (mass of erodent required per unit area of coating removal). Calculate the critical dose, Q_c , for coating removal under these conditions.

Design and selection of materials for tribological applications

8

8.1 INTRODUCTION AND GENERAL PRINCIPLES

Wear and friction, as factors in the performance of an engineering system, are often neglected in the design process. Yet as we saw in [Chapter 1](#), wear leads in many cases to important maintenance or replacement costs and to associated loss of production or availability, which also carry economic penalties. In addition, friction often causes reduced efficiency in the operation of a machine and can lead to unacceptable temperature rises, and wear can result in increased vibration before it has progressed far enough to justify replacement or overhaul. Premature, and especially unpredicted, tribological failure can in some cases lead to damagingly high costs for the product supplier and even to the tragic consequences of human injury. In the case of medical, power-generation and transport applications these risks are especially acute. Tribological performance should therefore always be considered early in the design process for any mechanical system, and not just as an afterthought once problems have emerged in trials or even in service. It should be integrated into the process alongside geometrical and mechanical design and materials selection, as illustrated in [Fig. 8.1](#).

We have discussed methods for reducing friction in [Chapters 3](#) and [4](#); our main focus here is on wear. The designer has two main concerns: to establish whether significant wear will occur during the design life of the machine or component, and if so, to take steps to reduce the rate of wear to acceptable levels within the economic and other constraints imposed on the design. Here we include within the term ‘wear’ all types of tribological degradation, including rolling contact fatigue as well as surface roughening leading to excessive vibration. The meanings of the concepts ‘significant wear’ and ‘acceptable levels’ clearly depend on the details of the system being designed, as also do the steps that can be taken to reduce wear. In order to propose methods of reducing wear, the designer needs to establish the mechanism by which it will occur, and also to understand the factors that control wear by that mechanism. One of the aims of the earlier chapters of this book has been to provide a basic understanding of wear mechanisms.

[Chapters 5](#) and [6](#) have described several types of wear that commonly occur in mechanical systems. Where solid surfaces slide over each other, wear can occur by different mechanisms, depending on the state of lubrication, the nature of the sliding surfaces, the chemical environment and the operating conditions, such as normal load and sliding speed. As we have seen, the relative displacement of the surfaces need not be intentional or at all large: even when they are nominally fixed, slight relative oscillatory motion may occur leading to fretting and associated wear. Wear under all these conditions has been discussed in [Chapter 5](#).

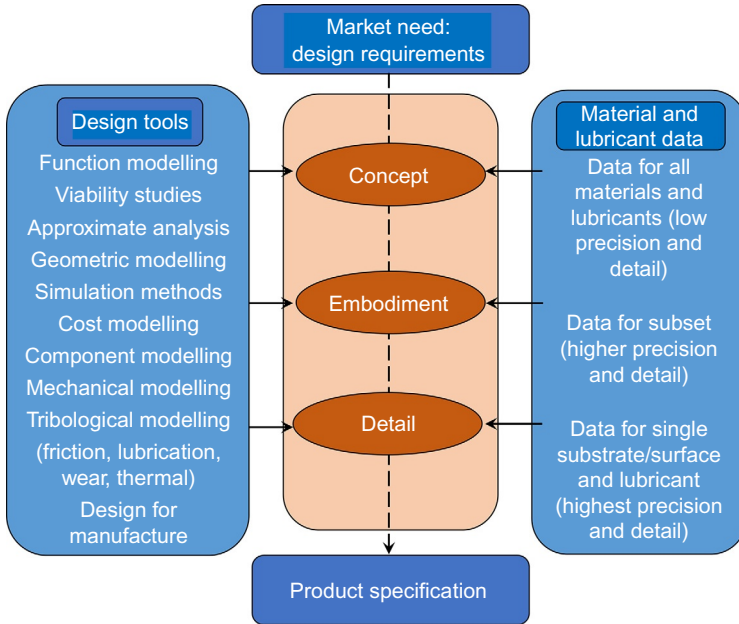


Fig. 8.1 The process of mechanical design, including consideration of tribological performance and surface engineering (adapted from Ashby, M.F., 2005. *Materials Selection in Mechanical Design*, third ed. Elsevier)

When the contact predominantly involves rolling rather than sliding, as in rolling element bearings, surface damage and wear occur by rolling contact fatigue, which is discussed in [Section 9.2](#) together with the selection of materials for bearing applications. If hard particles are present between sliding surfaces then the wear mechanisms are somewhat different and were discussed in [Chapter 6](#), together with the similar processes that occur when hard particles impinge on to a surface, causing wear by erosion.

In the present chapter we shall review the methods available to estimate wear rates in design, and then examine the measures that can be taken to reduce wear in a range of situations.

8.2 ESTIMATION OF WEAR RATES

Three methods are available to the designer to predict rates of wear. The first, not a useful method in initial design but nevertheless valuable as a means of predicting the ultimate lifetime of a system in operation, is to measure the wear rate in an actual system in service. For example, measurements of erosive wear in a slurry transport pipeline in use can be employed to estimate the time (or total

material throughput) before the pipe wall will be reduced to a certain critical thickness (perhaps determined by its ability to withstand internal pressure, in conjunction with a suitable safety factor), and hence to predict its useful life. The wear in a bearing bush in a truck suspension or an aircraft control surface can be measured after a certain period of use measured in distance travelled or flying hours, and the result extrapolated to predict the point at which the bush will need replacement.

There are pitfalls in this type of approach. As we saw in [Chapters 5 and 6](#) and also discuss in [Appendix C](#), wear rates in some systems are not constant but change with time: they may either decrease after a running-in period (as in many cases of lubricated sliding wear) or they may increase after an initial incubation period (as in solid particle erosion under certain conditions). To predict lifetimes by extrapolation, the designer must be confident that the wear rate will remain effectively constant, and that the operating variables of the system will not change and affect the wear rate, over its lifetime. Further, some knowledge of the mechanism of final failure of the system must exist, and the wear rate measured must be related to that mechanism. In the case of the slurry pipeline, for example, wall perforation by wear usually occurs on the outside curve of a bend, and it is the wear rate at that point that will determine the lifetime of the system, not the average wear rate at other locations. The consequences of wear may be quite different in different contexts, and wear that is tolerable in an off-highway truck suspension may be completely unacceptable in a safety-critical aircraft component.

The second method of estimating wear rates in service is to obtain data from tests on components of the system, under conditions that simulate those expected in service. This method is widely used to predict wear in bearings, which can be exposed in laboratory test rigs to loads, speeds, temperatures and states of lubrication that closely simulate service conditions. If an accelerated test is used to produce wear more rapidly than it occurs in service, for example by increasing the sliding speed of a bearing, then care must be taken (as remarked in [Section 5.2](#)) to ensure that the wear mechanism is not changed and that the other relevant operating conditions remain the same. As with wear measurements on complete systems in service, data from components in rig tests should be extrapolated only with caution.

Theoretical or empirical equations relating wear rate to the operating variables such as load and speed provide the third means of predicting wear rates. Simple equations for sliding wear (the Archard equation), abrasion and erosion were given in [Chapters 5 and 6](#). These can provide only a crude estimate of wear rate, but may well be of value in initial design calculations provided that values of the operating variables are known. In the case of equation [5.7](#) for sliding wear, equation [6.5](#) for abrasion and equation [6.21](#) for erosion, the absolute value of the wear rate depends on the value of the dimensionless wear coefficient K . Typical values of K have been discussed in [Chapters 5 and 6](#), and are summarized here in [Table 8.1](#). Although the ranges listed here are wide, sufficient experimental data exist for rather narrower ranges to be quoted for some specific materials and conditions. Sources of such information are listed at the end of this chapter.

Table 8.1 Typical values of dimensionless wear coefficient *K* for different types of wear (values apply to metals; values for ceramics may be similar)

	<i>K</i>
Sliding wear	
Unlubricated: severe	10^{-4} – 10^{-2}
Unlubricated: mild	10^{-6} – 10^{-4}
Hydrodynamic and squeeze film lubrication	$<10^{-13}$
Elastohydrodynamic lubrication	10^{-13} – 10^{-9}
Boundary lubrication	10^{-10} – 10^{-6}
Solid lubrication	$\sim 10^{-6}$
Abrasive wear	
Three-body (rolling abrasion)	5×10^{-4} – 5×10^{-3}
Two-body (grooving abrasion)	5×10^{-3} – 5×10^{-2}
Erosive wear	5×10^{-3} – 10^{-1}

8.3 THE SYSTEMS APPROACH

Where relative motion takes place between surfaces, wear can never be completely eliminated, although in some circumstances it may be reduced to an insignificant level. If the expected (or, in the case of redesign of a functioning system, actual) rate of wear is unacceptable, the designer must change the system in some way to reduce it. It is important to appreciate that the wear rate in a particular system is determined by the interaction of many factors. In the so-called *systems approach* to wear analysis, these influences are recognized and divided into two groups: the *structure* of the mechanical system, and the *operating variables* imposed on that structure.

The structure can be described by the materials making up the surfaces that are in relative motion, the nature of any interfacial material present (including any lubricants and abrasive particles), the environment (e.g., a surrounding gas or liquid), and by the geometrical relationships between these components. The operating variables are the conditions imposed on the system during use, such as operating speed, load and temperature. All these factors can influence the wear rate, and the designer can control most of them. We shall consider in the remainder of this chapter how wear in a system can be reduced by changing the operating variables, by lubrication, and by suitable selection of materials (including surface engineering).

8.4 REDUCING WEAR BY CHANGING THE OPERATING VARIABLES

There may be only limited scope for altering the operating variables, since these are usually determined by the overall purpose of the system. The simple equations given in [Chapters 5](#) and [6](#) indicate the likely effect of changing the operating variables on

several mechanisms of wear, and can provide some guidance. For example, in sliding wear with or without the presence of abrasive particles, equations 5.7 and 6.5 suggest that the wear rate will be reduced by lowering the normal pressure (applied load per unit nominal contact area); this may be achieved either by reducing the load (which will often be impractical within the constraints of a particular design) or by increasing the area over which it is carried. As we saw in Chapter 5, there is also the possibility that a reduction in normal pressure, or in sliding velocity, may lead to a transition in the mechanism of sliding wear. This can lead to a beneficial reduction in wear rate, but may alternatively (as illustrated in Section 5.7) cause a catastrophic increase. Knowledge of the regime in which wear is occurring is clearly very helpful in this context, and if the system is so well characterized that a wear mechanism map can be constructed, this may provide a valuable tool.

Fretting wear results from relative movement of two surfaces in contact (see Section 5.8). Two classes of fretting problem can be identified: *displacement-controlled* and *stress-controlled*. In the former, which occurs most commonly between parts that are required to move, such as bearing surfaces, the relative displacement cannot be eliminated and the designer must try to reduce the surface traction produced by the displacement. Lowering the normal force between the surfaces or the coefficient of friction between them will have this effect, as will design measures that reduce any concentration of contact stress caused by geometrical factors. In the second case, described as stress-controlled, the object of measures to reduce wear must be to reduce or eliminate the relative movement. Here an *increase* in the clamping force or coefficient of friction can lead to reduced displacement, and thus to a lower rate of fretting wear. Judicious use of suitable adhesive bonding or sealing materials can be helpful in these respects, as well as in preventing the access of the atmospheric oxygen necessary for the fretting wear mechanism to operate. The optimum design solution for a particular fretting problem therefore depends on the origins of the movement. In all cases, major reductions in fretting wear can be achieved by attention to the source of the fretting motion, whether it is caused by differential thermal expansion, by out-of-balance forces in rotating components, or by other sources of vibration. Unlike many other types of wear, fretting can often be completely eliminated by the use of suitable design measures.

Surface damage by contact fatigue, common in gear teeth, cam followers and rolling bearings (see Section 9.2.1), occurs as a result of cyclic surface loading. It is usually impractical to reduce the number of loading cycles to which the system is exposed, but steps can sometimes be taken to reduce the maximum contact stress through changes in geometry or reduction in load with beneficial results.

If wear is due to hard particles, as in abrasive or erosive wear, then the phenomena discussed in Chapter 6 suggest ways in which the wear rate may be reduced. In some cases, the particles can be removed from the system, either partially or completely, and this will usually lead to reduced wear. Since larger particles tend to cause proportionately more wear than smaller ones, methods of separation that remove the larger particles selectively (such as filtration or inertial separation from a fluid) can be especially useful. If the particles are present as contaminants in a fluid lubricant, then reducing the maximum size of the particles present to below the minimum lubricant

film thickness, so that no load is carried by the particles, will result in greatly reduced abrasive wear.

In some systems the presence of abrasive particles is inevitable, and there is no scope for reducing their concentration or size. In wear by solid particle erosion (see [Section 6.4](#)), the most powerful factor controlling the wear rate is the particle impact velocity, and any measures to reduce this (if possible within the design limitations) will be beneficial. Advances in computational fluid mechanics now allow particle trajectories through fluid flow systems to be predicted (which can include pipework, valves and chokes for both liquid, gas and two-phase flows, pumps and flow paths through gas turbine engines). The computed flow trajectories can then be combined with empirical models for erosive wear to predict the distribution and severity of wear within the system, forming a valuable design tool. Wear in pipelines carrying gas-or liquid-borne particles (e.g., slurries) tends to be concentrated at the outside radius of bends (the ‘extrados’) or at constrictions (e.g., valves), since it is at these points, where the flow changes direction, that particles strike the surface rather than sliding along it. As we have seen, the angle of impact that causes the most rapid wear depends on the nature of the surface material; the impact angle of particles at a right-angled bend lies typically in the range 20° – 45° , depending on the shape of the bend, the nature of the particles and that of the carrier fluid. There is evidence that due to these geometrical factors the wear rate is particularly severe in slurry pipe bends with a ratio of bend radius to pipe diameter of between 2 and 3.5, and wear rates can be reduced by using bends of different, preferably larger, radius. In the pneumatic transport of powders the use of a blind T-piece rather than a conventional bend is often successful; particles pack into the redundant arm of the bend and protect the material from damage.

8.5 EFFECT OF LUBRICATION

Lubrication provides a powerful method of reducing wear, as well as friction, in many sliding systems. Full-film hydrodynamic lubrication (see [Section 4.4](#)) leads to the lowest rates of sliding wear, with a corresponding value of K typically less than 10^{-13} . Such low wear rates are truly insignificant. But hydrodynamic conditions cannot be maintained for ever, and under the conditions of boundary lubrication that will pertain during starting or stopping the value of K may rise as high as 10^{-6} , depending on the properties of the lubricant used. If conditions are so severe that the boundary film is penetrated and sliding occurs between essentially unlubricated surfaces, then the wear coefficient may become 10^{-4} to 10^{-2} , values that would be unacceptable in nearly all engineering applications. Values of K applicable to these and other sliding conditions are listed in [Table 8.1](#).

Clearly, hydrodynamic (or thick squeeze-film) lubrication is the most desirable state for a sliding system, and if feasible the designer should try to ensure that the system remains in this regime under all operating conditions. The most important factor determining the regime of lubrication is the minimum lubricant film thickness

compared with the roughness of the surfaces, as discussed in [Section 4.5](#). The value of λ , defined by equation 4.20, can be increased either by reducing the roughness of the sliding surfaces or by increasing the thickness of the lubricant film. The surface roughness is determined by the final surface finishing process used in manufacturing, as modified by wear during the running-in process. The lubricant film thickness can be raised, for a given load and sliding speed, by increasing the bearing area or the viscosity of the lubricant (although at the expense of greater frictional power dissipation). The design of fluid film bearings is a complex subject and outside the scope of this book; sources of further information are listed at the end of [Chapter 4](#), while materials for all types of bearings are discussed in [Section 9.2](#).

8.6 SELECTION OF MATERIALS AND SURFACE ENGINEERING METHODS

8.6.1 Introduction

[Figure 8.1](#) outlines the process involved in the mechanical design of a product. It includes reference to the tribological aspects of performance (such as frictional forces, temperature rise and wear rate) as well as the more traditional mechanical aspects (such as load-carrying capacity and deflection or distortion). The choice of materials from which the components of a system are made is frequently restricted by factors that have little or nothing to do with tribology. Cost, for example, is a vital factor in most applications. Overall weight may be important, and so may corrosion resistance. Mechanical properties such as strength, stiffness and toughness are usually of primary concern in most mechanical engineering applications. Electrical or magnetic properties are important in other cases. Although these requirements may limit the range of usable materials, they usually provide some scope for choice. Furthermore, since most of the properties listed above (except perhaps corrosion resistance) are determined by the bulk of the material, there is ample scope for modifying those surface properties that are of major concern to the tribologist. Methods for modifying the wear and frictional behaviour of surfaces by surface engineering have been reviewed in [Chapter 7](#). The numerous processes that are available should be considered as an integral part of the process of overall design and material selection, alongside the nature of any possible lubrication. Comments on surface engineering processes in the context of sliding wear, fretting wear, and wear by hard particles are included in [Sections 8.6.2–8.6.4](#) below, while more general points that need to be considered in selecting a process are covered in [Section 8.6.5](#).

Metals form by far the most common choice of materials for mechanical components. The compositions and microstructures of metallic alloys are largely standardized, often internationally, and their properties can be more or less accurately predicted. Non-metallic materials are less well standardized, and the properties of these materials, even with nominally identical compositions, tend to vary more widely than those of metals. The extent of tribological information available for

all these materials varies in a similar way. For metals, more data are generally available, and these data show less variability, than those for other materials. Unlike conventional mechanical and physical properties, however, the response of materials in tribological applications often cannot be expressed by simple numbers. As we saw in [Chapters 5 and 6](#), there is no single quantity representing the ‘wear resistance’ of a material that is applicable over a range of mechanisms of wear, or even in many cases for a single mechanism over a wide range of operating conditions. The selection of materials and surface engineering methods for tribological performance is therefore inevitably less quantitative, and more dependent on generalized rules and orders of merit, than for example selection for fracture toughness, specific stiffness or yield stress.

The relative performance of different materials varies in different tribological applications, depending largely on the particular mechanism of wear that dominates. In the following discussion we shall concentrate on the selection of materials (both bulk materials and engineered surfaces) for resistance to wear involving sliding (lubricated and unlubricated, and including fretting), abrasion and erosion. The properties required in the important tribological context of bearings (both plain and rolling element) are rather different, and this topic is treated separately in [Section 9.2](#).

8.6.2 *Sliding wear*

The relative ranking of materials for resistance to sliding wear depends on the precise conditions under which sliding occurs. As we saw in [Chapter 5](#), sliding wear in all types of material can occur by several different mechanisms, and transitions between different dominant mechanisms can be induced by small changes in normal load, sliding speed, temperature or environment. General statements can, however, be made about the sliding wear behaviour of materials. Most of the experimental data available relate either to relatively low wear rates, often associated with some lubrication and with the mild wear regime, or alternatively to severe wear and surface damage under heavy load and low sliding speed (i.e., *galling*: see [Section 5.1](#)).

The Archard wear equation (see [Section 5.3](#)) relates the wear rate Q (volume removed per unit sliding distance) to the normal load W

$$Q = \frac{KW}{H} \quad (8.1)$$

but provides only a crude indication of how the properties of a material affect its resistance to sliding wear. The only material property that explicitly enters the equation is the hardness, H , of the softer of the two surfaces. The other properties of the material that also influence its wear rate are hidden in the dimensionless wear coefficient K . The value of K also depends strongly on the state of lubrication of the system. The effects of these two factors are separated in [Fig. 8.2](#), which shows how the value of K depends on the material of the sliding surfaces and on the quality of lubrication. As discussed in [Section 8.5](#) above, the potential benefits of lubrication are large.

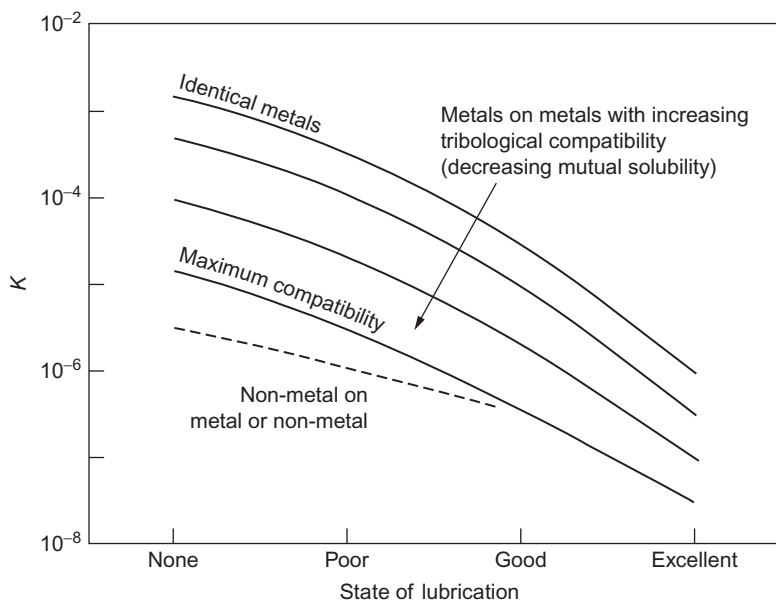


Fig. 8.2 Typical values of the wear coefficient K for a wide range of materials sliding under varying conditions of lubrication (drawn using data from Rabinowicz, E., 1980. In: Peterson, M.B., Winer, W.O. (Eds.), *Wear Control Handbook*. ASME, pp. 475–506)

In general, higher values of K are found for metals sliding against metals than for non-metals sliding against either metals or non-metals. Most metals sliding against counterfaces of the same metal show high values of K . If the metals are dissimilar, the value of K is lower and it has been suggested that it depends on the degree of *tribological compatibility* of the two metals. The term ‘compatibility’ in this context denotes a reluctance of the opposing surfaces to form a strong interfacial bond, which would lead to a high wear rate under unlubricated conditions. The tribological compatibilities of metals do not correlate perfectly with other properties, although the extent of mutual solid solubility, deduced from the equilibrium phase diagram for the materials of the sliding couple, is often suggested as a guide.

Figure 8.3 represents in chart form the mutual solubilities of pairs of pure metals. Those combinations marked as completely insoluble show negligible solid solubility in each other, and also even form two distinct co-existing phases in the liquid state; they usually form tribologically compatible pairs. Identical pairs of metals are, of course, completely mutually soluble and show poor compatibility. Other pairs show varying extents of solid solubility, as indicated on the chart. In general, sliding pairs with high mutual solubility show low tribological compatibility and hence relatively high values of K ; low mutual solubility, leading to good compatibility, is needed for low wear rates.

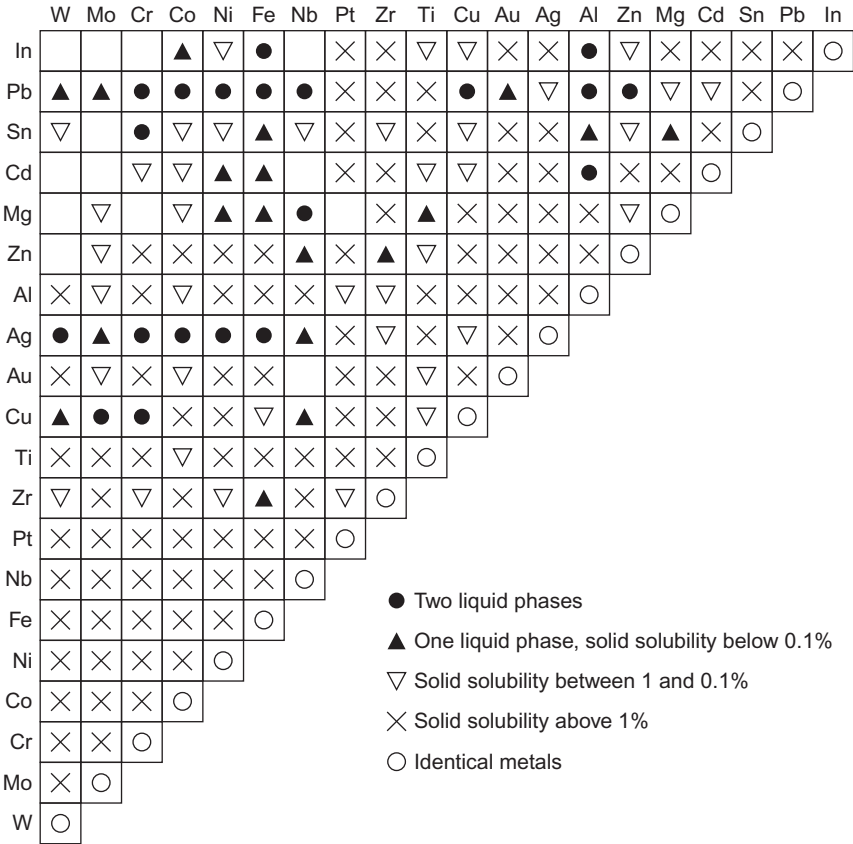


Fig. 8.3 Chart showing the relative mutual solubilities of pairs of pure metals, deduced from their binary phase diagrams (after Rabinowicz, E., 1995. *Friction and Wear of Materials*, second ed. John Wiley and Sons, New York)

Mutual solubility is not the only factor influencing tribological compatibility. It is also associated with the properties of the surface films (usually oxide) on the sliding materials. The absence of appreciable oxide films on the noble metals such as gold, silver, platinum and rhodium tends to be associated with low wear rates for these metals. Oxidative mechanisms probably play an insignificant role. Values of K are then typically about one third of the values predicted from Fig. 8.2 for other metals.

Some metals with hexagonal close-packed crystal structures also show anomalous behaviour, associated both with their limited ductility compared with the cubic metals, and also with chemical factors. Titanium, zirconium and hafnium, for example, show the behaviour expected from Fig. 8.2 when sliding unlubricated against themselves, but exhibit only slightly reduced wear rates when lubricated by most hydrocarbon

lubricants. Other h.c.p. metals generally show lower wear rates than those predicted by Fig. 8.2.

The hardness of steels and other metals that form an oxide film during sliding is important in determining the mechanical stability of that film and hence the dominant mechanism of wear. If the metal is hard enough to provide sufficient support for the oxide, mild wear may occur at a comparatively low rate by an oxidative mechanism. In this way, hardness can have a strong influence on the sliding wear of some metals. However, although increasing the hardness of a particular alloy usually leads to a reduction in wear rate, hardness is not a good predictor of the relative wear resistance of different alloys. Other factors, especially the presence of microstructural features such as carbides in steels or graphite in cast irons, are often much more important.

The resistance of metals to severe adhesive wear and associated surface damage (galling) under high normal loads does not always correlate with their wear resistance under less severe conditions. Several factors have been found to control the susceptibilities of metals to galling: the effectiveness of surface films in preventing adhesion, the strength of the adhesion once these films break down, and the extent of junction growth (see Section 3.4.3) that occurs on further sliding. Mutual solid solubility, as an indicator of the strength of the adhesive force, plays some role; metals that adhere strongly are more liable to galling damage. Hexagonal metals with a restricted number of operative crystallographic slip systems show lower galling tendencies than cubic metals, presumably because of their lower ductility. Studies of the effect of stacking-fault energy in alloys have suggested that metals with low stacking-fault energy are less inclined to gall than those with high energy. Other investigators have found high work-hardening rates to be beneficial in several cobalt-, nickel- and iron-based alloys, although work-hardening rate is not an infallible indicator of resistance to galling; austenitic stainless steels, for example, are notoriously susceptible to galling yet work-harden rapidly by the formation of martensite. Hardness alone is a poor indicator of galling resistance; in steels, for example, a high concentration of carbides or nitrides is desirable, and confers greater wear resistance than similar hardness levels achieved with a lower volume fraction of these hard and brittle precipitates. Hard coatings or diffusion layers that are also of very limited ductility generally confer good resistance to this type of wear. Rough surfaces, preferably of a random nature (e.g., as produced by abrasive grit blasting) generally increase galling resistance because junction growth is limited; metals with polished surfaces are, in contrast, much more liable to suffer galling.

As we have seen in Section 5.10 and Fig. 8.2, ceramics undergoing mild sliding wear can exhibit wear coefficients as low as, or even lower than, those for dissimilar metals. This fact, coupled with their greater hardness, indicates that ceramics can offer significantly lower sliding wear rates than metals. However, there are drawbacks to the use of bulk ceramics in tribological applications. Their mechanical properties (especially fracture toughness) may not be adequate for the demands placed upon them, they may be difficult to fabricate in the shape required, and the possibility of small-scale surface fracture leading to severe wear (see Section 5.10.3) calls for cautious design. Nevertheless, bulk ceramic components can prove extremely durable for

some tribological purposes: examples include alumina bushes and face seals in water valves and circulating pumps, silicon nitride valve components, and alumina femoral heads and cups for artificial hip joint implants, as discussed in [Section 9.5](#).

Some of the disadvantages of bulk ceramics can be overcome by using the material in the form of a coating on a metallic substrate, and ceramic coatings produced by thermal spraying, physical vapour deposition (PVD) and chemical vapour deposition (CVD) form an important group of surface engineering methods. In all tribological use of ceramics, lubrication is highly advantageous since by reducing the surface tractions the occurrence of local fracture leading to severe wear can be avoided. However, the possibility of chemical interaction between an unsuitable lubricant and the ceramic, causing enhanced wear, must always be considered (see [Section 5.10.5](#)).

Polymers are rarely thought of as wear-resistant structural materials, being more usually used as bulk bearing materials or coatings, often under conditions of marginal or dry lubrication (see [Section 9.2.3](#)). Some polymers with sufficient strength may however be used as bulk structural components in tribological applications, notably nylons (polyamides), acetal, polyetheretherketone (PEEK) and polyethersulfone (PES), often reinforced with a suitable filler. Cages for rolling element bearings, for example, can be moulded from glass-filled nylon (see [Section 9.2.1](#)), and this material, as well as acetal, is also commonly used for lightly loaded gears. Selector forks for racing car gearboxes fabricated from carbon-fibre reinforced PEEK have been used as they combine good tribological performance with substantial weight savings compared with the conventional forged steel components.

The diversity of surface engineering methods available to the designer, and described in detail in [Chapter 7](#), permits the tribological properties of a surface to be selected independently, at least to some extent, from the properties of the bulk of the material. [Figure 8.4](#) illustrates the wide range of depth and hardness that can be achieved in the surface regions of the most common structural material, steel, by these methods.

From this diagram it will be clear that different methods will be most suited to different types of application. The depth of the modified region or the thickness of the coating is of primary importance; methods that produce only thin layers with modified properties, such as the vapour phase coating methods (CVD and PVD) or ion implantation will be useful only in applications where the maximum extent of wear will be comparably small, and where the maximum subsurface stress lies close to or at the surface. For such applications, e.g., small precision engineering components, milling cutters, twist drills and press dies, these methods can offer substantial benefits. PVD titanium nitride coatings, in particular, on high speed steel and cemented carbide tools are found to give greatly enhanced life; even better performance is claimed for similar coatings containing other elements (e.g., (Ti, Al)N or (Ti, Nb)N).

In other cases, where the surface stress imposed in service penetrates more deeply into the component, methods that produce a thicker modified layer or coating are required. In a heavily loaded gear wheel, for example, the surface material must have a high yield stress to ensure that it remains elastic when exposed to localized contact stresses and to reduce sliding wear. However, the core of the gear tooth and rest of the wheel require high fracture toughness and resistance to fatigue cracking, in order to

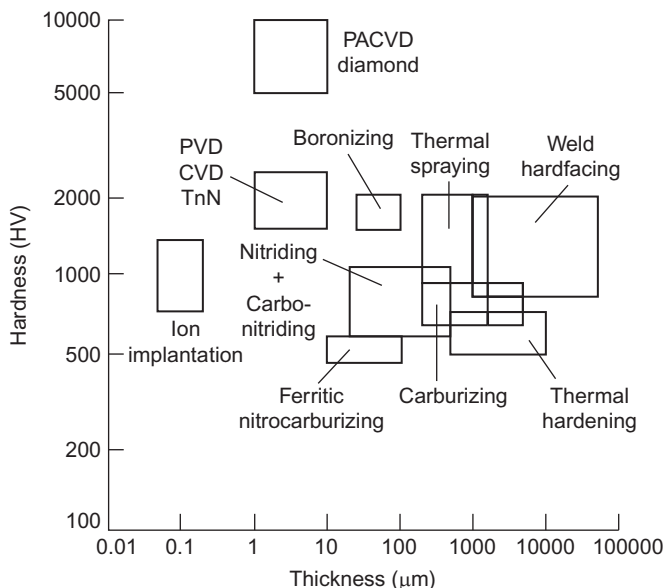


Fig. 8.4 Typical depths of treated layer or coating, and range of hardness achieved, for different methods of surface engineering. All the methods shown are applicable to steels, but the diagram is not exhaustive, and some of the coating methods can also be applied to other substrates

withstand the cyclic loading, perhaps applied impulsively, to which the gear train is exposed. In a steel, these two requirements of highest yield stress or hardness (desirable in the surface) and the greatest fracture toughness (for the core material) cannot be achieved together. The problem of optimizing both at an economical cost can be solved only by modifying the surface material, perhaps by carburizing or nitriding, so that a suitably hard surface can be combined with a core material of high toughness.

Further possibilities exist to combine different methods of surface treatment on a single component (by so-called *duplex surface engineering*). For example, nitriding, carburizing or thermal hardening may be carried out to produce a deep layer of modified material, followed by PVD or CVD coating or ion implantation to produce a thinner surface layer of yet different properties. Coatings can also be produced with graded compositions and properties, to improve the matching of mechanical properties and enhance adhesion at the interface with the substrate. It is clear that substantial development of these and other surface engineering processes will occur in the future.

Guidance on the selection of specific materials for particular applications would be out of place in a general text, and sources of such information are suggested at the end of this chapter. Databases and expert systems can provide valuable sources of information for the designer. However, some brief guidelines on the selection of materials for resistance to sliding wear can be given:

1. There is no general correlation between wear rate and coefficient of friction, although lubricants, which may be externally supplied or present as constituents

of one of the materials (e.g., graphite in cast irons or molybdenum disulfide in some filled polymer composites) will tend to reduce both wear and friction. Even poor lubrication is better than none in reducing wear.

2. Surfaces of identical metals in sliding contact should be avoided. High tribological compatibility, which is usually associated with low mutual solubility, is more likely to be achieved between dissimilar metals.
3. High hardness is often beneficial and in metals can be achieved locally by suitable surface engineering methods (see item 6 below and [Chapter 7](#)). However, other important factors such as mutual solubility and work-hardening rate may mask the influence of hardness.
4. In steels, high carbide or nitride contents are beneficial, even at the expense of slightly reduced hardness.
5. A high work-hardening rate in a metal usually gives good resistance to severe wear and galling. Initial surface finish correlates with resistance to galling. Random rough surfaces (as produced by abrasive blasting) provide greater resistance than smoother surfaces.
6. Hard surface layers, for example those produced in steels by carburizing, nitriding or nitrocarburizing, or hard non-metallic coatings, as deposited for example by PVD or CVD processes often give excellent resistance to sliding wear. High hardness and low ductility seem to be desirable mechanical properties in such coatings.

8.6.3 Fretting wear

As we saw in [Sections 5.8](#) and [8.4](#), fretting may be classified as either displacement-controlled or stress-controlled, and can often be effectively limited by correct design measures. The selection of materials can also play a part in reducing fretting wear.

Some metals, notably titanium and its alloys, are notoriously susceptible to fretting damage, and their use should be avoided if fretting is likely. If a reduction in friction is desirable, as in displacement-controlled fretting, then lubrication may be effective, either by means of an externally-supplied liquid lubricant (preferably a low viscosity oil that can penetrate between the surfaces) or a solid lubricant film (inorganic or polymeric) applied to one of the surfaces. A soft metallic coating, applied for example by electroplating or by physical vapour deposition to a hard substrate, may also reduce the frictional force.

Methods of surface engineering leading to increased hardness can also reduce fretting wear, in both displacement-controlled and stress-controlled applications. Hardening the surface may also introduce compressive residual stress, with the further beneficial effect of reducing the tendency to fatigue crack growth from the fretting damage (fretting fatigue). However, some surface treatments cause an increase in friction, while others lead to a decrease. Care must therefore be taken to choose a hardening method that gives an appropriate level of friction for the particular application.

8.6.4 Wear by hard particles

We saw in [Section 6.2.1](#) that the rate of wear of a material by abrasion or erosion falls sharply if its hardness can be made greater than that of the abrasive particles themselves. Desirable as this condition may be, it is not readily achieved with steel

components since very few alloys are as hard as naturally occurring abrasive particles. Silica, for example, typically has a hardness of some 800 HV, a value attained only by martensitic steels, untempered, of high carbon content (see Fig. 7.7). Although the results discussed in Sections 6.3.3 and 6.4.3 can be used to select engineering alloys with optimum resistance to abrasive and erosive wear respectively, the designer seeking the lowest wear rates will be forced to use harder materials, either in bulk form or as coatings. Candidate bulk materials include white cast irons (usually containing chromium, with a high volume fraction of carbides in a martensitic matrix), cemented carbides (e.g., tungsten carbide in a cobalt or nickel binder) and ceramics (e.g., alumina, silicon carbide, silicon nitride, boron carbide or sintered polycrystalline diamond). As discussed in Sections 6.3.3 and 6.4.3, these materials can offer excellent resistance to wear by hard particles and in many cases are available as prefabricated components (e.g., as tubes, tiles, slurry pump liners, grit blaster nozzles etc.) that can readily be incorporated into new or existing designs.

For some applications, however, bulk wear resistant materials may not be suitable, perhaps for reasons of cost, overall weight, difficulty of fabrication or mechanical properties. Surface engineering methods can then be used to apply a coating of a wear-resistant material to a substrate with lower wear resistance but with the desired bulk properties. As in the case of sliding wear (see Section 8.6.2), the thickness of the modified surface layer is a most important consideration. Both the expected rate of wear and the depth to which significant stresses are induced by particle contact must be taken into account. Weld hardfacing processes (see Fig. 8.4 and Section 7.4.2) must be used to produce the thickest coatings: metallic matrices (typically iron- or cobalt-based) with a high content of chromium carbide, tungsten carbide or other hard carbides can be deposited by welding methods to provide resistance to high stress abrasion by large particles, in applications such as rock drills, excavator teeth and ore-crushing machinery. Rather thinner coatings of these materials can be produced by thermal spray processes, some of which can also deposit ceramic materials such as alumina, chromia, zirconia or titania. Boronizing of steels also leads to a surface layer with sufficient hardness to resist some types of abrasive wear, but the benefits of the other methods of surface modification of steels discussed in Section 7.3 are generally not great because of the limited increase in hardness which can be achieved. The very thin but hard coatings formed by CVD and PVD processes provide useful abrasive wear resistance only if the abrasive particles are sufficiently small. For example, PVD titanium nitride coatings can increase the life of moulds and extrusion dies used for polymer processing, by reducing the rate of abrasive wear due to fine filler particles in the polymer melt.

In certain cases, polymers can provide viable alternatives to very hard materials. For example, ultra-high molecular weight polyethylene (UHMWPE) can be used to line hoppers and chutes carrying powdered materials, and offers low sliding friction together with adequate abrasion resistance. Polyurethanes are also used to line powder-handling equipment, and in aqueous slurry pipelines and pumps provide dual protection against both corrosion and erosive wear. Other elastomers, especially natural rubber, can also show good resistance to erosive wear by airborne particles, especially for rounded particles at high impact angles and low velocities (see Section 6.4.3).

8.6.5 General aspects of surface engineering processes

In selecting a surface engineering method (and where relevant, material), the function of the surface layer is the primary concern. It may depend upon intrinsic material properties such as hardness and fracture toughness along with extrinsic properties specifically related to the surface engineered layer (e.g., thickness, surface roughness, nature of the interface (sharp or diffuse), residual stress state and bond strength). Assuming that the required properties can be realized in a given surface engineered system, the selection of the process then turns on its technical feasibility and economic profitability. A number of features associated with the process and the process-workpiece compatibility need to be considered. This section seeks to outline the topics that must be taken into account when selecting a process, and to highlight possible constraints on processes in these areas. It also considers an important example of a further constraint: the need to modify or even completely replace a process to conform to new legislation.

Process—Substrate compatibility

For many processes, there are restrictions on substrate type. For processes where the surface engineered layer is formed either by modification of the existing component surface (such as thermal hardening) or by reaction between an applied species and the component (as in anodizing), there are obvious limitations associated with the reaction itself. For example, boronizing can be applied only to a limited range of materials which form hard boride layers (such as iron-based substrates). Some of the process constraints are rather less obvious. For example, in the case of hard anodizing of aluminium alloys, several factors associated with the substrate must be taken into account. Hard anodizing is difficult on aluminium alloys that contain more than 5% copper or 7% silicon; for heat treatable aluminium alloys, significant differences exist in the anodizing conditions depending upon the prior heat treatment; attachments or inserts of metals other than the base aluminium alloy must be masked off (both electrically and chemically) prior to anodizing to prevent excessive pitting (commonly referred to as ‘burning’) and corrosion.

Other restrictions on the choice of process imposed by the substrate type are associated either with successful operation of the process depending upon certain features of the substrate, or with unacceptable changes in the substrate that result from the coating process. For example, electroplating processes demand a substrate that is electrically conducting. To electroplate a non-conductor, a metallic coating is normally first applied by an electroless method (although not all surfaces are amenable to electroless plating) which then allows electroplating to proceed. Second, the temperature at which the process is conducted may cause changes in the substrate. For example, some CVD processes operate at temperatures above 850°C and so will modify the properties of many steel substrates, including tool steels, as a result of phase transformation or precipitation which may lead to dimensional changes (and loss of tolerance) or property changes (especially of hardness).

Process—Component size compatibility

Extremes of component size can make surface engineering by certain processes (but not all) so complex that they become uneconomical. It must be stressed that these are normally difficulties and not inherent constraints, and thus can often be overcome technically if a financial case can be made.

Very large components are easiest to coat by processes that do not require the component to be sealed in a chamber or immersed in a bath. This area is dominated by processes where the coating is deposited as a liquid, such as weld hardfacing, thermal spraying and the application of polymer paints and putties (see [Section 7.4.2](#)), along with selective thermal hardening (e.g., flame hardening). These can all be utilized readily on large structures without disassembly and are often used in these contexts for repair in industries such as those involving mineral handling where wear is generally unavoidable. Components up to around 3 m in their largest dimension are regularly surface engineered by processes where they are required to be either immersed in a bath (e.g., electroplating or electroless plating) or placed in a vacuum chamber (such as PVD or CVD).

Small components provide some difficulties in surface engineering simply because they need to be held in some way during the coating process. Small parts being coated by a CVD process may be placed loosely on trays, and components being electroplated may be placed in agitated drums or barrels. In contrast, parts to be coated by thermal spraying are normally individually clamped. The two groups of methods will thus involve very different set-up times (and costs).

Coverage

Components often have complex geometries, and if the area to be treated lies within a recess or an internal bore then process selection may be restricted by this requirement. The severest restrictions are for ‘line-of-sight’ processes where the surface engineered layer can only be formed efficiently on surfaces that can be directly exposed to the material source (e.g., a thermal spray torch) or energy source (e.g., a laser beam). Electrolytic processes (such as anodizing and electroplating) can provide better coverage (with suitable agitation of the electrolyte) but tend to form thicker coatings on surfaces facing the counter-electrode and on external corners and edges. In electroplating, deep internal bores can be coated by use of a wire counter-electrode positioned centrally within the bore. Processes where a coating is formed at the surface by chemical reactions involving the species in the environment (such as CVD or electroless plating) provide the most uniform coverage, provided that the environment is not locally depleted of reacting species by poor flow or stagnation (which may occur in very narrow features or blind holes).

Replacement of hard chromium: An example of a legislative constraint

Chromium electroplating has been widely used for many years as a surface engineering method to provide damage, wear and corrosion resistance to a wide variety of components ranging from hydraulic actuators to threaded fasteners. There are, however,

significant concerns over the use of salts of hexavalent chromium (Cr(VI)) in the process, in terms of damage to the environment and potential hazards to the health of the workforce, since Cr(VI) is an acknowledged carcinogen. Within the European Union, a regulation known as REACH (Registration, Evaluation, Authorization and Restriction of Chemicals) came into force in 2007, with the 'sunset date' of 2017 for the use of hexavalent chromium-containing compounds in production. These environmental and health concerns (along with regulations) have required industry to consider substitute materials.

Replacement of hard chromium plating by substitution is not straightforward. Alternatives such as HVOF-sprayed WC-Co-Cr, electroless-deposited nickel-SiC and PVD chromium have all been suggested and tested for certain applications, but none has been identified as the universal replacement. The reasons for this are that a material is not defined by a single attribute (such as wear resistance in sliding), but by a long list of attributes (such as hardness, thickness, residual stress, corrosion resistance and many more), and to be able to replace a material with another effectively requires identification of all the attributes that are significant in the application, followed by identification of an alternative material which has these attributes at the level needed to meet the requirements. For example, electro-deposited chromium has a hardness between 800 and 1000 HV, but this does not mean that any other material with the same hardness will necessarily be an effective substitute. Moreover, in some applications where electrodeposited chromium is currently employed, there is no requirement for such high hardness, and identification of a material with similar hardness will therefore not be an effective way to select a replacement. The difficulties illustrated here remind us that tribology is a complex subject in which potential material failure mechanisms must be understood before any design decisions are taken.

SOURCES AND FURTHER READING

- ASM Handbook, 1992. Friction, Lubrication and Wear Technology. vol. 18 ASM International.
- Budinski, K.J., 1988. Surface Engineering for Wear Resistance. Prentice Hall.
- Czichos, H., 1978. Tribology: A Systems Approach to the Science and Technology of Friction, Lubrication and Wear. Tribology Series No. 1, Elsevier.
- HMSO, 1986. Wear Resistant Surfaces in Engineering. UK Department of Trade and Industry.
- Lansdown, A.R., Price, A.L., 1986. Materials to Resist Wear. Pergamon.
- Neale, M.J. (Ed.), 1995. Component Failures, Maintenance and Repair. Butterworth-Heinemann.
- Peterson, M.B., Winer, W.O. (Eds.), 1980. Wear Control Handbook. ASME.
- Roy, M., Davim, J.P. (Eds.), 2015. Thermal Spray Coatings and Their Tribological Performances. Engineering Science Reference.

QUESTIONS FOR CHAPTER 8

Question 8.1

Use the Internet or other resources to find out the shape and dimensions of a typical connector plug used in a smartphone or other consumer electronics device (for example, to connect it to a charging source or data cable). Write a report for the device manufacturer about the tribological issues that may be important in the design and use of the plug. Your answer should include consideration of the following aspects:

- (a) The conditions under which sliding between the plug and its socket occurs.
- (b) Possible methods to determine the wear rates of the sliding components.
- (c) Possible materials that might be used for these components.
- (d) Comments on why the connector has been designed as it has.

Question 8.2

Many different metal alloys are used for making coins. Some coins are made with a core of one metal, coated with another. Major requirements are low cost, ease of manufacture (including both the formation of the coin 'blank' and the die-pressing of detailed surface features), low toxicity, and good corrosion resistance. It is also important that the coins retain their surface detail for many years while in circulation, which implies resistance to wear.

- (a) How would you assess the wear rates and wear mechanisms of coins in use?
- (b) And how would you attempt to predict the wear rate of a new alloy proposed for use in coinage?

Question 8.3

A long horizontal conveyor belt is used to transport granular iron ore. The ore, initially at rest, is fed gently on to the moving rubber-covered belt at one end, covering its whole width, and falls off at the other. Assume that abrasive wear of the belt surface occurs due to the relative sliding between the belt and the layer of ore as it is accelerated up to the belt speed, and follows the simple relationship of equation 6.5.

- (a) Show that the thickness of rubber removed from the belt by wear in transporting a mass M of ore is given by:

$$t = \frac{kMV^2}{4\mu wL}$$

where k is the specific wear rate for abrasion of the rubber by the iron ore, V is the speed of the belt, μ is the coefficient of friction between the bed of ore and the belt surface, w is the width of the belt, and L is the overall (end-to-end) length of the conveyor. Make your assumptions clear.

- (b) In one particular conveyor, 1100 m long, 1.8 wide, and moving at 4 m s^{-1} , 10 mm thickness of rubber was worn from the belt after it had transported 80×10^6 tonnes of ore. Laboratory tests showed that the coefficient of friction between the ore bed and a sample of the rubber from the belt was 0.8. Estimate the value of the specific wear rate, k .
- (c) What approaches might be used to reduce the wear rate of the belt?

Question 8.4

Explain the following choices of surface treatment processes for tribological applications. In your answers, discuss the advantages and disadvantages of the processes.

- (a) Drill bits made from high-speed steel are coated by PVD methods and not by weld hardfacing.
- (b) DLC coatings are used on some car engine camshafts, and not electroplated chromium.
- (c) Weld hardfacing is applied to the teeth of excavator buckets used in earthmoving, rather than flame-sprayed ceramic coatings.
- (d) Steel injection moulding dies used to make glass-filled polypropylene parts are coated with titanium nitride by PVD, rather than sputtered molybdenum disulfide.

9.1 INTRODUCTION

In this chapter we shall review some practical applications of the tribological principles that have been covered in the earlier parts of the book. We start by considering bearings, which form fundamental and vital components of nearly all mechanical engineering systems. In the rest of the chapter we consider applications of tribology in automotive engineering, manufacturing technology (including grinding and polishing, cutting tools and chemical–mechanical polishing), bio-tribology (with particular attention to natural and artificial hip joints) and magnetic data storage.

9.2 BEARINGS: ROLLING AND SLIDING

A *bearing* is a device that permits two parts of a mechanism to move freely relative to each other in either one or two dimensions, while constraining them in the remaining dimensions. We can classify them according to the motion they allow. The simplest arrangement is the *linear bearing*, which allows linear motion but can nevertheless support substantial normal loads. Figure 9.1(a) represents a linear bearing in idealized form. More common in engineering use is the *rotary bearing*, which permits a cylindrical shaft to rotate freely while transmitting radial loads and maintaining the axis of the shaft in a fixed position (Fig. 9.1b). A *thrust bearing* (Fig. 9.1c) also allows rotation of a shaft, but transmits a load along the axis of rotation. Some designs of rotary bearing can carry both radial and axial loads.

The most important requirement of a bearing is that it operates with low friction. This is usually achieved in one of two ways: with smooth lubricated sliding surfaces in a *plain bearing*, or by interposing balls or rollers between the moving surfaces in a *rolling element bearing*.

Both linear and rotary motion (with axial or radial loads) can be achieved with either plain bearings or rolling bearings. Most types of bearing will function satisfactorily only over limited ranges of operating conditions, and the load carried by the bearing and the sliding speed at which it operates are the two most important factors that determine its life. Figure 9.2 illustrates how the limiting operating conditions depend on rotational speed and radial load for plain rotary (journal) bearings of various types, for one particular shaft diameter. At low speeds, marginally lubricated and ‘dry’ plain bearings perform as well as rolling element bearings, but as speed is increased, rolling element bearings offer progressively greater load-carrying capacity. Once the sliding speed is sufficiently high to maintain a hydrodynamic film, hydrodynamically lubricated plain bearings will carry greater loads than rolling bearings, and can be used at considerably higher speeds, ultimately limited by the strength of the rotating shaft itself.

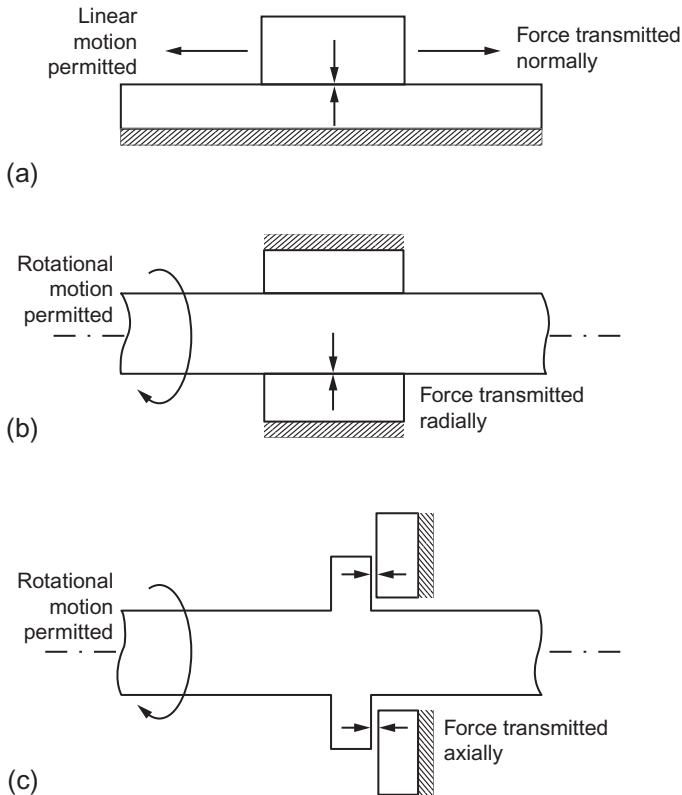


Fig. 9.1 Schematic illustrations of three different types of bearing: (a) linear bearing; (b) journal (shaft) bearing; (c) thrust bearing

The choice of bearing type for a given application will usually be based on many factors apart from speed and load, and is the subject of detailed discussion in some of the references listed at the end of this chapter. It is sufficient here to note that among other factors which may also be important in the selection and design of bearings are size, weight, cost, precision of shaft location, lubricant requirements and frictional drag. The differing demands of the wide range of applications for which bearings are used has resulted in a correspondingly wide range of types of bearing. In this section we shall review these types, paying particular attention to the materials employed in their construction.

9.2.1 Rolling element bearings

In a rolling element bearing, the load is carried by a set of balls or rollers, located between a pair of inner and outer tracks or *races*. The rolling elements are often held in place, regularly spaced around the tracks, by means of a *cage*, which is free to move with the balls or rollers between the tracks. Figure 9.3 shows typical single-row radial

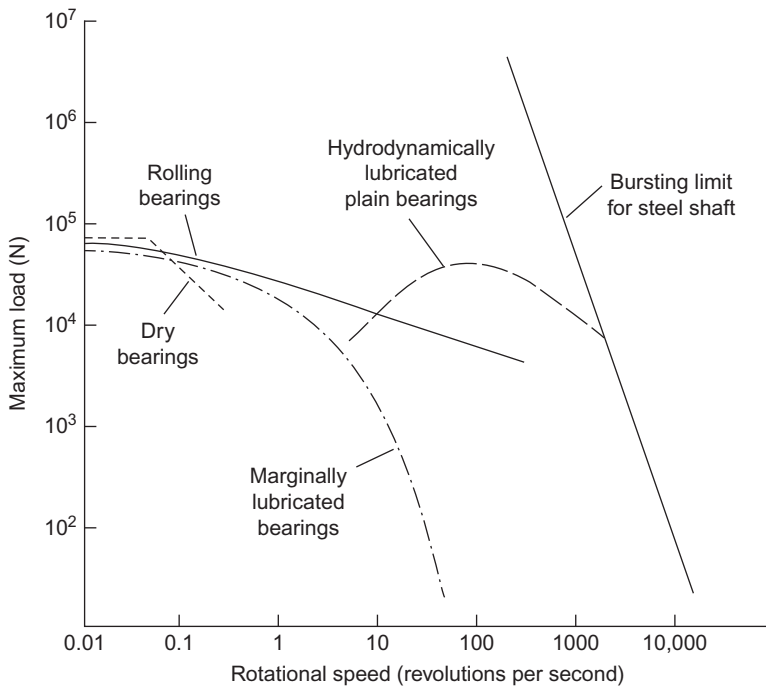


Fig. 9.2 Diagram showing the limiting values of load and rotational speed for journal bearings of various types, for a shaft diameter of 50 mm and a bearing life of 10^4 h. The curve relating to hydrodynamically lubricated bearings applies to a bearing of length 50 mm (data from Neale, M.J. (Ed.), 1973. *Tribology Handbook*. Butterworths, London)

ball and roller bearings. These are the simplest designs. Parallel roller bearings can carry high radial loads but offer no resistance to forces in the axial direction, whereas radial ball bearings, although primarily intended to carry radial loads, can also support a smaller axial load. Many other designs of bearing are used, which provide for a wide range of loading conditions: ball thrust races, for example, can carry large axial loads, while tapered roller bearings, used as opposed pairs, combine good radial and axial load capacity. Rolling bearings are manufactured to standardized dimensions, to very high precision and at moderate cost as a result of volume production.

When a purely radial load is applied to one of the types of bearing illustrated in Fig. 9.3, the load is borne to some extent by all the balls or rollers in the load-carrying half of the bearing. The most heavily loaded rolling element carries about $5/Z$ times the applied load, where Z is the total number of balls or rollers in the bearing ($Z > 5$). Figure 9.4 shows the geometry of the contact between a ball and its track. For both balls and rollers, the contact with the track is counterformal, and the contact pressures resulting from the applied loads are therefore high. These pressures, together with the associated subsurface stresses, can be calculated from Hertzian elasticity theory, provided that plastic flow does not occur.

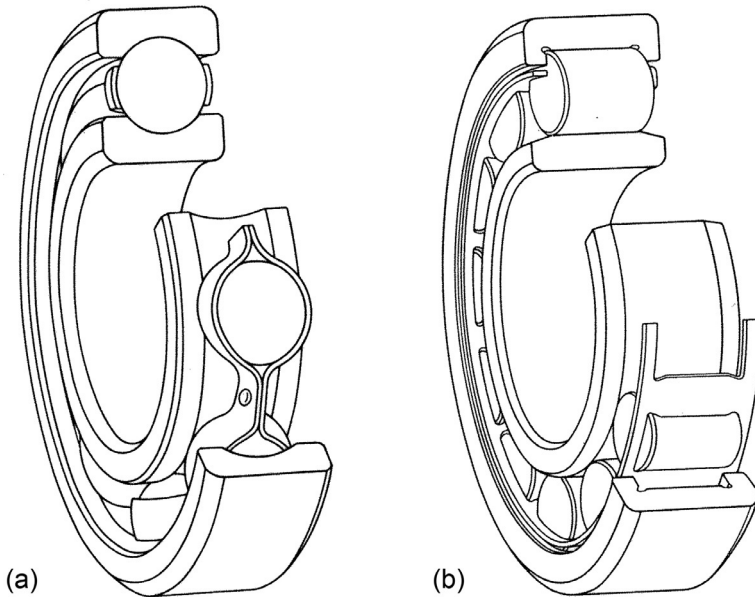
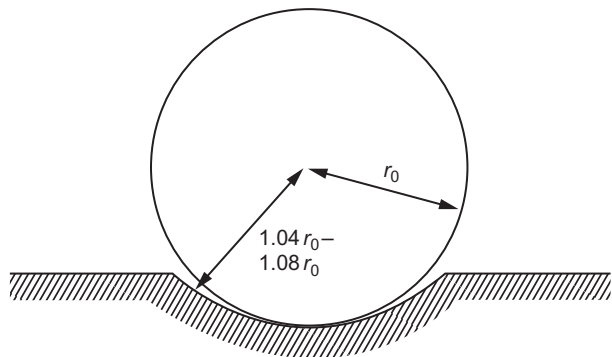


Fig. 9.3 Typical constructions of simple rolling element bearings: (a) single-row radial ball bearing; (b) single row cylindrical roller bearing (from Hjertzen, D.G., Jarvis, R.A., 1983. Rolling element bearings. In: Jones, M.H., Scott, D. (Eds.), *Industrial Tribology*. Elsevier (Chapter 6))

Fig. 9.4 Illustration of contact between a bearing ball and its track. Typically, the radius of curvature of the track is 4%–8% greater than that of the ball



Materials for rolling elements and races

A primary requirement for the operation of a rolling element bearing is that deformation of its components should at all times be elastic, in view of the high stresses that result from contact between the balls or rollers and the tracks. The static load rating for a steel ball bearing is typically chosen so that the maximum contact pressure at the centre of the most heavily-loaded ball-race contact is 4.2 GPa. The materials chosen

for the ball and race must therefore have a high yield stress, and high resistance to fatigue failure is also needed.

In most applications, steels are used for both the balls or rollers and the tracks, with typical hardnesses of 700–800 HV (60–64 HRC). The rolling elements may be up to 10% harder than the tracks. The standard alloy, used for general applications, is a through-hardened 1% carbon, 1.5% chromium steel, designated 534A99 or 535A99 (to British Standard BS970: previous designation En 31). Its US designation is 52100 (SAE/AISI). The high yield stress of this steel after heat treatment results from a microstructure of high carbon martensite, containing primary carbides. Bearings fabricated from this alloy are suitable for service at up to 125 °C; for non-critical applications they can be used at up to 200 °C. Case-hardening steels (typically chromium–nickel or manganese–chromium low alloy steels with 0.15% carbon) are sometimes used where heavy sections render through-hardening impossible, or where the higher toughness of the core of a case-hardened component is beneficial in resisting shock loading.

For high temperature applications, as in gas turbine engines, high-speed steels (secondary-hardening) are used: for example in the UK a high tungsten alloy, 18-4-1 (0.7% C, 0.3% Mn, 0.25% Si, 4% Cr, 1% V, 18% W, designated BT1 to British Standard BS4659), and in the USA a molybdenum steel, AISI M50 (0.8% C, 0.3% Mn, 0.25% Si, 4% Cr, 1% V, 4.25% Mo). With these alloys, bearings can operate suitably lubricated at up to 500 °C; both these steels give similar performance as rolling bearing materials.

Bearings required to operate in corrosive environments may be made from stainless steels: AISI 440C (1% C, 1% Mn, 1% Si, 18% Cr, 0.75% Mo), a martensitic steel, is a common choice, and can be hardened almost to the strength attained by standard bearing steels.

For very light loads, bearings with thermoplastic polymer races, typically injection moulded from acetal or polypropylene, can be used with stainless steel balls. These can be significantly cheaper than bearings with steel races as well as being lighter, operating more quietly, needing no lubrication and resisting corrosion.

Excellent corrosion resistance in many environments, together with high load-carrying capacity, can be achieved with rolling bearings fabricated from engineering ceramics, which are also of considerable interest for operation at very high temperatures. Hot pressed or hot isostatically pressed (HIPed) silicon nitride (Si_3N_4) is the most widely favoured ceramic, and with a good surface finish and careful fabrication this material can show a very good fatigue life, significantly better than M50 tool steel. The lower density of ceramic materials (typically 40% that of tool steel) results in another advantage for high speed use: a substantial fraction of the loading on the outer race of a high speed bearing is due to centrifugal forces, which are significantly lower for a material of lower density. Hybrid ball bearings in which silicon nitride balls run against steel races are used in some applications, and these also show good high-speed performance but at lower cost than all-ceramic bearings. Hybrid bearings can also show advantages over all-steel bearings in terms of lower friction and thus lower frictional heating, and are more tolerant of the ingress of abrasive particles. Because the ceramic

balls are electrically insulating, the possibility of bearing damage due to stray currents in electrical machines can be avoided.

The fatigue behaviour of bearing steels under rolling contact is well characterized, and rolling bearings tend to be designed on the basis of fatigue lifetime predictions, on the assumption that the bearing life will be limited by fatigue cracks that initiate at subsurface flaws and propagate under the cyclic stress field associated with rolling contact. Substantial improvements in the fatigue lives of rolling bearings have resulted from improved steelmaking practice, in particular from the reduction of oxide particles and other inclusions since these act as nuclei for subsurface fatigue cracking. Striking benefits have resulted from the introduction of vacuum induction remelting (VIR), vacuum arc refining (VAR) and electroflux remelting (EFR), as illustrated in Fig. 9.5. Vacuum degassing is now routinely applied to bearing steels, and results in twice the fatigue life of an air-melted steel at little extra cost; the use of the more expensive techniques of VIR/VAR and EFR is restricted to steels for critical applications such as gas turbine shaft bearings. Over the past fifty years or so, there has been such an improvement in the quality of bearing steels that in most applications bearings now only seldom fail by rolling contact fatigue. More commonly, their life is determined by failure mechanisms involving surface phenomena, which can be collectively termed

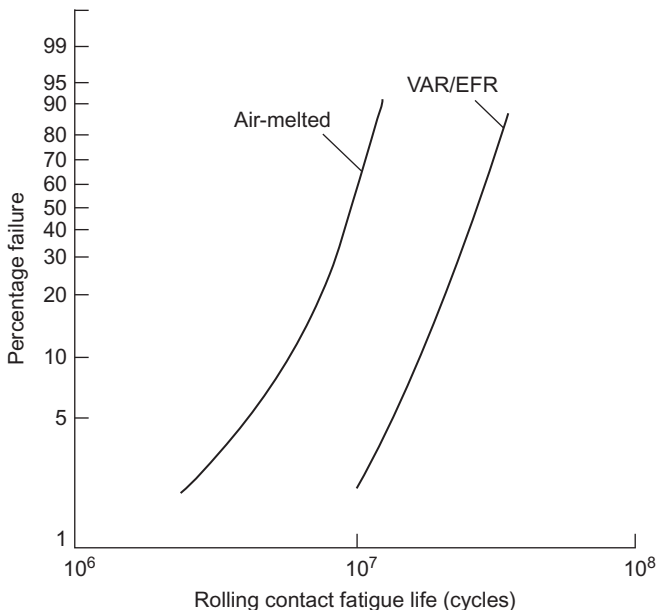


Fig. 9.5 Influence of steel processing methods on the rolling contact fatigue lives of rolling bearings made from AISI 52100 steel. The reduction of oxide and other inclusion contents by the use of vacuum arc refining (VAR), electroflux remelting (EFR) and vacuum induction remelting (VIR) has led to significant improvements in fatigue life (from Oakes, G., Barraclough, K.C., 1981. In: Meetham, G.W. (Ed.), *The Development of Gas Turbine Materials*. Applied Science Publishers)

‘surface distress’. These include pitting, surface-initiated fatigue and various modes of wear, and can be ascribed to defects in the original surface finish of the components, or to poor lubrication, or to contamination by abrasive particles or wear debris.

Surface damage to the race (termed *Brinelling* after the hardness test of the same name—see [Appendix A](#)) can sometimes occur, caused by impact loading as the harder ball or roller plastically indents the softer race; in a large bearing, such damage is often associated with mishandling upon installation. Another form of surface damage, termed *false-Brinelling*, is caused by fretting between the rolling element and the race in a highly loaded bearing that is essentially stationary. The fretting damage (discussed in detail in [Section 5.8](#)) results from vibratory motion and is typically observed in bearings in machines that are being transported by road or rail (with associated vibration) or in wind turbines that are not rotating but are still experiencing the transients of wind loading. Restricting the vibrational motion and the use of suitable lubricant additives are possible methods of prevention. Both Brinelling and false-Brinelling can lead to surface spalling of the race in service and thus to premature failure of the bearing.

Materials for bearing cages

The cage or separator that retains the balls or rollers between the races is an important component in most designs of rolling bearing (see [Fig. 9.3](#)). If a cage is not present, neighbouring balls or rollers will rub against each other with a high local sliding speed, and damage may result. By separating the rolling elements, the cage can lead to a substantial increase in the safe operating speed of a bearing.

Since the cage rubs against both the inner and outer races and the rolling elements, it must be made from a material that can slide against these components under poorly lubricated conditions without a high wear rate. Mild steel pressings, riveted or spot-welded together, are commonly used in standard bearings, sometimes with a tin or silver plating or a phosphate coating to lower the friction. Cages machined from 60/40 brass, bronze, or aluminium alloy are also sometimes used. Polymer materials can provide the advantages of low running noise, lightness and low wear rates. Fabric-reinforced phenolic resin is used in some applications. Cages in thermoplastic polymers can be fabricated very economically by injection moulding. Nylon 66, often reinforced with glass fibres, has been extensively used, although limited to operating temperatures below 120°C and above -40°C. Polyethersulfone, also reinforced by glass fibres, offers good performance at temperatures up to 180°C.

Friction and lubrication in rolling bearings

The frictional torque imposed by a rolling bearing supporting a shaft is very low. Expressed as a coefficient of friction, it has a typical value of 0.0015 for a radial ball bearing with a single row of balls, and can be as low as 0.0010 for a cylindrical roller bearing with short rollers. A rolling bearing is therefore a very efficient machine component. Only a well-designed hydrodynamically lubricated plain bearing operating near the minimum of the Stribeck curve (see [Fig. 4.7](#)) can offer a comparably low

coefficient of friction; the rolling bearing, however, maintains its low friction even from a standstill, whereas a plain bearing will exhibit a much higher starting torque.

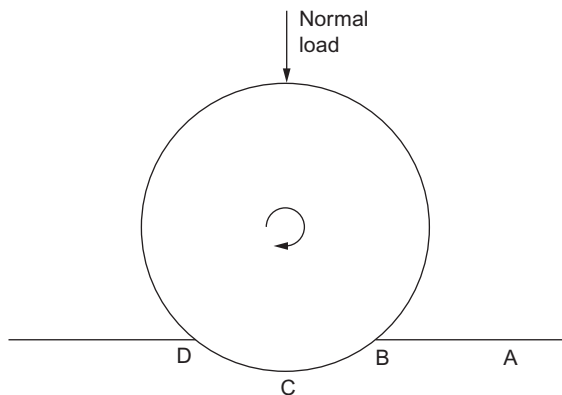
The frictional drag in a rolling bearing originates from several sources. The most important, in a lubricated bearing, is the energy dissipated in the thin lubricant films between the rolling elements and the tracks. At moderate speeds in a cylindrical roller bearing lubricated with oil, about 60% of the total energy lost is dissipated in these EHL films. Nearly all the remaining energy is accounted for by the lubricated sliding of the cage against the other components of the bearing. Only a very small fraction (of the order of 1%) of the energy is lost in the hysteresis associated with the cyclic elastic deformation of the balls or rollers and the track (as illustrated for a ball rolling on a polymer surface in Fig. 3.29).

Lubrication of rolling bearings is essential, mainly in order to form protective films between the various components and thus prevent excessive wear. Sliding contact occurs between the cage and the rolling elements and races; it also occurs to some extent between the rollers or balls and the races. Although under zero load a cylinder or sphere will roll over an elastic plane surface without slipping, when it supports a load some local slippage occurs in the following ways.

Cylindrical rollers and spheres will both experience *Reynolds slip* (named after Osborne Reynolds, 1842–1912), due to the progressive stretching of the surface within the contact region. Figure 9.6 illustrates a cross-section through a cylinder, rolling over an originally plane surface under load. The surface between points B and D has been stretched to accommodate the cylinder. Reynolds envisaged that the strain in the surface would be greater at point C than at B or D, and that there would therefore have to be some differential movement between the surface of the plane and the surface of the cylinder as the cylinder rolled. He suggested that frictional energy dissipation associated with this slip was the cause of rolling resistance. It is now clear, however, that Reynolds slip is of minimal importance even in the rolling of steel spheres on rubber where the surface strains can be quite high; in practical rolling bearings it plays an even smaller role.

The second type of slip occurs only with spherical rolling elements and not with cylindrical rollers. Figure 9.7 shows both in plan view and in section a sphere rolling

Fig. 9.6 Cross-section through a sphere or cylinder rolling over an elastic surface, illustrating the origin of Reynolds slip



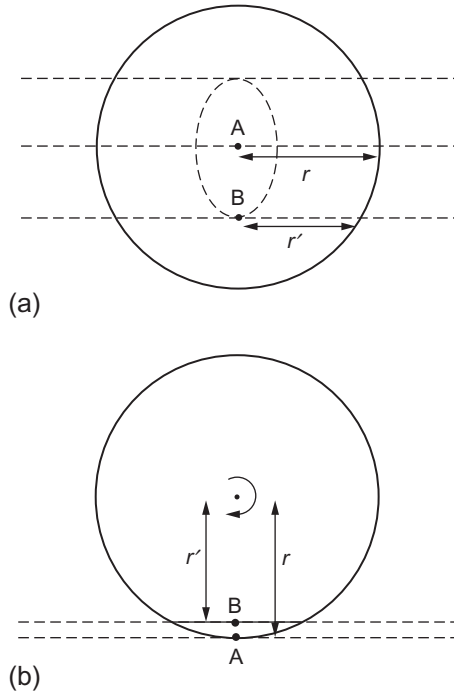


Fig. 9.7 (a) Plan view and (b) section through a sphere rolling over a bearing track, to illustrate the origin of Heathcote slip

under load in a cylindrical groove. The contact zone is elliptical. If the sphere rotates through one complete revolution, then point A on its surface will have travelled a distance $2\pi r$, while point B has travelled the smaller distance $2\pi r'$. Since the sphere moves as a rigid body, slip must therefore have occurred within the contact region in order to accommodate this difference; this differential slippage is called *Heathcote slip*. While this simple picture illustrates the origin of Heathcote slip, more detailed analysis is needed to establish exactly which areas within the contact patch do indeed slip, and in which areas the displacements are accounted for by elastic strains.

Friction in the areas undergoing Heathcote slip makes a considerably greater contribution to rolling resistance than does Reynolds slip, but neither is important in the overall energy balance in a rolling bearing. Slip in the contact region will, however, cause surface damage which can lead to severe wear or early failure through surface-initiated fatigue if the bearing is unlubricated; an important function of the lubricant in a rolling bearing is therefore to provide a film thick enough to separate the surfaces, usually in the elastohydrodynamic regime, and thereby reduce wear and surface damage. Grease and oil (see [Section 4.3](#)) are both used to lubricate rolling bearings. Oil offers lower frictional drag and also transfers heat away from the bearing, a valuable function at high speeds, while grease needs less complex means of sealing and replenishment and is widely used. Both oils and greases give some protection

against corrosion. Suitable synthetic oils can be used with tool steel bearings at high temperatures, while for even higher temperatures, with ceramic bearings, solid lubricants may be useful (see [Section 4.7](#)).

9.2.2 Fluid film lubricated bearings

Introduction

Plain bearings can be broadly classified into two types: those in which a lubricant (usually oil) is supplied continuously in order to produce a hydrodynamic (or hydrostatic) lubricating film during operation, and those operating under more marginal conditions of lubrication, where the lubricant is either oil or grease, supplied intermittently and retained in the bearing structure, or a solid lubricant incorporated into or even constituting the bearing material. We shall consider the first type, the fluid film lubricated bearing, in this section, while marginally lubricated and dry bearings are discussed in [Section 9.2.3](#).

As we have seen in considering hydrodynamic lubrication ([Section 4.4](#)), in a journal bearing operating under fully hydrodynamic (or hydrostatic) conditions the journal and shaft will be widely separated by the lubricant film, and there should ideally be no contact between them. It might therefore be thought that the materials of the bearing housing and shaft would be unimportant. When the shaft is not rotating, however, the lubricant film is squeezed out from the bearing contact, and the bearing system must be able to operate under very poorly lubricated conditions on starting, stopping, and possible extreme overloading. The bearing material must therefore exhibit *compatibility* with the journal material, to tolerate limited direct contact without gross seizure. Furthermore, the lubricant will inevitably become contaminated with solid particles, perhaps from atmospheric dust, or from wear debris. If these particles are larger than the thinnest lubricant film encountered in the bearing, then abrasive wear may occur. Plain journal bearing materials can tolerate some abrasive contamination if they have good *embeddability*: that is, if they will allow abrasive particles to embed into their surface, so that they do not circulate repeatedly with the lubricant causing abrasive wear of both journal and bearing. In many applications, journal bearings are subject to repetitive loading which results in high cyclic pressures in the lubricant film; these can lead to fatigue failure of the bearing material, which should therefore possess good *fatigue strength* as well as high *compressive strength*. Since bearings, shafts and machine structures are all built to finite tolerances, and all will distort to some extent under load, there may be some misalignment between a journal and its housing. A good plain bearing material will tolerate some misalignment if it shows *conformability*. Finally, bearing materials should resist *corrosion* by the lubricant, even though it may be contaminated with water or acidic oxidation products, or, for example, with combustion products in an internal combustion engine.

These properties of compatibility, embeddability, fatigue resistance, compressive strength, conformability and corrosion resistance are needed by all plain bearings, although their relative importance depends considerably on the conditions of load and speed under which the bearing is required to operate. No single material offers

ideal properties in all these areas, since in some cases they are mutually incompatible: a high compressive strength, for example, implies a high hardness, which in turn implies poor embeddability. The choice of a bearing material involves a compromise, and different materials are therefore suitable for different applications.

We can separate applications of lubricated bearings into two types: those of low severity, in which the average bearing pressure is low or moderate and in which the bearing often carries a steady load, and those characterized by high pressures and large cyclic loads. Conditions of the first type are found in the shaft bearings of large rolling mills, steam turbines and low-speed diesel engines, often for marine use, while the second more severe conditions are associated with the crankshaft bearings of high speed petrol (gasoline) and diesel engines.

In nearly all applications the journal running in the bearing is steel; it is a more expensive component to replace than the bearing sleeve, and wear must be avoided. It is therefore often hardened, particularly if it runs against one of the harder bearing materials. Small diameter shafts, for example, can be nitrided, which increases both the surface hardness of the journal and its fatigue resistance (see [Section 7.3.2](#)).

Low stress applications

In applications involving low bearing pressures and large bearings, some degree of misalignment is inevitable, and it is particularly important that seizure of the bearing on starting and stopping is avoided. Stresses, however, are low. Compatibility and conformability are therefore more important than strength in the bearing material, and the *whitemetals* are very widely used. Whitemetals, also known as *babbitts* after Isaac Babbitt who patented the use of pewter (a tin alloy) as a bearing metal in 1839, are *tin-based* and *lead-based* alloys with unrivalled compatibility, conformability and embeddability.

Tin-based whitemetals, historically the earliest compositions but still widely used, are alloys of tin with antimony and copper. A typical composition is 89% tin, 7.5% antimony and 3.5% copper (British Standard BS3332/1, ASTM B23 Alloy 2); several other compositions are also used and are covered by national standards. Cast alloys containing up to 8% copper and less than about 8% antimony contain needles of Cu_6Sn_5 , often in a characteristic star-like formation, together with some particles of SbSn , in a solid-solution (tin–antimony) matrix. The proportion of the copper–tin constituent increases with copper content. Primary precipitation of the copper–tin needles tends to form a network which prevents the antimony–tin particles from segregating. Higher antimony contents (over about 8%) lead to the formation of primary cuboids of SbSn , although the precise microstructure depends on cooling rate and will also, due to segregation effects, vary with location within the casting. The lead content in all tin-based whitemetals is kept low (typically below 0.5%) to avoid the formation of the low melting point lead–tin eutectic (M.P. 183 °C).

Lead-based whitemetals, while historically important, are progressively being replaced by tin-based alloys for environmental reasons. They contain tin and antimony, with arsenic in some compositions. A typical arsenic-free alloy contains 10% antimony and 6% tin (ASTM B23 Alloy 13, SAE 13); its microstructure comprises

primary cuboids of SbSn, embedded in a matrix of lead–antimony–tin. Segregation of the antimony–tin cuboids can occur in casting. The addition of arsenic strengthens the alloy, particularly at high temperatures. A typical arsenic-containing lead-based alloy contains 16% antimony, 1% arsenic and a nominal 1% tin (ASTM B23 Alloy 15). Its microstructure contains fine particles of antimony in a solid-solution matrix.

Both tin-based and lead-based whitemetals commonly contain hard particles (e.g., antimony–tin cuboids or copper–tin needles) in much softer solid-solution matrices. The coefficients of friction for the unlubricated sliding of these materials against steel do not, however, appear to be influenced by the presence of the particles. Measured friction values against steels are quite high without lubrication: typically $\mu = 0.5$ to 0.8.

Whitemetals show very good compatibility and conformability, with the tin-based alloys being somewhat more resistant to seizure. Tin-based whitemetals can be susceptible to thermal fatigue if subjected to wide-range thermal cycling; this results from the anisotropy of the thermal expansion of tin (which has a body-centred tetragonal structure) and does not occur in lead-based alloys. Lead-based whitemetals, on the other hand, can corrode severely in the presence of acids (although to some extent protected by alloying with tin and antimony), whereas tin-based whitemetals are more corrosion-resistant.

The use of whitemetal bearings is limited largely by their strength, and, particularly in applications involving cyclic loading, by their fatigue strength. Although the strengths of whitemetals depend to some degree on their composition, they vary more strongly with the thickness of the bearing metal employed. Whitemetal bearings are cast, usually by gravity or centrifugally; replaceable steel bearing shells lined with whitemetal are commonly used, although in some applications in heavy machinery the alloy may be directly cast in-situ. As a result of the plastic constraint introduced by the backing, a thin layer of whitemetal bonded as a lining on to a steel backing has a considerably higher compressive strength than a thicker lining. Similar effects are seen on the fatigue strength: Fig. 9.8 shows the increased fatigue strength obtained as the thickness of the whitemetal layer is reduced.

In more demanding applications, very thin linings of whitemetal on steel or bronze backings are used. However, as the lining becomes thinner, so its tolerance to abrasive particles and to shaft misalignment diminishes. In situations where the bearing pressures are high, and where there is a large dynamic component to the loading, whitemetal bearings are inadequate and the materials described in the next section must be used.

High stress applications

Perhaps the most demanding application in which hydrodynamically lubricated plain bearings operate is as crankshaft bearings in the modern high speed internal combustion engine, discussed further in Section 9.3.2. This environment is severe: the bearings carry large cyclic loads at high rotational speeds. To achieve good thermodynamic efficiency, the running temperature of the engine is high; the bulk lubricating oil temperature may be up to 130 °C, and still higher temperatures (perhaps

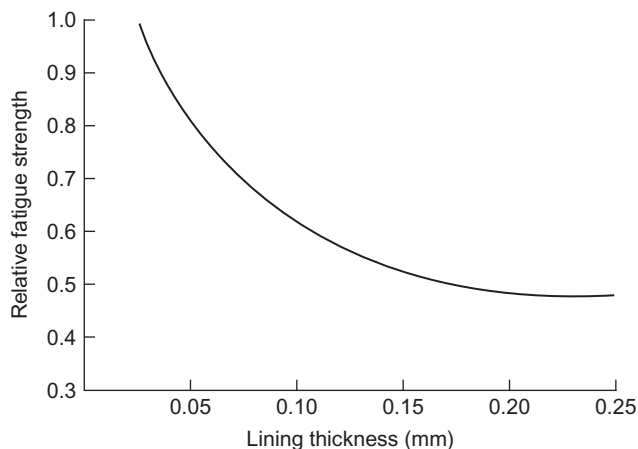


Fig. 9.8 Illustration of the increase in relative fatigue strength associated with a reduction in thickness of a whitemetal bearing alloy supported on a steel backing (data from Neale, M.J. (Ed.), 1973. *Tribology Handbook*. Butterworths, London)

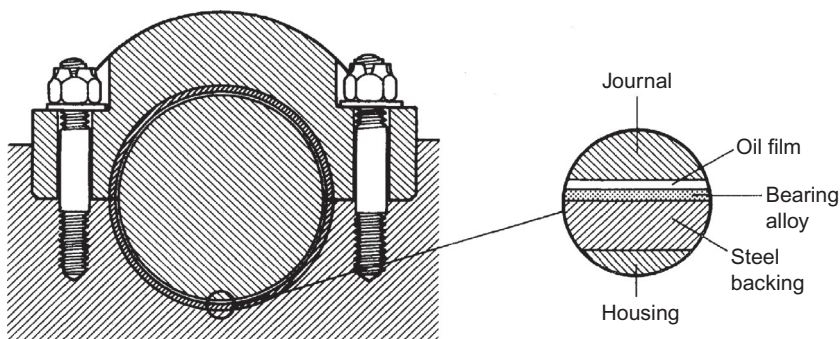


Fig. 9.9 Typical construction of a thin-wall crankshaft bearing in a reciprocating internal-combustion engine. The bearing consists of a thin layer of bearing alloy supported on a replaceable steel shell

180°C) will be attained within the hydrodynamic film. The maximum oil film pressures generated on the power stroke can be over 200 MPa, and the bearing material requires good compressive strength and fatigue resistance. Embeddability, conformability and compatibility are also important, since the minimum oil film thickness during running can be less than 1 μm , which may be significantly smaller than the size of the abrasive contamination, or the elastic deflections of the crankshaft.

To meet these stringent requirements composite bearings are used as shown in Fig. 9.9. The bearing, known as a *thin wall bearing*, consists of a thin steel shell, typically 2 mm thick, lined with one or more layers of bearing material with a total thickness of about 300 μm . The shell is split into two to allow assembly, and is formed from

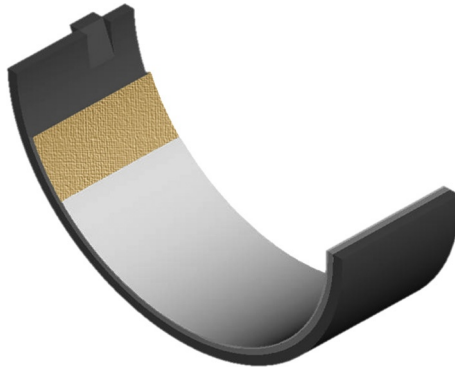


Fig. 9.10 Structure of a typical thin-wall bearing shell. The steel backing carries a layer of bearing alloy (in this case, copper-based), with a thin overlay coating (courtesy of MAHLE Engine Systems UK Ltd.)

low- or medium-carbon steel strip to standard dimensions for interchangeability. [Figure 9.10](#) shows an example of a typical bearing shell. The increased strength resulting from the use of a thin lining on a more rigid backing has been mentioned in the previous section. Bearing alloys commonly used as linings in these applications are *copper-based* (traditionally copper–lead or leaded bronze, but being rapidly supplanted for environmental reasons by bismuth bronze) or *aluminium-based* (aluminium–tin or aluminium–tin–silicon). The copper alloys have higher fatigue strength than the aluminium alloys, and therefore used for crankpin (big-end) bearings which experience higher cyclic pressures than main bearings, and for main bearings in diesel engines. Aluminium alloys perform well at the lower bearing pressures encountered in the main bearings of petrol (gasoline) engines (and also in crankpin bearings in low-powered engines).

Copper–lead bearing alloys typically contain about 10%–25% lead, while bismuth bronzes contain about 4% bismuth. The alloy is cast or sintered on to steel strip which is then press-formed into the bearing shell. Lead and bismuth are both miscible with copper in the liquid state in the proportions used, but their solid solubility is effectively zero. The structures of these alloys therefore consists of a relatively strong continuous copper-rich phase interspersed with regions of much softer lead or bismuth, as illustrated in [Fig. 9.11\(a\)](#). The alloy in this form exhibits low sliding friction against steel in the dry state ($\mu \approx 0.2$), due to the lubricating action of the thin film of soft metal which becomes smeared out over the surface (see [Section 4.7](#)).

Thin wall bearings lined only with copper alloys are used in some undemanding applications such as small end bushes (between the connecting rod and the gudgeon pin, as described in [Section 9.3.2](#)). Substantial improvements in bearing performance can be gained from the application of a further thin *overlay coating*, typically 10–12 μm thick, on top of the copper-based lining, and the bearings used for engine crankshafts are usually overlay coated. The overlay is chosen to be much softer than the underlying bearing alloy, and a variety of materials and processing routes is used. Pure tin can be applied by electroplating (see [Section 7.4.1](#)), and aluminium–tin alloy,

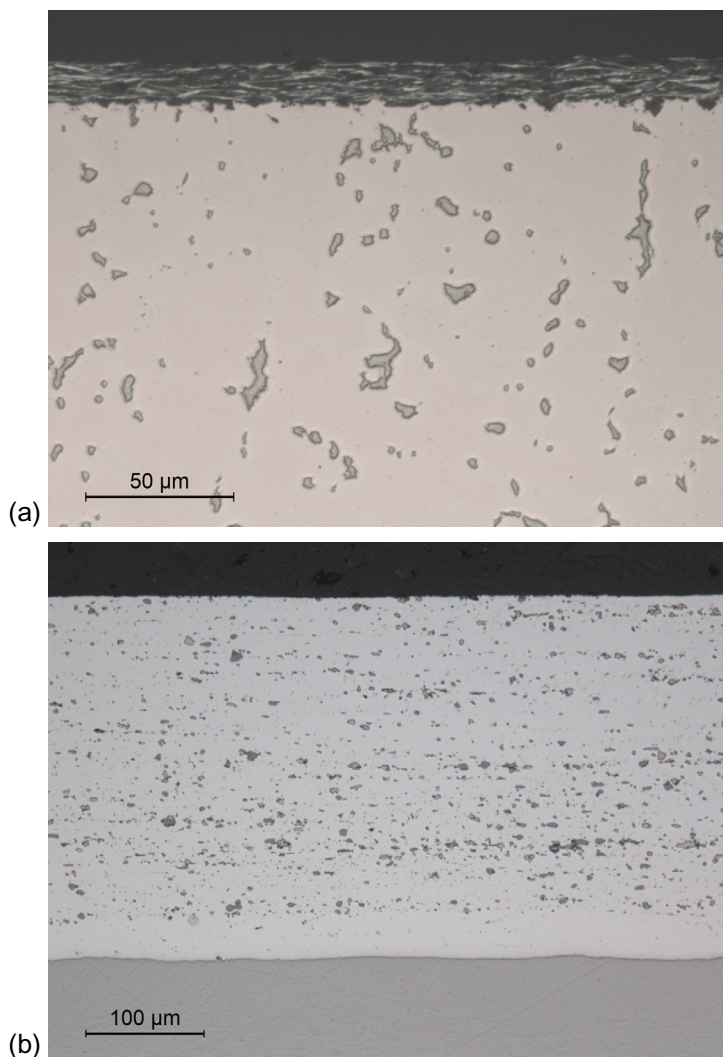


Fig. 9.11 (a) Example of a cast bismuth bronze bearing alloy (CuSn+4% Bi+4% Ni) with a polymer composite coating (PAI resin+metallic flakes+PTFE). (b) Example of an aluminium–tin bearing alloy (Al+6% Sn+2% Si) roll-bonded on to a steel support. A thin intermediate layer of aluminium or aluminium alloy is used to enhance the bond strength and is visible in this cross-section above the steel shell (courtesy of MAHLE Engine Systems UK Ltd.)

which offers the best performance, can be deposited by PVD (sputter) coating (see [Section 7.4.3](#)). The most recent development is the use of polymer-based coatings, typically based on polyamide-imide resin (PAI), harder particulate filler (e.g., metal or metal oxide particles) and a solid lubricant (e.g., PTFE, MoS₂ or hexagonal boron nitride); they are applied by spraying from a solvent carrier followed by curing.

An example is shown in Fig. 9.11(a). Overlay coatings improve the compatibility and embeddability of the bearing, and also appreciably increase its fatigue strength; they may also provide some corrosion protection. The polymer overlays offer certain benefits over electroplated tin in terms of performance, while being less expensive to apply than the sputtered coatings.

For lower loads *aluminium-based* alloys are used, such as aluminium–20% tin. Tin and aluminium have very low mutual solubility in the solid state, and the structure again consists of a very soft phase dispersed in a stronger matrix, as illustrated in Fig. 9.11(b). Higher matrix strength and greater wear resistance is achieved by alloying with copper (up to 2%) or silicon (up to 6%). These bearings have better embeddability and seizure resistance than the copper-based alloys, and will function satisfactorily without an overlay coating; furthermore, they are very resistant to corrosion. However for demanding engine applications, such as stop-start or hybrid technology involving many more starting cycles than traditional operation, polymer-based overlay coatings are used to reduce the wear rate associated with repeated exposure to boundary-lubricated conditions.

9.2.3 Marginally lubricated and dry bearings

Introduction

In many applications, adequate bearing performance can be obtained from plain bearings that do not operate under fully hydrodynamic conditions. These bearings, which may be only intermittently lubricated, are termed *marginally lubricated*, the mechanisms of lubrication being partly hydrodynamic, but also involving contributions from mixed, boundary and sometimes solid lubrication. The lubricant may be oil or grease. For some uses, plain bearings containing solid lubricants are used without any external supply of lubricant; these are called *dry bearings*. Some marginally lubricated and dry bearings can carry loads as high as rolling element bearings and fluid film plain bearings, but they are more limited in speed, as shown in Fig. 9.2; their major advantages lie in their small dimensions and low cost, both for the bearing itself and for the means of supplying the lubricant. Because these types of bearings operate with low lubricant film thicknesses (whether oil, grease or solid lubricant), embedment of large hard contaminant particles into the bearing surface can lead to rapid abrasive wear of the shaft, and care should be taken to avoid contamination in their installation and use.

Marginally lubricated bearings are constructed either as solid components or as porous metal parts. Dry bearings are nearly always solid. A frequently used guide to the capacity of a marginally lubricated or dry bearing is the value of the product PV , where P is the mean bearing pressure (i.e., load divided by projected bearing area) and V is the sliding velocity; maximum values of PV are often quoted by manufacturers of proprietary bearing materials. In the absence of more detailed design information, the PV value is useful, but it can be misleading; a better guide is provided by a plot of P against V for some specified lifetime or total wear, as shown for example in Fig. 9.12. At high load, the life of the bearing will be limited by its strength, whereas at high speed it tends to be limited by frictional heating; the value of PV multiplied by the

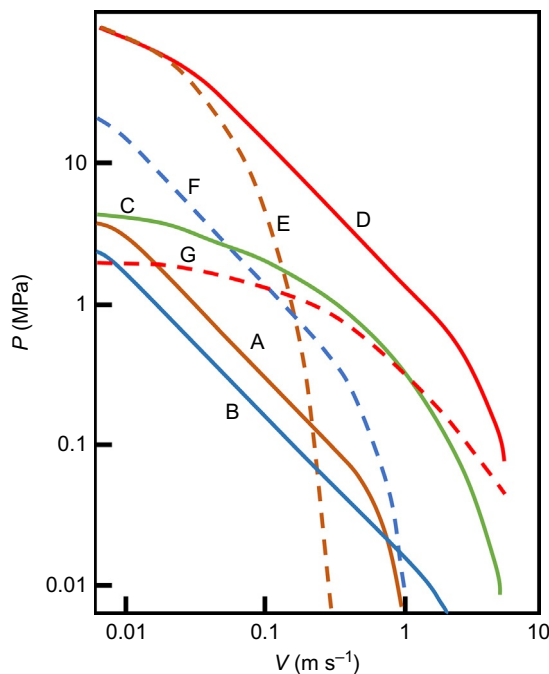


Fig. 9.12 Examples of P - V curves showing the limits of operation for various types of dry bearing bushes. The curves relate to journal bearings with a wear rate of $25 \mu\text{m}$ per 100 h (under unidirectional load). A: thermoplastics, B: PTFE, C: PTFE + fillers, D: Porous bronze + PTFE + Pb, E: PTFE/glass/thermoset composite, F: reinforced thermoset + MoS_2 , G: Thermoset/carbon-graphite + PTFE (from Neale, M.J. (Ed.), 1993. *Bearings: A Tribology Handbook*. Butterworth Heinemann, London)

coefficient of friction is the frictional power dissipated per unit area. Only in the intermediate regime does the specification of a maximum PV value provide a reasonably reliable criterion for design.

Marginally lubricated bearings

A wide range of solid bearing materials is used under marginally lubricated conditions; selection must depend on many factors, including the load, speed and nature of the movement (whether continuous or oscillatory), the means of lubrication, cost and methods of fabrication and installation.

Solid bronze (copper-tin) bushes are commonly used, often containing lead or, increasingly, bismuth as a replacement for lead for environmental reasons. These alloys are readily available in bar and tube form and are therefore suited to low-volume production of bearing bushes by simple machining operations. The load capacity of leaded bronze bearings decreases with increasing lead content, while their conformability and embeddability increase. The higher strength alloys with lower lead content tend to be favoured for low speed, grease lubricated operation, while higher lead contents are used for higher speeds with oil lubrication. Typical compositions are 89% copper–10% tin–0.5% phosphorus (unleaded phosphor bronze), 80% copper–10% tin–10% lead (medium-lead tin bronze), 70% copper–5% tin–25% lead (high-lead tin bronze) and 83% copper–8% tin–7% bismuth–2% zinc (bismuth tin bronze); many other compositions are also used, including the copper-lead alloys and

lead bronzes described in [Section 9.2.2](#) above. The lead or bismuth phase present in the leaded or bismuth bronzes results in a low coefficient of dry friction (see [Sections 3.5.4](#) and [4.7](#)), and is therefore beneficial under marginally lubricated conditions. Further improvements in bearing capacity under conditions of poor lubrication can be achieved by incorporating graphite as a solid lubricant into the bearing material. Proprietary materials are available containing from 4% to 14% by mass (10% to 40% by volume) of graphite, finely dispersed in a leaded or unleaded bronze matrix. These materials are fabricated by powder metallurgical techniques, and can be used at up to 200 °C. For higher temperature use, where conventional lubricants are unusable, similar composite materials containing graphite and iron or graphite and nickel are available, which can be used without external lubrication at up to 600 °C.

Other bearing materials described in [Section 9.2.2](#), such as the aluminium–tin alloys and the tin-based and lead-based whitemetals, are also sometimes used for marginally lubricated applications. All these materials can be used in the form of solid bushes, or at higher loads as relatively thin linings on steel or bronze backings.

Porous bronze or iron bushes, impregnated with oil, are widely used. These are pressed to shape from powder and then sintered, to give an interconnected porosity of 10%–25%. They are then impregnated with oil under vacuum before use. Bronze is a common material, often containing some graphite as a solid lubricant (e.g., 89% Cu, 10% Sn, 1% graphite); porous bushes in this material can be run against unhardened steel shafts, and on account of their good thermal conductivity function well at high rotational speeds. Porous iron bushes (e.g., pure iron, with 2%–25% copper) are stronger, but have lower thermal conductivity; they are therefore more suitable for higher load, lower speed applications. Because the material is harder, a hardened steel shaft must be used to avoid heavy wear. The mechanism of lubrication in an oil-impregnated porous bush is, ideally, hydrodynamic; the hydrodynamic pressure distribution in the bearing when it is running leads to continuous circulation of oil through the pores. They are therefore better suited to running at high speeds (and correspondingly low loads) than to lower speeds where the lubrication will be poorer. As the porosity of the bush is decreased, so its load capacity increases; on the other hand, high porosity produces a large inbuilt reservoir of oil and allows it to circulate more readily, resulting in better lubrication. [Figure 9.13](#) illustrates the regimes of speed and load over which porous metal bushes are useful. A maximum PV value of $1.75 \text{ MPa m s}^{-1}$ is sometimes specified for these types of bearing.

Solid polymer bushes are used in some marginally lubricated, lightly loaded applications. Unfilled thermoplastics, particularly polyamides (nylons), polyacetals (POM) and PBT, are suitable only for low loads; PV limits of $0.05 \text{ MPa m s}^{-1}$ are typical. Better performance under marginal lubrication results from the incorporation into these polymers of a solid lubricant. Polytetrafluoroethylene (PTFE) or molybdenum disulfide (MoS_2) are common additives, while the bulk strengths of these materials can be increased by the use of inert fillers such as glass fibres. Reinforced thermosetting polymers (e.g., polyester or phenolic resins) are also used in some applications. Solid thermoplastics bearings can be injection moulded and are therefore cheap to produce; the performance of these materials can be considerably improved, at the expense of more complex fabrication, by bonding them as a thin layer to a metallic backing.

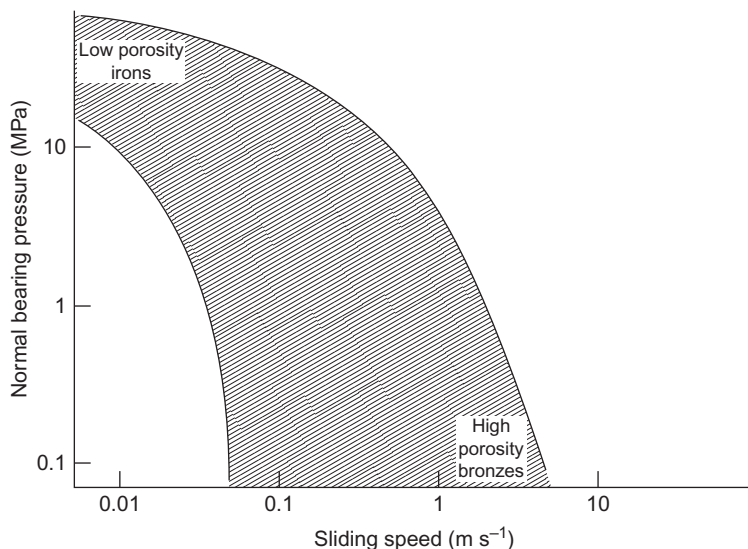


Fig. 9.13 Regime of speed and pressure over which porous metal bushes are useful as marginally lubricated bearings (data from Waterman N.A., Ashby M.F. (Eds.), 1991. Elsevier Materials Selector, vol. 1. Elsevier.)

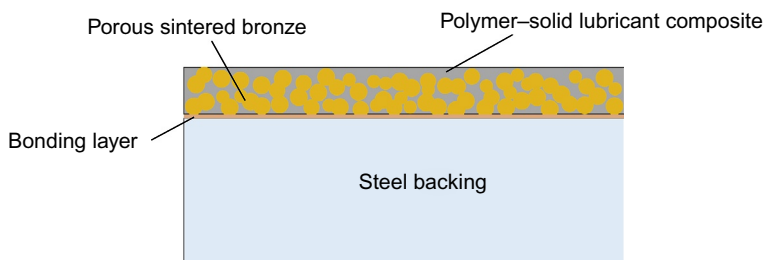


Fig. 9.14 Schematic diagram showing the structure of a typical marginally-lubricated metal-polymer composite bearing. A porous bronze network sintered on to a steel backing strip is infiltrated with a composite material containing a thermoplastic polymer and solid lubricants

Commercial composite bearings of this type are made by sintering a porous layer of bronze powder (typically 0.2–0.3 mm thick) on to a steel backing strip, previously copper plated to provide a good bond, and then infiltrating the bronze layer with a thermoplastic material. Polyacetal is a common choice; a mixture of polyvinylidene fluoride (PVDF), PTFE and lead powder is also used, or a lead-free composite of polyetheretherketone (PEEK), PTFE and graphite. The microstructure of a typical material of this type is shown schematically in Fig. 9.14. Such bearings can operate at PV values up to 10 MPa m s^{-1} , and show good embeddability.

Dry bearings

All the bearing materials discussed in the previous section are lubricated to some extent by externally supplied grease or oil. They may be lubricated only before use, with the lubricant retained within the bearing sufficing for its whole life, or they may be relubricated at intervals. In some applications, for example in the food processing industry, or for a bearing operating in vacuum or at high temperature, the use of oil or grease may not be possible, or the avoidance of periodic lubrication may be attractive. For these and other reasons *dry bearings* are used. These incorporate a lubricant in their composition, and no external supply of lubricant is required. PTFE, molybdenum disulfide and graphite are commonly used solid lubricants (see [Section 4.7](#)), and dry bearing materials are usually composites containing one or more of these lubricants.

PTFE is a good solid lubricant, but extremely soft, and must be reinforced for use as a bulk bearing material. Common fillers are glass fibres, bronze particles, MoS_2 or graphite; the composite material must be prepared by powder processing methods since PTFE cannot be processed as a melt. Solid bearings made from these materials (e.g., a composite of PTFE, graphite and bronze powder) exhibit low friction and good wear behaviour when not externally lubricated; however, the maximum loading is low, with limiting PV values of about 0.3 MPa m s^{-1} for reasonable life. Substantially better performance is given by composite thin films on steel backing shells, with structures very similar to that shown in [Fig. 9.14](#). A typical material, for static bearing pressures up to $\sim 250 \text{ MPa}$ and temperatures up to 280°C , consists of an intimate mixture of PTFE and lead powder, impregnated into a porous bronze layer sintered to a steel strip. Here the normal load is largely supported by the bronze particles, and sliding occurs in a thin film of PTFE at the surface. The lead is thought to play a role, together with copper from the bronze, in the formation of a stable transfer film of PTFE on the steel counterface (see [Section 5.11.3](#)). Newer formulations are available that eliminate the lead content. Low values of μ are attained, typically $0.02\text{--}0.1$, although these composite materials, like bulk PTFE, do not obey Amontons' Laws. There is a substantial dependence of μ on both load and sliding speed and the lowest coefficient of friction is found at low sliding speed and at high load. Sliding speeds are usually limited to $\sim 2 \text{ m s}^{-1}$ to avoid excessive frictional heating. However, as a result of the compressive strength imparted by the bronze matrix, composite bearings of this type can be used at PV ratings of up to $\sim 2 \text{ MPa m s}^{-1}$.

Many other composite dry bearing materials containing PTFE are used. Some contain other solid lubricants as well (e.g., graphite or MoS_2) and a wide variety of other constituents have been employed (e.g., epoxy resins, polyimide in bulk or fibre form, glass or carbon fibres, phenolic resins, polyetheretherketone—PEEK).

For operation at high temperatures, where organic materials would decompose, graphite and molybdenum disulfide are useful lubricants. The graphite-bronze, graphite-iron and graphite-nickel composites discussed above for marginally lubricated conditions will also operate dry at high temperatures, and bearings of solid graphite are also sometimes used. Many other proprietary composite materials have been developed for high temperature use; some for example, incorporate gold as a

solid lubricant, while others use molybdenum disulfide in a refractory metal matrix (e.g., molybdenum, tungsten, tantalum or niobium). Materials of this latter type can be used in air at up to 400°C, their operation being limited by oxidation, while in vacuum or inert gas they will function at up to 1200°C. Other metal–matrix composites, and carbon–carbon composites, also offer exciting potential for the future.

9.3 AUTOMOTIVE TRIBOLOGY

9.3.1 Introduction

Passenger cars provide the dominant mode of road transport in much of the world. About 70 million new cars are produced each year, with nearly all (in 2017) being fuelled by petrol (gasoline) or diesel. They account for about 45% of the global energy used in transport, which itself accounts for about 20% of primary energy consumption. At the time of writing, widespread use of vehicles powered by alternative energy sources such as hydrogen or stored electrical energy seems to be some way off.

A significant proportion of the energy used by a passenger car is dissipated in friction, as shown in Fig. 9.15. Of the chemical energy generated by combustion of the fuel, typically 38% is converted into mechanical work and most of this is used to overcome frictional resistance. A relatively small amount (some 5% of the initial fuel energy) is used to overcome air drag. Frictional losses in the engine and transmission represent about 16% of the total fuel energy, with a further 11% being accounted for by the rolling resistance of the tyres and 5% dissipated in braking. While improvements in aerodynamic design and innovations such as regenerative braking may lead to some reduction in hydrocarbon fuel use, it is clear that there is also significant

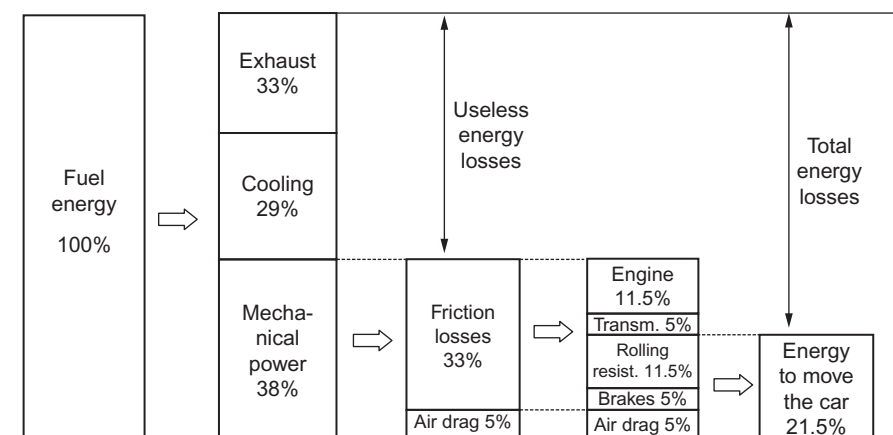


Fig. 9.15 Breakdown of passenger car energy consumption (from Holmberg, K., Andersson, P., Erdemir, A., 2012. Global energy consumption due to friction in passenger cars. *Tribol. Int.* 47, 221–234)

potential for tribological improvements. In this section we shall examine some important applications of tribology in the context of automotive engineering.

9.3.2 Engine components

The reciprocating internal combustion engine used in the vast majority of passenger cars has evolved considerably over the past century. It contains a rich variety of sliding and rolling contacts, all of which dissipate energy and experience wear. The engine designer needs an understanding of tribology to reduce these effects to acceptable levels.

Figure 9.16 shows the structure of a typical four-cylinder direct-injection petrol (gasoline) engine; the basic construction of a diesel engine is very similar. In the standard four-stroke cycle, downward motion of the piston in the *induction stroke* draws air into the cylinder; in the upward *compression stroke* the air is compressed and mixed with fuel injected under high pressure; chemical reaction of this fuel–air mixture, initiated by a spark in a petrol engine or by the heating caused by compression in a diesel (compression-ignition) engine, is followed by expansion of the resulting hot combustion products which drives the piston downwards in the *power stroke*; and in the subsequent upward *exhaust stroke* the combustion products are expelled from the cylinder. The *connecting rod* links the piston to the crankshaft, so that as the piston reciprocates in the cylinder it causes the crankshaft to rotate. Valves sliding in guides

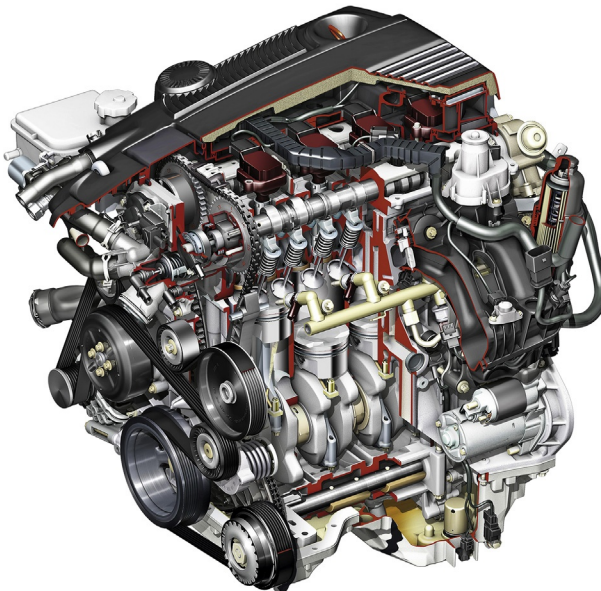


Fig. 9.16 Cutaway image of a modern 4-cylinder direct-injection petrol (gasoline) engine (Courtesy of Daimler-Benz AG)

in the cylinder head control the inlet of air and the outlet of exhaust gases at the correct points in the cycle; in this example there are two *inlet valves* and two *exhaust valves* per cylinder. The valves are opened and closed by the action of cams on one or more continuously rotating *camshafts* which act either directly on the valves or operate them indirectly through cam-followers and rocker arms, collectively known as the *valve train*. Other (but by no means all) components within the engine and its accessories through which frictional energy is dissipated include the roller chain or timing belt which drives the camshafts, the oil, coolant and fuel pumps, seals, electrical generator, power steering pump, air conditioning compressor and associated drive belts. The internal components of the engine are lubricated by oil stored in the sump at the bottom of the engine and circulated to the various regions by a pump. As discussed in [Section 4.3](#), the fully-formulated automotive lubricant used in an engine consists of a base oil and several additives with different functions to ensure that it provides effective lubrication in all the tribological contacts.

The dominant source of frictional energy dissipation in an engine (about half of the total) is associated with the piston-cylinder contact. Of the remaining power dissipated by friction, roughly one third can be ascribed to the valve-train, and two thirds to the crankshaft bearings. As shown in [Fig. 9.17](#), the clearance gap between the piston and the cylinder is sealed by a set of (usually) three rings, carried in grooves in the piston, which press against the cylinder bore. In order to reduce lubricant consumption caused by oil from the engine sump (beneath the piston) traveling up past the piston into the combustion chamber, a good seal is needed, but close contact has the undesirable effect of increasing the frictional losses. The ring and the groove geometry are designed so that high pressure in the combustion chamber provides additional radial force on the top two rings (known as the compression rings) to increase their contact pressure. Axial load in the connecting rod results in piston-cylinder side loading since

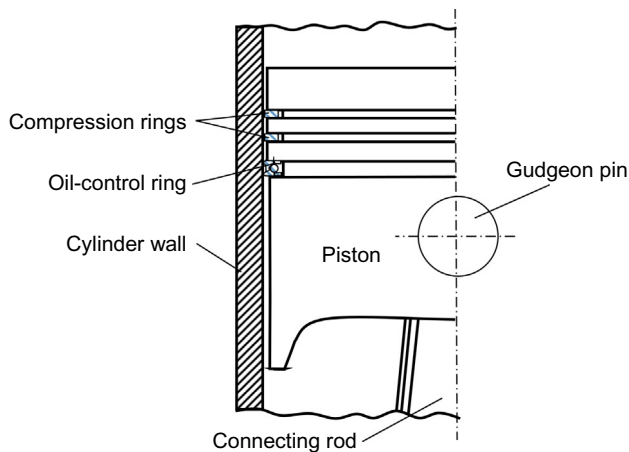


Fig. 9.17 Piston assembly and piston ring—cylinder contacts (after Tung, S.C., McMillan, M.L., 2004. Automotive tribology overview of current advances and challenges for the future. *Tribol. Int.* 37, 517–536)

(for almost all of the cycle) the axis of the connecting rod is not parallel to the axis of the cylinder. The axial loads vary with the power being developed and the engine speed and also with the position in the stroke, with the largest forces occurring in the combustion stroke. The angle between the connecting rod and cylinder axes also varies with position in the stroke, with this being largest around the middle of the stroke; however, in the middle of the piston stroke, the piston is also moving quickly, and the lubrication of the ring-cylinder contact can be hydrodynamic. At the extremes of the stroke where the motion reverses (termed top dead-centre and bottom dead-centre), the sliding speed falls to zero and the contact experiences mixed lubrication. The film thickness can then be $<0.2\text{ }\mu\text{m}$, giving a film thickness ratio λ (see [Section 4.5](#)) of <1 . Additional sliding contact, at much lower contact pressure, occurs between the skirt of the piston and the cylinder wall, and here the lubrication is hydrodynamic; there is also some squeeze-film effect.

Effective lubrication, good sealing and low frictional losses are all desirable in the piston ring-cylinder contact; careful attention to the geometry of the rings and grooves, the materials of all the components and the topography of the bore surface (which often carries a diagonal grooved pattern formed by a *plateau honing* process) can all play a part in improving engine efficiency and reducing wear. Typical materials for piston rings are grey cast iron or nodular iron for solid rings, as well as steel for top compression rings and oil control rings. In some cases they are coated, for example by chromium electroplating or flame-sprayed molybdenum to reduce wear. Cylinder liners are usually grey cast iron or an aluminium alloy, while aluminium–silicon alloy is the most common piston material.

[Table 9.1](#) lists typical tribological parameters for a petrol engine, and [Fig. 9.18](#) shows a modified Stribeck curve, plotted in terms of the film thickness ratio λ , which indicates the range of operating conditions for various engine components.

The crankshaft is supported by several *main bearings*, and the connecting rods rotate about the offset crankpins on *big-end bearings*. The crankshaft is usually made

Table 9.1 Typical tribological parameters for a petrol engine (data from Priest, M., Taylor, C.M., 2000. Automobile engine tribology—approaching the surface. *Wear* 241, 193–203)

	Crankshaft bearing	Piston ring/liner (top compression ring)	Cam/follower
Minimum oil film thickness	$<1\text{ }\mu\text{m}$	$<0.2\text{ }\mu\text{m}$	$0.1\text{ }\mu\text{m}$
Maximum temperature	180°C	200°C groove, 120°C liner	150°C
Minimum oil viscosity	2.5 mPa s	6.5 mPa s	EHL
Maximum shear rate	10^8 s^{-1}	10^7 s^{-1}	10^7 s^{-1}
Composite surface roughness (R_a)	$0.35\text{ }\mu\text{m}$	$0.2\text{ }\mu\text{m}$	$0.3\text{ }\mu\text{m}$
Maximum pressure	60 MPa	70 MPa	600 MPa
Power loss (typical)	0.08 kW per bearing	0.15 kW per ring	0.04 kW per cam

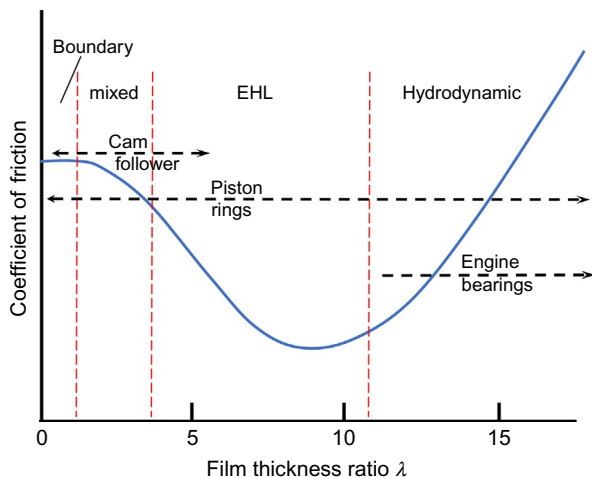


Fig. 9.18 Modified Stribeck diagram, plotted in terms of film thickness ratio λ , showing typical ranges of lubrication regime for various engine components (after Priest, M., Taylor, C.M., 2000. Automobile engine tribology—approaching the surface. *Wear* 241, 193–203)

from forged steel or, for less highly stressed engines, spheroidal graphite cast iron; the bearing surfaces of steel crankshafts are typically induction hardened or nitrided (see [Section 7.3.2](#)). Crankshaft bearings are examples of plain journal bearings, and form the most important example of the high-stress application discussed in [Section 9.2.2](#) above. Although crankshaft bearings are exposed to continuous rotational motion, the loads on them are not steady but vary with the crank position during the engine cycle and the varying forces on the connecting rods which depend on the power being generated by the engine and its speed. Analysis of the lubricant film thickness and local pressure within the bearings is therefore complex. As shown in [Fig. 9.18](#), the lubrication regime is predominantly hydrodynamic, although film thicknesses as low as $1\text{ }\mu\text{m}$ or even less can be reached which put the contact into the mixed regime over at least part of the cycle.

The small end of the connecting rod is joined flexibly to the piston through a hollow cylindrical *gudgeon pin* (US *wrist pin*, *piston pin*) made from hardened steel. The pin may be fixed to the connecting rod and rotate in an oscillatory motion in two bearing bushes within the piston, or alternatively be fixed in the piston and rotate in a bush in the connecting rod, or in some designs can rotate in both components. With relatively low sliding speeds, high loads and oscillatory rather than continuous rotation, these journal bearings operate under boundary or mixed lubrication conditions.

The camshaft rotates in plain journal bearings that are hydrodynamically lubricated. The most severe conditions in the valve train occur at the contact regions between the cams and their followers, where the high load and concentrated contact geometry lead to very high contact pressures and to elastohydrodynamic lubrication (EHL), as discussed in [Section 4.5](#). As noted for piston-cylinder contacts, the tribological conditions are most severe at the reversal of the point of contact between the cam and follower where the lubricant entrainment velocity falls to zero. Materials with high yield stress are required for the cams and their mating components, and

hardened steels are common. Low friction and wear in the cam-follower contact depend on the presence of both friction modifiers and anti-wear additives in the lubricant (see [Section 4.6](#)) and surface coatings such as diamond-like carbon (DLC—see [Section 7.4.3](#)) are also sometimes used to reduce friction under these conditions of poor lubrication.

9.3.3 Transmission components

Details of the design of automotive gearboxes are outside the scope of this book, but both manual and automatic gearboxes contain numerous bearings, typically both rolling bearings and plain bushes, and other sliding surfaces as well as the contacts between the teeth of the gears themselves. In addition, automatic transmissions contain multi-plate or band clutches and manual gearboxes have synchronizer rings whose smooth action is important in achieving satisfactory changes of gear ratio. These diverse components are lubricated by formulated oils (*transmission fluids*) containing appropriate additives, designed to be compatible with the materials used and to achieve both low wear and frictional losses in the sliding and rolling contacts. The clutches that are actuated to change the overall gear ratio in automatic gearboxes, in which steel surfaces slide against counterbodies carrying a layer of *friction material*, require particular friction characteristics for optimum operation. These combine a low static coefficient of friction with a high and stable dynamic coefficient of friction, and are achieved by the use of friction modifiers in the transmission fluid (see [Section 4.6](#)) and non-metallic composite friction materials that are similar in some respects to those used in brakes and discussed in [Section 9.3.4](#).

Other transmission components may include final drive gears and differential gears, as well as driveshaft joints such as *constant velocity joints* as illustrated in [Fig. 9.19](#). Constant velocity joints play an essential role in transmitting the driving torque (and in some cases braking torque) to the wheels whose positions and orientations are varied by steering and suspension movements, and can accommodate articulation angles (between the input and output shafts) of more than 45°. The joint consists of steel inner and outer races with accurately profiled hardened tracks. The torque is applied via a splined shaft which fits into the inner race, and is transferred to the outer race by hardened steel balls located in tracks between the two components, and held in place by a cage. Material requirements for the balls and races are similar to those for rolling bearings, but the contact stresses can be even higher: typically up to 8 GPa. The relatively slow reciprocating motion of the balls and the high pressures mean that these contacts operate under conditions of boundary lubrication. The joint is lubricated with a grease containing both friction modifier and antiwear additives to reduce operating power loss and wear, and sulfurized esters and organic phosphates may also be included as EP additives. The joint is surrounded and protected from the environment by a flexible elastomeric polymer ‘boot’, and compatibility with this imposes further demands on the lubricant.



Fig. 9.19 Example of an automotive constant velocity joint, which allows torque to be transmitted efficiently through a rotating, articulating coupling (courtesy of GKN Driveline plc)

9.3.4 Brakes

Automotive disc brakes decelerate a vehicle by dissipating its kinetic energy in sliding contacts between pads of *friction material* and *brake discs* or *rotors* that rotate with the wheels. An example of a typical brake assembly is shown in Fig. 9.20; sliding between the pads and the discs occurs at about half the speed of the vehicle, and the pads typically cover 10%–15% of the area of the track they describe on the disc.

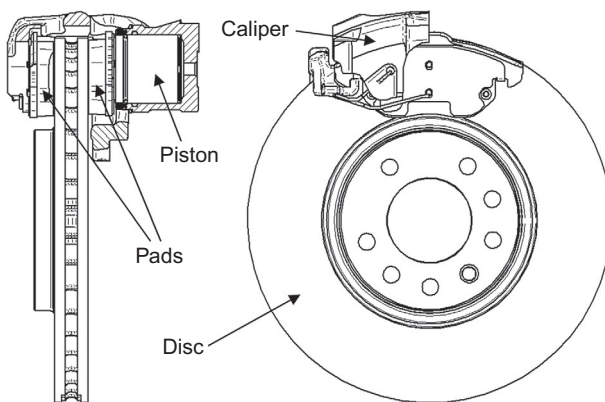


Fig. 9.20 Disc brake assembly with a single-piston floating caliper and a ventilated disc (from Wahlström, J., 2011. A study of airborne wear particles from automotive disc brakes (Doctoral thesis). Department of Machine Design Royal Institute of Technology, Stockholm, ISBN 978-91-7415-871-7, with permission)

The coefficient of friction between the pad and disc in a road car is typically about 0.4. Particularly high friction is not important; what is desirable is a stable and consistent value that does not fall significantly as the temperature of the pad is increased (a phenomenon known as *brake fade*). Large amounts of energy are dissipated at the sliding interface, at up to 30 kW under severe braking conditions. While temperature rises of 150–300 °C are common, significantly higher temperatures occur in both the pads and discs in steep mountain descents or competitive use. In normal passenger cars the discs are usually made from grey cast iron, while the brake pads consist of composites containing many (sometimes 10 or more) different constituents, compacted by hot-pressing into a solid mass with 5%–10% porosity. There is wide variety in the composition of these composite friction materials, but the constituents can be broadly classified as binders, fibres, frictional additives (lubricants and abrasives) and fillers. Table 9.2 provides a summary of the functions of each constituent, as well as some examples.

The nominal pressure between the brake pad and the disc during braking is typically between 1 and 10 MPa. However, the true contact area is only a small proportion of the total pad area, typically 15%–20% for moderate braking loads, distributed over a number of small islands (plateaux) that protrude by a few µm above the level of the rest of the pad surface, as shown in Fig. 9.21. The islands are typically 50–500 µm in size and are formed from protruding hard constituents (such as fibres) surrounded by softer pad materials and wear debris. Isolated contact spots form and disappear rapidly within each island. The frictional behaviour of the pad-disc system thus depends on the composition, microstructure and properties of small areas on the pad as well as the nature and properties of the surface of the disc; both will vary with the history of the contact and the complex response provides a good example of the need for a systems approach in analyzing tribological problems (see Section 8.3).

Table 9.2 Constituents of typical automotive brake friction materials, their functions, and typical examples (information from Chan, D., Stachowiak, G.W., 2004. Review of automotive brake friction materials. Proc. Inst. Mech. Eng. D J. Automob. Eng. 218, 953–966; Eriksson, M., Bergman, F., Jacobson, S., 2002. On the nature of tribological contact in automotive brakes. Wear 252, 26–36)

Description	Function	Examples
Binder	Form thermally stable matrix	Thermoset phenolic resin
Fibres	Provide mechanical strength	Brass, steel, Kevlar, glass, ceramics (formerly asbestos)
Lubricant	Stabilize friction, especially at high temperature	Graphite, metal sulfides
Abrasive	Increase friction, clean surface film from disc	Zirconia, alumina, chromium oxide, metal silicates
Filler	Improve manufacturability and reduce cost	Mica, vermiculite, barium sulfate

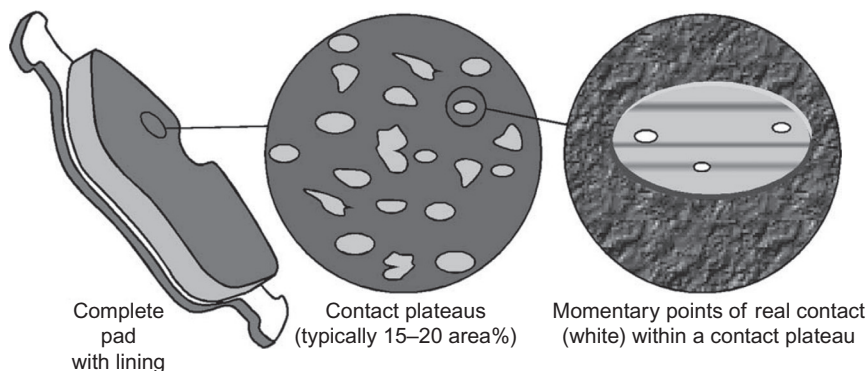


Fig. 9.21 Illustration of the real area of contact between an automotive brake pad and disc (from Eriksson, M., Bergman, F., Jacobson, S., 2002. On the nature of tribological contact in automotive brakes. *Wear* 252, 26–36)

9.3.5 Tyres

The pneumatic tyre that is universally used for motor vehicles has evolved considerably since its invention in the nineteenth century. Figure 9.22 shows a radial-ply tyre in cross-section. It is a composite structure made up from several components with different functions and properties. Most important from a tribological viewpoint are the tread which makes contact with the road surface, and the sidewalls through which the driving, braking and steering forces are transmitted from the wheel to the tread. Rubber compounds (natural rubber and various synthetic elastomers, blended with a range of fillers and additives) are used extensively in the tyre structure,

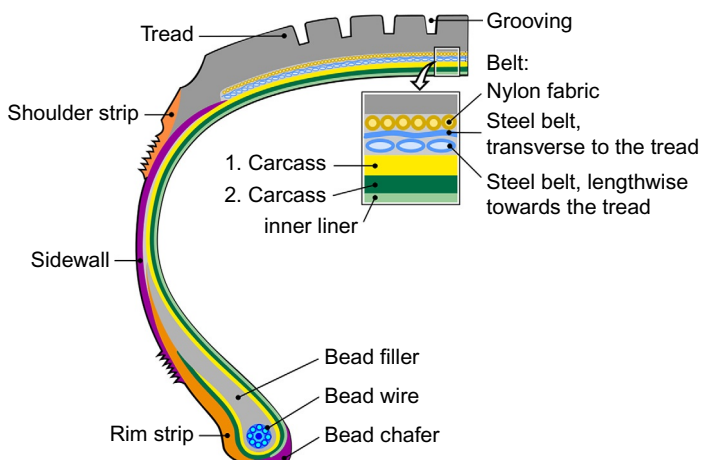


Fig. 9.22 Cross-section showing the structure of a radial-ply car tyre (from MagentaGreen, Wikimedia Commons, CC BY-SA 3.0)

together with reinforcing fibres and cords which may include steel wires, drawn polymers and glass fibres, in some cases woven into fabrics. In a radial-ply tyre, the predominant orientation of the reinforcing cords in the carcass is close to 90° to the circumferential centre-line of the tyre, so that within the sidewall they are situated radially with respect to the circle of the wheel.

With a typical inflation pressure of about 0.2 MPa the tyre provides a relatively large contact area with the road surface which allows high accelerating, braking and cornering forces to be transmitted, and also adds compliance to the suspension of the vehicle. On a smooth hard road surface the coefficient of rolling friction for a modern passenger car tyre when correctly inflated is about 0.01, although it is quite common for tyres in use to be under-inflated, which in combination with a rough road surface gives an average coefficient of perhaps 0.02. There is continuing effort to reduce this value, since a 10% reduction in rolling resistance has been estimated to lead to a 2% reduction in energy demand or fuel consumption.

Figure 9.23(a) shows a schematic diagram of a tyre in contact with the road surface. As the tyre rolls towards the left, the initially undistorted sidewall at A moves down towards the bottom of the tyre (point B) and deforms elastically. As the tyre rolls further, the strains in the sidewall then relax, and by point C it has regained its original geometry. This cyclic deformation of the tyre structure, nearly all accommodated by flexing of the sidewall, is not perfectly elastic but involves viscoelastic dissipation of a small amount of energy which appears as heat and contributes to the rolling resistance of the tyre. The process is analogous to the energy dissipation that is responsible for the deformation component of friction in polymers, as discussed in Section 3.8.2.

As shown in Fig. 9.23(b), the tread deforms macroscopically as it flexes to enter and leave the contact patch, leading to viscoelastic losses that make up the second contribution to the tyre's rolling resistance (typically more than half the total). Low rolling resistance thus requires *low* viscoelastic loss. The surface of the tread also distorts to some extent when it is loaded against the protruding aggregate particles, typically a few mm in size, that make up the surface of the roadway. The real area of contact between the tread

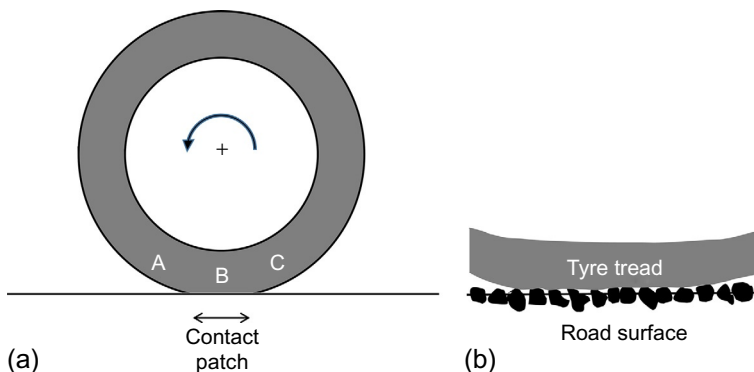


Fig. 9.23 Contact between a pneumatic tyre and a road surface: (a) overall view showing contact patch; (b) deformation of the tyre tread as it contacts the road

and the road surface is usually significantly less than the apparent area. A high coefficient of sliding friction is desirable between the tread surface and the road, since this allows high braking and cornering forces to be transmitted while avoiding skidding; to achieve this, *high* viscoelastic losses are desirable at the tread-aggregate contacts.

The apparent conflict between these two requirements of low viscoelastic loss to reduce the rolling resistance, and high viscoelastic loss to achieve high sliding friction against the road surface can be resolved, however, because the characteristic frequencies of deformation in the two cases are different. The frequency (1/timescale) associated with flexing of the tread as it passes through the contact patch is V/L where L is the length of the contact patch (typically ~ 100 mm for a passenger car) and V is the speed of the vehicle. The frequency of deformation of the tyre tread skidding at the same speed V over hard protuberances of length d is V/d where d is typically 1–10 mm. Thus the two frequencies differ by a factor of 10–100. In tyre design, not only are the tread compounds selected to have low loss at the lower frequency and high loss at the higher, but rubber compounds with different, tailored properties are used for the various structural elements within the tyre; in the case of the tread compound, wear rate is another important consideration.

9.4 TRIBOLOGY IN MANUFACTURING

9.4.1 Metal-cutting

Subtractive processes are very widely used in manufacturing, in which material (usually metal) is removed from a crudely-shaped initial part (perhaps a bar or slab, or a forging or casting with a more complex shape) to achieve the final shape and dimensions. Lathe-turning, milling, drilling, sawing, thread-cutting and gear-cutting are some examples of subtractive processes involving metal-cutting that are traditionally carried out in specialized machines. The modern trend is toward integration of several processes into a single machining centre under digital control, in which loading and unloading of the workpiece is avoided between the individual sequential steps.

Figure 9.24 shows a schematic diagram of *orthogonal machining* which illustrates the essential elements involved in all these cutting processes. In some cases such as simple turning the operation involves a single cutting edge, while in others several cutting edges are engaged with the workpiece simultaneously. The geometry is defined by the depth of cut and the rake angle; in this diagram the rake angle has a small positive value, but zero or negative rake angles are also used. The tool moves to the left relative to the workpiece and as it does so a layer of material is removed from the top surface of the workpiece and flow up the rake face of the tool, forming a chip. This material is deformed by plastic shear as it passes through the *primary shear zone*, which in the simplest models is represented by a single shear plane (represented by a broken line in the diagram). The cutting process is exactly the same as we have seen in Fig. 6.11(a) in the context of abrasive wear, except that in that diagram the rake angle is negative.

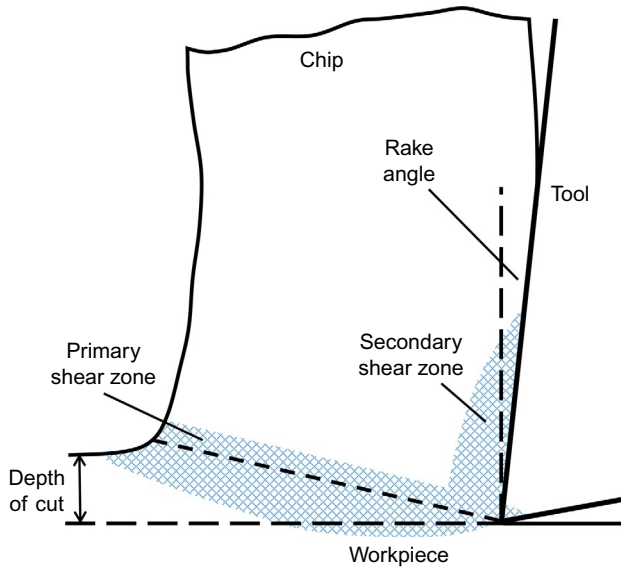


Fig. 9.24 Simplified diagram showing the geometry of orthogonal machining by a single-point tool, and associated chip flow and shear zones (after Childs, T.H.C., et al. 2000. *Metal Machining*. Butterworth Heinemann, London)

Typically one half to three-quarters of the mechanical work involved in cutting is dissipated as heat through plastic work in the primary shear zone, and of that perhaps 85% is carried away in the chip, with the remainder being conducted into the workpiece. The rest of the mechanical work is associated with friction at the interface between the rake face of the tool and the chip. Conditions there are severe: the surface of the chip is freshly cut from the bulk workpiece and therefore at least initially free from any oxide film, and the forces exerted by the tool are sufficiently high to cause plastic flow in the metal in the primary shear zone. Depending on the geometry and material of the cutting tool, the lubrication conditions (if any), the workpiece material and the cutting conditions, there may be adhesion between the chip and the tool over the lower part of the rake face (in which case the chip material shears close to the tool surface), or there may be sliding over the whole contact length. These conditions result in additional plastic work in a *secondary shear zone* extending along the chip from the tip of the tool. Some of the heat generated in the secondary shear zone is conducted into the chip, but much of it is conducted into the tool. The temperature distribution in the tool depends on the cutting conditions and the thermal properties of the tool and workpiece material. [Figure 9.25](#) shows typical results from a finite-element model. A relatively small temperature rise occurs in the workpiece, with much higher values in both the chip and the tool. In this example the highest temperature in the tool is some distance from its tip (the cutting edge), but under other conditions (for example with a low-conductivity workpiece material such as a titanium alloy) it is possible for the

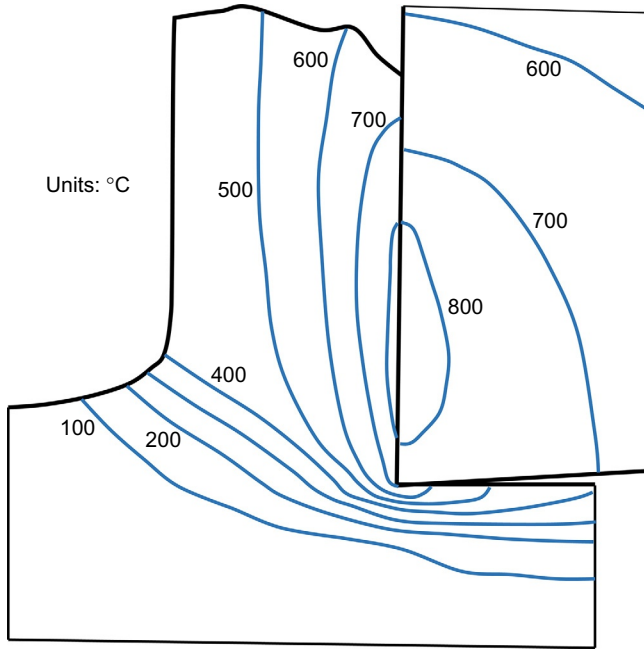


Fig. 9.25 Example of computed temperature contours within the workpiece, chip and tool during single-point machining (after Childs, T.H.C., et al. 2000. *Metal Machining*. Butterworth Heinemann, London)

hottest region to be at the tip. As the cutting speed increases so also does the power dissipated in the process, and the tool temperature rises.

The life of a cutting tool is, ideally, determined by wear processes that cause a steady, progressive and predictable change in its dimensions that can be incorporated into production planning, so that the tool can be replaced or adjusted before the quality of the machining process becomes unacceptable. Tool failure by fracture, in contrast, is less predictable and if it occurs can lead to unplanned downtime and scrapped products. The life of a cutting tool (in terms of the time t for which it can be used before reaching some threshold of wear) is often empirically found to be related to the cutting speed V by Taylor's tool life equation

$$Vt^n = C \quad (9.1)$$

where C is the cutting speed at which $t = 1$, and n is the tool-life exponent. Figure 9.26 shows some examples of tool-life curves, on a log-log plot, for the machining of a low alloy steel with three different types of cutting tool. The values of n are shown. The Taylor equation is obeyed quite well, but at low speeds the ceramic tools failed by a chipping mechanism leading to a much higher value of n . If tool wear followed the

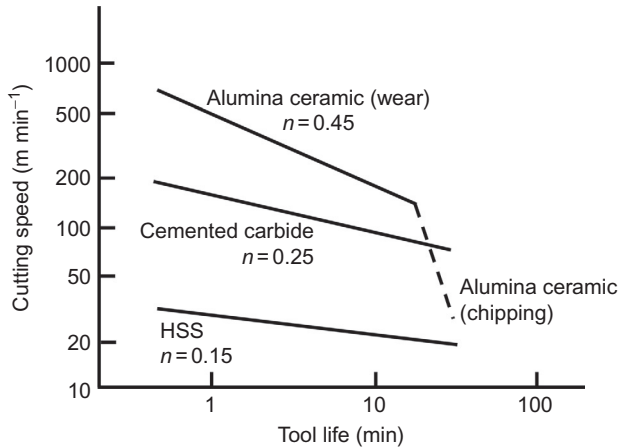


Fig. 9.26 Representative tool-life curves for turning a low alloy steel (after Childs, T.H.C., et al. 2000. *Metal Machining*. Butterworth Heinemann, London)

Archard wear equation, we would expect the tool life to be inversely proportional to the cutting speed so that $n = 1$, but for most cutting operations n is lower, between 1 and 0.2. This is because, as we have seen above, as the cutting speed is increased the temperature of the tool material also increases, with an associated reduction in its hardness (or yield stress) and an increased rate of surface chemical reaction (such as oxidation).

As illustrated in Fig. 9.27, several different mechanisms of tool wear can operate which show different dependence of wear rate on temperature. Abrasion (as discussed in Chapter 6), is a mechanical process caused by hard particles in the workpiece, while larger-scale attrition, chipping and fracture may also occur; material loss from these is relatively independent of temperature.

The group of mechanisms classed as ‘thermal damage’, however, becomes rapidly more important as the tool temperature rises. Thermal damage by plastic deformation occurs when the cutting edge becomes hot and too weak to sustain the forces applied to it; tool materials which retain their strength at high temperature are needed to avoid this. Diffusion can occur when the tool and workpiece material are mutually soluble, and leads not only to direct loss of tool material by diffusion into the chip, but also to weakening of the surface material which is then more susceptible to other wear processes. Chemical reaction, for example oxidation of the tool surface, can lead to rapid removal of the reaction product and is an example of a tribochemical wear process (see Appendix B). Thermal softening, diffusion and chemical reaction are all thermally activated processes, and so although they may have different activation energies, we expect tool wear by all of these mechanisms to increase rapidly with temperature.

The combination of a high contact pressure and the lack of oxide on the fresh chip surface favours adhesion (cold welding) of workpiece material on to the rake face of

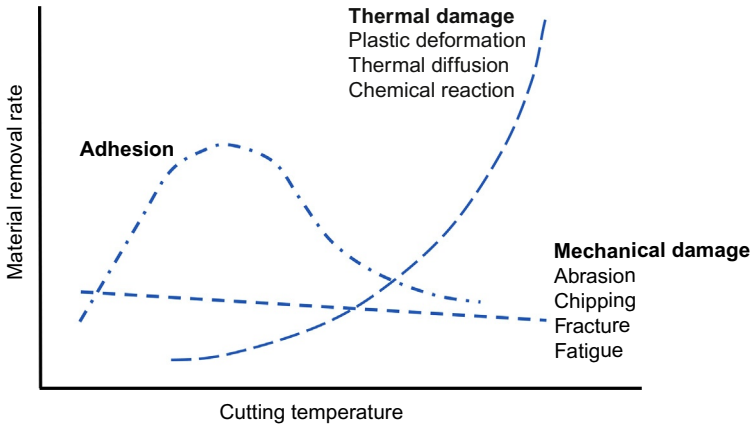


Fig. 9.27 Variation of material removal rate (i.e., tool wear rate) with cutting temperature for different tool damage mechanisms (after Childs, T.H.C., et al. 2000. *Metal Machining*. Butterworth Heinemann, London)

the tool. This adhesion can be so strong that subsequent detachment occurs within the tool material rather than within the chip or at the interface, leading to adhesive wear. This mechanism does not occur so readily at low temperature since the ductility of the workpiece is then lower, while at high temperature detachment occurs preferentially in the chip or at the interface. Thus, as shown in Fig. 9.27, we see a peak in adhesive wear at an intermediate temperature.

The long tool life at high cutting speed desirable for high productivity is therefore favoured by a hard, strong tool material (which will sustain high cutting forces and resist abrasion) that also retains its strength at high temperature, is not soluble in the workpiece material and to which the workpiece material will not adhere. These requirements are difficult to satisfy simultaneously and it is not surprising that a wide range of materials is used in practice. Economic considerations, of course, also play a part. High speed steel (tool steel) is still widely used for tools operating under moderately demanding conditions, while higher loads and cutting speeds, as well as stronger workpiece materials, require more expensive cemented carbides (e.g., WC/Co and more complex materials such as Ti(C,N)/TaC/Co/Ni) or ceramics (e.g., alumina, silicon nitride and various ceramic-matrix composites). Diamond (polycrystalline and single-crystal) and cubic boron nitride, examples of 'super-hard' materials, are used for some specialized applications. Carbide and ceramic tools are often coated with materials such as TiN, TiC or (TiAl)N or more complex multicomponent/multilayer systems by CVD or PVD processes (see Chapter 7), in order to reduce tool wear by adhesion and diffusion, as well as reducing friction between the tool and the chip. The benefits of tool coatings can be very substantial and most carbide tools are currently coated.

Liquid cutting fluids are employed in many cutting operations, with several benefits. At high cutting speeds, especially for operations in which the tool cuts continuously, their major role is in cooling the tool and to a lesser extent, the workpiece. At lower speeds, and for interrupted cutting, they provide some lubrication at the tool–chip interface, although the continuous creation of fresh chip surface and its movement along the rake face of the tool, coupled with the intimate contact between the two surfaces, make penetration of the lubricant into this region difficult. Other functions of cutting fluids include flushing the chip fragments away from the cutting zone, and providing some corrosion protection (by inhibition) to the workpiece. Cutting fluids used mainly as coolants are predominantly emulsions consisting of very small oil droplets in a continuous water phase, with the high latent heat of evaporation of the water providing a potent mechanism for heat absorption. Fluids used mainly as lubricants are usually oils containing additives such as EP additives, antioxidants and corrosion inhibitors (see [Chapter 4](#)). Handling, recycling and disposal of used cutting fluids involves environmental considerations, and there are benefits in reducing their use and in using environmentally-friendly lubricants wherever possible. Some cutting operations can be performed dry (i.e., without the use of a liquid coolant or lubricant), especially with a suitably coated tool. There are also applications where gas-jet cooling can be used, sometimes combined with lubrication by a fine mist of oil droplets (so-called *minimal quantity lubrication*, MQL) rather than by a continuous jet of lubricant.

9.4.2 Grinding, polishing and chemical–mechanical planarization

Subtractive manufacturing processes involving the use of small abrasive particles share many features with the processes of abrasive wear discussed in [Chapter 6](#).

In *grinding* processes the abrasive particles are bonded into a composite counterbody, usually a wheel, which moves against the workpiece. [Figure 9.28](#) shows

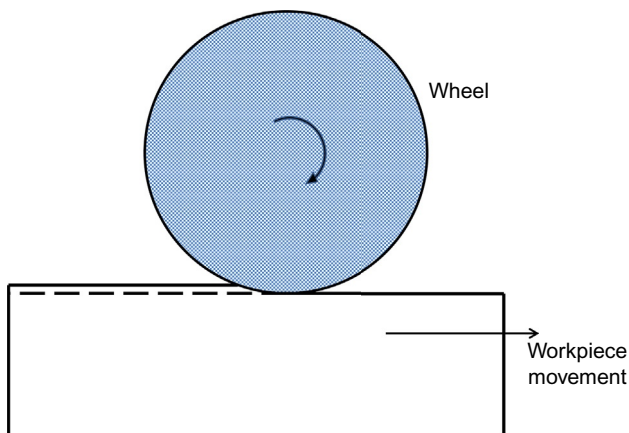


Fig. 9.28 Geometry of a typical grinding process

the geometry of a typical grinding process, in which the surface of a plane workpiece is moved to the right beneath a rapidly rotating cylindrical grinding wheel. A thin layer of material is removed by processes that in the context of wear we would classify as sliding abrasion. The purpose of grinding is either to remove material rapidly in a relatively rough shaping operation (*stock removal grinding*), or to achieve close control of the final component shape and surface finish (*form and finish grinding*—sometimes called *fine grinding*). In fine grinding the depth of cut is small compared with the radius of curvature of the tips of the abrasive particles, which despite their angular shapes act more like rounded particles and present high negative rake angles (in the metal-cutting terminology discussed in [Section 9.4.1](#)) or, equivalently, low attack angles (as defined in [Section 6.3.1](#)). Removal of material is thus relatively inefficient compared with the case of stock removal grinding in which the abrasive particles act like miniature cutting tools, albeit with a range of orientations and rake/attack angles. This difference is reflected in the specific energy associated with the two types of process: the specific energy, defined as the work done per unit volume of material removed (see [Section 6.1](#)), may be four or five times greater for fine grinding than for stock removal grinding. In fine grinding, therefore, very little of the mechanical work is carried away as heat energy by the chips, and especially if no coolant is used, a large proportion of the heat is conducted into the workpiece.

Grinding wheels typically contain about 50% by volume of abrasive particles, with 10% binder material and 40% porosity. Silicon carbide and alumina are widely used as abrasives, with the much harder diamond and cubic boron nitride (CBN) employed for specialized purposes. The hardnesses of these materials are listed in [Table 6.1](#). Depending on the application, the grit particles are between 10 μm and 1 mm in size. Hardness is not the only important property of the abrasive particles, however; angularity, tendency to fracture in use (friability), strength at high temperature and reactivity with the workpiece material also play key roles. Glassy ceramics (vitreous binders), thermosetting resins, other polymers including rubbers, and metals are all used as binder materials. The porosity of the wheel material accommodates detached chips of workpiece material and aids access by the coolant. Fracture of abrasive grains and their detachment from the binder leads to continuous renewal of the cutting edges on the wheel surface. But while this process can maintain good grinding performance it may also change the wheel geometry, and therefore, where close dimensional tolerance is required in form grinding, necessitate re-dressing and re-shaping of the wheel.

Polishing is performed to reduce surface roughness, and *lapping* to achieve precise form; both these finishing processes use loose, fine abrasive particles (typically smaller than 10 μm , and often much smaller) suspended in a liquid to form a slurry. [Figure 9.29](#) provides a generic illustration of these processes as applied to flat components. The workpiece is rotated under controlled downward pressure against a rotating polishing pad or lapping plate. In polishing the pad material, typically a woven or non-woven cloth, is porous and compliant, so that the abrasive particles are free to change their orientation but not translate over the surface of the pad. Lapping is carried out with a more rigid counterface, typically of metal, in which the grit particles become embedded. Shaped laps can be used to achieve complex profiles in the finished workpiece. When sufficiently small abrasive particles are used, material is

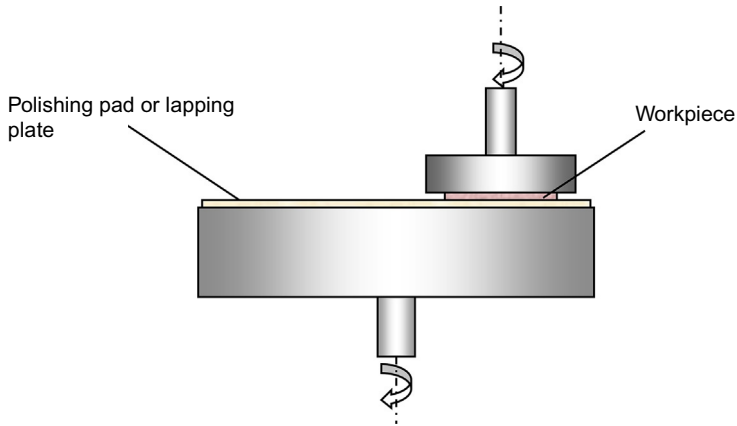


Fig. 9.29 Generic illustration of polishing and lapping processes

removed by ductile mechanisms even from macroscopically brittle workpiece materials such as inorganic glasses and ceramics. One of the earliest quantitative studies of the polishing and lapping of glass was carried out by FW Preston, who in 1927 suggested that the material removal rate (volume removed per unit time) was proportional to the applied pressure on the polishing pad and to the relative sliding speed between the pad and the glass surface, exactly as predicted by equation 6.5 for abrasive wear; in this context it is often called the *Preston equation*.

Although polishing is effective at removing material by purely mechanical abrasive processes, in some applications the polishing rate can be enhanced by the addition of chemical agents to the slurry, so that both chemical and mechanical activity contribute synergistically to a tribochemical process as discussed in [Section 6.3.4](#) and in [Appendix B](#). *Chemical mechanical planarization* (CMP) is an important process in the semiconductor industry that uses this principle, combining simultaneous chemical etching with fine abrasive polishing against porous polymeric pads. As the density of features on semiconductor chips increases, their spacing is reduced and the resolution of the methods used to reproduce the device patterns must increase. In order to reproduce such fine features by photolithography at very short (UV) wavelengths with very small depth of focus, an extremely flat surface is required on the wafer. CMP is used in several essential processing steps in the manufacturing of these devices, and in modelling the process with the objective of predicting and controlling the material removal rate, the ‘Preston equation’ is usually used as an empirical starting point.

9.5 BIO-TRIBOLOGY: NATURAL AND ARTIFICIAL HIP JOINTS

The field of *biotribology*, tribology in the context of living organisms, is very broad. It includes the study of friction between soft tissues within the living body (for example between the heart and lungs and their surrounding structures, and between the cornea

of the eye and the eyelid), sliding motion in the many mobile skeletal joints, and numerous interactions at the surface of the body. Examples of the latter include the behaviour of skin and hair, the tribological phenomena that occur in the mouth between tongue, teeth and food, and the frictional behaviour of footwear. Bio-tribologists have also studied tribological phenomena in non-human animals as well as in plants, including the remarkable adhesive properties of some insect and reptile feet that allow them through a combination of materials and structure to move on vertical and even overhanging surfaces in apparent defiance of gravity. Other topics include friction of surgical gloves and catheters, the lubrication of contact lenses, and the cleaning and wear of teeth. Here we shall restrict our attention to the subject of the synovial joints that are vital to animal locomotion, and examine how, in humans, highly successful methods have been developed to replace diseased joints and restore mobility.

Figure 9.30 illustrates the important features, from a tribological perspective, of a mammalian *synovial joint*. Although their geometry is different in different parts of the human body, numerous similar structures are present in the skeleton, especially in the upper and lower limbs including the hands and feet. The joint articulates by the relative sliding between two bony components; the sliding surfaces are covered with a smooth, soft and porous composite material called *articular cartilage*, composed of collagen (a protein), proteoglycans (proteins bonded to polysaccharides) and water (which makes up about 80% of the total). The thickness of the cartilage layer ranges typically from 0.5 to 1.5 mm in the hand and shoulder to 1 to 6 mm in the hip, knee and ankle joints, reflecting the greater loads carried by the lower extremities. The joint is bathed in *synovial fluid* which lubricates the joint and is retained by a *capsule*. The major components of synovial fluid are water, long-chain proteins, hyaluronic acid

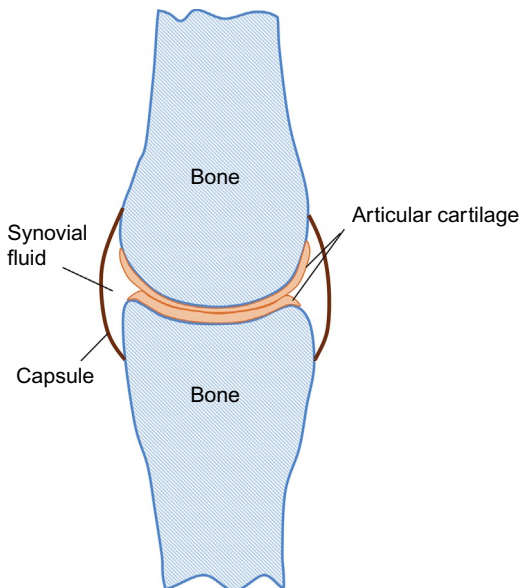


Fig. 9.30 Tribologically important features of a synovial joint

and phospholipids; this natural lubricant is highly non-Newtonian and shear-thins significantly.

Synovial joints exhibit remarkably low friction, with coefficients of friction in the range 0.002–0.006. In a typical engineering context such a low value would only be expected with full-film, hydrodynamic lubrication, but the fact that the low friction in natural joints is observed at very slow sliding speeds, significantly less than 1 mm s^{-1} , and also even after prolonged static loading, suggests that there must be a major contribution from other mechanisms. It seems likely that the lubrication of synovial joints involves several mechanisms that dominate under different conditions: micro-elastohydrodynamic and boundary lubrication probably both play important roles.

The lubricant film thickness in a natural synovial joint predicted from conventional soft EHL theory as described in [Section 4.5](#) is typically $0.1\text{--}1 \text{ }\mu\text{m}$, which is much less than the roughness of the cartilage ($2\text{--}5 \text{ }\mu\text{m}$). This would be consistent with a λ ratio below 1 and suggest the presence of appreciable direct surface contact implying relatively high friction, in contrast to the behaviour actually observed in the joint. However, if the elastic distortion of the compliant surface asperities within the contact region is taken into account in a *micro-elastohydrodynamic* model, a quite different picture emerges: in the region where the lubricant film is thinnest and the pressure highest, the surface roughness is smoothed out substantially, and the relevant value of λ now lies well within the full-film lubrication regime.

While micro-EHL can explain the low friction seen in moving joints, an additional explanation is needed for the persistence of low friction in synovial joints under conditions of prolonged static loading. This further contribution probably comes from the boundary lubricating action of surface-active phospholipid (SAPL) present in the synovial fluid which forms thin layers on the cartilage surfaces. SAPL has been observed to form lamellar layers up to $0.2 \text{ }\mu\text{m}$ thick. Further understanding of the lubrication mechanisms in natural joints will undoubtedly emerge from research into the interactions between their material properties (which are highly complex at various different structural scales) and the properties and behaviour of the synovial fluid.

Well-functioning synovial joints provide free mobility of limbs, hands, feet, fingers and toes, but in a significant proportion of the population their performance is compromised by degradation of the cartilage in the condition known as *osteoarthritis*. Worldwide, 10% of men and 18% of women over the age of 60 have symptomatic osteoarthritis. It is one of the ten most disabling diseases in developed countries, and causes debilitating loss of mobility if it affects the hip or knee. Perhaps the most striking example of a tribological advance in medicine, *total hip replacement* (THR) can restore the function of a diseased hip and is now a routine procedure. About 100,000 patients receive total hip replacements each year in the UK, and more than 300,000 in the USA; annual rates of THR worldwide range up to 300 per 100,000 population. In the natural hip joint the rounded head of the femur (thigh bone) forms a spherical bearing within a conforming socket (the acetabulum) in the pelvis, as seen in [Fig. 9.31](#). The contacting surfaces are coated with articular cartilage. Modern hip prostheses, as also shown in [Fig. 9.31](#), have evolved considerably since the mid-1960s when they were first introduced, but most still replace both components of the natural ball-and-socket joint with a very smooth metallic (stainless steel or cobalt–chromium)

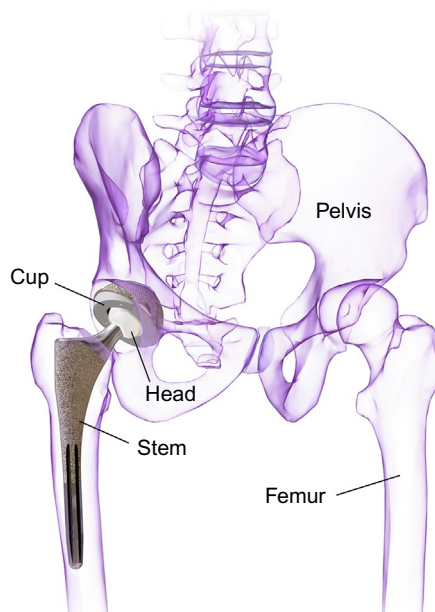


Fig. 9.31 Diagram showing a total hip replacement (hip prosthesis) in situ. The head of the femur is replaced by a ball-shaped head at the end of a stem which is fitted into the femur. The ball articulates in a mating polymeric cup implanted in the pelvis. The stem and head are separate components held together by a taper joint (adapted from Blausen Medical Communications Inc., Wikimedia Commons, CC-BY-SA-4.0)

ball at the head of the femur, articulating in a cup made from ultra-high molecular weight polyethylene (UHMWPE). Smaller, but still substantial, numbers use a ceramic head with a polyethylene cup, or ceramic for both components. Some designs have also used metallic heads sliding in metallic cups. The choice of materials, the detailed design of the prosthesis and the surgical skills with which the joint is selected and implanted all play important roles in the success of the replacement joint.

The hip prosthesis is an application that makes stringent demands on the materials, geometry and construction of its components. The load applied to the hip joint during walking varies in a complex way during the gait cycle, with the characteristic three-peaked shape shown in Fig. 9.32. The first peak in the load occurs when the heel of the foot touches the ground ('heel-strike') at the beginning of the 'stance' phase, followed shortly by a second similar peak as the other foot leaves the ground. The third peak, sometime later, occurs just before heel-strike of the other foot. These peak loads are about four times body weight, leading to a typical peak mean pressure on the hip joint bearing of around 5 MPa. Much higher loads, of course, are imposed by running or jumping. The peak loads coincide with periods of slow movement of the joint, so that it is exposed to a high load combined with low fluid entraining velocity; as we have seen in Section 4.4, this favours a thin fluid film. During the 'swing' phase when the foot is out of contact with the ground the load is much lower, and the entraining velocity is higher, favouring a thicker film.

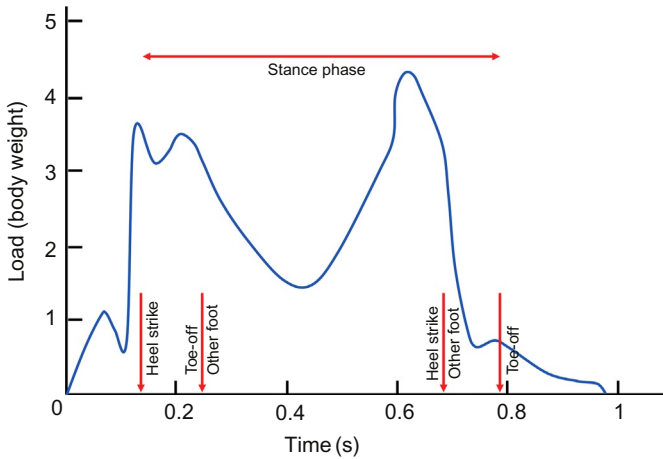


Fig. 9.32 Typical record of total force acting on the hip during walking (after Paul, J.P., 1967. Forces transmitted by joints in the human body. Proc. Inst. Mech. Eng. 181 (Pt 3J), 8–15)

Modern THRs have an annual failure rate of 0.5–1%, with a typical lifetime of 15–20 years. Gross failure of the joint components is avoided by correct mechanical design and by quality control of the materials used; the main reason for joint replacement is associated with wear, specifically the body's reaction to debris produced by wear over a long period. Laboratory experiments can be carried out in which joints are tested under conditions that simulate the complex loading and movement history in the human body, with a protein-containing lubricant (usually diluted bovine serum) which simulates synovial fluid. Results from these, and also measurements of material loss in implanted joints, suggest that metal/UHMWPE joints have typical volume wear rates of the order of 20–30 mm³ per million cycles. The particles of polyethylene debris produced by wear of the acetabular cup are typically 0.1–1 µm in size, and can cause biological reactions in the tissues surrounding the joint that lead to resorption of the bone (*osteolysis*) and thus to loosening of the joint.

The lubrication mechanisms in THRs differ from those in the natural joint. Joints with hard/soft couples (metal or ceramic on polymer) operate toward the boundary end of the mixed lubrication regime, with significant asperity contact, while hard/hard combinations (ceramic/ceramic or metal/metal) show thicker films and for some of the loading cycle may exhibit full asperity separation, with mixed lubrication occurring otherwise. The designer of the prosthesis aims to achieve both low friction and low wear rate. Low friction is needed to limit the torque on the implant; excessive torque would hazard the attachment to the surrounding bone. For a fixed coefficient of friction and applied load, a smaller femoral head will lead to a lower torque on the outer surface of the acetabular cup and hence to a lower stress at this interface. Wear can be expected to increase with relative sliding distance, and thus to reduce the total wear at a fixed specific wear rate, a small femoral head would also be desirable.

However, the friction and wear rate can be expected to vary with fluid film thickness, and to achieve low values a thick film is desired; this will be favoured by a low bearing pressure and a high entraining velocity, both of which are associated with a large head diameter. A larger head is also beneficial in reducing the risk of dislocation. The selection of head size clearly involves a compromise.

Various methods have been used to reduce the wear rate in THR's. In order to increase the λ ratio, the surfaces of the components are made extremely smooth, with a metallic femoral head typically having a roughness of only a few nm *Ra*. Ceramic femoral heads (typically high-purity alumina or zirconia-toughened alumina) may be polished to even lower roughness and offer some advantages. Modification of the UHMWPE is another route: for example, cross-linking of the polymer (to produce XLPE) by gamma or electron-beam irradiation may be beneficial in reducing the wear rate. The irradiation process introduces free radicals into the polymer structure, however, and measures must be used to prevent these leading to oxidative damage and premature failure: methods include melting of the cross-linked polymer or blending of the polymer with vitamin E which acts as an anti-oxidant. Ceramic/ceramic or metal/metal couples have been associated with significantly lower volumetric wear rates than metal/UHMWPE joints, but in all cases the size of the debris particles and their biological activity must also be taken into account. It has been suggested, for example, that the lower wear rate for XLPE than for conventional UHMWPE may be offset to some extent by a greater osteolytic potential from the debris. Despite the total volume of debris being less, the smaller particles may nevertheless present a greater surface area per unit volume, with greater potential for biological activity.

9.6 MAGNETIC DATA STORAGE

The magnetic hard disk drive, a key component for data storage in desktop and laptop computers as well as larger-scale data servers, has increased in capacity by a phenomenal extent since it was first introduced in the 1950s. The binary data is stored as information encoded into magnetized regions on the surfaces of rapidly spinning disks. In modern drives the data is written to the disk and read from it by a read/write head that is supported on a very thin film of air—a gas bearing—as the disk spins beneath it. The first disk drive to use aerodynamically-supported heads, the IBM 1302 of 1965, was the size of a small car and stored 120 MB (1 megabyte = 10^6 bytes) at a cost (in 2017 terms) of almost US\$2 million; in contrast, a palm-sized drive of the present day stores several TB (1 terabyte = 10^{12} bytes) for around \$100, and data capacity is likely to continue to increase. Current data densities exceed 1 terabit per square inch (approximately 1.5 gigabits per square millimetre; 1 byte = 8 bits), equivalent to each bit of information taking up a square area 25 nm by 25 nm. This dramatic increase in storage capacity has been achieved by a progressive reduction in the size and a change in orientation of the magnetized domains in which the data is stored as well as by advances in the physics of the read–write process, but it has also involved a remarkable reduction in the ‘flying height’ of the head over the disk surface. A small gap between the

head and the magnetic material is needed to achieve high lateral resolution in the reading and writing processes. The earliest drives had a head flying height of about $6\text{ }\mu\text{m}$, but in modern drives this has been reduced to less than 10 nm , a feat that demands not only precise modelling of the gas film and head dynamics but also the production of disk surfaces of extreme flatness and smoothness.

Figure 9.33 shows the typical internal arrangement of a hard disk drive, with the head at the end of its support arm; each surface of the disks in a multi-disk stack is accessed by a separate head. The disk typically has a diameter of 63.5 or 88.9 mm (2.5 or 3.5 inches) and rotates at 7200 rpm . The read/write function is performed by a very small active region embedded into a larger slider, $1\text{--}2\text{ mm}$ across. The design of the slider can be complex, as illustrated in Fig. 9.34 which shows one example. The surface of the slider is patterned with steps and ledges typically $0.5\text{--}2\text{ }\mu\text{m}$ in height that are designed to create a distribution of aerodynamic pressure across the slider surface which raises the slider above the disk in a stable manner. Aerodynamic lubrication acts in the wedge-shaped air film between the stationary slider and the moving disk surface. The net pressure difference between the lower and upper surfaces of the slider results in a vertical force that is balanced by the bending of a flexural element in the support arm, and the flying height and attitude are controlled by the equilibrium between the aerodynamic and bending forces. Feedback mechanisms are used to control the head position very accurately. Although most drives currently operate in air, there are potential advantages in terms of reduced flying height and lower frictional drag in using a gas with lower viscosity, and helium is therefore used in a few applications.

The disk platters themselves provide an excellent example of surface engineering for functional applications, combining mechanical, data storage and tribological properties. Multilayer structures, as shown in Fig. 9.35, are deposited on a substrate made from rolled aluminium–magnesium alloy (AA5083; 96% Al, 4% Mg) which is highly polished, with a typical roughness R_a of $15\text{--}25\text{ nm}$. Electroless plating (see Section 7.4.1) is used to deposit a relatively thick ($10\text{ }\mu\text{m}$) nickel–phosphorus layer (90% Ni, 10% P) which acts as a hard substrate for the functional magnetic layers

Fig. 9.33 Internal arrangement of a magnetic hard disk drive, showing a read/write head attached to the end of its support arm. The arm is moved across the surface of the disk so that the head can access circular tracks at various distances from the centre of the disk. Each surface of the three disks in this example has its own read/write head (Eric Gaba, Wikimedia Commons, CC-BY-SA-3.0)



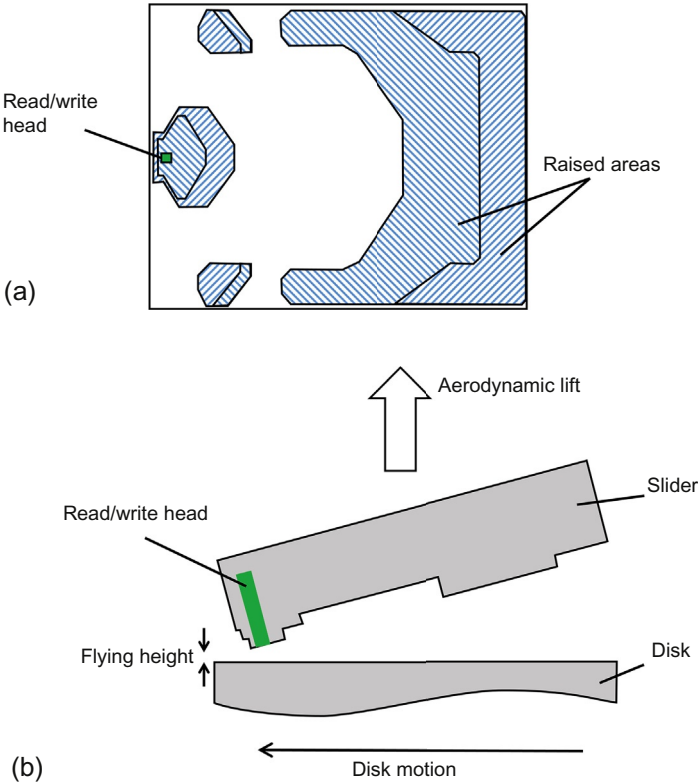


Fig. 9.34 Example of a slider and magnetic disk read/write head, showing: (a) the lower surface of the slider, patterned with raised areas and recesses to control the airflow across its surface and the resulting pressure distribution and aerodynamic lift; and (b) a cross-section through the slider and head, showing (exaggerated) the wedge-shaped air film formed between the slider and the disk (based on patent US 8174795)

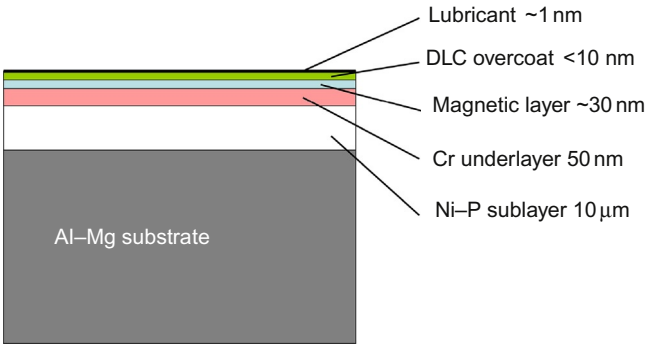


Fig. 9.35 Schematic cross-section through the surface region of a typical hard disk for magnetic data storage (not to scale), showing the multiple coatings required for magnetic, mechanical and tribological functions

and is also polished (to 0.5–2 nm *Ra*). The phosphorus increases the hardness of the deposit. A sequence of very thin layers is then deposited by PVD processes: a sublayer of chromium (~50 nm) to bond the magnetic layer to the Ni–P layer; the magnetic layer in which the data is stored (~30 nm of a complex alloy, often Co-based, sometimes itself with a multilayer structure); and an even thinner (~15 nm) hydrogenated diamond-like carbon (DLC) overcoat which protects the magnetic layer from mechanical damage. Finally an extremely thin (~1 nm) layer of lubricant (typically perfluoropolyether, with very low vapour pressure) is deposited by dip-coating from solution, again to protect the disk surface from damage if the head touches it. Both the lubricant layer and the DLC layer provide some corrosion protection to the magnetic film.

SOURCES AND FURTHER READING

- ASM, 1991. 10th ed. *Metals Handbook: Friction, Lubrication and Wear Technology*, vol. 4. ASM International.
- Barwell, F.T., 1979. *Bearing Systems: Principles and Practice*. Oxford University Press.
- Bhushan, B., 2001. Macro- and micro-tribology of magnetic storage devices. In: Bhushan, B. (Ed.), *Modern Tribology Handbook*, vol. 2. CRC Press, pp. 1413–1513.
- Brockett, C., Williams, S., Jin, Z., Isaac, G., Fisher, J., 2006. Friction of total hip replacements with different bearings and loading conditions. *J. Biomed. Mater. Res. B Appl. Biomater.* <http://dx.doi.org/10.1002/jbm.b.30691>.
- Chan, D., Stachowiak, G.W., 2004. Review of automotive brake friction materials. *Proc. Inst. Mech. Eng. D J. Automob. Eng.* 218, 953–966.
- Childs, T.H.C., Maekawa, K., Obikawa, T., Yamane, Y., 2000. *Metal Machining*. Butterworth Heinemann.
- Cullum, R.D., 1988. *Handbook of Engineering Design*. Butterworths.
- Dowson, D., 2012. Bio-tribology. *Faraday Discuss* 156, 9–30.
- Eriksson, M., Bergman, F., Jacobson, S., 2002. On the nature of tribological contact in automotive brakes. *Wear* 252, 26–36.
- Holmberg, K., Andersson, P., Erdemir, A., 2012. Global energy consumption due to friction in passenger cars. *Tribol. Int.* 47, 221–234.
- Jones, M.H., Scott, D. (Eds.), 1983. *Industrial Tribology, Tribology Series No. 8*, Elsevier.
- Kapoor, A., Tung, S.C., Schwartz, S.E., Priest, M., Dwyer-Joyce, R.S., 2001. Automotive tribology. In: Bhushan, B. (Ed.), *Modern Tribology Handbook*, vol. 2. CRC Press, pp. 1187–1229.
- Kirkham, R.O., Park, G., Groves, W.S., Stobo, J.J., 1983. Plain bearing alloys. *Metals Forum* 6, 149–160.
- Liang, H., Craven, D., 2005. *Tribology in Chemical–Mechanical Planarization*. CRC Press.
- Mufti, R.A., Priest, M., 2009. Effect of engine operating conditions and lubricant rheology on the distribution of losses in an internal combustion engine. *Trans. ASME J. Tribol.* 131, 041101.
- Neale, M.J. (Ed.), 1993. *Bearings: A Tribology Handbook*. Butterworth Heinemann.
- Nisbet, T.S., 1974. *Rolling Bearings, Engineering Design Guide No. 4*. Oxford University Press.
- Peterson, M.B., Winer, W.O. (Eds.), 1980. *Wear Control Handbook*. ASME.
- Pratt, G.C., 1973. Materials for plain bearings. *Int. Metall. Rev.* 18, 62–88.

- Rajpura, A., Kendoff, D., Board, T.N., 2014. The current state of bearing surfaces in total hip replacement. *Bone Joint J.* 96-B, 147–156.
- Schmid, S.R., Wilson, W.R.D., 2001. Tribology in manufacturing. In: Bhushan, B. (Ed.), *Modern Tribology Handbook*, vol. 2. CRC Press, pp. 1385–1411.
- Shaw, M.C., 1996. *Principles of Abrasive Polishing*. Clarendon Press.
- Tung, S.C., McMillan, M.L., 2004. Automotive tribology overview of current advances and challenges for the future. *Tribol. Int.* 37, 517–536.
- Wong, V.W., Tung, S.C., 2016. Overview of automotive engine friction and reduction trends—effects of surface, material and lubricant-additive technologies. *Friction* 4 (1), 1–28.

QUESTIONS FOR CHAPTER 9

(Some of these questions require familiarity with material covered in earlier chapters.)

Question 9.1

- (a) Wheel bearings for passenger cars are designed to last for 300,000 km. Approximately how many revolutions of the bearing does this distance correspond to? Assume that the car tyre has an overall diameter of 0.6 m. Why is your answer not exactly correct?
- (b) A simple equation used to predict the life of rolling bearings is:

$$L = \left(\frac{C}{P} \right)^\alpha$$

where L is the life in millions of revolutions, P is the load applied to the bearing, and C is the load for which the life is one million revolutions. Rolling contact fatigue tests on the bearing steel show that it fails after N cycles at a peak Hertzian contact pressure $p_{\max}$, where the two are related by the empirical equation

$$p_{\max} \propto N^{-0.11}$$

Estimate the value of the exponent α for a radial ball bearing.

- (c) Two ball bearings are manufactured from the same steel with identical geometries, but one is 1.5 times the linear dimensions of the other. If the smaller bearing has a life of 10^7 revolutions at a load of 5 kN, estimate the load that could be applied to the larger bearing for the same life. What load applied to the larger bearing would give a life of 10^8 revolutions?

Question 9.2

- (a) Explain why rolling bearings need to be lubricated.
- (b) Give reasons for the following:
 - (i) Bearings should not be tightly packed with grease.
 - (ii) Lubricating oils must be kept free from particulate contamination.
 - (iii) Hybrid ball bearings, in which silicon nitride balls run between steel races, are used in some high-speed rotating machine tools.

Question 9.3

In hydrodynamic (fluid film) lubrication the sliding surfaces are fully separated by a lubricant film. Why, then, are the materials used for hydrodynamically lubricated plain journal bearings important? Discuss the properties required in these bearing materials. Explain how the requirements for marginally lubricated and dry bearings

differ from these. In your answer, provide examples of typical bearing materials for all three cases.

Question 9.4

The motor of a hand-held electric drill has a steel shaft 6 mm in diameter, which is supported in a plain bearing bush, 6 mm long, at each end and rotates at 3000 rpm (revolutions per minute). The shaft carries a drive gear at one end, which results, under typical running conditions, in a sideways force of 40 N acting between the shaft and the bush. The maximum acceptable wear in the bush is judged to be 75 μm after 300 h running time.

- (a) Use the data provided in Fig. 9.12, which relates to bushes of various materials running against a steel shaft with a wear rate of 25 μm per 100 h, to consider the suitability of various candidate materials for the bushes. Hence recommend a suitable material for the bushes.
- (b) In order to save money, it is proposed to use bushes made from acetal (a thermoplastic). What will happen if these bushes are used in the drill?

Question 9.5

As the oil temperature in a car engine is varied between 25 and 85°C, the relative magnitudes of the frictional power loss in the various components change. The power dissipated in the crankshaft bearings decreases steadily with increasing temperature, and that in the valvetrain increases. The contribution from the piston assembly generally shows little variation with temperature, although with a low-viscosity oil it increases at the highest temperatures. Discuss and explain these observations.

Question 9.6

- (a) Estimate the temperature rise in the front brake disks in a passenger car after it has been rapidly stopped from 120 km h^{-1} . Assume that the car has a mass of 1500 kg, that 70% of the braking effort occurs at the front disks and that all the heat energy passes into the disks. The surface of the disk in contact with the brake pads is an annulus with internal diameter 170 mm and external diameter 290 mm. The thickness of the solid disk is 25 mm and it is made from grey cast iron with a density of 7200 kg m^{-3} and specific heat capacity of 460 $\text{J kg}^{-1} \text{K}^{-1}$. State your assumptions clearly, and explain why your answer provides an over-estimate.
- (b) Compare your answer in part (a) to an estimate of the maximum possible temperature rise in the disks due to the descent of a steep mountain pass through a vertical height of 300 m, and comment on your result.

Question 9.7

Estimate the following, for a pneumatic tyre of 0.6 m outside diameter fitted to a road vehicle travelling at 80 km h^{-1} . Assume that the vehicle mass is 1200 kg, that the tyre is inflated to an internal pressure of 2 bar, and that the roughness of the road surface has a wavelength of 10 mm.

- (a) The frequency with which a point on the rim of the tyre passes through the tyre-road contact patch.
- (b) The frequency with which the tread of the tyre is deformed in passing through the tyre-road contact patch.
- (c) The frequency of the tread deformation associated with skidding over a rough road surface at this speed.

Explain how the tyre materials can be selected to achieve both low rolling resistance and high resistance to skidding.

Question 9.8

- (a) A metal-on-polymer artificial hip joint with a femoral head diameter of 22 mm is tested in a laboratory rig, which simulates the load and movement cycles associated with walking. After 10^6 cycles in the simulator the volume of material lost by wear from the polymer cup is 23.0 mm^3 . A joint of identical design but with a head 32 mm in diameter is subjected to exactly the same conditions in terms of load and angular displacement cycles. Assume that the behaviour follows the Archard wear equation, and hence estimate the volume of wear debris produced after the same number of cycles by the joint with larger head. Why might the actual wear rate measured in the simulator test be significantly different from your estimate?
- (b) Explain why the wear volume alone does not provide enough information to assess the biological impact of wear debris from an artificial joint.
- (c) Explain why, from a tribological perspective, for an obese patient with an artificial hip joint there may be advantages in losing weight.

Appendix A: Indentation hardness—Measurement and meaning

‘Hardness’ is a mechanical property often encountered in tribology, and it is helpful to be familiar with the various ways by which it is measured and quantified. The theoretical basis of indentation-induced plastic flow has been discussed in [Section 2.5.2](#).

Most methods of measuring hardness are based on indentation, and involve pressing an indenter of well-defined shape into the surface under a known load. They can be classified into those in which the hardness is deduced from measurement of the area of the indentation that remains in the surface of the sample after the load has been removed, and those where it is based on the indentation depth.

A.1 METHODS BASED ON INDENTATION AREA

The Vickers hardness test, first described in 1922, is still one of the most widely used methods. The indenter is a square-based diamond pyramid as shown in [Fig. A.1\(a\)](#), with an angle of 136° between the opposite faces. A load W is applied to the indenter and then removed, and the resulting indentation in the surface is measured by optical microscopy. The surface trace of the indentation is square, as shown in [Fig. A.1\(b\)](#). If the mean length of the two diagonals of the square is d , then the Vickers hardness HV is defined by

$$HV = 1.854W/d^2 \quad (\text{A.1})$$

The factor of 1.854 arises because Vickers hardness is defined as the indentation load W divided by the pyramidal surface area of the indentation (rather than the projected area, which would give a factor of 2.0).

It is important to ensure that the correct units are used in equation [A.1](#). Vickers hardness is conventionally quoted in units of kg force per mm^2 , so that the load W must be in kg force (1 kgf = 9.81 N) and the diagonal length d in mm. The ISO standard method is to state the hardness in kgf mm^{-2} but without the symbols for the units, followed by the letters HV, and then a number which represents the load W in kgf. For example 125 HV 5 denotes a Vickers hardness of 125 kgf mm^{-2} measured with an applied load of 5 kgf (=49.1 N). A hardness of 100 HV corresponds to an indentation pressure (load divided by pyramidal area) of 981 MN m^{-2} or 981 MPa. Typical values of Vickers hardness are: aluminium alloys 50–150 HV; steels 100–900 HV; engineering ceramics 1000–3000 HV.

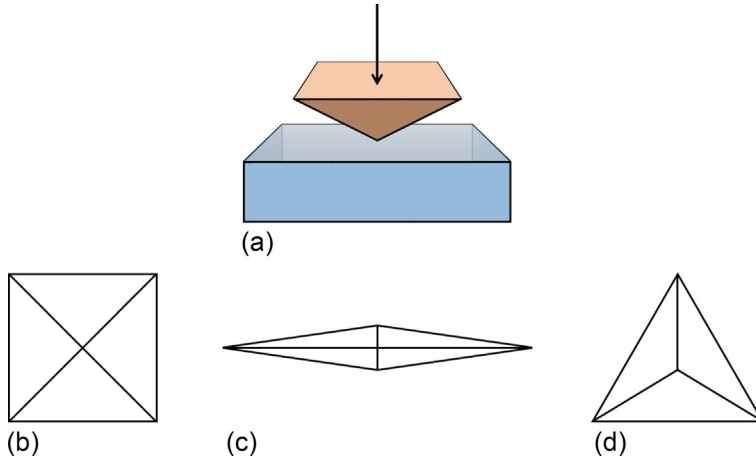


Fig. A.1 (a) The Vickers hardness test in which a square-based diamond pyramid is pressed into the surface under a known load. Plan views of indentations formed by (b) a Vickers indenter; (c) a Knoop indenter; (d) a Berkovich indenter.

Vickers measurements can be made with a wide range of loads, from a few grams force to tens of kg force. Conventionally, with $W \geq 5$ kgf the measurement is simply termed ‘Vickers hardness’, while testing with $0.2 \text{ kgf} < W < 5 \text{ kgf}$ is termed ‘low load Vickers hardness’ and $10 \text{ gf} < W \leq 0.2 \text{ kgf}$ is the regime of ‘microindentation hardness’ or ‘microhardness’.

Indenters with other shapes can be used to measure hardness in a similar manner to the Vickers test. The most common are the Knoop test which uses a rhombic-based diamond pyramid and produces an indentation that in plan view is lozenge-shaped with the long diagonal 7.114 times the length of the shorter one, as shown in Fig. A.1(c), and the Brinell test which uses a spherical indenter and produces a circular indentation.

In contrast to the Vickers method that is based on surface area, the Knoop hardness HK is defined as the mean pressure over the *projected* area of the indentation, and is given by:

$$HK = 14.229W/d^2 \quad (\text{A.2})$$

where W is the applied load and d is the length of the long diagonal of the indentation. The standard method of stating the Knoop hardness is analogous to that for Vickers, by giving the value in kgf mm^{-2} without the units, followed by the letters HK and then the load in kgf, e.g., 550 HK 0.2.

For very hard materials the Knoop method has an advantage over the Vickers test for which the indentations would be small: with one diagonal much longer than the other, at least one dimension of the indentation can be measured more accurately.

Knoop indentation can be used to explore material anisotropy, by performing a series of tests over which the orientation of the indentation in the surface is varied. It is also better suited to hardness measurement of thin sheets and coatings, since for the same load the indentation produced by a Knoop indenter is shallower than that produced by the Vickers method, and the region of plastic deformation does not extend so deeply beneath the sample surface. As a general rule, to avoid errors caused by the penetration of the plastic zone through the sample the depth of the indentation should be less than one tenth of the thickness of sheet material. The depth of a Vickers indentation is one seventh of the length of the diagonal, whereas for the Knoop geometry the depth is one thirtieth of the long diagonal.

The standard Brinell indenter is a cemented tungsten carbide (WC-Co) sphere, 10 mm in diameter, although smaller balls can also be used. Large indenters require high loads, and the Brinell test uses loads up to 3000 kgf (29.4 kN). The Brinell hardness HB is defined as the load divided by the curved surface area of the indentation, with the assumption that it has the same geometry as the ball:

$$HB = \frac{2W}{\pi D (D - \sqrt{D^2 - d^2})} \quad (\text{A.3})$$

where D is the diameter of the ball and d is the diameter of the indentation measured in plan view. The Brinell hardness is stated as the value in kgf mm^{-2} without units, followed by the symbol HBW (signifying a tungsten carbide indenter), the ball diameter in mm and the load in kgf. For example 350 HBW 10/3000 denotes a hardness of 350 kgf mm^{-2} , measured with a 10 mm diameter WC-Co ball and a load of 3000 kgf.

A.2 METHODS BASED ON INDENTATION DEPTH

The area-based methods all require the size of the residual indentation to be measured after the indenter has been removed. In depth-based methods, in contrast, the hardness is derived from measurements made while the indenter is still in contact with the surface. The Rockwell method (introduced in 1914) involves high loads and unlike the area-based methods described above it does not require a skilled operator to measure the indentation by optical microscopy. Under much smaller loads, in the regime of *nano-indentation* the indentation becomes much too small for accurate optical measurement, and depth-sensing methods are widely used.

The principles of the Rockwell method are illustrated in Fig. A.2. The indenter is pressed into the surface with an initial (minor) load of W_0 , and then a further load of W_1 is applied. The total load ($W_0 + W_1$) at this point is described as the ‘major’ load. The load is then reduced to W_0 , and the distance x that the indenter has sunk into the surface, relative to its position after the initial application of the minor load, is used to calculate the hardness. In the Rockwell method a depth displacement x of zero is taken to correspond to a hardness of 100 units, while a displacement of $200 \mu\text{m}$ corresponds to a hardness of 0. There is not a single Rockwell scale, but several standard scales,

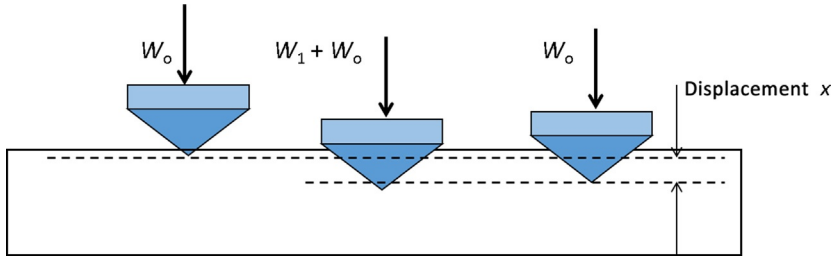


Fig. A.2 Illustration of the Rockwell hardness testing method.

identified by letters. The scales differ in the types of indenters used, and in the loads applied. In order to achieve a hardness value within the range 0–100 for a given material (i.e., so that $x < 200\ \mu\text{m}$) an appropriate scale must be chosen. The Rockwell C scale, for example, is commonly used for hardened steels. The indenter for this scale is a diamond cone (120° included angle) with a spherically-rounded tip with a radius of 0.2 mm (called a ‘Brale’ indenter): the minor load is 10 kgf and the major load 150 kgf. The resulting hardness is written as the hardness number followed by a symbol to specify the particular Rockwell scale: e.g., 55 HRC denotes a hardness of 55 on the Rockwell C scale. In quoting Rockwell hardness values, it is essential to specify the scale. For example a hard steel, measured with the Rockwell C method, might have a hardness of 62 HRC, while a soft copper alloy, measured with the Rockwell F scale could have a hardness of 62 HRF. The Vickers hardness values of these two materials would be quite different: about 750 HV and 60 HV, respectively.

In depth-sensing nano-indentation, also known as instrumented indentation, the depth of penetration of the indenter into the sample surface is continually recorded as the load on the indenter is progressively increased, as illustrated by the loading curve in Fig. A.3. The indenter is usually a triangular-based diamond pyramid known as a Berkovich indenter, with a geometry chosen so that it has the same ratio of projected area to indentation depth as the Vickers indenter. Figure A.1(d) illustrates an indentation made by a Berkovich indenter. The use of a triangular pyramid has a significant advantage over a square-based pyramid for very small indentations, in that the three faces of the pyramid formed by grinding and polishing must meet at a point. This cannot be ensured with four faces, and it is therefore much more difficult to achieve such a sharp indenter tip (rather than an edge) in the Vickers geometry.

In principle, if the relationship between the indentation area A (i.e., the area over which the load is supported) and the depth x is known, then it is straightforward to deduce the hardness H at any point on the loading curve shown in Fig. A.3, from $H = W/A$. For an ideal system in which the test material and indenter were all rigid, the load would drop to zero as soon as the indenter was raised any distance from its maximum depth. In practice the system will exhibit some elastic deformation and unloading curve will not be vertical; its initial slope can then be used to determine the reduced modulus E^* of the sample and the indenter material. The initial slope

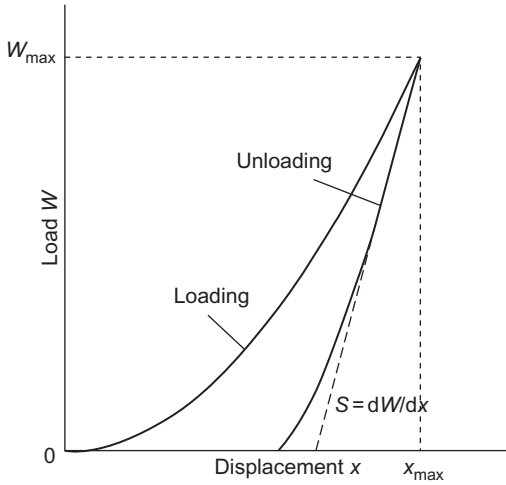


Fig. A.3 Example of a loading–unloading curve produced in a depth-sensing indentation test.

of the unloading curve, the contact stiffness $S = dW/dx$, is related to E^* by the equation

$$S = 1.05 \frac{2}{\sqrt{\pi}} E^* A_{\max} \quad (\text{A.4})$$

where $A_{\max}$ is the area of the indentation corresponding to the maximum depth $x_{\max}$. The reduced modulus E^* is related to the Young's modulus E and Poisson's ratio ν of the indenter and sample, E_1, ν_1, E_2, ν_2 , respectively, by:

$$\frac{1}{E^*} = \frac{1 - \nu_1^2}{E_1} + \frac{1 - \nu_2^2}{E_2} \quad (\text{A.5})$$

In practice, extracting accurate values for hardness and elastic modulus from depth-sensing indentation data requires a well-designed and operated testing machine and demands stringent attention to many factors which become more challenging as the load, displacement and indentation size becomes smaller. These factors include vibration isolation, calibration of the load and displacement measurements, correction for thermal expansion and contraction in the test apparatus during the test (thermal drift), correction for compliance of the test apparatus, accurate characterization of the geometry of the particular indenter (the depth-area relationship, sometimes called the *shape function*), and accurate identification of the precise point at which the indenter first touches the surface. Methods have been developed by which a small cyclic displacement is imposed on the main indenter displacement, in order to provide a continuous measurement of the contact stiffness as a function of depth during the indentation process. 'Instrumented indentation' can be used to probe mechanical properties in very small volumes of materials, with indentation sizes from a few μm to sub- μm in width.

For materials such as elastomers which do not experience plastic deformation on indentation, instrumented indentation is also used to measure ‘hardness’, but in this case the term now refers to the response of the material to elastic indentation. There are standard scales (for example Shore and IRHD = International Rubber Hardness Degrees) which define the indenter geometry, loading method and measurement protocol. The Shore A scale is quite widely used for rubbers, and employs a hardened steel indenter in the form of a truncated cone, with a tip diameter of 0.79 mm and an included angle of 35°; the load is applied by a calibrated spring and varies considerably with the depth of the indentation. The IRHD methods in contrast use a fixed (dead weight) load. Hardness values measured by the various Shore and IRHD scales can be related empirically to values of elastic modulus.

A.3 INTERPRETATION OF HARDNESS AND COMPARISON OF HARDNESS SCALES

As discussed in [Section 2.5.2](#), when a rigid indenter is pressed into the surface of a material that deforms plastically (after negligible elastic strain), the hardness H defined as the average pressure resisting indentation is about three times the uniaxial flow stress Y of the material. This applies to most metallic materials and for any shape of indenter (e.g., a flat-ended punch, a cone, a pyramid or a sphere). Expressed in general terms, we find empirically that

$$H \approx cY \quad (\text{A.6})$$

where c is the ‘constraint factor’. The factor c of ‘about three’ varies slightly with indenter shape, but a ratio of 3 is useful to remember and provides a simple way of estimating the flow stress of a metal from a hardness value expressed as an indentation pressure. For materials with a higher yield strain (such as some inorganic glasses and glassy polymers) the value of c is smaller, perhaps as low as 1.5.

The Vickers, Knoop, and Berkovich indenters are pyramidal and thus have the important property of geometrical similarity: indentations made by the same indenter but with different depths have exactly the same shape. The strain distribution around an indentation thus remains the same whatever its size, and the extent of strain hardening induced by the process of hardness testing (in a material which strain hardens) would be expected to be independent of the load used and hence the size of the indentation. A Vickers indenter induces a ‘representative strain’ of about 8%, meaning that a sample of a work-hardening material deformed by this uniaxial strain would have a flow stress equal to about one third of the measured hardness. Indentations of different depth made by a sphere, in contrast, are not self-similar, and the strain induced in the material is greater for larger indentations; the representative strain associated with a Brinell indentation is about $20 d/D\%$.

The strain fields (elastic and plastic) induced in a material by indenters of different shapes will in general be different. So for a general material that exhibits both elastic

and plastic deformation, the hardness defined as the indenter load divided by the area (projected or surface) of the residual indentation will be a unique property of the particular method used to generate the indentation. That said, all the methods based on measurement of a residual indentation (Vickers, Knoop, or Brinell) give quite similar hardness values, and for restricted classes of materials (e.g., non-austenitic steels, austenitic stainless steels, wrought aluminium alloys or copper alloys) they can be related consistently to values derived from Rockwell tests. Tables are available for this purpose (e.g., ASTM standard E140), and [Table A.1](#) shows an example of some data. Hardness values derived from well-calibrated depth-sensing measurement with a Berkovich indenter are also similar to those derived by the Vickers or Knoop micro-indentation methods.

Table A.1 Comparison of hardness values for non-austenitic steels (data from ASTM standard E140: Standard hardness conversion table for metals)

Rockwell C HRC	Vickers HV	Brinell HBW	Knoop HK
60	697	654	732
50	513	481	542
40	392	371	402
30	302	286	311
20	238	226	251

Appendix B: Fundamentals of corrosion and tribocorrosion

B.1 OXIDE GROWTH IN AIR

The reaction of a metal with oxygen to form an oxide is one example of *corrosion*. Almost all metals (with the exception of gold) are thermodynamically unstable with respect to oxidation in air at room temperature, and so naturally become covered with films of oxide. The overall reaction can be written (for the example of a metal M with valency 2 forming the oxide MO) as:



The change in free energy on oxidation (ΔG°) for several metals is listed in [Table B.1](#). A negative value means that the metal is thermodynamically unstable in the presence of oxygen: i.e., the oxide has a lower free energy than the metal plus oxygen from which it forms.

However, the magnitude of ΔG° tells us nothing about the rate at which oxidation will occur. In air at room temperature, oxide films typically grow rapidly to a thickness of the order of nanometres or a few tens of nanometres, but then cease to grow further. Only at high temperatures do they grow progressively to form thicker films (of the order of micrometres or more in thickness) or even thicker *scales*.

[Figure B.1](#) illustrates an oxide film on a metal surface in air and shows two different ways by which it can grow further. At the outer surface, oxygen molecules dissociate to form atoms that become ionized with the addition of electrons:



and at the metal-oxide interface, metal ions form from metal atoms, releasing electrons:



Reaction [B.2](#), which involves the consumption of electrons, is called *reduction*, and [B.3](#) which produces electrons is called *oxidation* (although oxygen need not be involved).

Table B.1 Free energy of oxidation at 273 K, per mol of oxygen

Material	Oxide	ΔG° (kJ mol ⁻¹ O ₂)
Magnesium	MgO	-1162
Aluminium	Al ₂ O ₃	-1045
Zirconium	ZrO ₂	-1028
Titanium	TiO ₂	-848
Silicon	SiO ₂	-836
Niobium	Nb ₂ O ₅	-757
Chromium	Cr ₂ O ₃	-701
Zinc	ZnO	-636
Molybdenum	MoO ₂	-534
Tungsten	WO ₃	-510
Iron	Fe ₃ O ₄	-508
Tin	SnO	-500
Nickel	NiO	-439
Cobalt	CoO	-422
Lead	Pb ₃ O ₄	-309
Copper	CuO	-254
Silver	Ag ₂ O	-5
Gold	Au ₂ O ₃	+80

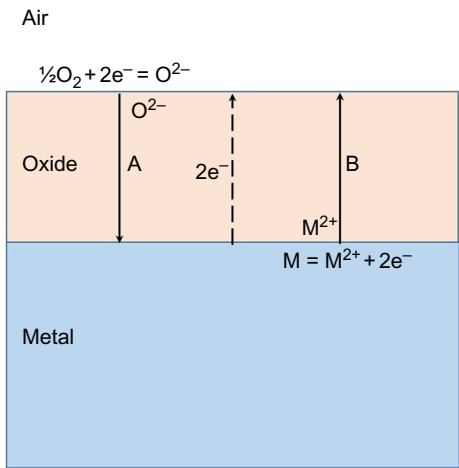


Fig. B.1 Key processes in the growth of an oxide film on a metal in air

The oxide MO is formed when the oxygen ions (O^{2-}) and metal ions (M^{2+}) combine. For the film to continue to grow, either oxygen ions must travel inwards from the outer surface to combine with metal ions (path A), or metal ions must travel outwards from the metal/oxide interface to combine with oxygen ions (path B). In the first case, new oxide forms at the metal-oxide interface, while in the second it forms at the outer surface. For either mechanism to operate, electrons (e^-) must also move

from the metal to the outer surface to maintain electrical charge balance. The rate of growth of the oxide film will therefore depend on the thickness of the film, on its electrical conductivity (involving the movement of electrons), on the ease of formation of the metal and oxygen ions, and on the rates at which they move through the film.

The initial growth of the very thin films (a few nm thick) which form on metals at room temperature is driven by very high electric fields which drag ions through the film. The high field arises from the potential difference between the inner and outer surface, but as the film becomes thicker the field falls and growth by this mechanism, which is exponentially proportional to the field, slows abruptly.

At higher temperatures, growth occurs by the diffusion of ions through the film (which is enhanced, for example, by the presence of vacancies and grain boundaries). In a simple model for the growth of the oxide film by this mechanism, we can assume that the diffusion rate is proportional to the concentration gradient of the ions through the film, which will be inversely proportional to the thickness of the film. So for oxygen ion diffusion in a film of thickness x , if the oxygen ion concentration at the metal-oxide interface is assumed to be zero,

$$\frac{dx}{dt} \propto D \frac{C_o}{x} \quad (\text{B.4})$$

where D is the diffusion coefficient for oxygen ions in the film and C_o is their concentration at the outer surface. This equation can be integrated to give the thickness of the film as a function of time:

$$x^2 = k_p t \quad (\text{B.5})$$

This *parabolic* growth law is typical for the oxidation of metals at high temperatures, and applies whether growth happens by diffusion of metal or oxygen ions. We can see immediately that temperature will have a strong effect on the oxide growth rate, since the *parabolic rate constant* k_p is proportional to D which typically shows an Arrhenius dependence on temperature:

$$D = D_o \exp(-Q/RT) \quad (\text{B.6})$$

where Q is the activation energy for diffusion of the relevant ions.

If the oxide is sufficiently insulating, then electron transport rather than ion diffusion limits the oxide growth rate. For example, the electrical resistivity of Al_2O_3 at 1000°C is about 10^9 times that of FeO , which explains the very low oxidation rate of aluminium even at high temperatures.

B.2 AQUEOUS CORROSION

If the metal is immersed in water containing some dissolved oxygen, oxidation of the metal (i.e., the formation of ions by reaction B.3) can occur much more rapidly than in the case of oxide film growth in air. Figure B.2 illustrates the key processes for the

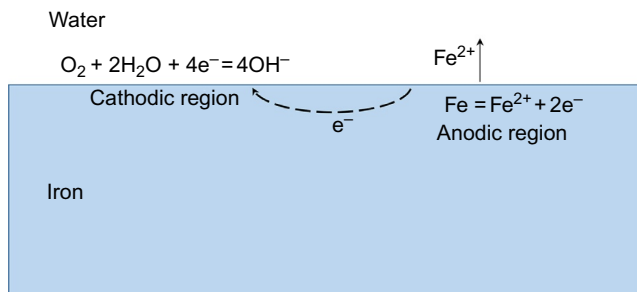


Fig. B.2 Key processes in the corrosion of iron in oxygenated water

progressive oxidation (corrosion) of iron in oxygenated water. Iron atoms lose electrons at one region of the metal surface (termed the *anodic* region) to form ions (Fe^{2+}) which dissolve in the water. The electrons are conducted through the metal to a nearby *cathodic* region where they are consumed in the reduction of oxygen in the presence of water, forming hydroxyl (OH^-) ions. The hydroxyl ions react with the iron ions in the water to form $\text{Fe}(\text{OH})_2$, which is poorly soluble and forms a solid precipitate. Unlike oxidation in air, in this case the corrosion products do not form a continuous barrier film on the metal surface, which therefore continues to corrode at a steady rate.

Because the anodic and cathodic reactions involve the transfer of electrons, it is convenient to describe the associated energy changes in terms of the electrical potential through which the electrons are moved. This approach is used to generate the list of *standard electrode potentials* shown in Table B.2. The values in the table are shown relative to the potential for the hydrogen oxidation reaction, and all apply to a state of equilibrium under standard thermodynamic conditions in which the reactants have unit activity (i.e., gases are present at a pressure of 1 atm and ions at 1 M concentration). At equilibrium, the rates of the forward and backward reactions (releasing and consuming electrons) are equal.

The effect of changing the applied potential is to accelerate one of these reactions at the expense of the other, as depicted graphically in the *Evans diagram* shown in Fig. B.3. The potential is plotted on the vertical axis, and the current (which represents the overall rate of the reaction) is plotted on a logarithmic scale on the horizontal axis. In this graphical presentation, all currents, whether the process is oxidation or reduction, are assumed to be positive.

In this example, we have assumed the same anodic and cathodic reactions as in Fig. B.2 under standard conditions, purely for illustration. When immersed in the electrolyte the iron surface adopts a potential at which the anodic reaction (oxidation of iron) and cathodic reaction (reduction of oxygen) generate and consume electrons at the same rates; this *corrosion potential* E_{corr} lies between the *equilibrium* (open-circuit) potentials E_0 for the two reactions. The rate of the corrosion reaction can be described by the corrosion current i_{corr} , since the rate at which the metal is converted to ions is directly proportional to the rate of charge flow.

Table B.2 Standard electrode potentials at 300 K (for molar concentrations of ions and 1 atm partial pressure of gases). The two values for the oxygen reduction (cathodic) reaction marked * differ because they relate to conditions of molar concentrations of hydroxyl and hydrogen ions: i.e., pH 14 and pH 0, respectively

Reaction	SEP (V relative to standard hydrogen electrode)
$\text{Mg} = \text{Mg}^{2+} + 2\text{e}^-$	-2.36
$\text{Al} = \text{Al}^{3+} + 3\text{e}^-$	-1.66
$\text{Zn} = \text{Zn}^{2+} + 2\text{e}^-$	-0.76
$\text{Cr} = \text{Cr}^{3+} + 3\text{e}^-$	-0.74
$\text{Fe} = \text{Fe}^{2+} + 2\text{e}^-$	-0.44
$\text{Ni} = \text{Ni}^{2+} + 2\text{e}^-$	-0.25
$\text{Sn} = \text{Sn}^{2+} + 2\text{e}^-$	-0.14
$\text{Pb} = \text{Pb}^{2+} + 2\text{e}^-$	-0.13
$\text{H}_2 = 2\text{H}^+ + 2\text{e}^-$	0
$\text{Cu} = \text{Cu}^{2+} + 2\text{e}^-$	+0.34
$\text{O}_2 + 2\text{H}_2\text{O} + 4\text{e}^- = 4(\text{OH}^-)^*$	+0.40
$\text{Ag} = \text{Ag}^+ + \text{e}^-$	+0.80
$\text{O}_2 + 4\text{H}^+ + 4\text{e}^- = 2\text{H}_2\text{O}^*$	+1.23
$\text{Au} = \text{Au}^{3+} + 3\text{e}^-$	+1.42

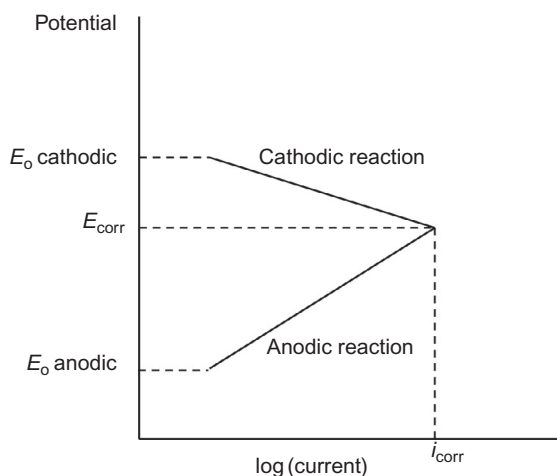
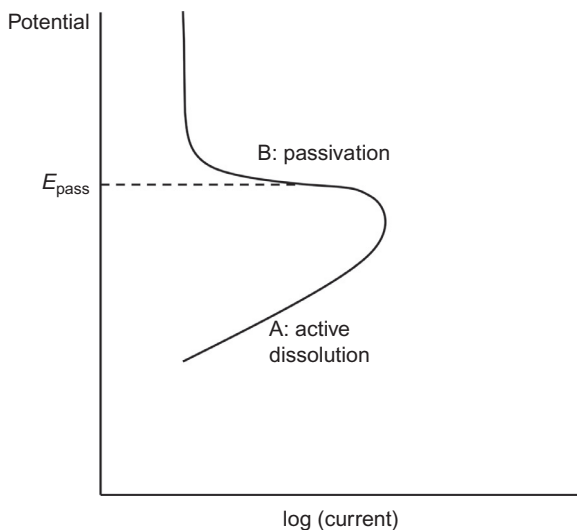


Fig. B.3 Current-potential diagram illustrating the fundamental processes operating in aqueous corrosion

The slopes of the lines in Fig. B.3, upwards for the anodic reaction and downwards for the cathodic, are associated with various *polarization* processes; although they are shown as straight lines, in practice they may be curved. The polarization curve, which describes the change in current with the change in potential away from E_o (the *overpotential*), is determined by many factors, including the charge carried

Fig. B.4 Current-potential diagram for an anodic reaction illustrating the phenomenon of passivation



by the ions, their mobility and concentration in the solution, temperature, the resistivity of the electrolyte, the nature of any surface films formed etc. Polarization curves are usually measured in experiments with a *potentiostat*, in which the potential of the metal is held constant or changed in a controlled way while the current is measured.

A more complex example of an anodic polarization curve is shown schematically in Fig. B.4. In regime A the oxidation rate increases steadily as the potential is raised, a state known as *active dissolution*, while in regime B, above a certain potential E_{pass} the current falls again to a low value. This behaviour is known as *passivation*, and is typically caused by the formation of a thin (often a few nm thick) adherent oxide film which slows down the dissolution process. Stainless steels are important examples of metals which show passivity, in their case because of the formation of passive films of chromia (Cr_2O_3).

B.3 TRIBOCORROSION

Tribocorrosion is a general term for degradation processes that involve significant contributions from both mechanical and chemical mechanisms: subdivisions include *wear-corrosion*, *corrosion-wear*, *erosion-corrosion*, *corrosion-erosion* and *erosion-oxidation*. Although the presence of oxide films often plays an important role in the tribological behaviour of metals in air, the term *tribocorrosion* is usually reserved for cases where significant corrosion would occur in the absence of any tribological effects.

Tribocorrosion thus involves the presence of mechanical stimuli as well as an environment that favours corrosion, and material can be removed from a surface by several different mechanisms as discussed in detail in [Chapters 5 and 6](#). Here we shall define some key concepts.

The total rate of material loss T under conditions of tribocorrosion is in general not simply the sum of the rate of wear that would result from the same mechanical action in the absence of corrosion, W_o , and the rate of corrosion that would occur in the absence of mechanical action, C_o . There will usually also be some additional *synergistic* contribution, S (which may be either positive, or negative—termed *antagonistic*) so that we can write:

$$T = W_o + C_o + S \quad (\text{B.7})$$

The synergistic contribution S can be split into the change in corrosion rate caused by mechanical (wear) processes, ΔC_w , and the change in wear rate caused by corrosion, ΔW_c , so that:

$$T = W_o + C_o + \Delta C_w + \Delta W_c \quad (\text{B.8})$$

For a passive metal (i.e., one that forms a protective film of corrosion product on its surface which inhibits further corrosion as discussed in [Section B.2](#)), the most common contribution to synergy is through the continuing removal of the protective film by mechanical wear processes. Once removed, partially or completely, the film then re-grows, and is again removed. This repeated cycle, in which the average thickness of the film is less than its equilibrium thickness under corrosion conditions alone, leads to an enhancement of the corrosion rate by the mechanical action of wear, and therefore to a positive value for ΔC_w . We discuss other mechanisms in [Chapters 5 and 6](#).

In some cases it is possible to design experiments to measure the various contributions to materials loss separately and compare their importance. For example, in the case of tribocorrosion in an aqueous medium, we can use electrochemical methods (see [Section B.2](#)) to determine the corrosion current i_{corr} from which, by Faraday's law, we can calculate the total contribution from corrosion, $C (=C_o + \Delta C_w)$. By repeating the experiment under (identical) corrosion conditions but in the absence of mechanical action, we can similarly determine C_o . By applying a sufficiently large negative potential to the sample (typically, 1 V below the open-circuit potential), or by substituting a carrier liquid in which significant corrosion does not occur, corrosion can be suppressed and the measured material loss rate will then be W_o . ΔW_c is then derived from equation [B.8](#), the values of W_o , C_o , ΔC_w and the measured value of T . Although these measurements appear straightforward in principle, in practice they require careful planning and execution.

The relative importance of mechanical processes (wear) and corrosion can be described by the ratio C/W where $W = W_o + \Delta W_c$, $C = C_o + \Delta C_w$ and

$$\frac{C}{W} = \frac{C_o + \Delta C_w}{W_o + \Delta W_c} \quad (\text{B.9})$$

Regimes of tribocorrosion are often defined as follows, depending on the value of C/W :

- wear-dominated: $C/W < 0.1$
- wear-corrosion: $0.1 \leq C/W < 1$
- corrosion-wear: $1 \leq C/W < 10$
- corrosion-dominated: $C/W \geq 10$

FURTHER READING

- Ashby, M.F., Jones, D.R.H., 1996. Engineering Materials I. Butterworth Heinemann.
- Landolt, D., Mischler, S. (Eds.), 2011. Tribocorrosion of Passive Metals and Coatings. Woodhead Publishing.
- McCafferty, E., 2010. Introduction to Corrosion Science. Springer.
- Newey, C., Weaver, G. (Eds.), 1990. Materials Principles and Practice. Butterworth Scientific.
- Scully, J.C., 1990. The Fundamentals of Corrosion, third ed. Pergamon Press.
- Wood, R.J.K., 2007. Tribo-corrosion of coatings: a review. J. Phys. D. Appl. Phys. 40, 5502–5521.

Appendix C: Definition of wear rate

In considering the development of wear, a *measure of wear* is often plotted against what may be called the *exposure to wear*.

The *measure of wear* (W) is typically the *volume* of material removed, or alternatively the *mass* of material removed. In the latter case we must take care in comparing the wear of materials of different densities, since it is usually the change in dimensions of a component rather than its mass loss that defines its loss of functionality. Other measures are possible, such as the *depth* of material removed, or the *area* over which a coating has been removed from a substrate.

The *exposure to wear* (E) can be described in many ways. Sometimes, the exposure to wear is a simple measure of the duration of the exposure of the material to the conditions which are resulting in wear, but more complex measures which describe not only the exposure but also the severity of that exposure may be employed. Typical measures of exposure to wear are:

- sliding distance (common for sliding wear and abrasive wear testing);
- mass of abrasive particles impinged (common for solid particle erosion testing);
- time (common for cavitation erosion testing);
- frictional energy dissipated (common for fretting wear testing).

From these two, a wear rate can be defined, which is dW/dE .

The development of the amount of wear with the exposure to wear can take many forms, with typical examples schematically illustrated in [Fig. C.1](#). In all the cases shown, there is a period where the wear rate, dW/dE is constant (termed the steady state wear rate), but this sometimes only develops after an initial transient period in which the wear rate is very different. [Figure C.1\(a\)](#) illustrates the case where there is no initial transient. [Figure C.1\(b\)](#) shows a period of rapid wear preceding the development of a lower steady-state rate of wear; this type of behaviour is commonly observed across a wide range of test types, and is associated with the prepared surface of a test specimen having different characteristics to the bulk of the test specimen. For example a ground surface of a ceramic specimen may contain cracks and be more rapidly removed than the underlying bulk material. [Figure C.1\(c\)](#) illustrates a case where there is a period of negligible wear preceding the development of a higher steady-state rate of wear; this type of behaviour is observed if damage must accumulate in the material before any of it is actually removed. Finally, [Fig. C.1\(d\)](#) illustrates a period of ‘negative wear’ preceding the development of a positive steady state rate of wear;

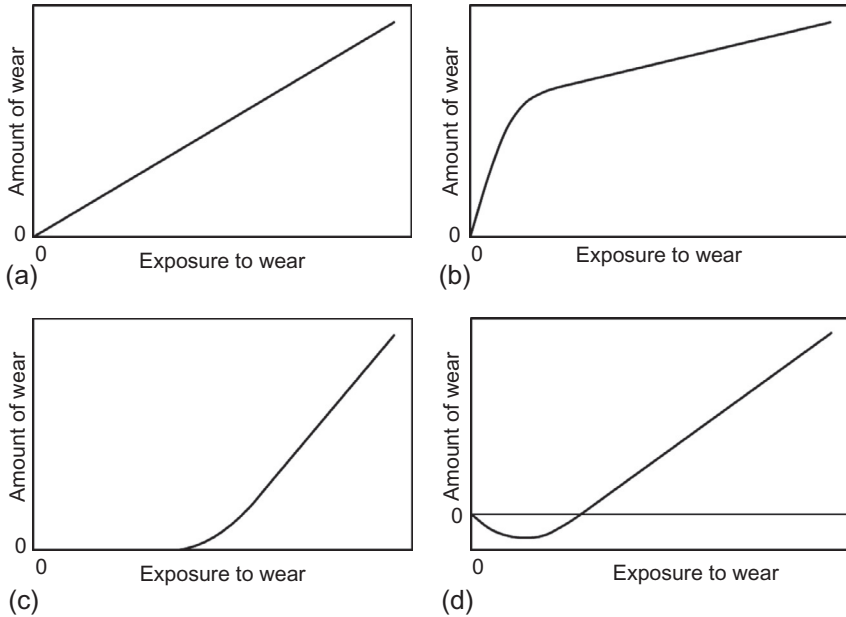


Fig. C.1 Schematic plots showing the development of the amount of wear with the exposure to wear: (a) constant wear rate; (b) constant wear rate following an initial period of rapid wear; (c) constant wear rate following an initial period of damage incubation; (d) constant wear rate following a period during which the wear process itself adds material to the surface

this type of behaviour can occur in the solid particle erosion of soft metals, where the first stage is characterized by the embedment of erodent particles into the surface, before a steady state rate of material removal is subsequently developed.

Materials are often compared against each other in terms of their wear rates under a certain set of test conditions. The example shown in Fig. C.2 indicates that in cases where dW/dE varies with E , then some care needs to be exercised. If W is measured at various value of E , then the amount of wear as a function of exposure in the steady state regime can be described as follows:

$$W = E \left(\frac{dW}{dE} \right)_{\text{true}} + W_{\text{transient}}$$

A wear rate defined simply as (W/E) either at one measured point (or indeed for many points), is a misleading value. In the example shown in Fig. C.2, it can be seen that $(dW/dE)_{\text{false}}$ is larger than $(dW/dE)_{\text{true}}$, but it also can be seen that as the exposure is increased, the ‘false’ value of wear rate will decrease and tend towards the true steady-state wear rate. After the initial transient there is no change in the mechanism or rate of wear, and the gradual decrease of $(dW/dE)_{\text{false}}$ with increasing exposure should not be

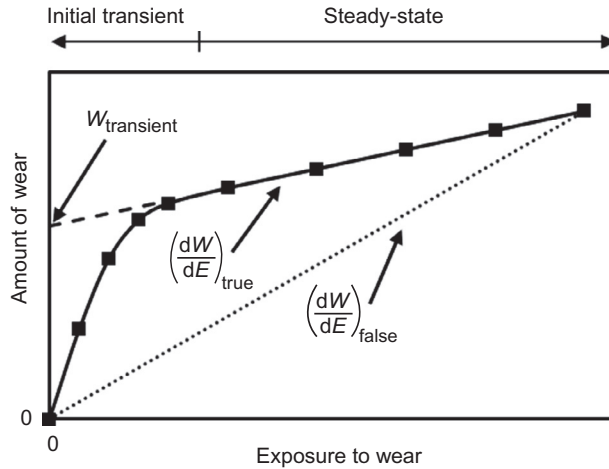


Fig. C.2 Schematic diagram illustrating the need to take multiple measurements of the amount of wear if a true wear rate is to be identified, unless the amount of wear is known to increase linearly with exposure

interpreted as suggesting there are such changes. To avoid these errors of interpretation, it is essential to measure the amount of wear after a number of different exposures, and to plot the data in the form suggested in Fig. C.1 (i.e., the amount of wear versus the exposure). Only in this way can the evolution of wear be accurately observed, and a steady-state wear rate determined with confidence.

FURTHER READING

Blau, P.J., 2015. How common is the steady-state? The implications of wear transitions for materials selection and design. *Wear* 332–333, 1120–1128.

Author Index

Note: Page numbers followed by *f* indicate figures, and *t* indicate tables.

A

Adams, M.J., 231
Adewoye, O.O., 62*f*
Adler, W.F., 214*f*, 231
Allsopp, D.N., 203*f*
Alpas, A.T., 122*f*
Andersson, P., 323*f*, 348
Archard, J.F., 69*f*, 114, 115*f*, 118*f*, 131*f*
Arnell, R.D., 103
Arnold, J.C., 225*f*
Ashby, M.F., 132*f*, 161, 284*f*, 321*f*, 368
Atkins, A.G., 156*f*

B

Baker, I., 74
Ball, A., 224*f*
Barraclough, K.C., 308*f*
Barwell, F.T., 348
Batchelor, A.W., 103
Begelinger, A., 142*f*
Bell, J.C., 85*t*
Bell, T., 85*t*
Bellon, P., 121*f*
Berasetegui, E.G., 32
Bergman, F., 330*t*, 331*f*, 348
Bhansali, K.J., 161
Bhushan, B., 4, 103, 348
Biggers, J.V., 198*f*
Blau, P.J., 74, 371
Board, T.N., 349
Boelli, G., 263*f*
Booser, E.R., 103
Bowden, F.P., 39*f*, 41, 45*t*, 53*f*, 56*f*, 74, 97*f*, 102*f*
Bradt, R.C., 198*f*
Briscoe, B.J., 67*f*, 74, 154*f*, 161, 231
Brockett, C., 348
Brunton, J.H., 132*f*
Bryggman, U., 137*f*
Buckley, D.H., 51*f*, 57*f*, 65*f*, 74, 161

Budinski, K.J., 277, 300
Bull, S.J., 32, 277
Burstein, G.T., 227*f*

C

Cai, W., 121*f*
Cameron, A., 103
Carrera-Espinoza, R., 255*f*
Challen, J.M., 180*f*
Chan, D., 330*t*, 348
Child, H.C., 252*t*, 277
Childs, T.H.C., 30*t*, 54*f*, 334*f*, 348
Clauer, A.H., 223*f*
Clauss, F.J., 74
Clyne, T.W., 258*f*, 267*f*
Conrad, H., 220*f*
Cousens, A.K., 211*f*
Cranmer, D.C., 63*f*
Craven, D., 348
Cullum, R.D., 348
Curran, J.A., 258*f*
Czichos, H., 141*f*, 300

D

Davies, P.B., 103
Davim, J.P., 300
de Gee, A.W.J., 142*f*
Deakin, M.J., 273*f*
Dearnley, P.A., 277
Donnet, C., 278, 61*f*, 74
Donovan, D.M., 278
Dowson, D., 3, 141*f*, 152*f*, 348
Dwyer-Joyce, R.S., 348

E

Eldis, G.T., 195*f*
Erdemir, A., 61*f*, 74, 278, 323*f*, 348
Eriksson, M., 330*t*, 331*f*, 348
Evans, A.G., 185–186*f*
Evans, D.C., 153*f*, 158*f*, 160*f*, 190*f*

F

Finnie, I., 173*f*, 177*f*, 212*f*, 216*f*, 232
 Fiore, N., 195*f*
 Fischer, T.E., 146*f*, 161
 Fisher, J., 348
 Florek, J.J., 55*f*
 Fox, M.F., 103
 Friedrich, K., 67*f*, 161, 231
 Fulcher, J.K., 195*f*

G

Gabel, M.K., 278
 Gay, J.G., 250*f*
 Gee, M.G., 277*f*
 Giovinnazzo, R.J., 33
 Glaeser, W.A., 123*f*, 198*f*, 204*f*
 Greenwood, J.A., 26–29, 27–28*f*, 32
 Groves, W.S., 348
 Grube, W.L., 250*f*
 Gundlach, R.B., 204*f*

H

Habig, K.-H., 141*f*
 Hainsworth, S.V., 278
 Halling, J., 103
 Hamilton, G.M., 32
 Hamrock, B.J., 94, 103
 Hannink, R.H.J., 62*f*
 Hines, J.E., 198*f*
 Hirst, W., 118*f*
 Hjertzen, D.G., 292*f*
 Hockey, B.J., 185*f*, 219*f*
 Hokkirigawa, K., 181*f*
 Holmberg, K., 323*f*, 348
 Houdková, Š., 246*f*
 Houmid-Bennami, H., 150*f*
 Hsu, S.M., 148*f*
 Hutchings, I.M., 74, 193*f*, 203*f*, 206*f*, 211*f*,
 225*f*, 227*f*, 267*f*
 Hutchings, R., 224*f*

I

Isaac, G., 348
 Ishigaki, H., 60*f*
 Ives, L.K., 214*f*

J

Jacobson, S., 247*f*, 330*r*, 331*f*, 348
 Jahanmir, S., 161
 Jarvis, R.A., 292*f*
 Jennett, N.M., 277*f*
 Jiang, J., 128*f*
 Jin, Z., 348
 Johnson, K.L., 25*f*, 32
 Jones, D.R.H., 368
 Jones, M.H., 103, 306*f*, 348
 Jost, H.P., 3*r*, 4

K

Kabil, Y., 216*f*
 Kapoor, A., 348
 Kato, K., 123*f*, 149*f*, 161, 181*f*
 Kayaba, T., 123*f*
 Kendoff, D., 349
 Kennedy, F.E., 74
 Khrushchov, M.M., 178*f*
 Kirkham, R.O., 348
 Kosel, T.H., 195*f*
 Krim, J., 74

L

Lancaster, J.K., 74, 118*f*, 153*f*, 156*f*, 158*f*,
 160*f*, 190*f*
 Landman, U., 42*f*
 Landolt, D., 231, 368
 Lansdown, A.R., 84*r*, 85*f*, 103, 300
 Lardner, T.J., 32
 Larsen-Basse, J., 196*f*, 232
 Lawn, B.R., 184*f*
 Leach, R.K., 33
 Lee, R.E., 55*f*
 Li, Y., 227*f*
 Liang, H., 348
 Lim, S.C., 132*f*, 161
 Ling, F.F., 74, 161
 Lu, Y., 74
 Ludema, K.C., 60*f*, 68*f*, 120*f*, 123*f*,
 147*f*, 152*f*
 Luedtke, W.D., 42*f*

M

Maekawa, K., 348
 Mairey, D., 150*f*
 May, P.W., 269*f*

Mazumder, J., 243*f*
McCabe, L.P., 220*f*
McCafferty, E., 368
McCool, J.I., 33
McFadden, D.H., 212*f*
McMillan, M.L., 325*f*, 349
Merchant, H.D., 161
Miller, J.E., 232
Mischler, S., 231, 368
Misra, A., 173*f*, 177*f*, 232
Miyoshi, K., 74, 161
Moore, M.A., 167*f*, 179*f*
Mortier, R.M., 103
Mufti, R.A., 348
Muhr, A.H., 157*f*
Mulhearn, T.O., 182*f*

N

Neale, M.J., 45*t*, 103, 300, 305*f*, 315*f*, 319*f*,
348
Newey, C., 368
Nicholls, J.R., 273*f*
Nisbet, T.S., 348
Nosonovsky, M., 4
Nowotny, H., 120*f*

O

Oakes, G., 308*f*
Obikawa, T., 348
Omar, M.K., 156*f*
Orszulik, S.T., 103
Oxley, P.L.B., 180*f*

P

Page, T.F., 32, 62*f*
Pan, C.H.T., 74, 161
Park, G., 348
Parks, J.L., 204*f*
Paul, J.P., 344*f*
Pennefather, R.C., 224*f*
Persson, B.N.J., 74
Peterson, M.B., 55*f*, 161, 278, 291*f*, 300, 348
Petterson, U., 247*f*
Pollock, H.M., 74
Pratt, G.C., 348
Preece, C.M., 232
Price, A.L., 300
Priest, M., 326*t*, 327*f*, 348

Q

Quinn, T.F.J., 127*t*, 161

R

Rabinowicz, E., 39*f*, 74, 232, 291*f*
Rajpura, A., 349
Ramalingam, S., 278
Rice, R.W., 185*f*
Rice, S.L., 120*f*
Rickerby, D.S., 277
Rigney, D.A., 74, 161, 232
Ringer, E.M., 42*f*
Ringlein, J., 74
Ritter, J.E., 33
Robbins, M.O., 74
Roberts, A.D., 157*f*
Robertson, J., 274*f*, 274*t*
Rogers, P.M., 226*f*
Rowson, D.M., 127*t*
Roy, M., 300
Ruff, A.W., 120*f*, 214*f*

S

Salamon, G., 142*f*
Samuels, L.E., 182*f*
Sargent, G.A., 220*f*
Sasada, T., 124*f*
Savage, R.H., 64*f*
Schaefer, D.L., 64*f*
Schmid, S.R., 349
Schmidt, F., 232
Schofield, J., 3*t*
Schwartz, S.E., 348
Scott, D., 103, 153*f*, 158*f*, 160*f*, 161, 190*f*,
306*f*, 348
Scully, J.C., 368
Shaw, M.C., 349
Sheldon, G.L., 215*f*, 221*f*
Shen, M.C., 148*f*
Shetty, D.K., 223*f*
Shipway, P.H., 136*f*, 260*f*
Singer, I.L., 74
Söderberg, S., 137*f*, 138*f*, 222*f*
Sparks, A.J., 172*f*
Spikes, H., 103
Stachowiak, G.W., 103, 330*t*, 348
Stack, M.M., 128*f*
Steen, W.M., 243*f*

Stobo, J.J., 348
Stott, F.H., 128*f*
Stout, K.J., 18*t*
Sullivan, J.L., 127*t*, 161
Sutton, D.C., 60*t*
Swain, M.V., 184*f*

T

Tabor, D., 39*f*, 41, 45*t*, 53*f*, 56*f*, 68*f*, 74, 97*f*,
102*f*, 161, 169*f*
Takadoun, J., 150*f*
Taylor, C.M., 326*t*, 327*f*
Thomas, T.R., 33
Towler, B., 278
Trezona, R.I., 203*f*
Tripp, J.H., 28*f*
Tung, S.C., 325*f*, 348–349

U

Uhure, N.J., 278

V

Vingsbo, O., 138–139*f*

W

Wahlström, J., 329*f*
Waterhouse, R.B., 161
Waterman, N.A., 321*f*
Wayne, S.F., 120*f*

Weaver, G., 368
Wellman, R.G., 273*f*
Weston, D.P., 260*f*
Whomes, T.L., 103
Wiederhorn, S.M., 219*f*
Williams, J.A., 33, 103
Williams, S., 348
Williamson, J.B.P., 26–29, 27*f*, 32
Wilshaw, T.R., 186*f*
Wilson, R.W., 122*f*
Wilson, S., 122*f*
Wilson, W.R.D., 349
Winer, W.O., 161, 278, 291*f*,
300, 348
Wolak, J., 216*f*
Wong, V.W., 349
Wood, R.J.K., 200*f*, 232, 368
Wright, I.G., 223*f*

X

Xu, J., 149*f*, 161

Y

Yamane, Y., 348

Z

Zhang, X.S., 267*f*
Zum Gahr, K.-H., 187*f*, 189*f*, 191*f*, 195*f*,
220*f*, 232

Subject Index

Note: Page numbers followed by *f* indicate figures and *t* indicate tables.

A

- Abbot-Firestone curve, 15
- Abrasion. *See also* Abrasive wear
 - hard and soft, 168, 194–196
 - patterns in rubber, 157, 157*f*
- Abrasion-corrosion, lubrication effects, 197–199
- Abrasion-corrosion, tests, 202
- Abrasive particles
 - degradation, 197, 199
 - properties, 167–174
 - shape, 169–172, 205
- Abrasive wear
 - abrasion-corrosion, 197–199
 - brittle fracture mechanisms, 183–188
 - degradation of abrasive particles, 197, 199
 - effect of hardness, 176–178, 189–192
 - effect of lubrication, 181–182, 197–199
 - effect of oxygen and humidity, 197–199
 - elastomers and rubbers, 152–153, 153*f*, 155, 157
 - high and low stress, 165
 - plastic deformation in, 174–182
 - polymers, 152–154, 153*f*
 - reduction, 287–288, 296–297
 - testing methods, 199–205
 - two and three body, definition, 165
- Accelerated wear tests, 113, 230, 285
- Acetal
 - abrasive wear, 190*f*
 - erosive wear, 224
 - in plain bearings, 320–321
 - sliding wear, 293–294
- Adhesion
 - metals and ceramics, 41–42, 41*f*
 - polymers, 66, 69–71, 151
- Adhesive wear, 107
- Air plasma spraying (APS), 265–266, 267*f*
- Alumina
 - abrasive particles, 202–204, 204*f*
 - abrasive wear, 198, 198*f*, 297
 - coatings, 265–266, 267*f*, 297
 - erosion-oxidation, 226*f*
 - erosive wear, 219–220, 219*f*, 222–223, 266–267, 296–297
 - formation in fretting, 140
 - friction, 58–59, 61–62, 62*f*
 - lubricated wear, 148–150, 148*f*
 - sliding wear, 145, 148–150, 148*f*
 - tribological applications, 294
- Aluminium and alloys
 - abrasive wear, 177*f*, 178
 - anodizing, 257
 - erosion-oxidation, 226–227, 227*f*
 - erosive wear, 211*f*, 215, 221–222
 - friction, 39*f*, 45*t*, 52
 - in plain bearing alloys, 314–318
 - surface hardening, 256
 - surface melting, 245
- Amontons, 37–38, 40
- Amplitude density function, 13, 15
- Anodizing, 257
- Antioxidants in lubricants, 85–86
- Anti-wear additives, 85–86, 96, 98–99, 99*f*
- Archard wear equation, 114–118
- Arc spraying, 265–266
- Areal average roughness (*S_a*), 17
- Articular cartilage, 341–342
- Asperities
 - definition, 18
 - deformation, 23–26, 48–49
- Atomic force microscopy (AFM), 7
- Attack angle of abrasive particles, 180–182, 180*f*
- Austenite, retained, 144, 190–191, 244–245
- Austenitic stainless steels. *See* Steels
- Autocatalytic, 261
- Autocatalytic plating, 261
- Autocorrelation function, 15–16

Automotive tribology

- brakes, 329–330
- engine components, 324–328
- tyres, 331–333
- transmission components, 328

B

- Babbitts. *See* Whitemetals
- Ball bearings. *See* Rolling bearings
- Ball-cratering/dimple grinding, 276–277
- Bearing ratio curve, 14*f*, 15
- Bearings
 - big-end/crankpin, 90–92
 - gas, 91
 - geometries, 304
 - plain, 312–318
 - rolling, 303–323
 - selection, 319–321, 321*f*
- Beryllium-copper. *See* Copper-beryllium
- Beryllium, friction, 55
- Big-end bearings, 90–92, 326–327
- Biotribology
 - synovial joint, 341–342
 - total hip replacement (THR), 342–345
- Bismuth
 - erosive wear, 216*f*
 - friction, 56*f*
- Bolted joints, fretting, 135
- Bond coats in thermal spraying, 264
- Boriding. *See* Boronizing
- Boron carbide, 296–297
- Boronizing, 242*f*, 248, 252, 297
- Boron nitride as lubricant, 101
- Boundary lubrication, 79, 96–100, 141–144, 141*f*, 288
- Bowden and Tabor model for friction, 41–44
- Brake disks, 329, 329*f*
- Brake fade, 330
- Brakes, 329–330
- Brakes, aircraft, 150
- Brasses
 - abrasive wear, 176–178
 - erosive wear, 222*f*
 - friction, 45*t*, 53–54
 - sliding wear, 117, 118*f*
 - surface hardening, 256
- Breaking-in. *See* Running-in

Brittle fracture

- in erosive wear, 217–220
- in wear of ceramics, 145–147, 183–188

Brittleness index, 183–184

Butyl stearate, 97

C

- Cadmium
 - erosive wear, 215–216*f*
 - friction, 45*t*, 52
- Cadmium chloride, friction, 65
- Cage, rolling bearing, 304–305, 309
- Calcium fluoride, as lubricant, 101
- Calcium hydroxide, friction, 65
- Capacitance, 113
- Carbides in abrasive wear, 178, 192–197, 196*f*
- Carbon-carbon composites, 150, 322–323
- Carbon fillers in polymers, 159
- Carbonitriding, 242*f*, 248, 295*f*
- Carboxylic acids, 96–97, 97*f*
- Carburizing, 242*f*, 294–295
- Case hardening. *See* Carburizing
- Cast irons
 - abrasive wear, 189
 - grey, 295
 - grey, abrasive wear, 187–188
 - grey, friction, 45*t*, 53–54
 - surface melting, 246
 - transformation hardening, 244
 - white, 263, 296–297
 - white, abrasive wear, 191–192, 194, 195*f*, 203–204, 204*f*, 296–297
- Castor oil, 97
- Cemented carbides, 192, 194–196, 223–224, 263, 296–297
 - abrasive wear, 192–197, 296–297
 - boronizing, 255
 - erosive wear, 219–220, 223–224, 296–297
 - hardness, 194–196
- Centrifugal accelerator, erosive wear, 229
- Ceramics. *See also* Specific ceramics
 - abrasive wear, 184–186, 185–186*f*
 - erosive wear, 218–220, 219*f*
 - fracture in sliding contacts, 61, 62*f*
 - friction, 58–62
 - hardness values, 170–171*t*
 - lubricated wear, 150–151, 294
 - rolling bearings, 307–308

- sliding wear, 145–151, 293–296
 - tribochemical reactions, 59, 62
 - unlubricated sliding wear, 145
 - Cermets. *See* Cemented carbides
 - Cetyl alcohol, 97
 - Chemical mechanical planarization (CMP), 338–340
 - Chemical vapour deposition (CVD), 268–270, 294
 - Chemisorption of boundary lubricants, 96–97
 - Chemomechanical effects, 145
 - Chlorine, effect on friction, 55
 - Chromia, coatings, 297
 - Chromium
 - carbide, 256, 297
 - friction, 45*t*, 52
 - plating, 260
 - Chromizing, 248, 252, 256
 - Coated surfaces, contact of, 31
 - Coatings, 258–276
 - Cobalt and alloys, 263
 - abrasive wear, 192–197
 - friction, 55–56
 - galling, 293
 - Coefficient of friction
 - in abrasive wear, 176
 - definition, 37
 - rolling bearings, 309–312
 - typical values for metals, 45*t*
 - Coefficient of wear. *See* Wear coefficient
 - Cold spraying/kinetic metallization, 275
 - Compatibility
 - in plain bearing materials, 312
 - in sliding wear, 293, 295–296
 - Composites
 - abrasive wear, 192–197, 193*f*
 - carbon-carbon, 150, 322–323
 - erosive wear, 219–220, 228–229
 - Compression rings, 325–326
 - Concentrated contact. *See* Counterformal contact
 - Cone cracking. *See* Hertzian fracture
 - Conformability, in plain bearing materials, 312–313
 - Conformal contact, 86, 87*f*, 111–112, 111*f*
 - Constant velocity joints, 328
 - Contact resistance, 113, 118*f*
 - Copper and alloys
 - abrasive wear, 173*f*, 176–178, 177*f*
 - erosive wear, 214, 214*f*, 221–222
 - friction, 45*t*, 51–52, 56*f*
 - plain bearing alloys, 314–318
 - surface hardening, 256
 - Copper-beryllium, 118*f*, 168–169, 170–171*t*
 - Copper-gold, friction, 57
 - Copper-lead bearing alloy, 45*t*, 53–54, 316–318
 - Copper oxide, filler in HDPE, 159
 - Corrosion, fundamentals, 361–368
 - Corrosion, in plain bearings, 312–313
 - Coulomb, 38, 40
 - Coulomb model for friction, 40
 - Counterformal contact, 92, 111–112, 111*f*
 - Crankpin bearing, 90–92
 - Crankshaft bearings, 314–315, 315*f*, 326–327
 - Cubic boron nitride (CBN), 339
 - Cutting deformation
 - in abrasive wear, 178–180, 180*f*
 - in erosive wear, 209–210, 210*f*, 212–213, 212*f*
 - Cutting tools, 270, 294
 - CVD coatings, 294, 297
- D**
- Delamination mechanism of sliding wear, 124–125, 131
 - Delsun process, 256
 - Design, tribological aspects, 283–300
 - Detonation-gun (D-gun) spraying, 265–267
 - D-gun process, 265
 - Diamond
 - bulk polycrystalline, 296–297
 - film deposition by CVD, 294
 - friction, 59–60
 - sliding wear, 150
 - Diamond-like carbon (DLC) coatings, 273–275, 327–328, 346–348
 - Dry bearings, 318–323
 - Duplex surface engineering, 295
 - Dynamic friction, 39
 - Dynamic viscosity, 80
- E**
- Economic factors in tribology, 2–3*t*
 - EHL. *See* Elastohydrodynamic lubrication (EHL)
 - Elastic line contact, 21–22

- Elastic point contact, 19–21
 Elastic stress fields, tractional load, 24
 Elastohydrodynamic lubrication (EHL), 79, 92–95, 140–144, 327–328
 Elastomers
 abrasive wear, 152–153, 153*f*, 155, 157, 198, 200–201
 erosive wear, 224–225, 297
 sliding wear, 152–153, 153*f*, 155, 157
 Electrical discharge machining (EDM), 246–247
 Electric arc spraying, 263–264
 Electrodeposition. *See* Electroplating
 Electroless plating, 261
 Electron beam hardening, 243–244
 Electron-beam PVD (EBPVD), 272–273
 Electron beam surface treatments, 243–245
 Electroplating, 259, 261
 Embeddability, in plain bearing materials, 312–313
 Energy dissipation in friction, 40
 Energy savings through tribology, 3*t*
 EP (extreme pressure) additives, 97–98, 143–144
 Epoxy resins, erosive wear, 224
 Erosion-corrosion, 225–227, 231
 Erosive particles, properties, 167–174
 Erosive wear, 206–231
 by brittle fracture, 217–220
 definition, 165–166
 effect of impact velocity, 214, 214*f*
 effect of material hardness, 213–220
 by plastic deformation, 206–217
 reduction, 287–288, 297
 test methods, 227–231
 Evaporation (coating method), 270
- F**
- Fatigue
 fretting, 135, 296
 in plain bearings, 312–314
 in rolling contacts, 284, 306–309, 311–312
 in sliding wear of metals, 125, 130
 thermal, in whitemetal, 314
 in wear of polymers, 152–157
 Fatty acids, 96–97, 97*f*
 Fillers in polymers, effect on wear, 159–160
 Film thickness ratio. *See* Lambda ratio
- Fine grinding, 338–339
 Flame hardening, 241
 Flame spraying, 264–266, 265*f*
 Flash temperature, 72–73, 131, 133
 Food processing machinery, lubrication and bearings, 100, 322
 Form and finish grinding, 338–339
 Form error, definition, 9
 Fracture
 in erosive wear, 217–220
 in wear of ceramics, 145–147, 183–188
 Fretting, 134–140
 corrosion, 140
 fatigue, 135, 296
 wear, reduction, 286–287, 296
 Fretting wear, 134–140, 296
 Friction
 adhesion contribution, 43
 Bowden and Tabor model, 41–44
 coefficient of, 37
 cutting, 133
 definition, 37
 deformation contribution, 43
 effect of ion implantation, 251–252
 heating, 71–74
 in hydrodynamic lubrication, 89
 junction growth, 45–48
 laws of, 37–39
 material, 328–329
 metals, 50–57
 stir welding, 275–276
 views of, 48–50
 work hardening in, 44
 Fusion processes, 262
- G**
- Galling, 107, 293, 296
 Garnet, abrasive particles, 170–171*t*, 203–204, 204*f*
 Gas
 bearing, 91
 nitriding, 242*f*, 248, 253–254
 Gas-blast erosion test, 228–229
 Gas carburizing, 249
 Gas nitriding, 253–254
 Gears, surface engineering, 294–295
 Glass
 abrasive wear, 188

- deformation and fracture in sliding contact, 147*f*, 188
- erosive wear, 219*f*, 221, 221*f*
- indentation fracture, 188
- Gold
 - friction, 45*t*, 51
 - sliding wear, 292
 - solid lubricant, 102, 322–323
- Gouging abrasion, 165, 204–205
- Graphite
 - fluoride, 65
 - friction, 63–65, 63*f*, 65*f*
 - in plain bearings, 319–323
 - solid lubricant, 100–101
- Greases, compositions and properties, 86
- Grinding, 338–340
- Grinding
 - form and finish, 338–339
 - geometry of, 338–339, 338*f*
 - stock removal, 338–339
- Gudgeon pin, 327
- Gun barrel, friction in, 133
- H**
- Hafnium, sliding wear, 292–293
- Hardfacing. *See* Weld hardfacing
- Hardfacing processes, 262
- Hardness
 - correlation with abrasive wear in polymers, 154
 - effect on abrasive wear, 176–178, 189–192
 - effect on erosive wear, 213–220
 - effect on sliding wear, 293, 296
 - ion-implanted surfaces, 251
 - measurement, 23, 353–359
 - particles, 167–169
 - typical values, 170–171*t*
- Hard phases, in abrasive wear, 192–194, 193*f*
- HDPE
 - friction, 66, 71
 - sliding wear, 159
- Heathcote slip, 310–312, 311*f*
- Heating, frictional, 71–74
- Hertzian contact theory, 31, 146, 147*f*
- Hertzian fracture, 145–147, 183, 217–218
- Hexadecanol, 97
- Hexagonal close packed metals, sliding wear, 292–293
- High velocity oxy-fuel (HVOF) spraying, 263–265
- Hot-spots, 125–126, 133
- Humidity effects
 - in abrasive wear, 197–199
 - in fretting wear of metals, 140
 - in friction of ceramics, 59, 60*t*
 - in wear of ceramics, 147–150
- Hydrodynamic lubrication, 79, 86–92
- Hydrogen embrittlement, 261
- I**
- Impact angle in erosive wear, 209–210, 210–211*f*, 212–214, 215*f*, 220–221
- Impact velocity in erosive wear, 214, 214*f*, 217*f*, 218
- Incubation period in erosive wear, 208–209, 227, 285
- Indentation fracture, 183–184, 184*f*, 217–218
- Indentation, plastic, 23
- Indium
 - demonstration of adhesion, 41
 - friction, 45*t*, 52
 - thin film lubricant, 102
- Induction hardening, 241
- Interferometry, 8
- Ion
 - implantation, 251, 294–295
 - nitriding, 253–254
 - plating, 272
- IRG transition diagram, 142, 142*f*
- Iron
 - friction, 51, 51*f*, 54*f*
 - oxides, 53–54, 126, 140, 144, 166
- J**
- Jaw-crusher test, 204–205, 205*f*
- Jet-impingement erosion test, 228–229
- Johnson-Kendall- Roberts (JKR) theory, 31–32
- Journal bearings, pressures, 91
- Junction growth, 45–48
- K**
- Kinematic viscosity, 81
- Kurtosis, 14, 17

L

- Lambda ratio
 - measure of fluid film thickness, 95, 140–141, 287–288
- Lamellar solids
 - friction, 63–65
 - as lubricants, 101
- Lapping, 339–340
- Laser cladding, 262
- Laser hardening, 242–243
- Laser surface treatments, 243–244
- Lateral fracture, 183–185, 217–220
- LDPE, 66
 - abrasive wear, 190*f*
 - sliding wear, 159–160
- Lead
 - erosive wear, 215*f*
 - friction, 45*t*
 - in plain bearing alloys, 313–314, 316
 - thin film lubricant, 102
- Lead bronzes, bearing alloys, 319–320
- Lead iodide, friction, 65
- Lead oxide, filler in HDPE, 159
- Leonardo da Vinci, 37–38
- Liquid carburizing, 251
- Lubricant
 - additives, 86
 - definition, 79
- Lubricated wear
 - ceramics, 150–151, 294
 - metals, 140–145
 - polymers, 160–161
- Lubrication, 79–102, 104–105
 - boundary, 79, 96–100, 141–144, 141*f*, 288
 - ceramics, 150–151
 - effect on abrasive wear, 181–182, 197–199
 - effect on sliding wear, 288, 290–296
 - EHL, 79, 92–95, 142–144
 - hydrodynamic, 79, 86–92
 - plain bearings, 312
 - polymers, 160
 - rolling bearings, 309–312
 - solid, 79, 100–102, 141, 144
 - squeeze-film, 90
- Lubricity, 97

M

- MacCoull-Walther equation, 81–82
- Machined surfaces, topography, 17–18
- Machining
 - effect of environment, 199
 - slip-line field, 180*f*
- Magnesia, erosive wear, 219–220, 219*f*
- Magnesium
 - erosive wear, 216*f*
 - friction, 45*t*
- Magnetic data storage, 345–348
- Magnetic tape, wear by, 199
- Manganese steels, 263
- Manual metal arc (MMA) welding, 262
- Manufacturing, tribology in
 - grinding, 338–340
 - metal-cutting, 333–338
 - polishing and chemical–mechanical planarization, 338–340
- Maps, wear-regime, 132–134, 142, 148, 286–287
- Marginally lubricated plain bearings, 318–323
- Martensite
 - formation in sliding, 133–134
 - hardness, 244–245, 244*f*, 296–297
- Materials selection, 289–300
- Median fracture, 183
- Melting in sliding of metals, 56, 133
- Metal inert gas (MIG), 262–263
- Metallizing, 256
- Metals
 - abrasive wear, 174–182, 189–199
 - adhesion, 41–42, 42*f*
 - contact, 30
 - erosive wear, 206–217, 221–223
 - friction, 50–57
 - hardness values, 170–171*t*
- Mica
 - friction, 65
 - smoothness, 7
- Micro-arc oxidation (MAO), 257–258
- Micro-elastohydrodynamic model, 342
- Micro-scale abrasion test, 201–202, 276–277
- Microslip, 135
- Mild wear
 - ceramics, 147–148
 - metals, 127–132, 142
 - polymers, 159–160

Minimal quantity lubrication (MQL), 338

Mixed EHL, 95, 141

Mohs hardness scale, 168, 169*f*

Molecular dynamics, modelling of adhesion, 42

Molybdenum disulfide, 295

friction, 63–65

in plain bearings, 320–321

solid lubricant, 100

erosive wear, 215

Multiple asperity contact

simple theory of, 26

statistical theories of, 26–30

N

Nickel and alloys, 263

erosive wear, 215–216*f*, 221–222

friction, 55, 55*f*

galling, 293

plating, 259–260

Niobium carbide, 256

Nitriding, 242*f*, 248, 252–254, 253*f*, 294–295

Nitrocarburizing, 242*f*, 248, 252, 254

Nylons, 101

abrasive wear, 190*f*

erosive wear, 222*f*, 224

lubrication, 295

in plain bearings, 320–321

rolling bearing cages, 309

sliding wear, 158*f*, 160, 160*f*, 293–294

O

Octadecanoic acid. *See* Stearic acid

Oil

compositions and properties, 83–86

viscosity, 81

Oiliness, 97

Operating variables, 286–288

Order-disorder transformation, effect on

friction, 57, 58*f*

Osteoarthritis, 342–343

Overlay coatings in plain bearings, 316–318, 316*f*

Oxidation

in abrasive wear, 197–199

in fretting wear of metals, 140

in friction of ceramics, 59–60

in friction of metals, 50–52

in sliding wear of ceramics, 145, 148–150

in sliding wear of metals, 125–127,

142–143, 293–294

P

Pack carburizing, 249

Paraffinic oils, 84–85

Partial EHL, 95, 141–144

Particle

degradation in abrasion testing, 197, 199

degradation in erosion testing, 229

flux, in erosion testing, 230–231

hardness, influence on abrasive or erosive wear, 167–169

shape, in abrasive and erosive wear,

169–172, 205, 209–212, 210*f*

size, in abrasive wear, 173–174, 186–187

velocity, in erosion testing, 229–230

Péclet number, 73

PEEK. *See* Polyetheretherketone (PEEK)

Phase transformations, effect on friction, 56–57

Physical vapour deposition (PVD), 268,

270–273, 294

evaporation, 270

ion plating, 272

Pin-on-disc test, 110*f*, 111

Pipelines, bend design, 288

Plain bearings, 303, 304*f*, 312–318

dry, 318–323, 319*f*

high temperature, 319–320, 322–323

marginally lubricated, 318–323

materials for high stress applications,

314–318

materials for low stress applications,

313–314

PV values, 318–322

Plasma

carburizing, 242*f*, 250

nitriding, 242*f*, 250, 253–254

spraying, 263–267, 267*f*

Plasma-assisted CVD (PACVD), 241*f*

Plasma carburizing, 250

Plasma electrolytic oxidation (PEO), 257–258

Plasma nitriding, 253–254

Plasma-sprayed coatings, 294

Plasma transferred arc (PTA) hardfacing, 262

Plastic deformation, contacts, 22–23

- Plasticity
 in abrasive wear, 174–182
 in erosive wear, 206–217
 index, 29, 30*t*
 in sliding wear of metals, 122–125, 128–129
- Plastics materials. *See* Polymers
- Plateau honing process, 326
- Platinum, sliding wear, 292
- Ploughing deformation
 in abrasive wear, 178–181, 180*f*
 in erosive wear, 209–210, 210*f*
- Point contact, definition, 19
- Polishing and lapping processes
 generic illustration of, 339–340, 340*f*
- Polishing planarization, 338–340
- Polyamides. *See* Nylons
- Polycarbonate
 abrasive wear, 190*f*, 224
 erosive wear, 224
- Polyetheretherketone (PEEK), 101, 294
 plain bearings, 320–322
- Polyethersulfone (PES)
 rolling bearing cages, 309
 sliding wear, 293–294
- Polyethylene
 high density (*see* HDPE)
 low density (*see* LDPE)
 ultra-high molecular weight (*see* Ultra-high molecular weight polyethylene (UHMWPE))
- Polyimides, 101
- Polymers. *See also* Specific polymers
 abrasive wear, 152–154, 153*f*, 189–190, 190*f*, 198
 adhesion component of friction, 66, 69–71, 151
 cohesive wear mechanisms, 151
 deformation component of friction, 66–68
 erosive wear, 222*f*, 224
 friction, 66–71
 interfacial wear mechanisms, 151, 158–160
 lubricated wear, 160–161
 nature of contact, 29
 plain bearings, 320–322, 321*f*
 rolling bearings, 308–309
 sliding wear, 151–161
- Polymethylmethacrylate (PMMA)
 abrasive wear, 190*f*
 erosive wear, 222*f*, 224
 friction, 66, 69*f*
 sliding wear, 158, 158*f*
- Polypropylene
 abrasive wear, 190*f*, 224
 erosive wear, 224
- Polystyrene, sliding wear, 158
- Polytetrafluoroethylene (PTFE)
 abrasive wear, 190*f*
 friction, 66, 68, 68*f*, 71
 influence of water on wear, 161
 in plain bearings, 320–322
 sliding wear, 158*f*, 159, 161
 solid lubricant, 100–101
- Polyurethanes, 297
- Polyvinylchloride (PVC), 158, 224
 abrasive wear, 190*f*
 erosive wear, 224
 sliding wear, 158
- Pour-point depressants, 85–86, 85*t*
- Power spectral density, 16
- Preston equation, 339–340
- Primary shear zone, 333
- Profilometry, 7–9
- PVD. *See* Physical vapour deposition (PVD)
- PVD coatings, 270–273, 294–295
- PV rating in bearing selection, 318–319, 322
- ## Q
- Quartz. *See* Silica
- ## R
- Ra*, definition, 12–13
- Radial-ply tyre, 331–333, 331*f*
- Rapeseed oil, 97
- Ratner-Lancaster correlation, 154, 154*f*
- Reactive
 evaporation, 270
 ion plating, 272
 sputtering, 272
- Recirculating loop test, erosive wear, 229
- Registration, Evaluation, Authorization and Restriction of Chemicals (REACH), 299–300
- Rehbinder effect, 145, 148–150
- Residual stress, in fretting wear, 296
- Retained austenite, 144, 190–191, 244–245
- Reynolds equation, 87–91

- Reynolds slip, 310–312, 310*f*
Rhodium, sliding wear, 292
Riveted joints, fretting, 135
Roller bearings. *See* Rolling bearings
Roll formation in rubber, 157
Rolling bearings, 303–323
 cages, 294, 309
 fretting damage, 135
 friction and lubrication, 309–312
 high temperature, 307–308
 lifetimes, 308–309
Rolling contact fatigue, 284, 306–309, 311–312
Rolling friction, 309–312
Roughness
 definition, 12
 effect on galling, 293
 measures, 12
 typical values, 18*t*
Roundness of particles, 171–172
Rq, definition, 13
Rubber
 abrasive wear, 152–153, 153*f*, 155, 157, 201, 201*f*
 EHL, 92
 erosive wear, 222*f*, 224–225, 231, 297
 sliding wear, 152–153, 153*f*, 155, 157
Rubber wheel abrasion test, 201–203, 201*f*, 205
Running-in, 113, 117, 134, 142–144, 285, 288–289
- S**
Sapphire, erosive wear, 219–220, 219*f*
Scoring, 107. *See also* Scuffing
Scratch hardness, 168
Scratch test, 276
Scuffing, 95, 107, 142–144
Secondary shear zone, 334–335
Seizure, metals, 132*f*
Selection of materials, 289–300
Severe wear
 ceramics, 147–148
 metals, 127–132
 polymers, 159–160
Silica
 abrasive particles, 168–169, 170–171*t*, 172*f*, 296–297
 erosive wear, 219–220, 219*f*
Silicon carbide
 abrasive particles, 202–204, 204*f*
 erosive wear, 219–220, 219*f*, 223, 296–297
 friction, 58–59, 61, 62–63*f*
 lubricated wear, 150
 sliding wear, 149
Silicon, erosive wear, 219–220, 219*f*
Silicon nitride
 erosive wear, 219–220, 219*f*, 223, 296–297
 friction, 58–59, 60*f*, 62, 63*f*
 lubricated wear, 150
 sliding wear, 149, 149–150*f*
 tribological applications, 294
Silver
 erosive wear, 216*f*
 friction, 45*t*
 sliding wear, 293
Skewness, 14
Skis, friction of, 56
Sliding speed, effect on friction, 39, 56, 56*f*, 61–62, 63*f*
Sliding wear, 290–296
 ceramics, 145–151, 294
 debris nature, 121
 definition, 3
 dependence on load and distance, 116–117, 118*f*
 elastomers and rubbers, 152–153, 153*f*, 155, 157
 hard coatings, 294
 lubricated metals, 107, 140–145
 nature of surface layers, 120–121
 oxidative mechanisms, 125–127, 142–143
 plasticity-dominated mechanisms, 122–127, 147–148
 polymers, 151–161
 reduction, 287, 293
 surface layer changes, 119–120
 unlubricated metals, 107, 128–129, 132
Slurry erosion, 165
 in pipelines, 284–285
 reduction, 284–285
 test methods, 228–230
Soft EHL, 92
Solid lubrication, 79, 100–102, 141, 144
Solid particle erosion. *See* Erosive wear
Solvents, effect on wear of polymers, 156–157
Space, lubrication in, 100, 102

- Specific energy for abrasive wear, 176
- Sputtering, 271
- Sputtering (coating method), 271
- Squeeze-film lubrication, 90
- Stacking fault energy, in galling wear, 293
- Standards, sliding wear tests, 113
- Standoff distance, 263–264
- Static friction, coefficient of, 39
- Stearic acid, 97, 150
- Steels
- abrasive wear, 176, 177–178*f*, 178, 190–193, 191*f*, 203–204, 204*f*, 296–297
 - erosion-oxidation, 226*f*
 - erosive wear, 211*f*, 215, 216*f*, 220–223, 220*f*, 222*f*, 296–297
 - fretting wear, 138–139, 139*f*
 - friction, 45*t*, 53–54, 55*f*
 - galling, 293
 - hardness values, 170–171*t*
 - lubricated sliding wear, 142–144
 - nitriding, 252–253
 - oxides on, 53–54, 125–127
 - rolling bearings, 306–309
 - surface melting, 245
 - transformation hardening, 240–245
 - unlubricated sliding wear, 128–129, 132–134, 132*f*
- Stellites
- abrasive wear, 192
- Stock removal grinding, 338–339
- Strain in abrasion of metals, 176–178, 179*f*
- Strain rates in erosive wear, 216–217, 217*f*
- Stribeck curve, 89–90, 90*f*, 141, 326, 327*f*
- Surface
- areal measures, 16–17
 - coated, 31
 - contact, 18–32
 - elastic contact, energy contributions in, 31–32
 - engineering, 289–300
 - melting, 245–246
 - slope, 9
 - topography, measurement, 9–16
- Surface-active phospholipid (SAPL), 342
- Surface melting, 245–246
- Surfaces texturing, 246–247
- Synovial joint, 341–342
- Systems approach to wear, 286
- T**
- Talc, friction, 65
- Tantalum
- erosive wear, 216*f*
 - friction, 56
- Taylor's tool life equation, 335–336
- Temperature
- effect on friction, 54–57, 61–62, 62*f*
 - effect on sliding wear of metals, 129–131
 - effects in polymer wear, 159–160
 - rise in sliding contact, 131
- Testing methods
- abrasive wear, 199–205
 - erosive wear, 227–231
 - sliding wear, 109–114
- Thermal
- hardening, 295*f*
 - spraying, 262–267, 295*f*
- Thermal barrier coatings (TBCs), 272–273
- Thermal spraying processes, 263
- Thermochemical reactions, 252–256
- Thermochemical surface treatments, 248–249, 252–253, 256
- Thermoelastic instability, 151
- Thin metal films as lubricants, 102
- Thin wall bearings, 315–316
- Tin
- erosive wear, 215*f*
 - friction, 45*t*, 52, 56–57
 - in plain bearing alloys, 313–314
 - thin film lubricant, 102
- Tyres, 331–333
- Tyres, wear of, 198
- Titania coatings, 297
- Titanium and alloys
- boronizing, 255
 - fretting wear, 140, 296
 - friction, 55
 - sliding wear, 137, 138*f*, 293
 - surface hardening, 256
- Titanium carbide, 256
- CVD, 268
 - fracture, 146–147
 - friction, 59
- Titanium nitride, 294
- friction, 58–60
 - PVD, 268
- Topography, surface, 7–18

- Total hip replacement (THR), 342–343
 annual rates of, 342–343
 lubrication mechanisms, 344–345
 wear rate in, 345
- Toyota diffusion (TD) process, 256
- Transfer films in polymer sliding, 70–71,
 158–159, 322
- Transformation hardening, 240–245
- Transition diagram, IRG, 142, 142*f*
- Transitions
 ductile-brittle, 188, 221
 sliding wear, 117, 127
 wear mechanism in ceramics, 147–148,
 286–287
 wear mechanism in metals, 113, 132–134,
 286–287
- Tribochemical reactions, 59, 62, 145, 149,
 149*f*
- Tribocorrosion, 366–368
- Tribologically transformed zone (TTZ), 119
- Tribology, definition and etymology, 1
- Tribometer, 109
- Tribotester, 109
- Tufftride process, 254
- Tungsten
 CVD coatings, 268
 erosive wear, 215
- Tungsten carbide, 297
- Tungsten carbide/cobalt. *See* Cemented
 carbides
- Tungsten disulfide
 friction, 65
 solid lubricant, 101
- Tungsten inert gas process (TIG), 245,
 262–263
- U**
- Ultra-high molecular weight polyethylene
 (UHMWPE), 297, 342–343
 wear and friction, 152*f*, 297
- V**
- Vacuum
 carburizing, 242*f*, 250
 friction of metals in, 50
 lubrication in, 100, 102
 plain bearings for, 322
 plasma spraying, 266–267, 267*f*
- Vacuum carburizing, 250
- Vacuum plasma spraying (VPS), 265–267
- Valve train, 324–325
- Vanadium carbide, 256
- Vegetable oils, 97
- Velocity exponent in erosive wear, 213–214,
 218–219
- Velocity measurement in erosion testing,
 229–230
- Vickers hardness, 23
- Viscoelasticity in polymer friction, 66–67
- Viscosity, 79–83
- Viscosity index (VI), 82
- W**
- Water. *See also* Humidity
 effect on fretting wear of metals, 140
 effect on wear of ceramics, 145,
 148–150
 effect on wear of PTFE, 161
 viscosity, 80
- Waviness, definition, 9
- Wear
 abrasive (*see* Abrasive wear)
 accelerated tests, 230, 285
 coefficient, 116
 coefficient, estimation, 285, 286*t*
 coefficient for abrasive wear, 176,
 182
 debris, metals, 122–127, 122*f*
 definition, 1
 erosive (*see* Erosive wear)
 fretting (*see* Fretting wear)
 rates, 284–285, 369–371
 reduction, 286–287, 293, 296
 sliding (*see* Sliding wear)
 testing methods (*see* Testing methods)
 value of laboratory tests, 285
- Wear coefficient, 116, 209, 286*t*, 291*f*
- Wedge-formation in abrasive wear, 179,
 180*f*
- Weld hardfacing, 262–263, 295*f*, 297
- Whirling arm erosion test, 229
- White layer, 254
- Whitemetals, 313–314
 friction, 45*t*, 53–54, 314
- Wire ropes, fretting, 135
- Wood, friction against steel, 39*f*

Z

Zinc dialkyldithiophosphate (ZDDP),
98–100, 99*f*

Zinc, sliding wear, 128–129

Zirconia
coatings, 297

friction, 58–59, 61–62

lubricated wear, 151

sliding wear, 149–150

Zirconium

friction, 55

sliding wear, 292–293