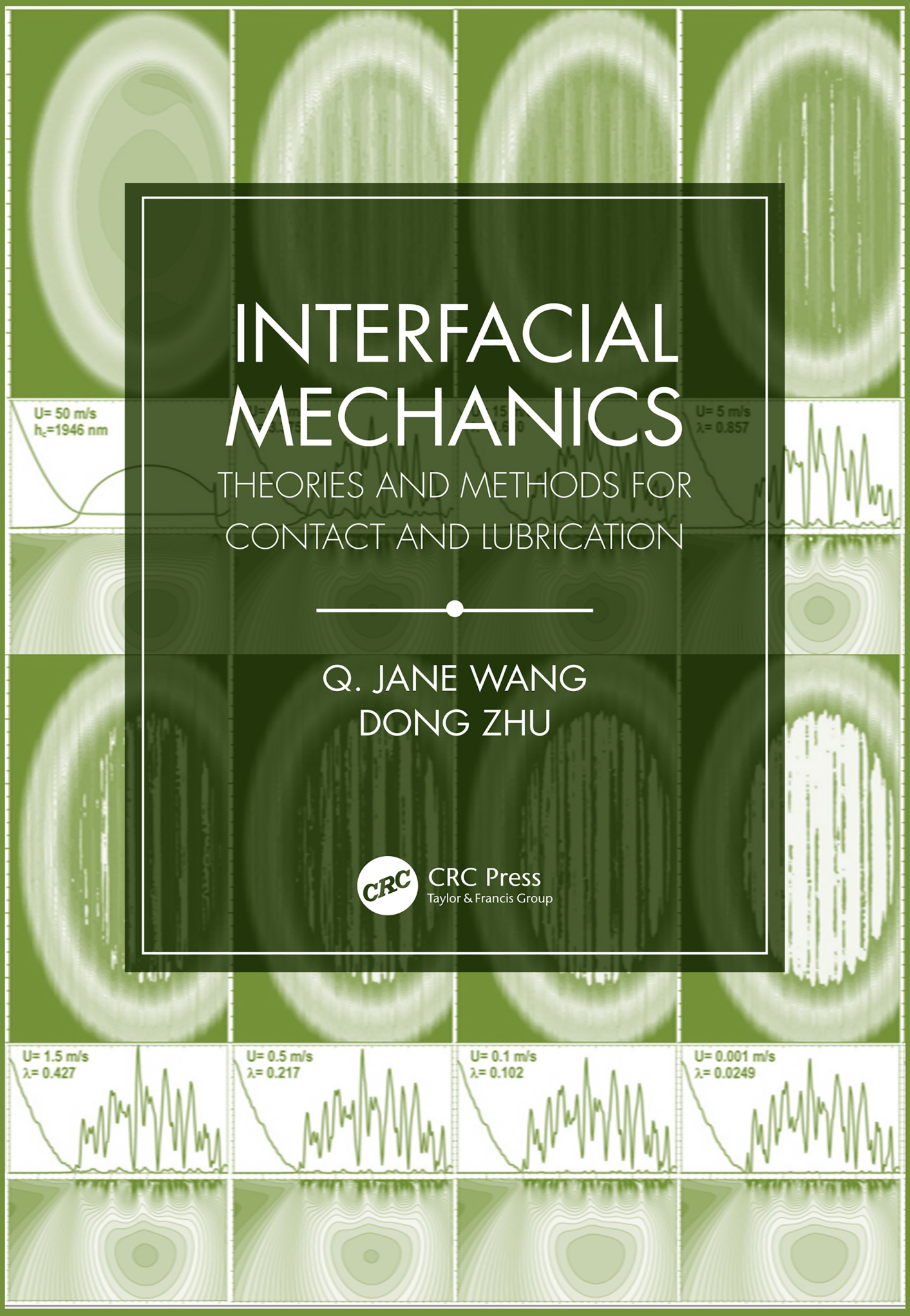


INTERFACIAL MECHANICS

THEORIES AND METHODS FOR CONTACT AND LUBRICATION

Q. JANE WANG
DONG ZHU

CRC Press
Taylor & Francis Group

Interfacial Mechanics

Interfacial Mechanics

Theories and Methods for Contact and Lubrication

Q. Jane Wang and Dong Zhu

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed on acid-free paper

International Standard Book Number-13: 978-1-4398-1510-6 (Hardback)

Library of Congress Cataloging-in-Publication Data

Names: Wang, Q. Jane- author. | Zhu, Dong, author.
Title: Interfacial mechanics : theories and methods for contact and lubrication / Q. Jane Wang, Dong Zhu.
Description: First edition. | Boca Raton, FL : CRC Press/Taylor & Francis Group, 2019. | Includes bibliographical references and index. | Summary: "This book discusses "tribological interface" that consists of two solid surfaces in contact with or without fluids in between. This specific type of interface is commonly seen in reality and extremely important in engineering applications. This book is written for engineering researchers and design engineers as well as graduate and senior undergraduate students. Mathematical treatments are tailored to a first degree in engineering often without rigorous descriptions and proofs. It focuses on the basic concepts, mathematic models, numerical solution procedures, major results and their physical meanings, as well as engineering applications"— Provided by publisher.
Identifiers: LCCN 2019026355 (print) | LCCN 2019026356 (ebook) | ISBN 9781439815106 (hardback) | ISBN 9780429131011 (ebook)
Subjects: LCSH: Tribology—Mathematics. | Interfaces (Physical sciences)—Mathematical models. | Lubrication and lubricants.
Classification: LCC TJ1075.W36 2019 (print) | LCC TJ1075 (ebook) | DDC 621.8/90151—dc23
LC record available at https://lccn.loc.gov/2019026355
LC ebook record available at https://lccn.loc.gov/2019026356

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

Preface

Power and motion are transmitted through interfaces (lubricated or dry) at surface contact locations in various machine components. According to the estimates presented in the well-known Jost report (1966) and other publications, roughly one third of energy produced globally from different sources (petroleum, coal, hydraulic, nuclear, solar, wind, agricultural, and others) is dissipated by frictional loss that takes place largely with surface interaction. Also, it has been found that a majority of mechanical components failures originate at the interfaces due to severe surface contact and insufficient lubrication. A good understanding of surface contact and lubrication characteristics, therefore, is vital to components design and product development. Its importance can never be overestimated, as it is fundamental to the improvement of machine performance, efficiency, durability, and reliability, as well as to the conservation of energy and materials that underpins environmental protection.

It is commonly agreed that contact mechanics, as a branch of engineering science, originated from the study by Heinrich Hertz in 1881, when he published his classic paper, *On the Contact of Elastic Solids*. Over the last 130+ years, the Hertzian theory has been enjoying wide application in engineering practice due to its simplicity and satisfactory accuracy for frictionless elastic counterformal contact of smooth surfaces. Many researchers have made contributions to the field of contact mechanics since then, but, surprisingly, there have been only a small number of books found in the literature, despite the importance of this subject. These include the books by L. A. Galin (1953), G. M. L. Gladwell (1980), K. L. Johnson (1987), D. A. Hills et al. (1993), and I. G. Goryacheva (1998).

It is also widely accepted that the foundation of the modern lubrication theory was laid by Osborne Reynolds in 1886. Under certain conditions, a complete separation of surfaces could happen in a lubricated contact due to hydrodynamic action at the interface. In practice, lubrication with a full hydrodynamic lubricant film is often the best means to reduce friction and prevent surface failure when possible. The Reynolds equation mathematically describes the hydrodynamic lubrication mechanisms and has been a governing equation in lubrication analyses. A number of books have been published on this subject, including, but not limited to, those by O. Pinkus and B. Sternlicht (1961), A. A. Cameron (1966), B. J. Hamrock (1994), Y. Hori (2002), and others.

In the middle of the last century, it was found that even in non-conformal contacts under heavy loading conditions, it might still be possible to form a lubricant film between the two surfaces, and the film formation mechanism could still be described by the Reynolds equation if surface elastic deformation and dependence of lubricant viscosity on pressure are considered in the analysis. This revolutionized lubrication theory is named the elastohydrodynamic lubrication (EHL), which has been receiving great attention since 1960s, and is now still an attractive and challenging area of research. A number of books were published on the EHL, including those by Dowson and Higginson (1966), Hamrock and Dowson (1981), and later by R. Gohar (2001).

Since the early 1990s, computer and information technologies have been explosively advancing. Today, a regular personal computer could be much more powerful than a mainframe supercomputer in 1970–1980s. The tremendous improvements in computational speed and data storage space enable researchers to conduct sophisticated modeling and numerical simulations that were inconceivable in the past. This has greatly accelerated theory technology development in all branches of science and engineering, including tribology. Perhaps the influences of advanced computer and numerical simulation technologies on research and development in the fields of surface contact and lubrication can be summarized mainly in the following three aspects:

1. The ways to model and solve practical contact-lubrication problems have been irreversibly changed. It is well known that, in fact, a vast majority of problems found in engineering reality are not analytically solvable. In the past, engineering analyses were based on assumptions of simple/artificial contact geometry, ideally smooth surfaces, rigid or perfectly elastic and homogeneous solid materials, and Newtonian fluids, and others, due to limited computational ability. There have been simplified analytical solutions developed, which might present some basic trends and rough estimates, but could omit details in many realistic aspects, such as the effects of non-Hertzian arbitrary contact geometry, inhomogeneous inelastic materials, non-Newtonian fluid properties, and rough surface topography on the interface behaviors under practical conditions.

   One of the key problems, for example, has long been how to handle surface roughness. Early studies either assumed ideally smooth surfaces or employed stochastic models based on simplified surface asperity geometry and height distribution. In the early to mid-1990s, advanced metrology technologies, such as optical profilometry, were developed, making accurate digitization of engineering surfaces fast and easy. High-speed computers and their large memory space enable deterministic solution methods for simulating sophisticated contact and lubrication problems with digitized machined topography. Now full-scale numerical simulations gradually become the mainstream in theoretical studies and problem solving, and efforts are currently being made to remove many simplifying assumptions, e.g. those for perfectly elastic and homogeneous materials stated above.
2. Advanced computer and numerical solution technologies have bridged contact and lubrication recently.

In the past, studies on contact mechanics were parallel to those on lubrication, due to their apparently different fundamental assumptions, formulations, and solution methods. In other words, contact mechanics did not consider possible lubrication, and lubrication theories completely ignored possible surface contact. In reality, however, most functional components usually operate in a mode called mixed lubrication, in which both surface asperity contacts and hydrodynamic lubricant films coexist, and neither can be ignored.

Actually, a "dry" contact is a special case of "lubricated" contact under extreme conditions, such as ultralow viscosity and density (e.g. those of air) and/or ultralow speed (e.g. zero speed). In principle, contact and lubrication are actually on the same physical ground, and there is no theoretical barrier in between, thus, we should be able to solve contact problems with lubrication models as long as surface deformation is considered in the analysis.

Efficient numerical algorithms have been developed and discretization meshes much densified since mid-1990s. Accordingly, unified model approaches for handling both hydrodynamic lubrication and surface contact simultaneously have been developed. One can now simulate the entire transition from full-film hydrodynamic lubrication to mixed lubrication and all the way down to the state of dry contact with a unified equation system and numerical solution procedure. The assumption of "no lubrication" (that leads to contact mechanics) or "no surface contact" (that has been the foundation of classical lubrication theories) can be removed.

3. Advanced computer technologies bring sophisticated analyses to design engineers. In the past, contact and lubrication models and design procedures were developed based on simplified analyses for people who were using manual calculation tools, such as slide rules, abacuses, and daisy wheel calculators, providing simplified analytical and empirical solutions, formulae, and datum charts. Some models and formulae of this type have been proven to be acceptable in some applications, but many others may be dissatisfactory to today's more demanding and challenging engineering requirements.

   With the assistance of advanced computers, it is now possible to make sophisticated analysis programs available to engineers, so that more precise numerical calculations and predictions can be conducted quickly by design engineers on personal computers. Many simplified methods, empirical formulae, and datum charts may soon become unnecessary or obsolete in the time of science-based artificial intelligence.

The developments and trends of research since mid-1990s have experienced irreversible changes of research methods and witnessed fusion of contact and lubrication theories in the studies of interfaces subjected to contact and relative motion. This book intends to review these changes and reflect the merging of these two branches of studies. Actually, new challenges and new frontiers are pushing for more in-depth field integrations, and rapid research developments and wide expansions of related fields are more than what we can cover in this book. We hope that our efforts presented can help promote further discussion and stimulate more indepth investigations.

It is important to note that "interface" is a general term that concerns the boundary between two different materials, which can be in any combination of solids, liquids, and gases. The research on interfacial phenomena is often multi-scale and interdisciplinary in nature. What we will focus on in this book is a special type of interfaces, called "tribological interface" that consists of two solid surfaces in contact with or without fluid(s) in between. This type of interface widely exists in nature and engineering reality, and extremely important in various scientific and industrial applications.

In this book, we will focus on basic concepts, mathematic models, numerical solution procedures, major results and their physical meanings, as well as engineering applications of modeled solutions. We hope engineering researchers, design engineers, and graduate and undergraduate students find this book helpful for their work and studies.

Q. Jane Wang and Dong Zhu
Mount Prospect, Illinois, USA

Acknowledgments

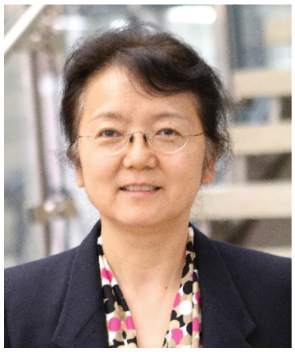
In our studies of interfacial mechanics, which have led to the writing of this book, we have collaborated with many research partners, colleagues, students, and friends, who have made invaluable contributions and provided insightful suggestions and assistance in different ways. It has been a great honor for us to work with the following people, listed in alphabetical order:

Ms. Jannat Ahmed, Northwestern Univ., USA
Dr. Xiaolan (X. L.) Ai, the Timken Co., USA
Dr. Girma Biresaw, US Dept. of Agriculture, USA
Mr. Michael P. Bujold, Eaton Corp., USA
Dr. Wayne W. Chen, Schlumberger Corp., USA
Prof. Herbert S. Cheng (Ph. D. thesis advisor of Q. Jane Wang), Northwestern Univ., USA
Prof. Qingbing (Q. B.) Dong, Chongqing Univ., China
Dr. Kyugo Hamai, Hitachi, Ltd., Japan
Prof. H. P. Evans, Cardiff Univ., UK
Dr. Tao He, Northwestern Univ., USA
Prof. Yuanzhong (Y. Z.) Hu, Tsinghua Univ., China
Prof. Xiaoqing (X. Q.) Jin, Chongqing Univ., China
Prof. Leon M. Keer, Northwestern Univ., USA
Mr. Donglong (D. L.) Li, Chongqing Univ., China
Mr. Bohdan Lisowsky, Eaton Corp., USA
Prof. Geng Liu, Northwestern Polytechnical Univ., China
Dr. Shuangbiao (S. B.) Liu, Caterpillar, Inc., USA
Dr. Yuchuan (Y. C.) Liu, General Motors, USA
Prof. Jianbin (J. B.) Luo, Tsinghua Univ., China
Prof. Ashlie Martini, Univ. of California, Merced, USA
Dr. Toshikazu Nanbu, Nissan Motor Co., Japan
Prof. Daniel Nélias, INSA Lyon, France
Dr. Ning Ren, Valvoline Inc., USA
Prof. Wei Pu, Sichuan Univ., China
Dr. Fanghui (F. H.) Shi, General Motors, USA
Dr. Xiujiang (X. J.) Shi, Harbin Engineering Univ., China
Mr. Linlin (L. L.) Sun, Northwestern Polytechnical Univ., China
Prof. Wenzhong (W. Z.) Wang, Beijing Institute of Technology, China
Prof. Xiaopeng (X. P.) Wang, Three Gorges Univ., China
Dr. Yansong (Y. S.) Wang, Baker Hughes, USA
Prof. Zhanjiang (Z. J.) Wang, Southwestern Jiaotong Univ., China
Prof. Shizhu (S. Z.) Wen (Ph. D. thesis advisor of Dong Zhu), Tsinghua Univ., China
Prof. Lechun (L. C.) Xie, Wuhan Univ. of Technology, China
Dr. Chengjiao (C. J.) Yu, Baker Hughes, USA
Dr. Mengqi (M. Q.) Zhang, Southwestern Jiaotong Univ., China
Dr. Xin Zhang, Northwestern Univ., USA
Prof. Kun Zhou, Nanyang Technological Univ., Singapore
Prof. Qinghua (Q. H.) Zhou, Sichuan Univ., China,
and many others.

We would also like to acknowledge the following publishing companies who have kindly granted copyright permissions for reproducing a number of figures and tables from the cited papers and books. The full information of each original paper or book has been given in the reference list.

AIP Publishing
American Chemical Society (ACS)
American Society of Mechanical Engineers (ASME)
CRC Press
Elsevier
John Wiley & Sons
Marcel Dekker
Royal Society, London
Royal Society of Chemistry
SAGE Publishing
Science China Press and Springer
Society of Tribologists and Lubrication Engineers (STLE)
Springer

Authors



Q. Jane Wang received her Ph.D. in mechanical engineering from Northwestern University, IL, USA, in 1993. She taught for 5 years at Florida International University, Miami, FL, USA, and is now a professor in the Mechanical Engineering Department at Northwestern University, USA. She was elected fellow of the American Society of Mechanical Engineers (ASME) in 2009 and Society of Tribologists and Lubrication Engineers (STLE) in 2007. Her research interests are mainly in the areas of contact and interfacial mechanics and tribology of advanced materials and novel lubricants.

Dong Zhu received his Ph.D. in mechanical engineering from Tsinghua University, China, in April of 1984. He started to work at the Center for Engineering Tribology, Northwestern University, USA, as a research fellow in January of 1986. He joined the Technical Center of Aluminum Company of America in the beginning of 1991 then Eaton Innovation Center in 1994, doing tribology and surface engineering related research and product development. After his retirement from Eaton, he was appointed to a professor position at Sichuan University, China, and adjunct professor at Tsinghua University. He is now an adjunct professor at Harbin Engineering University. He was elected fellow of the American Society of Mechanical Engineers (ASME) in 2007 and Society of Tribologists and Lubrication Engineers (STLE) in 2006. His research interests mainly include elastohydrodynamic lubrication (EHL), mixed lubrication, surface engineering, and tribological testing.

Nomenclature

Listed below are widely used symbols and abbreviations in this book.
Those used only in specific chapters/sections will be defined in the text.

a	semi-axis of the Hertzian contact ellipse in the x-direction, or radius of the Hertzian circle, or half-width of the Hertzian contact zone for a line contact
a_c	critical contact radius at the onset of plastic deformation
A_c	contact area ratio (total area of surface asperity contacts divided by the nominal contact area), or real surface contact area
b	semi-axis of Hertzian contact ellipse in the y-direction
c	radial clearance of a journal bearing
C_p	specific heat
C_{ijkl}	elastic stiffness tensor, $C_{ijkl} = \lambda\delta_{ij}\delta_{kl} + \mu\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right)$
D^*	$D^* = K/(\rho C_p)$, thermal diffusivity of material
e	eigenstrain
E	Young's (elastic) modulus
E_1, E_2	Young's (elastic) moduli of Body 1 and Body 2, respectively
E'	$E' = 2[(1 - \nu_1^2)/E_1 + (1 - \nu_2^2)/E_2]^{-1}$, effective Young's (elastic) modulus
E^*	$E^* = [(1 - \nu_1^2)/E_1 + (1 - \nu_2^2)/E_2]^{-1}$, equivalent Young's (elastic) modulus
E_T	tangent modulus for the linear hardening law
ET	$ET = E_T/E$, dimensionless tangent modulus for the linear hardening law
f	friction coefficient or coefficient of friction (COF)
f_b or f_c	boundary friction coefficient or contact friction coefficient
f_h	hydrodynamic friction coefficient
G^*	$G^* = \alpha E'$, dimensionless material parameter
G_s	$G_s = E/[2(1 + \nu)]$, elastic shear modulus
h	local film thickness (or gap)
H	$H = h/R_x$ or h/a, dimensionless film thickness
h_a	average film thickness (or average gap)
h_c	central film thickness
H_c	$H_c = h_c/R_x$, dimensionless central film thickness
h_{cs}	central film thickness from smooth surface solution
h_m	minimum film thickness
H_m	$H_m = h_m/R_x$, dimensionless minimum film thickness
J_1	$J_1 = \sigma_1 + \sigma_2 + \sigma_3$, the first stress invariant
J_2	$J_2 = S_{ij}S_{ij}/2$, the second stress invariant
k or K	$k = b/a$, ellipticity of the Hertzian contact zone
K	thermal conductivity of solid material
K_f	thermal conductivity of fluid lubricant
l_e	nominal or effective length of line contact or roller contact
p	pressure
p_c	critical maximum Hertzian pressure at the onset of plastic deformation
p_c	contact pressure
p_h	maximum Hertzian pressure
p_{max}	maximum pressure
q, q_x, q_y	tangential traction and its components in the x- and y-directions, respectively
R	radius (or effective radius) of curvature at contact
r_1, r_2	radii of curvature of Surfaces 1 and 2, respectively
R_a	mean surface roughness height
R_q	root mean square (RMS) roughness
R_x	$R_x = (1/r_{1x} + 1/r_{2x})^{-1}$ for point contact, or $R_x = (1/r_1 + 1/r_2)^{-1}$ for line contact, effective radius of curvature in the x–z plane
R_y	$R_y = (1/r_{1y} + 1/r_{2y})^{-1}$, effective radius of curvature in the y–z plane
S	$S = (u_2 - u_1)/u$, slide-to-roll ratio
S_{ij}	$S_{ij} = \sigma_{ij} - \frac{1}{3}\sigma_{kk}\delta_{ij}$, deviatoric stress tensor
S_y	initial yield strength of material
t	time
T	dimensionless time
T	temperature
T_1, T_2	temperature of Bodies 1 and 2 (or Surfaces 1 and 2), respectively
T_{s1}, T_{s2}	temperature of Surfaces 1 and 2, respectively
U or u	$U = (u_1 + u_2)/2$, rolling (entraining) velocity in the x-direction
U^*	$U^* = \eta_o U/(E'R_x)$, dimensionless speed parameter
u_1, u_2	velocities of Surfaces 1 and 2, respectively
U_1, U_2	velocities of Surfaces 1 and 2, respectively
u_x, u_y, u_z	displacements in the x-, y-, and z-directions, respectively
u_z^p, u_z^q	displacements in the z-direction due to normal pressure p and tangential traction q, respectively
u, v, w	velocity components in the x-, y-, and z-directions, respectively
v	surface deformation
v	$v = u_2 - u_1$, sliding velocity
V	$V = (v_1 + v_2)/2$, rolling (entraining) velocity in the y-direction
v_e	surface elastic deformation

v_p	surface plastic deformation
w or W	applied load for point contact
W	$W = w/l_e$, applied load per unit contact length for line contact
w	surface wear
W^*	$W^* = w/(E'R_x^2)$ for point contact, or $W^* = w/(E'R_x l_e)$ for line contact, dimensionless load parameter
W_c	contact load ratio (load supported by surface asperity contact divided by total load)
W_c	critical load at the onset of plastic deformation
W_x, W_y	load components in the x- and y-directions, respectively
W_x, W_y	wave lengths of sinusoidal roughness in the x- and y-directions, respectively
x, y, z	Cartesian coordinates (x is often selected to be parallel to the rolling direction)
X, Y, Z	$X = x/a$, $Y = y/b$, and $Z = x/a$, dimensionless x-, y- and z-coordinates, respectively

Greek Letters

α	pressure–viscosity exponent used in the pressure-viscosity equation (2.26)
γ	roughness orientation
δ_1, δ_2	roughness heights of Surfaces 1 and 2, respectively
δ_{ij}	the Kronecker Delta, $\delta_{ij} = 1$ when $i = j$, or $\delta_{ij} = 0$ if $i \neq j$
Δx, ΔX	$\Delta X = \Delta x/a$, mesh size and dimensionless mesh size in the x-direction
Δy, ΔY	$\Delta Y = \Delta y/b$, mesh size and dimensionless mesh size in the y-direction
ε	total strain
ε_{ij}	total strain components
ε_{ij}^e	elastic strain components
ε_{ij}^p	plastic strain components
ε^p	$\varepsilon^p = \sqrt{\frac{2}{3}\varepsilon_{ij}^p \varepsilon_{ij}^p}$, effective plastic strain
ε_p	prescribed small positive value for checking pressure convergence
ε_w	prescribed small positive value for checking load convergence
η	viscosity
η^*	equivalent (or effective) viscosity of non-Newtonian lubricant
η_o	viscosity under ambient condition
λ	$\lambda = h_a/\sigma$, film thickness ratio, or λ ratio, or specific film thickness
Λ	$\Lambda = h_{cs}/\sigma$, hydrodynamic roughness parameter, or nominal film thickness ratio
μ	material shear modulus
ν	Poisson's ratio
ν_1, ν_2	Poisson's ratios of Body 1 and Body 2, respectively
ρ	density of lubricant
ρ_1, ρ_2	densities of solid Bodies 1 and 2, respectively
ρ_o	density of lubricant under ambient condition
σ	$\sigma = \sqrt{R_{q1}^2 + R_{q2}^2}$, composite RMS roughness of the two surfaces
σ_1, σ_2, σ_3	principal stresses
σ_{ij}	stress components
σ_{ij}^e	elastic stress components
σ_{ij}^r	residual stress components
σ_{VM}	$\sigma_{VM} = \sqrt{\frac{3}{2}S_{ij}S_{ij}}$, equivalent von Mises stress
σ_Y	yield strength of material
σ_{Y0}	initial yield strength of material
τ	shear stress
ω	rotational velocity
ϕ_x, ϕ_y	pressure flow factors
ϕ_s	shear flow factor
Φ_s	shear flow factor for a single surface
ψ	dimensionless clearance of a journal bearing

Abbreviations

1B	first-order backward differential scheme
1D	one-dimensional
2B	second-order backward differential scheme
2C	second-order central differential scheme
2D	two-dimensional
3D	three-dimensional
BEM	boundary element method
CAD	computer aided design
CC-FT	continuous convolution and Fourier transform
CFD	computational fluid dynamics
CGM	conjugate gradient method
COF	coefficient of friction
DC-FFT	discrete convolution and fast Fourier transform
DC-CC-FFT	discrete-continuous convolution and FFT
DCD-FFT	DC-FFT with duplicated padding
DCR-FFT	discrete correlation and fast Fourier transform
DCS-FFT	DC-FFT with IC summation
DCSS-FFT	DC-FFT with IC double summations
DET	distortional energy theory
DFT	discrete Fourier transform
DS	direct summation
EHL	elastohydrodynamic lubrication
EIM	equivalent inclusion method
EPP	elastic-perfectly plastic
EPS	effective plastic strain
FDM	finite difference method
FEA	finite element analysis
FEM	finite element method
FFT	fast Fourier transform
FGM	functionally graded material

FRF	frequency response function
FT	Fourier transform
GF	Green's function
G–S	Gauss–Seidel
H–D	Hamrock and Dowson
IC	influence coefficient
IDFT	inverse discrete Fourier transform
IFFT	inverse fast Fourier transform
IFT	inverse Fourier transform
JFO	Jakobsson–Floberg–Olsson
MEMCT	mechanical–electrical–magnetic–chemical–thermal
MG	multi-grid
MLMI	multi-level multi-integration
MSST	maximum shear stress theory
PEHL	plasto-elastohydrodynamic lubrication
P–H	Pan and Hamrock
PMD	progressive mesh densification
RCF	rolling contact fatigue
RMS	root mean square
SAM	semi-analytical method
SOR	successive over-relaxation
TCM	traction cancellation method
TEHL	thermal elastohydrodynamic lubrication

1 Introduction

1.1 SIGNIFICANCE OF THE TOPICS

"Interface" is a general term that concerns a boundary between two different materials, which can possibly be in any combination of solid(s), liquid(s), and gas(es). Interfacial phenomena are complex in nature. Relevant studies on interfacial behaviors have been in a wide range from different aspects, often multi-scale and interdisciplinary that may involve many branches of science and engineering. Great efforts have been made in order to understand interfacial mechanics, physics, and chemistry in different fields. However, so far, there has not been a well-developed branch of science that covers all the different types of interfacial phenomena in a general sense. Interfacial mechanics, in fact, is still an evolving field of study generally in its infancy.

In this book, our discussion will focus on a special type of interface system, called "tribological interface", or "interface" in short, that consists of two solid body surfaces in contact and possible relative motion with or without fluid(s) in between. This specific type of interface system is widely seen in reality and extremely important in science and engineering practice. Figure 1.1 gives a sketch showing two solid surfaces in contact and relative motion, but, in reality, fluid(s) (often acts as lubricant), boundary films and a small quantity of debris/particles may also be observed in this interface system. For commonly used metallic materials, thin oxide films, as well as some other surface layers and coatings, may often exist in engineering reality. Because the boundary and oxide films are usually extremely thin, and the size and quantity of possible debris very small, they are often ignored in most interfacial analyses. Therefore, a basic model used in this book is constructed with two solid bodies having smooth or rough surfaces in contact and possible relative motion with or without fluid lubricant in between.

Power and motion are transmitted through interfaces, which are often lubricated in one way or another, at surface contact locations of various components that are basic elements of all kinds of vehicles, industrial machineries, engineering equipment, and scientific devices. As is well known, a large portion of energy produced globally is dissipated through frictional loss that occurs mainly due to the interaction of surfaces of machine components. Also, it has been found that, today, roughly more than 70%–80% of mechanical component failures take place or originate at surfaces due to severe contact, rubbing, and insufficient lubrication. Machine performance, loading capacity, efficiency, durability, and reliability appear to be macro-scale events that always need to be well controlled and optimized in engineering practice. However, they are dependent strongly upon micro- and nano-scale interfacial characteristics. A deep understanding of the nature of such interfacial mechanisms, therefore, is vital to components design and product development. In the 21st century, critical issues associated with energy shortage, environmental pollution, and global warming, as well as their impact to economic growth and stability, impose increasingly strong challenges to researchers and engineers, thus striving for better efficiency and reliability becomes more urgent than ever before.

In reality, solving interfacial problems may involve multi-scale multidisciplinary science and engineering, including continuum mechanics, rheology, mechanical design theory, materials science, thermal dynamics, physics, chemistry, and others. Associated studies may span from macro- to micro- all the way down to nano- and subnano-scales, as depicted in Figure 1.2.

For such complicated multi-scale problems, however, continuum mechanics-based contact and lubrication analyses are fundamental, providing basic information, such as interfacial pressure due to solid contact and/or hydrodynamics, lubricant film thickness or gap, surface deformation, friction/traction, temperature, and subsurface strain, and stress distributions, which are necessary for performance, efficiency, and failure predictions. Also, it serves as a foundation of further in-depth investigations in the areas of interfacial physics, chemistry, and materials failure mechanisms.

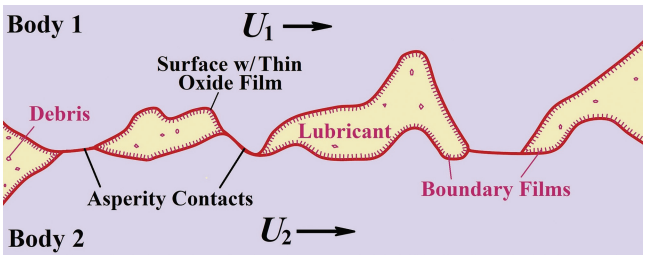


FIGURE 1.1 Schematic of a tribological interface system.

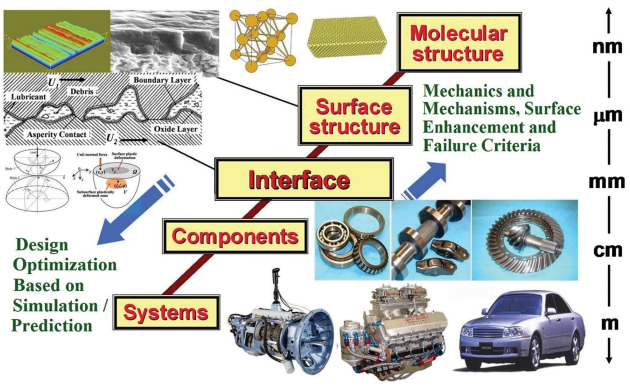


FIGURE 1.2 Multi-scale system problems.

It is well known that material failures found in mechanical components can generally be categorized into two types: structure (or bulk) failures and surface failures. For structure failures, predicting and preventing technologies have been much better developed. For example, not long ago, people still relied on photoelasticity, strain gauge experiments, and destructive tests to evaluate structure strengths. In the last few decades, as the FEM/CAD technologies have been well developed with commercial software packages readily available, system and component structure strength can now be quickly and accurately predicted by using computers. Product development cycles have been greatly shortened, and photoelasticity and destructive tests can be avoided in most cases. Today structure failures have been significantly reduced in engineering reality.

On the other hand, however, the majority of component failures fall in the surface failure category. These include sliding and fretting wear, scuffing, pitting due to contact fatigue, and others. Unfortunately, so far, interfacial sciences and problem-solving technologies are still mostly premature, and few commercial software packages are conveniently available for engineers to accurately evaluate surface strength of various components.

In fact, today modeling and analyzing contact and lubrication problems, predicting interface performance, and evaluating surface strength and life often appear to be a bottleneck in advanced product design and development. Moreover, critical problems found during laboratory and field tests, machine operations, and customer warranty services are usually associated with malfunction and failure of interface systems, statistically much more than any other problems. There are great challenges, together with vast opportunities, for further development in the field of interfacial sciences, and the importance of understanding these interfaces can never be overestimated.

1.2 TRIBOLOGICAL INTERFACE SYSTEMS

The interface in contact with relative motion is often referred to as a tribological interface. The word "tribology" originated from the Greek word, *tribos*, which means rubbing between two surfaces or simply refers to a science about "rubbing" (Halling, 1971). Tribology, in other words, is a branch of science that develops theories and technologies for improving functionality, efficiency, and durability of the surfaces in contact and relative motion.

It is important to note that a tribological interface usually involves two solid surfaces in contact with or without lubricating media in between that can possibly be liquid(s), solid(s), gas(es) or their combination(s). Also, the two solid bodies may have surface oxide layers, adsorbed layers, boundary layers, non-uniform layers due to various fabrication processes and surface treatments. Within the solid bodies, there may be impurities, particles, fibers, inclusions, defects, voids, and other types of inhomogeneity. Typically, ultrathin oxide and boundary layers may be ignored in most interfacial analyses, but lubricants are often included because of the significant roles they play. Surface layers/coatings and other types of inhomogeneity may be taken into account when needed.

In a tribological interface system, physical (including mechanical and thermal) and chemical phenomena may simultaneously occur. The subject of this book will be focused mainly on mechanical properties, but attention will also be paid to thermal behaviors, because there is often a strong mutual dependence between the mechanical and thermal characteristics. The tribological interface systems to be discussed can primarily be categorized based on geometry, relative motion, lubrication status, and materials.

1.2.1 Interface Systems Defined Based on Geometry

An interface system may be subjected to external loading and relative motion. The contact geometry at the interface can be categorized generally into two types as "conformal contact" and "non-conformal contact" (also called "counterformal contact" or "concentrated contact"). In a conformal contact, the two surfaces have their curvatures closely matching each other so that the area of surface interaction is large, typically comparable to the dimensions of the contacting bodies (e.g. machine elements). Conformal-contact components include journal and thrust bearings, washers, piston skirt/ring/cylinder liner systems, many types of joints and seals, clutches, brakes, and many others.

A non-conformal contact, on the contrary, is formed by two surfaces whose curvatures do not closely match. As a result, the interaction area is usually small in both principal directions (called "point contact"), or at least in one direction ("line contact"), often resulting in a localized high pressure concentration. The size (or width) of the contact zone is usually much smaller than the dimensions of the contacting bodies and the radii of curvature of the surfaces at the contact location. Non-conformal contact can be found in various gears, rolling element bearings, cam/follower systems, ball screws, some pumps and metal-forming tools, traction drives and continuously variable transmissions, and so on.

Figure 1.3 shows some examples of conformal contact components commonly seen in engineering practice, and Figure 1.4 presents those of non-conformal contacts.

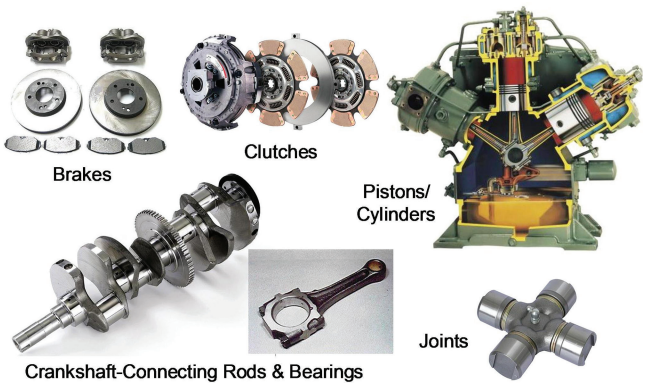


FIGURE 1.3 Examples of conformal contact components.

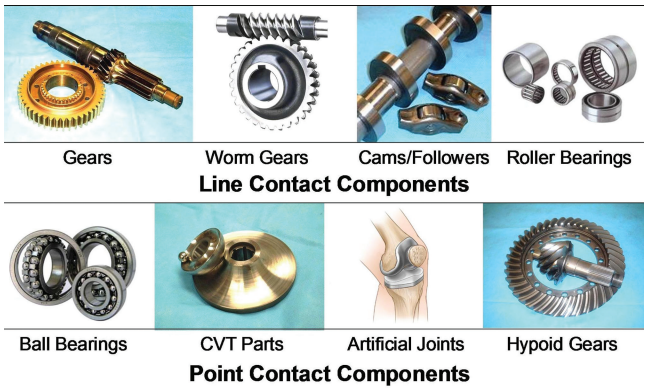


FIGURE 1.4 Examples of non-conformal contact components.

It is important to note that for conformal contact problems the contact geometry, the resultant area of surface interaction, and tribological behaviors depend largely upon the macro structure deformation of specific components and their operating conditions. The surface strength, therefore, often needs to be analyzed together with the structure strength (unless the macro structure is assumed to be rigid and the body deformation is not considered). Since there is a large variety of component structure design, conformal contact problems often have to be solved case by case through a combination of the interface and structure analyses, typically by using the finite element method (FEM) for the latter. Analyzed examples will be presented in Chapter 6.

For counterformal contact problems, on the other hand, the situation is different. Usually the size of surface interaction zone in a counterformal contact is much smaller (or narrower) than the dimensions of the contacting bodies and the radii of curvature of the surfaces at the contact. The pressure and stresses concentrated within or around the small contact zone are typically much higher than other stresses distributed in the macro structure. It is justified, therefore, to isolate the localized interfacial problem within the vicinity of the small contact zone and ignore the influences of the body structure and macro-scale operating conditions. That is why the counterformal contact problems can be simplified and solved locally by considering the contacting bodies as infinitely large half-space solids.

Since in a counterformal contact problem the contact zone is so small or narrow that the complicated body structure is not involved, the local interface geometry can often be simplified to either a line or point contact with satisfactory accuracy. An example has been illustrated in Figure 1.5 for a pair of spur gear teeth. At the contact point, the radii of curvature of pinion and gear due to the involute profiles, r_1 and r_2, can be readily calculated. Because the width of surface interaction zone is much smaller than the length, the contact geometry can be simplified to two infinitely long parallel cylindrical bodies against each other under a uniformly distributed normal load, w, as depicted in the figure.

For the two parallel cylinders, the gap between the two surfaces before deformation can be calculated as follows (refer to Figure 1.6):

$$h(x) = h_o + \left(r_1 - \sqrt{r_1^2 - x^2}\right) + \left(r_2 - \sqrt{r_2^2 - x^2}\right)$$

$$= h_o + r_1\left[1 - \sqrt{1 - \left(\frac{x^2}{r_1^2}\right)}\right] + r_2\left[1 - \sqrt{1 - \left(\frac{x^2}{r_2^2}\right)}\right] \tag{1.1}$$

in which h_o stands for the separation between the two bodies before deformation is considered. $h_o = 0$ means that the two cylinders are in touch under a zero load (so there is no deformation), and usually, h_o is a small negative value if deformation due to a normal load occurs.

Equation (1.1) is sufficiently simple according to today's computational ability. This type of contact geometry without further simplification has been used in some studies, e.g. the work by Zhu et al. (2012), in order to accurately handle complex geometric details when necessary. However, since the width of possible surface interaction area is usually much smaller than the radii of curvature of the contacting bodies, r_1 and r_2, traditionally the geometry in the vicinity of contact zone can be further simplified by means of Taylor's expansion series:

$$h(x) = h_o + r_1\left[1 - \left(1 - \frac{x^2}{2r_1^2} - \frac{x^4}{8r_1^4} - \frac{x^6}{16r_1^6}\cdots\right)\right]$$

$$+ r_2\left[1 - \left(1 - \frac{x^2}{2r_2^2} - \frac{x^4}{8r_2^4} - \frac{x^6}{16r_2^6}\cdots\right)\right]$$

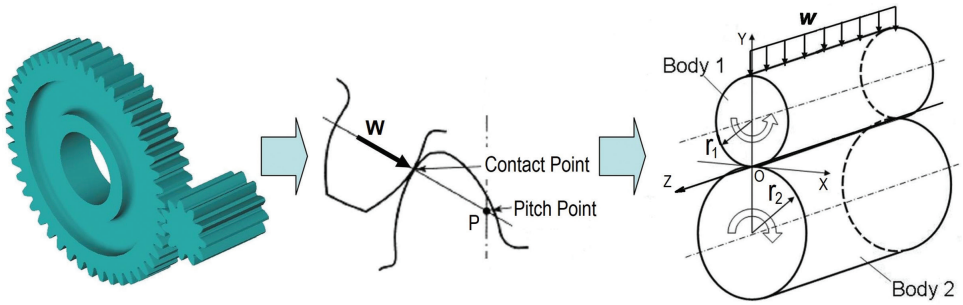


FIGURE 1.5 Typical line contact between spur gear teeth.

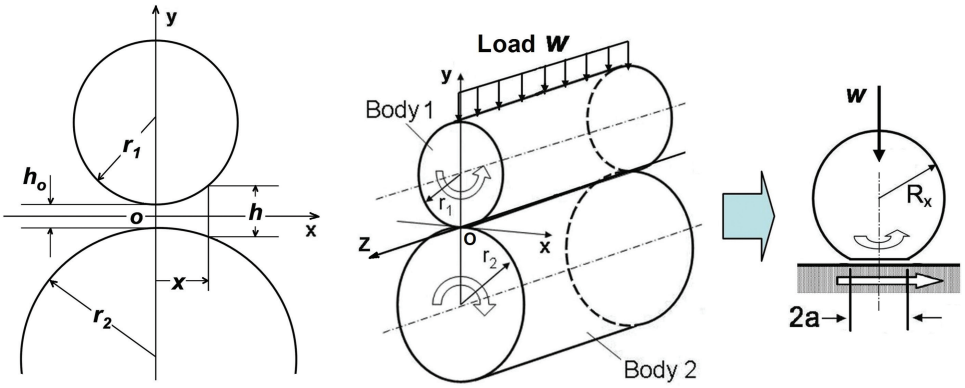


FIGURE 1.6 Model for line contact geometry.

Considering $x \ll r_1$ and $x \ll r_2$, we can simplify the above expression by neglecting the high orders of infinitesimal and then get

$$h(x) \approx h_o + \frac{x^2}{2r_1} + \frac{x^2}{2r_2}$$

that is

$$h(x) \approx h_o + \frac{x^2}{2R_x} \tag{1.2}$$

where

$$\frac{1}{R_x} = \frac{1}{r_1} + \frac{1}{r_2} \tag{1.3.1}$$

$$R_x = \frac{r_1 r_2}{r_1 + r_2} \tag{1.3.2}$$

R_x is defined as the effective (or equivalent, or reduced) radius of curvature at the contact location. Based on this mathematic simplification, the line contact between the two parallel cylindrical bodies can be considered equivalently as a cylinder of radius R_x against a flat half-space, as shown by the graph on the right-hand side of Figure 1.6. This simplification is satisfactory when the deformation is limited and the width of contact zone is small.

A typical example of point contact is illustrated in Figure 1.7 that is from a ball bearing. The ball has a constant radius in all directions, but at contact point *A*, the inner race has a positive curvature in the circumferential direction and a negative one in the axial direction. Both radii of curvature do not match that of the ball, so that a small elliptical contact zone is formed under a normal load, as illustrated in Figure 1.7. In the same way, a small elliptical contact area *B* can be found between the ball and the outer race, but the radii of curvature of the outer race are negative in both circumferential and axial directions.

Figure 1.8 shows a model for a general point contact, where r_{1x}, r_{1y}, r_{2x}, and r_{2y} are radii of curvature of Bodies 1 and 2 at the contact point in the *x*- and *y*-directions, respectively. Origin of coordinates *O* is chosen to be the initial contact point between the two bodies at the very onset of the deformation, and the *o*–*x*–*y* plane is tangential to both surfaces in contact. Note that curvature is defined as positive if the profile is convex (that means the curvature center is located inside the body) or negative when the profile is concave (the curvature center is outside the body).

Generally, in the vicinity of small contact zone, the geometry formed by two curved surfaces can be approximated by two ellipsoids against each other, and the gap in between can be expressed as follows if deformation is not considered.

$$h(x,y) = h_o + \left(r_{1x} - \sqrt{D_1{}^2 - x^2}\right) + \left(r_{2x} - \sqrt{D_2{}^2 - x^2}\right) \tag{1.4}$$

where

$$D_1 = \sqrt{r_{1y}{}^2 - y^2} - \left(r_{1y} - r_{1x}\right)$$

$$D_2 = \sqrt{r_{2y}{}^2 - y^2} - \left(r_{2y} - r_{2x}\right)$$

Under the assumption that the two surfaces are ideally smooth having continuous first- and second-order derivatives in both

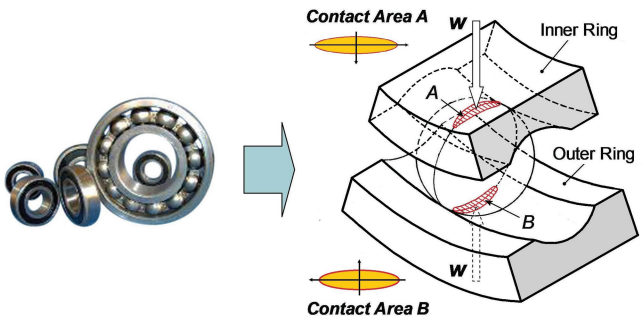


FIGURE 1.7 Typical point contact in a ball bearing.

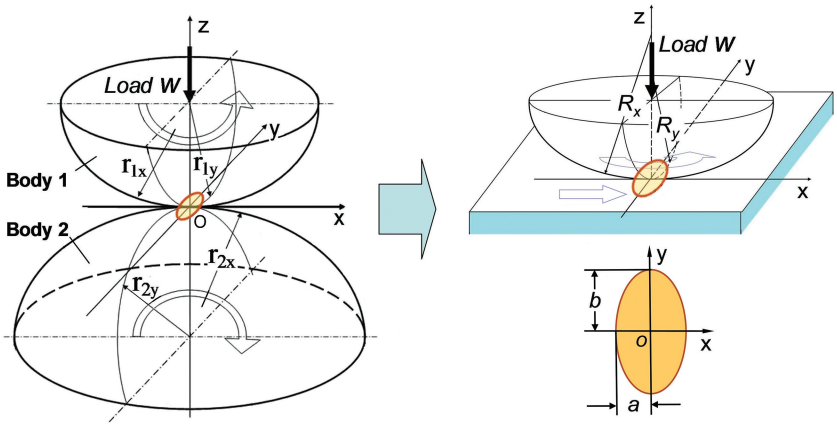


FIGURE 1.8 Model for general point contact geometry.

the x- and y-directions, the expression of gap can be simplified by means of Taylor's expansion series:

$$h(x,y) = h_o + z_1 + z_2 = h_o + \left(A_1x^2 + B_1y^2 + C_1xy + \cdots\right) + \left(A_2x^2 + B_2y^2 + C_2xy + \cdots\right)$$

in which A_i, B_i, C_i (i = 1, 2) are constants. Properly adjusting the x- and y-axes to be parallel to the principal directions of the contact zone, we can always make C_1xy and C_2xy disappear. The above expression can then be reduced if the terms of high-order infinitesimal are ignored.

$$h(x,y) = h_o + z_1 + z_2 \approx h_o + \left(A_1x^2 + B_1y^2\right) + \left(A_2x^2 + B_2y^2\right) = h_o + \frac{x^2}{2r_{1x}} + \frac{y^2}{2r_{1y}} + \frac{x^2}{2r_{2x}} + \frac{y^2}{2r_{2y}}$$

$$\text{i.e. } h(x,y) \approx h_o + Ax^2 + By^2 = h_o + \frac{x^2}{2R_x} + \frac{y^2}{2R_y} \tag{1.5}$$

where

$$\frac{1}{R_x} = \frac{1}{r_{1x}} + \frac{1}{r_{2x}} \tag{1.5.1}$$

$$\frac{1}{R_y} = \frac{1}{r_{1y}} + \frac{1}{r_{2y}} \tag{1.5.2}$$

$$R_x = \frac{r_{1x}r_{2x}}{r_{1x} + r_{2x}} \tag{1.5.3}$$

$$R_y = \frac{r_{1y}r_{2y}}{r_{1y} + r_{2y}} \tag{1.5.4}$$

in which R_x and R_y are the effective radii of curvature in the x- and y-directions, respectively.

It is obvious that the two ellipsoids in contact can be considered equivalently as one ellipsoid with the radii of curvature of R_x and R_y against a flat half-space, as shown in Figure 1.8. Also, It is apparent that, according to Equation (1.5), contour line for a constant original gap, $h(x, y)$, should be an ellipse. It will be proven later that after elastic deformation, the contact zone with a zero gap will also be an ellipse. That is why point contact is generally called "elliptical contact". A typical elliptical contact zone is illustrated in Figure 1.8. When $R_x = R_y$, the contour lines for constant gaps and the border of contact zone become round circles, so it can be called "circular contact".

If one of the principal radii of curvature, e.g. R_y, approaches infinity, or $R_y \gg R_x$, the contour lines approach straight lines parallel to the y-axis, and the contact zone becomes close to a straight uniform narrow band. This indeed turns to be a line contact. Therefore, line contact is actually a special case of point contact when one principal radius of curvature is infinity or much greater than the other.

It is important to note that, in some cases, the situation may be more complicated and the geometry in the vicinity of the surface interaction zone sometimes could not be properly simplified to either 2D line contact or 3D elliptical contact. Care must be taken when modeling a realistic complex problem. An example is shown in Figure 1.9. As it is well known, many line contact components, such as spur gear teeth, bearing rollers, and cams, are often purposely designed with crowning in the axial direction in order to accommodate possible fabrication errors, misalignment and unevenly distributed load. Also, the contact zone is always finite in length, and at its two ends, there are usually round corners or chamfers to reduce possible stress concentrations. This type of realistic geometry can often be modeled with a crowned finite cylindrical roller with round corners (or chamfers) at the two ends against a flat half-space, as depicted in Figure 1.9.

There may be different ways to model this type of interface geometry. First, if the crowning radius, R_y, is much greater than the nominal roller radius, R_x, we can consider R_y as infinitely large and simplify the geometry to that of a two-dimensional (2D) line contact. When R_y is not sufficiently large, the problem may be handled as an elliptical contact. Both line contact and elliptical contact problems are relatively

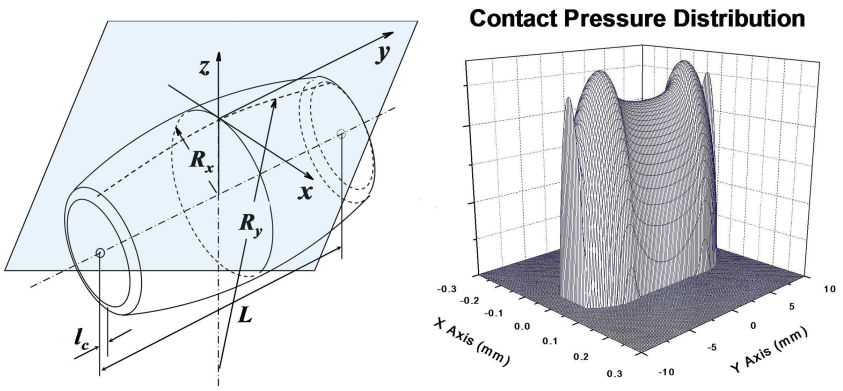


FIGURE 1.9 A crowned finite roller against a flat half-space. (Right graph is from Hu et al., 1999.)

simple that can be readily solved by the Hertzian theory to be discussed in Chapter 3. However, if the load is so heavy that the stress concentrations due to round corners of small radius and/or sharp edges at the two ends of the roller become significant, we must adopt a more complicated 3D model shown as follows in order to predict the interfacial behavior more precisely.

$$h(x,y) = h_o + R_x - \sqrt{D^2 - x^2} \tag{1.6}$$

where

$$D = \sqrt{R_y^{\,2} - y^2} - \left(R_y - R_x\right), \quad \text{if } |y| \le \frac{L}{2} - l_c$$

or $D = R_c + \sqrt{r^2 - d^2}$

$$d = |y| - y_c$$

$$y_c = \left(\frac{L}{2} - l_c\right)\left(1 - \frac{r}{R_y}\right)$$

$$R_c = \sqrt{\left(R_y - r\right)^2 - y_c^{\,2}} - \left(R_y - R_x\right), \quad \text{if } \frac{L}{2} - l_c < |y| \le \frac{L}{2}$$

In Equation (1.6), L is the nominal length of roller, l_c the width and r the radius of the round corner. It is difficult to solve such crowned finite roller contact problem analytically, but with an advanced numerical solution algorithm, this may be an easy task. See a sample solution of contact pressure distribution in Figure 1.9, obtained from a numerical approach presented by Hu et al. (1999). Numerical examples for lubricated contacts can be found in Section 8.7.

Generally, analytical solution methods depend greatly upon the specific geometry, material properties, and operating conditions, and most engineering problems are not analytically solvable. Numerical solution methods, on the other hand, are universally applicable for various types of geometry under different conditions, which will be discussed later in this book.

1.2.2 Interface Systems Defined Based on Relative Motion

Relative motion often occurs between a pair of contacting surfaces which usually result in normal pressure and tangential friction in the contact zone. Actually, tendency of relative motion also causes friction even though macroscopically the surfaces are relatively stationary. Friction due to the former is called kinetic friction while that due to the latter is static friction. Accordingly, the interfaces are the sliding and static interfaces, respectively.

Generally, if the coordinate origin is fixed to the center of the surface interaction zone, and the surface velocity of Body 1, relative to the coordinate system, is u_1 and that of Body 2 is u_2, as shown in Figure 1.10, the sliding speed is defined as the difference between u_1 and u_2, expressed as

$$V = u_2 - u_1 \tag{1.7.1}$$

or in vector form

$$\vec{V} = \vec{u}_2 - \vec{u}_1 \tag{1.7.2}$$

In the case of a counterformal interface, a type of motion called "rolling" is commonly seen, which means one surface "rolls" over the other with no (or little) relative sliding. The rolling motion can be found in many mechanical systems, such as wheel/rail contact and ball/race interface in a bearing. When the coordinate system is fixed to the contact center between the two surfaces, the rolling velocity is defined as

$$u = \left(u_1 + u_2\right)/2 \tag{1.8.1}$$

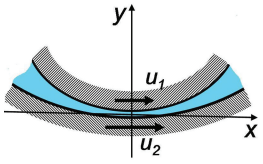


FIGURE 1.10 Relative motion between two surfaces in contact.

$$\text{or } \vec{u} = \left(\vec{u}_2 + \vec{u}_1\right)/2 \tag{1.8.2}$$

In a pure rolling interface system, both surface velocities, $\vec{u}_1$ and $\vec{u}_2$, are the same; hence, there is no sliding motion, $\vec{V} = 0$. In many engineering applications, however, rolling and sliding coexist, $\vec{u}_1 \neq \vec{u}_2$ and $\vec{V} \neq 0$. This is called "rolling-sliding interface" that can be found in various gears and many other components. Since high rolling speed is often favorable for hydrodynamic lubrication formation and surface failure prevention, but large sliding may have negative influences, it is often needed to look at the relative importance between the two. A parameter called "slide-to-roll ratio" is defined as follows for this purpose.

$$S = V/u = 2\left(u_2 - u_1\right)/\left(u_2 + u_1\right) \tag{1.9}$$

$S = 0$ stands for pure rolling, while $S = \infty$ or $-\infty$ means pure sliding with $u_2 = -u_1$ ($u = 0$). If one of the surfaces is stationary, $u_1 = 0$ or $u_2 = 0$ (i.e. $S = 2$ or -2), it is called simple sliding.

A more complicated situation can also be seen in engineering reality, in which the surface velocity vectors may have not only their tangential components but normal components as well, which refer to "squeeze motion" between the two surfaces. A positive squeeze tends to reduce the gap that may also be favorable for the hydrodynamic pressure generation and lubrication formation, which can be found in gears and cam/follower systems if dynamic effect is considered.

Sometimes the relative motion between the two bodies may have an angular velocity component, which is typically around an axis perpendicular to the tangential plane at the contact location. This is usually called "spinning" motion, which, to a various extent, widely exists in some machine elements.

When two surfaces are in contact and subjected to both normal and tangential loads, stick and slip zones may coexist, which means a "partial slip" contact interface. Strictly speaking, many static interfaces are actually with partial slip. Even at a static interface in a bolted connection, partial slip may still take place in an asperity contact level (see Z. J. Wang et al., 2013a). Figure 1.11 schematically shows a partial slip contact of an elastoplastic ellipsoid with an elastic half-space, where tangential load is smaller than the static friction. In the stick region, shear traction is below a limiting value determined by the static friction, and the slip distance between the two mating points are equal to zero (see Johnson, 1987). The direction of the shear traction is always opposite to the direction of slip tendency. Because no macroscopic lateral body motion is accomplished in partial slip, the static friction coefficient is applicable to this model.

When two bodies are in a pure rolling contact with friction and transmit tangential force to each other, the surface elastic deformations at the interface between the two contacting surfaces usually do not match, and there exist a stick area, where the two mated points share the same speed, and a slip area, where the two mated points are subjected to different velocities (see Z. J. Wang et al., 2012a). The unmatched deformations will lead to the velocity difference between the two bodies, which is called "creep".

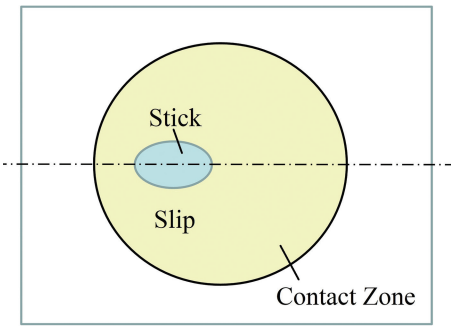


FIGURE 1.11 A partial slip interface, where stick and slip zones coexist.

In addition, there is a specific type of reciprocating motion of usually small amplitude at the interface called "fretting", which is commonly observed in many engineering applications, such as bolted, riveted, tight fitted, and some electrically engaged connections under vibration condition. Amplitude of fretting motion is often tiny and, sometimes, even invisible macroscopically, but it may result in significant wear due to fatigue and fracture, as well as surface property degradation.

1.2.3 Interface Systems Defined Based on Lubricating Media

Interfaces may be subject to different lubricating media, and most commonly used lubricants are liquids and low friction solids, but sometimes gases. Magnetic field, electric and electromagnetic levitation can also be used to support the load and separate the surfaces in order to reduce friction and prevent the surfaces from failures. Both liquid and gas lubrication are called fluid lubrication, in which the fluid flows around the interface are governed by fluid mechanics. Compared with a liquid, the viscosity and density of a gas are much lower, but its compressibility is much higher. Solid lubrication, on the other hand, is different, for which either lubricious coating(s) or solid powder(s) is often employed.

There have been extensive investigations of different lubricious solid materials that can substantially reduce friction. Molybdenum disulfide (MoS_2), graphite, polytetrafluoroethylene (PTFE), and diamond-like carbon (DLC) coatings are among the most commonly used solid lubricants. In this book, however, we will focus on fluid lubrication whose functionality is dominated mainly by continuum mechanics. Detailed discussion on solid lubricants is largely related to materials science and chemistry and is beyond the scope of this book.

A magnetic bearing utilizes magnetic suspension as the load-supporting medium to avoid direct surface contact. The magnetic field allows relative motion in the bearing

components under an extremely low friction and theoretically zero mechanical wear. Permanent magnets and electromagnets are both feasible for magnetic bearing design. However, the latter provides a convenience of better performance control. Magnetically levitated (MAGLEV) train is currently being commercialized that has great potential especially for the future high-speed transportation, using magnets to replace the conventional wheels, axles, bearings, and rails, as shown in Figure 1.12. The magnetic levitation lifts the vehicle from the guide ways with a small clearance and permits extremely smooth high-speed motion.

Electrical contacts widely exist in relays, switches, brushes, and actuators commonly seen in electrical equipment, appliances, and locomotives. They are also in mechanical systems that are subject to electrical current, such as bearings in a generator or motor. Electrical contact interfaces may involve different relative motions, such as impact, sliding, rotation, and vibration. Like other mechanical interfaces, generally the surfaces for electrical contact are not ideally smooth, which makes the interface imperfect, and the current actually passes through the spots of surface asperity contacts, whose total area is typically only a small portion of the nominal contact area. The gap at an electrical interface may change due to its physical function, which may induce arcing. At such an interface, a conductive fluid may be used to enhance interaction and reduce arcing.

A sketch of an electrical interface is given in Figure 1.13. At the interface, the materials may be subject to electrical, mechanical, and thermal loading. As a result, the material may experience melting, arcing, shearing, thermal degradation, and deformations, which could be plastic and/or elastic at the locations of different depths.

The interface of the armature and the launching rail of a rail gun is a good example of sliding electromagnetic contact (see Figure 1.14). The rail gun consists of two parallel rails and an armature that carries a projectile mounted between the rails. Electrical current, applied at the end of the upper rail, flows through the rail, passes through the armature, and then returns via the bottom rail. The high current flowing through rails and the armature creates strong magnetic fields in the region between the rails, according to the Biot–Savart law and the Ampere law. The interaction between magnetic field and the electric field due to current flow through the armature generates the Lorentz force, which corresponds to the source of armature motion (see Kim, 2006). In the configuration of the rail gun illustrated here, the applied current is to go through sliding imperfect contact regions. These are indicated by two gray areas between the rails and the armature that compose a closed circuit and generate a strong magnetic field.

Figure 1.15 shows how the interfacial friction coefficient changes depending mainly on the lubricating medium. This figure gives only approximate ranges as a reference in common engineering practice. It is not for accurate evaluation/prediction in any specific case. Friction is a very complex phenomenon, and it is not difficult to find exceptional cases.

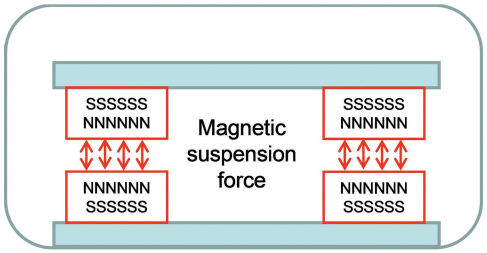


FIGURE 1.12 Schematic of the MAGLEV by magnetic levitation.

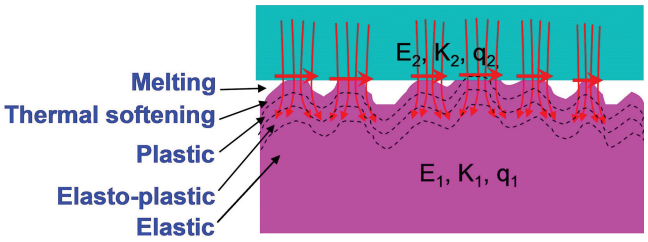


FIGURE 1.13 An imperfect electrical or electromagnetic contact. *E*, *K*, and *q* are elastic modulus, thermal diffusivity, and heat flux, respectively.

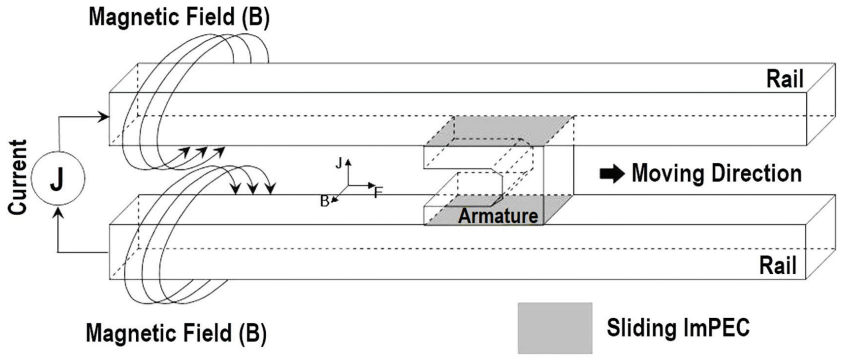


FIGURE 1.14 Sketch of electromagnetic contacts in a rail gun.

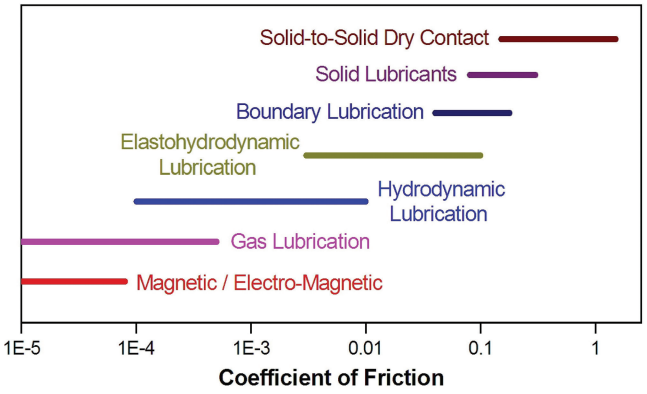


FIGURE 1.15 Approximate ranges of friction coefficient.

1.2.4 Interface Systems Defined Based on Lubrication Status

It is well understood that in most cases, the best way to reduce friction and minimize risk of surface failure is lubrication, adding thin layer(s) of fluid and/or solid lubricant(s) to separate the surfaces and avoid direct surface contact whenever possible. In industries, liquid lubricants, such as mineral and synthetic oils, are widely used, as they are effective, inexpensive, and abundantly available in the market. For various reasons, however, a complete separation of the contacting surfaces is often difficult or even impossible in many cases. Judging by the degree of surface separation (or lubricant film thickness) and the load-supporting mechanism, we can categorize tribological interfaces into the following three basic modes:

1. **Full-film lubrication** for a complete (or nearly complete) separation of the surfaces with a fluid-lubricant film that supports the entire (or nearly entire) load. It can be achieved either by applying an external static pressure to support the load and separate the surfaces (called "hydrostatic lubrication") or by hydrodynamic action due to specific gap geometry and motion of the two surfaces (called "hydrodynamic lubrication"). Rolling and squeezing motion usually help the hydrodynamic pressure generation that may result in the complete separation of the surfaces if the interfacial geometry and operating conditions are properly arranged and the surface roughness is small.
2. **Mixed lubrication** in which both surface separation by the lubricant film(s) and direct surface contact coexist in one interface system, and the load is shared between the two. In reality, no surface is ideally smooth, and the surface roughness is often of the same order of magnitude as, or even greater than, the average thickness of possible lubricant film, thus, surface asperity contact is sometimes difficult to avoid. In fact, most functional components operate in the mixed lubrication regime.
3. **Dry contact or boundary lubrication** in which hydrodynamic influence due to fluid lubricant is none or negligibly weak, and the load is fully or mainly supported by the surface contact. Note that there may be nanometer-scale thin boundary or adsorbed layers covering the surfaces. It is called boundary lubrication when the boundary film does exist and play a favorable role for friction reduction and surface protection. (This will be discussed in more detail in Section 10.3.) Because boundary films are usually ultrathin, their influences on contact pressure and surface-subsurface stresses are negligibly small, thus, they are often ignored when analyzing the interface by means of continuum mechanics. In this case, the boundary film-lubricated interface is treated in the same way as that of a dry contact. In other words, surface contact, in the sense of mechanics, does not necessarily mean a mathematical zero gap. It is defined as the state in which no hydrodynamic flow is possible or worth consideration due to the sufficiently small gap between the two surfaces.

Figure 1.16 shows the basic modes described above and the sketch for a typical interface system involving both lubricant films and surface asperity contacts. Generally, the contacting surfaces of Body 1 and Body 2 may be covered by ultrathin oxide layers that are common for many metals and alloys. On top of the oxide layer there may be a boundary layer possibly consisting of a few (or several) layers of molecules typically from either the lubricant or additives blended in the lubricant. These molecules may be orderly arranged and adsorbed to the surfaces, in comparison with the moving lubricant molecules randomly distributed in the gap and sheared due to the relative motion.

The gap between the two surfaces is often filled with a fluid lubricant, whose rheological behavior could be another complicated research subject. Various types of contaminates, wear debris and other foreign particles may also more or less exist in the system. However, because the fundamental studies of interfacial phenomena are based on the

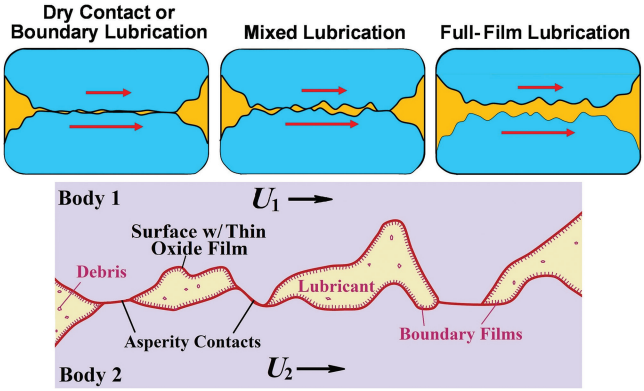


FIGURE 1.16 Three basic modes of interfacial lubrication.

continuum mechanics, ultrathin oxide, and boundary layers and tiny particles in small quantity are usually ignored, and the solid bodies and lubricants are often assumed to be homogeneous continua, unless otherwise specifically noticed.

In fact, dry contact is nothing but a special case of lubricated contact under some extreme conditions (such as ultralow lubricant viscosity and density). Sometimes when the surface velocities, U_1 and U_2, as illustrated in Figure 1.16, are sufficiently low or zero, the fluid lubrication can be ignored and the surface interaction can be approximated as a "dry" contact due to vanishing hydrodynamic action.

Theoretically, therefore, both lubricated and dry contact problems can be physically and mathematically modeled in the same way with a unified equation system. This gives a primary reason why we believe that contact and lubrication can be unified and described in the same book. However, while solving contact and lubrication problems analytically or numerically, technical treatments may still be different, as dry contact problems are usually simpler due to omission of lubricant and its rheology in the model approach, if otherwise under exactly the same conditions. Detailed explanations will be given later in the relevant chapters.

1.3 BRIEF HISTORIC REVIEW

The purpose of this section is not to present a comprehensive review of all the previous studies in the fields of contact and lubrication. Readers may refer to the book by D. Dowson (1998) and some other literatures for a more complete coverage of the development history if needed. Instead, this section only intends to outline the basic stages of the science and technology development in this field, in order to better understand their relation with other branches of science and engineering. This will be beneficial for a better comprehension of basic theories and practices and will also help gain a strategic vision into the future.

1.3.1 Empirical Knowledge Accumulated in Early Years

Generally, the development in the field of tribology has corresponded to that of human history and civilization. Perhaps, primitives started to face contact and lubrication problems as soon as they stood up and began to use tools. It has been discovered that during the palaeolithic period, hundreds of thousands of years ago, utilization of friction by rubbing wood on wood for making fire completely changed the way of the human life, and hand tools/weapons for breaking and changing shapes of other objects were widely in use based on a preliminary understanding that breakage, deformation, and material removal could be made if a heavy force is concentrated on a tiny area and, especially, with relative motion and a large difference in hardness between the tool/weapon and the object. No one knows exactly when people initially gained the experience that rolling friction is usually much lower than that of sliding, and lubrication with animal greases or plant liquids could dramatically reduce resistance while moving heavy objects.

Some researchers suggest that simple hand-held bearings made of stone or bone for the spindles of drills were developed in the Stone Age period. Firm evidences have shown that applications of potter's wheels and wheeled vehicles with bearings were started about 3,000–7,000 years ago in ancient Sumer, Egypt, China, and Indus Valley. The use of lubricants was recorded almost as early as the written human history started, e.g. about 2500–3000 BC when the pyramids were built. An example of liquid lubricant application in ancient Egypt is shown in Figure 1.17. During the early civilizations, preliminary knowledge on contact and lubrication was gradually accumulated from and contributed back to human activities and daily lives, but was purely empirical based on experience gained after countless failures.

Roughly starting from 1000 BC or so, more and more mechanisms, machineries, and transportation vehicles were developed. Early examples reported in history include lifting

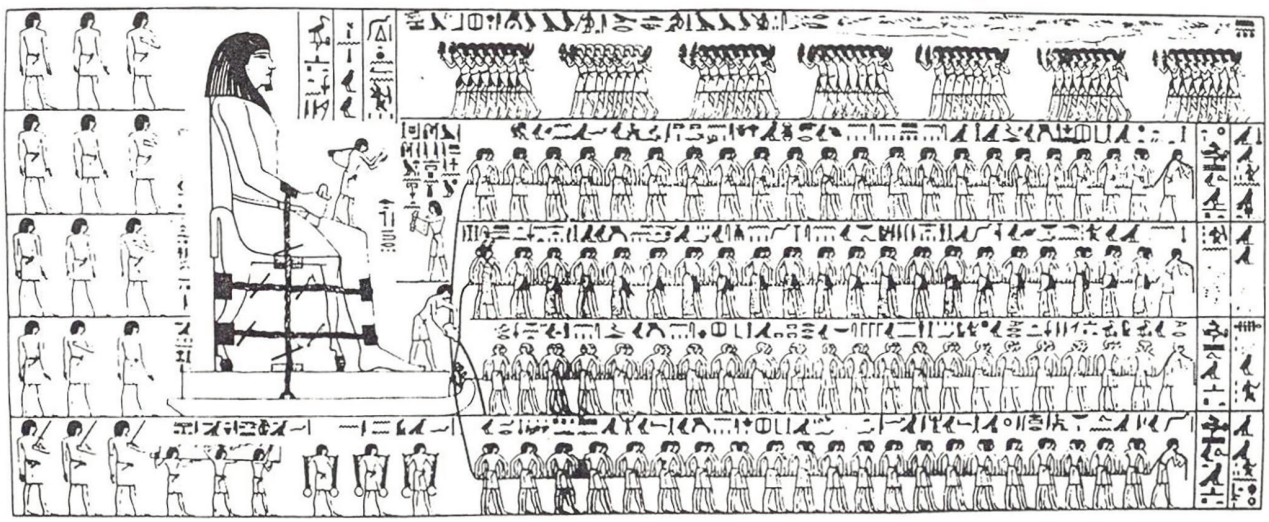


FIGURE 1.17 Transporting an Egyptian Colossus assisted with liquid lubrication (1900 BC).

pulleys (800 BC, Assyrian), tapered bronze bearing with leather seals on vehicles (300 BC, China), rotary querns or mills (about 150 BC, China and Mediterranean), lathes (25 BC, Rome), bronze ball and roller bearings (44–54 AD, Rome), pivot bearing systems for model temple and aeolipile (60s AD, Greece), the differential gearing system for a south-pointing chariot (255 AD, China), and more.

Later on, during 400–1400 AD (the medieval period in the west), significant engineering developments were reported and various machineries gradually became widespread. Typical examples include wheelbarrow and windmills (230–500 AD, China), water clocks (around 500 AD, Greece, Rome, and Middle East), watermills (100–1000 AD, China and Europe), swing and wheel ploughs with hard stone pebbles for anti-wear protection (1335–1340 AD, medieval Denmark), and, most interestingly, mechanical clocks with gear, bearing, spring, foliot, and escapement mechanisms (1386 AD, England). In this long period of time, various metal bearings, gears, and organic lubricants such as animal fats and vegetable oils were increasingly used. Preliminary knowledge on contact, lubrication, friction, and wear was furthered but still mainly empirical. No recorded scientific research was found in the literature until the Renaissance (about 1450–1600 AD).

1.3.2 Pioneering Studies

Leonardo da Vinci (1452–1519) was doubtlessly a genius having made milestone contributions not only in art but in science and engineering as well. His pioneer studies on tribology-related problems were perhaps first recorded in history, covering various aspects, including friction, wear, lubrication, bearing materials, and gear and bearing design. Further scientific investigations on friction and lubrication were inspirited by rapid development of basic science in the 16th through early 18th centuries mainly in Europe, pioneered by Francis Bacon (1561–1626), Galileo Galilei (1564–1642), Isaac Newton (1642–1727), and others.

Two distinguished studies in this field deserve special mentioning: one is Robert Hooke's law, which is considered as the birth of the elasticity theory, and the other is the concept of fluid viscous flow proposed by Isaac Newton (1687), both employing linear mathematic models to describe fundamental mechanical properties of solid and liquid. Knowledge on rolling and sliding friction was continuously furthered by Robert Hooke, Guillaume Amontons (1699), Gottfried W. Leibnitz (1706), Leonhard Euler (1750), and others, and then a summit was reached by Charles Coulomb (1780–1785), who presented the most comprehensive work at that time. Coulomb's friction law was another simple linear model, approximately describing basic frictional behavior (which will be discussed in Section 10.2.2). It has enjoyed wide application in engineering practice since then.

1.3.3 Establishment of Contact Mechanics and Lubrication Theory

The Industrial Revolution (around 1750–1850) greatly accelerated the development of contact and lubrication theories and practices, as many new mechanical innovations (such as steam engines, railways, machine tools, metal-forming and agriculture machineries, and others), and their rapid applications constantly imposed new and urgent challenges to interfacial research. Ball and roller bearings of various types were invented or significantly improved. Mineral oils and greases were produced as lubricants initially in 1850s, and they gradually gained a dominant market share due to their excellent functionality, abundant availability and low prices.

Lubricated journal bearings became widespread in railroad vehicles and other equipment, in which friction did not seem to follow Coulomb's law. Beauchamp Tower was then assigned by the Institution of Mechanical Engineers in the UK to investigate friction in journal bearings. His accidental discovery of hydrodynamic pressure in the oil film was published in his remarkable reports (1883, 1885). Based on Tower's experimental observations, Osborne Reynolds (1886) conducted his magnificent mathematical analysis and derived the Reynolds equation for fluid-film lubrication.

Tower's and Reynolds' work revolutionized the state of knowledge on lubrication and also presented a best example how experimental and theoretical investigations could collaborate successfully answering questions raised from engineering practice. Today, it is commonly agreed that Tower's and Reynolds' contributions symbolized the birth of modern lubrication theory, as a branch of engineering science, and the Reynolds equation has been its foundation since then.

Parallel to the lubrication studies mentioned above, research on dry contact also achieved significant breakthroughs in 1880s. A brilliant study was conducted by Heinrich Hertz, who published his milestone paper, "On the Contact of Elastic Solids" in 1881. In the meantime, the classical approaches based on the potential theory for calculating displacements and stresses due to traction forces applied on the surface of an elastic half-space were presented by V. Cerruti (1882) and J. Boussinesq (1885). It is now widely accepted that contact mechanics has become an independent branch of engineering science since 1880s. Over the last 130+ years, the well-known Hertzian theory has been enjoying broad application in engineering practice due to its simplicity and satisfactory accuracy for frictionless elastic counterformal contact of smooth surfaces. However, the significance of Cerruti and Boussinesq's contributions was not fully recognized until the recent years when model-based numerical solution methods have gradually become the mainstream in research and problem solving. The establishment and application of contact and lubrication theories started from the 1880s are symbolically sketched in Figure 1.18.

It is interesting to note that both contact mechanics and lubrication theory originated in the decade of 1880s, but they were completely irrelevant at that time, as the study of contact did not consider possible existence of any lubricant, and the lubrication analyses ignored contact and deformation of the solid bodies. This situation was continued for a long period of time until roughly late 1970s–1980s. It is also important to notice that, after that very fruitful decade during which theoretical foundation was laid, research progress was indeed

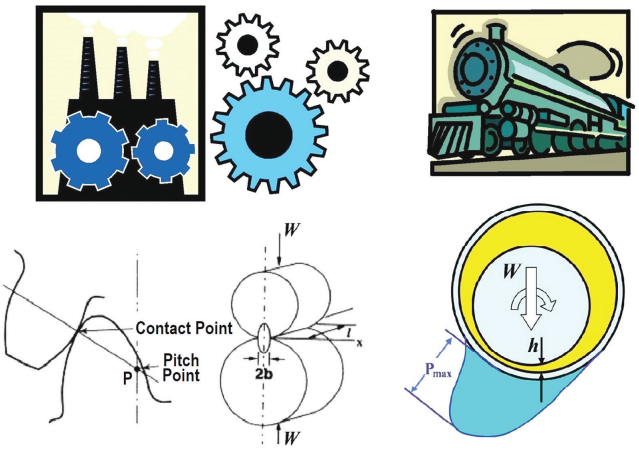


FIGURE 1.18 Establishment and application of contact and lubrication theories (1880s~). Left: Hertzian contact and right: fluid-film lubrication.

relatively slow until late 1950s–1970s when digital computers gradually became available to engineering researchers.

Actually, in the field of contact mechanics, limited progress was made other than the widespread application of the Hertzian theory in rapid-developing manufacturing industries and further development of the potential theories. The reason for the slow theoretical development was primarily due to the fact that most engineering problems are not analytically solvable, and numerical solutions were either impossible or greatly limited by manual calculation tools at that time.

For the lubrication theories, on the other hand, achievements appeared to be more significant, including hydrodynamic lubrication with compressible fluids such as air (Kingsbury, 1897, Harrison, 1913); explicit solutions of the Reynolds equation for infinitely long journal bearings (Sommerfeld, 1904); tilting pad bearings (Michell, 1905); hydrostatic bearings (Rayleigh, 1917); study on bearing dynamics, stiffness, and damping (Harrison, 1919, Stodola, 1925); short bearing solutions (Ocvirk, 1952); and others. However, most representative bearing studies during this period were analytical solutions under the assumptions of simplified geometry, such as infinitely long or short bearings. A small number of numerical solutions published, such as the one by Michell (1905), were approximate in one way or another, e.g. assuming a polynomial pressure distribution or neglecting high-order terms in a mathematical series. Since only simple manual calculation tools were available, complex problems could not be solved as they usually require massive computation.

1.3.4 Rapid Development Assisted by Digital Computers

After the Second World War, numerical solution methods, such as the finite difference method (FDM) for discretizing differential equations and relaxation procedure for finding converged solution, attracted increasingly more attention. Efforts on developing various algorithms for model-based numerical simulations were greatly encouraged by digital computers that gradually became available to researchers and engineers in the 1960s through 1980s. During this time period, the development was significantly accelerated in nearly every aspect of contact mechanics and lubrication theory.

However, among the most remarkable advancements, perhaps the following two achievements deserve special mentioning. The first is the numerical solution for fluid-film bearings of finite length, and the second is the establishment of elastohydrodynamic lubrication (EHL) theory. Both were on the basis of rapid advancing computer technologies.

Early numerical studies on finite journal bearings were conducted by Cameron and Wood (1949), Sassenfeld and Walther (1954), Pinkus (1958), Raimondi and Boyd (1958), and others. The new approach could cover wide ranges of design parameter and operating condition variations and take into account detailed complicated geometric structures without impractical simplifications. Thermal effect on bearing performance was studied by Dowson and Hudson (1964), and others, and dynamically loaded bearings were numerically investigated by Booker (1965, 1971), and others. Closely following these pioneer studies on journal bearings, the effect of journal housing deformation was investigated with the FEM, and many other components were also analyzed through numerical solutions, including thrust bearings, pistons and rings, and dynamic rotors, which will not be reviewed here.

The EHL theory, which considers lubrication enhancement due to surface elastic deformation and viscosity increase caused by high pressure mainly in non-conformal contacts, has been commonly recognized as the most important advancement in the field of lubrication theory during the second half of the 20th century. Attempts to analyze fluid-film lubrication in gears and rolling element bearings could be traced back to the beginning of the 20th century, and a remarkable example was the gear lubrication study by Martin (1916), whose results indicated that we could not reasonably predict lubricant film thickness between gear teeth by considering hydrodynamics alone. The first successful simplified EHL solution, taking into account the effects of both elastic deformation and viscosity increase, was obtained by Grubin (1949), and the first full numerical solution by Petrusevich (1951).

When the Reynolds equation is coupled with the elasticity equation and the pressure–viscosity relationship, the whole equation system becomes highly nonlinear, bringing tough challenges to numerical simulation. That is why the EHL research was greatly accelerated in 1960s and thereafter with the assistance of digital computers after Dowson and Higginson published their milestone paper, "A Numerical Solution to the Elastohydrodynamic Problem", in 1959. Based on the obtained numerical solutions, various EHL film thickness formulae were derived through curve fitting based on obtained numerical solutions by Dowson and Higginson (1961), Dowson and Toyoda (1978), et al. for line contact problems, and Hamrock and Dowson (1976–1977), and others for point contacts. Thermal EHL in line contacts was numerically tackled by Cheng and Sternlicht (1965), Cheng (1965), Dowson and Whitaker (1966), et al., and that in point

contacts by Zhu and Wen (1984) and Kim and Sadeghi (1992). In the meantime, transient solution, starvation effect, and non-Newtonian behavior of the EHL were also investigated. The EHL numerical solution was indeed the most attractive and challenging subject in the field of lubrication theory, and there were a large number of papers published in 1960s–1980s.

When solving contact and lubrication problems, one of the major difficulties is to appropriately handle surface roughness and non-Hertzian irregular contact geometry, with which there is little chance for getting successful analytical solutions. Early studies were based on stochastic models, employing a small number of statistical parameters to describe rough surface characteristics and their influences on interfacial behaviors. Representative contact studies include rough surface contact models by Greenwood and Williamson (1966), Greenwood and Tripp (1970–1971), and others. Later empirical formulae by means of statistical regression were developed by Lee and Ren (1996), and others based on the obtained results of numerical contact simulations. Stochastic lubrication models were presented by Christensen (1970), Elrod (1973), Tonder (1977), Patir and Cheng (1978a), et al. Stochastic mixed EHL analyses with rough surfaces, using Patir and Cheng's average Reynolds equation and Greenwood and Tripp's model to separately handle lubrication and asperity contact, were accomplished by Patir and Cheng (1978b), Majumdar and Hamrock (1982), Prakash and Czichos (1983), et al. for line contact problems, and by Zhu and Cheng (1988) for point contacts. These stochastic analyses could demonstrate some basic trends and predict mean values of pressure and film thickness (or gap) but failed to give detailed physical parameter distributions and maximum/minimum values, which are often critical in performance evaluation and failure analysis. Also, these stochastic models seem to be incapable of handling cases under heavy loads with severe asperity contacts and deep asperity penetration.

1.3.5 Recent Advancements

Started roughly from the mid-1990s, computer and information technologies have been explosively advancing that have irreversibly changed the ways of science and engineering research. Many difficult tasks impossible to accomplish in the past have become possible today. Any real machined roughness and irregular contact geometry, e.g. can now be directly digitized and used in computer simulations at much improved discretization accuracy, thus, some simplified, but often controversial, stochastic models may no longer be needed in many cases, especially for counterformal contact problems.

Early deterministic contact analyses with real machined roughness and/or arbitrary contact geometry include those by Ju and Farris (1996), Stabley and Kato (1997), et al. Improved numerical solution algorithms for contact problems were developed by Hu et al. (1999), Ai and Sawamiphakdi (1999), Polonsky and Keer (1999), and others.

In order to accelerate the computation in solution process while ensure the numerical accuracy, efficient numerical algorithms have been developed that include the multi-level multi-integration (MLMI) method by Lubrecht and Ioannides (1991) and the discrete convolution and fast Fourier transform (DC-FFT) by S. B. Liu et al. (2000) and Liu and Wang (2002) based on the discrete convolution theorem.

In recent years, great efforts have been made to solve contact and lubrication problems with elastoplastic and inhomogeneous materials. For example, although contact study with layered and functionally graded materials were pioneered by Burmister (1945), W. T. Chen (1971), Chen and Engel (1972), et al., the research progress has greatly been accelerated recently, yielding fruitful results that include those from O'Sullivan and King (1988), Nogi and Kato (1997), Polonsky and Keer (2000a), S. B. Liu and Wang (2001, 2005), Cai and Bhushan (2005), C. J. Yu et al. (2013, 2014), and others, in which thermo-elasticity may also be considered. On the other hand, contact modeling has been extended to take into account possible plastic deformation and material hardening by Jacq et al. (2002), F. Wang and Keer (2005), Nelias et al. (2007), W. W. Chen et al. (2007), and Chen and Wang (2008a,b), through a semi-analytical method (SAM). The equivalent inclusion method (EIM) by Eshelby (1957) was used to build contact models for handling inhomogeneous materials (W. W. Chen et al., 2010, Leroux et al., 2010, Leroux and Nélias, 2011, K. Zhou et al., 2011a,b, and many others). Later, S. B. Liu et al. (2012) published a set of analytical formulae for elastic fields caused by eigenstrains in a half-space solid, providing the analytical core for fast and effective FFT-based numerical simulation tool in order to analyze arbitrarily distributed eigenstrains and material inhomogeneities in various contact problems. Contact modeling efforts have also be extended to tackle multiple field electromagnetic-elastic problems (Ding et al., 1996, Giannakopoulos and Suresh, 1999, X. Zhang et al., 2017).

In the field of lubrication study, significant breakthroughs were made also in the mid-1990s and thereafter. First, deterministic EHL solutions in point contacts with real machined rough surfaces were presented by Xu and Sadeghi (1996), Zhu and Ai (1997), and others. Generally, this type of solution requires advanced high-speed computers with large memory space, which are necessary for sufficiently accurate discretization of the Reynolds equation in association with high-frequency machined roughness usually in motion that makes the solution time-dependent even under steady-state operating conditions.

Second, upon improving the so-called "semi-system approach" originally presented by Ai (1993), Zhu and Hu (1999, 2001a,b), and Hu and Zhu (2000) developed a unified numerical approach to handle both hydrodynamic lubrication and surface asperity contact simultaneously in the same system of mixed EHL. This approach is capable of simulating the entire transition from full-film to mixed EHL all the way down to dry contact under severe operating conditions using either smooth or machined 3D rough surfaces. This continuous transition was further investigated and the numerical method modified by Zhu (2003b, 2007), Y. C. Liu et al. (2006a), and others. A set of numerical solutions for a mixed EHL sample case with two machined rough surfaces in a rolling-sliding

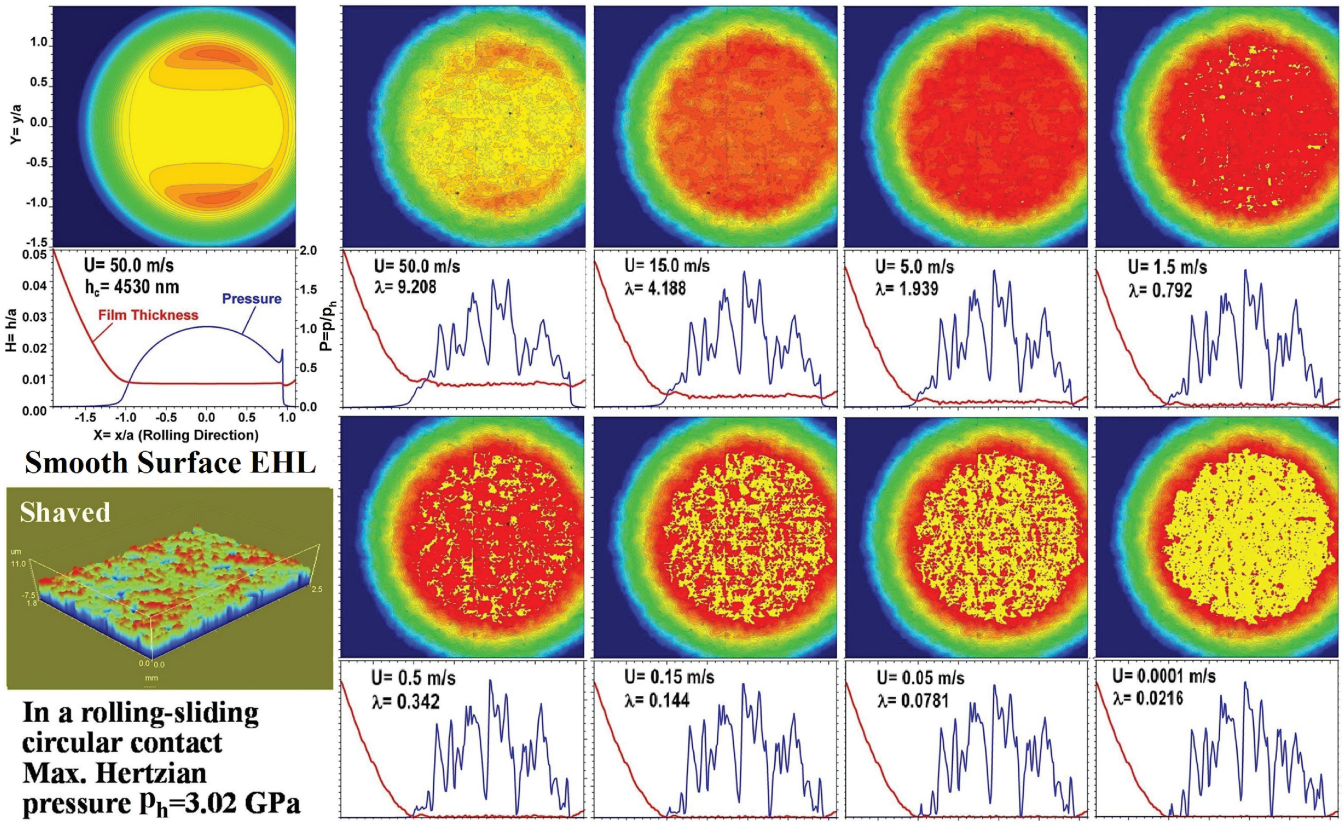


FIGURE 1.19 Simulation of entire transition from full-film and mixed EHL to dry contact with a unified model and numerical solution approach.

point contact, demonstrating the entire transition of lubrication status, simulated by this unified model approach, is presented in Figure 1.19, where the yellow-colored areas in the film-thickness contour maps indicate micro-scale rough-surface asperity contacts, which increase significantly as the rolling speed decreases. Refer to Sections 7.3.7 and 7.3.8 for detailed description.

The recent work on the mixed EHL by Zhu, Hu, and others mentioned above has successfully proven that we can use a unified mixed EHL model approach to simulate both hydrodynamic lubrication and solid-to-solid asperity contacts. This has bridged contact and lubrication for the first time. As described in the previous section, dry contact is a special case of lubricated contact under some extreme conditions. Therefore, the lubrication formulation should be appropriate also for solving contact problems as long as it takes into account surface deformation.

Although theoretically there is no barrier in between, contact and lubrication problems have long been tackled separately by different researchers with different methods since 1880s due to technical difficulties. The separate treatments may be good in many cases, as the dry contact problems are usually simpler than those of lubrication under otherwise the same conditions, thus, more efficient special handling of dry contact is often justified. However, in a mixed lubrication system, both hydrodynamic films and direct surface asperity contacts coexist, and the borders between these two different areas are not pre-defined and often irregular and transient, thus, separate treatments in a deterministic simulation could be difficult or even inappropriate. The unified approach provides a powerful tool for solving mixed lubrication problems under severe conditions, and it has demonstrated great potentials in various engineering applications.

1.3.6 Conclusion Remarks

We are not, once again, trying to present a comprehensive review of the history. There are certainly some important studies not mentioned here (some representative investigations will be referenced later in relevant chapters). The main objective of the above review is to make the following points:

The theoretical foundation of contact and lubrication mechanics was laid in the 1880s, but rapid development did not take place until late 1950s–1980s when digital computers gradually became available to researchers and engineers. The primary reason is that most contact and lubrication problems found in engineering reality are not analytically solvable, and numerical solutions were greatly limited by manual calculation tools in the past.

Dry contact is nothing but a special case of lubricated contact, and there is no theoretical barrier in between. However, they have long been treated separately until recently due to technical difficulties. Today, with the assistance of advanced

computer and numerical solution technologies, they have been bridged together, and the entire transition from full-film and mixed lubrication down to dry contact can now be simulated with a unified model and numerical approach that are capable of analyzing the effect of machined surface roughness under severe operating conditions.

Numerically solving practical problems requires advanced high-speed computers and proper numerical solution technologies in order to satisfactorily handle non-Hertzian contact geometry, real machined roughness, non-Newtonian lubricant rheology, thermal behavior, inhomogeneous and inelastic material properties, as well as severe/extreme operating conditions, and more. These solutions for practical cases have recently become possible. However, there are still great challenges together with vast opportunities for further development in the field of interfacial mechanics.

1.4 INTERFACIAL MECHANICS

The above discussion helps justify the statement that the classic contact and lubrication theories originated in the 1880s have been promoted to a new stage, and the ways of conducting research have been irreversibly changed. The contact mechanics and lubrication theories are merging into "Interfacial Mechanics" (as originally stated in the paper by Zhu and Wang, 2011), which has indeed a long development history of nearly 140 years but has been renewed recently due to rapid advancements of computer and numerical simulation technologies in an extended scope. The concept of interfacial mechanics may possibly provide a scientific platform for multi-scale interdisciplinary research, helpful for collaborating with other closely related fields, as schematically illustrated in Figure 1.20.

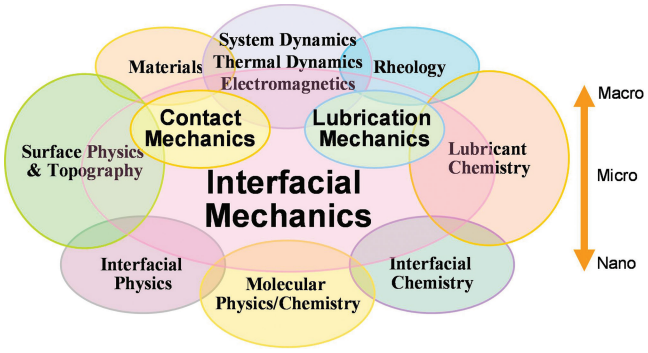


FIGURE 1.20 Interfacial mechanics and its related fields of study. (From Zhu and Wang, 2011.)

1.5 COVERAGE OF THIS BOOK

This book intends to cover the following aspects:

1. Fundamental principles and basic equations of both contact and lubrication theories. Although there is theoretically no barrier between contact and lubrication, and their formulations and numerical approaches can be unified, in many cases, the basic mathematic treatments for each will still be separately described, because contact problems are considerably simpler under otherwise the same conditions due to neglected lubricant and its rheology. The unified approach will be applied mainly in solving mixed lubrication problems, in which both hydrodynamic lubricant films and direct surface asperity contacts coexist and separate treatment may be difficult or even inappropriate.
2. Basic numerical solution techniques. Analytical solutions will not be the major focal points, as the vast majority of practical problems are not analytically solvable. However, some of the basic analytical solutions will still be provided for completeness and model validation through comparison, especially as cornerstones for building sophisticated numerical algorithms when needed.
3. Most recent and representative numerical results and their physical meanings, such as deterministic solutions of contact and lubrication problems with rough surfaces, computer simulations of mixed lubrication, which can be extended to those also covering the dry contact and full-film lubrication, and contact and EHL with elastoplastic and inhomogeneous materials presented recently, and more. These updated results aim at demonstrating basic trends and mechanisms of interfacial behaviors in extended ranges of operating conditions and the influences of materials and surface geometric/topographic properties.
4. Some special topics, such as interface failure criteria and failure prediction methods, virtual texturing techniques, mechanisms of emulsion lubrication and mixed lubrication in gears, the effect of multiple fields on interfacial characteristics, and so on.

We tried to make our best efforts to reflect the latest research developments in the last a few decades and to present practical and efficient numerical methods for modeling tribological interfaces. We hope these efforts, together with the technical contents included in this book, can stimulate further in-depth investigations and discussions.

2 Properties of Engineering Materials and Surfaces

This chapter describes some basic properties of engineering materials and surfaces that are fundamental for physical and mathematical model development. The materials discussed here are referred mainly to solid materials such as metals and ceramics, as well as commonly used fluid lubricants. We will briefly review several basic physical concepts, parameters, and rules pertaining to the behaviors of materials for clarity and convenience in the use in interfacial modeling and analyses without detailed explorations of the material properties.

2.1 MECHANICAL PROPERTIES OF TYPICAL SOLID MATERIALS

Most solid engineering materials demonstrate elastic behavior if the deformation (or strain) is small. In a test under uniaxial stress condition, when the load starts to increase gradually, the material is initially in its elastic region, thus, a straight (or nearly straight) line relationship between stress σ and strain ε is usually followed, as shown in Figure 2.1. This linear relationship is known as Hooke's law, and the slope of straight line OA is elastic modulus E, also called Young's modulus. The end of this linear region marks the proportional limit at A, corresponding to stress σ_p. When the load is further increased, the stress–strain relationship may deviate slightly from the straight line, but the deformation may still be elastic, because if the load is slowly released in this region, the material will backtrack to its original shape with no considerable residual deformation.

For a typical ductile material such as steel, as the load continuously increases, the initial yielding limit, σ_{y0} (later often expressed as S_y), will be reached and the material will start to show noticeable plastic (permanent) deformation. The plastic strain may grow drastically beyond the yield point B. After reaching the ultimate strength limit, σ_u, at point D,

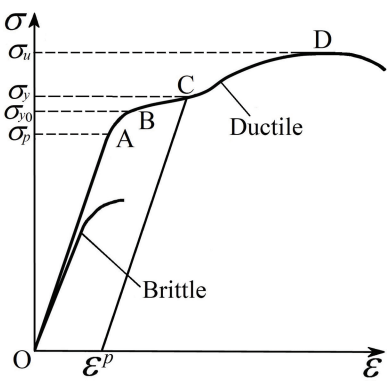


FIGURE 2.1 Typical stress–strain relationships.

the material may soon break. Note that the initial yield limit, σ_{y0}, or the yield strength, is an important property of ductile materials, which is commonly defined at 0.2% plastic strain based on the stress–strain curve under the uniaxial stress condition. Usually the yield strengths under tension and compression are about the same for a ductile material.

In the plastic zone beyond the yielding limit, the material behavior becomes more complicated. Generally, if the load is released at any point C in the plastic zone, stress σ will return back to zero or very small but strain ε will partially remain. The remaining portion of the strain, ε^p, is called plastic strain. When the load is applied again, the stress may need to reach a higher yielding limit, σ_y, to gain additional plastic deformation due to material strain hardening, as shown in Figure 2.1.

The slope of the ε–σ curve in the plastic zone, E_T, may or may not be a constant. It is known as the tangent modulus that is resulted from both elastic and plastic strains. Because the strain increment after yield consists of both elastic and plastic strains, $d\varepsilon = d\varepsilon^e + d\varepsilon^p$, similar to the situation of two springs connected in series, the plastic modulus, H, which considers only the portion of pure plastic strain, is defined as follows (see Chakrabarty, 2006):

$$\frac{1}{E_T} = \frac{1}{E} + \frac{1}{H} \tag{2.1}$$

For a brittle material such as cast iron or ceramic or concrete, on the other hand, rupture may take place much earlier without significant plastic deformation especially under a tensile strain condition, as schematically shown in Figure 2.1.

Material hardening in the plastic zone beyond the yield limit is a complicated phenomenon. A number of empirical laws for strain hardening have been experimentally obtained. Two commonly used hardening laws are schematically depicted in Figure 2.2 in the form of simplified stress–strain relations for the purpose of describing different material behaviors approximately. The linear hardening law can be expressed as (see Hill, 1950, Simo and Hughes, 1998)

$$\sigma_y\left(\varepsilon^p\right) = \sigma_{y0} + \frac{E_T}{1 - E_T/E}\varepsilon^p \tag{2.2.1}$$

where E is the elastic (Young's) modulus, E_T the tangent modulus, σ_{y0} the initial yield limit, and ε^p the plastic strain. Note that materials with a negligible strain hardening rate, $E_T/E = 0$, are regarded as ideally plastic or elastic-perfectly plastic (EPP), while those with a very strong hardening, $E_T/E \to 1.0$, can be considered as ideally elastic with the yield strength $\sigma_y \to \infty$.

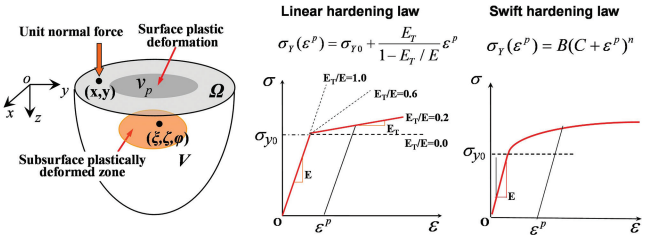


FIGURE 2.2 Different hardening laws.

A nonlinear hardening law, the Swift power hardening law (see Swift, 1952), is written as

$$\sigma_y\left(\varepsilon^p\right) = B\left(C + \varepsilon^p\right)^n \tag{2.2.2}$$

where B, C, and n are three work-hardening constants, which offer more flexibility for accurate description of the material property. Obviously, the initial yield strength is $\sigma_{y0} = B \cdot C^n$. More discussion on the plasticity will be given in Chapter 11.

Poisson's ratio ν is also an important property of solid material, which is defined as the ratio of lateral strain to longitudinal strain under the condition of uniaxial loading. Table 2.1 shows the mechanical property data for some commonly used materials as references.

It is important to note that, although solid materials are traditionally assumed to be homogeneous continuua, in reality, various types of inhomogeneity, such as impurities, inclusions, voids, and non-uniform surface layers, widely exist, often significantly affecting interfacial characteristics. Today, diverse material and surface treatment technologies have been developed and commonly used that may involve favorable material inhomogeneities.

For example, surface enhancement techniques are widely employed today to improve the strength of surfaces and possibly to reduce friction as well. Typical techniques include: (1) hardfacing through either fusion hardfacing or thermal spray; (2) coating by means of chemical vapor deposition or physical vapor deposition, etc.; (3) diffusion processes such as carburizing, nitriding, boriding, and others; (4) selective hardening such as flame hardening, induction hardening, and laser hardening; (5) ion implantation; (6) laser cladding and weld overlay; and many others (see Schneider and Chatterjee, 2013).

The major advantage of surface hardening is that the treatment may only enhance the strength within a certain case depth, while keeping the bulk of the material intact. Figure 2.3 presents the hardness variations in the depth direction for several steels case hardened through different carburizing methods and an inductive heat treatment. For these typical examples, the case depth is about 1–1.5 mm. Increased surface hardness can help significantly improve the rolling contact fatigue (RCF) life, while the toughness of the bulk material still remains.

2.2 TOPOGRAPHIC PROPERTIES OF ENGINEERING SURFACES

2.2.1 Engineering Surfaces

In engineering reality, no surface is ideally smooth. Surface roughness is usually a natural inevasible result from a surface finishing process such as machining. The overall surface appearance is referred to as topography and the pattern and structure of a surface geometry are considered as the surface texture. Generally, topographic peaks of a surface are called asperities.

In most cases, better finished surfaces are usually preferred for the purpose of improving the performance, appearance, efficiency, and reliability of components, but in the meantime, they may also lead to a considerable increase in fabrication cost. In practice, therefore, engineers are often required to find a good compromise between surface finish improvement and cost reduction. For this purpose, the surface roughness effect on interfacial characteristics needs to be investigated and well understood.

Three-dimensional (3D) images of selected machined surfaces, acquired by using an optical profilometer, are shown in

TABLE 2.1
Mechanical Properties of Some Engineering Materials

Material	Elastic Modulus E (GPa)	Elastic Shear Modulus G (GPa)	Poisson's Ratio ν	Density (g/cm^3)
Steel	190–210	76–83	0.25–0.3	7.4–7.9
Cast iron	118–126	44–46	0.3	7.0
Copper	108–127	39–48	0.31–0.34	8.9
Bronze	108–113	39–41	0.32–0.35	8.7–8.9
Brass	89–97	34–36	0.32–0.40	8.4–8.8
Aluminum	68–70	25–27	0.32–0.35	2.7
Zinc	82	31.4	0.27	7.1
Lead	16	6.8	0.42	11.37
Ceramics	350–550		0.17–0.25	2.3–5.1
Glass	55–70		0.25–0.27	2.2–2.4
Concrete	13.7–39.0	5–15.7	0.1–0.18	1.8–2.5
Rubber	0.0078		0.47	0.93

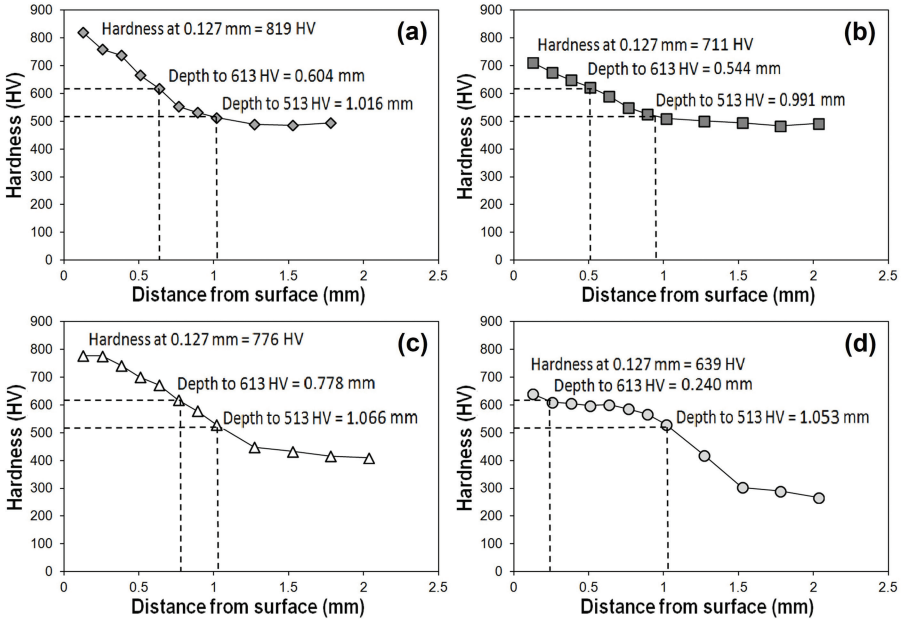


FIGURE 2.3 Microhardness profiles in the depth direction. (a) vacuum carburized AISI 8620, (b) atmosphere carburized AISI 8620, (c) vacuum carburized AISI 9310, and (d) induction hardened AISI 4140. (From Xie et al., 2015.)

Figure 2.4, which cover a good variety of surfaces commonly found in engineering reality. It can be seen clearly that both surface roughness height and topography vary a great deal. The statistical parameters, R_a, R_q, and R_t, will be defined in the next section.

These surfaces, shown in Figure 2.4, will be used repeatedly in analyzed numerical sample cases throughout this book.

In reality, a tribological process may often alter the topography of a surface. Running-in, e.g. may smoothen the surfaces, resulting in a better conformity, while excessive wear and pitting due to contact fatigue may roughen the surfaces and cause even more severe failures and seizure in a possible chain reaction. Figure 2.5 presents an example, showing the topographic evolution of a boron carbide-coating surface against a 52100 steel ball, caused by wear. The images are plotted from the data digitized by an atomic force microscope (AFM).

2.2.2 Surface Characterization by Statistical Parameters

Surface topographic feature is complex in nature, often demonstrating considerable randomness due to the machining process used, as shown in Figure 2.4. Generally it cannot be satisfactorily described by only one or a few parameters. In engineering practice, however, characteristic parameters for surface property should be simply specified in one way or another, e.g. in engineering drawings, blue prints, and technical documents. Also, statistical analyses are often needed in some interfacial studies. Under the circumstances, the following statistical parameters are commonly employed for general purposes while discussions and arguments constantly remain on which parameters to use especially for some specific applications.

Two-dimensional (2D) surface profiles taken along a traverse distance (x) are used below to illustrate statistical parameter definitions, but the concepts of these parameters are also applicable to 3D topography. Figure 2.6 illustrates such a profile as an example. Note that in order to see the tiny roughness height variation clearly, a large magnification, e.g. 1,000–2,000X, is usually applied for the height in the z-direction, in comparison to the sampling distance in the x-direction.

Roughness Average, or Centerline Average (CLA), R_a

This is the arithmetic average of the absolute values of the profile height deviations recorded within the evaluation length L (or area) and measured from the mean line (also called the centerline, or sometimes the nominal smooth surface). If the mean line is chosen to be the x-axis, as shown in Figure 2.6, R_a can be analytically defined as

$$R_a = \frac{1}{L}\int_0^L |z(x)|\,dx \tag{2.3}$$

or, for measured discrete roughness heights $|z_i|$ ($i = 1, 2, 3, \ldots, N$):

$$R_a = \left(|z_1| + |z_2| + |z_3| + \cdots + |z_N|\right)/N = \frac{1}{N}\sum_{i=1}^{N} |z_i| \tag{2.4}$$

where N is the total number of discrete datum points. Note that R_a is most commonly used for specifying the surface finish in industries.

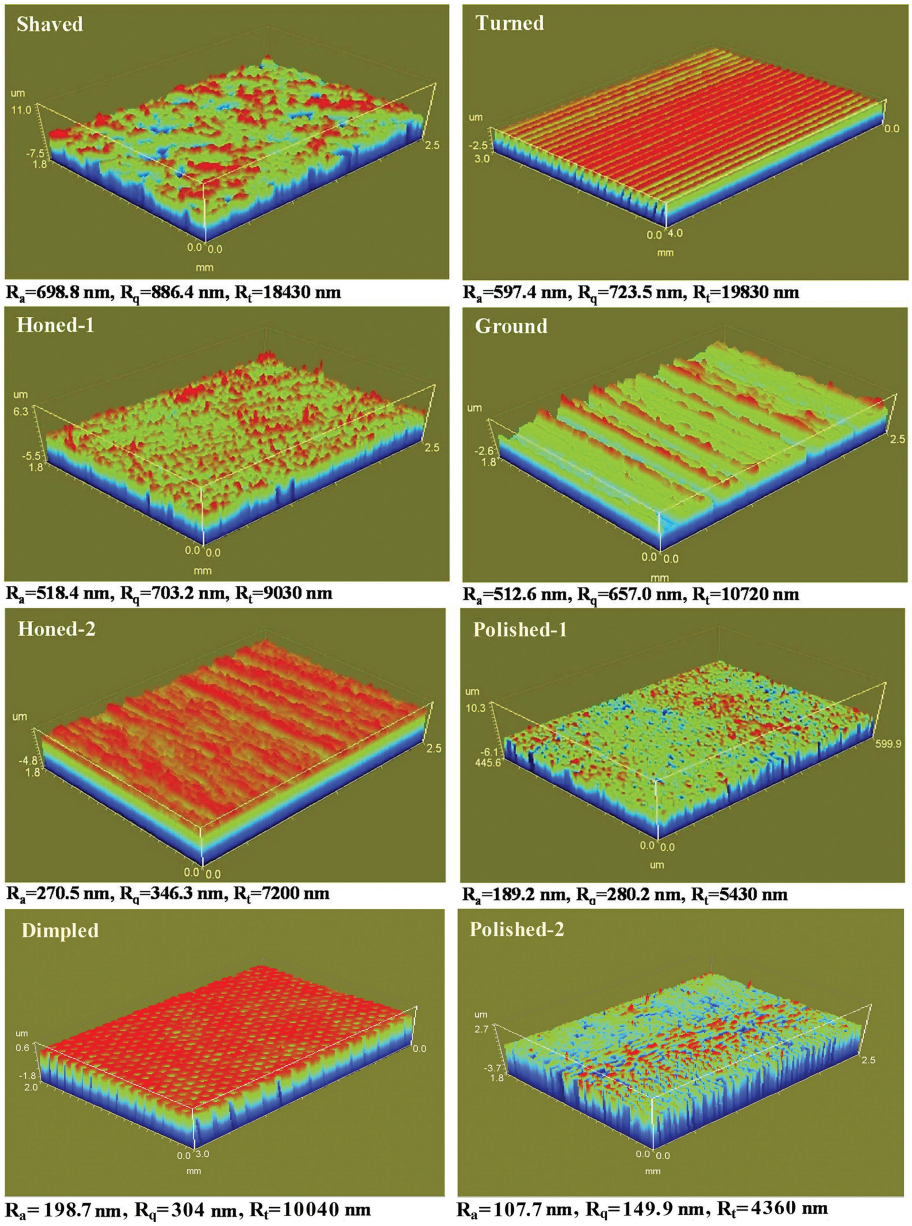


FIGURE 2.4 Typical machined surfaces. (From Zhu, 2003c, modified.)

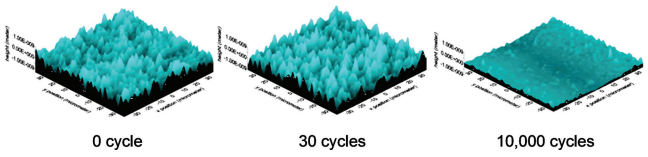


FIGURE 2.5 Topographic evolution of a boron carbide coating surface due to wear in a ball-on-disk test. (From Harris et al., 2002.)

Standard Deviation or Root Mean Square (RMS) Roughness, R_q

R_q is defined as the root mean square average of the profile height deviations taken within the evaluation length (or area) and measured from the mean line, expressed as follows:

$$R_q = \left[\frac{1}{L} \int_0^L z(x)^2 \, dx \right]^{1/2} \tag{2.5}$$

The digital approximation can be given by

$$R_q = \left[\left(z_1^2 + z_2^2 + z_3^2 + \cdots + z_N^2 \right) / N \right]^{1/2} \tag{2.6}$$

Maximum Profile Peak Height, R_p

This parameter means the distance between the highest point of the profile and the mean line within the evaluation length, as shown in Figure 2.6.

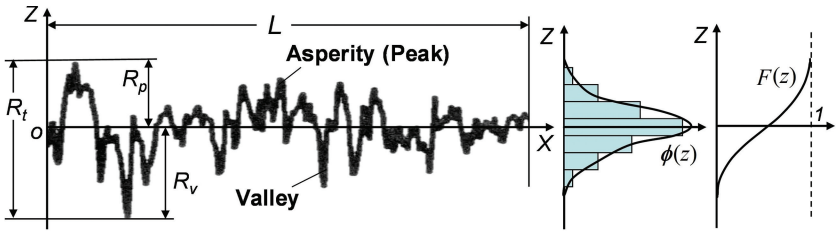


FIGURE 2.6 Surface profile, its height histogram, and cumulative distribution function.

Maximum Profile Valley Depth, R_v

It is the distance between the lowest point of the profile and the mean line within the evaluation length (see Figure 2.6).

Maximum Height of the Profile, R_t

This is defined as the vertical distance between the highest and the lowest points within the evaluation length.

$$R_t = R_p + R_v \tag{2.7}$$

Probability Density Function (PDF), $\phi(z)$

$\phi(z)$ is often utilized to depict surface statistical characteristics as most engineering surfaces are random to a large extent due to the finishing process used. For many freshly machined surfaces, the probability density function (PDF) of roughness height distribution is often Gaussian or nearly Gaussian that can be expressed as

$$\phi(z) = \frac{1}{R_q\sqrt{2\pi}} e^{-\frac{z^2}{2R_q^2}} \tag{2.8}$$

It is also called the normal distribution.

Cumulative Distribution Function, $F(z_o)$, and Bearing Area, BA

For a given PDF, the probability, Pr, for roughness height z to be between z_1 and z_2 can be calculated by integrating the PDF from z_1 to z_2:

$$\Pr(z_1 \le z \le z_2) = \int_{z_1}^{z_2} \phi(z)\,dz \tag{2.9}$$

The cumulative distribution function, F, and the bearing area, BA, at a particular roughness height, z_0, can also be determined. The distribution function, F, at z_0 is the accumulative value of the PDF for $z \le z_0$:

$$F(z_0) = \int_{-\infty}^{z_0} \phi(z)\,dz \tag{2.10}$$

Clearly, integration of the PDF over the full range of z results in one:

$$F(\infty) = \int_{-\infty}^{\infty} \phi(z)\,dz = 1 \tag{2.11}$$

The bearing area, BA, accounts for percentage of asperities in "contact" with a flat surface of a rigid body approaching from summits of a rough surface.

$$BA(z_0) = \int_{z_0}^{\infty} \phi(z)\,dz = 1 - F(z_0) \tag{2.12}$$

which is the area from z_0 to infinity in the probability density function plot.

Skewness, R_{sk}, and Kurtosis, R_{ku}

These are two commonly used parameters that are employed to express the shape of a distribution function plot. Skewness is defined as the third statistical moment of height distribution. Both continuous and discrete forms are given below.

$$R_{sk} = \frac{\int_0^L z(x)^3\,dz}{R_q^3 L} = \frac{1}{R_q^3 N}\sum_1^N z_i^3 \tag{2.13}$$

Kurtosis, R_{ku}, is defined as the fourth statistical moment of height distribution. Both continuous and discrete forms can be written as

$$R_{ku} = \frac{\int_0^L z(x)^4\,dz}{R_q^4 L} = \frac{1}{R_q^4 N}\sum_1^N z_i^4 \tag{2.14}$$

These two statistical parameters further describe the patterns of a height distribution. For a surface with a Gaussian distribution, the skewness should be 0 due to symmetry of the height distribution probability, and the kurtosis should be 3. Figure 2.7 shows the relationships between the values of skewness and kurtosis and the shapes of the distribution functions by using the Gaussian distribution as a comparison reference. Negative skewness is likely due to deep valleys of the surface profile, which result in large negative terms in the skewness calculation after the application of the power of 3. The deep valleys cause a tail to the left in the PDF plot. Similarly, positive

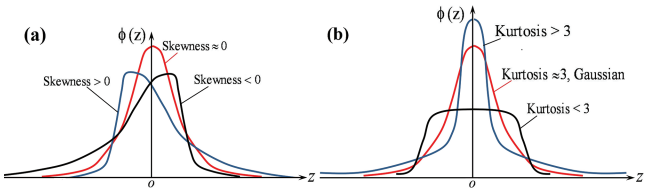


FIGURE 2.7 Comparing surfaces with different skewness and kurtosis.

skewness is due to tall asperities, which result in a tail to the positive right side of the corresponding PDF.

On the other hand, a surface with kurtosis larger than that of a Gaussian distribution, which is 3, may have high peaks and deep valleys and is called leptokurtic, while a surface with kurtosis smaller than 3 may have far less sharp peaks and shallower valleys and is called platykurtic.

Autocorrelation Function

The probability density function, $\phi(z)$, has yielded statistical information about a surface, mainly on how the heights deviate from the mean. However, surfaces having the same centerline average or RMS roughness values may possess very different topographic structure. Sinusoidal functions of the same amplitude but different frequencies, triangular, and trapezoidal functions of the same peak-valley value but different spans are some of the examples. Autocorrelation function is a spatial structural function that is commonly used to reveal the structural information of a surface. Mathematically, autocorrelation function $R(l)$ of function $f(x)$ is the integration of the product of the latter by its shifted form over the entire domain of x. l is the distance of shift along the x-direction. The autocorrelation function, $R(l)$, is defined as follows in Figure 2.8 and by Equation (2.15). Its non-dimensional form, $r(l)$, is given in Equation (2.16), which is $R(l)$ normalized by the square of the RMS roughness. Equation (2.15) gives both the summation and integration forms for the use of discrete and continuous surface data, respectively.

$$R(l) = E\left[z(x)z(x+l)\right] = \frac{1}{N-N_l}\sum_{1}^{N-N_l}\left[z_i(x)z_i(x+l)\right]$$

$$= \lim_{L\to\infty}\frac{1}{L}\int_0^L z(x)z(x+l)dx \qquad (2.15)$$

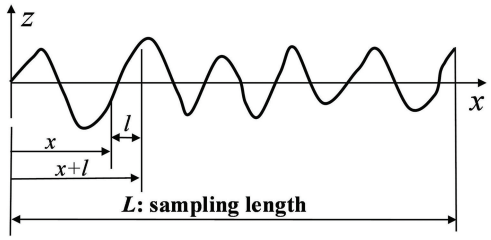


FIGURE 2.8 Calculation of autocorrelation function.

where L is the sampling length, l the shift in x, and N_l the number of data within l.

$$r(l) = R(l)/R_q^2 \qquad (2.16)$$

It is easily seen that when $l = 0$, the autocorrelation function, $R(l)$, becomes RMS roughness R_q and the non-dimensional autocorrelation function, $r(l)$, becomes one.

Correlation Length, l^*, and Asperity Orientation

The correlation length, l^*, in Figure 2.9, measures how quickly a random event decays. It is the value of l over which $r(l)$ drops to $A\%$ of its value at the origin (which is one or the square of the RMS roughness when dimensional). Here, $A\%$ may be 10%, 20%, 37% (or 36.8%, which is $1/e$), or 50%. These values are all commonly seen in research publications.

The correlation length may be used to determine the orientation of surface roughness. A parameter (Peklenik number), γ, is defined as the ratio of the correlation lengths in the two orthogonal directions:

$$\gamma = l_x^*/l_y^* \qquad (2.17)$$

The Peklenik number indicates surface topographic (or texture) orientation. γ equals one, or around one, implies an isotropic surface without a directional prereference, while a transverse surface can be defined by $\gamma \ll 1$ and longitudinal by $\gamma \gg 1$. The term "transverse" means that the "ridges" of asperities characterized by the x-direction correlation length are perpendicular to the direction of the relative motion of the surfaces, and the "longitudinal" indicates the ridges parallel to the relative motion.

Figure 2.10 illustrates the asperity footprints of surfaces with different orientations. Note that the footprints of longitudinal and transverse surfaces are similar. Rotating one

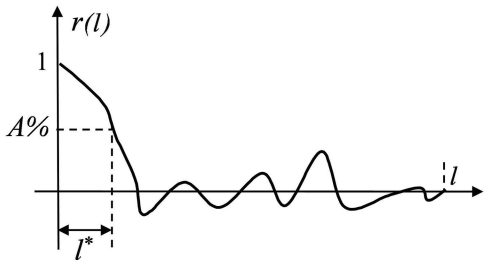


FIGURE 2.9 Definition of correlation length l^*.

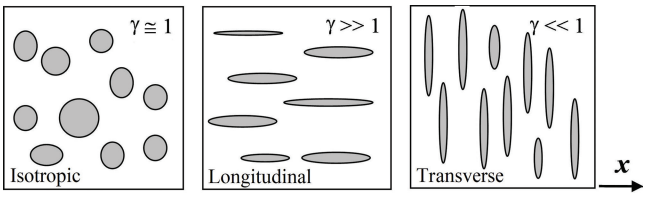


FIGURE 2.10 Various asperity orientations indicated by different values of γ.

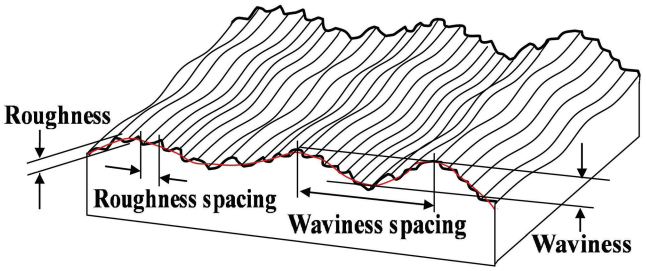


FIGURE 2.11 Waviness and roughness.

for 90° would result in the other. Patir and Cheng (1978a) used $\gamma = 9$ as the required very large number to define a longitudinal rough surface, and $\gamma = 1/9$ to be the one for transverse texture. Later, Lee and Ren (1996) found that $\gamma = 6$ would already be sufficient to see clear longitudinal orientation.

Waviness and Roughness

A surface may have irregularities at different scales. Waviness and roughness are the common terms to explain some of the geometric properties. Figure 2.11 simply illustrates a typical sample that has the appearance of surface waviness and roughness. In fact, if we look at small-scale surface features, the surface profile height variation shown in Figure 2.11 may be considered as the "waviness" in comparison to the "roughness" observed at an even smaller scale. In other words, the so-called waviness is a kind of roughness at a larger scale of observation. The resolution of the measurement instrument used for surface digitization and the parameters employed in the measurement determine the level of surface geometric features to be observed.

2.2.3 Surface Characterization by Direct Digitization

It is important to note that, as mentioned before, rough surface topography is usually complex and there would be more than a hundred statistical parameters that could be selected to describe its characteristics. Those given above are only a small portion of them, which are more frequently utilized in practice. Although in the last several decades great efforts have been made to develop stochastic models, it has been proven that a limited number of parameters are generally insufficient to describe the topographic features and their influences on interfacial characteristics, especially in engineering practice where only a few parameters (or less) are usually allowed. However, the further development of data science may help make the statistical models more powerful tools for surface characterization.

An alternative is to fully digitize the topography and use the obtained data matrix as a deterministic representation of the surface. It requires, however, advanced surface measurement technology with a sufficiently high resolution, as well as data-processing software capability with a large memory space, which were historically infeasible until recently. In the last 2–3 decades or so, a number of techniques have become available for the measurement and digitization of engineering surface topography with a satisfactory resolution. These include scanning electron microscopy (SEM), scanning tunneling microscopy (STM), atomic force microscopy (AFM), 3D stylus profilometry and optical profilometry of various types, and others.

Basically, all the techniques can be divided into two groups: contact measurement and non-contact measurement. Traditionally, surfaces are characterized by using two-dimensional measurement techniques with stylus, and sample profiles obtained can be found in Figures 2.6 and 2.12. Based on a measured profile, required statistical parameters can be readily obtained.

However, it is apparent that 2D measurements are often insufficient to describe/differentiate surfaces. In Figure 2.12, e.g. the two machined surface profiles look similar and their R_a and R_q values are close. In fact, these two surfaces are also illustrated in 3D in Figure 2.4, showing completely different topographic properties. This indicates that 2D profiles are often inappropriate, and 3D measurement is highly preferred in practice.

Caution should be exercised to be aware of the limitations of some surface measurement instruments. The stylus type of profilometers, e.g. probes a surface with a sharp stylus tip and scans the surface by the tip motion. The measurement accuracy is limited and distortion of the measured profile is inevitable, as shown in Figure 2.13, because of the limitations with the stylus tip size and radius of curvature. However, a stylus profilometer may be made portable and small enough to be attached to a machine for in situ inspection. With a stylus device, it is convenient to measure internal surfaces non-destructively. The stylus technology has been updated for 3D measurement through scanning line by line.

In the last 30 years or so, non-contact measurement techniques have been rapidly developed. Optical profilometry, for instance, has been widely employed in engineering practice. It is claimed that the vertical resolution can be as high as 0.5 nm

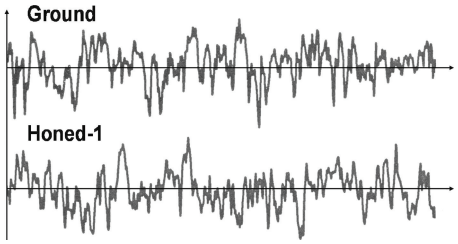


FIGURE 2.12 Two machined surface profiles from 2D stylus measurements.

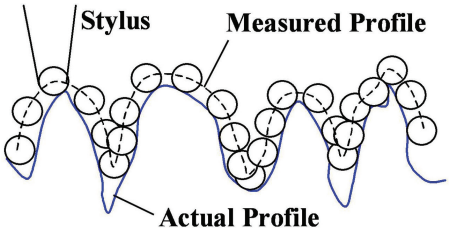


FIGURE 2.13 Distortion of measured profile due to stylus tip radius.

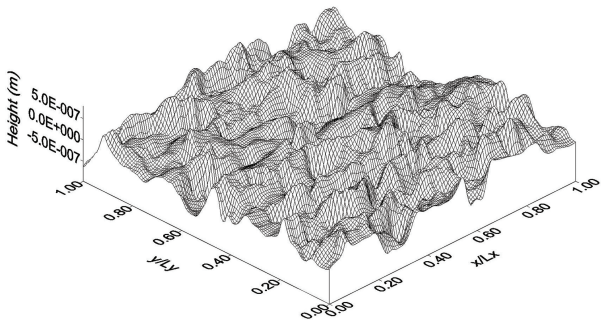


FIGURE 2.14 An optically measured sample surface topography.

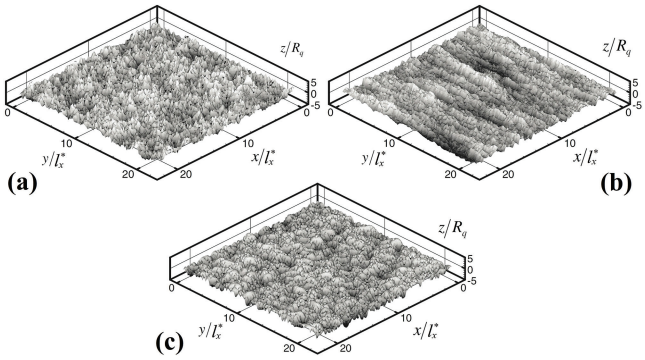


FIGURE 2.15 Sample rough surfaces generated by computer. (a) an isotropic Gaussian surface, (b) an anisotropic surface, and (c) a surface with a negative skewness of Rsk = −0.7. (From Chen et al., 2007.)

or better and the lateral sampling length as small as a couple of microns. With the assistance of advanced data processing technologies, the measurement time can be as little as 5–30 s, and the large-sized data matrices for 3D topography can be easily stored, processed, and transferred. Figures 2.4 and 2.14 demonstrate some sample 3D surface topographies obtained from optical profilometry.

As mentioned above, surface topographic feature is usually so complicated that one can hardly employ one or a few statistical parameters or even a measured 2D profile to satisfactorily describe its characteristics. In recent years, the rapid developments of computation and information technologies encourage the direct use of digitized 3D topography for deterministic interfacial analysis. However, files storing fully digitized surfaces are often large in size, and analysis technologies are still under development. For problems in nonconformal contacts, the area of surface interaction is usually small, thus, the use of fully digitized surfaces is feasible. For conformal contacts, on the other hand, stochastic models are sometimes more realistic, because the involved surfaces may be so large that the data matrices from their full digitization become impractically complex and difficult to handle.

2.2.4 Rough Surfaces Generated by Computer

In rough surface contact and lubrication analyses, computer-generated surfaces are sometimes used. This type of surfaces may exhibit 3D random topography and also possess prescribed statistical roughness properties, e.g. following a certain required roughness height distribution pattern (such as Gaussian). Since this type of computer-generated rough surfaces has well-controlled statistical features, they are often employed in parametric studies for relating interfacial characteristics to surface statistical properties.

In order to create a random surface topography with prescribed statistical parameters, specific numerical procedures and relevant computer programs have been developed, such as those by Patir (1978), Hu and Tonder (1992), and others. Description of the mathematical algorithms for random surface generation is not a topic to be covered in this book. Readers may refer to the original papers for details if interested.

Figure 2.15 displays three typical examples of computer-generated rough surfaces, obtained by utilizing the numerical procedure developed by Hu and Tonder. Their RMS roughness is prescribed to be $R_q = 0.667\,\mu m$, and the correlation length in the y-direction $l_y^* = 40\,\mu m$. Figure 2.15a shows an isotropic Gaussian surface with its correlation length ratio $\gamma = 1.0$, skewness $R_{sk} = 0$ and Kurtosis $R_{ku} = 3.0$. Figure 2.15b illustrates an anisotropic surface because of the correlation length ratio of $\gamma = 7.0$ prescribed, but the other parameters remain the same as those in Figure 2.15a. Figure 2.15c gives a surface with a negative skewness of $R_{sk} = -0.7$, demonstrating smoother asperity crests together with deeper valleys, as compared with the Gaussian surface shown in Figure 2.15a.

Application examples using computer-generated rough surfaces will be presented in Chapters 3–5 and thereafter.

2.3 LUBRICANT PROPERTIES

It has long been known that lubrication is among the topmost effective and economic means to reduce friction and prevent component failures in machineries. Various lubricants are frequently used at the locations of surface interaction between machine components. The most extensively used lubricants are those refined from crude oils, which are abundant in nature so they are relatively inexpensive. Other commonly employed lubricants include synthetic oils, plant oils, and animal greases. Sometimes even water-based emulsions and gases can be utilized as lubricants under certain conditions. In recent years, different bio-lubricants are increasingly used in industries. In addition, various additives are specially formulated for different purposes and blended into base oils. In commercial market, we can hardly find any liquid lubricants without additives.

Under the circumstances, understanding of lubricant properties becomes a multidisciplinary topic and detailed explanation and extensive discussion may not be allowed in this book. We will only focus on fundamentals of tribological properties of commonly used liquid lubricants in this section for

the purpose of accommodating interfacial mechanics model development to be discussed in this book. Readers may refer to other books, such as the one edited by Mang and Dresel (2001), as well as *Encyclopedia of Tribology* edited by Wang and Chung (2013), for more complete coverage of the lubricant properties if necessary.

2.3.1 Viscosity

Many industrial liquid lubricants exhibit Newtonian behavior under low pressure and low shear strain rate conditions. This means that, based on a laminar flow assumption, the lubricant internal shear stress, τ, caused by the relative motion between each two adjacent layers of lubricant, is proportional to the shear strain rate, du/dz, as depicted in Figure 2.16.

The constant of proportionality, η, depends on the rheological property of lubricant. It is called "dynamic viscosity" or simply "viscosity". This is often considered as the most important tribological property of lubricant, representing the internal resistance to the relative motion.

$$\tau = \eta\dot{\gamma} = \eta\frac{du}{dz} \tag{2.18}$$

$$\eta = \frac{\tau}{\dot{\gamma}} = \frac{\tau}{du/dz} \tag{2.19}$$

The SI unit of dynamic viscosity is N s/m^2 or Pa s. In industries, a different unit, centipoise or *cP* in short, is also commonly used, centipoise = Pa s $\times 10^{-3}$.

In practice, viscosity is usually measured by using a viscometer that allows a specific type of relative motion of the fluid through a confined space or gap of certain geometry under well-controlled temperature and pressure. Measurement results are often in the form of "kinematic viscosity" that is defined as

$$\nu = \frac{\eta}{\rho} \tag{2.20}$$

where ν is the kinematic viscosity and ρ the fluid density at the prevailing temperature. The SI unit of kinematic viscosity is m^2/s or cm^2/s. However, viscosity charts provided by lubricant suppliers often employ centistokes or cSt instead, and the conversion is centistokes = (m^2/s) $\times 10^{-6}$.

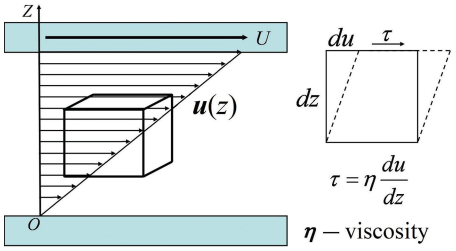


FIGURE 2.16 Definition of viscosity.

2.3.2 Effect of Temperature on Viscosity

It has been well known that viscosity of most industrial lubricants is significantly affected by temperature. A common practice for finding the viscosity–temperature relationship is to measure the viscosity with a capillary tube viscometer under controlled temperatures. Based on a huge amount of testing data, the following Ubbelohde–Walther equation has been generally accepted and also forms the basis of ISO, ASTM, and DIN calculation guidelines.

$$\log\log(\nu + c) = a - b\log T \tag{2.21}$$

where a, b, and c are constants and T the temperature in °K. For mineral oils, c is often between 0.6 and 0.9. Since c plays only a minor role in viscosity calculation and its change is usually limited, it is commonly assumed to be 0.6 or 0.7, based on which a and b can be readily determined by providing two points of measured viscosity at two different temperatures. Table 2.2 shows a sample set of lubricant data provided by a supplier, and Figure 2.17 illustrates typical viscosity–temperature relationships on a log–log scale for some lubricants.

Viscosity–temperature behavior of a lubricant can also be characterized by its "viscosity index (VI)", which has been widely used in industries since its introduction in late 1920s, thus, familiarity with VI is essential when working with industrial lubricants. In the VI calculation, for each measured sample oil viscosity Y at 100°C, we can find two parameters from the tables in ASTM Method D2270, called L and H. L is the 40°C viscosity of a reference oil that has a VI of zero and the same 100°C viscosity Y, and H is the 40°C viscosity of the second reference oil that has its VI of 100 and also the same 100°C viscosity. Assuming U is the 40°C viscosity of the sample oil, its VI can be evaluated by

$$VI = 100\frac{L - U}{L - H} \tag{2.22}$$

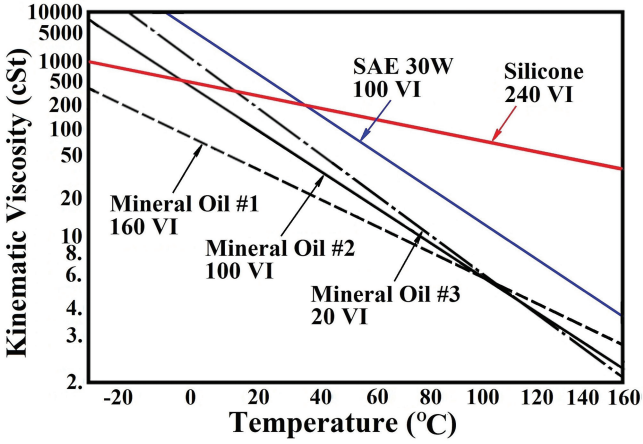


FIGURE 2.17 Sample viscosity–temperature curves.

TABLE 2.2
A Set of Typical Lubricant Property Data

Name		Mobil 1 (SAE Grade: 10W-30)
Viscosity in cSt (ASTM D 445)	40°C	62.0
	100°C	10.0
Viscosity index (ASTM D 2270)		147
Sulfated ash in Wt% (ASTM D 874)		1.0
Pour point in °C (ASTM D 97)		–45
Flash point in °C (ASTM D 92)		224
Density at 15°C in kg/liter (ASMT D 4052)		0.86

Data from Exxon-Mobil.

It is obvious that the greater the *VI*, the less significant the viscosity change with temperature. Note that this calculation method is old and quite awkward, suitable primarily for lubricants with *VI* below 100. Today many good lubricants are available in the market with their *VI* values way above 100, for which the evaluation method has been modified as follows.

$$VI = 100 + 140\left\{\log^{-1}\left[\frac{\log H - \log U}{\log Y}\right] - 1\right\} \tag{2.23}$$

More detailed explanation can be found in ASTM Method D2270. Usually lubricant suppliers provide VI values of their products, as shown in Table 2.2. Typical VI values of some lubricants can be found in Figure 2.17.

Effect of temperature on dynamic viscosity can be approximated by various equations, among which a simple one originally proposed by Reynolds (1886) is given below

$$\eta = \eta_o e^{-\beta \Delta T} \tag{2.24}$$

where η_o is the viscosity at environment temperature T_o, $\Delta T = T - T_o$ the temperature increase, and β the temperature–viscosity exponent. A more accurate evaluation is expressed by the Vogel equation that reads

$$\eta = K e^{\frac{b}{T+c}} \tag{2.25}$$

where three constants, K, b, and c, can be determined by three points of measured viscosity at different temperatures. More detailed discussion can be found in Cameron's book (1966).

2.3.3 Effect of Pressure on Viscosity

Viscosity change due to increase in pressure may be negligible when the pressure is low, but it may become significant if the pressure is elevated. This is extremely important in elastohydrodynamic lubrication (EHL), in which the pressure in the contact zone can possibly be as high as 0.5–4 GPa or so. Unfortunately, measuring lubricant viscosity at such high pressures is a tedious or even difficult task, and lubricant suppliers often do not provide such data for their products. There are several pressure–viscosity models that have been established based on experimental investigations and commonly used mainly in the EHL studies.

The pioneer work by Barus (1893) reported the viscosity data of marine glue as a function of the average pressure in a linear model. Today, however, the following exponential relationship is somehow recognized in tribology as the Barus equation:

$$\eta = \eta_o e^{\alpha p} \tag{2.26}$$

where η_o is the viscosity at ambient pressure and α the pressure–viscosity exponent, which is a property of the lubricant. According to Cameron (1966), et al., the Barus exponential relationship describes the viscosity behavior quite well up to the pressure range of 200–400 MPa at low temperatures. Also, this equation is convenient to use and has been yielding EHL simulation results in reasonably good agreement with measured film thickness data.

When the pressure goes beyond, however, Equation (2.26) may considerably overestimate the viscosity. Modified models have been developed that often contain more than one lubricant property parameters in order to better fit experimental data. One of the modified relationships is a power function presented by Chu and Cameron (1962), which employs two constants, θ and n, to gain more flexibility. It can be expressed as follows:

$$\eta = \eta_o (1 + \theta p)^n \tag{2.27}$$

In principle, constants θ and n can be estimated experimentally, and this relationship may be considerably more accurate if reliable testing data are available. Due to the lack of testing data, Cameron suggested $n = 16$ for commonly used mineral oils, and θ can be estimated with the following formula based on the assumption that Equations (2.26) and (2.27) agree with each other and yield the same viscosity if the pressure is low:

$$\theta = \alpha / n \text{ when } p \to 0 \tag{2.28}$$

Another commonly used relationship was presented by Roelands in his Ph.D. thesis in 1966. Under the isothermal condition, it can be expressed in the current SI unit system as

$$\eta = \eta_o \mathrm{EXP}\left\{\left(\ln\eta_o + 9.67\right)\left[\left(1 + 5.1\times10^{-9}p\right)^Z - 1\right]\right\} \tag{2.29}$$

where Z is the pressure–viscosity index, which can be obtained experimentally. When the measured Z value is not available, it can be estimated with the following formula based on the assumption that both Equations (2.29) and (2.26) are in good agreement at low pressures:

$$Z = \frac{\alpha}{5.1\times10^{-9}\left(\ln\eta_o + 9.67\right)} \tag{2.30}$$

Equations (2.26) and (2.29) have been widely employed by researchers in EHL studies.

For high pressure in excess of 1 GPa, however, Equations (2.27) and (2.29) might still overestimate the viscosity. Based on experimental results, Allen et al. (1970) proposed a two-slope exponential model that reads,

$$\frac{\eta}{\eta_o} = \begin{cases} \mathrm{EXP}(\alpha p), & p < p_a \\ \mathrm{EXP}\left(c_0 + c_1 p + c_2 p^2 + c_3 p^3\right), & p_a \le p \le p_b \\ \mathrm{EXP}\left[\alpha p_t + \alpha_2\left(p - p_t\right)\right], & p > p_b \end{cases} \tag{2.31}$$

where p_t is a "break pressure", beyond which the increase in viscosity changes slope, $p_a = 0.7p_t$ and $p_b = 1.4p_t$. Allen et al. suggested a p_t value of 380 MPa for some typical mineral oils. However, this value may be considerably lower if the initial viscosity is high. Coefficients c_0 through c_3 are so determined that the transition between two slopes is smooth. This model actually employs three lubricant parameters, α, α_2, and p_t, to describe the pressure–viscosity relationship. It may be able to give good predictions for a wide range of pressure if the adequate parameters are available. Unfortunately, its application appears to be quite limited mainly due to its complexity and the lack of reliable testing data for each specific lubricant.

Recently, free-volume models have attracted considerable attention. Doolittle (1951) explored the relationship between the viscosity and the fractional free volume using an exponential function and developed the first free-volume viscosity model, showing that the resistance to flow depends on the relative volume of molecules present per unit free volume. Based on this, improved free-volume viscosity models were developed and the one presented by Cook et al. (1993), is given below.

$$\overline{\eta} = \exp\left\{B\frac{V_{occ}}{V_0}\left[\frac{1}{\dfrac{V}{V_0} - \dfrac{V_{occ}}{V_0}} - \frac{1}{1 - \dfrac{V_{occ}}{V_0}}\right]\right\} \tag{2.32}$$

where V is the volume, V_0, the volume at ambient pressure, V_{occ} the occupied volume, and B the Doolittle parameter. It was found by Y. C. Liu et al. (2006b), and others that the free-volume model appears to be able to yield EHL simulation results closest to experimental data. However, it is not convenient to use this model in engineering practice due to the lack of lubricant property parameters. Therefore, so far, most commonly used viscosity models are still those given by Equations (2.26) and (2.29).

It is important to note that, because experimental evaluation of pressure–viscosity coefficient is a tedious task requiring specific instruments, only a small number of systematic studies have been published, including the well-known testing work reported by the ASME Research Committee on Lubrication in 1953. For convenience, various empirical formulae have been developed in order to estimate lubricant pressure–viscosity coefficient based on commonly available lubricant properties measured at ambient pressure. One of the formulae by Wu et al. (1989) is given below as an example.

$$\alpha = \left(0.1657 + 0.2332\log\nu_o\right)b \tag{2.33}$$

where α is the pressure–viscosity coefficient of lubricant defined in Equation (2.26), ν_o the kinematic viscosity in mm^2/s at ambient pressure, and b the temperature–viscosity property in Equation (2.21). This formula correlates the pressure–viscosity property with the lubricant viscosity and the temperature–viscosity relationship and is easy to use.

Since both Equations (2.24) and (2.26) are exponential relations, they can be readily combined to form an equation describing the viscosity as a function of temperature and pressure:

$$\eta = \eta_o e^{\alpha p - \beta(T - T_o)} \tag{2.34}$$

where η_o is the viscosity at ambient pressure and temperature. This model is simple, based on an assumption that both α and β are independent constants. Figure 2.18 shows a set of sample viscosity data for Mil L23699 lubricant measured by

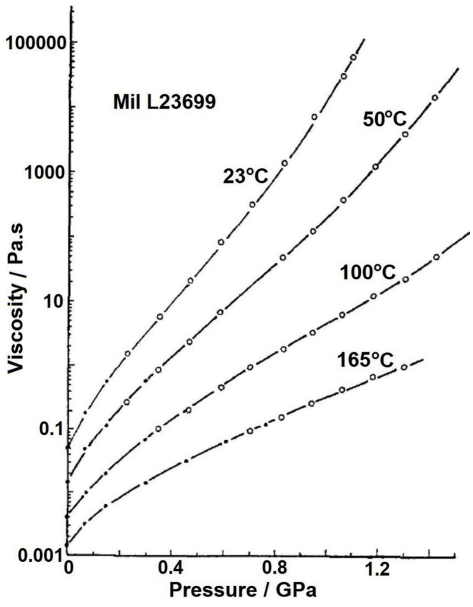


FIGURE 2.18 Typical pressure–viscosity data for Mil L23699 lubricant. (From Bair, 2001.)

Bair (2001), demonstrating the viscosity variations as functions of pressure and temperature.

It is observed that actually pressure–viscosity coefficient α does not appear to be constant. In fact, it not only depends on pressure but also on temperature (or viscosity). A more adequate viscosity model is often preferred that reads,

$$\eta = \eta_o \mathrm{EXP}\left[\alpha p + (\beta + \gamma p)\left(\frac{1}{T} - \frac{1}{T_o}\right)\right] \tag{2.35}$$

where the third lubricant property parameter, γ, is the pressure–temperature–viscosity coefficient. With these three parameters, the formula can better fit testing data such as those shown in Figure 2.18. The Roelands viscosity formula, Equation (2.29), can also be extended to include the temperature effect, expressed as follows (see Roelands, 1966).

$$\eta = \eta_o \mathrm{EXP}\left\{(\ln \eta_o + 9.67)\left[\left(1 + 5.1\times 10^{-9} p\right)^Z \times \left(\frac{T - 138}{T_o - 138}\right)^{-S} - 1\right]\right\} \tag{2.36}$$

where T and T_o are in °K and S is the temperature–viscosity index, which can be estimated by the following formula if there is no measured data available:

$$S = \frac{\beta(T_o - 138)}{\ln \eta_o + 9.67} \tag{2.37}$$

2.3.4 Density

Attention paid to density of liquid lubricants has not been as much as that to viscosity, because density variation and its influence on lubrication are relatively mild. However, knowledge of how the lubricant density varies with pressure and temperature is still necessary for modeling interfacial phenomena, especially those of the elastohydrodynamic lubrication (EHL). Ignoring this property may result in an inaccurate estimate of lubrication characteristics. There are only a limited number of models available for the pressure–density relationship. A widely accepted one proposed by Dowson and Higginson (1966) is written as follows.

$$\rho = \rho_o\left(1 + \frac{C_1 p}{1 + C_2 p}\right) \tag{2.38}$$

where ρ is the density, ρ_o the density under ambient condition, p the pressure in GPa, and C_1 and C_2 are constants. Usually we can take $C_1 = 0.6$ and $C_2 = 1.7$.

If the effect of temperature on density is also considered, Equation (2.38) can be extended to the following form:

$$\rho = \rho_o\left(1 + \frac{C_1 p}{1 + C_2 p}\right)\left[1 - C_3(T - T_o)\right] \tag{2.39}$$

in which T and T_o are in °K and one can often assume $C_3 = 6.5 \times 10^{-4}$ °K^{-1}.

2.3.5 Non-Newtonian Behaviors

The Newtonian assumption results in a constant parameter of viscosity independent of lubricant shear strain rate (but dependent upon temperature and pressure). It is generally acceptable for most industrial lubricants, including mineral and synthetic oils, under low pressure and low shear strain rate conditions. These conditions can be found in many machine elements with surface interaction in conformal contacts, such as journal and thrust bearings, seals, joints, piston and cylinder systems, and so on. Under high pressure and high shear conditions, on the other hand, the Newtonian assumption may no longer be valid, and some polymeric contents in lubricants may increase the possibility of non-Newtonian behaviors such as shear thinning. This type of high pressure and high shear strain rate conditions can be found from components with counterformal contacts, such as gears, rolling bearings, cams and followers, and many others.

It is important to note that even for these components with counterformal contacts, the Newtonian fluid models may still be acceptable in some cases for some specific purposes. A good example is the prediction of EHL lubricant film thickness, which is dominated by lubricant flow conditions in the inlet zone where the pressure and shear rate are relatively low, so that the film thickness predicted by using a Newtonian model may still be satisfactorily accurate. However, the Newtonian models are invalid for predicting friction and flash temperature, which are mainly influenced by the lubricant characteristics under high pressure and high shear conditions in the contact zone. Relevant discussions will be given later in Chapters 7–10.

The non-Newtonian behavior of lubricants is a complex topic still under investigation. The difficulty is mainly due to the fact that in a counterformal contact, the pressure can possibly reach 1–4 GPa, the lubricant shear strain rate can be as high as 10^8 1/s, and the time for lubricant passing through the interaction zone as little as a small fraction of a millisecond. In addition, at high shear rates, the interfacial transient flash temperature rise can possibly be several hundred degrees centigrade. So far, these conditions combined cannot be reproduced in a laboratory with any rheometers or viscometers other than in a highly transient tiny EHL contact. In other words, today it is still very difficult, or even impossible, to directly measure the lubricant rheological properties in a laboratory precisely under the abovementioned EHL conditions. All the rheological models available today, in fact, are still to a large extent hypothetic. Great efforts are still needed in the field of lubricant rheology especially under the EHL conditions.

However, a fundamental understanding has been commonly accepted that, when the shear strain rate is very low, the shear stress is still approximately proportional to the shear strain rate and the constant of proportionality is the viscosity. As the shear rate increases, the viscosity may gradually decrease demonstrating a shear thinning behavior. If the

shear strain rate is further increased, the shear stress will gradually reach its asymptotic limit τ_L, called the "limiting shear stress", as schematically illustrated in Figure 2.19. It is believed that when the shear strain rate goes beyond the limit, slippage occurs either within the lubricant film or at the interface between the lubricant and one of the bounding surfaces, so that the shear stress stops increasing. Note that the limiting shear stress is a property of lubricant and is supposed to be a function of pressure and temperature.

There have been numerous models describing the non-Newtonian phenomena, and detailed discussions on the lubricant rheology are actually more than we can cover in this book. However, a brief review is certainly helpful. One of the commonly used non-Newtonian shear stress–shear strain rate relationship was proposed early by Eyring (1936) and then endorsed by Hirst and Moore (1974), Johnson and Tevaarwerk (1977), and others, which reads,

$$\dot{\gamma} = \frac{du}{dz} = \frac{\tau_E}{\eta}\sinh\left(\frac{\tau}{\tau_E}\right) \tag{2.40}$$

where τ_E is known as the characteristic stress or the Eying stress, beyond which the non-Newtonian behavior becomes significant. This nonlinear viscous model will reduce to the linear Newtonian model, Equation (2.18), when τ_E approaches infinity.

The Eyring model has been widely used in practice. However, in Equation (2.40), the lubricant shear stress increases monotonically with no limitation if the shear strain rate is continuously increased. Bair and Winer (1979) proposed a viscous model based on the limiting shear stress τ_L measured experimentally, which can be expressed as follows:

$$\dot{\gamma} = -\frac{\tau_L}{\eta}\ln\left(1-\frac{\tau}{\tau_L}\right) \tag{2.41}$$

This model has also been enjoying wide application. In 1980, Gecim and Winer proposed a different model expressed as follows

$$\dot{\gamma} = \frac{\tau_L}{\eta}\tanh^{-1}\left(\frac{\tau}{\tau_L}\right) \tag{2.42}$$

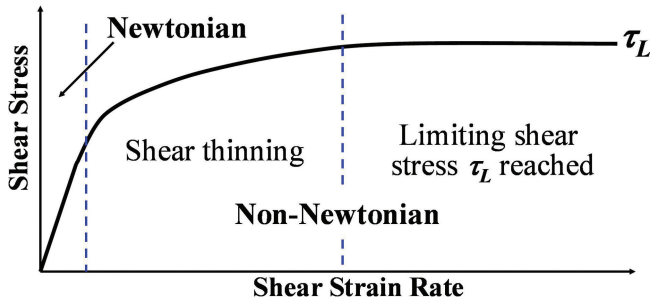


FIGURE 2.19 Schematic of shear stress–shear strain rate relationship.

Later, the research group led by Hamrock presented a few different models in similar forms that include the one by Ivonen and Hamrock (1989):

$$\dot{\gamma} = \frac{\tau_L}{\eta}\left[\left(1-\frac{\tau}{\tau_L}\right)^{-1}-1\right] \tag{2.43}$$

and that by Lee and Hamrock (1990a):

$$\dot{\gamma} = \frac{\tau_L}{\eta}\left[1-\left(\frac{\tau}{\tau_L}\right)^{2}\right]^{-1/2} \tag{2.44}$$

A general model was proposed by Elsharkawy and Hamrock (1991):

$$\dot{\gamma} = \frac{\tau_L}{\eta}\left[1-\left(\frac{\tau}{\tau_L}\right)^{n}\right]^{-1/n} \tag{2.45}$$

All the models mentioned above are illustrated in Figure 2.20 for comparison. These models all employ the concept of limiting shear stress except the one by Eyring.

Recently, considerable attention has been paid to the Carreau model, in which the lubricant is assumed to be an isotropic shear-thinning fluid and its composite shear strain rate, $\dot{\gamma}_e$, and the composite shear stress, τ_e, obey the following relationship (see Carreau, 1972, Tanner, 1985):

$$\dot{\gamma}_e = \frac{\tau_e}{\eta}\left[1+\left(\frac{\tau_e}{G_\infty}\right)^{2}\right]^{(1-n)/(2n)} \tag{2.46}$$

where $\dot{\gamma}_e = \sqrt{\dot{\gamma}_x^2+\dot{\gamma}_y^2}$, $\tau_e = \sqrt{\tau_x^2+\tau_y^2}$, η is the low-shear viscosity, and exponent n is a fluid property independent of operating conditions. Here, the principal axes of stress and strain rate are assumed to be the same.

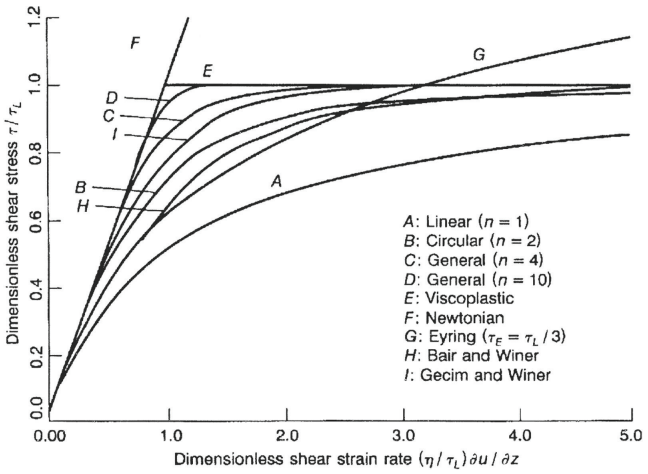


FIGURE 2.20 Non-Newtonian rheological models. (From Hamrock et al., 2004.)

Equations (2.40) through (2.46) describe the viscous behavior of lubricants. It is believed, based on experimental observations, that if the pressure continuously increases, lubricants may start to behave as a viscous-elastic and eventually an elastic solid material at extremely high pressures. The transition from viscous to elastic may be determined by a parameter called the Deborah number, defined as:

$$D = \frac{\eta U}{2aG_\infty} \tag{2.47}$$

where U is the average velocity of the two surfaces, a the half-width or radius of Hertzian contact zone and G_∞ the elastic shear modulus of lubricant. The Deborah number can be recognized as the ratio of the relaxation time for a fluid (η/G) to the time of passage of the fluid through the contact zone ($2a/U$). For Deborah numbers much smaller than 1.0, the lubricant behaves as a viscous fluid, while for Deborah numbers much larger than 1.0, the elastic behavior becomes dominant. When modeling friction (or sometimes called traction) under typical high pressure–high shear rate EHL conditions both viscous and elastic behaviors need to be considered, and the following Maxwell model is so far commonly used:

$$\dot{\gamma} = \dot{\gamma}_e + \dot{\gamma}_v = \frac{1}{G_\infty}\frac{d\tau}{dt} + F(\tau) \tag{2.48}$$

in which a shear stress–shear strain rate relation chosen from Equation (2.40) through (2.46) or others can be applied to the viscous term $F(\tau)$. For example, the following Bair and Winer's model will often be used in this book usually for analyzing frictional and thermal behaviors:

$$\dot{\gamma} = \dot{\gamma}_e + \dot{\gamma}_v = \frac{1}{G_\infty}\frac{d\tau}{dt} - \frac{\tau_L}{\eta}\ln\left(1 - \frac{\tau}{\tau_L}\right) \tag{2.49}$$

Readers may refer to Tanner (1985), Bair (2007), and other publications for more detailed discussion on lubricant rheology.

2.3.6 Additives in Lubricants

A lubricant possibly performs some or all of the following important functions during its engineering application:

1. Fluid-film lubrication. It reduces friction and wear and protects the mechanical components by introducing a fluid film between the surfaces to avoid or minimize possible contact.
2. Boundary film lubrication. It helps to form a stable boundary film that is desirable for reducing friction at asperity contact locations and prolonging components life against surface failures such as wear, pitting, and scuffing.
3. Cooling. It helps to dissipate heat away from the critical interfaces in the system.
4. Cleaning and suspending. These functions facilitate a smooth operation of the system by removing and suspending foreign particles, dirt, deposits, oxides, rust, and wear debris.
5. Anti-corrosion and anti-oxidation.

Untreated base oils without any additive do not possess the necessary properties to be effective in today's demanding industrial environment. Additives can improve the lubricating ability of the base oil by either enhancing the desirable properties already present or adding new functionalities. Proper application of additives may significantly enhance the performance of lubricants in most, or all, of the function categories listed above. Today, we can hardly find any neat base oils without additives in the commercial market.

Usually, many lubricant additive molecules, except perhaps those of some viscosity improvers and pour point depressants, consist of an oleophilic hydrocarbon group (that helps to improve the solubility) and a hetero-atom (nitrogen, oxygen, sulfur, and phosphorus)-based polar surface active group that make it easy for the additive molecules to be adsorbed to the surface, and form a boundary film (see the sketch shown in Figure 2.21). A proper balance between polar and non-polar characteristics is critical to the additive performance. Refer to Section 10.3 for more discussion.

A typical additive package may contain the following:

1. Boundary film additives, including wear inhibitors, lubricity agents, friction modifiers, extreme pressure (EP) additives, and others. These additives increase the probability of boundary film formation and therefore help to reduce friction and the risk of various surface failures.
2. Corrosion and oxidation inhibitors. These additives aim at preventing surface corrosion (the former) and retarding oil degradation (the latter).
3. Detergents and dispersants. The former are used to clean and neutralize oil impurities to avoid deposits, while the latter keep contaminants suspended in the oil to prevent particles from coagulating.
4. Viscosity improvers, which are usually polymeric molecules of various sizes and shapes, sensitive to temperature. They are used to improve VI of a lubricant. It makes possible to use less viscous base oils to reduce viscous dissipation, especially at low temperature.

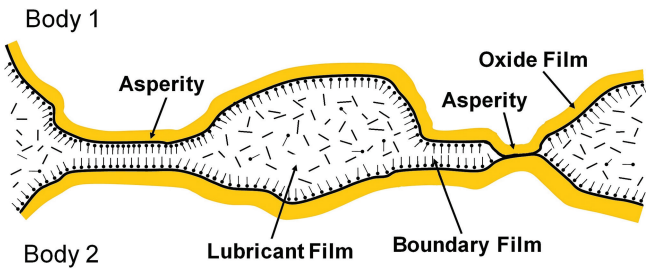


FIGURE 2.21 Boundary film formation.

5. Emulsifiers and demulsifiers. The former helps to form emulsion lubricants, while the latter inhibits the production of air bubbles and foams in the lubricant to prevent cavitation-induced problems.
6. Biocides, used for controlling the activity of microorganisms and extending the useful life of a lubricant.
7. Others, including pour point depressant, foam inhibitors, seal-swell agents, and dyes.

Usually additives are well blended and dissolved into the base oils, and they take typically several percent or even 10%–15% in volume. In many cases, the additives do not change the lubricant viscosity and density significantly. Moreover, the boundary films are often only a few or several layers of molecules adsorbed to the surface. That is why traditionally they are ignored in macro-scale contact and hydrodynamic lubrication modeling, which is based on continuum mechanics.

However, additives may often have a great influence on the boundary lubrication behaviors. At present, more detailed discussion on additives is beyond the scope of this book.

3 Fundamentals of Contact Mechanics

3.1 INTRODUCTION

As explained in Chapter 1, a dry contact is a special case of lubricated contact. Elastic dry contact is an important and fundamental problem not only because its mechanism is relatively simple that is considered as a foundation for understanding more complex interfacial phenomena but also due to the fact that many practical issues can be simplified to and treated as those of dry elastic contacts when the deformation is small and the influence of hydrodynamic lubrication is weak or none at the interface.

It is well accepted that the foundation of contact mechanics was laid nearly 140 years ago symbolized by the work of Hertz (1881, 1882), who investigated frictionless counterformal dry contact between two homogeneous elastic solid bodies of simple geometry with ideally smooth surfaces. The Hertzian theory has been widely used in engineering practice due to its simplicity and satisfactory accuracy for contact pressure and deformation calculations. Also, the classic solutions for displacements and stresses caused by normal and tangential tractions applied on the surface of an elastic half-space were derived originally by Cerruti (1882) and Boussinesq (1885) based on the potential theory. These solutions were later well described and advanced by Love (1927, 1929, 1952), Ling (1973), Johnson (1987), Hills et al. (1993), and others.

Over the years, a number of analytical solutions to non-Hertzian elastic contact problems have been developed based on the classic theories stated above. These analytical solutions (many of which are summarized in the book by Johnson, 1987) are usually for problems with simple contact geometries and idealized material properties and operating conditions. Moreover, the theories of contact yield have been established and the approximate stochastic models of rough surface interaction derived by Greenwood and Williamson (1966), Greenwood and Tripp (1970–1971), and others.

The classic theories have also been extended to solve contact problems involving layered materials, such as solid bodies with coatings. In this area, the pioneer studies were conducted by Burmister (1945), W. T. Chen (1971), Chen and Engel (1972), et al., and most of the solutions were pursued by means of Fourier transform. Many researchers have contributed to the solution methods for contact problems with single-layered materials (e.g. O'Sullivan and King, 1988, Nogi and Kato, 1997, Polonsky and Keer, 2000a, S. B. Liu and Wang, 2002), bi- or tri-layered materials (Cai and Bhushan, 2005, C. J. Yu et al., 2013, and others), and multilayer coatings (Kuo and Keer, 1992, Plumet and Dubourg, 1998, et al.). More recently, C. J. Yu et al. (2014) derived the first set of explicit expressions for the contact elastic field (stresses and displacements) in the frequency domain with analytical coefficients given in reoccurring formulae for multilayered material systems. H. B. Zhang et al. (2018a, b) extended this method to the frequency-domain solution for frictional heat conduction in multilayered materials.

Today, many analytical formulae are the kernels of criteria for advanced mechanical component design, and many core solutions in the form of Green's function and convolution integrals have become the backbone of accurate and efficient numerical solution methods for complicated engineering problems.

This chapter focuses on the basic principles of elastic contact mechanics, the classic Hertzian theories, the analytical solutions to several typical non-Hertzian problems, the statistical model approach and the simulation-based curve-fitting formulae for rough surface contacts, the analyses of initial contact yield for homogeneous and case-hardened materials, and the derivation of the solution for the elastic field in contacting multilayered materials. In this chapter, the influences of hydrodynamic lubrication, material inhomogeneity and inelasticity, boundary layers, wear, and thermal behaviors are not considered. The Cartesian coordinate system is used in terms of (x, y, z), double subscripts i, j refer to x, y, z, while the single subscript, i, may indicate x, y, z, or a contact body, or may be used to number functions.

3.2 BASIC HALF-SPACE ELASTICITY THEORIES

In a non-conformal contact between two elastic bodies, the size of surface interaction zone is usually small in comparison with the dimensions of the contacting bodies and the radii of curvature of the two surfaces at the contact. The pressure is possibly high, concentrated in this small contact area, and its influence rapidly decays with distance. Therefore, each of the bodies can be simply treated as a half-space solid. The contact area on one surface is the same as that on the other surface, so are the contact pressures as the action and the reaction. Generally, the fundamental problems include how to analyze interaction between the two bodies based on the elastic half-space assumption and how to calculate interfacial pressure, deformations, and stresses in the contacting bodies due to the application of concentrated or distributed normal and tangential loads.

3.2.1 Potential Equations

The surface of an elastic half-space solid, having its Young's modulus E and Poison's ratio ν, may be under arbitrarily distributed normal and tangential loading. The area subject to load application is marked by S, as shown in Figure 3.1. In other words, there is no load outside S. The normal pressure, $p(\xi,\eta)$, and the tangential tractions in the two orthogonal directions, $q_x(\xi,\eta)$ and $q_y(\xi,\eta)$, applied at point (ξ,η) on

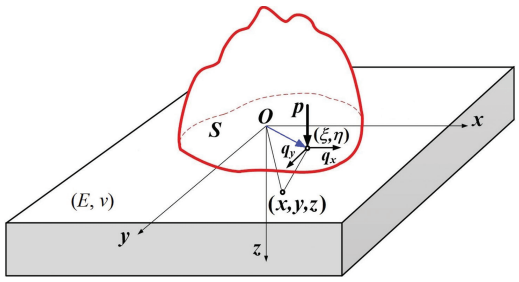


FIGURE 3.1 Elastic half-space under normal and tangential tractions.

the surface, called the loading point, are also illustrated in the figure. The load can be expressed as $\boldsymbol{q} = \{q_1, q_2, q_3\}$ or $\{q_x, q_y, -p\}$. Now the problem is how to calculate displacements and stresses at any point $x_j = (x, y, z)$, called the response point, within the half-space solid due to the distributed load, $\boldsymbol{q}$, applied in domain S on the surface.

Note that the distance between the loading point $(\xi, \eta, 0)$ and the response point (x, y, z) is $R = \sqrt{(\xi - x)^2 + (\eta - y)^2 + z^2}$. In the following, the elastic field at the response point is analyzed in terms of material displacements and stresses.

The load applied in surface area S will cause the material to deform and to be stressed. If the displacements are u_i, with $i = x, y, z$, and the stresses are σ_{ij}, with $i, j = x, y, z$, the displacements and stresses are related in the way as follows according to the theory of linear elasticity (see Timoshenko and Goodier, 1970). Basic stress and strain expressions in linear elasticity are briefly summarized in Appendix A.

$$\sigma_{xx}(x,y,z) = \frac{2\nu G_s}{1-2\nu}\left(\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} + \frac{\partial u_z}{\partial z}\right) + 2G_s\frac{\partial u_x}{\partial x} \quad (3.1)$$

$$\sigma_{yy}(x,y,z) = \frac{2\nu G_s}{1-2\nu}\left(\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} + \frac{\partial u_z}{\partial z}\right) + 2G_s\frac{\partial u_y}{\partial y} \quad (3.2)$$

$$\sigma_{zz}(x,y,z) = \frac{2\nu G_s}{1-2\nu}\left(\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} + \frac{\partial u_z}{\partial z}\right) + 2G_s\frac{\partial u_z}{\partial z} \quad (3.3)$$

$$\sigma_{xy}(x,y,z) = \tau_{xy}(x,y,z) = G_s\left(\frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x}\right) \quad (3.4)$$

$$\sigma_{yz}(x,y,z) = \tau_{yz}(x,y,z) = G_s\left(\frac{\partial u_y}{\partial z} + \frac{\partial u_z}{\partial y}\right) \quad (3.5)$$

$$\sigma_{xz}(x,y,z) = \tau_{xz}(x,y,z) = G_s\left(\frac{\partial u_x}{\partial z} + \frac{\partial u_z}{\partial x}\right) \quad (3.6)$$

where $G_s = \dfrac{E}{2(1+\nu)}$ is the elastic shear modulus.

The displacement solution through the potential theories provides a systematic approach to solve the elastic field with the model shown in Figure 3.1. According to Love (1952) and Johnson (1987), several potential functions, $F_i = \{F_1, F_2, F_3\} = \{F_1, G_1, H_1\}$, are defined and used in contact elasticity:

$$F_i(x,y,z) = \iint_S q_i(\xi,\eta)\Omega\, d\xi\, d\eta$$

$$\text{or } F_1(x,y,z) = \iint_S q_x(\xi,\eta)\Omega\, d\xi\, d\eta \quad (3.7)$$

$$F_2(x,y,z) = \iint_S q_y(\xi,\eta)\Omega\, d\xi\, d\eta \quad (3.8)$$

$$F_3(x,y,z) = \iint_S p(\xi,\eta)\Omega\, d\xi\, d\eta \quad (3.9)$$

where $\Omega = z\ln(R + z) - R$ is a harmonic function of the coordinates and $\dfrac{\partial \Omega}{\partial z} = \ln(R+z) + \dfrac{z}{R+z}\left(\dfrac{z}{R} + 1\right) - \dfrac{z}{R} = \ln(R+z)$. Each of the functions F_i satisfies the Laplace equation $\nabla^2 F_i = 0$. Moreover, we can define potential functions, $G_i = \{F, G, H\} = \left\{\dfrac{\partial F_1}{\partial z}, \dfrac{\partial G_1}{\partial z}, \dfrac{\partial H_1}{\partial z}\right\} = \left\{\dfrac{\partial F_i}{\partial z}\right\}$, as follows

$$G_i(x,y,z) = \frac{\partial F_i}{\partial z} = \iint_S q_i(\xi,\eta)\ln(R+z)\, d\xi\, d\eta$$

$$\text{i.e. } G_1(x,y,z) = \frac{\partial F_1}{\partial z} = F = \iint_S q_x(\xi,\eta)\ln(R+z)\, d\xi\, d\eta \quad (3.10)$$

$$G_2(x,y,z) = \frac{\partial F_2}{\partial z} = G = \frac{\partial G_1}{\partial z} = \iint_S q_y(\xi,\eta)\ln(R+z)\, d\xi\, d\eta \quad (3.11)$$

$$G_3(x,y,z) = \frac{\partial F_3}{\partial z} = H = \frac{\partial H_1}{\partial z} = \iint_S p(\xi,\eta)\ln(R+z)\, d\xi\, d\eta \quad (3.12)$$

By using the potentials given above, the following are further defined:

$$\psi_1 = \frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z} = \frac{\partial F_1}{\partial x} + \frac{\partial G_1}{\partial y} + \frac{\partial H_1}{\partial z} \quad (3.13)$$

$$\psi = \frac{\partial \psi_1}{\partial z} = \frac{\partial G_1}{\partial x} + \frac{\partial G_2}{\partial y} + \frac{\partial G_3}{\partial z} = \frac{\partial F}{\partial x} + \frac{\partial G}{\partial y} + \frac{\partial H}{\partial z} \quad (3.14)$$

The displacements at any point (x, y, z) within the elastic half-space solid can be expressed as follows according to Love (1952):

$$u_x(x,y,z) = \frac{1}{4\pi G_s}\left(2\frac{\partial G_1}{\partial z} - \frac{\partial G_3}{\partial x} + 2\nu\frac{\partial \psi_1}{\partial x} - z\frac{\partial \psi}{\partial x}\right) \quad (3.15)$$

$$u_y(x,y,z) = \frac{1}{4\pi G_s}\left(2\frac{\partial G_2}{\partial z} - \frac{\partial G_3}{\partial y} + 2\nu\frac{\partial \psi_1}{\partial y} - z\frac{\partial \psi}{\partial y}\right) \quad (3.16)$$

$$u_z(x,y,z)=\frac{1}{4\pi G_s}\left(\frac{\partial G_3}{\partial z}+(1-2\nu)\psi-z\frac{\partial \psi}{\partial z}\right) \quad (3.17)$$

The displacement solutions to Equations (3.15)–(3.17) can be used to calculate stresses at any point (x, y, z) within the half-space solid via Equations (3.1)–(3.6).

Generally, the loading on the surface may consist of normal pressure p and tangential tractions q_x and q_y. However, q_x and q_y can often be combined into a resultant traction along a properly selected coordinate axis, proper selection of the coordinate system, thus, the solution expressions can be simplified. Assuming that the traction is in the x-direction, we can have $G_2 = F_2 = 0$. Therefore, Equations (3.13) and (3.14) can be re-written as:

$$\psi_1(x,y,z)=\frac{\partial F_1}{\partial x}+\frac{\partial F_3}{\partial z} \quad (3.18)$$

$$\psi(x,y,z)=\frac{\partial \psi_1}{\partial z}=\frac{\partial G_1}{\partial x}+\frac{\partial G_3}{\partial z} \quad (3.19)$$

3.2.2 Displacement Due to Normal Loading

If only normal pressure $p(x, y)$ is applied, one has $F_1 = F_2 = G_1 = G_2 = 0$. One also has $\frac{\partial}{\partial z}\ln(R+z)=\frac{1}{R+z}\left(\frac{z}{R}+1\right)=\frac{1}{R}$, hence, Equations (3.18) and (3.19) become

$$\psi_1(x,y,z)=\frac{\partial F_3}{\partial z}=G_3=\iint_S p(\xi,\eta)\ln(R+z)\,d\xi\,d\eta \quad (3.20)$$

$$\psi(x,y,z)=\frac{\partial \psi_1}{\partial z}=\frac{\partial G_3}{\partial z}=\frac{\partial}{\partial z}\iint_S p(\xi,\eta)\ln(R+z)\,d\xi\,d\eta$$

$$=\iint_S \frac{p(\xi,\eta)}{R}\,d\xi\,d\eta \quad (3.21)$$

With $\frac{\partial G_3}{\partial x}=\frac{\partial}{\partial x}\frac{\partial F_3}{\partial z}=\frac{\partial \psi_1}{\partial x}$ and $\frac{\partial G_3}{\partial y}=\frac{\partial}{\partial y}\frac{\partial F_3}{\partial z}=\frac{\partial \psi_1}{\partial y}$, displacements at the response point, (x, y, z), can be given by

$$u_x^p(x,y,z)=\frac{1}{4\pi G_s}\left\{-\frac{\partial G_3}{\partial x}+2\nu\frac{\partial \psi_1}{\partial x}-z\frac{\partial \psi}{\partial x}\right\}$$

$$=-\frac{1}{4\pi G_s}\left\{(1-2\nu)\frac{\partial \psi_1}{\partial x}+z\frac{\partial \psi}{\partial x}\right\} \quad (3.22)$$

$$u_y^p(x,y,z)=\frac{1}{4\pi G_s}\left\{-\frac{\partial G_3}{\partial y}+2\nu\frac{\partial \psi_1}{\partial y}-z\frac{\partial \psi}{\partial y}\right\}$$

$$=-\frac{1}{4\pi G_s}\left\{(1-2\nu)\frac{\partial \psi_1}{\partial y}+z\frac{\partial \psi}{\partial y}\right\} \quad (3.23)$$

$$u_z^p(x,y,z)=\frac{1}{4\pi G_s}\left\{\frac{\partial G_3}{\partial z}+(1-2\nu)\psi-z\frac{\partial \psi}{\partial z}\right\}$$

$$=\frac{1}{4\pi G_s}\left\{2(1-\nu)\psi-z\frac{\partial \psi}{\partial z}\right\} \quad (3.24)$$

The surface displacements can be determined by setting $z = 0$, expressed as follows:

$$u_x^p(x,y,0)=-\frac{1}{4\pi G_s}\left\{(1-2\nu)\frac{\partial \psi_1}{\partial x}+z\frac{\partial \psi}{\partial x}\right\}_{z=0}$$

$$=-\frac{1-2\nu}{4\pi G_s}\left\{\frac{\partial \psi_1}{\partial x}\right\}_{z=0} \quad (3.25)$$

$$u_y^p(x,y,0)=-\frac{1}{4\pi G_s}\left\{(1-2\nu)\frac{\partial \psi_1}{\partial y}+z\frac{\partial \psi}{\partial y}\right\}_{z=0}$$

$$=-\frac{1-2\nu}{4\pi G_s}\left\{\frac{\partial \psi_1}{\partial y}\right\}_{z=0} \quad (3.26)$$

$$u_z^p(x,y,0)=\frac{1}{4\pi G_s}\left\{2(1-\nu)\psi-z\frac{\partial \psi}{\partial z}\right\}_{z=0}=\frac{(1-\nu)}{2\pi G_s}\{\psi\}_{z=0}$$

$$=\frac{(1-\nu)}{2\pi G_s}\iint_S \frac{p(\xi,\eta)}{R}\,d\xi\,d\eta$$

Because $G_s = E/[2(1+\nu)]$, the above equation for normal displacement u_z^p can be written as

$$u_z^p(x,y)=\frac{\left(1-\nu^2\right)}{\pi E}\iint_S \frac{p(\xi,\eta)}{\sqrt{(x-\xi)^2+(y-\eta)^2}}\,d\xi d\eta \quad (3.27.1)$$

This equation can also be expressed in the following form

$$u_z^p(x,y)=\iint_S G^p(x-\xi,y-\eta)\,p(\xi,\eta)\,d\xi\,d\eta \quad (3.27.2)$$

in which G^p is Green's function defined as

$$G^p(x,y)=\frac{\left(1-\nu^2\right)}{\pi E r}=\frac{\left(1-\nu^2\right)}{\pi E}\cdot\frac{1}{\sqrt{x^2+y^2}} \quad (3.28)$$

where $r=\sqrt{x^2+y^2}$. Note that any calculation through the superposition principle for linear operator problems, such as a convolution integration based on the influence of a point source phenomenon, can adopt the concept of Green's function, which was originally introduced by an English mathematician, George Green, in 1828. Sometimes Green's function is also called the "influence function".

Equation (3.27) is a convolution integral, commonly known as the Boussinesq integral that is widely used for surface normal deformation calculation. It is important to note that this convolution equation usually involves a large amount of numerical computation if the entire distribution of surface deformation, $u_z^p(x,y)$, is sought. Detailed numerical algorithms for efficient computation will be discussed later in the next chapter.

When two surfaces with different material properties, E_1, ν_1, and E_2, ν_2, are in contact, producing a distributed contact pressure, p, the total normal displacement is the sum of those

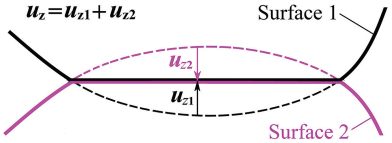


FIGURE 3.2 Total displacement of two surfaces in a contact.

of the two surfaces, as sketched in Figure 3.2 and expressed by the equation below.

$$u_z^p(x,y) = \frac{\left(1-\nu_1^2\right)}{\pi E_1}\iint_S \frac{p(\xi,\eta)}{\sqrt{(x-\xi)^2+(y-\eta)^2}}\,d\xi d\eta + \frac{\left(1-\nu_2^2\right)}{\pi E_2}\iint_S \frac{p(\xi,\eta)}{\sqrt{(x-\xi)^2+(y-\eta)^2}}\,d\xi d\eta = \left[\frac{\left(1-\nu_1^2\right)}{\pi E_1}+\frac{\left(1-\nu_2^2\right)}{\pi E_2}\right]\iint_S \frac{p(\xi,\eta)}{\sqrt{(x-\xi)^2+(y-\eta)^2}}\,d\xi d\eta$$

$$\text{i.e. } u_z^p(x,y) = \frac{2}{\pi E'}\iint_S \frac{p(\xi,\eta)}{\sqrt{(x-\xi)^2+(y-\eta)^2}}\,d\xi d\eta \quad (3.29)$$

where E' is called the effective elastic modulus or effective Young's modulus, defined as

$$E' = 2\left(\frac{1-\nu_1{}^2}{E_1}+\frac{1-\nu_2{}^2}{E_2}\right)^{-1} \quad (3.30.1)$$

Sometimes a slightly different parameter, E^*, may be used. It is often called equivalent Young's modulus and defined as follows:

$$E^* = \left(\frac{1-\nu_1{}^2}{E_1}+\frac{1-\nu_2{}^2}{E_2}\right)^{-1} = E'/2 \quad (3.30.2)$$

For line contact problems under the plane strain condition, assuming p is only a function of x (or ξ) but constant in the y- (or η-) direction, and is distributed within the domain S_x, as sketched in Figure 3.3, we can have an incremental

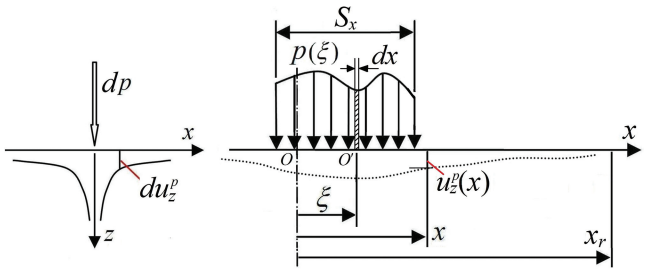


FIGURE 3.3 Deformation due to distributed pressure within domain S_x. At reference point x_r, the deformation approaches zero.

displacement, du_z^p, at response point x, given as follows, if an elemental line loading $dp = pdx$ is applied at the coordinate origin (see Timoshenko and Goodier, 1970):

$$du_z^p(x) = -\frac{2\left(1-\nu^2\right)pdx}{\pi E}\ln|x| + C'$$

where constant C' can be determined by choosing a reference point x_r at the surface that is sufficiently far away from the loading point. Thus, we have $du_z^p(x_r) = 0$, and then

$$C' = \frac{2\left(1-\nu^2\right)pdx}{\pi E}\ln|x_r|$$

$$du_z^p(x) = -\frac{2\left(1-\nu^2\right)pdx}{\pi E}\ln|x| + \frac{2\left(1-\nu^2\right)pdx}{\pi E}\ln|x_r| = -\frac{2\left(1-\nu^2\right)pdx}{\pi E}\ln\left|\frac{x}{x_r}\right|$$

Integrating the above equation over entire domain S_x yields

$$u_z^p(x) = -\frac{2\left(1-\nu^2\right)}{\pi E}\int_{S_x} p(\xi)\ln\left|\frac{x-\xi}{x_r-\xi}\right|d\xi \quad (3.31)$$

Equation (3.31) is recognized as the Flamant solution (see Flamant, 1892), which is widely used to calculate surface normal deformation for line contact problems. For two surfaces in a line contact, the total surface displacement can be written as

$$u_z^p(x) = -\frac{4}{\pi E'}\int_{S_x} p(\xi)\ln\left|\frac{x-\xi}{x_r-\xi}\right|d\xi \quad (3.32.1)$$

The sign of the absolute value can be removed for convenience in calculation, thus, the above equation can be expressed in a different form as follows, which is frequently used in practice:

$$u_z^p(x) = -\frac{2}{\pi E'}\int_{S_x} p(\xi)\ln(x-\xi)^2\,d\xi + C \quad (3.32.2)$$

Note that integral constant C can be determined during an iterative procedure for solving a contact problem subjected to load balance.

3.2.3 Displacement Due to Tangential Traction

Under the action of a pure tangential traction, e.g. $q_x(\xi, \eta)$, we have $G_2 = F_2 = G_3 = F_3 = 0$, then Equations (3.18) and (3.19) reduce to

$$\psi = \frac{\partial G_1}{\partial x} = \frac{\partial^2 F_1}{\partial x\,\partial z} \quad (3.33)$$

$$\psi_1 = \frac{\partial F_1}{\partial x} \quad (3.34)$$

Again, $\Omega = z\ln(R+z) - R$, $\frac{\partial \Omega}{\partial z} = \ln(R+z)$, and $\frac{\partial^2 \Omega}{\partial z^2} = \frac{\partial}{\partial z}\ln(R+z) = \frac{1}{R}$; we can further calculate the following Ω derivatives:

$$\frac{\partial \Omega}{\partial x} = -\frac{z(\xi - x)}{R(R+z)} + \frac{\xi - x}{R}$$

$$\frac{\partial^2 \Omega}{\partial z \partial x} = -\frac{1}{R+z}\cdot\frac{\xi - x}{R}$$

$$\frac{\partial^3 \Omega}{\partial z^2 \partial x} = \frac{\xi - x}{R^3}$$

Similarly, other derivatives can be derived. With all the related derivatives above, and $G_2 = F_2 = G_3 = F_3 = 0$, the displacements at (x, y, z) due to tangential traction q_x can be expressed as

$$u_x^{q_x}(x,y,z) = \frac{1}{4\pi G_s}\left(2\frac{\partial G_1}{\partial z} + \frac{\partial G_3}{\partial x} + 2\nu\frac{\partial \psi_1}{\partial x} - z\frac{\partial \psi}{\partial x}\right) = \frac{1}{4\pi G_s}\left(2\frac{\partial G_1}{\partial z} + 2\nu\frac{\partial \psi_1}{\partial x} - z\frac{\partial \psi}{\partial x}\right)$$

$$= \frac{1}{4\pi G_s}\left(2\frac{\partial^2 F_1}{\partial z^2} + 2\nu\frac{\partial^2 F_1}{\partial x^2} - z\frac{\partial^3 F_1}{\partial x^2 \partial z}\right) \tag{3.35.1}$$

$$= \frac{1}{4\pi G_s}\iint_S q_x(\xi,\eta)\left[\frac{1}{R} + \frac{1-2\nu}{R+z} + \frac{(\xi - x)^2}{R^3} - \frac{(1-2\nu)(\xi - x)^2}{R(R+z)^2}\right]d\xi d\eta$$

$$u_y^{q_x}(x,y,z) = \frac{1}{4\pi G_s}\left(2\nu\frac{\partial \psi_1}{\partial y} - z\frac{\partial \psi}{\partial y}\right) = \frac{1}{4\pi G_s}\left(2\nu\frac{\partial^2 F_1}{\partial y \partial x} - z\frac{\partial^3 F_1}{\partial y \partial x \partial z}\right)$$

$$= \frac{1}{4\pi G_s}\iint_S q_x(\xi,\eta)\left[\frac{(\xi - x)(\eta - y)}{R^3} - \frac{(1-2\nu)(\xi - x)(\eta - y)}{R(R+z)^2}\right]d\xi d\eta \tag{3.35.2}$$

$$u_z^{q_x}(x,y,z) = \frac{1}{4\pi G_s}\left[\frac{\partial G_3}{\partial z} + (1-2\nu)\psi - z\frac{\partial \psi}{\partial z}\right] = \frac{1}{4\pi G_s}\left[(1-2\nu)\frac{\partial^2 F_1}{\partial x \partial z} - z\frac{\partial^3 F_1}{\partial x \partial z^2}\right]$$

$$= -\frac{1}{4\pi G_s}\iint_S q_x(\xi,\eta)\left[\frac{(\xi - x)z}{R^3} + \frac{(1-2\nu)(\xi - x)}{R(R+z)}\right]d\xi d\eta \tag{3.35.3}$$

The surface normal displacement is obtained through setting $z = 0$ in Equation (3.35.3).

$$u_z^{q_x}(x,y,0) = -\frac{1}{4\pi G_s}\iint_S q_x(\xi,\eta)\left[\frac{(\xi - x)z}{R^3} + \frac{(1-2\nu)(\xi - x)}{R(R+z)}\right]_{z=0}$$

$$= \frac{(1-2\nu)}{4\pi G_s}\iint_S q_x(\xi,\eta)\frac{(x - \xi)}{R^2}d\xi d\eta$$

$$\text{i.e. } u_z^{q_x}(x,y) = \frac{(1-2\nu)(1+\nu)}{2\pi E}\iint_S \frac{(x - \xi)q_x(\xi,\eta)}{(x-\xi)^2 + (y-\eta)^2}d\xi d\eta \tag{3.36.1}$$

This equation can also be written as

$$u_z^{q_x}(x,y) = \iint_S G^s(x - \xi, y - \eta)q_x(\xi,\eta)d\xi d\eta \tag{3.36.2}$$

in which Green's function is defined as

$$G^s(x,y) = \frac{(1-2\nu)(1+\nu)x}{2\pi E r^2} = \frac{(1-2\nu)(1+\nu)}{2\pi E}\cdot\frac{x}{x^2 + y^2} \tag{3.37}$$

The solution to a concentrated tangential traction at a point on the surface of a half-space is called the Cerruti solution, and Equation (3.36) is commonly known as the Cerruti integral that is widely used for calculating surface normal displacement due to a distributed tangential load. When two surfaces of different material properties, E_1, ν_1 and E_2, ν_2, are in contact and subject to tangential traction q_x, the total normal displacement is the sum of those of the two surfaces, as given below. It is important to note that the tractions at a certain location (ξ, η) on the two mating surfaces are usually the same in absolute value but opposite in direction.

$$u_z^{q_x} = u_{z1}^{q_x} + u_{z2}^{q_x}$$

$$= \frac{(1-2\nu_1)(1+\nu_1)}{2\pi E_1}\iint_S \frac{(x - \xi)q_x(\xi,\eta)}{(x-\xi)^2 + (y-\eta)^2}d\xi d\eta$$

$$+ \frac{(1-2\nu_2)(1+\nu_2)}{2\pi E_2}\iint_S \frac{(x - \xi)\left(-q_x(\xi,\eta)\right)}{(x-\xi)^2 + (y-\eta)^2}d\xi d\eta$$

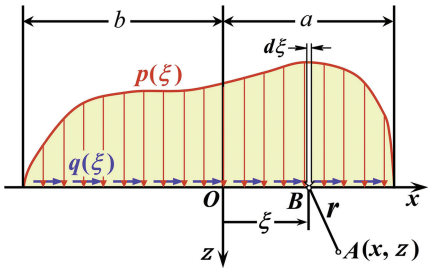


FIGURE 3.4 Normal pressure $p(x)$ and tangential traction $q(x)$ distributed on the surface within the strip ($-b \le x \le a$).

$$\text{i.e. } u_z^{q_x} = \left[\frac{(1-2\nu_1)(1+\nu_1)}{2\pi E_1} - \frac{(1-2\nu_2)(1+\nu_2)}{2\pi E_2}\right] \times \iint_S \frac{(x-\xi)q_x(\xi,\eta)}{(x-\xi)^2+(y-\eta)^2}\,d\xi\,d\eta \tag{3.38}$$

The equation above indicates that the total displacement is zero when the two materials have the same elastic properties.

When both normal and tangential loads are applied at the interface, the surface deformation in the z-direction can be computed numerically through the following Boussinesq–Cerruti integral equation:

$$u_z = u_z^p + u_z^{q_x} = \frac{2}{\pi E'}\iint_S \frac{p(\xi,\eta)}{\sqrt{(x-\xi)^2+(y-\eta)^2}}\,d\xi d\eta + \left[\frac{(1-2\nu_1)(1+\nu_1)}{2\pi E_1} - \frac{(1-2\nu_2)(1+\nu_2)}{2\pi E_2}\right] \times \iint_S \frac{(x-\xi)q_x(\xi,\eta)}{(x-\xi)^2+(y-\eta)^2}\,d\xi\,d\eta \tag{3.39}$$

For line contact problems, if a distributed tangential traction $q(x)$ is applied on the surface of a half-space within a strip $[-b, a]$ under the plane strain condition, as sketched in Figure 3.4, the normal displacement u_z^q at response point $x \in (-b, a)$ on the surface can be expressed as follows (see Johnson, 1987):

$$u_z^q(x) = \frac{(1-2\nu)(1+\nu)}{2E}\left[\int_{-b}^{x} q(\xi)\,d\xi - \int_{x}^{a} q(\xi)\,d\xi\right] + C \tag{3.40}$$

3.2.4 General Equations for Surface Displacements

More generally, the surface displacements, expressed below as u_x, u_y, and u_z in the x-, y-, and z-directions, respectively, due to tangential tractions q_x and q_y and normal pressure p over the contact area can be formulated through Equations (3.22)–(3.24) and (3.35) by setting $z = 0$ and summing up the effects of the surface tractions (see Hills et al., 1993, W. W. Chen and Wang, 2008a). The deformations due to q_x and q_y can both be determined through Equation (3.35) by exchanging x and y.

$$u_x(x,y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left\{G_{xx}(x-x',y-y')\cdot q_x(x',y') + G_{xy}(x-x',y-y')\cdot q_y(x',y') + G_{xz}(x-x',y-y')\cdot p(x',y')\right\}dx'dy'$$

$$\text{i.e. } u_x(x,y) = G_{xx} * q_x + G_{xy} * q_y + G_{xz} * p \tag{3.41.1}$$

Similarly,

$$u_y(x,y) = G_{yx} * q_x + G_{yy} * q_y + G_{yz} * p \tag{3.41.2}$$

$$u_z(x,y) = G_{zx} * q_x + G_{zy} * q_y + G_{zz} * p \tag{3.41.3}$$

where symbol "*" means continuous convolution operation, and G_{mn} ($m, n = x, y, z$) are Green's functions given below:

$$G_{xx} = \frac{1}{\pi r^3}\left(\frac{x^2}{E^\#} + \frac{2y^2}{E'}\right),\quad G_{xy} = \frac{xy}{\pi\mu^\# r^3},\quad G_{xz} = -\frac{x}{\pi\mu' r^2},$$

$$G_{yx} = \frac{xy}{\pi\mu^\# r^3},\quad G_{yy} = \frac{1}{\pi r^3}\left(\frac{2x^2}{E'} + \frac{y^2}{E^\#}\right),\quad G_{yz} = -\frac{y}{\pi\mu' r^2},$$

$$G_{zx} = \frac{x}{\pi\mu' r^2},\quad G_{zy} = \frac{y}{\pi\mu' r^2},\quad G_{zz} = \frac{2}{\pi E' r},$$

where

$$r = \sqrt{x^2+y^2},\quad \frac{1}{E^\#} = \frac{1+\nu_1}{E_1} + \frac{1+\nu_2}{E_2},\quad \frac{2}{E'} = \frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2},$$

$$\frac{1}{\mu'} = \frac{(1+\nu_1)(1-2\nu_1)}{2E_1} - \frac{(1+\nu_2)(1-2\nu_2)}{2E_2},$$

$$\frac{1}{\mu^\#} = \frac{\nu_1(1+\nu_1)}{E_1} + \frac{\nu_2(1+\nu_2)}{E_2}$$

Equations (3.41.1)–(3.41.3) above can also be written in a matrix form as follows. Again, operator "*" means the continuous convolution integration rather than just a simple matrix multiplication.

$$\begin{bmatrix} u_x(x,y) \\ u_y(x,y) \\ u_z(x,y) \end{bmatrix} = \begin{bmatrix} G_{xx} & G_{xy} & G_{xz} \\ G_{yx} & G_{yy} & G_{yz} \\ G_{zx} & G_{zy} & G_{zz} \end{bmatrix} * \begin{bmatrix} q_x(x,y) \\ q_y(x,y) \\ p(x,y) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\pi r^3}\left(\frac{x^2}{E^{\#}}+\frac{2y^2}{E'}\right) & \frac{xy}{\pi\mu^{\#}r^3} & \frac{-x}{\pi\mu' r^2} \\ \frac{xy}{\pi\mu^{\#}r^3} & \frac{1}{\pi r^3}\left(\frac{2x^2}{E'}+\frac{y^2}{E^{\#}}\right) & \frac{-y}{\pi\mu' r^2} \\ \frac{x}{\pi\mu' r^2} & \frac{y}{\pi\mu' r^2} & \frac{2}{\pi E' r} \end{bmatrix} * \begin{bmatrix} q_x(x,y) \\ q_y(x,y) \\ p(x,y) \end{bmatrix} \tag{3.41'}$$

Application of Equation (3.41) or (3.41′) can be found in the papers by W. W. Chen and Wang (2008a,b), W. W. Chen et al. (2008a,b), and others, in which numerical solutions for various contact problems are presented.

3.2.5 Subsurface Stresses

When a unit normal force is applied at the coordinate origin on the surface of an elastic half-space, the induced subsurface stresses G^n_{ab} (subscripts a, $b = x$, y, z) at any response point (x, y, z) within the elastic half-space solid can be derived by substituting Equations (3.22)–(3.24) into Equations (3.1)–(3.6), and using the Boussinesq potentials, Equations (3.20) and (3.21), that further reduce to $\psi_1 = \ln(\rho + z)$ and $\psi = 1/\rho$. The resultant expressions for stresses G^n_{ab}, Equations (3.42.1) through (3.42.6), are given below with $r^2 = x^2 + y^2$ and $\rho = \sqrt{x^2 + y^2 + z^2}$. Note that Equations (3.42.4) and (3.42.4′) given below for G^n_{xy} are slightly different from those in the books by Johnson (1987) and Hills et al. (1993) (refer to S. B. Liu, 2001).

$$G^n_{xx}(x,y,z) = \frac{1}{2\pi}\left\{\frac{1-2\nu}{r^2}\left[\left(1-\frac{z}{\rho}\right)\frac{x^2-y^2}{r^2}+\frac{zy^2}{\rho^3}\right]-\frac{3zx^2}{\rho^5}\right\} \tag{3.42.1}$$

$$G^n_{yy}(x,y,z) = G^n_{xx}(y,x,z) = \frac{1}{2\pi}\left\{\frac{1-2\nu}{r^2}\left[\left(1-\frac{z}{\rho}\right)\frac{y^2-x^2}{r^2}+\frac{zx^2}{\rho^3}\right]-\frac{3zy^2}{\rho^5}\right\} \tag{3.42.2}$$

$$G^n_{zz}(x,y,z) = -\frac{3}{2\pi}\cdot\frac{z^3}{\rho^5} \tag{3.42.3}$$

$$G^n_{xy}(x,y,z) = \frac{1}{2\pi}\left\{\frac{1-2\nu}{r^2}\left[\left(1-\frac{z}{\rho}\right)\frac{2xy}{r^2}-\frac{xyz}{\rho^3}\right]-\frac{3xyz}{\rho^5}\right\} \tag{3.42.4}$$

$$G^n_{xz}(x,y,z) = -\frac{3}{2\pi}\cdot\frac{xz^2}{\rho^5} \tag{3.42.5}$$

$$G^n_{yz}(x,y,z) = G^n_{xz}(y,x,z) = -\frac{3}{2\pi}\cdot\frac{yz^2}{\rho^5} \tag{3.42.6}$$

Since we have $\left(1-\frac{z}{\rho}\right)\frac{f(x,y)}{r^2} = \frac{(\rho-z)f(x,y)}{\rho(\rho^2-z^2)} = \frac{f(x,y)}{\rho(\rho+z)}$

Equations (3.42.1), (3.42.2), and (3.42.4) can also be written as follows

$$G^n_{xx}(x,y,z) = \frac{1}{2\pi}\left\{\frac{1-2\nu}{r^2}\left[\frac{x^2-y^2}{\rho(\rho+z)}+\frac{zy^2}{\rho^3}\right]-\frac{3zx^2}{\rho^5}\right\} \tag{3.42.1'}$$

$$G^n_{yy}(x,y,z) = \frac{1}{2\pi}\left\{\frac{1-2\nu}{r^2}\left[\frac{y^2-x^2}{\rho(\rho+z)}+\frac{zx^2}{\rho^3}\right]-\frac{3zy^2}{\rho^5}\right\} \tag{3.42.2'}$$

$$G^n_{xy}(x,y,z) = \frac{1}{2\pi}\left\{\frac{1-2\nu}{r^2}\left[\frac{2xy}{\rho(\rho+z)}-\frac{xyz}{\rho^3}\right]-\frac{3xyz}{\rho^5}\right\} \tag{3.42.4'}$$

Similarly, if the shear traction $q_x(x, y)$ in Equations (3.35.1)–(3.35.3) is taken to be a concentrated unit force applied on a vanishingly small area at the origin, the subsurface stress components G^{sx}_{ab} due to a unit tangential force along the x-direction can be derived by differentiating (3.35) and then substituting the result into (3.1)–(3.6). The resultant stress components are expressed as follows.

$$G^{sx}_{xx}(x,y,z) = \frac{1}{2\pi}\left\{(1-2\nu)\left[\frac{x}{\rho^3}-\frac{3x}{\rho(\rho+z)^2}+\frac{x^3}{\rho^3(\rho+z)^2}+\frac{2x^3}{\rho^2(\rho+z)^3}\right]-\frac{3x^3}{\rho^5}\right\} \tag{3.43.1}$$

$$G^{sx}_{yy}(x,y,z) = \frac{1}{2\pi}\left\{(1-2\nu)\left[\frac{x}{\rho^3}-\frac{x}{\rho(\rho+z)^2}+\frac{xy^2}{\rho^3(\rho+z)^2}+\frac{2xy^2}{\rho^2(\rho+z)^3}\right]-\frac{3xy^2}{\rho^5}\right\} \tag{3.43.2}$$

$$G^{sx}_{zz}(x,y,z) = -\frac{3}{2\pi}\cdot\frac{xz^2}{\rho^5} \tag{3.43.3}$$

$$G^{sx}_{xy}(x,y,z) = \frac{1}{2\pi}\left\{(1-2\nu)\left[-\frac{y}{\rho(\rho+z)^2}+\frac{x^2y}{\rho^3(\rho+z)^2}+\frac{2x^2y}{\rho^2(\rho+z)^3}\right]-\frac{3x^2y}{\rho^5}\right\} \tag{3.43.4}$$

$$G^{sx}_{xz}(x,y,z) = -\frac{3}{2\pi}\cdot\frac{x^2z}{\rho^5} \tag{3.43.5}$$

$$G^{sx}_{yz}(x,y,z) = -\frac{3}{2\pi}\cdot\frac{xyz}{\rho^5} \tag{3.43.6}$$

Interestingly, we can find

$$G_{xx}^{sx} + G_{yy}^{sx} + G_{zz}^{sx} = -(1+\nu)x/\left(\pi\rho^3\right) \tag{3.43.7}$$

Stress components G_{ab}^{sy} due to a unit shear force along the y-axis can be obtained from the expressions of G_{ab}^{sx} by exchanging x and y in the superscripts and the coordinates. For example, $G_{ab}^{sy}(x,y,z) = G_{ab}^{sx}(y,x,z)$.

Each component of the above G_{ab}^{n}, G_{ab}^{sx}, and G_{ab}^{sy} represents the influence of a selected unit point source of traction (indicated by the superscript n or sx or sy), applied at the loading point of $x = 0$ and $y = 0$, on the targeted subsurface stress component (marked by subscripts a and b) at any response point (x, y, z). G_{ab}^{n}, G_{ab}^{sx}, and G_{ab}^{sy} are Green's functions (or influence functions). Generally, subsurface stresses at any point (x, y, z) due to any known distributed tractions, $p(x, y)$, $q_x(x, y)$, and $q_y(x, y)$, applied over the surface area S can be written as the following convolution integrals according to the superposition principle.

$$\begin{aligned}\sigma_{ab}(x,y,z) = \iint_S \Big[& G_{ab}^{sx}(x-x',y-y',z)\cdot q_x(x',y') \\ & + G_{ab}^{sy}(x-x',y-y',z)\cdot q_y(x',y') \\ & + G_{ab}^{n}(x-x',y-y',z)\cdot p(x',y')\Big]dx'dy', \\ & (a,b = x,y,z)\end{aligned} \tag{3.44}$$

which can also be expressed as

$$\sigma_{ab}(x,y,z) = G_{ab}^{sx} * q_x + G_{ab}^{sy} * q_y + G_{ab}^{n} * p, \quad (a,b = x,y,z) \tag{3.44'}$$

In most practical cases, Equations (3.29), (3.38), (3.41), and (3.44) involve convolution integrations that may have no analytical solution. Therefore, numerical solution should be pursued, and a large amount of computation is often expected. Various numerical algorithms have been developed to accelerate the computation process. An efficient approach, called the DC-FFT (discrete convolution and fast Fourier transform) method, has been developed by S. B. Liu et al. (2000) and S. B. Liu and Wang (2002) that will be discussed in the next chapter.

For line contact problems under the plane strain condition, in principle, subsurface stresses can-be obtained from Equations (3.1)–(3.17) through the displacement results. Johnson (1987) presented the stresses due to arbitrarily distributed pressure $p(x)$ and tangential traction $q(x)$ applied on the surface of a half-space within the strip $(-b < x < a)$, as shown in Figure 3.4. The stress components at any point $A(x, z)$ in the solid body are given below.

$$\begin{aligned}\sigma_x(x,z) = & -\frac{2z}{\pi}\int_{-b}^{a} p(\xi)\frac{(x-\xi)^2 d\xi}{\left[(x-\xi)^2+z^2\right]^2} \\ & -\frac{2}{\pi}\int_{-b}^{a} q(\xi)\frac{(x-\xi)^3 d\xi}{\left[(x-\xi)^2+z^2\right]^2}\end{aligned} \tag{3.45.1}$$

$$\begin{aligned}\sigma_z(x,z) = & -\frac{2z^3}{\pi}\int_{-b}^{a} p(\xi)\frac{d\xi}{\left[(x-\xi)^2+z^2\right]^2} \\ & -\frac{2z^2}{\pi}\int_{-b}^{a} q(\xi)\frac{(x-\xi)d\xi}{\left[(x-\xi)^2+z^2\right]^2}\end{aligned} \tag{3.45.2}$$

$$\begin{aligned}\tau_{xz}(x,z) = & -\frac{2z^2}{\pi}\int_{-b}^{a} p(\xi)\frac{(x-\xi)d\xi}{\left[(x-\xi)^2+z^2\right]^2} \\ & -\frac{2z}{\pi}\int_{-b}^{a} q(\xi)\frac{(x-\xi)^2 d\xi}{\left[(x-\xi)^2+z^2\right]^2}\end{aligned} \tag{3.45.3}$$

These are also convolution integrals over Green's functions with $p(x)$ or $q(x)$.

3.3 LINE CONTACT HERTZIAN THEORY

3.3.1 Basic Model

Many mechanical components can be treated as those with counterformal contacts, which are generally categorized into line and point contacts, as stated in Section 1.2.1. For these machine elements, the Hertzian theory (Hertz, 1881, 1882) was almost the only practical means to provide reasonable estimates of contact stress and deformation in industrial applications for many decades until recently. Even today, the Hertzian theory is still widely in use because of its simplicity and satisfactory accuracy for contact problems with homogeneous elastic materials without significant influences of lubrication, surface roughness, and friction. The available Hertzian formulae provide classic solutions for interfacial pressure and contact deformation at the contact location, where the geometric shapes of the two bodies are approximated by either cylinders or ellipsoids.

In Section 3.3, discussion is focused on line contacts. Point contact problems will be tackled later. Figure 3.5 shows a basic model that consists of two parallel cylinders of infinite length under a normal load per unit contact length, W, uniformly distributed along the axial direction. When the

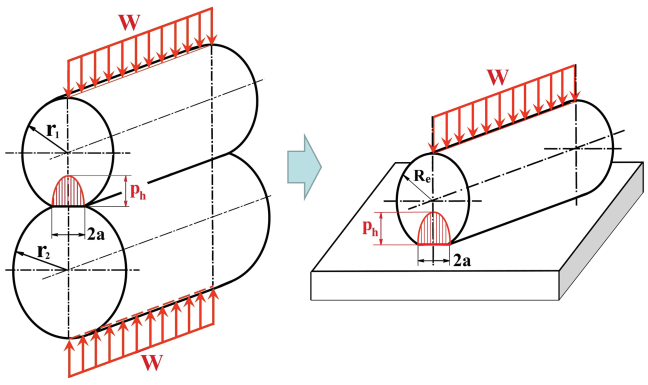


FIGURE 3.5 Basic model for a line contact between two parallel cylinders. The equivalent cylinder is shown on the right.