(2)

Final Report

N00014-80-C-1736

Tribological Technology

ONR/STI

September 1982

Volume I

AD A121035

DTIC FILE COPY

DTIC
ELECTE
NOV 3 1982
S D

DISTRIBUTION STATEMENT A
Approved for public release
Distribution Unlimited

82 10 25 080

TRIBOLOGICAL TECHNOLOGY

edited by

PETER B. SENHOLZI

Mechanical Technology Incorporated
1656 Homewood Landing Road
Annapolis, Maryland, USA

Proceedings of the NATO Advanced Study Institute
on Tribological Technology, held at
Maratea, Italy, September 13 - 26, 1981

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

The Tribological Technology ASI Program was cosponsored by NATO and the U.S. Office of Naval Research. Dr. M. di Lullo has served as the NATO sponsor and Messrs. H. Martin and K. Ellingsworth have served as Office of Naval Research sponsors. Credit is given to these individuals for their generous assistance and guidance that was provided during the organization and conductance of the ASI Program.

A special acknowledgement is extended to the respective ASI technical contributors who have taken time from their busy schedules in order to ensure the successful conductance of the ASI Program and compilation of the text (especially Brian Reason and Horst Czichos who served on the ASI Organizing Committee). My staff members, Nanette Brown and Alan Maciejewski, are also to be acknowledged for their assistance in the areas of program organization, material compilation, and text editing. Appreciation for the extensive manual typing responsibilities is extended to Susan G. Rohn. Special recognition and thanks is extended to my dedicated assistant, Phyllis Pittman, who expended considerable time and energy in the successful completion of all facets of the ASI Program. Last but not least, I would like to extend my gratitude to my wife, Patricia, who provided considerable support as well as exhibited extreme patience throughout the conductance of this Program.

INTRODUCTION

Peter B. Senholzi
Manager, Applications Engineering
Mechanical Technology Incorporated

In most discussions concerning tribology, authors felt compelled to introduce the topic by first defining the term. In this discussion, however, the term will not be defined until after the circumstances that spawned its existence have been summarized.

This discussion will revolve around two facets of mechanical system design, that of structural integrity and wear integrity. In the not too distant past, the evolutionary process of machine design had in most cases attained a stage of development where the structural integrity of a system was ensured. Simply stated, this means that the machine components would not "break." However, at this development phase, little emphasis was placed on component wear integrity. This meant that although machinery components would not break, they would be subject to high wear/wear-out rates. Frequent component replacements were used in order to compensate for this situation. A "throw away" philosophy was thus adopted. This philosophy, however, resulted in many undesirable secondary effects. Mechanical systems were characterized by expensive logistics, costly maintenance, limited availability, limited lives, and substantial energy consumption. This energy consumption resulted from primary effects of high equipment frictional levels (causing the high wear rate) and such secondary effects as spare parts production and transportation. As a result of these factors, the "throw-away" philosophy was a very vulnerable luxury.

During the last fifteen or so years, environmental factors with respect to mechanical system operation have changed

drastically. Respective operating organizations have been faced with such factors as limited natural resources (material and energy), limited economic resources, spiraling manpower costs, and resulting spiraling capital equipment costs. These factors are compounded by a substantial inflation rate. This "new" operation environment comes into direct conflict with the existing "throw away" philosophy.

In order to resolve this conflict, the machine design process was forced to address the facet of wear integrity. However, the resolution is not as simple as it appears. Machine wear and friction are a function of such assorted variables as lubrication, additives, materials, surfaces, contamination, design, etc. No existing discipline was fully and effectively qualified to address the total of these variables and, hence, the problem of component wear integrity. An unresolvable situation was thus created under the existing technological structure.

One of the first to verbalize this situation were the British in the Jost Report published in 1966, which summarized the findings of the now famous Jost Committee. Established in 1964 by the UK Minister of State for Education and Science, the Committee was directed to define the current position of lubrication education and research in the UK and to give an opinion on the needs of industry in this field.

The resulting Jost Report outlined this wear integrity problem and recommended a multidisciplinary approach to its solution. The document labeled the proposed approach "tribology," and defined it as "the science and technology of interacting surfaces in relative motion and of the practices related thereto." This term included the subjects of friction, lubrication, and wear, e.g.

- The physics, chemistry, mechanics, and metallurgy of interacting surfaces in relative motion including the phenomena of friction and of wear.
- Fluid film lubrication, e.g., hydrostatic, hydrodynamic, aerostatic, and aerodynamic.
- Lubrication other than fluid film, e.g., boundary and solid lubrication.
- Lubrication in special conditions, e.g., during metal deformation and cutting processes.
- The properties and operational behavior of bearing materials.

- The engineering of bearings and bearing surfaces (e.g., plain and rolling bearings, piston rings, machine slides, gear teeth, etc.) including their design, manufacture, and operation.

- The engineering of bearing environments.

- The properties and operational behavior of fluid, semi-fluid, gaseous, and solid lubricants and of allied materials.

- The quality control and inspection of lubricants.

- The handling, dispensing, and application of lubricants.

- The management and organization of lubrication.

The report went on to provide justification in support of its recommendations by quantifying the potential cost savings accrued through the application of the multidisciplinary tribology approach. It was probably this quantification effort that was the most startling aspect of the Jost Report. Cost savings for British industry at that time (1966), were estimated to be approximately $500 million per year.

Although this quantification effort has raised much discussion as to the legitimacy of several included factors, the mere magnitude emphasizes the potential of tribology. Based on the findings of the Jost Report and motivated by the projected savings figure, Britain has since initiated a substantial tribological technology implementation program. The remainder of Europe, to a slightly lesser extent, has followed suit.

The United States is several years behind Europe in the adoption of tribological technology. This lag results in part, from the fact that U.S. organizations were not confronted with the conflict between the "throw away" philosophy and contradictory operating environmental trends until the early 1970's, several years after the British encounter. The difference in encounter dates probably results from economic factors, an explanation of which is beyond the scope of this discussion.

The relatively new nature of tribological technology, coupled with its multidisciplinary foundation and its varying adoption schedule, renders it a fertile area for academic treatment. It is this scientific basis that formed the lecture

content of the NATO Advanced Study Institute held in Maratea, Italy, in September of 1981.

The Tribological Technology ASI Lecture Program was designed to develop multidisciplinary tribological technology from a fundamental basis through an applications viewpoint. This development approach consists of eleven discreet lectures. These lectures form the foundation of this text and are organized as follows:

(1) Scope of Tribology;

The first lecture, chapter two, provides a description of the scope of tribology. Such aspects as background, economics, justification, terminology, applications, and implications are covered.

(2) Surface Interactions;

The third chapter provides a foundation for the Tribological Technology. It is comprised of fundamental discussions concerned with interaction theory and interaction types.

(3) Surface Interaction Elements;

Chapters four through eight provide in-depth discussions of individual interaction elements. Respective element discussions include:

A. Materials
B. Surfaces
C. Lubrication
D. Lubricants
E. Contamination

(5) Tribological Failures and Mechanical Design.

The ninth chapter provides discussions of tribological failure definition, failure types, consequences, and failure control and prevention measures.

(6) Tribo-Testing;

Chapter ten covers testing approaches available for interaction element research, development, or quality assurance. This testing segment includes mechanism, component, and full system approaches.

(7) Monitoring

Chapter eleven, "Monitoring," presents discussions concerned with the tracking of the wear process. Primary emphasis under this chapter involves equipment diagnostic approaches, diagnostic equipment, philosophy, and effectiveness.

(8) Tribological Multidisciplinary Approach

The final chapter discusses a total tribological approach from equipment design through manufacture and operation. The previous chapters are integrated under this discussion.

A schematic of this lecture program is illustrated in Figure 1. This approach and content provides a logical, comprehensive treatment of tribological technology.

The above respective lectures have been compiled into the present volume. Each author is an expert of international standing and was subjected to minimal constraints with respect to the formulation of individual written chapters. As a result, individual author styles and emphasis have not been suppressed in this volume.

The attached appendix contains several technical papers addressing relevant tribological research efforts. These papers were solicited and presented as part of the ASI program.

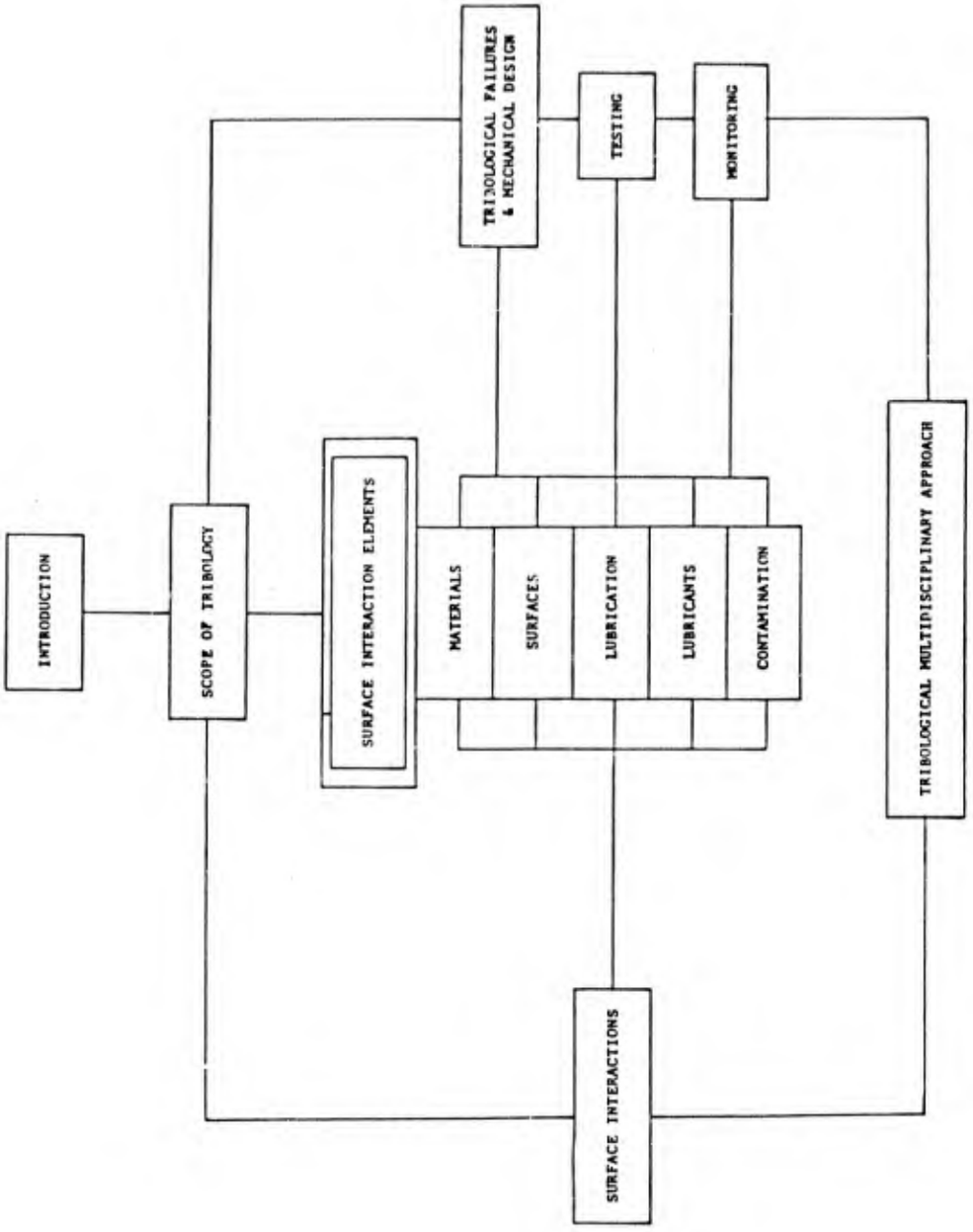


FIGURE 1 TRIBOLOGICAL TECHNOLOGY ASI PROGRAM APPROACH

SCOPE OF TRIBOLOGY

Horst Czichos
Berlin-Dahlem, West Germany

1. INTRODUCTION

The term "TRIBOLOGY" was coined by a British committee in 1966 from the word "tribos", which means "rubbing" in classic Greek. Tribology is defined as: "The science and technology of interacting surfaces in relative motion and of related subjects and practices." Since its definition tribology has been widely recognized as a new general concept embracing all aspects of the transmission and dissipation of energy and materials in mechanical equipment including the various topics of friction, wear, lubrication, and related fields of science and technology. This discussion provides a general description of tribology, its scope, applications, and implications. The material is presented under the following headings: background, economic impact, terminology, fields of application, tribological parameters, and tribological system effectiveness.

2. BACKGROUND

The term "TRIBOLOGY" was coined by a British committee in 1966 from the word "tribos", which means "rubbing" in classic Greek[19]. Since its definition 15 years ago, the term tribology has been widely recognized as a new general concept embracing all aspects of the transmission and dissipation of energy and materials in mechanical equipment including the various topics of friction, wear, lubrication, and related fields of science and technology. As friction is responsible for a major loss of useful mechanical energy, and wear is a major reason for replacing equipment; a better understanding and utilization of

the principles of tribology is particularly important for the conservation of energy and materials in engineering design.

The historical development of tribology is described comprehensively in the famous book by Dowson "History of Tribology," for which the reader is referred to[8]. In chronological terms, the development of the studies of the main areas of tribology - friction, lubrication, and wear - fall into two periods of time with the end of the Second World War marking the demarcation. In terms of subject-matter there seems little doubt that, if in the previous 75 years the main interest was on "lubrication" and in the time before on "friction", the present main interest seems to be on "wear." This development is connected with the increasing awareness that the complex friction and wear problems in contemporary mechanical engineering can only be treated successfully by a unifying approach which takes into account systematically all important influencing parameters and processes[7].

After the Second World War, a general trend in mechanical engineering towards higher loads, higher velocities, and higher operating temperatures was observed in connection with attempts to reduce the weights of the moving parts in machinery. Since that time there has been a steady increase of interest in the problems of friction, lubrication, and wear in connection with the increasing worldwide attempts on the conservation of energy and materials. Tokens of this increasing interest were, among other things, the appearance of the first international journal entitled WEAR (since 1957) entirely devoted to the science and technology of friction, wear, and lubrication. A further new journal, the TRANSACTIONS of ASLE, has been published (since 1958) by the American Society of Lubrication Engineers. In West Germany a research programme ("Schwerpunkt-Programm") had been launched in 1961 sponsored by the Deutsche Forschungsgemeinschaft (DFG). About the same time in England, a Working Group was set up to investigate the present state of lubrication education and research and to give an opinion on the needs of industry thereof. In 1966 this Working Group published a report (the so-called "Jost Report"), in which for the first time the term "tribology" appeared[19]. The origin of tribology was described in the report "The Introduction of A New Technology" (1973) as follows[20]:

In the early 1960's there was a steep increase in the reported failure of plant equipment and machinery due to wear and associated causes. At the same time, increased technology, increased capital intensity of plant equipment and the use of more continuous processes made breakdowns of such plant equipment and machinery more costly, competitively more serious, and therefore even less desirable than before. This trend was

recognized by specialists involved in the subjects of wear, friction, and lubrication. The situation seemed to call for more and better education in the subjects and for more and better coordinated research.

While trying to establish the reasons for the wide neglect of the subject in the past, despite its technological and economic importance, three principal reasons for this neglect may be mentioned:

(i) The interdisciplinary nature of the subject which included the disciplines of mechanical engineering, physics, metallurgy, and chemistry.

(ii) The fact that only with the advances in technological development in production methods of recent years had attention been focused on the importance and interdependence of the constituents of this interdisciplinary subject.

(iii) The term "lubrication," used in its narrower sense, had not only prevented many people from fully appreciating the economic significance of the subject matter, but it was also a misnomer for the description of the sphere of "transference of force from one moving surface to another" (whether the purpose of the transfer of such forces was associated with high friction, e.g., on brakes, clutches, conveyors or alternatively with low friction, e.g., with bearings, slides, etc.)

After consultation with the Editors of the Supplement of the Oxford English Dictionary, the term "Tribology" (Triboscience or Tribotechnology) was recommended for describing the subject matter. Tribology is defined as:

> "The science and technology of interacting surfaces in relative motion and of related subjects and practices."

3. IMPACT OF TRIBOLOGY

In considering the impact of tribology, three aspects may be emphasized: the scientific, the multidisciplinary, and the economic aspect of tribology.

From a scientific point of view, tribology is to be considered as being the discipline which tries to explain the most dominant irreversible processes in nature and technology. It is well known that all macroscopic processes in nature are irreversible. Science in its "pure" theories has largely omitted this irreversibility since the laws of "ideal" processes

were much easier to develop. In this connection, an interesting remark of the mathematician John V. Neumann should be quoted concerning an important branch of tribology, namely lubrication[7]. John V. Neumann pointed out that during most of the development of hydromechanics until about 1900, the scientific interest was mainly concentrated on the development of "pure" theories of fluid mechanics by solving beautiful mathematic problems neglecting terms of internal fluid friction. These approximations, however, had almost nothing to do with real fluids. He characterized the theorist making such analyses as a man who studied the flow of "dry" water. In contrast, tribology is attempting to investigate the irreversible processes of mechanics in detail and to explain the complex effects of friction-and-wear-induced energy and materials dissipation.

A second important aspect is the multidisciplinary nature of tribology. Since tribology is defined as "science and technology of interacting surfaces in relative motion", it includes not only the work of physicists, chemists, and material scientists interested in the surface properties of materials, but also the work of engineers who use "interacting surfaces" for the transmission of motion, forces, work etc., in various types of machinery. Therefore, tribology is connected with several branches of science and technology, such as physics, chemistry, materials science, mechanical engineering, and lubrication engineering. Consequently, the attempts for the solution of tribological problems require the combined effort of these branches.

The importance of the economic aspects of tribology has often been emphasized in the literature in recent years[23]. The 1966 Jost Report, in which the term tribology was coined, estimated that by the application of mainly known principles of science and technology of tribology, the economy of Great Britain could save in the region of £515 million p.a. (at 1965 values), as follows, see Figure 1[19].

		£ million
(a)	Reduction in energy consumption through lower friction	28
(b)	Reduction in manpower through better lubrication	10
(c)	Savings in lubricant costs	10
(d)	Savings in maintenance and replacement costs	230
(e)	Savings of losses consequential upon breakdown	115

(f)	Savings in capital expenditure due to higher utilization ratios and greater mechanical efficiency	22
(g)	Savings in investments through increased life of machinery	100

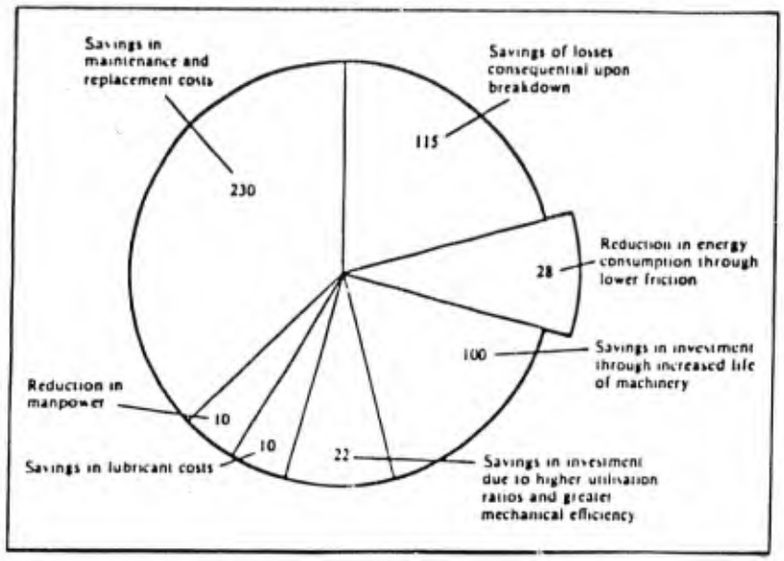


FIGURE 1 ESTIMATED SAVINGS IN TRIBOLOGY IN THE UK (JOST REPORT, 1966)

In West Germany a 280 page report on Tribology was published in 1976 by the Federal Ministry of Research and Technology (BMFT) as a basis of a pertinent research programme[21]. In this report, it was estimated that the economic losses due to friction and wear in West Germany amount to more than 10 billion DM p.a. (at 1975 values) which is equivalent to 1% of the GNP.

Also, in 1976 a plan for "energy conservation through tribology" was published by the American Society of Mechanical Engineers (ASME) supported by the US Department of Energy and the US Office of Naval Research[22]. The American investigation found that three major sectors, transportation, electric utilities, and industry, accounted for about 80%. Half of this energy was rejected in the course of its utilization. In the transportation and utility sectors, 75% and 65% respectively, resulted primarily from thermal cycle limitations but also from friction and sealing losses. In the industrial sector, waste of approximately 25% was caused primarily by mechanical losses, such as friction, wear, and leakage past seals. The total energy used, that rejected, the potential savings through tribology, and - in enlarged form - the areas in which such savings could be made are illustrated in Figure 2. Even if all

the figures cited in this section are taken as rough estimates only; they clearly indicate the importance of tribology for the conservation of energy and materials.

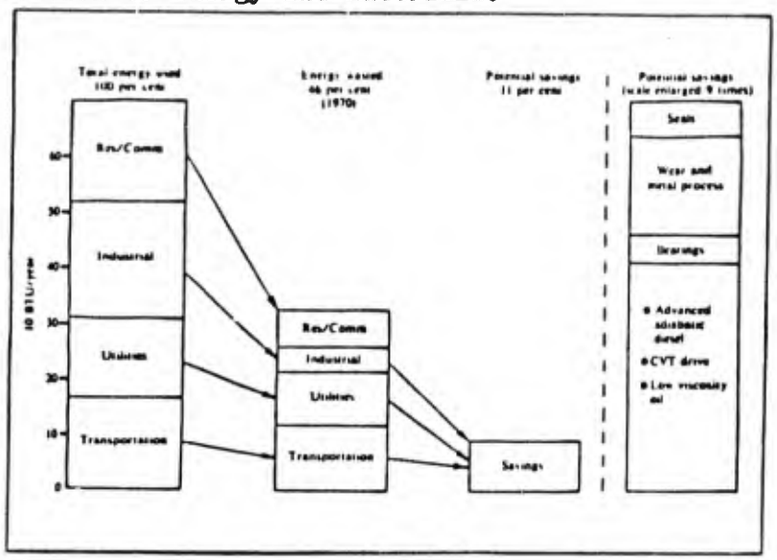


FIGURE 2 ESTIMATED LOSSES & SAVINGS IN THE US (ASME REPORT, 1977)

4. TERMINOLOGY

Owing to the multidisciplinary nature of tribology, the terminology used in this field has many origins. Thus, misleading or ambiguous terms are sometimes used. In 1969 the International Research Group on Wear of Engineering materials, working under the auspices of OECD, has compiled a "Glossary on Terms and Definitions" for the field of tribology. This Glossary has been recognized as a key word index and is included in the new ASME Wear Control Handbook published in 1980[16]. It contains definitions of the following items:

(i) 500 general tribological terms,
(ii) 93 terms on bearing types, and
(iii) 58 terms on oils and greases.

In addition, the Glossary contains an eight-language index consisting of translations of the tribological terms into the following languages: English, French, German, Italian, Spanish, Japanese (Romaji), Japanese (Kanji), Arabic, and Portuguese.

With this Glossary on Terms and Definitions, the basis for a multilingual terminology for the field of tribology has been laid. Because it is not possible to repeat details of the Glossary within the frame of this discussion, only the definitions of the three most important tribological terms will be cited here from the Glossary[16].

Friction: The resisting force tangenial to the common boundary between two bodies when, under the action of an external force, one body moves or tends to move relative to the surface of the other.

Wear: The progressive loss of substance from the operating surface of a body occurring as a result of relative motion at the surface.

Lubrication: The reduction of frictional resistance and wear or other forms of surface deterioration between two load-bearing surfaces by the application of a lubricant.

Because of the great variety of types, modes, or processes of friction, lubrication and wear which determine the behavior of any tribological entity, it is necessary to have a sub-classification of these basic terms. For these sub-classifications different criteria are used such as:

- kinematics or type of motion (sliding, rolling, impact, etc.)
- type of materials (solid, fluid, metal, polymer, etc.)
- type of interfacial tribological process (hydrodynamics, adhesion, abrasion, etc.)

Whereas in most cases it is possible to describe the basis of the kinematics or the type of material, the type of friction and lubrication mode of a given tribological entity (e.g. sliding friction or rolling friction; solid lubrication or fluid lubrication; etc.) there are considerable difficulties in attempting a logically consistent classification of wear phenomena. In 1979 the German Institute for Standardization (DIN) published a standard on wear (DIN 50 320 "WEAR") in which a classification of the terms of wear is given based on an analogy to the (classical) field of the strength of materials[18]. In order to classify the strength, deformation, or failure of a material, the following specifications must be given:

(i) type of material
(ii) type of external stress (e.g. compression, tension, bending, etc.)
(iii) type of internal deformation or fracture mechanism (e.g. ductile fracture, brittle fracture)

In analogy, in order to classify the type of wear occurring within a tribological entity, the following specifications must be given:

(i) the materials involved in wear (solid, fluid, etc.)
(ii) type of kinematic tribological action (sliding, rolling, impact, etc.)
(iii) type of inerfacial wear mechanism (adhesion, abrasion, surface fatigue, etc.)

The resulting classification of DIN 50 320 for the field of wear is shown in Figure 3. In the classification scheme shown in Figure 3, the types of wear are named according to the kinematics. This is in analogy both to the classification used in the field of the strength of materials, as mentioned above, and to the common classification of friction like sliding friction, rolling friction, etc. During the occurrence of the different types of wear, in the interface of the wear couple, one or more different wear mechanisms may act. It is generally accepted that there are four main wear mechanisms which are described in DIN 50 320 as follows[18]:

- Adhesion: Formation and rupture of interfacial adhesive bonds (e.g. "cold welded junctions," "scuffing")

- Abrasion: Removal of materials by a scratching process (micro-cutting process)

- Surface Fatigue: Fatigue and crack formation in surface regions by tribological stress cycles resulting in separations of materials (e.g. "pitting")

- Tribo-chemical Reactions: Development of reaction products as a result of chemical reactions taking place between the wear couple and the interfacial medium.

In conclusion with the terms and definitions of the Glossary as contained in the Wear Control Handbook and the classification of wear of the Wear Standard DIN 50 320, a multilingual terminology for the field of tribology is available[16, 18]. It is strongly recommended that this terminology be used in order to avoid ambiguities and to improve understanding and communication in the field of tribology on an international scale.

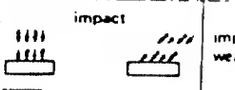
System structure	Tribological action (symbols)	Type of wear	Effective mechanisms (individually or combined)			
			Adhesion	Abrasion	Surface fatigue	Tribo-chemical reactions
Solid – interfacial medium (full fluid film separation) – solid	sliding rolling impact 	–			×	×
Solid – solid (with solid friction, boundary, lubrication, mixed lubrication)	sliding 	sliding wear	×	×	×	×
	rolling 	rolling wear	×	×	×	×
	impact 	impact wear	×	×	×	×
	oscillation 	fretting wear	×	×	×	×
Solid – solid and particles	sliding 	sliding abrasion		×		
	sliding 	sliding abrasion (three body abrasion)		×		
	rolling 	rolling abrasion (three body abrasion)		×		
Solid – fluid with particles	flow 	particle erosion (erosion wear)		×	×	×
Solid – gas with particles	flow 	fluid erosion (erosion wear)		×	×	×
	impact 	impact particle wear		×	×	×
Solid – fluid	flow oscillation 	material cavitation, cavitation erosion			×	×
	impact 	drop erosion			×	×

FIGURE 3 CLASSIFICATION OF WEAR PHENOMENA

5. FIELDS OF APPLICATION

The many technical aims and functions to be realized through "interacting surfaces in relative motion" lead to a great variety of tribo-technical components and tribo-engineering systems and processes. It can be shown that the technical functions realized through interacting material surfaces can be broadly classified in the following groups[7]:

a) Transmission of MOTION
b) Transmission of WORK (or POWER)
c) Generation or reproduction of INFORMATION
d) Transportation of MATERIALS
e) Forming of MATERIALS

The various engineering systems in which friction and wear processes occur can be easily classified according to their function in considering the pertinent inputs and outputs. A broad classification is given in Table 1. Invariably, motion is a characteristic of any tribo-mechanical system. This motion may constitute a transfer of work, materials, or information. In some instances, the purpose of a system may be to change a rate of motion or to eliminate it altogether. It is also often desired to restrict motion, i.e., to reduce the number of degrees of freedom a machine element may possess. In other instances, materials are not merely moved but also changed in state or form. Mechanical devices which produce or transfer information are still common, but are being steadily replaced by devices in which there is little or no mechanical motion; for example, the replacement of mechanical clocks by digital electronic clocks.

In all the engineering systems compiled in Table 1, friction, wear, and lubrication processes of various kinds occur. In order to optimize the functional technical behavior of these systems and to minimize friction-and- wear-induced energy and material losses, tribological knowledge must be applied with respect to all the basic engineering activities, namely:

- design
- manufacture
- operation
- condition monitoring
- repair
- recycling

6. TRIBOLOGICAL PARAMETERS

Any proper analysis and solution of tribological problems require the consideration of numerous parameters and influencing factors. Through the application of the recent methods of systems analysis, it is possible to compile and to classify the main tribological parameters[7].

The procedure of a systematic analysis of tribological parameters will be explained on the basis of the simplified representation of a tribological system, Figure 4, taken from DIN 50 320 [18].

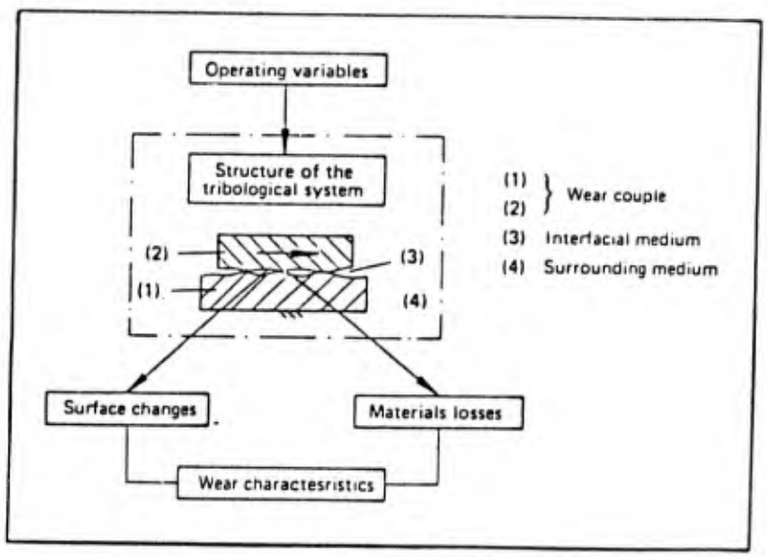


FIGURE 4 PARAMETER GROUPS OF A TRIBOLOGICAL SYSTEM

For the systematic analysis of the relevant tribological parameters, first the components and substances directly participating in the tribological process have to be conceptually separated from the other parts of the technical equipment considered. The components and substances directly participating in the tribological process are termed the "elements" or "components" of the tribological system. These elements, together with their tribologically relevant "properties" and the "interactions" between them, form the "structure of the tribological system." The operating quantities which act on the elements of the tribological system from outside form the "operating variables." The energy and material losses which occur through the action of the operating variables on the structure of the tribological system are to be described by tribological loss-characteristics.

It follows that for a systematic discussion of tribological parameters, the following four basic aspects must be considered:

I Functional technical purpose of the tribo-system
II Operating variables
III Structure of the tribo-systems, consisting of:
a) materials components involved
b) relevant properties of the components
c) interactions between the components
IV Tribological loss-characteristics

6.1 Functional Technical Purpose of Tribological Systems

The characterization of the functional technical purpose of a tribological system is a first needed overall description. The technical aims realized through interacting moving surfaces in tribological entities may range from aerospace applications to biomechanical joints. A broad classification has already been given in Section 5 (see Table 1).

Inputs and outputs needed for technical function		Primary technical function of the system	Examples
Main inputs {X}	Main outputs {Y}		
Motion + Work	Motion	Guidance of motion Coupling of motion Annihilation of motion	Bearings Clutches Brakes
	Work	Power transmission (mech., hydr., pneum.)	Gears
	Information	Generation of information	Clocks; Cams and followers
		Reproduction of information	Data transducer (audio, video; tape or record)
Motion + Materials	Materials	Transportation	Wheel/rail Pipeline
		Forming of materials	Wiredrawing

TABLE 1 CLASSIFICATION OF TECHNICAL FUNCTIONS OF TRIBOLOGICAL SYSTEMS

The most general technical purpose of a tribological system is the guidance of motion through various types of "bearings." The other basic groups are the transmission of mechanical work, the transmission of information - for instance the control of machine functions with cams - and the forming of materials. It follows that basically the technical function of tribological systems is connected with the transmission or transformation of one or more of the basic quantities:

- motion
- work
- information
- materials.

In using these four basic quantities or related quantities, the technical function of the various tribological systems may be classified in terms of the input-output relations of these quantities.

6.2 Operating Variables

Having identified the functional technical purpose of a tribological system under discussion, it is then necessary to characterize and to compile its basic operating variables.

The most characteristic operating variable of a tribological system is the type of relative motion between tribo-element (1) and tribo-element (2). The basic types of motion are:

- sliding
- rolling
- spin
- impact.

It can be shown that every type of relative motion between system components can be expressed as a superposition of these four basic types of motion. In addition to the characterization of the type of motion, its dependence on time should be specified, for example:

- continuous
- oscillating
- reciprocating
- intermittent.

The other basic operating variables are the following quantities:

- load F_N
- velocity v
- temperature T
- distance of motion s
- operating duration t.

For some tribological systems, these physical operating variables are accompanied by material inputs, e.g., the flow rate of the lubricant. Some disturbing inputs may also be present, e.g., vibration and radiation. In addition to these

parameters, it may often be necessary to also specify composite or derived quantities, e.g. contact pressure, temperature gradients, etc.

6.3 Structure of Tribological Systems

In addition to the description and classification of the various tribo-mechanical systems according to their external function and the compilation of pertinent operating variables, the question of the internal "structure" of these systems is to be discussed.

As described above, the structure of a tribological system is given by the systems elements (that is the material components of the system), their relevant properties and their inter-relations, described formally by the set S = { A, P, R}, see Figure 5.

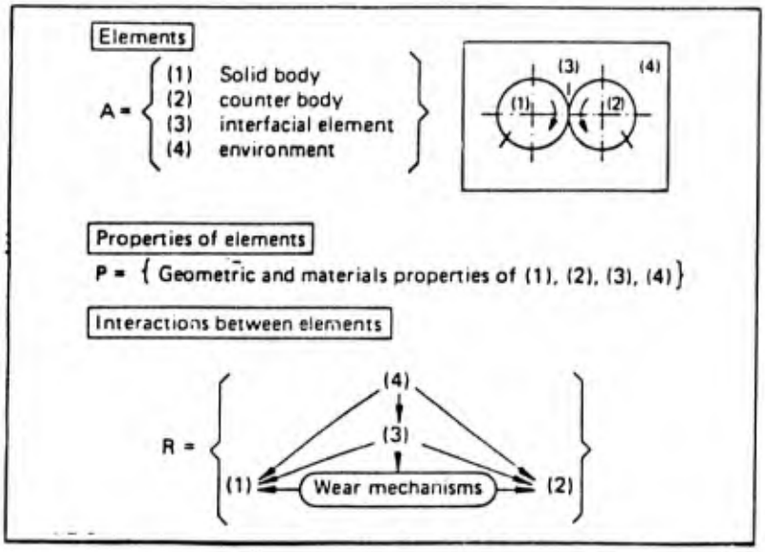


FIGURE 5 STRUCTURE OF A TRIBOLOGICAL SYSTEM S = { A, P, R}

6.3.1 Elements of the System, A = {a_i}

If the system's envelope is located as closely as possible around the "interacting surfaces in relative motion," it appears that in most tribological systems four different basic elements are involved in the friction and wear processes.

As illustrated in Figure 5 for a simple two-disc configuration, the components that form the pair of the "interacting surfaces in relative motion" are the moving elements, (1) and (2) respectively. The other two basic elements are the interfacial lubricant (3) (if any) and the

environmental atmosphere (4). These main elements are linked to others or may be composed of sub-constituents. For example, element (3), the lubricant, may consist of a base oil and additives. In Table 2 the elementary elements or components (1), (2), (3), (4) of tribological systems are listed as examples from every group of the basic tribological systems compiled in Table 1.

Tribological system (or process)	Elements of the system			
	Tribo-element (1) (moving or stationary)	Tribo-element (2) (moving or stationary	Interfacial medium (3)	Surrounding medium (4)
Sliding bearing	shaft	bushing	lubricant	air
Band clutch	shaft	band	–	air
Disc brake	disc	pad	contaminant	air
Worm gear set	worm	gear	gear oil	air
Cam and follower	cam	follower	lubricant	air
Printing unit	print-head	paper	dye	air
Audio pick-up	record	sapphire tip	–	air
Electrical contact	ring	brush	spray	cover gas
Locomotion	wheel	rail	contaminant	air
Pipeline	fluid	pipeline	–	–
Wiredrawing	wire	die	borax	air
Hot extrusion	billet	die	glass	air
Turning	workpiece	cutting tool	cutting fluid	air

TABLE 2 COMPILATION OF TYPICAL TRIBO-ENGINEERING SYSTEMS

6.3.2 Properties of the Elements, $P = \{P(a_i)\}$

The behavior of any tribological system is influenced by many properties of the basic elements (1), (2), (3), (4) all described above. Owing to the great variety of tribo-mechanical systems and tribological processes, it is very difficult to provide a comprehensive general compilation of the tribologically relevant properties of the systems elements. The following properties of the elements have been found to be of primary concern for the tribological behavior of the system:

(i) Properties of the Tribo-Elements (1) and (2) (moving elements)
The tribologically relevant properties of the elements (1) and (2) can be subdivided into "volume" properties and "surface" properties. At least the

following main properties of the elements (1) and (2) should be specified:
- Volume properties: geometry, chemical compositions, and metallurgical structure; materials data, including elastic modulus, hardness, density, and thermal conductivity.
- surface properties: surface roughness and surface composition.

(ii) Properties of the Lubricant (3)
The main relevant properties of the lubricant are the viscosity and its dependence upon temperature and pressure, together with the chemical composition of the lubricant.

(iii) Properties of the Environmental Atmosphere (4)
The main relevant properties of the atmosphere are its chemical composition and the amount and pressure of its components, especially water vapor.

To assist in the compilation of tribologically relevant parameters of the elements of tribo-mechanical systems, a data sheet has been developed and published as an appendix to the Wear Standard DIN 50 3[18].

6.3.3 Interactions Between the System's Elements, $R = \{R(a_i, a_j)\}$

The tribological interactions between the elements of a mechanical system, i.e. the contact, friction, lubrication, and wear processes, are of paramount interest in the description of any tribo-mechanical system. In Figure 6, basic tribological processes which are known today are expressed in the form of highly simplified schematic diagrams for systems of increasing complexity, i.e. increasing number of interacting elements.

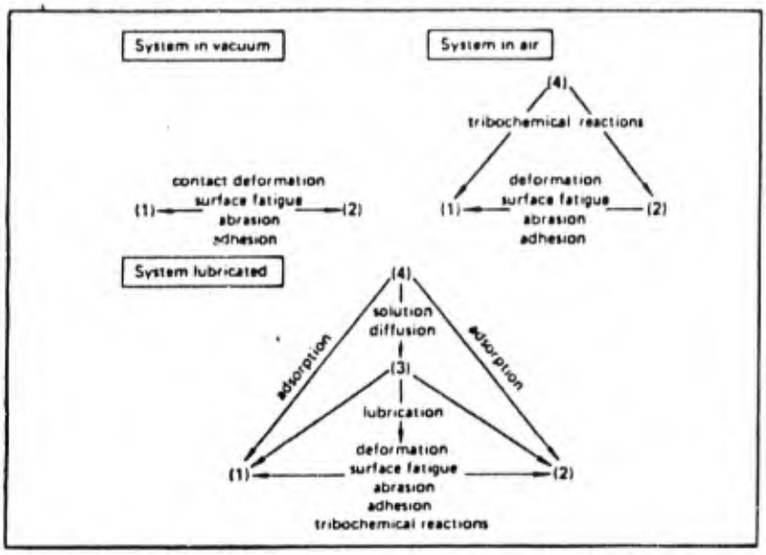


FIGURE 6 SCHEMATIC REPRESENTATION OF TRIBOLOGICAL INTERACTIONS BETWEEN SYSTEM COMPONENTS

In an ultrahigh vacuum, the simplest tribological system consists only of the two interacting partners (1) and (2). The main possible tribological interactions between moving and stationary elements are then covered by the terms contact deformation, surface fatigue, abrasion, and adhesion. In air, i.e. under dry friction condition, these processes are supplemented by interactions between the moving and the stationary partners (1) and (2) and the atmosphere (4). Through these interactions tribo-chemical reactions result. Finally, in a lubricated system, tribological processses between the elements are given by interactions between all four basic elements. In this case, the direct (contact) interactions between moving and stationary elements are prevented or influenced through the different mechanisms of lubrication. Depending on the thickness of the lubrication film, different lubrication regimes result.

Also here, interactions between (4) and (3) with (1) and (2) should be taken into account. For instance the diffusion of atmospheric oxygen into the lubricant (4)→(3), followed by oxidation processes between the lubricant and the moving and stationary partners (3)→(1), and (2), can distinctly influence the mechanisms of mixed and boundary lubrication.

6.4 Tribological Characteristics

The tribological characteristics, i.e. the characteristics that describe the dynamic changes of a tribo-mechanical system

as a consequence of the action of friction and wear processes, may be divided into the following three groups:

(a) Tribo-induced changes of system's structure
(b) Tribo-induced energy losses
(c) Tribo-induced material losses.

Depending on the tribological processes which occur within a tribo-mechanical system, the tribo-induced changes of a system's structure (a) concern:

(i) the destruction of elements or the creation of new elements in a tribo-system, e.g. the degradation of a lubricant or, on the contrary, the creation of "frictional polymers,"
(ii) changes in the properties of the elements, for instance, changes in contact topography and surface composition,
(iii) changes in the interrelations between the elements, for instance, changes of the wear mechanisms under the action of the operating variables, or changes in the lubrication mode.

The other two main tribological characteristics (b) and (c), the friction-induced energy losses and the wear-induced material losses, may be expressed formally as:

$$\text{friction losses} = f\,(\text{operating variables; system's structure})$$
$$\text{wear losses} = f\,(\text{operating variables; system's structure})$$

Consequently, friction coefficient, f, and wear rate, w, of a tribological system may be expressed formally as:

$$\boxed{f = f\,(X;S)} \qquad \boxed{w = f\,(X;S)}$$

where X : Operating variables

$S = \{A, P, R\}$: System's structure

Although the parameter groups X and S are not independent variables since they are connected with each other through the tribological interrelations R, the above symbolic representation of friction and wear characteristics can be conveniently used as a starting point for the practical application of a systems methodology in order to influence or to mitigate friction- and wear-induced energy and material losses.

7. TRIBOLOGICAL SYSTEM EFFECTIVENESS

The friction, lubrication, and wear processes involved in any tribological system determine its technical functional behavior. These processes may change the properties of the components, as for example the surface roughness of the interacting material elements, and may thus influence the performance of the whole system. In this final section, first the influence of tribological processes on the dynamic behavior of tribological systems will be considered. Then the aspects of failure, safety, and reliability of tribological systems will be discussed.

7.1 Dynamics of Tribological Systems

The functional behavior of any tribo-mechanical system is connected, by definition, with relative motion of one or more components of the system. As described in Section 5, this motion may constitute a transfer of work, information, or material through the tribo-mechanical system. In any case, the motion and the dynamics of the whole system are influenced by the interfacial friction processes between the moving components. Clearly, the details of the influences of tribological processes on the dynamics of motion may be quite different for the various tribo-mechanical systems, for example, ball bearings, slideways, or metal machining systems. Considering the functional transmission of motion through the various types of tribo-mechanical systems, the influences of the tribological processes may lead to unwanted vibrations of the moving parts and to "stick-slip" motion. These disturbing influences on the functional behavior can be observed in various tribo-mechanical systems; from the squealing of brakes to the chattering of machine tools during cutting processes.

Many tribo-mechanical systems whose functional purpose is connected with the transmission of motion can be modelled in a simplified manner by the configuration shown in Figure 7. The model system consists of a body (1) of mass m_1, moving relatively to its counterpart (2) of mass m_2 fixed to the ground via a spring with a spring constant C_{s2} and a damper with a damper constant C_d. The body (1) is driven via the spring C_{s1} at constant velocity $v_o = s/t$.

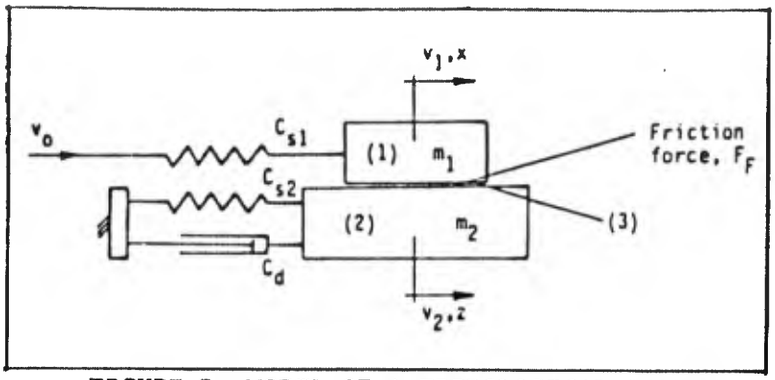


FIGURE 7 MODEL OF A TRIBOLOGICAL SYSTEM

The motion of body (1) of velocity v_1 and distance x relative to body (2) of velocity v_2 and distance z is influenced by the friction force F_F acting in the interface (3) between body (1) and body (2). From the following simple qualitative consideration it follows that the type of motion is determined by the value of the friction force at $v_{rel} = 0$ and the dependence of the friction force on the velocity $F_F = f(v)$. Let the initial state of the system, shown in Figure 7, be such that the springs C_{s1} and C_{s2} are uncompressed and m_1 and m_2 are at rest. When the motion of velocity v_o is introduced there will be no movememt of m_1 relative to m_2 ("stick phase") until the driving force on m_1 is high enough to overcome the (static) friction force between m_1 and m_2. If then the motion of m_1 relative to m_2 starts ("slip" phase), the springs decompress. Thus, the driving force is lowered by a certain amount. If now the driving force on m_1 falls below the (kinetic) friction force, a second "stick" phase may evolve.

This in turn leads to an increase of the driving force until the motion of the second slip phase starts, and so on.

Through appropriate modelling it may be possible to characterize the dynamic behavior of tribological systems by differential equations and to simulate it with the help of an analogue computer.

From the results of an analogue computer simulation, some general conclusions on the dynamic behavior of tribological systems may be drawn [7]

Generally speaking for a lubricated sliding system, depending on the "operating point" within the Stribeck curve, a different dynamic behaviour of the tribo- mechanical system can be observed. Without going into numerical details, the following three different general patterns of the motion behavior of the modelled tribo- mechanical system of Figure 7 can be distinguished:

(i) For the condition of friction around the minimum of the Stribeck curve, the system is unstable and the motion following a disturbance is divergent, i.e., the system excites itself to vibrations, as shown in Figure 8.

(ii) For the conditions of friction on the left part of the Stribeck curve the typical stick-slip motion diagram results, as in Figure 9.

(iii) For the conditions of friction on the right part of Stribeck's curve the system is stable, i.e. vibrations introduced to the system are damped automatically. This behavior can be seen in Figure 10 for five different slopes to the right part of the Stribeck curve and five operating values of the friction coefficient f.

The results confirm the experimental observation that stick-slip effects are likely to occur if the slope of the velocity curve is negative or equal to zero, $\frac{df}{dv} \leq 0$, as in the left part of the Stribeck curve. Thus, stick-slip effects may occur only under conditions of solid friction, boundary, or mixed lubrication, but are unlikely to occur under conditions of hydrodynamic lubrication.

7.2 Failure and Reliability of Tribological Systems

In reviewing the causes of failure of mechanical systems, Collacot, in his book on "Mechanical Fault Diagnosis and Condition Monitoring" distinguishes between the following main aspects[6]:

(a) Service failures
(b) Fatigue
(c) Excessive deformation
(d) Wear
(e) Corrosion
(f) Blockaages
(g) Design, manufacturing and assembly causes of failure

This compilation of the main classes of failure already indicates that there are several nontribological causes which may lead to failure of mechanical equipment. This can also be seen from the data of Tables 3 and 4 in which failure characteristics of typical tribo-mechanical systems, namely rolling and sliding bearings and mechanical clutches, are listed. The data have been compiled by the insurance company ALLIANZ[7]. Table 3 contains the results of investigations of the causes of 1400 rolling-bearing failures and 530 sliding-bearing

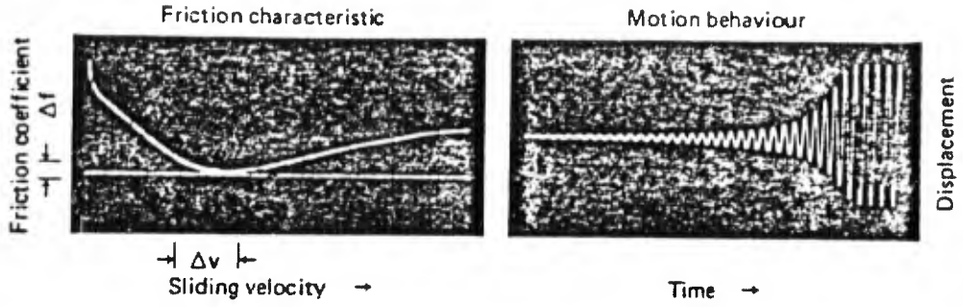


FIGURE 8 INSTABLE MOTION BEHAVIOR FOR CONDITIONS AT THE MINIMUM OF THE STRIBECK CURVE

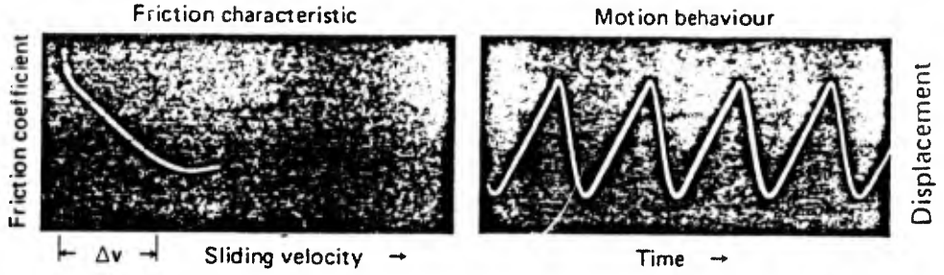


FIGURE 9 STICK-SLIP OSCILLATIONS FOR CONDITIONS ON THE LEFT PART OF THE STRIBECK CURVE

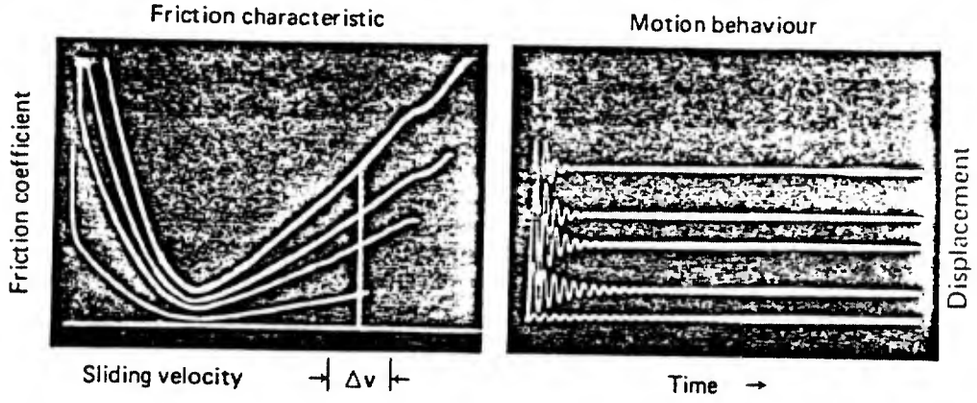


FIGURE 10 STABLE MOTION BEHAVIOR FOR CONDITIONS ON THE RIGHT PART OF THE STRIBECK CURVE

failures. It can be seen that for these tribo-mechanical systems, about 30% of the functional failures are due to wear processes. Table 4 contains the percentage of the main types of damage of mechanical clutches. Also in this case, the damage occurring at the load-transmitting surfaces has been found to constitute about 30% of the total failure causes. These examples show that, besides the tribo-induced causes of failure, various nontribological causes may lead to failure of mechanical equipment.

Causes of failure	Occurrence (%)	
	rolling bearings	sliding bearings
manufacturing faults	14.4	10.7
design and calculation faults	13.8	9.1
materials faults of components	1.9	3.6
service faults, maintenance faults, failure of monitoring equipment	37.4	39.1
wear	28.5	30.5
failure through external causes	4.0	7.0

TABLE 3 CAUSES OF FAILURE OF ROLLING AND SLIDING BEARINGS

Damage-types	Occurrence (%)
breakages due to overstressing	60
scuffing, seizure	18
mechanical and corrosive surface damage	15
cracks	5
deflections, deformations	2

TABLE 4 TYPES OF DAMAGE OF MECHANICAL CLUTCHES

Next, the question may be discussed, how the tribological processes, i.e. the friction and wear processes, influence and disturb the behavior of a given tribological system. This is illustrated schematically in Figure 11.

In the upper part of Figure 11, a typical tribological system, namely a gear box, is shown schematically. The technical function of the system is to transform certain inputs, namely angular velocity and torque, into outputs which are used technically. The technical function can then be described formally as a transformation of the inputs into the outputs via a certain transfer function.

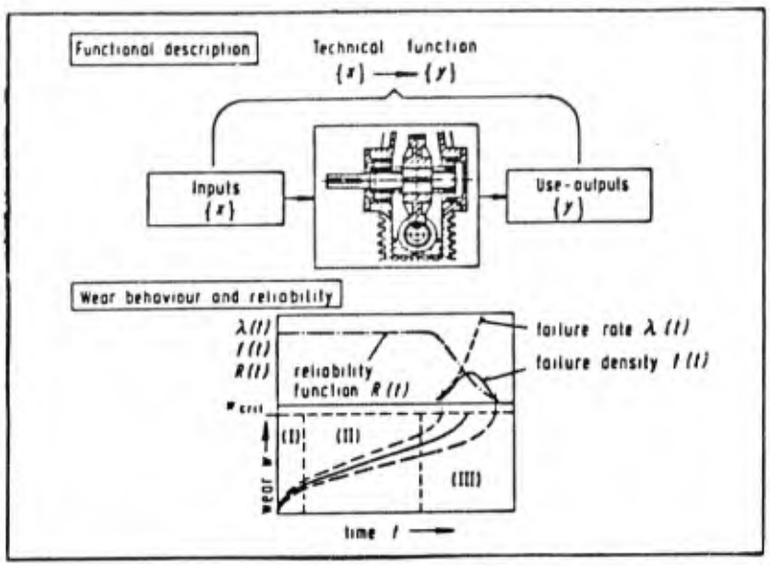


FIGURE 11 INFLUENCE OF WEAR-INDUCED CHANGES OF SYSTEMS STRUCTURE ON FUNCTION AND RELIABILITY

Now, the dynamic performance of the system is accompanied by perturbations on their function and structure. For example, through the action of wear processes, the properties of the moving components may be changed and a certain material loss-output may result. As illustrated in the lower part of Figure 11, for the loss-output of a tribo-mechanical system, three main different characteristics may be distinguished which are often observed experimentally:

(i) Self accommodation (running in)
(ii) Steady state
(iii) Self acceleration

These three modes of changes in the systems structure may follow each other in time, as indicated in the graph in the lower part of Figure 11. If then the wear rate reaches a maximum admissible level, the systems structure has changed in such a way that the functional input-output relations of the system are disturbed severely. Repeated measurements show random variations in the data, as indicated by the dashed lines in the wear diagram of Figure 11. From sample functions of the wear process, a distribution, a failure rate, and a corresponding reliability function results.

In a quantitative way the reliability of a mechanical system may be characterized mathematically as follows:

Generally, the reliability of a mechanical system is expressed by a probabilistic function R(t) based on the following definitions:

$F(t)$	:	probability distribution function of the time of failure
$f(t)=\frac{dF(t)}{dt}$	:	density function
$\lambda(t)=\frac{F(t)}{1-F(t)}$	:	failure rate (λ (t)dt is a conditional probability that the system will fail during the time t + dt under the condition that the system is safe until the time t)
$R(t) = 1-F(t)$	:	reliability function
MTBF	:	mean time to failure (measure of reliability for repairable equipment

In some cases, the failure rate (λt) of a component in a system can be estimated from the point of view of the physical behavior of the material used. Empirically and sometimes theoretically, the following probabilities are proposed:

(a) EXPONENTIAL DISTRIBUTION

$\lambda(t)$ = constant = c
$f(t) = C \cdot \exp(-Ct)$
$R(t) = \exp(-Ct)$

In this case the failure rate is constant. It means physically that any failure occurs accidentally without any accumulation of fatigue-like effects during its service time under certain stresses. Many kinds of electronic components follow this type of failure. Components in a machine break down in this mode when the failure is brittle fracture.

(b) RAYLEIGH DISTRIBUTION

$$\lambda(t) = Ct$$

$$f(t) = Ct \cdot \exp\left(\frac{-Ct^2}{2}\right)$$

In this case the failure rate increases with time. The constant, C, indicates the rate of deterioration of the component which depends upon the stress level applied to it.

(c) NORMAL DISTRIBUTION (TRUNCATED)

$$f(t) = (1/s\sqrt{2\pi}) \exp\left\{\frac{1}{2}\left(\frac{t-\mu}{s}\right)^2\right\}$$

Many components of machines obey this distribution, especially if the failure occurs due to wear processes. The failure rate of this distribution cannot be expressed in a simple form.

(d) WEIBULL DISTRIBUTION

$$\lambda(t) = \frac{C}{t_o} t^{C-1}$$

$$f(t) = \frac{C}{t_o} t^{C-1} \exp\left(-t^C/t_o\right)$$

This is a distribution with two paramenters, t_o, the nominal life and the constant C. The distribution is found to represent failure of many kinds of mechanical systems, such as fatigue in ball bearings.

(e) GAMMA DISTRIBUTION

$$f(t) = C\frac{(Ct)^{x-1}}{\Gamma(x)}\exp(-Ct)$$

where $\Gamma(x)$ is a Gamma function. This is also a distribution with two parameters. Theoretically, the importance of this distribution is attributed to the fact that the equation is an x-fold convolution of the exponential function. It means physically that a component fails at x-th shock which occurs as a Poisson statistical process.

These distributions are representative ones which appear in the failure process of various components and systems. As a

general overview, in Table 5, a compilation of the phenomena of deterioration and the mode of failure in connection with the underlying physical processes is given (the table is due to Yoshikawa)7.

Physical process	Deterioration (Mode of failure)						Probability distribution
	Topological change		Geometrical		Physical property		
	Fixing	Separation	Micro	Macro	Bulk	Surface	
Fracture		•					Exponential
Yielding				•			Exponential
Fatigue		•	•		•		Weibull
Creep		•		•	•		Normal
Diffusion					•		Normal
Corrosion						•	Rayleigh
Erosion						•	Rayleigh
Rusting	•					•	Rayleigh
Wear			•			•	Normal
Adhesion	•						Gamma
Staining			•			•	Exponential

TABLE 5 PHENOMENA OF DETERIORATION AND MODE OF FAILURE

It can be seen that different failure modes and different elementary failure processes are associated with different types of failure distribution functions. It follows, in turn, that conclusions on the type of failure mechanism may be drawn from the experimental determination of failure distribution curves. For a great deal of tribological systems failing as a consequence of wear processes, the failure behavior is characterized by the normal distribution or the Weibull distribution. If for a given type of tribological system the failure mode and the type of failure distribution are known, this knowledge can be used to improve the reliability of the system.

To conclude the discussion on the failure and reliability of tribological systems, the dependence of the failure rate on the operating duration of a system should be considered. If the failure rate is plotted as a function of time, a curve known as the "bath-tub-curve" is found, as shown in Figure 12.

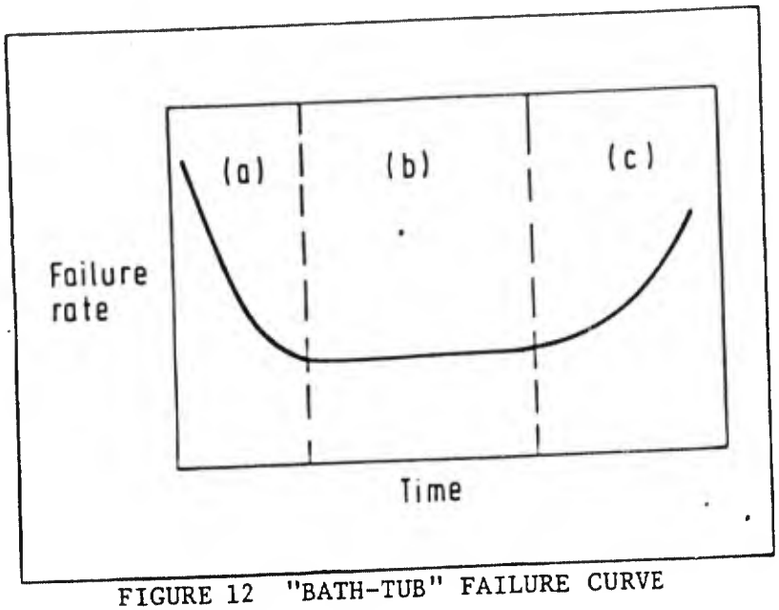


FIGURE 12 "BATH-TUB" FAILURE CURVE

In this curve three regimes can be distinguished:

(a) early failures
(b) random failures
(c) wear-out failures.

None of the distribution curves discussed above have this bath-tub-shaped failure curve, but an approximation may be obtained by selecting an appropriate probability density function for each of the three regimes. Regime (a) describes the region of the "infant death" of the system. This regime is characterized by a decrease of the failure rate with time, for example in running-in. The regime (b) of constant failure rate is the region of normal running. Here, failure occurs as a consequence of statistically independent factors. Regime (c) is characterized by an increase of the failure rate with time. Here, failure may be due to aging effects. As described above, for a great deal of tribo-induced failures the failure rate increases with time. Thus, region (c) of the "bath-tub-curve" of Figure 12 appears to be relevant for the normal mode of wear-induced failure of mechanical systems.

8. REFERENCES

The literature in the field of tribology is extremely voluminous owing to the vast number of topics included. At present, 6000 - 8000 new titles are published per year. An international and comprehensive bibliography is provided by the DOCUMENTATION TRIBOLOGY. Information about this DOCUMENTATION can be obtained from the Bundesanstalt fur Materialprufung (BAM), Dokumentations-stelle Rheologie und Tribologie, Unter den Eichen 87, D-1000 Berlin 45, West Germany.

In the following, only a selection of books in the field of tribology are listed, which were published after 1970.

8.1 Recent Books on Tribology (Selection)

1. Barwell, F.T., "Bearing Systems: Principles and Practices," (Oxford: University Press, 1978).

2. Benzing, R., Goldblatt, I., Hopkins, V., Jamison, W., Mecklenburg K., Peterson, M., "Friction and Wear Devices," (Park Ridge: American Society of Mechanical Engineers, 1976).

3. Bowden, F.P. and Tabor, D., "Friction - An Introduction to Tribology," (London: Heinemann, 1973).

4. Buckley, D.H., "Surface Effects in Adhesion, Friction, Wear, and Lubrication," (Amsterdam: Elsevier, 1977).

5. Cameron, A., "Basic Lubrication Theory," (London: Longman, 1971).

6. Collacott, R.A., "Mechanical Fault Diagnosis and Condition Monitoring," (London: Chapman and Hall, 1977).

7. Czichos, H., "Tribology - A Systems Approach to the Science and Technology of Friction, Lubrication and Wear," (Amsterdam: Elsevier, 1978).

8. Dowson, D., "History of Tribology," (London: Longman, 1979).

9. Dumbleton, J.H., "Tribology of Natural and Artificial Joints," (Amsterdam: Elsevier, 1981).

10. Engel, P., "Impact Wear of Materials," (Amsterdam: Elsevier, 1978).

11. Habig, K.H., "Hardness and Wear of Materials (in German)," (Munchen: Hanser, 1980).

12. Halling, J. (Ed.), "Principles of Tribology, "(London: MacMillan, 1975).

13. Ling, F.F., "Surface Mechanics," (New York: Wiley, 1973).

14. Moore, D.F., "Principles and Applications of Tribology," (Oxford: Pergamon, 1975).

15. Neale, M.J. (Ed.), "Tribology Handbook," (London: Butterworths, 1973).

16. Peterson, M.B. and Winer, W.O. (Eds.), "Wear Control Handbook," (New York: ASME, 1980).

17. Suh, N.P. and Saka, N. (Eds.), "Fundamentals of Tribology," (Cambridge, Mass: MIT Press, 1980).

8.2 Terminology

18. Standard DIN 50 320. "Wear: Terms, Systematic Analysis of Wear Processes, Classification of Wear Phenomena," (Berlin: Beuth Verlag, Dec. 1979).

8.3 Tribology Reports

19. "Lubrication (Tribology) Education and Research (Jost Report)," Department of Education and Science, HMSO, London, 1966.

20. "The Introduction of a New Technology," Department of Trade and Industry, HMSO, 1973.

21. "Research Report T 76-38 Tribologie (Code BMFT-FBT 76-38)." Bundesministerium fur Forschung und Technologie (Federal Ministry for Research and Technology), West Germany 1976.

22. "Strategy for Energy Conservation Through Tribology," ASME, New York, Nov. 1977.

23. Jost, H.P. and Schofield, J., "Energy Saving Through Tribology: A Techno-Economic Study", Proc. Inst. Mech. Engrs, London, Vol. 195 No. 16 (1981) 151-173.

SURFACE INTERACTIONS

Nam P. Suh
Massachussetts Institute of Technology
Cambridge, MA 02139
U.S.A.

1. INTRODUCTION

1.1 Three Apsects of Tribological Problems

Tribology is concerned with the science of the interface between two surfaces in relative motion. The nature and the consequence of the interactions that take place at the interface control the friction and wear behavior of the materials involved. During the interactions, forces are transmitted; energy is consumed; physical and chemical natures of the materials are changed; and the geometry of the surface topography is altered. All these consequences of the surface interactions satisfy the laws of nature. Therefore, the essence of tribology is two-fold: understanding the nature of these interactions and dealing with the technological problems created when such interactions occur. The purpose of this series of lectures is to expound on the basic mechanisms that govern the interfacial behavior.

There are three fundamental aspects to the tribological science that must be understood to deal with the technological problems:

(i) The effect of environment on surface characteristics through the physicochemical interactions.
(ii) The interaction of the surfaces in contact which results in force generation and transmission between the surfaces.

(iii) The behavior of the material near the surface in response to the external force acting on the surface.

An understanding of all three aspects of tribology requires a multidisciplinary background which is difficult to acquire. As a consequence, most researchers and practitioners only deal with a limited aspect of tribology. Perhaps, this may be one of the reasons why progress in the tribology field has been slow.

To tribologists who are mainly concerned with boundary lubrication, the first two aspects of tribology (i.e., physicochemical interactions of the surface with the environment and the asperity interactions) are of primary interest. However, the third aspect of tribology (i.e., the macroscopic deformation of the surface in response to the external force) cannot be ignored since it affects the two aspects. The physical and chemical interactions of lubricants (and gaseous environment) with the sliding surface produce compounds which are actually present at the interface. These materials in turn affect the nature of force generation and transmission at the asperity contact, and thus, the frictional force. Unfortunately, the mechanisms of boundary lubrication are yet to be fully understood. To date, lubricants and additives to lubricants have been developed through much trial and error. Although careful chemical analyses of the layers formed on the surface have been done, the mechanisms by which these additives impart beneficial effects have not yet been established.

Those interested in studying the frictional behavior of materials must comprehend the details of the chemical and physical interactions at the interface. This requires understanding of all three aspects of tribology. The contribution of each aspect of tribology to friction changes as a function of sliding distance (or time) and the environment. Therefore, the coefficient of friction of the same pair of materials differs depending on the specific application.

Wear is primarily a consequence of the response of materials to given surface tractions. In order to understand wear, it is therefore necessary to deal with the mechanics of deformation of the surface layer and the mechanisms of force transmission at the surface. In a given tribological situation, wear particles can be generated by different mechanisms. This is often confusing to those who have not been initiated into the tribology field, since they expect a single mechanism to be responsible for all wear particle

generation. What is important is to recognize the wear rate determining wear process in a given situation.

It should also be noted that the surface properties, both chemical and physical, are different from the bulk properties. The chemical composition of the surface layer may be different from the bulk; dislocations have different mobility and experience different forces; electronic configurations at the surface may be different from the bulk configurations; and even the atomic position at the surface differs from that of the bulk. Therefore, it is necessary to establish the fundamental understanding of the surface chemistry and physics in order to make future progress in tribology.

1.2 Phenomenological Observations Related to Friction and Early Theories

Friction exists between any two sliding surfaces in relative motion. This fact has been known to mankind from prehistoric days and has been used to man's advantage, albeit intuitively and empirically in most cases. Many of the macroscopic phenomenological observations made on frictional behavior still form the basis for much of the current engineering practice. In recent decades, the scientific efforts made have been attempting to provide rational explanations to these ancient observations through scientific investigations. Therefore, it is most appropriate to discuss the phenomenological aspect of friction as part of these introductory remarks. Frictional behavior is affected by the following factors:

a) Kinematics of the surfaces in contact, i.e., the direction and the magnitude of the relative motion between the surfaces in contact.
b) Externally applied load and/or displacements.
c) Environmental conditions such as temperature and lubricants.
d) Surface topography.
e) Material Properties.

The above list of important factors that control friction clearly indicates that the coefficient of friction (i.e, the ratio of tangential force to the normal load) is not a simple material property.

Most metals behave in such a manner that the coefficient of friction under normal sliding conditions is, to a first approximation, independent of normal load and sliding speed. However, as the normal load is increased to a very high value, such as found in metal cutting, the coefficient of friction

often decreases with the normal load and the sliding speed. In the intermediate normal load and speed ranges, the frictional coefficient may reach a peak value as shown in Figure 1.1.

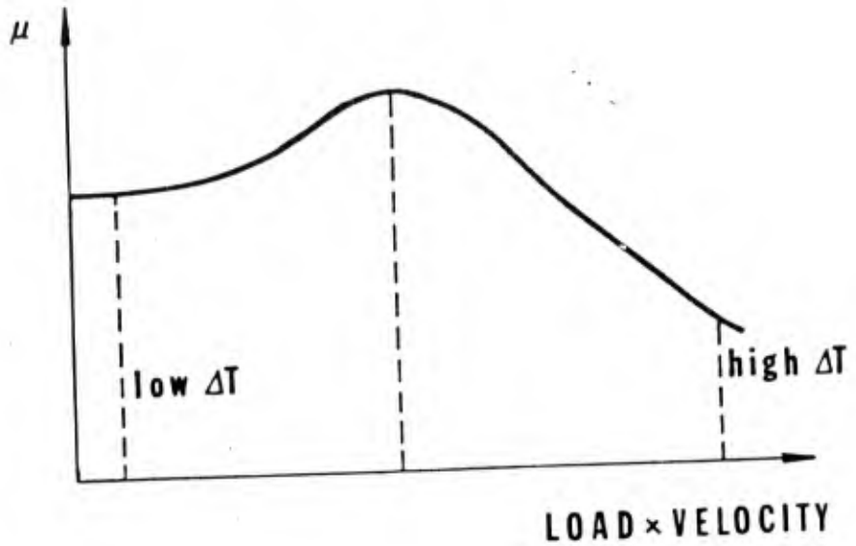
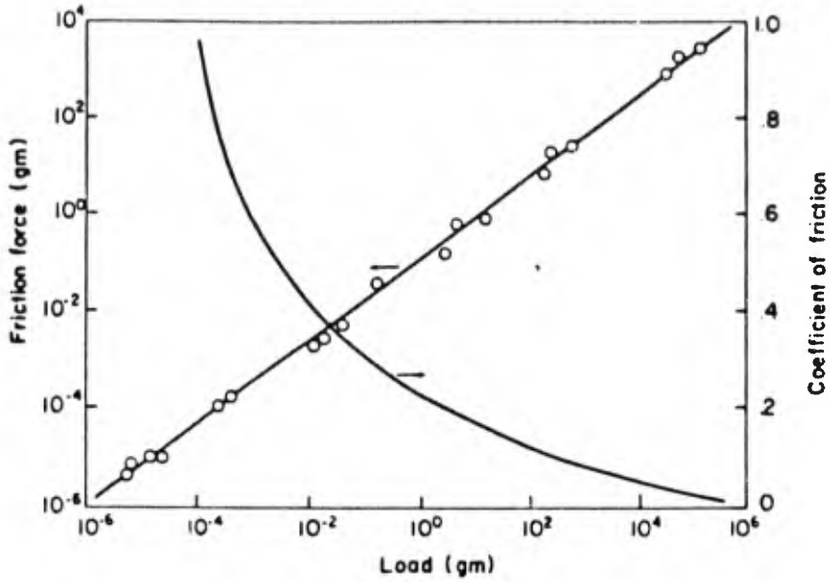


FIGURE 1.1 QUALITATIVE REPRESENTATION OF THE COEFFICIENT OF FRICTION OF METALS VERSUS LOAD TIME SPEEDS. ΔT IS THE INTERFACIAL TEMPERATURE RISE

Polymeric materials behave differently from metals in that their coefficient of friction is a function of the normal load and sliding velocity. As the load is increased, most polymers, be it a thermoplastic or a thermoset, exhibit lower friction coefficients, even at a low sliding speed as shown in Figure 1.2. As the sliding speed is increased, the coefficient of friction increases. At a very high speed, the coefficient of friction decreases with an increase in sliding speed, as shown in Figure 1.3.

In the past these observations were explained in terms of the adhesion model. In essence, this theory assumes that the surface consist of asperities and the interface consist of asperity contacts, as shown in Figure 1.4. The real area of contact is much smaller than the apparent area of contact in most cases, except in such applications as in metal cutting where the real area of contact approaches the apparent area of contact. When a relative motion is imparted to the interface by applying a tangential force, each asperity contact welds together and shears to accomodate the relative motion. Then,



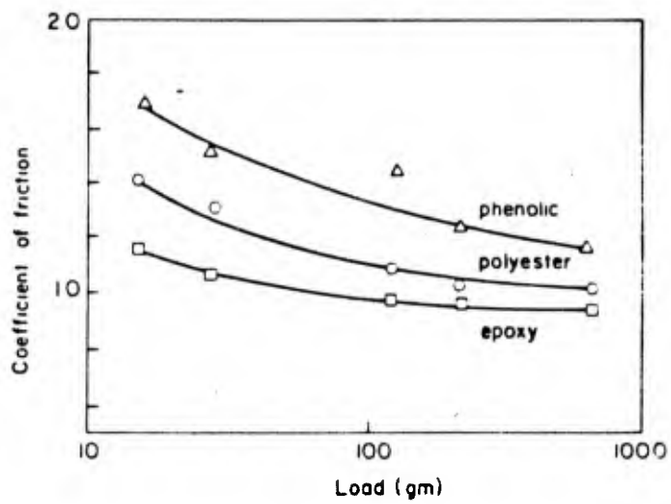


FIGURE 1.2 COEFFICIENT OF FRICTION OF THERMOPLASTICS AND THERMOSETTING PLASTICS

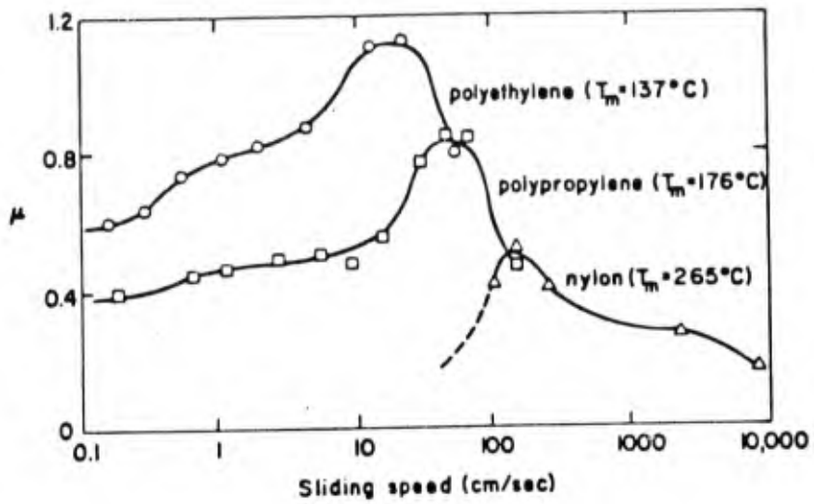


FIGURE 1.3 COEFFICIENT OF FRICTION OF THERMOPLASTICS AS A FUNCTION OF SLIDING SPEED

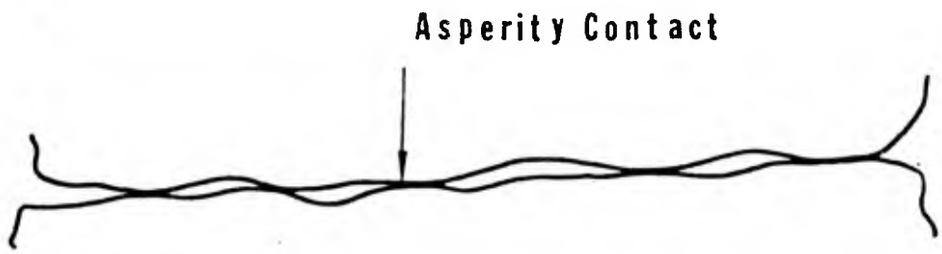


FIGURE 1.4 ASPERITY CONTACT

to a very rough first approximation, the frictional force is given by

$$F = A_r \tau \tag{1.1}$$

where τ is the shear stress at each junction and A_r is the real area of contact. The real area of contact was related to the hardness of the material based on the assumption that the real area of contact must be large enough to support the given normal load. That is,

$$A_r = \frac{L}{H} \tag{1.2}$$

where H is the indentation hardness of the metal. τ is the critical shear strength of the material which must be overcome to satisfy kinematics of the sliding motion. A great deal of discussion has taken place as to how τ is related to H. Many believers of the adhesion theory assumed that τ must be greater than the critical shear strength of the bulk material, due to the work-hardening of the surface layer during sliding. This type of argument was necessitated by the low predicted values of the friction coefficient. If we assume that τ is equal to the critical shear strength of the bulk k, then

$$F = A_r \tau = \frac{L}{H}\tau = \frac{L}{6k}k = \frac{L}{6}$$

$$\mu = \frac{F}{L} = \frac{1}{6} \tag{1.3}$$

which is much smaller than the typical observed values under steady state sliding conditions. In order to improve the correlation between the experimental results and the adhesion theory, a number of theories have been advanced.

One of these theories is due to Rabinowicz, who argued because of the surface energy of adhesion, the actual area of contact is much larger than that given by Eq. (1.2)[1]. If the overall surface energy change is denoted by W_{ab} (i.e., the surface energy of adhesion), then

$$W_{ab} = \gamma_a + \gamma_b - \gamma_{ab} \tag{1.4}$$

where γ_a and γ_b are the surface energies of the two contacting surfaces and γ_{ab} is the interface energy. The sum W_{ab} is always positive, i.e., the overall energy is decreased by bonding. Idealizing the indentation of asperities as an indentation by a conical indenter of material b penetrating

into a half-space of material a, as shown in Figure 1.5, the work done by the normal load L during an infinitesmal indentation dx may be equated to the difference in the work done in deforming the material plastically and the surface energy change, i.e.,

$$L\,dx = \pi r^2 H\,dx - (2\pi r) W_{ab} \frac{dx}{\sin\theta} \tag{1.5}$$

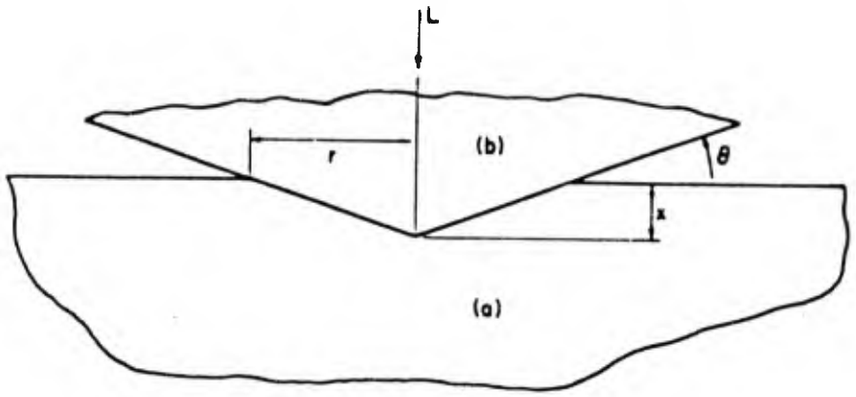


FIGURE 1.5 INDENTATION OF CONICAL INDENTER TO REPRESENT THE INDENTATION OF ASPERITIES

Eq. (1.5) may be rewritten for the area in contact as

$$\pi r^2 = \frac{L}{H} + \frac{2\pi r}{\sin\theta} \frac{W_{ab}}{H} \tag{1.6}$$

Eq. (1.6) states that when the interfacial energy change is included in considerations the projected area (πr^2) is larger than that given by Eq. (1.2) by an amount ($2\pi r/\sin\theta$) (W_{ab}/H). Substituting Eq. (1.6) into Eq. (1.3) the expression for the coefficient of friction gives

$$\mu = \frac{F}{L} = \frac{k}{H} \frac{1}{1 - 2\,W_{ab}/rH\sin\theta} = \frac{k}{H}\left(1 + K\frac{W_{ab}}{H}\right) \tag{1.7}$$

where k is a geometric factor. According to Eq. (1.7), the coefficient of friction is high when the ratio of the surface energy of adhesion W_{ab} to hardness H is large and when roughness angle θ is small. In support of this theory

Rabinowicz presented a correlation to μ vs. W_{ab}/H as shown in Figure 1.6.

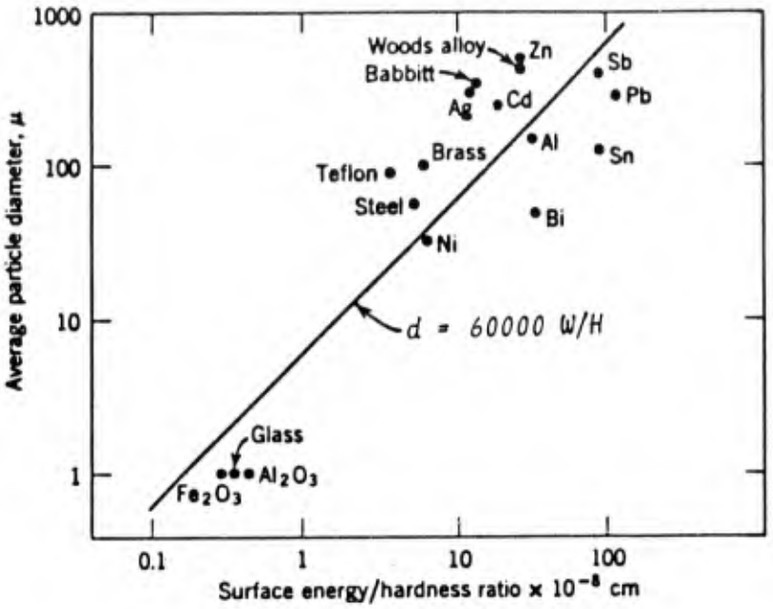


FIGURE 1.6 PLOT OF AVERAGE WEAR PARTICLE DIAMETER AGAINST THE W/H RATIO- FOR METALLIC AND NONMETALLIC MATERIALS

In spite of the correlation, some basic questions raise issues on the importance of surface energy considerations in determining friction. The most obvious difficulty is the relatively small magnitude of the surface energy change in comparison to the total work done. An order of magnitude analysis of Eq. (1.6) shows that the first term of RHS of Eq. (1.6) is much larger than the second term of RHS. Another difficulty is that most surfaces are so highly contaminated by absorbants and impurity atoms in the metal that the validity of experimentally measured W_{ab} for various surfaces is in doubt[2]. It will be shown in a later section that the correlation shown in Figure 1.6 can also be explained in terms of the mechanical properties of the surface.

Another adhesion model which also stresses the real area of contact is advanced by A. P. Green[3, 4]. Green analyzed the deformation of the surface asperity contact using the slip line field for a rigid-perfectly plastic material. The plasticity analysis of the junction shown in Figure 1.7 showed that the coefficient of friction can be larger than 0.17 as shown in Figure 1.8. For a typical surface, the angles δ and θ are less than 10°, yielding μ of approximately 0.6. Based on the analysis and experiments with plasticine, Green

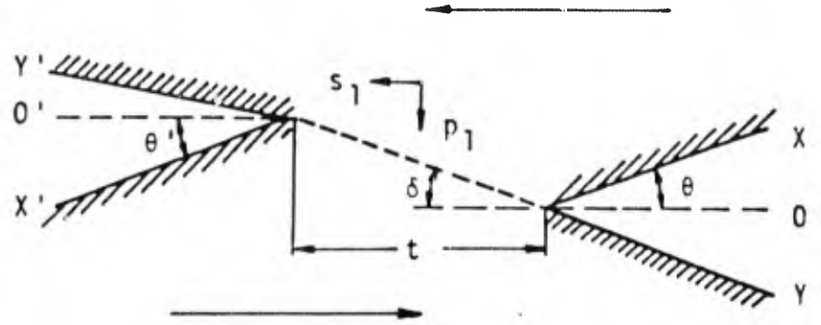


FIGURE 1.7 STRONG JUNCTION DURING STEADY SLIDING

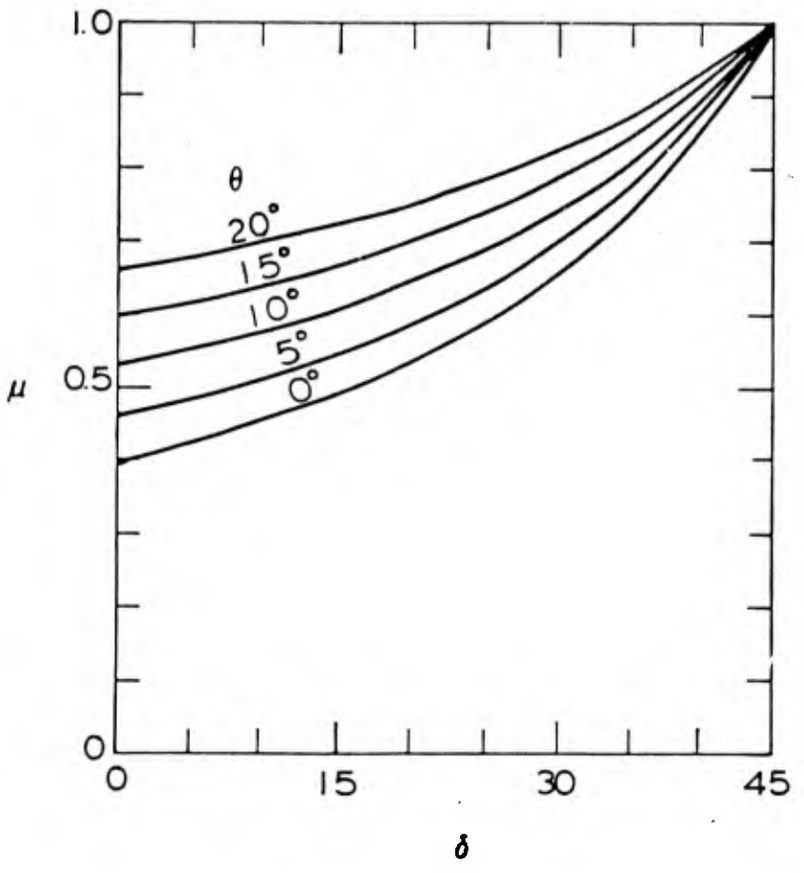


FIGURE 1.8 THEORETICAL RESULTS OF ARCHARD FOR STRONG JUNCTIONS FOR μ VS. SLOPE OF ASPERITIES

concluded that some of the strongly adhering junctions may also support a tensile stress during the deformation process. Since the total compressive normal load must be supported by the asperity junctions in compression, the equilibrium consideration requires the real area of contact under the steady state sliding situation be greater than when all the junctions are under compression. Therefore, he reasoned that the coefficient of friction can be extremely large. This line of thought cannot, however, explain large coefficients of friction observed, even when adhesion is absent.

One of the early attempts to explain friction was to relate it to surface roughness, because surface is not generally smooth, consisting of asperities (i.e., short range perturbations from the mean) and waviness (i.e., long range pertubations from the mean). The roughness theory assumed that the frictional force is equal to the force required to climb up the asperity of slope θ. Then, the coefficient of friction is given by:

$$\mu = \tan \theta \tag{1.8}$$

It is, however, clear that asperities undergo deformation due to the sliding action rather than simply sliding over each other. Moreover, asperities cannot continue to climb up asperities throughout the sliding action.

Another school of thought attributes the frictional force to a combined effect of adhesion, plowing, and roughness. Shaw and Macks expressed the frictional force to be the sum of the adhesion component, given by Eqs. (1.1) and (1.2), and the plowing component, which can be expressed as

$$\mu = \frac{\tau}{H} + \tan \theta + P \tag{1.9}$$

where P accounts for the plowing force[5]. Kragelskii also believes that plowing contributes to the frictional force as well as adhesion[6]. He expressed the coefficient of friction as

$$\mu = \frac{\tau_o}{P_r} + \beta + K\alpha_h \; h/r \tag{1.10}$$

where τ_o is the shear strength of the surface due to molecular bonds when the normal pressure is equal to zero; P_r is the contact pressure which, for a surface that has been run-in, depends on elastic modulus and the surface roughness parameter; α_h is the coefficient of hysteresis losses in friction; β is the molecular bond strengthening coefficient; h

is the height of an asperity; r is the asperity radius; and K is the friction parameter. In many sliding situations the ideas embodied in developing Eqs. (1.9) and (1.10) are closer to the real picture than the adhesion theory of friction. However, as shall be shown later, the frictional force is not a constant as implied by these equations; Eqs. (1.9) and (1.10) may not satisfy the equilibrium condition for a vertical component of forces; and contrary to Eq. (1.10), even after run-in, there can be plastic deformation due to adhesion which will affect the real area of contact. A new theory of friction which overcomes these difficulties is presented in a later section.

The friction behavior is greatly affected by the interfacial temperature and the environment. The interfacial temperature and environment govern the degree of adhesion, the nature of chemical reaction at the interface, and the hardness of the surfaces. In normal sliding situations at low speeds, the interfacial temperature rise is small, relative to its melting points. In high speed cutting of metals the interfacial temperature rise can be quite significant. As a consequence, the frictional behavior can be very different in two situations. At low temperatures, mechanical behavior of solids dominate the frictional behavior; whereas at high temperatures (i.e., when the homologous temperature is greater than 1/2), chemical interactions and mutual solubility may govern the frictional behavior.

1.3 Phenomenological Observations Related to Wear and Earlier Theories

Wear of materials occurs by many different mechanisms depending on the materials, the environmental and operating conditions, and the geometry of the wearing bodies. These wear mechanisms can be classified into two groups as shown in Table 1.1; those primarily dominated by the mechanical behavior of solids and those primarily dominated by chemical behavior of materials. In many wear situations there are many mechanisms operating simultaneously, but there is usually the rate determining process which must be identified to deal with the wear problem. What determines the dominant wear behavior are mechanical properties, chemical stability of materials, temperature, and operating conditions. More commonly known phenomenological aspects of the wear behavior of metals and polymers under sliding conditions will be described in this section.

When two materials slide against each other to a very rough approximation, the wear volume W is linearly proportional to the distance slide, S, and normal load, L,

A) Wear processes which are primarily dominated by the mechanical behavior of materials under a given loading condition

Type	Typical Characteristics & Definitions	Observed in
Sliding Wear (Delamination Wear)	Plastic deformation, crack nucleation and propagation in the subsurface	Sliders, bearings gears, cams where surfaces undergo relative motion.
Fretting Wear	The early stages of fretting wear is the same as sliding wear but depends on relative amplitude. The entrapped wear particles can have significant effect on wear. The relative displacement amplitude is important	Press fit parts with a small relative sliding motion.
Abrasive Wear	Hard particles or hard surface asperities plowing and cutting the surface in relative motion	Sliding surfaces, earth removing equipment.
Erosive Wear (Solid particle Impingement)	Due to solid particle impingement, large subsurface deformation, crack nucleation and propagation	Turbines, pipes for coal slurries, helicopter blades.
Fatigue Wear	Fatigue crack propagation takes place, normally perpendicular to the surface, without gross plastic deformation under cyclic loading conditions	Ball bearings, roller bearings, glassy sol.d sliders.

B) Wear processes which are primarily controlled by chemical processes and thermally activated processes

Type	Typical Characteristics & Definitions	Observed in
Solution Wear	Formation of new compounds of a lower free energy of formations; high temperature; no gross plastic deformation, atomic level wear process.	Carbide tools in cutting steel at high speeds.
Diffusive Wear	Diffusion of elements across the interface.	HSS tool in cutting steel at high speeds.
Oxidative Wear	Formation of weak, mechanically incompatible oxide layer	Sliding surfaces in highly oxidative environment (not common)
Corrosive Wear	Corrosion of grain boundaries and formation of pits	Lubricated and corrosive atmosphere

TABLE 1.1 CLASSIFICATION OF WEAR

but inversely proportional to the hardness of the material. This may be expressed as

$$W \propto \frac{LS}{3H} \tag{1.11}$$

Eq. (11) is normally written as

$$W = K \frac{LS}{3H} \tag{1.12}$$

where K is a dimensionless proportionality constant commonly known as the wear coefficient. The factor 3 is a result of Archard's model for adhesion theory of wear[7]. This will be discussed later. There are many exceptions to the above statement. For example, a soft, commercially pure copper can be much more resistant to wear than AISI 1045 steel which is much harder. In normal sliding situations the wear coefficient for most metals is in the range of 10^{-4} to 10^{-3}.

In abrasive wear, the surface of a softer metal is plowed by wear particles or hard asperities. Abrasive wear follows the relationship given by Eq. (1.12) reasonably well, i.e., the harder the material the less is the wear rate. The wear coefficient of typical abrasive wear is of the order of 10^{-2} to 10^{-1}.

The wear particles generated by the abrasive mechanism resemble metal chips generated by cutting action.

In fretting wear the interface undergoes a small oscillatory motion, which results in wear of materials. The wear coefficient in this case depends on the amplitude of oscillation, as shown in Figure 1.9, when the relative displacement at the interface is less than a critical value. At large amplitudes of oscillation, the fretting wear coefficient approaches those of unidirectional sliding wear.

At high sliding speeds and loads, such as in metal cutting, the wear rate depends sensitively on the chemical nature of the material. This is shown in Figure 1.10 for various carbides cutting steel. In this case, the wear rate correlates with temperature, but does not increase linearly with normal load and the sliding distance. The hardness of the tool material has little bearing on the wear life of these tools. On the other hand, an alumina tool will not wear due to chemical instability but rather by mechanical deformation of the surface layer[8, 9].

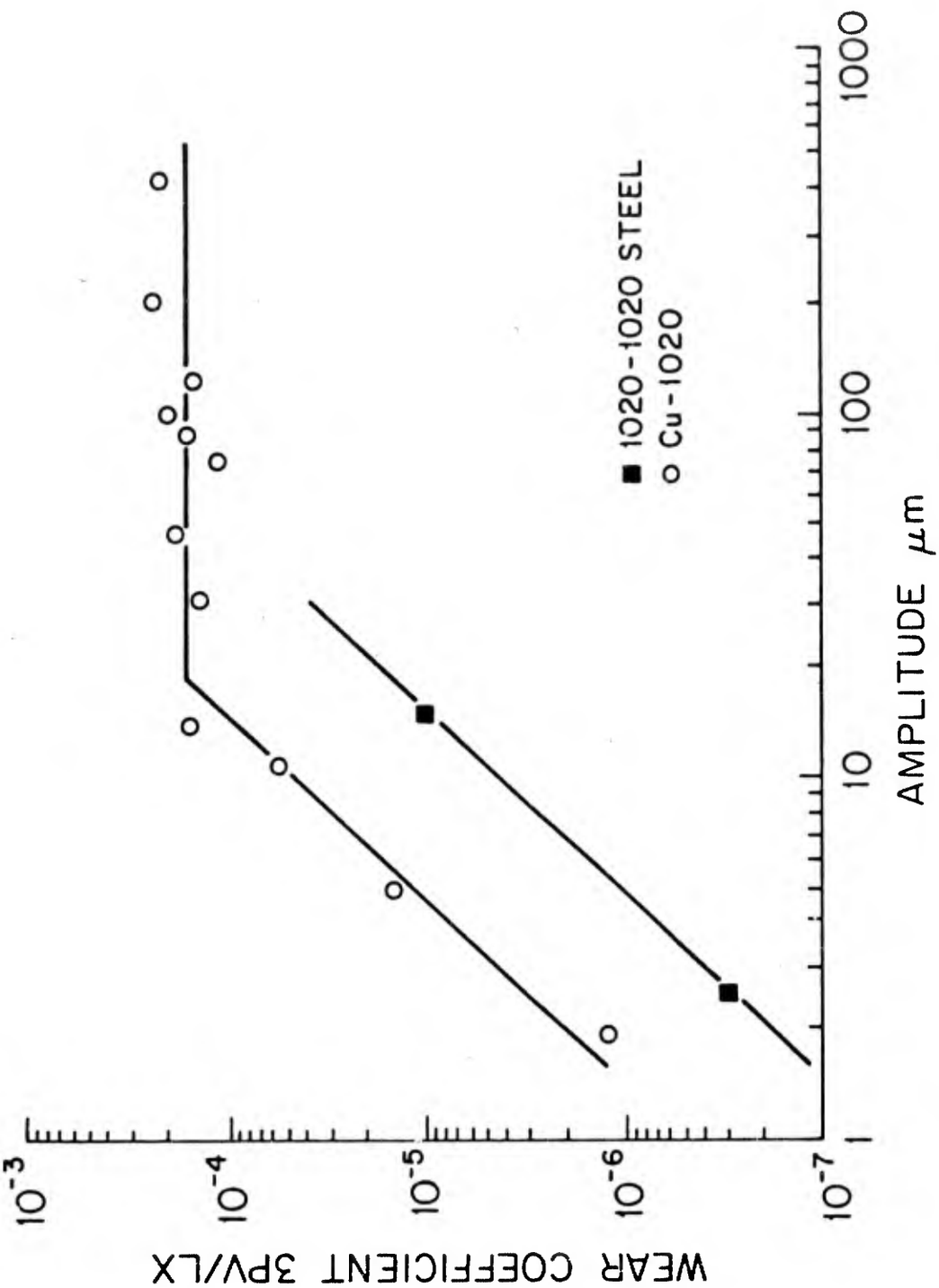


FIGURE 1.9 FRETTING WEAR COEFFICIENT AS A FUNCTION OF THE DISPLACEMENT AMPLITUDE

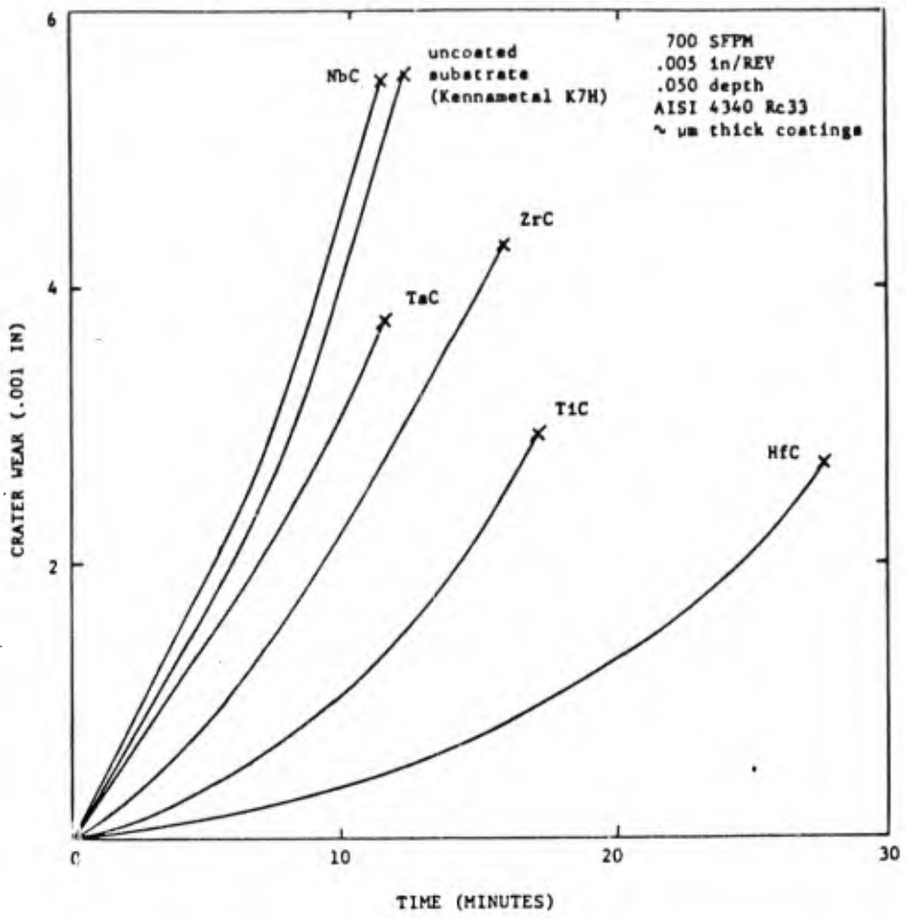


FIGURE 1.10 WEAR RATE OF VARIOUS CARBIDES

All these wear phenomena, except abrasive wear, were explained in terms of the adhesion theory of wear until the advent of the delamination theory of wear in 1972 for sliding wear and solution wear theory in 1979[10, 9]. The adhesion theory states that the wear of materials is due to the welding of asperity junctions which create a hemispheric wear particle when the weaker material near the welded junction fractures. Archard developed a mathematical model to describe this process, which yielded Eq. (1.12)[7]. The difficulty with this model was that there were too many exceptions and that this theory could not form the basis for improvement and development of wear resistant materials. It also violates the conservation law for energy because the actual work done is several orders of magnitude larger than the fracture energy.

The fact that this adhesion is not a complete description of the sliding wear behavior may be appreciated by examining the physical significance of the wear coefficient. The wear coefficient is a dimensionless quantity defined by Eq. (1.12) which may be rewritten as:

$$K = \frac{3WH}{LS} \tag{1.13}$$

Since L/H is the real area of contact and since the cross-sectional area of the plastically deformed subsurface zone under the asperity contact A_p is of the order of A_r, Shaw showed that Eq (1.13) may be rewritten as[11]:

$$K = \frac{3WH}{LS} = \frac{W}{A_p S} = \frac{\text{Worn Volume}}{\text{Volume of the plastically deformed zone}} \tag{1.14}$$

Therefore, the wear coefficient, K, for <u>sliding wear</u> may be interpreted as a dimensionless quantity that represents the ratio of the worn volume to the volume of the plastically deformed zone. Since K is the order of 10^{-4} to 10^{-3}, the volume of the material removed by wear is a very small fraction of the material undergoing plastic deformation below the asperity contact. Therefore, it is clear that sliding wear cannot be properly understood without comprehending the plastic deformation process at the subsurface. This process is the primary mode of energy dissipation during sliding wear.

The abrasive wear was modeled in the past as a cutting process. This assumes that an abrasve particle leaves a wear track of the same cross-sectional shape. For example, if the abrasive grain can be idealized as a cone, as shown in Figure 1.11, the wear volume, removed after the abrasive traverses a distance S, is the area shown by the shaded area. However, this type of over simplified model misrepresents the true picture. In order to illustrate this point, we may again examine the physical significance of wear coefficient but this time for abrasive wear, and compare its predictions with experimental results. It can be shown that when wear particles in the form of chips are generated without any plastic deformation, the specific energy μ (i.e., the work done to remove a unit volume of material by a cutting mechanism) is equal to the hardness of the material for the idealized conical model shown in Figure 1.11. Then, Eq. (1.12) may be rewritten as:

$$K = \frac{3WH}{\mu LS} \simeq \frac{Wu}{\mu LS} \tag{1.15}$$

assuming $3\mu \simeq 1$. It is then seen that the dimensionless quantity K is simply the ratio of the work done to generate wear particles, in the form of cut chips, to the total external work done. Therefore, when the entire work done is

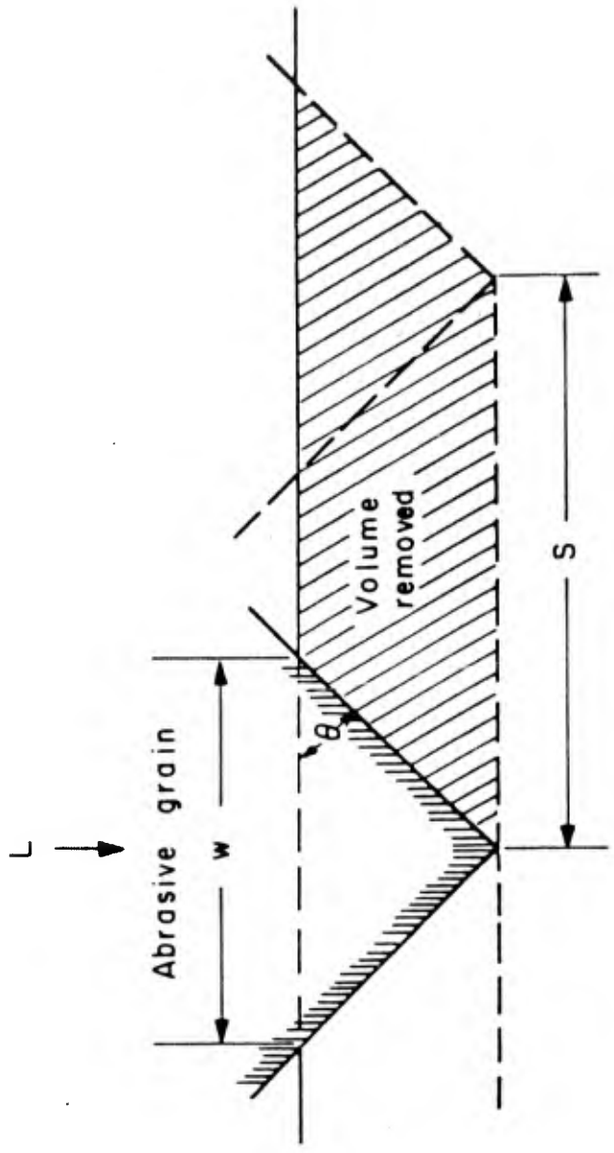


FIGURE 1.11 ABRASIVE WEAR MODEL

consumed to cut the surface, as the classical theories assumed, the wear coefficient should be nearly equal to unity. However, the experimentally determined maximum coefficients are one or two orders of magnitude less than unity.

The fact that a wear theory based on the cutting mechanism predicts too large a wear coefficient can also be seen from the results of the cutting test. Rabinowicz's theory, which is one of the more commonly cited works, states that the volume of material cut is equal to the volume displaced by an abraxise grain, i.e.[1],

$$K = \frac{3 \tan \theta}{\pi} \tag{1.16}$$

The same model for friction by plowing yields

$$\mu = \frac{\tan \theta}{\pi} \tag{1.17}$$

Comparing Eqs. (1.16) and (1.17) the friction coefficient is related to the wear coefficient as

$$K = 3\mu \tag{1.18}$$

When a diamond stylus with a cutting angle of $\theta = 35^{\circ}$ was used to cut AISI 1095 steel, the experimentally determined values of K and μ were 0.23 and 0.6, respectively[12]. The theoretically predicted value for K was 0.67.

The discrepancy between the simple cutting model and the actual abrasive wear process is primarily caused by the deformation of the subsurface layer. A large plastic deformation occurs at and below the surface during abrasion. Plowing also forms ridges which have to be removed subsequently. For this reason, the wear rate depends on the ductility of the material; less ductile materials have generally higher wear coefficients than more ductile materials. Two examples are commercially pure nickel and copper, see Figure 1.12.

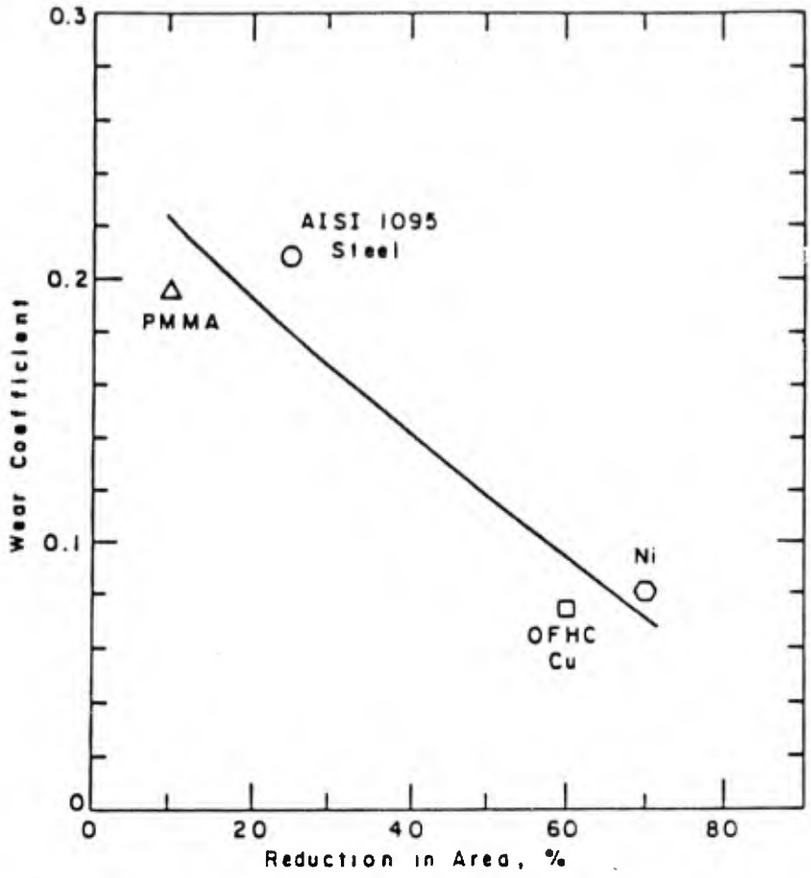


FIGURE 1.12 ABRASIVE WEAR RATE VERSUS DUCTIBILITY

The wear coefficient can be misleading because of its over-dependence on hardness. As shown in Table 1.2, the wear rate does not depend on the prior cold work, whereas the wear coefficient is very different between the annealed and cold-worked nickel. The large difference in numerical values of the wear coefficient, although the wear rate was the same, is due to the different values of the hardness used in computation. The cold-worked metal has the same wear rate as the annealed metal because the amount of cold work done during the wear process is much larger than the cold work done during metal processing and because the prior cold work might have induced "damages" which accelerates the wear process vis-a-vis annealed metals.

CONDITION	HARDNESS (kg/mm^2)	WEAR RATE (m^3/m)	WEAR COEFFICIENT	HARDNESS AFTER TEST
Annealed	88.5	8.07 x 10	0.053	240
Fully Cold Worked	242.0	8.66 x 10	0.157	242

* Material = Ni

Applied Load = 4 kg

Abrasive Size = 60 grit

TABLE 1.2 EFFECT OF COLD WORK ON ABRASION

A very useful piece of information to keep in mind in considering wear problems is that by knowing the wear coefficient, we can begin to speculate as to the cause of wear. With the caution that there are dangers in using numerical values indiscriminately, the following "ball park" figures are given for typical wear processes:

During sliding wear (be delamination or sometimes referred to as adhesive wear)	10^{-4} to 10^{-3}
Abrasive wear	10^{-2} to 1
Fretting wear (at large displacement amplitudes, > 100 μm)	10^{-4} to 10^{-2}

A rather extensive collection of wear data is available, but when special material combinations are to be used or when unique operating conditions exist, the most prudent thing to do is to perform actual tests in addition to using the wear coefficients available in handbooks[13].

1.4 Comments on Surface Topography

A great deal of work has been done to characterize the surface finish mathematically (see, for example, Whitehouse, Ref. 14). However, the relationships between the surface topography and the functional requirements for friction and wear are not yet fully understood. Based on the available information, it appears that the original surface finish is not very important in low speed dry sliding applications; whereas it is of importance when lubricants are used. In the former case, the surface geometry is drastically altered by

sliding actions; whereas in the latter case, not only does the initial surface geometry change rapidly, but also the pressure and temperature in the lubricant and the metal surface depend sensitively on the surface topography. In this series of lectures, the effect of surface topography will be neglected.

1.5 Introduction to the Lecture Notes

There are two kinds of surface interactions: chemical and physical. Therefore, we will first examine the chemical and physical state of solid surface, followed by discussions of the chemical and the physical behavior of solid surfaces in Section 2. Then, in Section 3 the generation and transmission of forces at the interface will be examined, in order to understand the genesis of friction. Section 4 will deal with the response of materials when external forces are applied to the metallic and polymeric surface by the asperities and wear particles. Based on the basic analytical results of Section 4, delamination wear will be covered in Section 5. Finally a few concluding remarks will be made in Section 6.

1.6 References - Section 1

1. Rabinowicz, E., "Friction and Wear of Materials", Wiley, New York, 1965.

2. Buckley, D.H., "Definition and Effect of Chemical Properties of Surfaces in Friction, Wear and Lubrication", Fundamentals of Tribology (Ed. N.P. Suh and N. Saka), MIT Press, 1980.

3. Green, A.P., "Friction Between Unlubricated Metals: A Theoretical Analysis of the Junction Model", Proceedings of the Royal Society of London, A228 (1955) 191-204.

4. Green, A.P., "The Plastic Yeilding of Metal Junctions Due to Combined Shear and Pressure", Journal of the Mechanics and Physics of Solids, 2(1955) 197-211.

5. Shaw, M.C. and E.F. Macks, "Analysis and Lubrication of Bearings", McGraw-Hill, New York, 1949.

6. Kragelskii, I.V., "Friction Interaction of Solids", Soviet Journal of Friction and Wear, 1(1980) 7-20 (Translated by Allerton Press, Inc.)

7. Archard, J.F., "Contact and Rubbing of Flat Surfaces", Journal of Applied Physics, 24(1953) 981-988.

8. Suh, N.P., "New Theories of Wear and Their Implications for Tool Materials", Wear, 62(1980) 1-20.

9. Kramer, B.M. and N.P. Suh, "Tool Wear by Solution: Quantitative Understanding", Journal of Engineering for Industry, Trans. A.S.M.E., 102 (1980) 303-

10. Suh, N.P., "The Delamination Theory of Wear", Wear, 25 (1973) 111-124.

11. Shaw, M.C., "Dimensional Analysis for Wear Systems," Wear, 43 (1977) 263-266.

12. Sin, H.-C., N. Saka, and N.P. Suh, "Abrasive Wear Mechanisms and the Grit Size Effect", Wear, 55(1979) 163-190.

13. Rabinowicz, E., "Wear Coefficients - Metals", Wear Control Handbook, A.S.M.E., 1980.

14. Whitehouse, D.J., "The Effects of Surface Topography On Wear", Fundamentals of Tribology (Ed. N.P. Suh and N. Saka), MIT Press, 1980.

2. CHEMICAL AND PHYSICAL STATE OF THE SOLID SURFACE

2.1 Brief Introduction to Metals, Polymers and Ceramics

Metals are characterized by their metallic bonding and crystal structures. Most metals have body centered cubic (b.c.c.), face centered cubic (f.c.c.), and hexagonal close packed (h.c.p.) structures. Some have other structures such as tetragonal. The mechanical and chemical behavior and properties depend on crystallographic orientation. Plastic deformation occurs on closely packed planes along the closely packed directions. These crystallographic planes and directions of plastic deformation are called slip planes and slip directions, respectively. Similarly, the surface energy depends on the crystallographic plane; the surface energy of the close packed planes is typically the lowest.

All metals have defects in the form of vacancies and dislocations. Dislocations, which are line defects of atomic arrangement, lower the stress required to cause plastic deformation because of their mobility under stress. Workhardening is a result of many dislocations interacting with each other when the dislocation density increases with plastic deformation. A well annealed solid typically has a dislocation density of 10^6 cm^{-2}. As the number of dislocations increases, they form dislocation cells which can

be as small as a few hundred angstroms in diameter. Dislocation behavior is of paramount importance in understanding the plastic deformation of crystalline solids. Although the dislocation behavior near the surface is extremely important in understanding all aspects of tribology, little is known about their structure and density of dislocations very near the surface (i.e., less than a hundred angstroms).

Most metals used in engineering applications are polycrystalline, made up of many grains of all orientations. Therefore, polycrystalline metals have isotropic properties on a macroscopic scale. Grain boundaries are interfaces between two grains of different orientation. Most grain boundaries are regions of random atomic arrangement of finite thickness and therefore, have higher energy than the bulk. Solutes, therefore, segregate to grain boundaries. In many situations polycrystalline solids with small grains have better mechanical properties; the yield strength and the toughness increases with decrease in the grain size.

Many metals used in engineering applications are alloys (both substitutional and interstitial) rather than pure elements. The alloying elements are added for grain refinement and to modify the chemical properites. Many metals used in tribological applications have hard ceramic phases which are introduced to increase hardness.

Polymers are covalently bonded solids. They are long chain molecules which consist mostly of hydrocarbons and other nonmetallic elements. Thermoplastics are linear long chain polymers, most of which either melt or soften as the temperature is increased, except "teflon" (polytetrafluroethylene) which decomposes before melting. Certain thermoplastics, such as polyethylene and polypropylene, are partially crystalline (i.e., crystalline regions are separated by amorphous regions). They have melting points as well as second order transition temperatures. They can undergo large plastic deformations. Glassy thermoplastics, such as polystyrene and polymethylmethacrylate (PMMA), are amorphous and brittle. Thermosetting plastics are those with 3-D network of covalent bonds and therefore amorphous and rigid at all temperatures, undergoing decomposition rather than melting at high temperatures. All these polymers are more corrosion-resistant than most metals and chemically inert under normal operating conditions.

Typical polymers have a broad molecular weight distribution. As a consequence, when these polymers are

solidified the molecular weight distribution near the surface or an interface is substantially different from those of the bulk. Therefore, in certain crystalline thermoplastics the surface layer has different mechanical properties from those of the bulk. The existence of such a layer is attributed to the preferential nucleation of high molecular species during the crystallation process at certain nucleation sites[1]. According to this reasoning, low molecular weight species of the polymer are rejected to the molten plastic-air interface during the crystallation process giving a weak surface layer. At a molten polymer-metal interface, a region of high cohesive strength is produced in the plastic if the metal surface provides nucleation sites. This is caused by rejection of the low molecular weight species from the interface into the bulk. In the absence of nucleation sites, a weak plastic layer results at the plastic-metal interface.

The first tribological application of polymers was about 50 years ago when phenolformaldelyde resin reinforced with fibers and fabrics was used to make bearings. Since then, particularly during the 1960's, a large number of new polymers were introduced for engineering uses, Table 2.1. Polymeric materials are increasingly becoming important in tribology, since in many applications it is not possible to establish hydrodynamic lubrication at all times due to the low sliding speeds (or oscillatory motions) or due to frequent stop-and-start situations. In some cases, such as in food and textile applications, the presence of lubricant is not acceptable due to the risk of contamination. In general, low cost, light weight, quiet operation, ease of fabrication, and resistance to corrosion have been the major reasons for use of polymeric bearing materials. In order to compensate for such properties as low thermal stability and low mechanical strength and to improve friction and wear properties, various fillers and fibers have been added to polymer matrix.

1839	Vulcanization of natural rubber
1907	Phenol-formaldehyde resin
1926	Alkyd resins
1931	Neoprene synthetic rubber
1937	Styrene-butadiene, acrylonitrile-butadiene rubbers
1938	Nylon 66
1941	Polyethylene
1942	Unsaturated polyesters for laminates
1943-45	Silicones, flurocarbon resins, polyurethanes, styreme-butadiene rubber
1947	Epoxy resins
1956	Linear polyethylene, acetals (polyoxymethylene)
1957	Polypropylene, polycarbonate
1959	cis-polyisoprene and cis-polybutadiene rubber
1960	Ethylene-propylene rubber
1962	Phenoxy resins, polyimide resins
1965	Polyphenylene oxide, polysulfones
af. 1965	Kenlar fibers

TABLE 2.1 INTRODUCTION OF PLASTICS USED IN TRIBOLOGY

Ceramic is a material which is a combination of one or more metals with a nonmetallic element, ussally oxygen, carbon, or nitrogen. The atoms are held together primarily by ionic bonding, with some covalent bonding. Ceramics, especially oxides and nitrides, have very low free energy of formation and therefore are very stable chemically. They do not readily react with other materials and have high melting points. Therefore, ceramics are used in high temperature tribology applications. Ceramics tend to be brittle and thus find limited use in applications that require toughness.

2.2 General Characteristics of a Solid Surface and Tribology.

A solid surface is created when a solid fractures or when a liquid solidifies. It is an asymmetric boundary between two regions: one region where the interatomic forces are greater than the thermal energy of each atom so that the atoms are closely packed, and the other region where there is gas with no "near neighbor" interatomic interactions. Because of the asymmetry, the atoms of the outer surface layer experience different forces than those in the bulk. Consequently, the

electronic arrangement near the surface is considerably different from that of the bulk. In the case of many covalently bonded solids and a few metals, even the atomic arrangment reconstructs to new equilibrium configurations[2]. An example of such reconstruction is shown in Figure 2.1. One would not expect such a reconstruction of the surface of a perfect single crystal, when the surface is a close-packed plane, because there is no atomic arrangement with a lower free energy. However, even a small amount of adsorbed gas atoms seems to affect the structure of the surface by dislodging the surface atoms and forming a periodic array of substrate atoms and chemisorbed gas atoms, see Figure 2.1[3]. There must be a certain degree of reconstruction on most polycrystalline solid surfaces because of the random orientation of grains, except when the surface deforms due to the sliding action and exposes close packed slip planes.

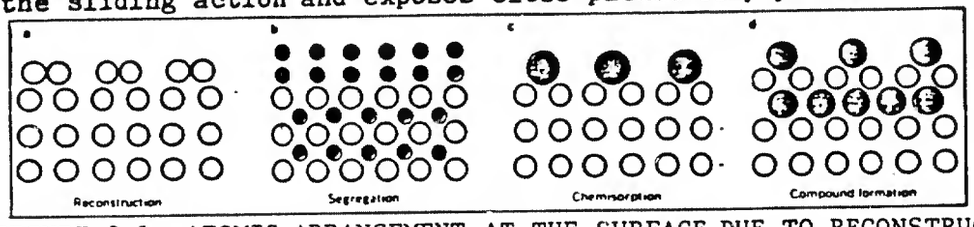


FIGURE 2.1 ATOMIC ARRANGEMENT AT THE SURFACE DUE TO RECONSTRUCTION OF THE SURFACE ATOMS. SEGREGATION OF SOLUTE ATOMS, CHEMISORPTION, AND DUE TO COMPOUND FORMATION BY MIXING OF ADSORBED ATOMS AND THE SUBSTRATE ATOMS (FROM ESTRUP, REF. 2)

In order to create a surface from the bulk, work must be done to break the atomic bonds. Consequently, the surface is at a state of higher free energy than the bulk; the surface energy is equal to the external work done in breaking the chemical bonds. The overall free energy of the surface and the surrounding is lowered when physisorption, chemisorption, and chemical reactions take place at the surface, see Figure 2.1. In physisorption, weak van der Waals-like forces act between an adsorbate and the substrate. Chemisorption bonds involve the mixing of the adsorbate and substrate wave functions, involving more reactive adsorbates such as H, O, CO and NO[4, 5]. Chemical reactions involve the formation of primary chemical bonds between the adsorbate and the substrate, resulting in a three-dimensional structure of a new compound. In tribology, chemisorption and chemical reactions at the surface are used to advantage in lowering frictional forces and minimizing the wear rates.

At the surface there is also an electrostatic potential difference associated with a charge double layer[6, 7]. Because of the asymmetry of repelling forces acting on the electrons of the outer-most shell of the surface atoms, one would expect to find excess electrons just outside the "original" surface and electron deficiency just inside the surface. Quantum

mechanical calculation shows that there is a finite probability of finding an electron outside the metal surface in a vacuum. This difference in electronic density about the surface creates electrostatic potential at the surface which is more negative outside than inside the surface. This is referred to as the electrical double layer. These double layers are present at all interfaces, including the solid-electrolyte interface. The electric field created by this double layer can only penetrate about 1 angstrom into the metal surface due to the high mobility of conduction electrons, but can penetrate in excess of 1 μm in the case of insulators.

It is not clear how the reconstruction of the atomic structure and the changes in the electronic state at the surface affect adhesion and tribological behavior. However, it may be reasonable to assume that the directionally bonded solids (i.e., covalently bonded and ionic solids) do not readily adhere to other solids during sliding. Even in the case of identical metals sliding against each other, the reconstructed surfaces cannot instantaneously reestablish the original chemical bond at asperity contacts although the free energy is lowered by going back to the bulk state. This is because there is a finite activation energy barrier that the atoms and electrons must overcome in order to return to the bulk state. Therefore, welding of moving asperity junctions may not be common occurrences for some materials when the temperature at the interface is low.

The change in chemical composition due to segregation can have a significant effect on the tribological characteristics of a surface. It has been shown experimentally that a freshly exposed surface of a solid solution establishes a new composition at the surface. Buckley showed that the surface of 1% Al-Cu solid solution has a 6.5 times higher concentration of aluminum atoms at the surface than in the bulk, see Table 2.2, increasing the adhesion of these solid solutions to gold five-fold over pure copper[8]. Surface enrichment by segregation was also observed in other systems: nickel in iron, silver in palladium, gold in copper, copper in nickel, silver in gold, tin in copper, aluminum in iron, and platinum in osmium. The segregation may be attributed to the need to lower the strain energy created by the difference in atomic size, see Table 2.3.

Alloy	Ratio of the Surface Conc. to Bulk Concentration
Cu - 1 a/o Al	6.5
Cu - 5 a/o Al	4.5
Cu - 10 a/o Al	3.1
Cu - 1 a/o Sn	15.0 ± 2
Fe - 10 a/o Al	8.0

TABLE 2.2 MAXIMUM COVERAGE OF MINOR CONSTITUENT ON ALLOY SURFACES (FROM BUCKLEY, REF. 8)

Aluminum	(f.c.c.)	1.431
Titanium	(h.c.p.)	1.458
Vanadium	(b.c.c.)	1.316
Chromium	(b.c.c.)	1.249
Manganese	(cubic comp.)	1.12
Iron	(b.c.c.)	1.241
Cobalt	(h.c.p.)	1.248
Nickel	(f.c.c.)	1.245
Copper	(f.c.c.)	1.278
Zinc	(h.c.p.)	1.332
Silver	(f.c.c.)	1.444
Cadmium	(h.c.p.)	1.489
Tin	(Bc tetragonal)	1.509

TABLE 2.3 ATOMIC RADIUS (IN ANGSTOMS)

Most real surfaces are ordinarily covered with adsorbates and oxides, which minimizes the metal-to-metal contact at asperity contacts. In this sense adsorbates and oxides act as lubricants. Even heating metals in a vacuum at high temperatures cannot get rid of all adsorbates. One has to resort to argon ion bombardment and the like to remove all the extraneous elements[8]. The important consequence of a coherent oxide layer which is stongly bonded to a substrate is that it alters the dilocation behavior near the surface, making the surface harder than the bulk upon plastic deformation. It should be noted that not all oxide layers formed on a metal surface are stable.

Even in the absence of adsorbates, the surface may be contaminated by migration of a few ppm of interstitial impurities to the surface. A few ppm of carbon in iron was shown to change the surface chemistry[8]. Similar segregation has been seen in such systems as oxygen in platinum, phosphorous in iron, sulfur and carbon in nickel, sulfur in molybdenum, and sodium in lithium. In view of these contamination problems, one has to be extremely careful in using the surface energy data in tribological applications.

An interesting correlation has been made between the cohesive energy density (CED) of crystalline solids and adsorption of albumin protein molecules[9]. The work was done to check the following hypothesis made on protein adsorption: (1) the negatively charged albumin molecules accumulate on a solid surface due to the surface charges created as a consequence of proton diffusion into the solid; and (2) the solid with high CED, which have higher interatomic potential, should make it more difficult for protons to diffuse into the solid. The results show that diamond, which has the highest CED, adsorbs the least amount of proteins and pure metals such as gold, which have a very low CED, adsorbs the most. The

adsorption of proteins is plotted as a function of CED of the substrate in Table 2.4.

MATERIAL	COHESIVE ENERGY DENSITY (CAL/CM^3)	SURFACE CONCENTRATION ($\mu G/CM^2$)
Diamond	49,700	0.1
Aluminum Oxide	34,200	0.1
Boron Nitride	31,000	0.9
Calcium Fluoride +	25,000	1.0, 3.4, 5.2
Magnesium Oxide +	15,800	3.5, 4.2, 10.0
Glass	15,000	1.5
Platinum	14,800	18.5
Stainless Steel	13,900	1.7
Gold	8,600	66.7
Aluminum	7,800	3.8
Silver	6,670	147.6

+ Adsorption dependent on plane of cleavage

TABLE 2.4 PROTEIN ADSORPTION OF VARIOUS MATERIALS AS A FUNCTION OF THEIR COHESIVE ENERGY DENSITIES

2.3 Chemical Interaction of Metal Surfaces with Lubricants

Metal surfaces are often lubricated with gases, liquids, and solids. Liquid lubricants with additives are the most commonly used means of lowering friction and wear. Lubricants are used to minimize the metal-to-metal contacts either by forming a new chemical compound on the surface or physically separating the surfaces, and thus lower the friction force. The mechanism of lubrication is in part dictated by the nature of interactions between the lubricant and the metal surface. The interactions can be in many different forms: chemical reaction which form new compounds on the surface; a physical interaction such as physisorption which attracts long chain molecules in the liquid to the surface due to van der Waal type electrostatic forces; or in the form of chemisorption

involving strong chemical bonds between the liquid components and the metal surface.

Clean metal surfaces react with gases and long chemical molecules dissolved in liquids (for a review of the subject, see Godfrey, Reference 10). This phenomenon has been used to advantage in developing lubricants that will form a stable compound on the surface by inducing chemical reactions which minimize adhesion at the asperity contacts. Oxygen is an important element in boundary lubricants. Klaus, et al., have shown, using microanalytical techniques, that dissolved oxygen reacts with super refined mineral oil and the metal surface to form organometallics which reduce friction and wear[11].

Sulfur compounds are commonly used in boundary lubricants which react with the metal surface to form metal sulfides upon decomposition of these compounds under sliding conditions. These compounds are typically iron-sulfide and form on scuffed surface plateaus[12]. When oxygen is also present, a mixture of iron sufide and iron oxide form[13, 14]. In a gaseous environment with oxygen, iron sulfide changes into iron oxide, indicating that iron oxide is more stable than iron sulfides[15].

Phosphorous compounds such as tricresyl phosphate in mineral oil have been used in many lubricants. It is believed that they form iron phosphates. Similarly, chlorides also form a stable surface layer.

The most commonly used additive is zinc organo dithiophosphates. When this additive was added to a paraffin white mineral oil and a light hydraulic oil, brown and blue films were found to form on highly deformed steel surfaces[16]. Auger analysis and secondary ion mass spectroscopy show that the blue film is 50 μm thick and is primarily a metastable protoxide of iron, $Fe_{1-x}O$[17]. The brown film contained P, S, and Zn, and was composed of minerals such as sulfides, sulphates, phosphates, and thiophosphates. Georges, et. al., claim that the brown film is the reaction between zinc dithiophosphates and iron and contains some polymers of type $[Zn(PSO_2)]_n$.

This adsorption of long chained molecules on the surface also affect the mechanical properties of metals and other solids. Rebinder effect is the most well known phenomenon which describes the adsorption - induced reduction of the surface hardness[18]. The implication of such chemo-mechanical effects as the Rebinder effect on the tribological behavior of materials is not clearly known. A review article on surface effects in crystal plasticity by Latanision gives a good

account of various surface phenomena under non-tribological loading conditions[6].

2.4 Mechanical Properties of Solid Surfaces

It was stated earlier that the chemical and physical properties of metallic and polymeric surfaces are different from the bulk. One of the physical properties tribologists are interested in is the flow strength of the materials near the surface. The question that has been investigated the most is whether the surface layer is softer or harder than the bulk when the specimen is subjected to uniform uniaxial deformation. This has been rather a controversial subject, since the experimental results support both views. According to Kramer the surface is harder, whereas Fourie holds the opposite view [19, 20]. Similar controversy also exists for highly worn surface layers[21, 22]. In order to understand the controversy, it is necessary to deal with dislocation dynamics near the surface, including generation and multiplication of dislocations and variation in dislocation density near the surface.

Kramer found that the original yield stress and the workhardening behavior of an aluminum monocrystal could be recovered, as shown in Figure 2.2, by chemically removing a 1 mm layer from the surface[19]. This experimental result was taken to mean that the Stage I workhardening is confined to the near-surface layer. Fourie, on the other hand, found the flow stress distribution of plastically deformed copper monocrystal to decrease near the surface, indicating that there was less workhardening near the surface[20]. Fourie's technique was first to deform a large specimen, slice them into thin sections ranging in thickness from 0.065 to 0.6 mm, and then reload them. The results are shown in Figure 2.3.

Many explanations have been advanced to deal with the soft-hard controversy. All these explanations are based on dislocation theory. One of the explanations for Kramer's results is that dislocations are generated near the surface layer by a Frank-Read type multiplication mechanism which can readily move into the interior of the solid, but cannot egress out of the surface because the surface (or a contaminated surface layer) acts as a barrier to dislocation motion and therefore the dislocations accumulate at the surface faster than the interior[6, 23]. Fourie's experimental results have been attributed to the fact that near the surface dislocations of opposite signs are not equally available, which are needed to form dislocation dipoles. Since dislocation dipoles are responsible for some of the workhardening in Stage II of

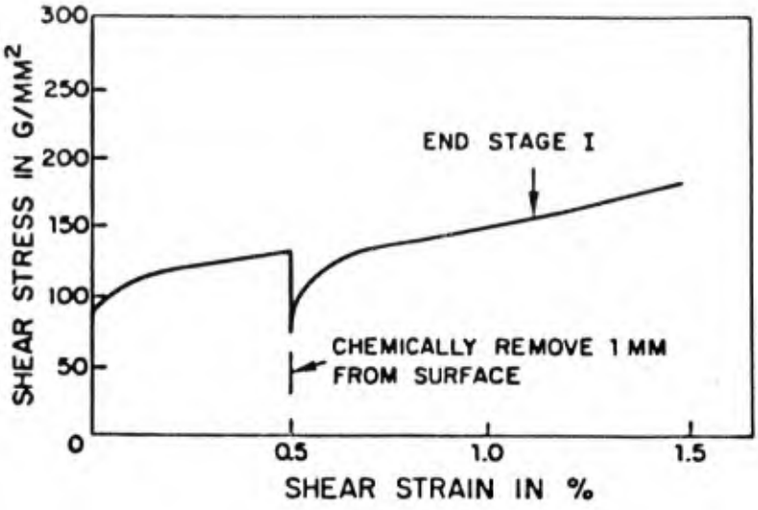


FIGURE 2.2 KRAMER'S EXPERIMENTAL RESULTS SHOWING THAT AN ALUMINUM MONOCRYSTAL RECOVERS THE ORIGINAL YIELD STRESS UPON RELOADING AFTER REMOVING A 1mm SURFACE LAYER ELECTRO-CHEMICALLY (FROM KRAMER AND DEWER, REF. 19B)

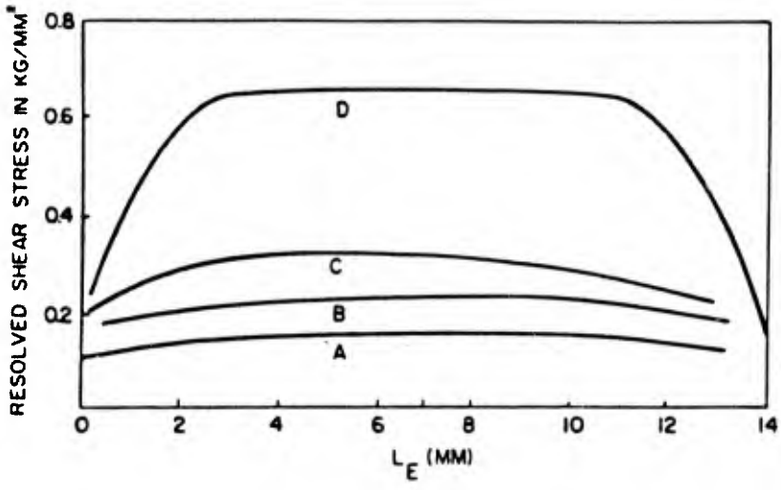


FIGURE 2.3 FOURIE'S EXPERIMENTAL RESULTS (20), WHICH SHOW THE FLOW STRESS DISTRIBUTION IN COPPER MONOCRYSTALS PLOTTED AS A FUNCTION OF THE LENGTH OF THE GLIDE PATH OF EDGE DISLOCATIONS, L_E, WHICH HAS THE VALUE 0 AND 14mm AT THE ORIGINAL SURFACES. CURVE A IS FOR AN AS-GROWN CRYSTAL: CURVES B,C, AND D ARE FOR PRESTRAINS OF 0.02, 0.029, AND 0.053, RESPECTIVELY (AFTER FOURIE, REF. 20)

f.c.c. metals, the surface layer does not workharden as much as the interior.

Dislocations very near and parallel to the surface experience image forces due to their proximity to the surface. When there is no continuous, coherent oxide layer adhering to the metal surface, the image force attracts the dislocations to the surface. When the image force is greater than the resisting drag force, commonly referred to as the dislocation friction stress, these dislocations are attracted to the surface and disappear, if the surface does not act as a barrier. Therefore, it is likely that the dislocations generated very near the surface during sliding may not be able to accumulate very near the surface, whereas the dislocations below this surface zone may be able to accumulate causing workhardening. When a hard, continuous oxide layer adheres to the metal surface, the boundary condition at the oxide-metal interface demands that a dislocation existing in the metal experiences a force which repels it from the surface. This would also be the case if a hard slider were moving over a soft metal surface. The magnitude of the image shear stress acting on a dislocation parallel to the surface is given by:

$$\tau_i = \frac{Gb}{4\pi\ (1-\nu)\ h} \qquad (2.1)$$

where G is the shear modulus, b the magnitude of the Burgers vector, ν Poissons' ratio, and h the distance from the surface. This force is opposed by the dislocation friction stress σ_f and the change in the surface energy as the dislocations emerge from the surface. Assuming that the dislocation friction stress dominates, Eq. (2.1) may be written for the maximum depth in which the dislocations will not be stable as:

$$h = \frac{Gb}{4\pi\ (1-\nu))\ \sigma_f} \qquad (2.2)$$

For silicon iron h is about 0.1 μm, whereas for pure copper it is about 10 to 20 μm.

Even if the lower dislocation density zone can exist very near the surface, it is expected that in polycrystalline metals only the dislocations in the outermost grains would be able to egress to the surface since the dislocations in the subsurface grains will encounter grain boundaries which are barriers to dislocation motion. Since the outermost grain of the worn surface undergoes a very large deformation, the resulting thickness of the grain can be extremely small. When it is less than the effective depth of image force given by

Eq. (2.2), the entire outer grain may be much softer than the subsurface grains. In no case can the soft layer be greater than the thickness of the outermost grain.

In order to estimate the thickness of the outermost grain, consider the deformation of a spherical grain of diameter D into an ellipsoid under the influence of shear stress. The thickness of a grain perpendicular to the surface c can be related to the equivalent strain $\bar{\varepsilon}$ of the surface layer as[24];

$$c = \frac{D}{\sqrt{3}\ \bar{\varepsilon}} \tag{2.3}$$

The maximum equivalent strain for f.c.c. metals is very large, being of the order of 100, while it is less for AISI 1020 steel, being of the order of 20[25]. The diameter of a typical f.c.c. metal may be about 50 μm. Then, the thickness of the deformed grain c is about 0.3 μm. This is less than the effective depth of the image force for pure copper. This means that the dislocation accumulation in the outermost layer of a clean surface may be less than those in the subsurface grains and that the outermost layer of 0.3 μm may remain soft. In this case, the metal is the hardest at subsurface, i.e., in the vicinity of 0.3 μm below the surface. According to Eq. (2.3) and experimental results, pure f.c.c. metals with larger grains should have a thicker softer surface layer than two phase metals, such as AISI 1020 steel, since the maximum equivalent strain is less for the latter[29].

The experimental results for hardness variation near the surface are contradictory. Savitskii, Kirk and Swanson presented experimental results which show that the hardness is the maximum at subsurface and decreases toward the surface. These results are shown in Figure 2.4[21, 22]. In general however, there are more papers presented in the literature which claim that the hardness is the greatest at the surface[26]. None of the work published so far considered the grain size effects and the maximum strain of the surface layer. This controversy cannot be resolved until precise experiments are done usting a reliable hardness be resovled until precise experiments are done using a reliable hardness measuring technique.

The significance of the hardness question is that if the outermost surface layer is the softest, the adhesion of the flat asperity contact at the surface cannot account for the observed friction coefficients in metals, since shearing will always occur through this soft layer at low stress. Then

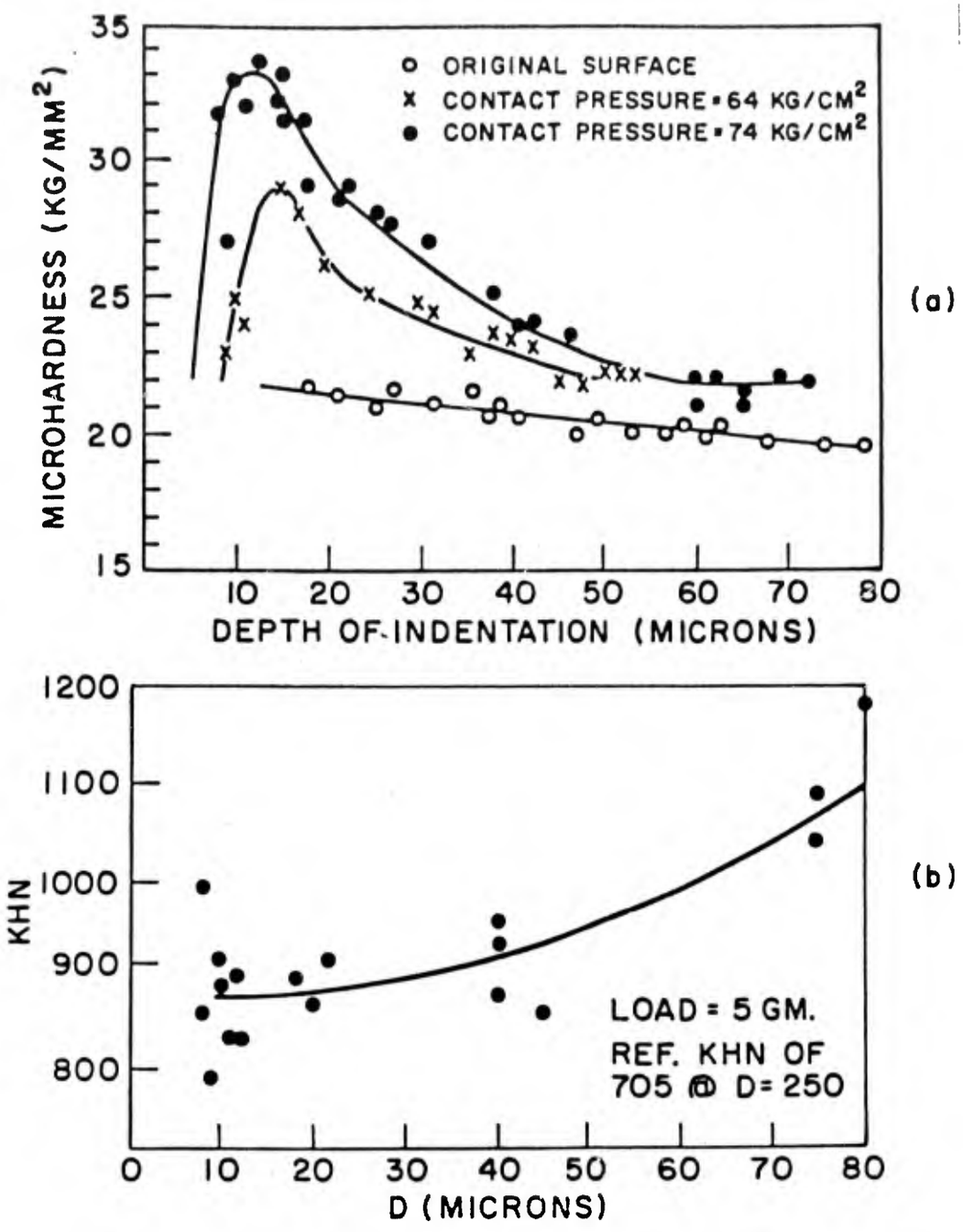


FIGURE 2.4 MEASUREMENTS OF THE SUBSURFACE HARDNESS VARIATION:
(A) INDENTATION HARDNESS OF THE WORN SURFACE OF ALUMINUM SPECIMENS UNDER VARYING INDENTATION LOAD (FROM REF. 21);
(B) INDENTATION HARDNESS OF COPPER SPECIMEN WEAR TESTED UNDER NORMAL LAOD OF 0.682kg (FROM REF. 22)

there must be other mechanisms that determine the frictional force and eventually the wear behavior.

2.5 Thermodynamic Analysis of an Interface

In earlier sections it was stated that the atomic structure of the surface of most metals (and probably other solids as well) is highly heterogeneous. The causes for the heterogeneity are the following:

a) Reconstruction of covalent solids and some metals,
b) Adsorbates due to physisorption and chemisorption,
c)) Oxide layers due to chemical reaction,
d) Segregation of solutes at the surface,
e) Reconstruction of the surface when adsorbates are present on the surface of metals, mixing the substrate atoms and adsorbate atoms.

The fact that these layers on the surface do not simply sit on the substrate but mix with the substrate atoms has been determined experimentally and through quantum mechanical arguments[27]. It can also be supported through thermodynamic arguments. In this section a thermodynamic analysis will be presented, following the treatment originally given by Cahn[28].

For the purpose of analysis, consider a thick oxide layer formed on a metal substrate. Then, the interface between the oxide layer and the substrate is the phase boundary between two different materials. The question we would like to answer is, "how thick is the interface?" If we assume that the interface has a finite thickness, the system consisting of the interface is locally a binary solution. If the mole fraction of oxygen, c, is a nonuniform property in the binary solution, the free energy per molecule, f, in the region of nonuniform composition will depend on both, the local composition and on the composition of the immediate environment. Assuming that the composition gradient is small compared with the reciprocal of the intermolecular distance, the concentration, c, and its derivatives may be taken as independent variables that can describe the local composition and the composition of the immediate environment. When f is a continuous function of these variables, it can be expanded in a Taylor series about f_o (the free-energy per molecule of a solution of uniform composition c) as:

$$f(c, \nabla c, \nabla^2 c, \ldots) = f_o(c) + \sum_i L_i \left(\frac{\partial c}{\partial x_i}\right)_o + \sum_{ij} K_{ij} \quad (1)$$

$$\left(\frac{\partial^2 o}{\partial x_i \partial x_j}\right) + \frac{1}{2} \sum_{ij} K_{ij}^{(2)} \left(\frac{\partial c}{\partial x_i}\right)\left(\frac{\partial c}{\partial x_j}\right) + \ldots , \qquad (2.4)$$

where

$$L_i = \left[\frac{\partial f}{\partial\left(\frac{\partial c}{\partial x_i}\right)}\right]_o$$

$$K_{ij}(1) = \left[\frac{\partial f}{\partial\left(\frac{\partial^2 c}{\partial x_i \partial x_j}\right)}\right]_o$$

$$K_{ij}(2) = \left[\frac{\partial^2 f}{\partial\left(\frac{\partial c}{\partial x_i}\right)\partial\left(\frac{\partial c}{\partial x_j}\right)}\right]_o$$

i, j represents the successive substitution of spatial coordinates x, y and z, and the subscript o represents the value of the parameter of the uniform composition. In general, $K_{ij}^{(1)}$ and $K_{ij}^{(2)}$ are tensors reflecting the crystal symmetry and L_i's are components of a polarization vector in a polar crystal. For a cubic crystal or an isotropic medium the free energy must be invariant due to the symmetry operators of reflection ($X_i \rightarrow X_j$) and of rotation about a fourfold axis ($X_i \rightarrow X_j$). Therefore,

$$L_i = 0$$

$$K_{ij}^{(1)} = K_1 = \frac{\partial f}{\partial \nabla^2 c}_o \qquad \text{if } i = j$$

$$K_{ij}^{(1)} = 0 \qquad \text{if } i = j \qquad (2.5)$$

$$K_{ij}^{(2)} = K_2 = \frac{\partial^2 f}{\partial(|\nabla c|)^2} \qquad \text{if } i = j$$

$$K_{ij}^{(2)} = 0 \qquad \text{if } i = j$$

Hence for a cubic lattice

$$f(c, \nabla c, \nabla^2 c, ..) = f_o(c) + K_1 \nabla^2 c + K_2 (\nabla c)^2 + \ldots \qquad (2.6)$$

Integrating over a volume V of the solution, we obtain for the total free energy F of this volume:

$$F = N_v \int f \, dv = N_v \int v \{ f_o(c) + K_1 \nabla^2 c + K_2 (\nabla c)^2 + \ldots \}_{dv} \quad (2.7)$$

where N_v is the number of molecules per unit volume. By applying the divergence theorem,

$$\int_v (K_1 \nabla^2 c) \, dv = -\int_v \left(\frac{dK_1}{dc}\right) (\nabla c)^2 dv + \int_s (K_1 \nabla_c \cdot n) ds \quad (2.8)$$

If we choose the boundary so that

$$\int_s (K_1 \nabla c \cdot n) \, ds = 0 \quad (2.9)$$

then

$$F = N_v \int \left[f_o(c) - \left(\frac{dK_1}{dc}\right) (\nabla c)^2 + K_2 (\nabla c)^2 + \right] dv \quad (2.10)$$

$$= N_v \int \left[f_o(c) + K (\nabla c)^2 + \ldots \right] dv$$

where

$$K = \frac{dK_1}{dc} + K_2 = -\left[\frac{\partial}{\partial c} \left(\frac{\partial f}{\partial \nabla 2_c} \right) \right]_o + \frac{\partial^2 f}{\partial (|\nabla c|)^2}$$

Eq. (2.10) states that the free energy is a sum of two contributions: one being the free energy that this volume would have in a homogenous solution and the other "gradient energy" which is a function of the local composition.

At a flat oxide/metal interface (two coexisting isotropic phases c_α and c_β), energy of nonequilibrium material of composition intermediate between c_α and c_β may be represented as a continuous function $f_o(c)$ of the form shown in Figure 2.5.

For one-dimensional composition change across the interface, and neglecting terms in derivatives higher than the second, we obtain for the total free energy F of the system as

$$F = AN_v - \int_{-\infty}^{\infty} \left[f_o(c) + K \left(\frac{dc}{dx}\right)^2 \right] dx \quad (2.11)$$

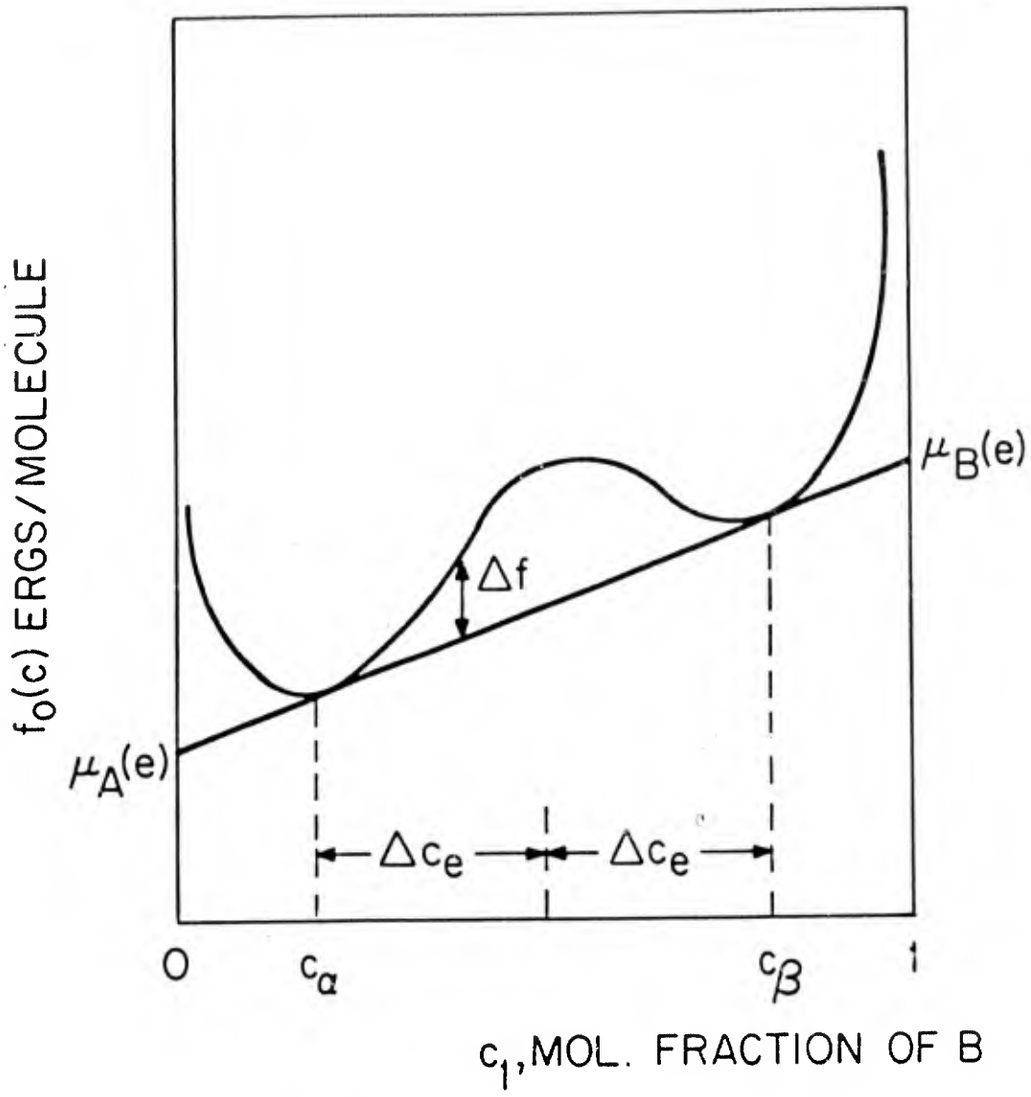


FIGURE 2.5 - FREE ENERGY OF A BINARY SYSTEM AS A FUNCTION OF COMPOSISTION (FROM CAHN, REF. 28).

The specific interfacial free energy, σ, is by definition the difference per unit area of interface between the actual free energy of the system and that which it would have if the properties of the phases are continuous throughout. Hence

$$\sigma = N_v - \int_{-\infty}^{\infty}\left[f_o(c) + K\left(\frac{dc}{dx}\right)^2 - c\mu_B(e) - (1-c)\,\mu_A(e)\right]dx \qquad (2.12)$$

$$= N_v - \int_{-\infty}^{\infty}\left[\Delta f(c) + K\left(\frac{dc}{dx}\right)^2\right]dx$$

where

$$\Delta f(c) = f_o - \left[c\mu_B(e) + (1-c)\,\mu_A(e)\right]$$

$$= c\left[\mu_B(c) - \mu_B(e)\right] + (1-c)\left[\mu_A(c) - \mu_A(e)\right]$$

In order to lower σ, the interfacial energy, highly diffused layer (i.e., small $K(dc/dx)^2$) is desirable, but it increases the $\Delta f(c)$ term by increasing the volume of the nonequilibrium composition, nonhomogeneous layer. Therefore, an optimum exists where $\{\Delta f(c) + K\left(\frac{dc}{dx}\right)^2\}$ term assumes a minimum value, i.e., the chemical potentials be constant through the system.

The foregoing thermodynamic treatment of the interface indicates that, in order to minimize the free energy, the atoms of adsorbates and oxide layers form a mixed layer of a finite thickness rather than simply sitting on the substrate. Therefore, most surfaces with contaminants are likely to be heterogeneous and thus may impede the egress of dislocations from the surface. Dislocation may still escape from the surface however, if they can move to the surface before adsorbates and oxides can form on the clean surface following the deformation of the surface layer.

2.6 Concluding Remarks

In this section the chemical and physical states of the outermost layer of the surface were examined. It was shown that the geometric arrangement, the chemical state, and the physical properties are highly heterogeneous. The substrates are normally covered with adsorbates, oxides and other chemical reaction products, and segregated solutes. These contaminants are likely to prevent adhesion in many sliding situations.

It was also shown that the hardness of the surface layer is different from that of the substrate, notwithstanding the controversy regarding the hardness/softness of the surface layer. The outermost grain of a polycrystalline metal can have a significantly different flow stress than the substrate if dislocation egress is not impeded by surface contaminants and heterogeneous structures. Further research must be done to clarify the mechanical properties of the outermost layer.

Finally, the state of the contaminant/substrate interface is examined from a thermodynamic point of view. It is shown that the atoms of contaminants are mixed with the atoms of the substrate in order to lower the free energy of the system. Because they are mixed, it is difficult to remove these contaminants by such techniques as heating. Consequently, it is safe to assume that the surface of most engineering materials is always contaminated. The high reactivity of solid surfaces can be used to minimize friction and wear. Additives are added to lubricants to form stable compounds on the surface, which lowers the frictional force by preventing plowing in addition to adhesion. This is discussed further in Section 3.

2.7 References - Section 2

1. Schonhorn, H., "Adhesion to Low Energy Polymers", Adhesion, Gordon and Breach Science Publishers, NY, 1969.

2. Estrup, P.J., "The Geometry of Surface Layers", Physics Today, April 1975, 33-41.

3. Somorjai, "Principles of Surface Chemistry", Prentice Hall, Englewood Cliffe, 1972.

4. Schrieffer, J.R., and P. Soven, "Theory of Electronic Structure", Physics Today, April 1975, 24-30.

5. Duke, C.B., "Electronic Structure of Clean Metal Interfaces", General Electric Report 68-C-33, Sept. 1968.

6. Latanision, R.M., "Surface Effects in Crystal Plasticity: General Review," Surface Effects in Crystal Plasticity, Noordhoff-Leyden, 1977.

7. Herring, C., in Metal Interfaces, ASM, 1, 1952.

8. Buckley, D.H., "Definition and Effect of Chemical Properties of Surfaces in Friction, Wear and Lubrication", Fundamentals of Tribology (Ed. N.P. Suh and N. Saka), MIT Press, 1980.

9. Suh, N.P. and L.T. Nguyen, "Correlation of Protein Adsorption with Cohesive Energy Density," to be published, 1981.

10. Godfrey, D., "Review of Usefulness of New Surface Analysis Instruments in Understanding Boundary Lubrication", Fundamentals of Tribology (Ed. N.P. Suh and N. Saka), MIT Press, 1980

11. Klaus, E.E., E.J. Tewksbury, and A.C. Bose, Proc. Japan Soc. of Lubrication Engineers - American Soc. of Lubrication Engrs., Int. Lub. Cong., Tokyo, June 9-11, 1975 (Ed. T. Sakurai, Elsevier Pub. Co., NY, 1976, p. 39).

12. Coy, R.C. and T.F.J. Quinn, in Tribology Convention 1972, London, Sept. 27-28, 1972; Institution of Mechanical Engineers, London, 1973, p. 62.

13. Godfrey, D., "Chemical Changes in Steel Surfaces During Extreme Pressure Lubrication", Trans. American Soc. of Lubrication Engineers, 5 (1962) 57-66.

14. Tomaru, M., S. Hironaka, and T. Sakurai, "Effects of Oxygen on the Load-Carrying Action of Some Additives," Wear, 41 (1977) 117-140.

15. Buckley,D.H., "Oxygen and Sulfur Interactions With a Clean Iron Surface and the Effect of Rubbing Contact on These Interactions", Trans. Am. Soc. of Lub. Engineers. 17, (1974) 206-212.

16. Jahanmir, S., "Wear of AISI 4340 Steel Under Boundary Lubrication", Wear of Materials, ASME, 1981.

17. Georges, J.M., G. Meille, M. Jacquet, B. Lamy, and T. Mathia, "A Study of the Durability of Boundary Films", Wear, 42 (1977) 217-228.

18. Rebinder, P.A., Reports to the VI Congress of Physicists, Moscow, 29, 1928.

19. Kramer, I.R., "The Effect of Surface Removal on the Plastic Flow Characteristics of Metals", Trans. AIME, 227 (1963) 1003-1010; Kramer, I.R., and L.J. Demer, Trans. AIME, 221 (1961) 780.

20. Fourie, J.T., "The Flow Stress Gradient Between the Surface and the Center of Deformed Copper Single Crystals", Phil. Mag., 17 (1968) 735-756.

21. Savitskii, K.B., cited in I.V. Kragelskii, Friction and Wear, Butterworth, Washington, 1965.

22. Kirk, J.A. and T.D. Swanson, "Subsurface Effects During Sliding Wear", Wear, 35 (1975) 63-67.

23. Nabarro, F.R.N., "Surface Effects in Crystal Plasticity - Overview from the Crystal Plasticity Standpoint", Surface Effects in Crystal Plasticity (Eds. R.M. Latanision and J.T. Fourie), Noordhoff-Leyden, 1977.

24. Dantzenberg, J.H. and J.H. Zaat, "Quantitative Determination of Deformation by Sliding Wear", Wear, 23 (1973) 9-19.

25. Augustsson, G., "Strain Field Near the Surface Due to Surface Traction", S.M. Thesis, MIT, 1974.

26. Ruff, A.W., "Deformation Studies at Sliding Wear Tracks in Iron", Wear, 10 (1976) 59-74.

27. Duke, C.B., "Atomic Geometry and Electronic Structure of Solid Surfaces", Surface Effects in Crystal Plasticity (Eds. R.M. Latanision and J.T. Fourie), Noordhoff-Leyden, 1977.

28. Cahn, J.W. and J.E. Hillard, "Free Energy of a Non-Uniform System. I. Interfacial Free Energy", Journal of Chemical Physics, 28 (1958) 258-267.

3. GENERATION AND TRANSMISSION OF FORCE AT THE INTERFACE - THE GENESIS OF FRICTION

3.1 Introduction

One of the three fundamental aspects of tribology is the genesis of friction between sliding surfaces, as discussed in Section 1. Beginning in the early 1940's, serious research has been done to understand the friction mechanism. Although various theories have been advanced, the adhesion theory has dominated other theories[1-4]. However, the experimental results often deviate from the adhesion theory. The experimentally observed coefficients of friction are generally much larger than those predicted; the prediction is worse when the experiment is done in vacuum or in an inert environment which best simulate the conditions assumed in deriving the

friction coefficient based on the adhesion theory. Further, the compatability argument for sliding surfaces, based on mutual solubility, cannot explain the large variations in the frictional behavior of metals with little difference in chemical solubility. Furthermore, there are some basic questions on adhesion, such as those discussed in Section 2, which cast doubt on the thesis that friction is entirely due to adhesion. It is clear that adhesion does exist in some sliding situations, but the question is "how important is it in determining the frictional force?"

As discussed in Section 2, the essence of the adhesion theory is that asperities of sliding surfaces come in contact with opposing asperities and form welded junctions which must be sheared to satisfy the kinematic requirements. Therefore, the frictional force depends directly on the actual area of contact which is a function of the applied normal and tangential load. In order to explain the discrepancy between the experimentally measured and the theoretically predicted friction coefficients, the adhesion theory has relied on the argument that the real area of contact is larger when some of these junctions are under tensile loading, requiring a correspondingly larger force to shear the interface.

Because the adhesion theory of friction emphasizes the importance of adhesion between asperities, a great deal of attention in the past has been devoted to the role of surface energy and the mutual solubility of the contacting materials [1]. It has been argued that metals with greater solubility will more readily form strong junctions and thus have higher friction and wear coefficients. Many experimental results were presented in support of this argument. However, questions have been raised on the validity of the argument since surfaces are easily contaminated by chemisorption and physisorption and the chemical composition of the surface is different from that of the bulk [5]. A more realistic model appears to be that the frictional behavior is controlled by plowing, adhesion and asperity deformation [6].

The frictional force depends on the history of sliding. The frictional force undergoes significant changes during the early stages of sliding before reaching steady state frictional behavior. The time dependent nature of frictional behavior is a rich source of information in understanding frictional behavior. The difference between the static and kinetic coefficients of friction is wellknown but the time dependent nature of the kinetic coefficient of friction is less well known [7]. This phenomenon is very important because when the interface is lubricated, the time scale is so expanded that the important frictional phenomenon of interest

between well lubricated surfaces, may in fact be confined to the early stages of the frictional behavior observed under the dry sliding conditions.

The frictional behavior of materials is important in tribology not only because of the frictional force between sliding surfaces of interest, but also because it generally affects the wear behavior[6, 8]. The delamination wear is clearly affected by surface traction due to its effect on crack nucleation and propagation processes at the subsurface. Also, plowing by wear particles and asperity deformation affect both the wear process and frictional behavior.

This section describes a different theory to explain the frictional behavior of materials. It will be demonstrated that adhesion does not play a significant role at the onset of sliding in typical sliding situations and that the frictional force is generated as a consequence of asperity deformation, plowing by wear particles and adhesion. Furthermore, the coefficient of friction is not a given material property, because it depends on the mechanical properties of the opposing surface and the environment. Experimental results will first be summarized before presenting theoretical support for the postulated genesis of friction. Finally, a "Friction Space" concept is presented as a means of representing the frictional behavior of materials.

3.2 Typical Friction Tests and Experimental Observations of Frictional Behavior of Metals

In Section 1 the phenonemological aspects of friction were discussed. In studying the fundamental mechanisms of friction however, it is necessary to investigate beyond the general phenomenological aspects of friction in great detail. The results of such a study will be described in this subsection before presenting a friction theory in a later subsection.

A series of experiments were conducted at MIT in order to study the friction and wear behavior of various combinations of the following materials: Armco iron, AISI 1020, 1045, and 1095 steel. These iron based metals with differing carbon contents, have large differences in hard-phase concentrations and hardness. Listed materials were chosen to minimize chemical differences, although the surface of Armco iron is expected to be substantially different chemically, from the steel specimens because of the absence of interstitials carbon atoms in Armco iron.

Armco iron was recrystallized at 973° K for one hour. AISI 1020, 1045 and 1095 steels were austenitized at 1173° K for 15 minutes, oil quenched, and then tempered at 673° K for one hour to obtain a spherodized microstructure. The hardness and the volume fraction of cementite are listed in Table 3.1.

Material	Heat Treatment	Vickers Hardness (MPa)	Volume Fraction of Cementite
Armco Iron	973 K, 1 hr; air-cooled	980 ± 50	0.0004
AISI 1020 steel	Austenitized at 1173 K, 15 min; oil-quenched; 673 K, 1 hr; air-cooled	1710 ± 100	0.020
AISI 1045 steel	Spheroidized: 1173 K, 15 min; oil-quenched; 673K, 1 hr; air-cooled	4120 ± 130	0.067
AISI 1095 steel	Spheroidized: 1173 K, 15 min; oil-quenched;673 K, 1 hr; air-cooled	6080 ± 350	0.142

TABLE 3.1 EXPERIMENTAL MATERIALS

Some tests were also conducted with OFHC copper, which was polished with 4/0 abrasive paper. These specimens were tested in air, sliding against AISI 1020 steel. The initial coefficient of friction, μ_1, was about the same as those obtained with steel iron specimens.

Samples of 6.35 mm in diameter were tested for friction and wear using crossed-cylinder geometry. The specimen (rotating cylinder) was rotated by the spindle of the lathe, and the slider (stationary cylinder) was held stationary in a holder attached to a lathe tool dynamometer which was mounted on the carriage of the lathe. Both normal and tangential forces were measured by a dynamometer-recorder assembly.

Tests were conducted in a purified argon atmosphere except for AISI 1020 steel where some samples were tested in air under both lubricated and unlubricated conditions. Water and light machine oil were used as lubricants. The experimental results were obtained under the following conditions: normal load of 1 kg (9.8 N), sliding speed of 0.02 m/s, total sliding distance of 36 m, and at room temperature.

Some of the specimens were sectioned along the sliding direction to measure the slope of the asperity by taking

micrographs of the asperities since the asperities of the machined surfaces were orientation dependent.

Extremely well polished surfaces were slid against each other to investigate the nature of the surface damage after predetermined amounts of sliding. These surfaces were observed using scanning electron microscopy.

The friction and wear coefficients of the iron-carbon system are tabulated in Tables 3.2 and 3.3 respectively. There are several important results worth considering in detail. First, the coefficient of friction changes as a function of the distance slid, especially at the early stage of sliding. It usually has a low initial value and gradually increases until reaching a steady state value. After it reaches a maximum value, the friction coefficient sometimes drops down if the stationary slider is much harder than the moving specimen. The same pair of materials does not show the drop in the coefficient of friction when their roles are reversed. The initial coefficient of friction is always in the range of about 0.1 to 0.2 regardless of the materials tested and whether or not lubricants are used. Second, the steady state coefficient of friction and the wear rates are higher when identical metals are slid against each other than when a harder stationary slider is slid against a moving softer specimen. However, when a softer stationary slider is slid against a harder moving specimen, the steady state coefficients of friction are nearly the same as those of the identical materials sliding against each other. In this case, the wear rates of unidentical pairs of metals are much greater than those of identical metals.

These changes in the friction and wear behavior are related to the changes in the surface topography, as shown in Figure 3.1 and 3.2, which are the micrographs of the slider surface and the specimen surface, respectively. These figures show that when the stationary slider is harder than the specimens, the hard surface is polished to a mirror finish and the high spots of the softer surface also acquire the same mirror finish. When the material underneath the polished surface fails, new high spots are created, which become polished to a mirror finish again. This does not happen when the stationary slider is softer than the specimen or when identical metals are slid against each other. In these cases, many plowing grooves are observed and the surface always stays rough. From these experimental results the following observations are made:

1. The coefficients of friction vs. sliding distance (or time) may be summarized in Figure 3.3a, when identical metals

			Specimen (rotating cylinder)			
			Armco iron	1020 steel	1045 steel	1095 steel
Slider (stationary cylinder)	Armco iron	μ_i	0.13	0.20	0.24	0.20
		μ_s	0.71	0.75	0.69	0.76
		μ^*	----	----	----	----
	1020 steel	μ_i	0.18	0.20	0.13	0.12
		μ_s	0.55	0.68	0.57	0.65
		μ^*	0.80	----	----	----
	1045 steel	μ_i	0.16	0.17	0.17	0.12
		μ_s	0.52	0.53	0.71	0.69
		μ^*	0.77	0.71	----	----
	1095 steel	μ_i	0.17	0.17	0.14	0.17
		μ_s	0.51	0.54	0.58	0.67
		μ^*	0.76	0.73	----	----

μ_i = initial coefficient friction

μ_s = steady state coefficient of friction

μ^* = peak value of the friction coefficient

TABLE 3.2 FRICTION COEFFICIENTS

			Specimen (rotating cylinder)			
			Armco iron	1020 steel	1045 steel	1095 steel
Slider (stationary cylinder)	Armco iron	K_{sp}	46.0 ± 37.2	127.0 ± 45.0	331.0 ± 268.0	1210.0 ± 120.0
		K_{sl}	25.1 ± 17.5	6.94± 2.33	16.4 ± 14.0	17.1 ± 0.5
	1020 steel	K_{sp}	6.10± 2.75	143.0 ± 88.0	25.3 ± 5.0	42.3 ± 17.6
		K_{sl}	5.23± 2.60	85.4 ± 48.4	2.57± 2.27	1.80± 1.33
	1045 steel	K_{sp}	3.59± 1.69	15.6 ± 5.9	94.2 ± 17.8	565.0 ± 269.0
		K_{sl}	2.89± 2.37	13.9 ± 7.2	49.4 ± 7.5	60.0 ± 17.4
	1095 steel	K_{sp}	2.11± 0.89	8.83 ± 4.04	7.32 ± 5.10	15.1 ± 7.6
		K_{sl}	2.35± 0.95	4.97 ± 2.72	5.24 ± 3.78	14.7 ± 10.3

K_{sp} = wear coefficient of specimen; K_{sl} = wear coefficient of slider

All K values are to be multiplied by 10^{-4}.

TABLE 3.3 WEAR COEFFICIENTS

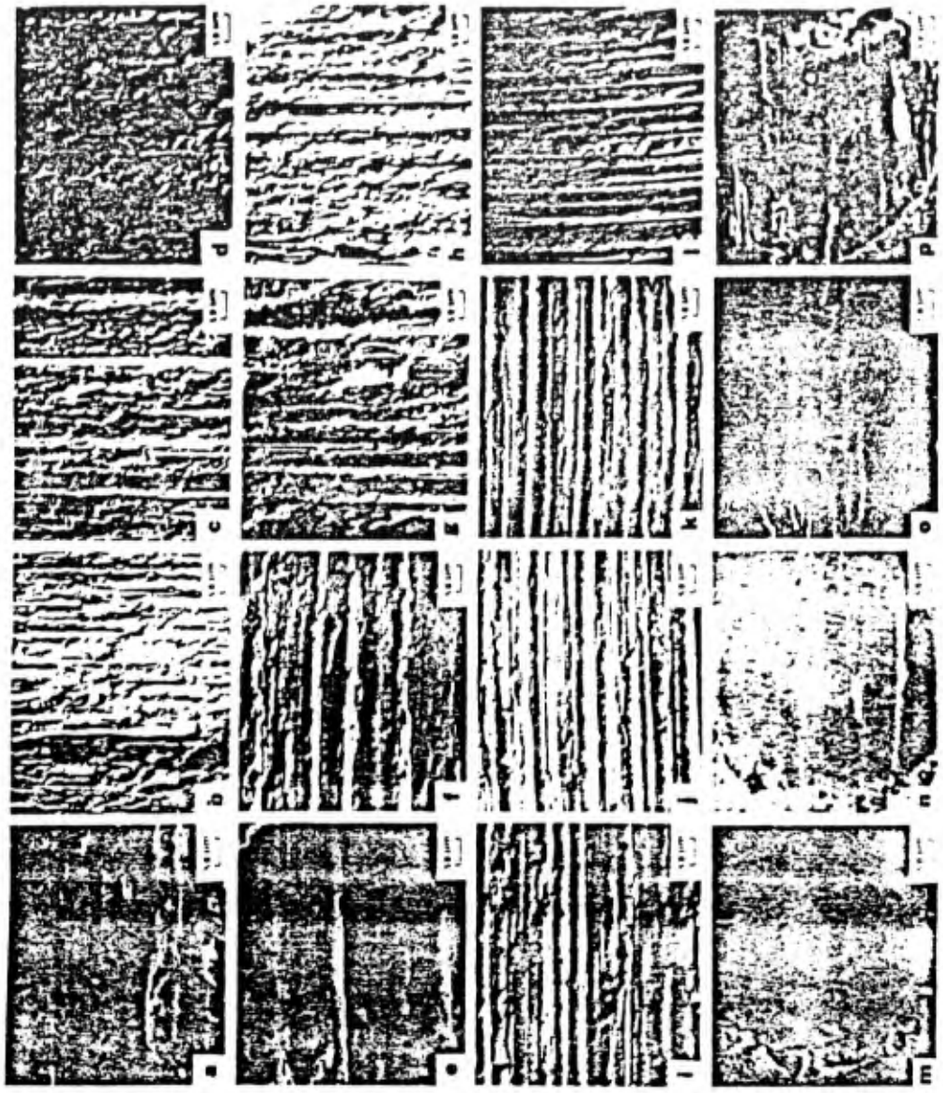


FIGURE 3.1 SCANNING ELECTRON MICROGRAPHS OF THE SURFACES OF WORN SLIDERS

(A), (B), (C), (D) IRON ON IRON, 1020, 1045, AND 1095 STEEL, RESPECTIVELY;

(E), (F), (G), (H) 1020 ON IRON, 1020, 1045, AND 1095 STEEL, RESPECTIVELY;

(I), (J), (K), (L) 1045 ON IRON, 1020, 1045, AND 1095 STEEL, RESPECTIVELY;

(M), (N), (O), (P) 1095 ON IRON, 1020, 1045, AND 1095 STEEL, RESPECTIVELY

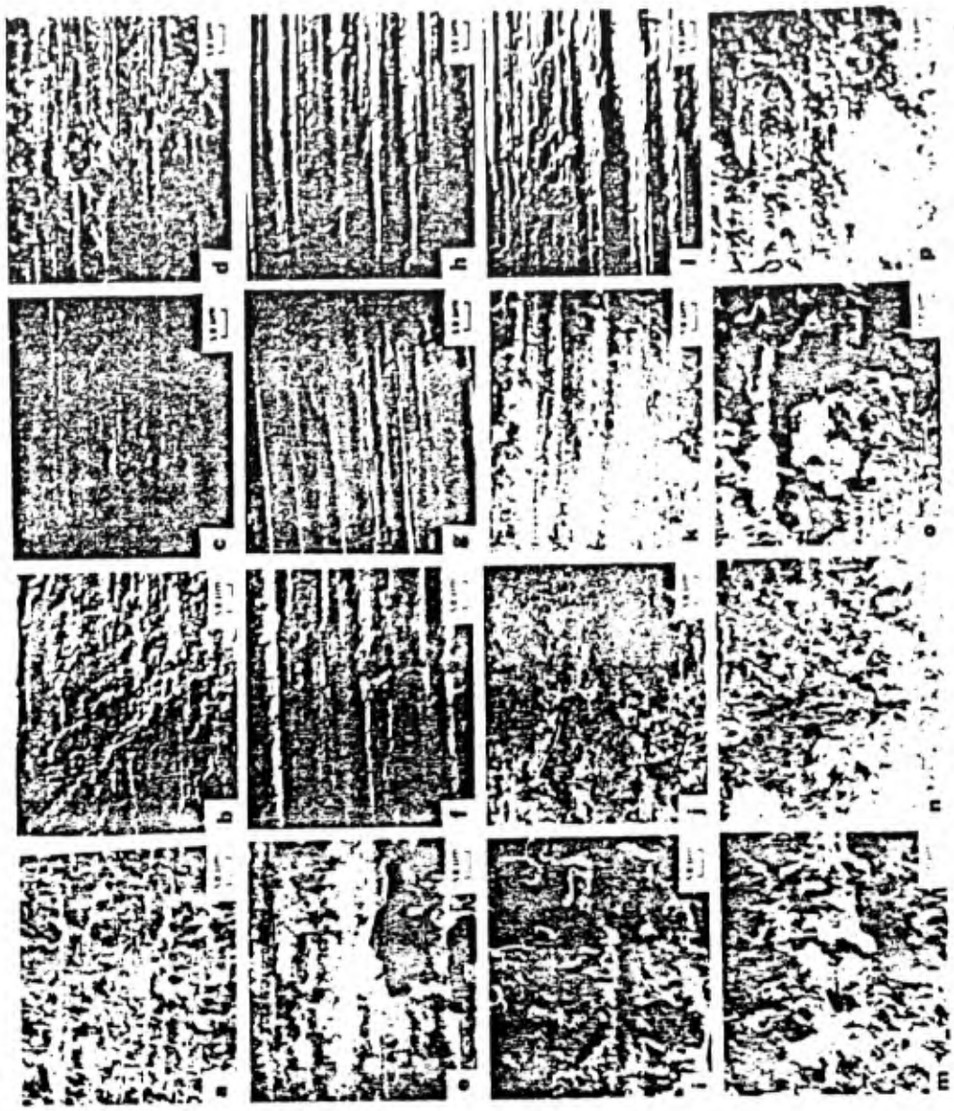


FIGURE 3.2 SCANNING ELECTRON MICROGRAPHS OF THE SURFACES OF WORN SPECIMENS;

(A), (E), (I), (M) IRON ON IRON, 1020, 1045, AND 1095 STEEL, RESPECTIVELY;

(B), (F), (J), (N) 1020 ON IRON, 1020, 1045, AND 1095 STEEL, RESPECTIVELY;

(C), (G), (K), (O) 1045 ON IRON, 1020, 1045, AND 1095 STEEL, RESPECTIVELY;

(D), (H), (L), (P) 1095 ON IRON, 1020, 1045, AND 1095 STEEL, RESPECTIVELY

are slid against each other. The drop in the coefficient of friction in Figure 3.3b is associated with mutual polishing of the mating surfaces[9]. The behavior shown in Figure 3.3b results primarily when the hardness of the stationary slider is much greater than the moving specimen.

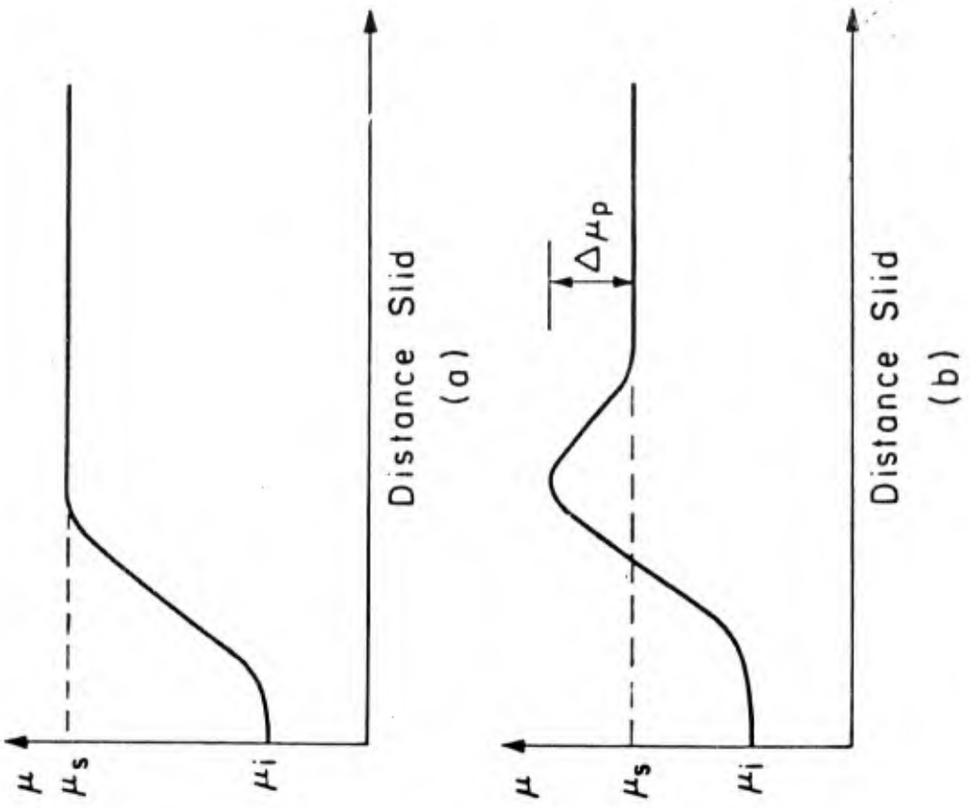


FIGURE 3.3 COEFFICIENT OF FRICTION VERSUS SLIDING DISTANCE:

(A) FOR AN ARMCO IRON SLIDER SLIDING AGAINST AN ARMCO IRON SPECIMEN (μ_i = 0.13, μ_s = 0.71);
(B) FOR AN AISI 1095 STEEL SLIDER SLIDING AGAINST AN ARMCO IRON SPECIMEN (μ_i = 0.17, μ_s = 0.51, $\Delta\mu_p$ = 0.25)

2. When wear particles are brushed from the sliding interface, the coefficient of friction decreases to a low value and gradually reaches a steady state value again, as schematically illustrated in Figure 3.4. (The effect of wear particles on the friction coefficient was also reported by

Suh; Jahanmir and Abrahamson and more recently by Kuwahara and Masumoto[7, 10].)

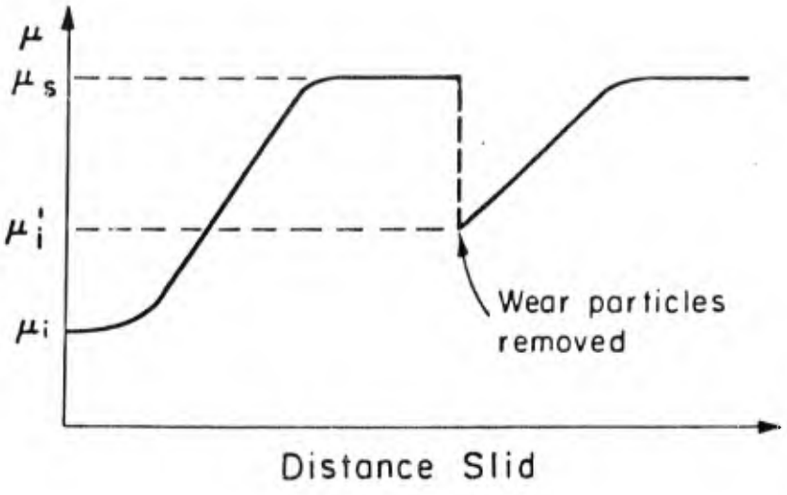


FIGURE 3.4 EFFECT OF REMOVING WEAR PARTICLES FOR AN ARMCO IRON SLIDER SLIDING AGAINST AN ARMCO IRON SPECIMEN (μ_i = 0.13, μ_s = 0.71, μ_i = 0.4)

3. The coefficient of friction can differ by as much as 0.2 even for the same pair of materials (which are chemically identical) depending on which is a stationary slider and which is a moving specimen, see Table 3.2.

4. The initial value of the kinetic coefficient of friction, μ_i, is in the neighborhood of 0.1 to 0.2 (but largely in the range of 0.12 to 0.17) for many materials tested, i.e., gold on gold, steel on steel, brass on steel, etc., and also regardless of whether or not lubricants are used[6, 9].

5. When the friction test is done with extremely well polished surfaces, plowing grooves are formed from the onset of testing.

6. Frictional behavior depends very much on experimental conditions.

3.3 Genesis of Friction

The experimental results clearly indicate that the observed friction coefficients cannot be explained in terms of

the adhesion theory. The effect of entrapped wear particles and the existence of μ_1, which is independent of environmental conditions and materials tested, cannot be explained by the adhesion theory. The theory is further defined by the dramatic changes in the coefficient of friction when the role of the slider and the specimen is reversed.

Based on the experimental results discussed in the preceeding subsection, the following postulate was advanced to explain the genesis of friction between the sliding surfaces[6]:

> "The coefficient of friction between the sliding surfaces is due to the various combined effects of asperity deformation, μ_d, plowing by wear particles and hard surface asperities, μ_p, and adhesion between the flat surfaces, μ_a. The relative contribution of these components depends on the condition of the sliding interface, which is affected by the history of sliding, the specific materials used, the surface topography and the environment."

In order to clarify the above postulate, the time dependent friction behavior of the materials tested, will be considered qualitatively by subdividing it into the following stages, see Figure 3.5.

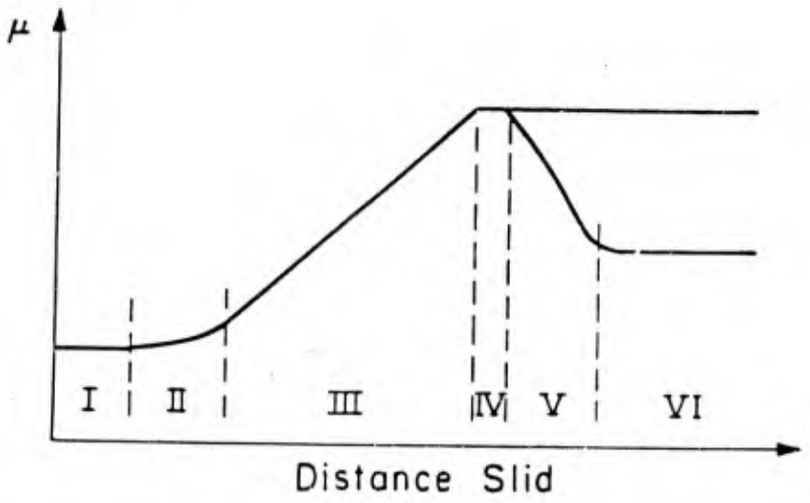


FIGURE 3.5 SIX STAGES IN THE FRICTIONAL FORCE VERSUS DISTANCE SLIDE RELATIONSHIP

a. Stage I -- In this early stage the coefficient of friction seems to be dictated by plowing of the surface by asperities. Adhesion does not play any significant role in this state due to the contaminated nature of the surface. The deformation of asperities does take place at the onset of sliding, which affects the static coefficient of friction. However, in Stage I, asperity deformation is not the major factor that determines the coefficient of friction, since they deform as soon as sliding commences and the surface is easily polished with the generation of new asperities in Stage I. Consequently, the coefficient of friction in this stage, μ_1, is largely independent of material combinations, the surface conditions, and the environmental conditions.

b. Stage II -- In this second stage, the frictional force begins to rise slowly due to increase in adhesion. When the interface is lubricated, Stage I persists for a long time and Stage II may not be present. The slope in Stage II can be steeper if the wear particles generated by the asperity deformation and fracture are entrapped between the sliding surfaces and proceed to plow the surfaces.

c. Stage III -- This stage is characterized by a steep increase in slope due to the rapid increase in the number of wear particles entrapped between the sliding surfaces as a consequence of higher wear rates. The slope is also affected by the increase in adhesion due to the increase in clean interfacial areas. The force required to deform the asperities will continue to contribute to the frictional force in this stage as long as surface asperities are present. The wear particles are generated when the process of wear particle formation by subsurface deformation, crack nucleation, and crack propagation postulated by the delamination theory or wear, is completed [1] Some of the wear particles get entrapped between the surfaces, causing plowing. The plowing will be greater when the wear particles are entrapped between metals of nearly equal hardness, because they will penetrate into both surfaces, preventing any slippage between the particle and the surface.

d. Stage IV -- This stage is reached when the number of wear particles entrapped between the interface remains constant. This occurs when the number of the newly entrapped particles equal the number of entrapped particles leaving the interface. The adhesion contribution to friction also remains constant is Stage IV. The asperity deformation continues to be important, since the wear by delamination creates new rough surfaces with asperities. However, in most cases asperity deformation is not as important as plowing since they deform readily and the frequency of new asperity generation is slow.

When two like metals are slid against each other or when the mechanism responsible for Stage V does not play a significant role, the coefficient between the two metals.

e. Stage V -- In some cases, such as when a very hard stationary slider is slid against a soft specimen, the asperities of the hard surface are gradually removed, creating a mirror finish as shown in Figure 3.6. In this case the frictional force decreases due to the decrease in plowing and asperity deformation. Plowing decreases since wear particles cannot anchor to a polished hard surface.

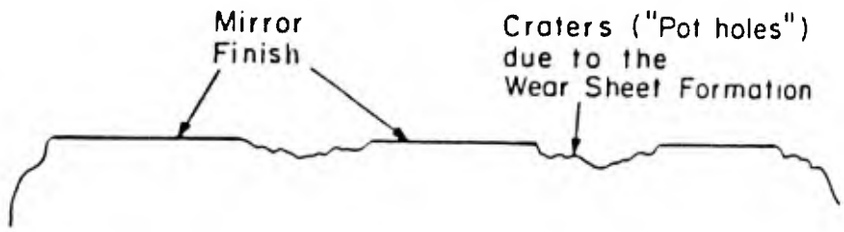


FIGURE 3.6 A HARD STATIONARY SURFACE POLISHED BY A SOFT SURFACE

f. Stage VI -- Eventually when the hard surface becomes mirror smooth to a maximum extent, the softer surface also aquires the same mirror finish and the frictional force levels off. The surfaces are never completely smooth since there are always "pot holes" due to the creation of delamination wear particles. These craters provide anchoring points for wear particles. When the hard surface is not stationary but moving against the softer surface, the hard surface remains rough, probably because polishing cannot take place due to geometric reasons. In this case, Stages V and VI are not present.

3.4 Analysis of the Friction Generating Mechanisms

The three basic mechanisms (i.e., asperity deformation, plowing, and adhesion) that are responsible for the generation of friction will be analyzed in this section. The asperity deformation determines the static coefficient of friction and also affects the dynamic coefficient of friction, since asperities are continuously generated due to delamination of wear sheets. However, the contribution of the asperity deformation to the dynamic coefficient of friction is not large relative to those by plowing and adhesion, since the asperities have to await the formation of delaminated wear

particles which often requires a large number of cyclic loading by the asperities of the opposing surface. On the other hand, plowing takes place continuously whenever wear particles are entrapped between the sliding surfaces or when the asperities of the counterface plow in all cases when clean flat surfaces come into contact during steady state sliding. The relative magnitude of these components will be determined approximately by using the slip-line field.

3.4.1 Analysis of the Asperity Deformation, μ_d

Consider two representative asperities approaching each other, as shown in Figure 3.7. When these asperities come into contact with each other, they have to deform in such a manner that the resulting displacement field is compatible with the sliding direction and that the sum of the vertical components of the surface traction at the contracting asperities must be equal to the applied normal load. A possible slip-line field that satisfies the kinematic condition is given in Figure 3.8. The solution demands that the shear stress along OA be whatever is necessary to satisfy the condition that $\alpha = \theta$, which is necessary to constrain the resulting deformation in the sliding direction. This would be possible even under the lubricated conditions if the interface, OA, is not perfectly smooth but rough enough to allow mechanical interlocking. The derivation of the normal and tangential force corresponding to the slip-line field shown in Figure 3.8 is given in the Appendix of this section.

The general solution is sketched in Figure 3.9. If it is assumed that the asperity deformation is the only phenomenon that takes place at the interface and is entirely responsible for the frictional force under a given load, the coefficient of friction due to asperity deformation μ_d varies from 0.39 to 1 as the slope of asperities increases from 0 to 45°. These values are closer to the static coefficient of friction than the dynamic friction coefficient measured during the early stage of sliding, i.e., Stage I[12].

In dynamic situations most of the normal load is carried by the entrapped wear particles and the flat contacts. Therefore, the actual contribution of the asperity deformation to the frictional force is expected to be a small fraction of the estimated value. This may be estimated indirectly from the experimental results.

It should be noted that the slip-line analysis done to determine μ_d is similar to Green's results which were discussed in Section 1[13]. However, in this case, μ_d, does not depend on adhesion since in the analysis the asperity

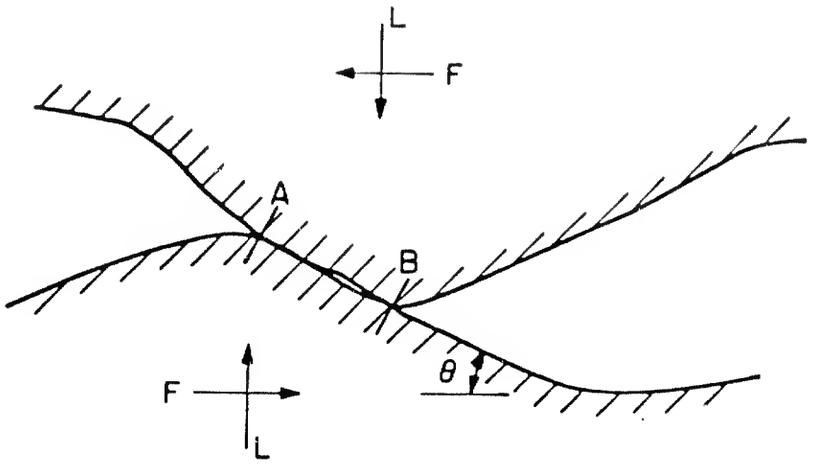


FIGURE 3.7 TWO INTERACTING SURFACE ASPERITIES

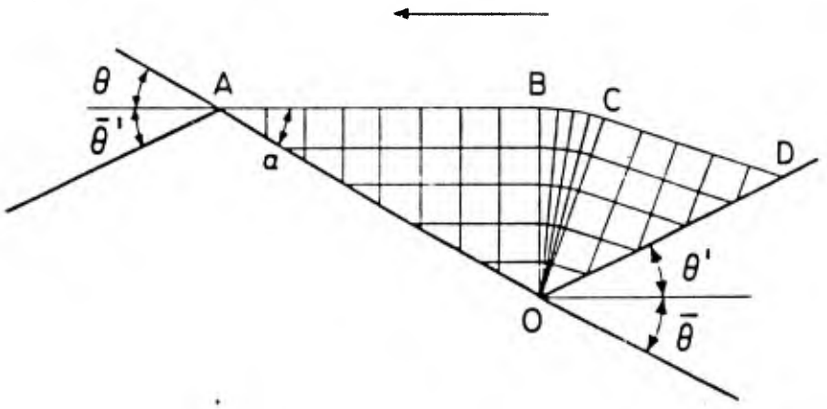


FIGURE 3.8 A GEOMETRICALLY COMPATIBLE SLIP-LINE FIELD. IT CAN BE SEEN THAT $\theta > \bar{\theta}$, $\theta' > \theta$, AND $\theta = \alpha$

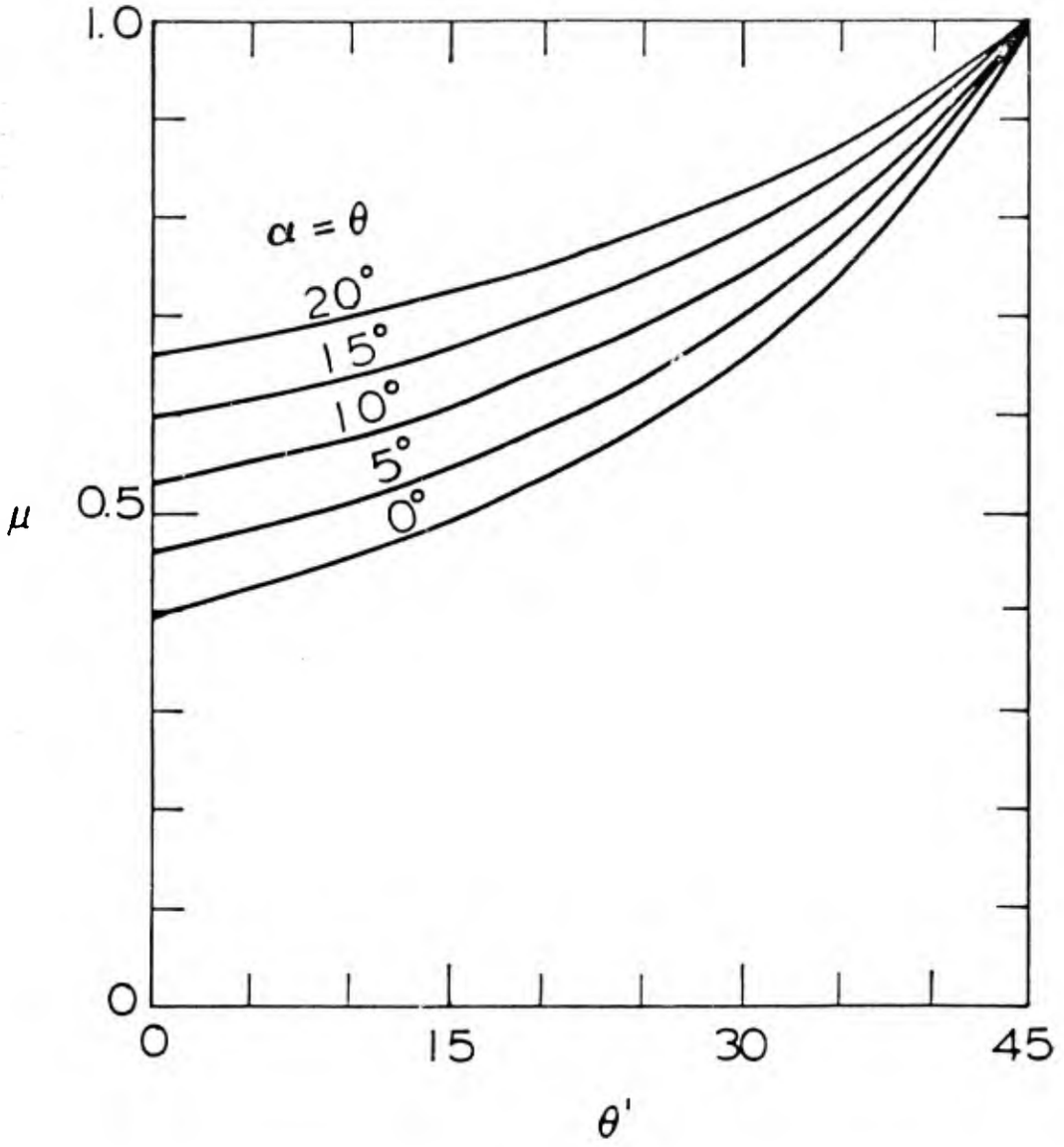


FIGURE 3.9 SLIP-LINE FIELD SOLUTION FOR FRICTION AS A FUNCTION OF THE SLOPE OF ASPERITIES

deformation was kinematically contrained. μ_d represents the frictional force due to the deformation of all interacting asperities. Before the onset of sliding between two surfaces, μ_d largely controls the static coefficient of friction.

3.4.2 Analysis of Adhesion Component of the Friction Coefficient, μ_a

A frictional force can arise due to the adhesion of two nearly flat surfaces. Unlike the deformation of asperities, this frictional force is a function of the adhesion between the two opposing surfaces. The adhesion force arises either due to the welding of two nearly flat portions of the surface or when the atoms are brought together in close proximity for interatomic interactions but without welding. The adhesion at the slopes of two interacting asperities has been included in deriving in μ_d as a subset of kinematically constrained deformation problems.

The specific experimental results presented in this section show that μ_a is not present (or is at least negligible) at the onset of sliding, probably due to the presence of contaminants on the surface. With the deformation of asperities and exposure of fresh new surfaces, the adhesion between nearly flat surfaces is expected to increase. The exact adhesion area cannot be determined a priori, since the applied normal load may also be carried out by interacting asperities and entrapped wear particles, although the limiting cases can be analyzed.

Consider two nearly flat surfaces coming into contact as shown in Figure 3.10. (Sometimes this type of contact is called a "rubbing" contact.) Depending on the nature of adhesion along the interface ED, the force required to move the rubbing surfaces with respect to each other varies. When there is no adhesion the force will be zero and when there is complete adhesion it will reach a maximum. The solution to this problem can be obtained again using the slip-line fields similar to that shown in Figure 3.10. The exact geometric shape of the slip-line field will depend on the boundary condition at ED. The solution sought can be adapted from the recent work of Challen and Oxley who derived an expression for the friction coefficient as[14];

$$\mu_a = \frac{A \sin \alpha + \cos (\cos^{-1} f - \alpha)}{A \cos \alpha + \sin (\cos^{-1} f - \alpha)} \qquad (3.1)$$

where

$$A = 1 + \frac{\pi}{2} + \cos^{-1} f - 2\alpha - 2 \sin \frac{-1 \sin\alpha}{1-f}$$

f = strength of the adhesion at ED as expressed at a fraction of the shear flow strength of the softer material (3.2)

α = the slope of the hard asperity

For nearly flat surfaces $\alpha \rightarrow 0$. Therefore, Eqs. (3.1) and (3.2) reduce to

$$\mu_a = \frac{f}{A + \sin(\cos^{-1} f)} \tag{3.3}$$

$$A = 1 + \frac{\pi}{2} + \cos^{-1} f \tag{3.4}$$

It can be seen that μ_a varies from 0 to 0.39 as f changes from 0 to 1.

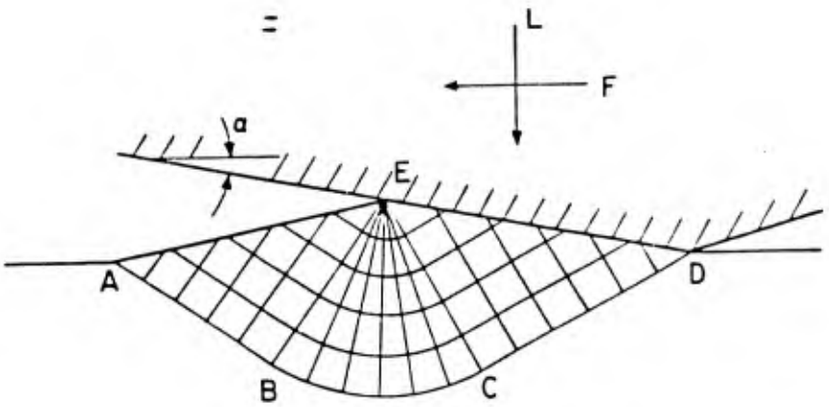


FIGURE 3.10 A SLIP-LINE FIELD FOR A RUBBING CONTACT

The friction coefficient determined by Eqs. (3.1) and (3.2) are based on the assumption that all the applied normal load is carried by the flat interfaces. However, since part of the normal load is also carried by purely elastic contacts, the interacting asperity junctions discussed in the preceeding

section and the entrapped wear particles, μ_a, under typical sliding conditions, should be less than 0.4. The experimental results obtained with the hard AISI 1095 steel slider and the soft Armco iron specimen showed that the steady state coefficient of friction reached a value of 0.51 when both surfaces were polished smooth and thus the friction was caused primarily by adhesion. The agreement between the theory and the experiment is reasonable since asperity interactions and plowing by wear particles must have also contributed to the frictional force.

3.4.3 Analysis of Plowing, μ_p

The plowing component of the frictional force can be due to the penetration of hard asperities or due to the penetration of wear particles. The plowing due to wear particles is schematically illustrated in Figure 3.11. When two surfaces are of equal hardness, the particle can penetrate into both surfaces. As the surfaces move with respect to each other, grooves will be formed in one or both of the surfaces. When one of the surfaces move with respect to each other, grooves will be formed in one or both of the surfaces. When one of the surfaces is very hard and smooth, the wear particle will simply slide along the hard surface and no plowing can occur. However, when the hard surface is very rough, wear particles can anchor in the hard surface and plow the soft surface.

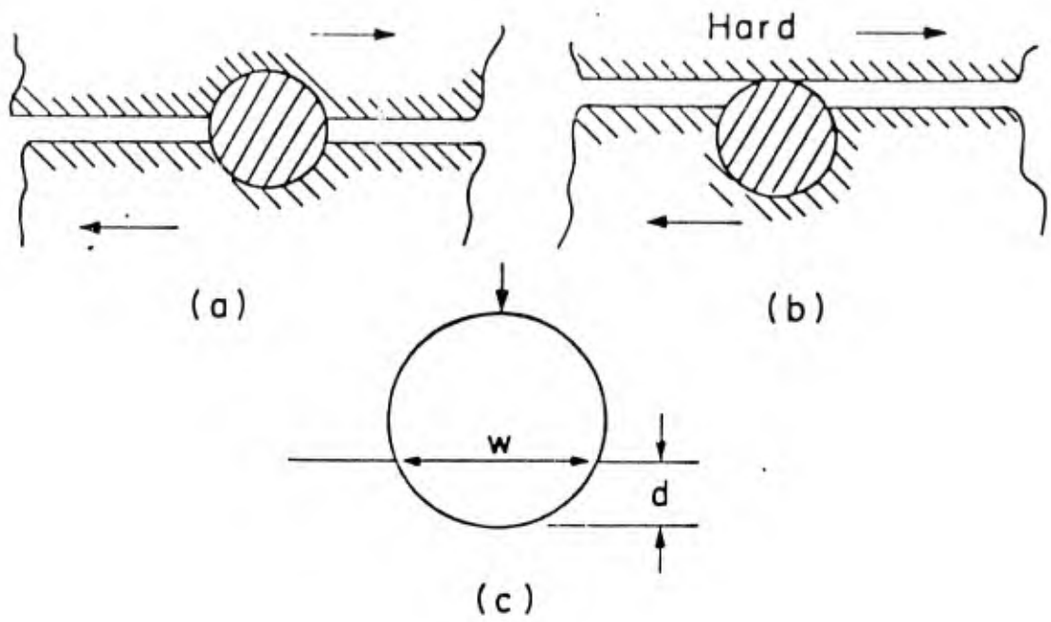


FIGURE 3.11 IDEALIZED MODEL OF WEAR PARTICLE INTERACTION BETWEEN TWO SLIDING SURFACES: (A) SURFACES OF EQUAL HARDNESS; (B) ONE SMOOTH VERY HARD SURFACE: (C) GEOMETRY OF THE WEAR PARTICLE

The friction due to plowing was investigated by Sin, et al., which showed that the contribution of plowing to the friction coefficient is very sensitive to the ratio of the radius of curvature of the particle to the depth of penetration[15]. The friction coefficient by plowing, μ_p, is given by:

$$\mu_p = \frac{2}{\pi}\left[\left(\frac{2r}{w}\right)^2 \sin^{-1}\frac{w}{2r} - \left(\frac{2r}{w}\right)^2 - 1\right]^{1/2} \qquad (3.5)$$

where w is the width of the penetration and r is the radius of curvature of the particle. The ratio of w/r, measured by sectioning the worn specimen is in the neighborhood of 0.8. Substituting this value into Eq. (3.5), the plowing coefficient of friction is found to be 0.2. This value is in the same range as the decrease in the friction coefficient observed by removing the wear particles from the Armco iron/Armco iron and the Armco iron/AISI 1095 steel interfaces, which were 0.31 and 0.16 respectively. The range of possible values of μ_p as a function of the ratio w/2r is shown in Figure 3.12.

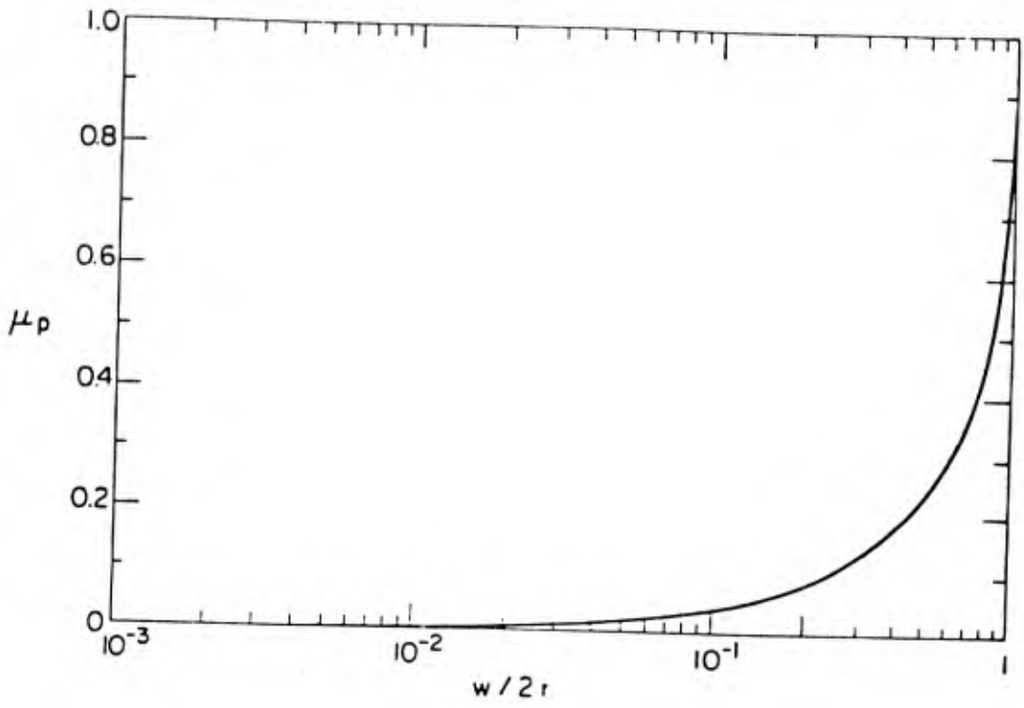


FIGURE 3.12 PLOWING COMPONENT OF THE FRICTION COEFFICIENT AS A FUNCTION OF THE RATIO OF THE WIDTH TO THE DIAMETER OF THE ENTRAPPED WEAR PARTICLE

Plowing not only increases the total frictional force and delamination wear, but also creates small wear particles which in turn affect the subsequent wear of sliding surfaces. Plowing action forms ridges along the sides of plowed grooves. When these ridges are deformed, flat, and subjected to repeated loading, some of these become loose wear particles with continued sliding. This is schematically illustrated in Figure 3.13.

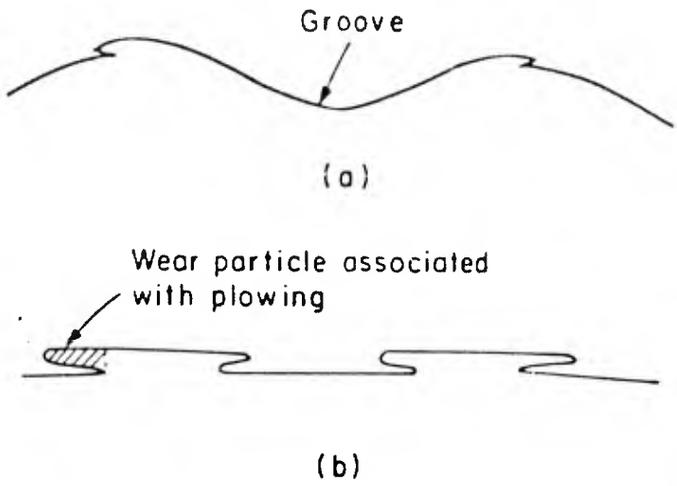


FIGURE 3.13 SCHEMATIC ILLUSTRATION OF WEAR PARTICLE FORMATION DUE TO PLOWING: (A) RIDGES FORMED ALONG THE SIDES OF THE PLOWED GROOVES; (B) FLATTENED RIDGES

3.5 Relative Contributions of μ_i, μ_a and μ_p to the Overall Friction Force

If the postulate for friction coefficient presented in this paper is correct, then the relative values of various friction components are as follows:

μ_d -- From 0.43 to 0.75, when the entire applied normal load is carried by typical surface asperities with a slope of 4° to 20°. It appears that μ_d is responsible for the static coefficient of friction μ_i in Stage I. The reason μ_d is not a major factor in Stage I is that once the original asperities deform, asperity interactions cannot take place. This friction component can contribute partially to the steady state coefficient of friction if new asperities are continuously generated as a consequence of delamination wear process.

μ_a -- From about 0 to 0.4, depending on the nature of adhesion between the flat part of the interacting surfaces. The low value is for a well lubricated surface with light lubrication, while the high value is for identical metals sliding against each other without any surface contaminants and oxide layers.

μ_p -- From nearly 0 to 1.0, from a theoretical point of view, depending on the depth of penetration, but normally less than 0.4 in a typical situation. The high values are

associated with two identical metals sliding against each other with deep penetration by wear particles, while the low value is obtained when either wear particles are totally absent from the interface or a soft surface is slid against a hard surface with a mirror finish.

The determination of the total friction coefficient in a given condition is complex. It is difficult to determine the relative contributions of μ_d, μ_a and μ_p to the total friction coefficient, because analyses for μ_d and μ_a were done assuming that the total normal load is carried by either asperities or flat contact areas. In real situations, the normal load will be apportioned among the asperity contacts, flat adhesion junctions, and the entrapped particles. However, it is quite plausible in many real situations, that each of these mechanisms which contribute to friction can take place sequentially rather than concurrently. Consider, for example, a flat junction and an asperity in contact as shown in Figure 3.14. When the flat areas first come into contact and form an adhesion junction, the analysis performed for μ_a is strictly valid. When the asperities also come into contact with further sliding, a large fraction of the normal load is still carried by the flat junctions due to their higher normal "stiffness," i.e., the force required to cause unit displacement along the vertical direction. In this case, in order to continue the sliding action the asperities will simply have to shear along the dotted line if the materials are identical, or along the crossed line if the top slider is much harder than the bottom slider. In this case, only a tangential force is required to deform the asperity plastically along the direction of sliding.

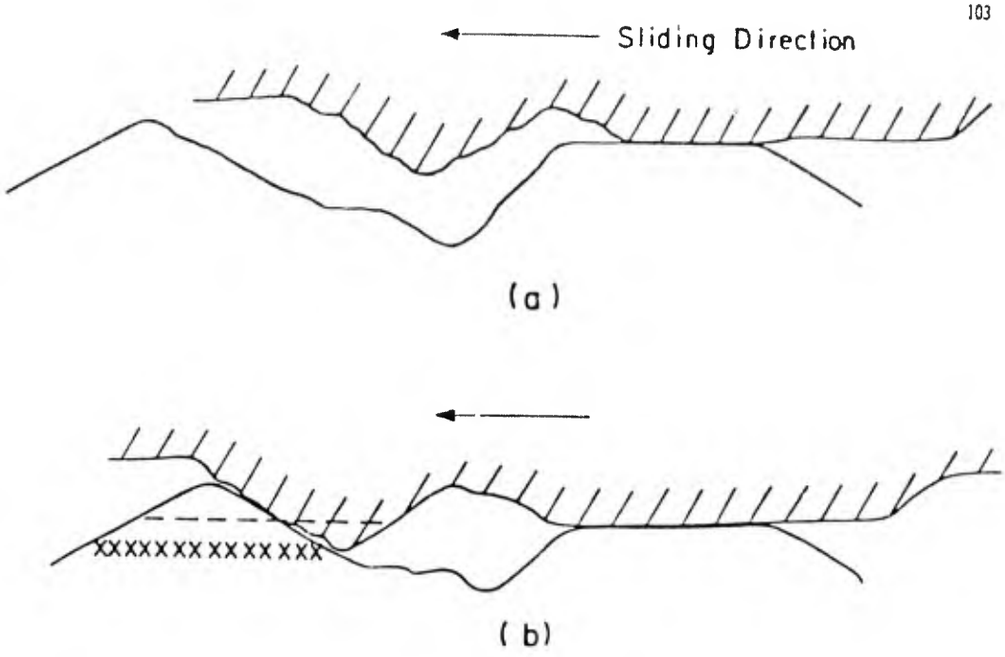


FIGURE 3.14 TWO SLIDING SURFACES IN CONTACT: (A) FIRST CONTACT OF FLAT SURFACES; (B) FLAT SURFACE CONTACT AND ASPERITY CONTACT

The coefficient of friction μ is represented in a "Friction Space" as a function of adhesion, plowing, and roughness as shown in Figure 3.15. The adhesion is expressed in terms of the nondimensional interfacial shear strength of the flat contacts which was defined by Eq. (3.2). The roughness is plotted in terms of the slope of the surface, while the plowing is given in terms of the ratio of the width of wear particle (or hard asperity) penetration to its radius. The lowest surface corresponds to the case of no asperities, i.e., $\theta = 0$, which forms the lower bound. The θ_i surface corresponds to the initial machined surface, while the θ^* surface represents the surface roughness when the peak of μ occurs. As the surface gets rougher and/or the number of the steady state asperities increase, μ will increase and the friction value will move in the Friction Space along the μ-axis. The friction surfaces 0^* is plotted from the actual experimental results obtained with Armco iron sliding against AISI 1095 steel. In this case, the asperity contribution to friction was 0.3.

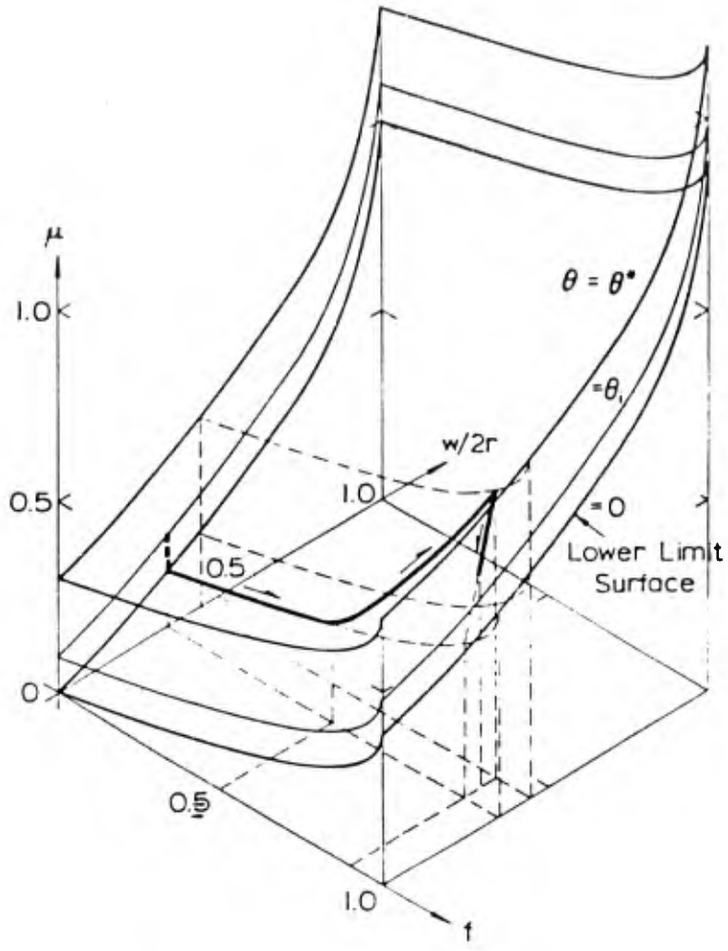


FIGURE 3.15 "FRICTION SPACE" SHOWING THE COEFFICIENT OF FRICTION AS A FUNCTION OF ADHESION BETWEEN FLAT CONTACTS, WEAR PARTICLE PENETRATION, AND SURFACE ROUGHNESS: $f = \tau_s/k$, WHERE τ_s IS THE SHEAR STRESS AT THE INTERFACE AND k IS THE SHEAR FLOW STRENGTH OF THE SOFT METAL; θ IS THE SLOPE OF ASPERITIES; w/2r IS THE RATIO OF THE WIDTH OF ASPERITY PENETRATION TO THE DIAMETER OF THE PARTICLE

Although this figure is not precise, it gives a reasonable picture of what happens in a given situation. The paths of the friction coefficient change, shown in Figure 3.3(a) and 3.3(b), are shown in this "Friction Space," Figure 3.15. The friction is shown to start from the initial roughness plane θ_i to nearly flat surface, i.e., $\theta = 0$ and traces along the paths indicated in Figures 3.3(a) and 3.3(b). In many cases the θ_i surface may be above the θ^* friction space, depending on the initial surface finish relative to the steady state surface roughness.

The foregoing argument may be applied to the specific case of gold sliding against gold. When unlubricated gold specimens without any oxide layer are slid against each other, the frictional force is the sum of μ_a and μ_p or the sum of a fraction of μ_d, μ_a and μ_p, depending on the situation. It is most likely that the μ_a term will always be present since two flat surfaces in contact may be the most stiff system in supporting the normal load. For example, if the normal load is first borne by flat contacts only, friction will be entirely due to adhesion, μ_a, which will reach a maximum value. Then if asperities of the opposing surfaces come in contact, an additional frictional force will be required to deform the asperities. In addition to these frictional forces, the third frictional component may also affect the frictional behavior if the wear particles become wedged in between the sliding surfaces. Therefore, the coefficient of friction between gold on gold may be as high as 1.4 to 1.6 and fluctuate between a maximum and a minimum value.

Lubricated surfaces can have a coefficient of friction whose magnitude will be determined by the degree of plowing and asperity interaction. Lubricated surfaces, without any asperities, are found to have a coefficient of friction of approximately 0.04 for a hard surface and 0.12 for a soft iron surface. However, when wear particles are entrapped between the sliding surfaces, the plowing component of the frictional force can be present raising the friction coefficient.

In the past, the high friction coefficient between like metals has been explained in terms of greater adhesion due to their greater solubility[1]. However, the evidence presented in this paper shows that the so-called compatibility of metals is dictated more by their mechanical behavior rather than by their chemical behavior. This is quite reasonable since the diffusion rate at the typical sliding junctions at room temperature is so low that the solubility between the metals cannot account for the observed wear rates[16].

3.6 Concluding Remarks

In this section it is shown that the classical adhesion theory of friction cannot explain experimental results and that the frictional coefficient is not an inherent material property. The coefficient of friction depends very much on the sliding conditions, material combinations, and geometry. The coefficient of friction is composed of three components: that due to the deforming asperities, μ_d, that due to plowing by wear particles entrapped between sliding surfaces and hard surface asperities, μ_p, and that due to adhesion, μ_a. The contribution to the overall coefficient of friction by plowing

and asperity deformation can be greater than that by adhesion. Typical values of μ_a can range from 0 to 0.4 and μ_p from 0 to 0.4 under typical conditions. However, μ_p can be as large as 1.0 when the depth of penetration by wear particles is large. The friction coefficient due to asperity deformation dictates the static friction coefficient can range from 0.43 to 0.75 depending on the slope of asperities.

The mechanisms responible for the genesis of friction presented in this section should be operative in the case of ceramics, thermosetting plastics, certain kinds of thermoplastics, and most metals. The frictional behavior of highly crystalline thermoplastics, such as teflon and high density polyethylene are affected also by the molecular orientation of molecules and their transfer to the counterface. The frictional behavior of crystalline thermoplastics will be treated in a separate section.

3.7 References - Section 3

1. Rabinowicz, E., "Friction and Wear of Materials," Wiley, New York, 1965. pp. 51-108.

2. Bowden, F.P. and D. Tabor, "Friction and Lubrication of Solids," Clarendon Press, Oxford, Part I, 1950, pp. 90-121, Part II, 1964, pp. 52 -86.

3. Ernst, H., and M.E. Merchant, "Surface Friction Between Metals - A Basic Factor in the Metal Cutting Process," Proceedings of the Special Summer Conference on Friction and Surface Finish, M.I.T. Press, Cambridge, 1940, pp. 76-101.

4. Shaw, M.C. and E.F. Macks, "Analysis and Lubrication of Bearings," McGraw Hill, 1949, pp. 457-461.

5. Buckley, D.H., "Definition and Effect of Chemical Properties of Surfaces in Friction, Wear, and Lubrication", Proceedings of the International Conference on the Fundamentals of Tribology, M.I.T. 1978, pp. 173-199.

6. Suh, N.P. and H.C. Sin, "The Gensis of Friction", Wear, 69(1981) 91-114.

7. Abrahamson, II, E.P., S. Jahanmir, and N.P. Suh, "The Effect of Surface Finish on the Wear of Sliding Surfaces," CIRP Ann. Int. Inst. Prod. Eng. Res., 24 (1975) 513-514.

8. Rabinowicz, E., Proceedings of the International Conference on Wear of Materials, St. Louis, Missouri, 1977, American Society of Mechanical Engineers, pp. 36-40.

9. Suh, N.P., H.C. Sin, M. Tohkai, and N. Saka, "Surface Topography and Functional Requirements for Dry Sliding Surfaces," CIRP Ann. Int. Inst. Prod. Eng. Res., 29 (1980) 413-418.

10. Kuwahara, K. and H. Masumoto, "Influence of Wear Particles on the Friction and Wear Between Copper Disk a nd Pin of Various Kinds of Metal," Journal of Japan Society of Lubrication Engineers, 25 (1980) 126-131.

11. Suh, N.P., "The Delamination Theory of Wear," Wear, 25 (1973) 111-124.

12. Kato, S., E. Mauri, A. Kobayashi, and T. Matsubayshi, "Characteristics of Surface Topography and Static Friction on Scraped Surface Slideway, Part I and Part II," Journal of Engineering for Industry, Trans. ASME, Vol. 103 (1980) pp. 97-108.

13. Green, A.P., "Friction Between Unlubricated Metals: A Theoretical Analysis of the Junction Model," Proceedings of Royal Society of London, A228 (19955) 191-204.

14. Challen, J.M. and P.L.B. Oxley, "An Explanation of the Different Regimes of Friction and Wear Using Asperity Deformation Models," Wear, 53 (1979) 229-243.

15. Sin, H.C., N. Saka, and N.P. Suh, "Abrasive Wear Mechanisms and the Grit Size Effect," Wear 55 (1979) 163-190.

16. Suh, N.P., "New Theories of Wear and Their Implications on Tool Materials," Wear, 1980, to appear.

3.8 Appendix - Section 3 - Slip-Line Field Solution for Sliding Contact

Figure 3.16 gives a possible slip-line field for the asperity contact between sliding surfaces. The interface AO between asperities and the stress free surface OD are both assumed to be straight with their directions defined by the angles θ and θ' measured from the sliding direction. From the figure it can be noticed that the slip-line ABCD is a β- line. Using the Hencky relations the stresses along the slip-line can be obtained.

At D, $\phi = \theta' + \frac{\pi}{4}$ and $p = k$ (1.A1)

Along AB, $\phi = \alpha - \theta + \frac{\pi}{2}$ and

$$P_{AB} = k\,(1 + \frac{\pi}{2} + 2\alpha - 2\theta - 2\theta') \quad (1.A2)$$

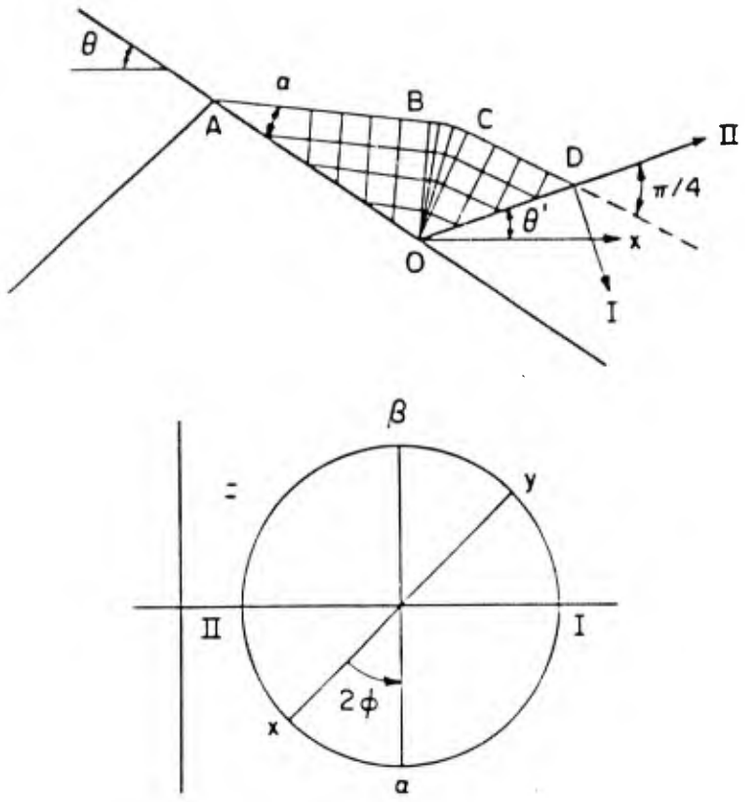


FIGURE 3.16 A SLIP-LINE FIELD FOR AN ASPERITY CONTACT

Isolating the junction along ABO as shown in Figure 3.17 we can find the resultant forces as

$$-F_y = = (AB)p \cos(\theta-\alpha) - (AB)k \sin(\theta-\alpha)$$

$$+ (OB)k \cos(\theta-\alpha) - (OB)p \sin(\theta-\alpha) \quad (1.A3)$$

$$-F_x = F = (AB)p \sin(\theta-\alpha) + (AB)k \cos(\theta-\alpha)$$

$$+ (OB)k \sin(\theta-\alpha) + (OB)p \cos(\theta-\alpha)$$

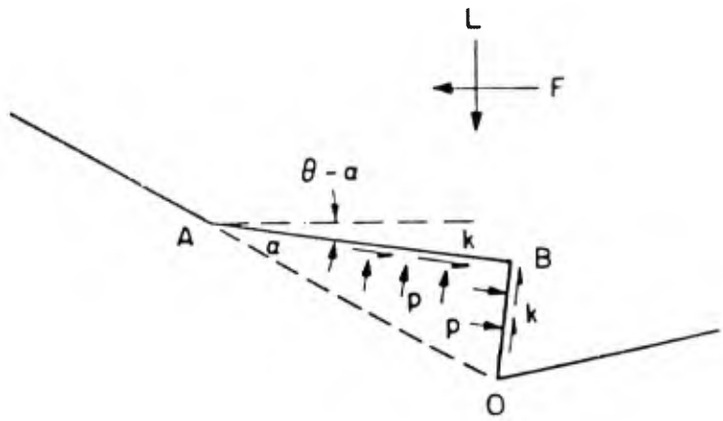


FIGURE 3.17 ISOLATION OF THE JUNCTION OF THE ASPERITY CONTACT ALONG ABO IN FIGURE 3.16

=

Using the geometric relations, L and F are expressed as

$$L = (AO)\ (p \cos\theta + k \sin\ (2\alpha - \theta)) \qquad (1.A4)$$

$$F = (OA)\ (p \sin + K \cos(2\alpha - \theta))$$

Therefore, the coefficient of friction is

$$\mu = \frac{F}{L} = \frac{(1 + \frac{\pi}{2} + 2\alpha - 2\theta - 2\theta')\sin\theta + \cos(2\alpha - \theta)}{(1 + \frac{\pi}{2} + 2\alpha - 2\theta - 2\theta')\cos\theta + \sin(2\alpha - \theta)} \qquad (1.A5)$$

When θ, θ', and α are small ($\theta, \theta', \alpha \rightarrow \theta$)

$$\mu = \frac{1}{1+\frac{\pi}{2}} = 0.39$$

For several values of α, θ, and θ', the coefficient of friction is plotted in Figure 3.9.

If the junction does not weld along OA, the shear stress along OA will be much less than k. The interfacial shear stress is related to the angle α as $\tau = k \cos 2\alpha$. When there is no shear stress along OA, i.e., $\alpha = \pi/4$, the junction will slide along OA until the junction can deform under the influence of the normal load alone. During this sliding the coefficient of friction can be otained by substituting $\alpha = \pi/4$, into Eq. 1.A5 as

$$\mu = \frac{(1 + \pi - 2\theta - 2\theta') \sin\theta + \cos\left(\frac{\pi}{2} - \theta\right)}{(1 + \pi - 2\theta - 2\theta') \cos\theta + \sin\left(\frac{\pi}{2} - \theta\right)}$$

or

$$\mu = \tan\theta$$

which is the same expression as that derived from the roughness theory of friction.

The slip-line field solution derived above is a general solution. When the sliding occurs, the slip-line field should satisfy the kinematic condition, which corresponds to the case of the slip-line AB being parallel to the sliding direction. Therefore, α is equal to θ.

4. RESPONSE OF MATERIALS TO SURFACE TRACTION

4.1 Introduction

In Section 3 the origin and the magnitude of surface traction at asperity contacts were given. It was shown that the force acting on an asperity contact is due to one of the following forces: plowing, adhesion, and asperity deformation. The problem to be examined in this section is how the material responds to the external force applied at the interface between the sliding surfaces. In order to answer this question the responses of the surface layer will be treated as a continuum mechanics problem.

There are three basic responses to the external force:

1) plastic deformation of the surface layer, including the formation of ridges on the surface and the subsurface deformation,
2) subsurface crack and void nucleation
3) subsurface crack propagation

Each of these constitutes a separate class of mechanic problems and therefore will be treated individually in the following subsections.

For the purpose of analysis the material will be assumed isotropic and homogeneous. These are not necessarily good assumptions, because the surface becomes highly anisotropic and in-homogeneous during sliding. However, the need to simplify the mathematic analysis justifies these assumptions. On the other hand, these assumptions may not be too unreasonable in studying crack nucleation and propagation, since the amount of deformation at the subsurface where cracks nucleate and propagate is much less than the outermost layer of the surface.

It will be also assumed that the interfacial temperature rise is negligible because of the low sliding speed. This is a very reasonable assumption in the case of metals sliding at low speeds. In a later section on metal cutting, the case of high interfacial temperature rise will be considered.

4.2 Deformation of the Surface Layer

4.2.1 Formation of Surface Ridges

When plowing occurs, ridges form along the plowed groove regardless of whether or not chips are formed by the plowing action. These ridges become flattened as shown in Figure 4.1, which eventually fracture upon repeated loading. This process of wear particle formation has not been analyzed, although many experimental results support the qualitative view.

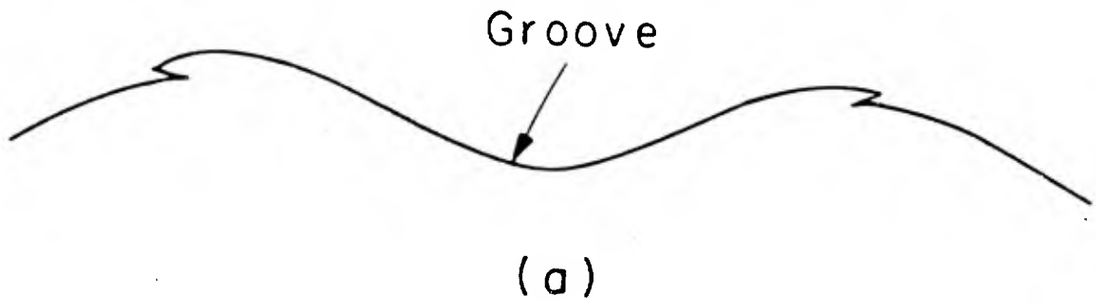
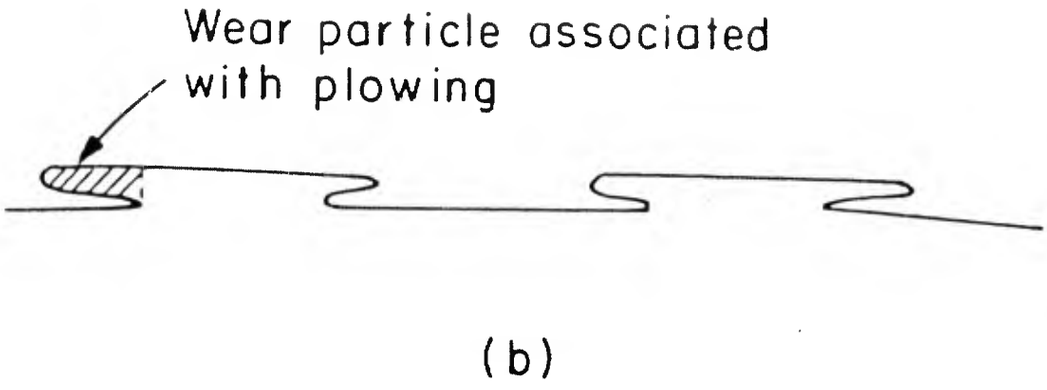


FIGURE 4.1



(b)

FIGURE 4.1 SCHEMATIC ILLUSTRATION OF WEAR PARTICLE FORMATION DUE TO PLOWING: (A) RIDGES FORMED ALONG THE SIDES OF THE PLOWED GROOVE; (B) FLATTENED RIDGES.

4.2.2 Deformation of a Semi-Infinite Elasto-Plastic Solid

The problem to be solved is the deformation of a semi-infinite elasto-plastic solid when it is loaded by moving asperities. The contact length at the asperity surface interface may be of the order of 10 to 100 μm and the distance between the asperity contacts may be 100 to 1000 μm apart. Therefore, the plastic deformation of the surface accumulates as the number of cyclic loading increases. The solutions to these problems can be obtained using the finite element method or using an approximate analytical/numerical technique. FEM solutions are possible for a variety of constitutive relationships but become very expensive to compute. The approximate technique used by Merwin and Johnson, Johnson and Jeffries, and Jahanmir and Suh will be outlined here[1, 2, 3]. This result will be compared to a FEM solution recently obtained at MIT.

When an elastic-plastic solid is loaded cyclically by moving asperities, residual stress and strain remain in the surface layer. This is of interest since these residual stress and strain fields may affect the subsequent crack nucleation and propagation processes. We will, therefore, show how the state of stress, due to the load applied by a moving asperity and the residual stress and strain field after repeated loading, can be determined. In the following subsections these results will be applied to solve crack nucleation and propagation problems.

Figure 4.2 shows the load exerted on a semi-infinite solid surface by an asperity. For convenience, it will be assumed that the asperity is stationary and the semi-infinite solid is moving with velocity U. The contact length is 2a. We will assume a plane strain condition, i.e., $\sigma_{zz} = 0$ and $\partial/\partial z = 0$, where the z axis is perpendicular to the plane of the paper. The stress distribution at the asperity contact will be assumed to be elliptic over the contact, which may be written as

$$\sigma_{xx} = \begin{cases} 2 q_o \left[\frac{x}{a} - \left(\frac{x^2}{a^2} - 1\right)^{1/2}\right] & \text{for } x \geq a \\ 2 q_o \left[\frac{x}{a} + \left(\frac{x^2}{a^2} - 1\right)^{1/2}\right] & \text{for } x \leq a \\ 2 q_o \left[\frac{x}{a} - P_o \left(1 - \frac{x^2}{a^2}\right)^{1/2}\right] & \text{for } x \leq a \end{cases} \quad (4.1)$$

$$\sigma_{yy} = \begin{cases} 0 & \text{for } -a \geq x \geq a \quad (4.2) \\ -P_o \left(1 - \frac{x^2}{a^2}\right)^{1/2} & \text{for } |x| \leq a \end{cases}$$

$$\sigma_{xy} = \begin{cases} 0 & \text{for } -a \geq x \geq a \quad (4.3) \\ q_o \left(1 - \frac{x^2}{a^2}\right)^{1/2} & \text{for } |x| \leq a \end{cases}$$

where σ's are stresses and the first subscript denotes the direction and the second subscript denotes the plane perpendicular to the axis. p_o and q_o are maximum normal stress and the maximum tangential stress at the contact.

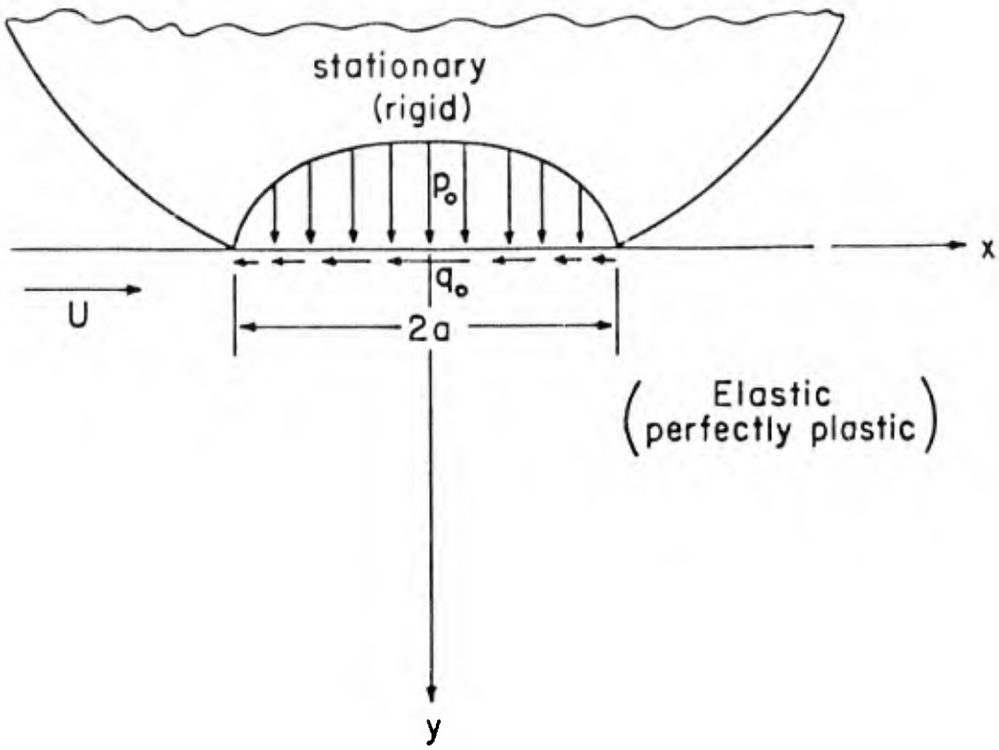


FIGURE 4.2 MODEL OF A CONTACT BETWEEN A STATIONARY RIGID ASPERITY AND A SLIDING ELASTIC-PERFECTLY PLASTIC PLANE.

1. Elastic case

We will first obtain a solution for an elastic solid which will then be used to obtain a solution for elasto-plastic solid. The governing equations are the equilibrium condition, geometric compatibility relations and the constitutive relations, which may, respectively, be written as:

$$\frac{\partial \sigma_{ij}}{\partial x_j} = 0 \qquad (4.4)$$

$$\varepsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \qquad (4.5)$$

$$\sigma_{ij} = \sigma_{ij} - \sigma = 2G\,(\varepsilon_{ij} - \varepsilon) = 2G\,\varepsilon_{ij}' \qquad (4.6)$$

where ε_{ij} is an incremental strain tensor; u's are displacement components, G is the shear modulus; σ_{ij} and ε_{ij}' are deviatoric stress and strain components, respectively; and σ and ε are the hydrostatic stress and the dilatational strain defined as:

$$\sigma = \sigma_{ii}/3 \tag{4.7}$$

$$\varepsilon = \varepsilon_{ii} \tag{4.8}$$

When the subscripts repeat, summation is implied.

The elastic solution for this contact problem was first obtained by Smith and Liu[5]. The stresses below the surface due to the both tangential and normal loads are given as:

$$\sigma_{xx} = \frac{q_o}{\pi}\left[(2x^2 - 2a^2 - 3y^2)\,\psi + 2\pi\,\frac{x}{a} + 2\,(a^2 - x^2 - y^2)\,\frac{x}{a}\,\bar{\psi}\right] - \frac{P_o}{\pi} y\left[\frac{a^2 + 2x^2 + 2y^2}{a} - \bar{\psi}\,\frac{2\pi}{a} - 3x\psi\right] \tag{4.9}$$

$$\sigma_{yy} = \frac{q_o}{\pi}\,y^2\psi - \frac{P_o}{\pi}\,y\left[a\bar{\psi} - x\psi\right] \tag{4.10}$$

$$\sigma_{xy} = \frac{q_o}{\pi}\left[(a^2 + 2x^2 + 2y^2)\,\frac{y}{a}\,\bar{\psi} - 2\pi\,\frac{y}{a} - 3xy\psi\right] - \frac{P_o}{\pi}\,y^2\psi \tag{4.11}$$

in which

$$\psi = \frac{\pi}{k_1}\,\frac{1 - \left(\frac{k_2}{k_1}\right)^{1/2}}{\left(\frac{k_2}{k_1}\right)^{1/2}\left[2\,\frac{k_2}{k_1}^{1/2} + \frac{k_1 + k_2 - 4a^2}{k_1}^{1/2}\right]} \tag{4.12}$$

$$\bar{\psi} = \frac{\pi}{k_1}\,\frac{1 + \left(\frac{k_2}{k_1}\right)^{1/2}}{\left(\frac{k_2}{k_1}\right)^{1/2}\left[2\left(\frac{k_2}{k_1}\right)^{1/2} + \frac{k_1 + k_2 - 4a^2}{k_1}\right]^{1/2}}$$

$$k_1 = (a + x)^2 = y^2$$

$$k_2 = (a - x)^2 + y^2$$

For plane strain the other stress components are given as:

$$\sigma_{zz} = \nu (\sigma_{xx} + \sigma_{yy}) \tag{4.13}$$

$$\sigma_{xz} = \sigma_{yz} = 0$$

where ν is the Poisson's ratio. In the calculation reported here a value of 0.3 is used.

2. Elasto-Perfectly Plastic Case

In order to find the stresses in the plastically deforming region and the cumulative plastic deformation after every passage of the asperity, it is necessary to trace the loading cycle of each point as it passes under the asperity. During plastic deformation it is assumed that the total strains remain identical with the elastic strains found from Eqs. (4.1) through (4.13) by applying Hooke's law. (The reasons behind this assumption will be discussed later.) The stresses at each point in the plastic region are then found by employing the incremental Prandtl-Reuss equations for an elasto-perfectly plastic material.

During plastic deformation the von Mises flow rule requires that the second invarient of the stress deviators, J_2', remains constant and equal to k^2, where k is the yield stress in shear. J_2 is defined as:

$$J_2 = \frac{1}{2} (\sigma_{ij}' \sigma_{ij}') \tag{4.14}$$

We may now proceed to find the stress-strain relations in the plastic region. The total strain rates, $\dot{\varepsilon}_{ij}^{p}$, are the summation of the elastic strain rates, $\dot{\varepsilon}_{ij}^{e}$, and the plastic strain rates, $\dot{\varepsilon}_{ij}^{p}$, or,

$$\dot{\varepsilon}_{ij} = \dot{\varepsilon}_{ij}^{e} + \dot{\varepsilon}_{ij}^{p} \tag{4.15}$$

The plastic strain rates can be found from the Prandtl-Reuss equations:

$$\dot{\varepsilon}_{ij}{}^{p} = \frac{\dot{W}^{p}}{2k^{2}} \sigma_{ij}' \tag{4.16}$$

where W^p is the plastic strain energy rate,

$$\dot{W}^{p} = \dot{\varepsilon}_{ij}{}^{p} \sigma_{ij}' = \bar{\sigma}\, \dot{\bar{\varepsilon}}^{p} \tag{4.17}$$

By substituting (4.13) and (4.16) in (4.15) we obtain,

$$\dot{\varepsilon}_{ij}' = \frac{\dot{\sigma}_{ij}'}{2G} + \frac{\dot{W}^{p}}{2k^{2}} \sigma_{ij}' \tag{4.18}$$

which can be rewritten as,

$$\dot{\sigma}_{ij}' = 2G \left(\dot{\varepsilon}_{ij}' - \frac{\dot{W}^{p}}{2k^{2}} \sigma_{ij}'\right) \tag{4.19}$$

Since the plastic strain energy is much larger than the elastic strain energy,

$$\dot{W}^{p} = \dot{W} \tag{4.20}$$

where W is the total strain energy rate. Therefore,

$$\dot{\sigma}_{ij}' \simeq 2G \left(\dot{\varepsilon}_{ij}' - \frac{W}{2k^{2}} \sigma_{ij}'\right) \tag{4.21}$$

where ε_{ij}, the total strain rate, was assumed to be identical with the elastic strains given by applying Hooke's law to (4.1) through (4.13). These relations apply during plastic deformation, as long as J_2 equals k^2 and W is positive. Otherwise, the elastic equations are used to find stresses.

In order to integrate (4.21), the time rates of change can be transformed to gradients with respect to x as follows:

$$\frac{d}{dt} (\sigma_{ij}'; \varepsilon_{ij}'; W) = U \frac{\partial}{\partial x} (\sigma_{ij}'; \varepsilon_{ij}'; W) \tag{4.22}$$

where U is the steady sliding speed of the lower plane. Therefore, by substituting (4.22) in (4.21) the time rates of change will be replaced with x-derivatives and consequently U is cancelled out from the equations.

Since the strains are assumed to be known (i.e., equal to the elastic strains), Eq. (4.21) gives the gradients with respect to x of the stress deviators. The gradients are then integrated by using Runge-Kutta-Gill method to calculate the stresses[6]. The step-by-step numerical anaylsis for the stress cycle of a point at any depth y is as follows:

> Starting from the elastic stress state at a position x_o where the stress at the point first satisfies the von Mises yield condition, on the entry side, the stress rates of change with x are found from Eq. (4.21). These stress rates are then used to predict the values of the stress components when the point has moved a small distance dx. Therefore, the state of stress of points at a constant depth is obtained as the point goes through the stress cycle.

The stresses which are found by the above procedure are inexact only to the extent that they do not satisfy the condition of equilibrium, since no attempt was made to maintain equilibrium during the loading cycle. The assumption of using strains identical to the elastic strains satisfies the compatibility condition. The stress boundary conditions (i.e., traction free surface outside the contact) were also satisfied. Therefore, the solution is only an approximation to the exact solution. However, it is possible to restore the condition of equilibrium at the end of the loading cycle, as discussed in the next subsection.

3. Residual Stresses and Strains

Had the loading been entirely elastic, the stresses would approach zero when the contact point approached $x = \infty$. However, as a result of plastic deformation each point must have a state of residual stress. The periodicity of the problem requires that residual stresses and strains be independent of x; i.e.,

$$
\begin{aligned}
(\sigma_{xx})_r &= f_1(y) \\
(\sigma_{yy})_r &= f_2(y) \\
(\sigma_{zz})_r &= \nu\ (f_1 + f_2) \\
(\sigma_{xy})_r &= f_3(y)
\end{aligned}
\qquad (4.23)
$$

However, the equilibrium equations

$$\frac{\partial (\sigma_{xx})_r}{\partial x} + \frac{\partial (\sigma_{xy})_r}{\partial y} = 0$$

$$\frac{\partial (\sigma_{xy})_r}{\partial x} + \frac{\partial (\sigma_{yy})_r}{\partial y} = 0 \qquad (4.24)$$

require that

$$\frac{df_3}{dy} = 0 \qquad \text{or} \qquad f_3 = c_1$$

$$\frac{df_2}{dy} = 0 \qquad \text{or} \qquad f_2 = c_2 \qquad (4.25)$$

The boundary conditions of $\sigma_{yy} = 0$ and $\sigma_{xy} = 0$ at $x = \infty$ and $y = 0$, then give $f_2 = f_3 = 0$ at $y = 0$ or $c_1 = c_2 = 0$. Finally, the only possible state of residual stress is

$$(\sigma_{xx})_r = f(y)$$

$$(\sigma_{yy})_r = 0$$

$$(\sigma_{zz})_r = \nu f(y) \qquad (4.26)$$

$$(\sigma_{xy})_r = 0$$

The numerical results of the preceeding subsection gives nonzero "pseudo-residual stresses" $(\sigma_{yy})_r$ and $(\sigma_{xy})'_r$ which violates equilibriium and stress boundary conditions. These conditions can be satisfied by permitting these stresses to relax elastically, which results in a state of residual strains. Carrying out this procedure, the residual strains at the end of a loading cycle are

$$(\varepsilon_{yy})_r = -\frac{(1 - 2\nu)}{2(1 - \nu)G} (\sigma_{yy})_r' \qquad (4.27)$$

$$(\gamma_{xy})_r = -\frac{(\sigma_{xy})_r'}{G} \qquad (4.28)$$

where $(\sigma_{yy})'_r$ and $(\sigma_{xy})'_r$ are the pseudo-residual stresses and $(\gamma_{xy})_r$ is the engineering shear strain. Furthermore, the residual stresses $(\sigma_{xx})_r$ and $(\sigma_{zz})_r$ become

$$(\sigma_{xx})_r = (\sigma_{xx})'_r - \frac{\nu}{(1-\nu)} (\sigma_{yy})'_r \tag{4.29}$$

$$(\sigma_{zz})_r = (\sigma_{zz})'_r - \frac{\nu}{(1-\nu)} (\sigma_{yy})'_r \tag{4.30}$$

Using the residual stresses, $(\sigma_{xx})_r$ and $(\sigma_{zz})_r$, as initial conditions in a repeated numerical integration of the last subsection, new values of residual stresses are calculated. This procedure is repeated until there is no further change in $(\sigma_{xx})_r$ and $(\sigma_{zz})_r$ (which corresponds to $(\sigma_{yy})'_r$ approaching zero). This condition is satisfied after 5 to 10 integration cycles. However, the residual shear strain per pass $(\gamma_{xy})_r$ approaches a constant value. Therefore, the steady state residual stress corresponds to the physical steady state of residual stresses and the residual shear strain corresponds to the physical steady state increment of shear strain for each passage.

4. The Procedure for Numerical Calculations

The preceeding method was programmed in DGC FORTRAN IV language and a Nova 2 minicomputer (which uses 32 bits for real numbers) and was used to perform the computations. The procedure is shown by a flow chart in Figure 4.3. The final state of stress and the residual stresses and strains calculated by the above procedure is an approximate solution for the case of steady state sliding (i.e., when the residual stress $(\sigma_{xx})_r$ and $(\sigma_{zz})_r$ become constant and the pseudo-residual stress $(\sigma_{yy})'_r$ becomes zero). Merwin and Johnson considered the integration cycles before steady state is reached to be the actual transient solution, but this is not necessarily correct since it is only after the last integration cycle that $(\sigma_{yy})'_r$ approaches zero.

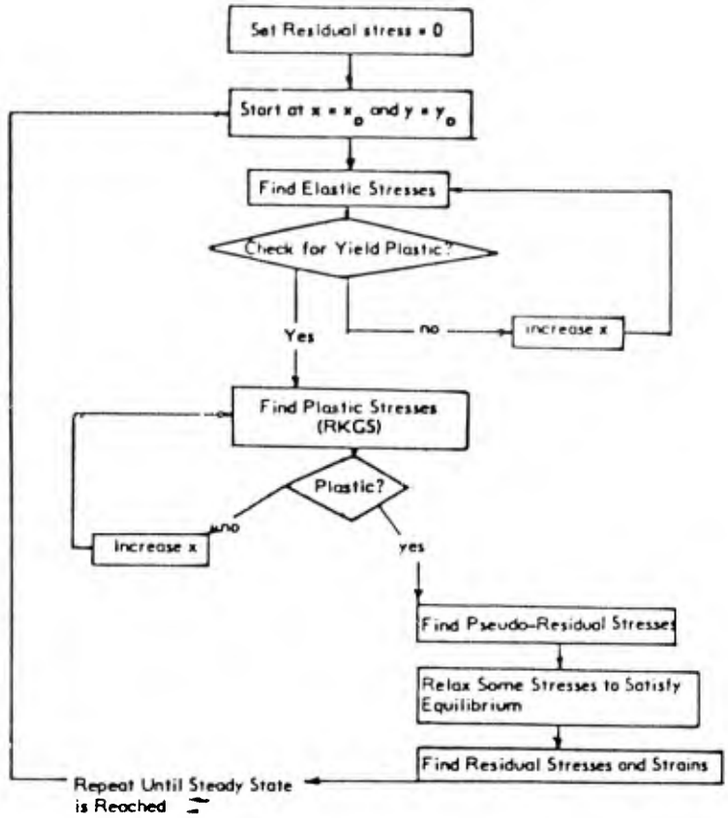


FIGURE 4.3 THE FLOW CHART FOR NUMERICAL CALCULATION OF THE STATE OF STRESS AND RESIDUAL STRESSES AND STRAINS UNDER A SLIDING CONTACT.

The assumption of allowing the total strains during plastic deformation to be identical with the elastic strains is reasonable for low friction coefficients (i.e., less than 0.5) since the region which continuously deforms plastically is contained by an elastic region around it. Therefore, the total strains cannot be much larger than the elastic strains. For larger friction coefficients, a boundary of the plastic region is at the surface, which may cause larger strain near the surface. Therefore, the solution becomes less exact at larger friction coefficients and near the free surface. However, it should still be reasonably deep below the surface near the elastic-plastic boundary.

The approximate solution found from the preceeding method has two types of instabilities. The stresses at the plastically deformed surface cannot be obtained due to the singularity of the elastic strain gradients at x = a. The

solution also becomes unstable in that the residual stresses do not converge to steady state values very near the steady state elastic-plastic boundaries for low friction coefficients (lower than 0.5). However, there is no problem at a small distance from the elastic plastic boundary inside the plastic region.

5. Discussion of the Numerical Solutions

The Merwin and Johnson method was used to find the size of the plastically deforming zone under a contact, the state of stress, the residual stresses, and the residual strains during steady state sliding. The result for the applied normal stress $P_o = 4k$ and different tangential stresses ranging from $q_o = 0$ to $q_o = 4k$ is given in Figure 4.4. It should be noted that for zero friction, a state of shakedown is reached and the steady state deformation is purely elastic. The size of the plastic region increases with increasing friction coefficient. For friction coefficients smaller than 0.5, the plastic region is below the surface, whereas, at larger friction coefficients the plastic region extends to the surface. The large size of the plastic zone in front of the slider for large friction coefficient is surprising. Perhaps this is due to the fact that plowing was assumed for the stress boundary condition, but the displacements for plowing (i.e., raised material in front of the slider) were not considered. If the solution would allow a raised surface in front of the contact, stresses would be relieved below the surface in front of the contact, and the size of the plastic zone would decrease.

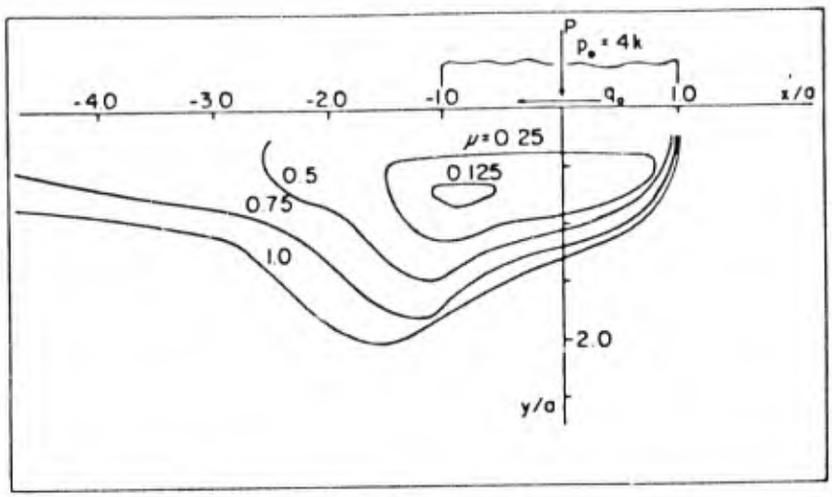


FIGURE 4.4 THE STEADY STATE PLASTIC DEFORMATION REGIONS IN AN ELASTIC-PERFECTLY PLASTIC MATERIAL UNDER A SLIDING CONTACT, FOR A MAXIMUM APPLIED NORMAL STRESS p_o = 4k AND DIFFERENT FRICTION COEFFICIENTS.

The steady state σ_{xx}, σ_{yy} and σ_{xy} components of stress at various depths are given in Figures 4.5 to 4.7. (The coordinates are described in Figure 4.2.) It is noted that σ_{xx} is always compressive whereas σ_{yy} becomes tensile behind the contact and close to the surface. This tensile zone has also been found in obtaining a similar solution by a finite element method. The distribution of steady state residual stress $(\sigma_{xx})_r$ is given in Figure 4.8 for different friction coefficients. (As discussed earlier, the only possible residual stresses are $(\sigma_{xx})_r$ and $(\sigma_{zz})_r$.) Figure 4.8 shows that the residual stress is compressive and its magnitude, for large friction coefficients, is largest near the surface.

The steady state increment of shear strain, $(\gamma_{xy})_r$, per each passage of the asperity, is given in Figure 4.9. It is observed that the increment of shear strain is largest at the surface, for large friction coefficients, and its magnitude increases with increases in the friction coefficient. It should be noted that during steady state sliding, the only nonzero plastic strain is the shear strain which accumulates with an amount $(\gamma_{xy})_r$ after each passage of the slider.

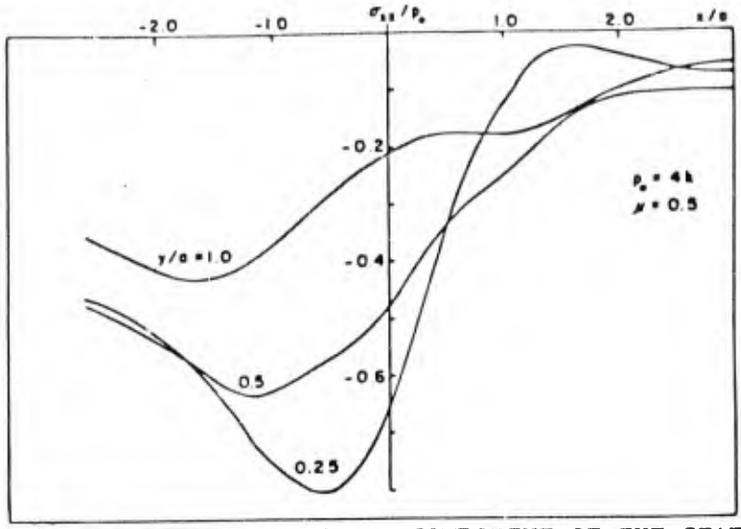


FIGURE 4.5 THE STEADY STATE σ_{xx} COMPONENT OF THE STATE OF STRESS AT DIFFERENT DEPTHS, NORMALIZED WITH RESPECT TO THE MAXIMUM APPLIED NORMAL STRESS p_o = 4k; FOR FRICTION COEFFICIENT μ = 0.5.

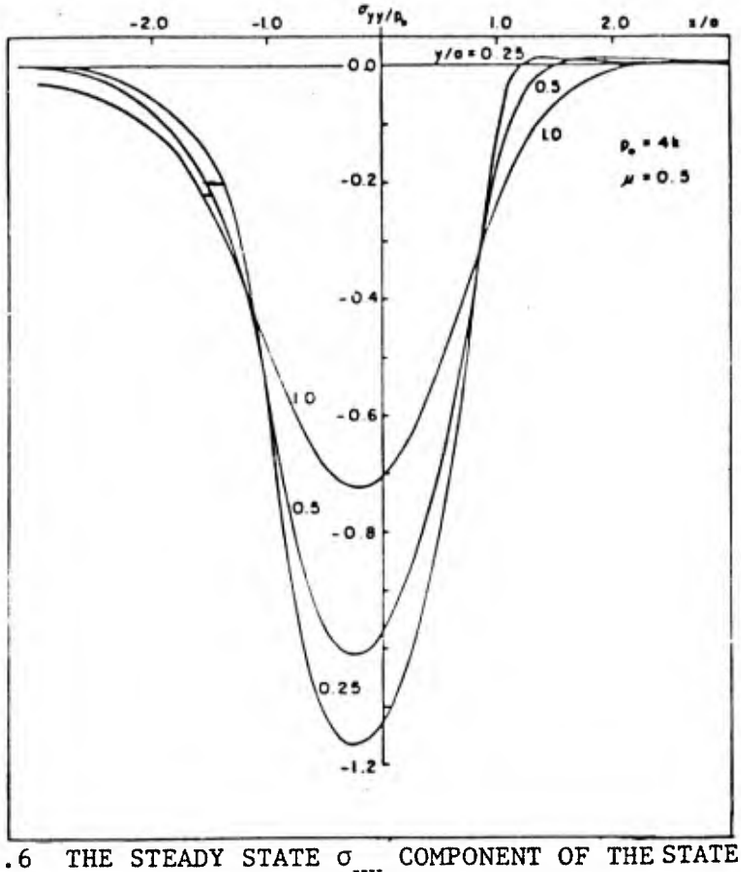


FIGURE 4.6 THE STEADY STATE σ_{yy} COMPONENT OF THE STATE OF STRESS AT DIFFERENT DEPTHS, NORMALIZED WITH RESPECT TO THE MAXIMUM APPLIED NORMAL STRESS p_o = 4k; FOR FRICTION COEFFICIENT μ = 0.5.

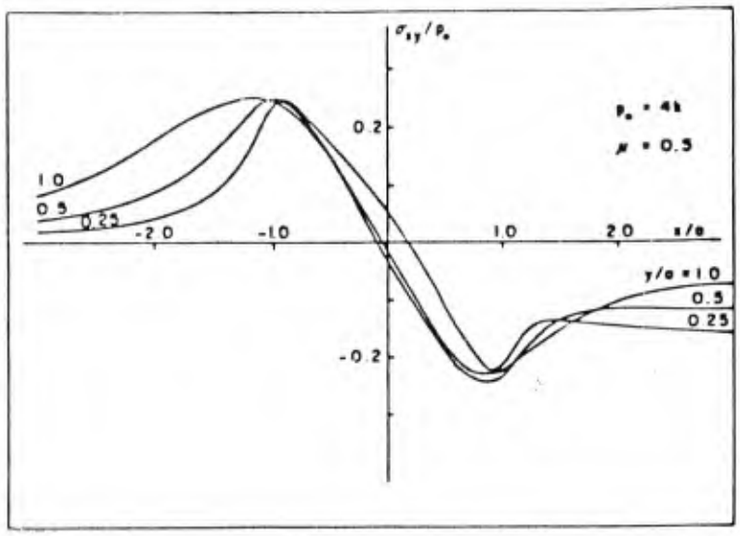


FIGURE 4.7 THE STEADY STATE σ_{xy} COMPONENT OF THE STATE OF STRESS AT DIFFERENT DEPTHS, NORMALIZED WITH RESPECT TO THE MAXIMUM APPLIED NORMAL STRESS $p_o = 4k$; FOR FRICTION COEFFICIENT $\mu = 0.5$.

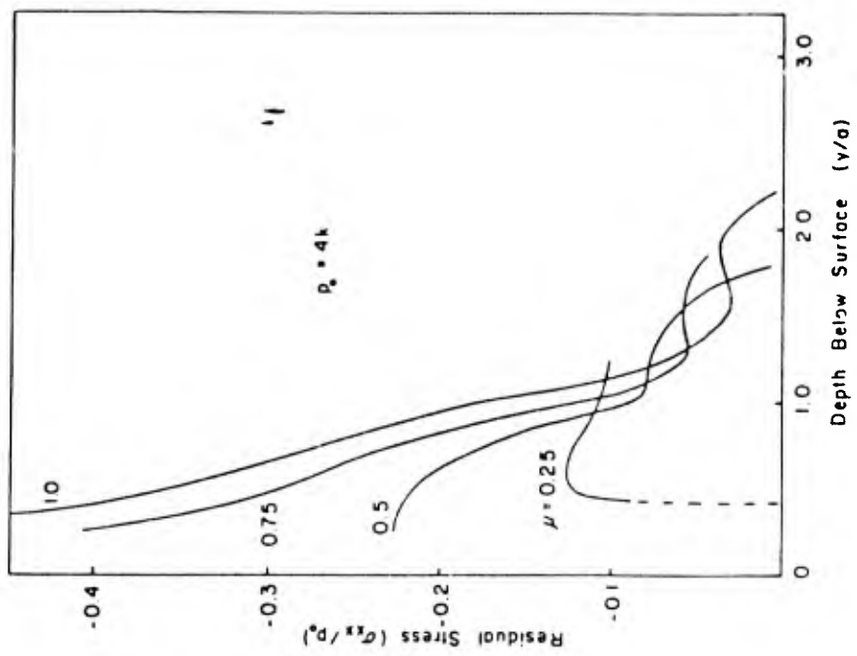


FIGURE 4.8 THE STEADY STATE RESIDUAL STRESS σ_{xx} FOR DIFFERENT FRICTION COEFFICIENTS, NORMALIZED WITH RESPECT TO THE MAXIMUM APPLIED NORMAL STRESS $p_o = 4k$.

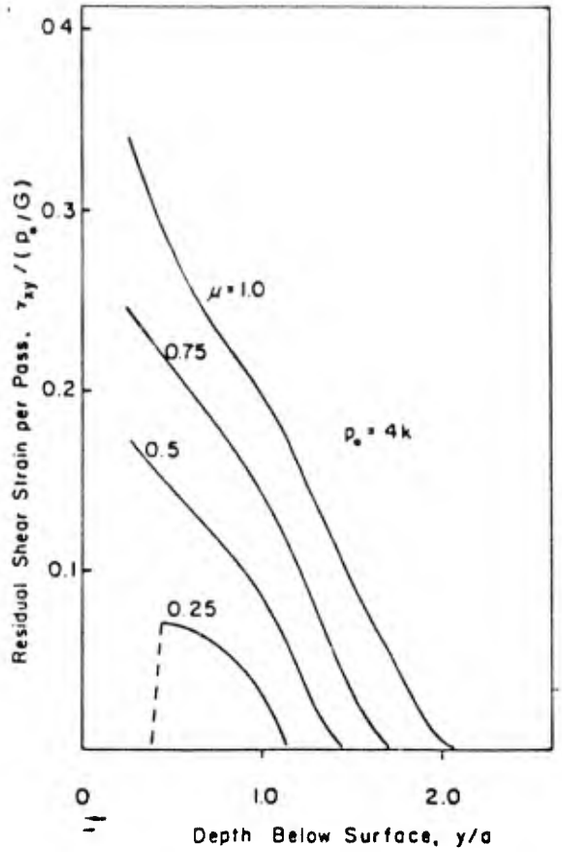


FIGURE 4.9 THE STEADY STATE RESIDUAL SHEAR STRAIN PER PASS, FOR DIFFERENT FRICTION COEFFICIENTS, NORMALIZED WITH RESPECT TO THE YIELD STRAIN IN PURE SHEAR AND THE MAXIMUM APPLIED NORMAL STRESS p_o = 4k.

4.2.3 Results of the Analysis Using FEM

The deformation of an elasto-plastic solid was also modeled and solved by using a finite element method[7]. The deformation of a semi-infinite, slightly work-hardening elasto-plastic solid was loaded by a moving asperity. The normal load was assumed to be 4k and the coefficient of friction 0.25. The material properties used were as follows: isotropic, slightly work-hardening (slope $d\sigma/d\varepsilon$ of the work-hardening region = 10^{-4} E where E is Young's modulus), E = 1.96 x 10^5 MPa = 2 x 10^4 kg/mm^2, ν = 0.28, $\sigma_y = \sqrt{3k}$ = 424 MPa = 43.3 kg/mm^2.

Figure 4.10 shows the plastically deformed region under a moving asperity during the first and fourth cycle, respectively. Under the moving load the material just in front of the load deforms plastically and part of the plastically deformed zone is much smaller than that under static load given in Figure 4.11. Furthermore, the repeated cyclic loading makes the plastic region smaller as shown in Figure 4.10(b). According to the numerical investigation of a rolling contact by Anand, the steady state deformation, after a few revolutions of a disk, would eventually reach a purely elastic state for a strain-hardening material, whereas elastic-perfectly plastic solids have elasto-plastic deformation at even high levels[8]. Therefore, the case shown in the figure may eventually reach a purely elastic state since it was assumed that the material is work-hardening.

The residual stress $(\sigma_{xx})_r$ is given in Figure 4.12 as a function of depth from the surface. It can be seen that the difference in $(\sigma_{xx})_r$ is almost negligible between third and fourth cycles. Also, the dominant residual stress is compressive in a direction parallel to the surface. The other components of residual stresses, $(\sigma_{yy})_r$ and $(\sigma_{xy})_r$ are very small. These two are the only components that can affect crack opening under sliding conditions. Therefore, it can be concluded that residual stresses do not affect crack propagation.

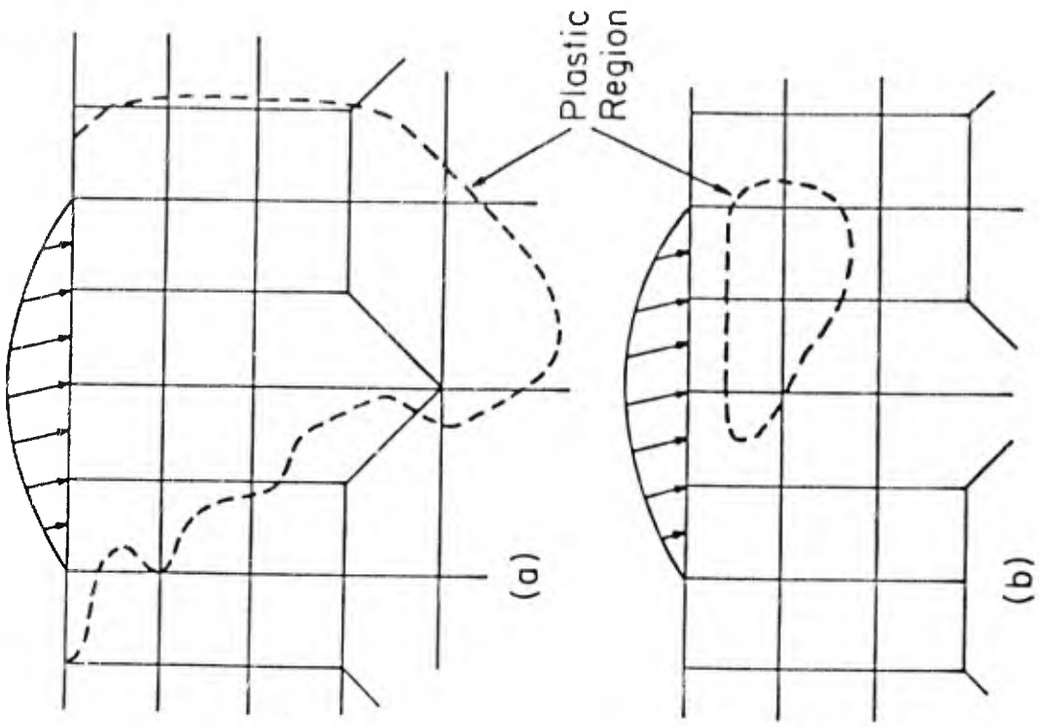


FIGURE 4.10 PLASTICALLY DEFORMED REGION UNDER A MOVING ASPERITY: (A) DURING THE FIRST CYCLE: (B) DURING THE FOURTH CYCLE.

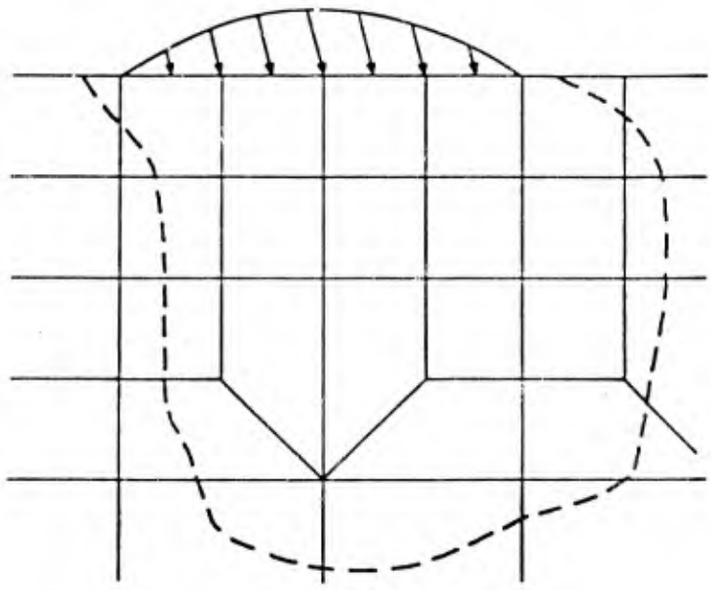


FIGURE 4.11 PLASTIC ZONE UNDER A STATIONARY ASPERITY.

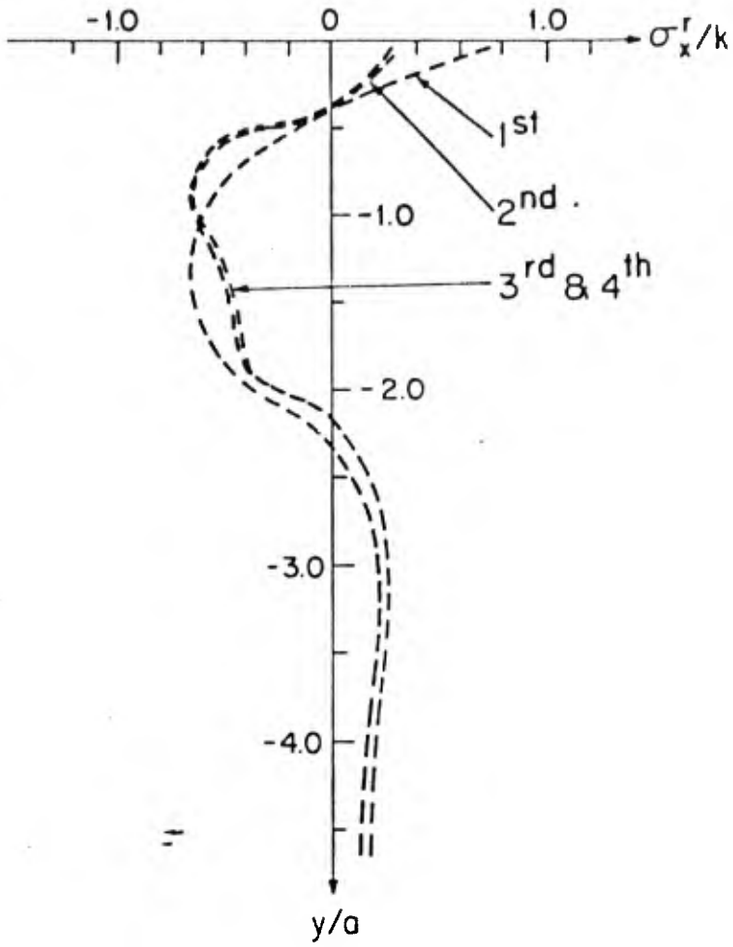


FIGURE 4.12 VARIATION OF RESIDUAL STRESS σ_x^r AS A FUNCTION OF DEPTH UNDER A MOVING ASPERITY.

These FEM results are in good agreement with the results obtained by Jahanmir and Suh using the technique developed by Merwin and Johnson[3, 1].

These theoretical results are also in agreement with the experimentally determined plastic strain fields which were obtained by measuring the grain deformation[9]. The plastic strain at the surface can be extremely large, especially on highly ductile copper, Figure 4.13.

4.3 Void and Crack Nucleation at the Subsurface

The failure of an initially flaw-free material proceeds in two sequential steps: (1) the initiation and nucleation of microcracks or microvoids and (2) crack propagation or void growth that leads to catastrophic failure. The first step of the failure process may not be an important process if flaws

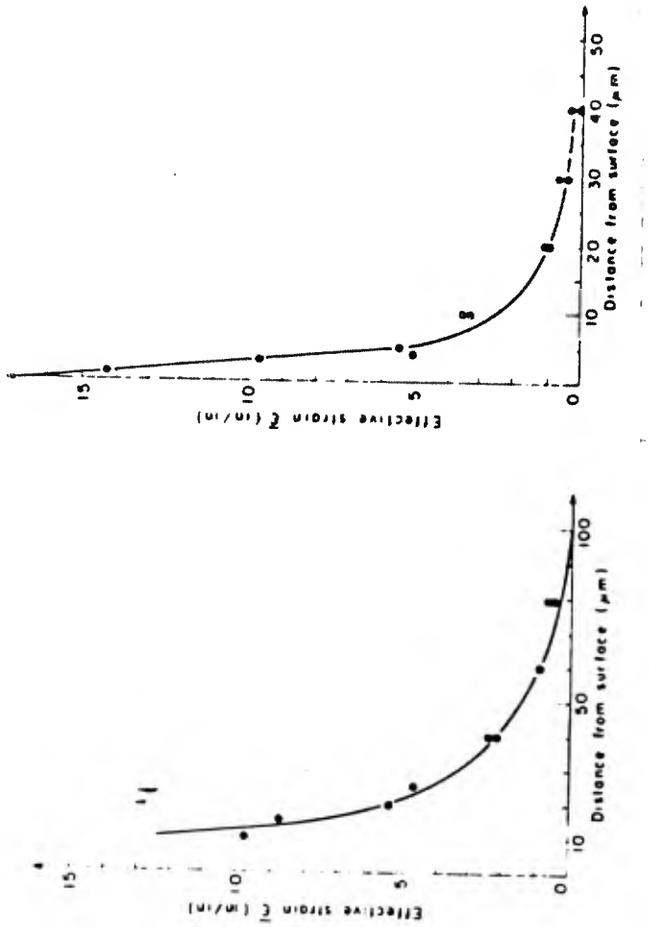


FIGURE 4.13 SUBSURFACE STRAIN VARIATION OBTAINED BY GRAIN SHAPE MEASUREMENTS: (A) COPPER SPECIMEN TESTED UNDER A NORMAL LOAD OF 2.1 kg AND AFTER A SLIDING DISTANCE OF 68 m IN ARGON; (B) AISI 1020 STEEL TESTED UNDER A NORMAL LOAD OF 2.4 kg AND A SLIDING DISTANCE OF 180 IN ARGON.

are initially present in the material. In this case, the crack propagation may control the failure process. However, in some materials, crack nucleation may determine the rate of wear rather than crack propagation. In this subsection, the mechanism of crack nucleation in a material without any flaws will be analyzed in order to understand the origin of deformation induced failure of elasto-plastic solids.

When the surface layer undergoes plastic deformation, voids and cracks can nucleate. Cracks in two phase metals are nucleated after repeated loading and consequent plastic deformation of the surface layer due to the displacement incompatibility between the hard particles and the matrix. Experimental results are shown in Figure 4.14. Even in single phase, metals cracks are present as shown in Figure 4.15. Although the exact mechanism for crack and void nucleation in single phase metals is not known, it has been suggested that interactions of dislocations and the formation of dislocations cells may be responsible [10 - 12]. It has also been suspected that even in single crystals, crack nucleation may be a result of displacement incompatibility between impurity inclusions and the matrix[10]. Intersection of twins may also create cracks[13].

The fact that the crack nucleation in two phase materials is due to the displacement incompatibility between spherical inclusions and the matrix can be illustrated as follows. Consider a hard rigid spherical particle surrounded by an elasto-plastic matrix which undergoes shear deformation. In the absence of the hard inclusion the matrix will deform in such a manner that an imaginary boundary will assume an ellipsoidal shape as shown in Figure 4.16. However, in the presence of the rigid sphere which is well bonded to the matrix, the interface cannot assume the ellipsoidal shape because of the geometric constraint imposed on the matrix by the sphere. Consequently, normal stresses are developed at the sphere-matrix interface. At location A the normal stress is tensile, while at location B is compressive. When the tensile normal stress at A exceeds the cohesive (or adhesive) strength of the bonding at the particle-matrix interface, a crack may nucleate, if the energy criterion is also satisfied.

The strength criterion may be expressed as

$$(\sigma_{kk})_{max} \geq \sigma_i \qquad (4.31)$$

where $(\sigma_{kk})_{max}$ is the maximum principal normal stress in tension and σ_i is the ideal cohesive strength at the interface. This is strictly a local criterion. When applied

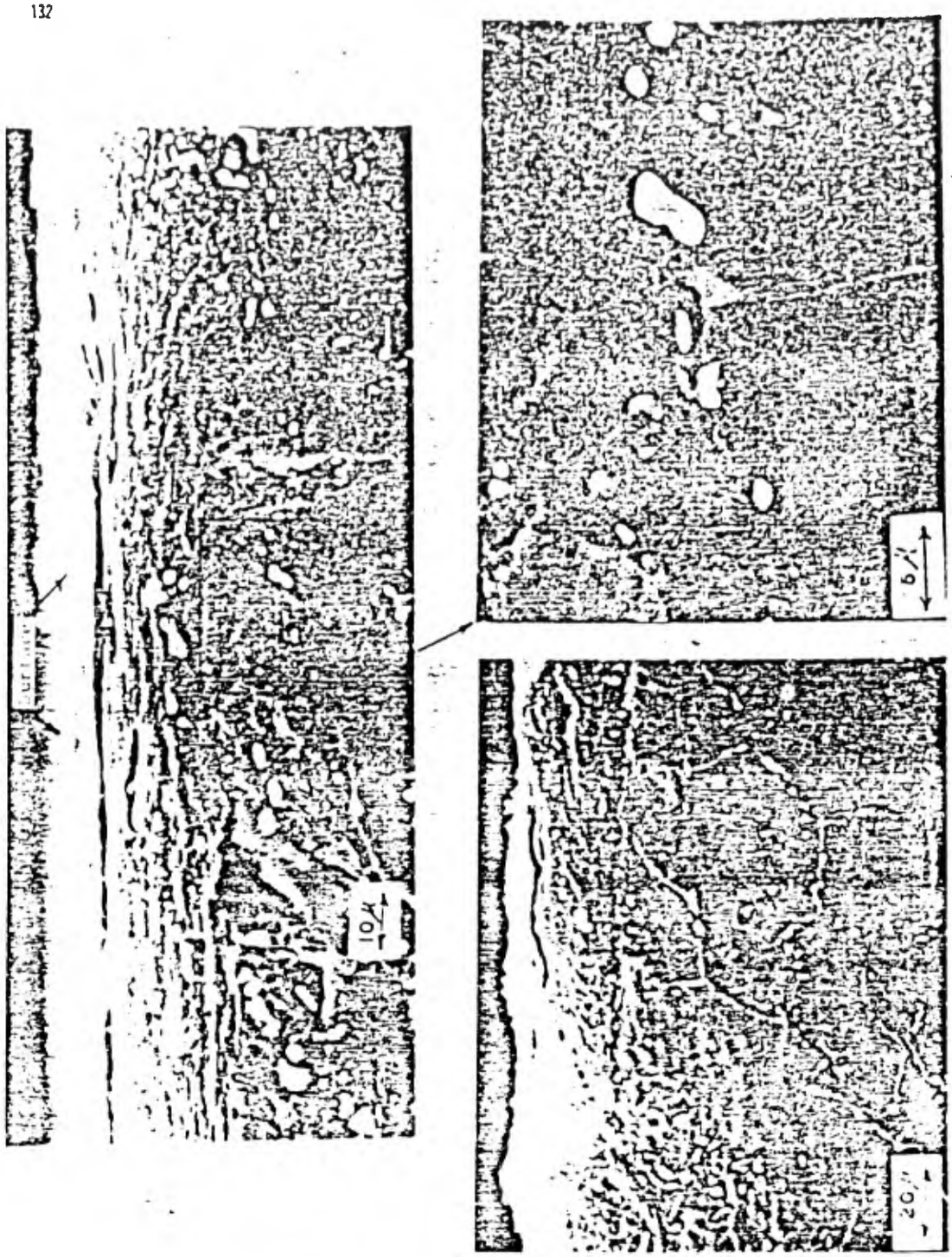


FIGURE 4.14 SUBSURFACE CRACKS AND DEFORMATION IN ANNEALED AISI 1020 STEEL.

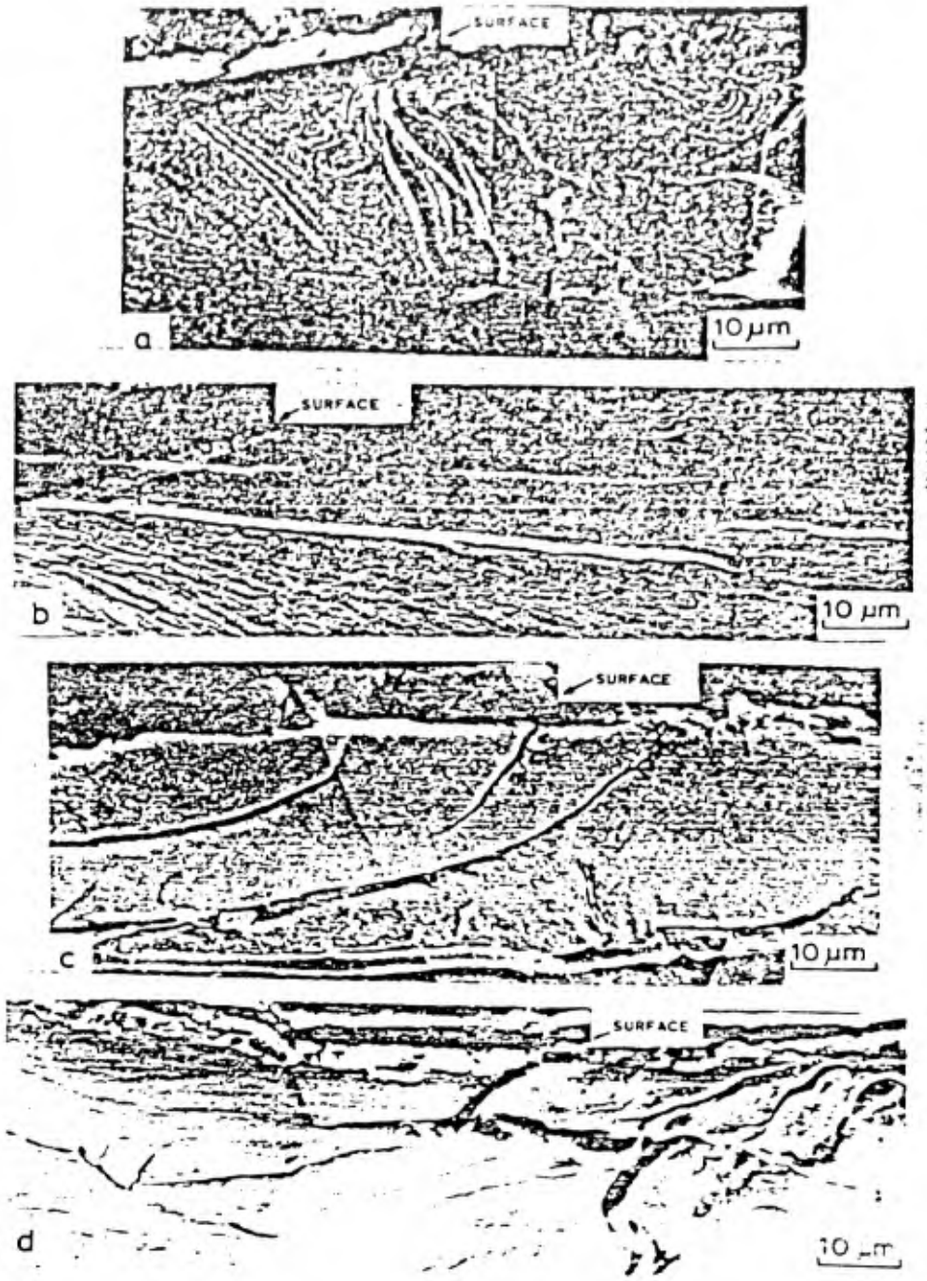


FIGURE 4.15 SCANNING ELECTRON MICROGRAPHS OF THE SUBSURFACE: (A) OFHC COPPER, (B) Cu - 5.7 at.% Sn, (C) Cu - 8.6 at.% Si and (D) Cu - 0.81 at.% Cr. THE NORMAL LOAD WAS 2 kg AT A SLIDING SPEED OF 2 m/min AND THE SLIDING DISTANCE WAS 200 m.

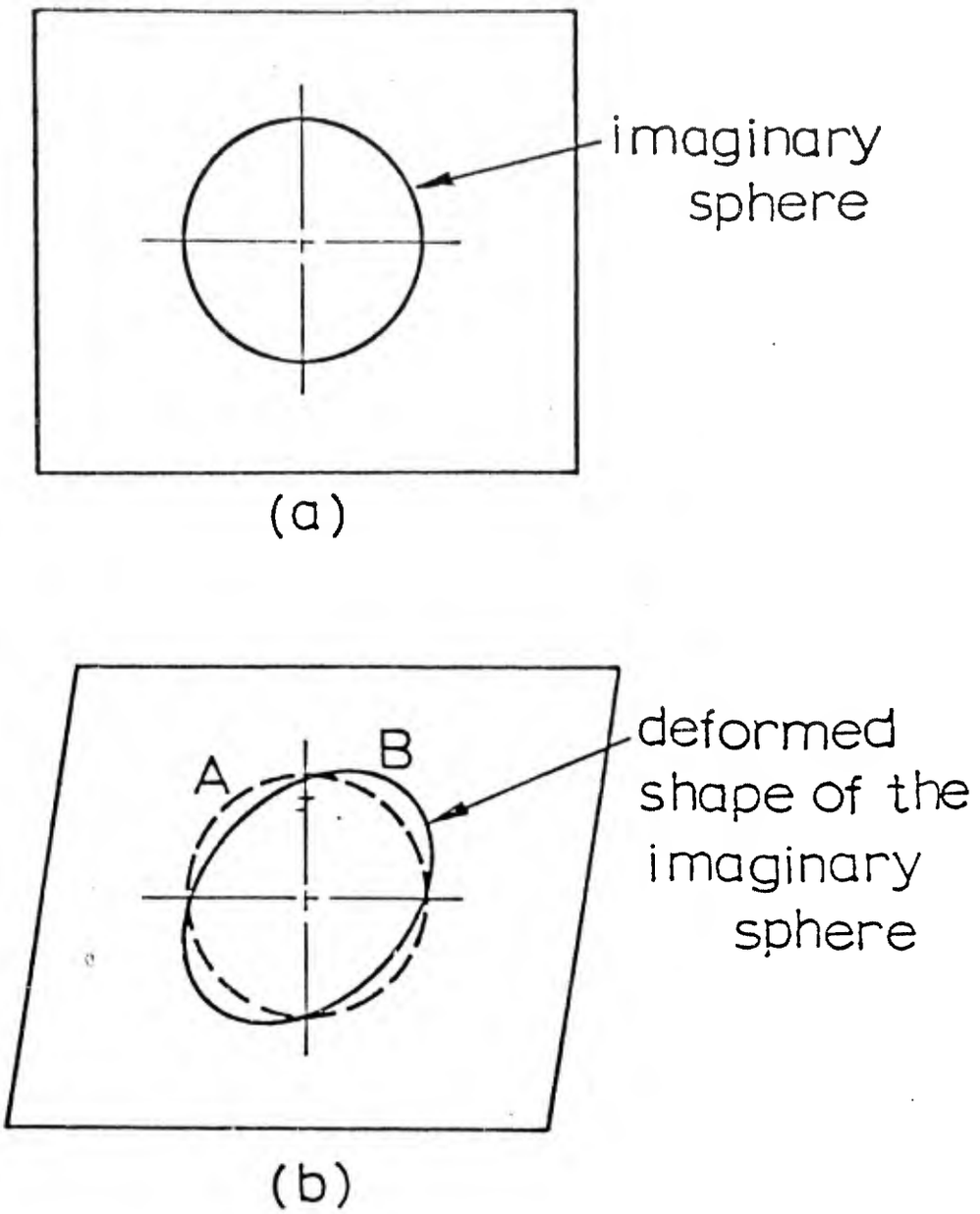


FIGURE 4.16 ILLUSTRATION OF THE DISPLACEMENT INCOMPATIBILITY BETWEEN THE MATRIX AND THE INCLUSION: (A) BEFORE DEFORMATION AND (B) AFTER DEFORMATION OF THE MATRIX.

to an inclusion filled material, the criterion itself bears no relationship to the size or shape of the inclusion. The shape of the inclusion will have to be considered in order to find the location of $(\sigma_{kk})_{max}$ and the relationship between the applied stress $(\sigma_{ij})_A$ and the interfacial stress $(\sigma_{kk})_{max}$. Once the stress concentration factor K is found, the stress criterion can be expressed in terms of the applied stress. In an elasto-plastic solid the interfacial stress concentration is limited by the flow strength of material $\sigma_y(\bar{\varepsilon}^p)$. For strain-hardening materials, the strain concentration around inclusions of different geometric shapes in inhomogeneous deformation fields are generally bound by two limiting idealizations of the plastic behavior of the material: a nonhardening rigid-plastic behavior and a linear hardening behavior with zero yield stress[14]. These results indicate that the interfacial stress, $(\sigma_{kk})_{max}$, at the surface of a cylindrical inclusion after yielding is bound by

$$\frac{3}{2}k \geq (\sigma_{kk})_{max} \geq 2k$$

where

$$(\sigma_{kk})_{max} = \sigma_{rr} - \sigma_t \tag{4.32}$$

k = flow stress in shear

σ_t = hydrostatic tensile component of the applied stress

Since the limits on $(\sigma_{kk})_{max}$ are very close to each other and nearly equal to $\sigma_y(\bar{\varepsilon}^p)$, the interfacial radial stress on the cylindrical inclusion may be taken as

$$\sigma_{rr} = \sigma_y\,(\bar{\varepsilon}^p) + \sigma_t \tag{4.33}$$

The energy criterion for void nucleation must be satisfied in addition to Eq. (4.31). When the inclusion filled material is subject to an external load, strain energy is stored in the elastic field within and around the inclusion. This strain energy will change as the elastic field changes during void nucleation. The energy in the matrix-inclusion system, E, should be sufficient to provide for the surface energy created by the void nucleation process. This may be expressed as

$$(E^s_{\text{ before}}) - (E^s_{\text{ after}}) \geq \Delta\gamma \cdot A \tag{4.34}$$

where A is the surface area of the nucleated void (or crack) and $\Delta\gamma$ is the surface energy change during void nucleation. If the void nucleates at the matrix-inclusion interface,

$$\Delta\gamma = -\gamma_{M-I} + (\gamma_M + \gamma_I) \tag{4.35}$$

where the subscripts M and I denote the matrix and inclusion, respectively, and γ the surface energy. Eq. (4.34) is inclusion size dependent, whereas the strength criterion is not. Figures 4.17 shows schematically both the strength and the energy criteria expressed in units of applied elastic strain as a function of the inclusion size. For inclusion size larger than d^*, the energy criterion is always satisfied, whenever the strength criterion is reached. However, for inclusion sizes smaller than d^*, satisfying the strength criterion does not necessarily guarantee the satisfaction of the energy requirement. Therefore, it may be argued that void nucleation will not be an energetically favorable process in this range of inclusion sizes.

In order to find σ_{rr} due to the accumulation of plastic shear strain, the solution of Eq. (4.33) is used. A point (x, y) is considered below the surface, subjected to a shear strain γ_{xy}. It is assumed that a rigid cylindrical particle is inserted at the center of a small volume element. The shear strain γ_{xy} develops an interfacial normal stress, σ_{rr}, around the particle, where σ_{rr} is a function of γ_{xy},

$$\sigma_{rr} = f(\gamma_{xy}) \tag{4.36}$$

The shear strain γ_{xy} is the total accumulated strain and it is increased by increments of $(\gamma_{xy})_r$ after each asperity pass. Maximum σ_{rr} can be developed when

$$\gamma_{xy} = \gamma_y = \frac{k}{G} \tag{4.37}$$

where γ_y is the yield strain in pure shear and G is the elastic shear modulus. The number of slider passages before maximum σ_{rr} is developed, can be found since the shear strain increment per pass, $(\gamma_{xy})_r$, is known from stress analysis. For an elasto-perfectly plastic solid, the maximum σ_{rr} due to accumulated plastic strain is equal to $\sqrt{3k}$ once the yield stress is exceeded. However, σ_{rr} depends on the position around the particle or,

$$(\sigma_{rr})_1 = \sqrt{3k}\ \mathrm{Sin}\ 2\theta \tag{4.38}$$

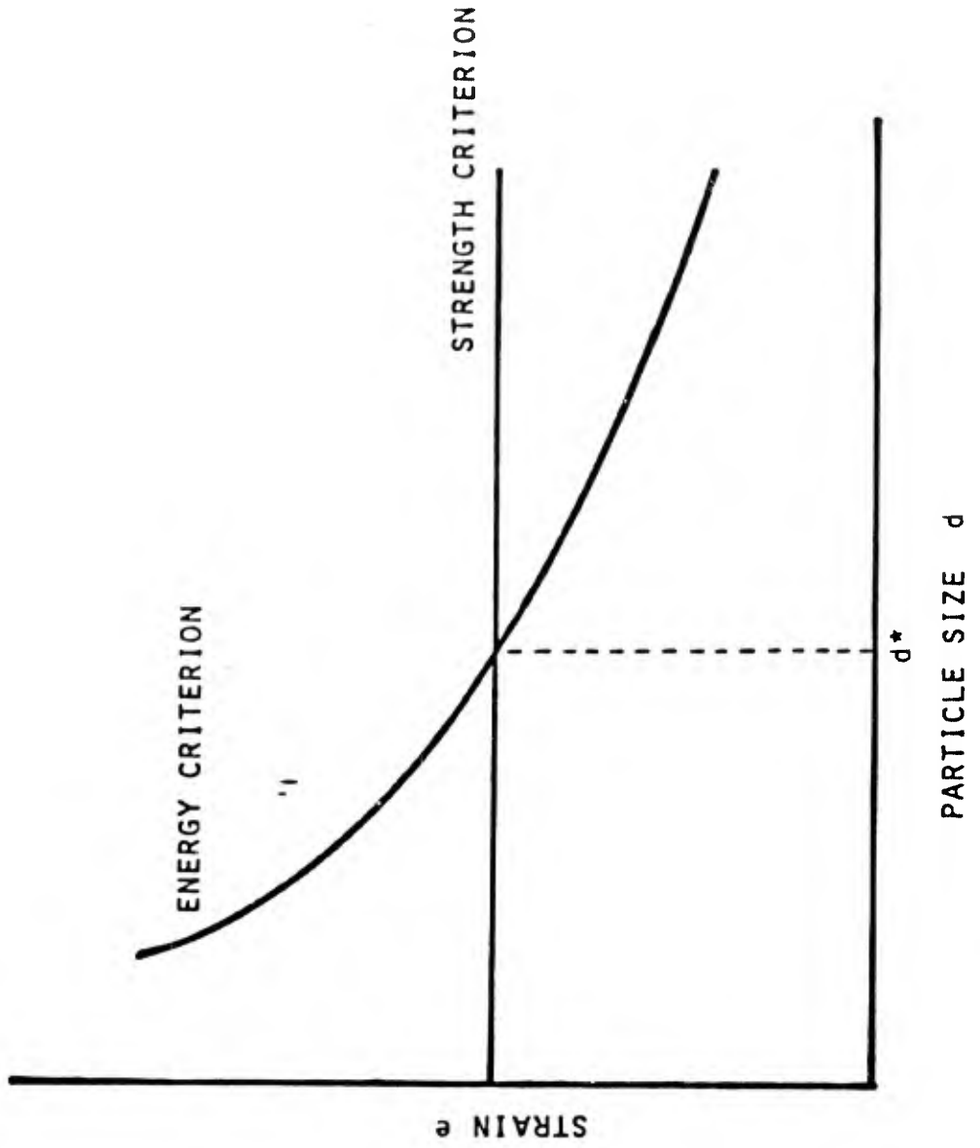


FIGURE 4.17 SCHEMATICS OF THE ENERGY AND STRESS CRITERION FOR VOID NUCLEATION.

where $(\sigma_{rr})_1$ is the normal interfacial stress due to the cumulative plastic deformation around a particle located at a depth y below the surface.

In order to find the maximum interfacial normal stress, σ_{rr}, due to the applied stress of the contact, Eq. (4.33) may be used. However, the stresses at each point must first be transformed to a state of maximum shear stress and hydrostatic stress by using Mohr's circle. Following the procedure of the last paragraph, it is assumed that the stresses act on an element and a particle is inserted at the center of the element, Figure 4.18. Therefore, using Eq. (4.33),

$$(\sigma_{rr})_2 = \sqrt{3}\tau_{max} \text{ Sin } 2(\theta - \phi) + \sigma_h \qquad (4.39)$$

where τ_{max} is the maximum shear stress at the point (x, y), θ is the angle from the x-axis to the axis of the maximum positive shear stress, and σ_h is the hydrostatic stress at (x, y); $\sqrt{3}$ appears in the equation since the von Mises yield criterion is used.

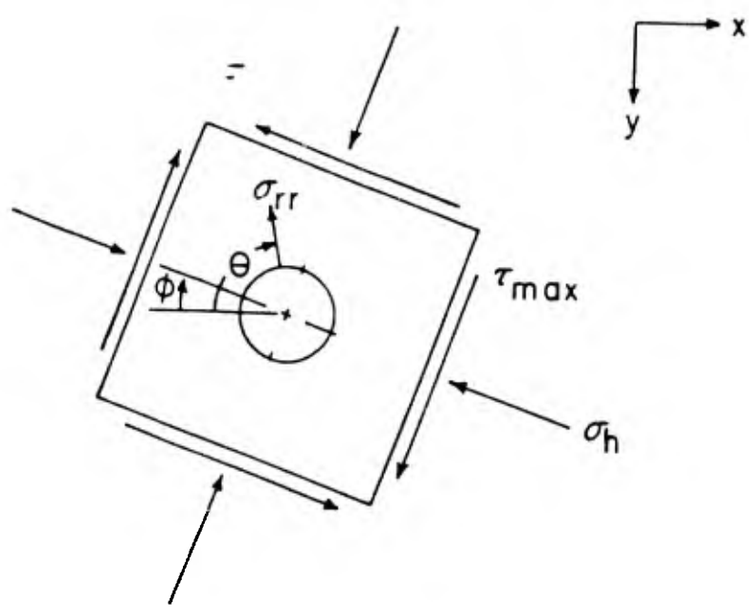


FIGURE 4.18 A RIGID CYLINDRICAL INCLUSION UNDER A GENERAL STATE OF STRESS.

The total σ_{rr} due to both the state of stress and the accumulation of plastic shear strain is found by adding Eq. (4.38) and Eq. (4.39), thus the maximum σ_{rr} which occurs at some angle $\theta = \theta_o$ can be obtained for each point (x, y)

$$(\sigma_{rr})_{max} = \sqrt{3k}\ \text{Sin}\ 2\theta_o + \sqrt{3\tau}_{max}\ \text{Sin}\ 2(\theta_o - \phi) + \sigma_h \quad (4.40)$$

The number of asperity passages before $(\sigma_{rr})_{max}$ can be developed at a given depth, can also be determined once the shear strain increment per pass is known at the respective depth.

Eq. (4.40) was solved numerically using Merwin and Johnson's method to calculate the state of stress and the residual stresses and strains. The contours of constant $(\sigma_{rr})_{max}$, found from Eq. (4.40), are plotted in Figures 4.19 to 4.21 for an applied normal stress at each asperity contact of 4k and friction coefficients of 1.0, 0.5, and 0.25. It should be noted that $(\sigma_{rr})_{max}$ is normalized with respect to k, the yield strength in shear, and all distances are normalized with respect to a, the half contact length. The figures show that $(\sigma_{rr})_{max}$ is compressive below the contact and attains its largest values in front of the slider, well below the surface.

If the particle-matrix bond strength is equal to 2k, Figures 4.22 to 4.24 show that the size of the region in which void nucleation is possible decreases with decreasing friction coefficient. The size of the void nucleation region is smaller for a stronger particle-matrix bond. It is not surprising to find that voids can only nucleate well below the surface ($\sigma_{rr} > 2k$), since the subsurface observations of wear samples in the last section showed that voids nucleate below the surface.

The minimum and the maximum depth of the void nucleation region is plottted in Figure 4.22, as a function of friction coefficient for the applied normal contact stresses, p_o and 4k and 6k. The range of depth for void nucleation increases with increasing friction coefficient for both applied normal stresses, but the voids can nucleate deeper below the surface for a larger applied normal stress. It is interesting to observe that at $p_o = 6k$ and zero friction (which applies to a case of pure rolling), void nucleation is possible in a small region. However, at $p_o = 4k$ and zero function coefficient, voids do not nucleate since the stresses shakedown to an elastic state during steady state condition.

The depth of the void nucleation region ($\sigma_{rr} \geq 2k$) is plotted in Figures 4.23 and 4.24 against the number of passes required to reach the maximum σ_{rr} at each depth, for different

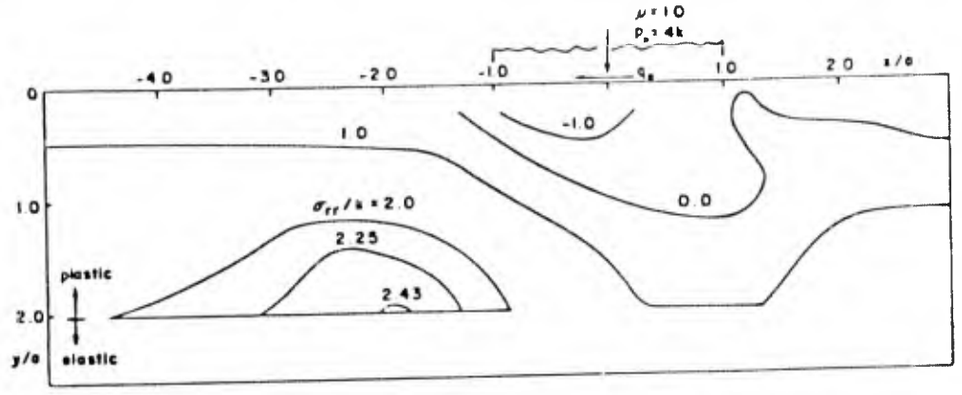


FIGURE 4.19 CONTOURS FOR σ_{rr} UNDER A SLIDING CONTACT, NORMALIZED WITH RESPECT TO k, THE SHEAR YIELD STRENGTH, FOR p_o = 4k and μ = 0.5.

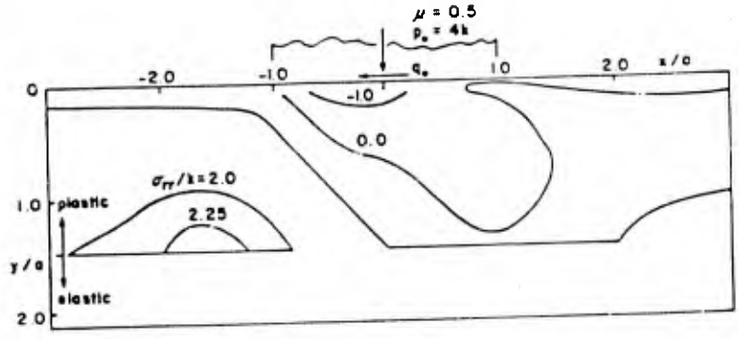


FIGURE 4.20 CONTOURS FOR σ_{rr} UNDER A SLIDING CONTACT, NORMALIZED WITH RESPECT TO k, THE SHEAR YIELD STRENGTH, FOR p_o = 4k AND μ = 0.25.

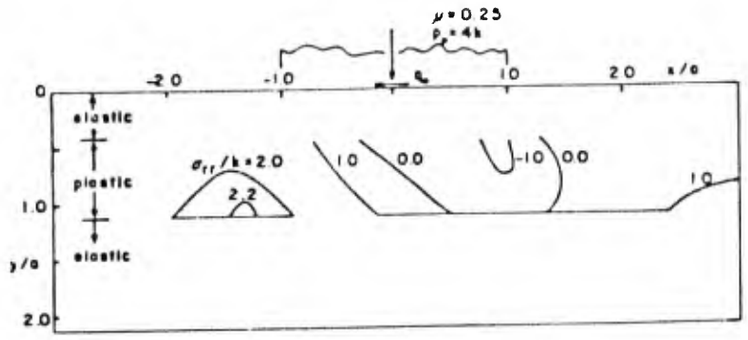


FIGURE 4.21 CONTOURS FOR σ_{rr} UNDER A SLIDING CONTACT, NORMALIZED WITH RESPECT TO k, THE SHEAR YIELD STRENGTH, FOR p_o = 4k AND μ = 0.25.

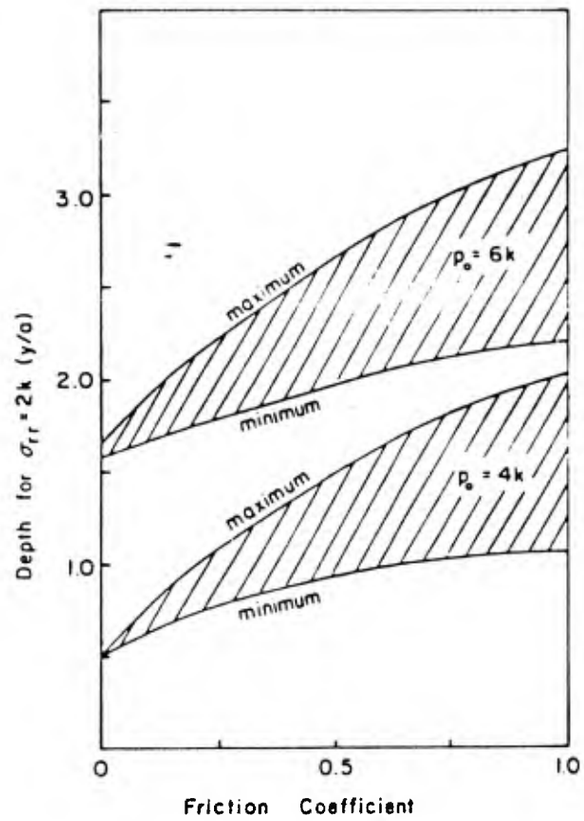


FIGURE 4.22 DEPTH OF VOID NUCLEATION REGIONS VERSUS FRICTION COEFFICIENT, FOR TWO DIFFERENT MAXIMUM APPLIED NORMAL STRESS p_o = 4k AND 6k.

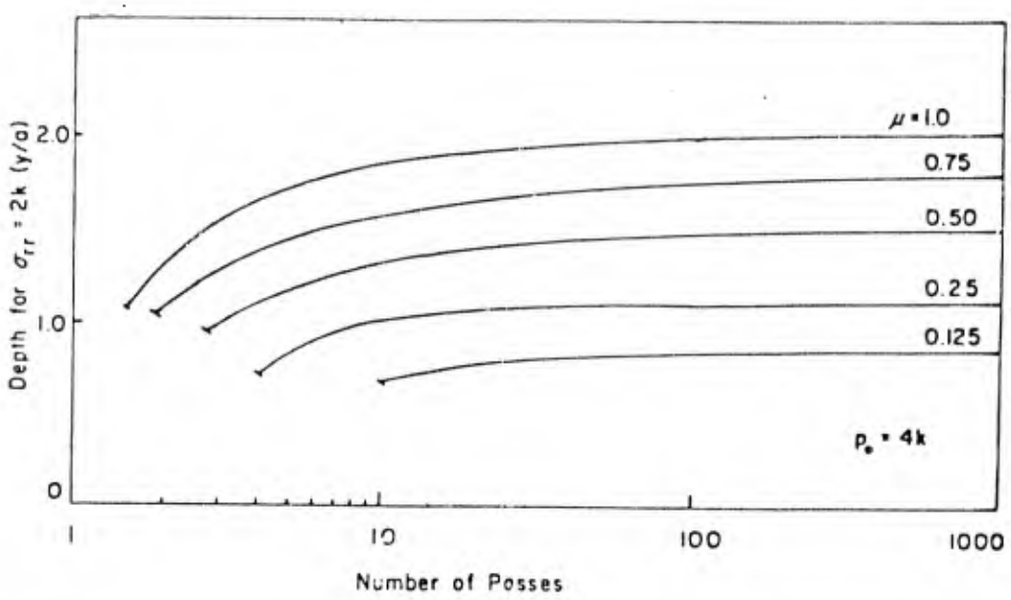


FIGURE 4.23 DEPTH OF VOID NUCLEATION REGIONS FOR DIFFERENT FRICTION COEFFICIENTS VERSUS THE NUMBER OF PASSES REQUIRED FOR VOID NUCLEATION, FOR A MAXIMUM APPLIED NORMAL STRESS, p_o = 4k

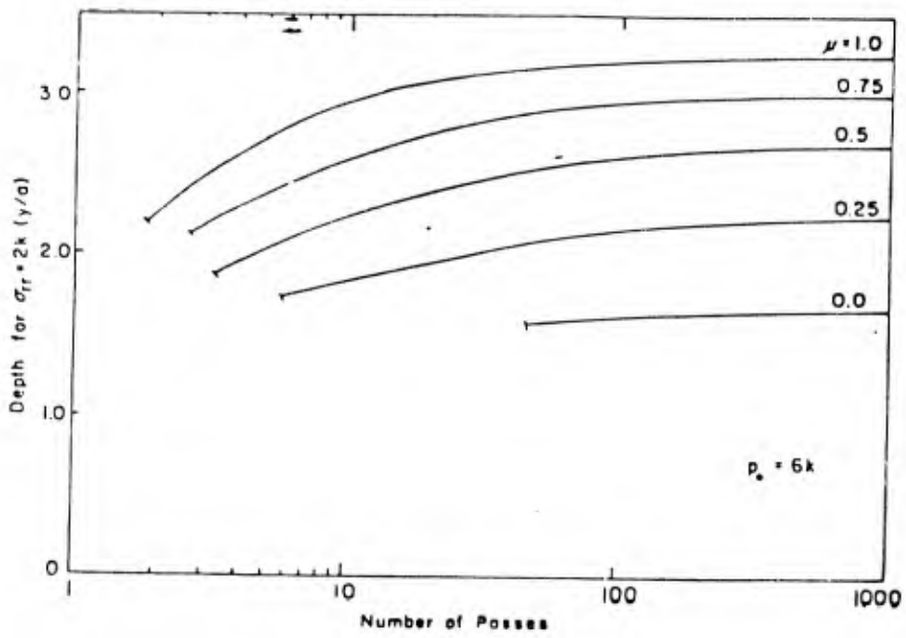


FIGURE 4.24 DEPTH OF VOID NUCLEATION REGIONS FOR DIFFERENT FRICTION COEFFICIENTS VERSUS THE NUMBER OF PASSES REQUIRED FOR VOID NUCLEATION, FOR A MAXIMUM APPLIED NORMAL STRESS p_o = 6k

friction coefficients and p_o of 4k and 6k. At a given depth, a smaller number of passes is required for void nucleation at larger friction coefficients. It should be noted that voids can nucleate at the minimum depths only after 1 to 10 passes; this implies that in many materials with hard inclusions, crack propagation may be the wear rate determining factor. It was estimated that about 10^5 asperity passes are required to create a wear sheet in AISI 1020 steel[15]. As the number of passes increases, the depth of void nucleation moves away from the surface.

It should be pointed out here that the addition of the interfacial normal stress at inclusions, due to the cumulative plastic deformation and due to the state of stress, violates the yield criterion. This violation occurs because an elastic-perfectly plastic material, σ_{rr}, cannot exceed the yield stress, i.e., $\sqrt{3k}$, and the solution may not be correct for $\sigma_{rr} \geq \sqrt{3k}$. However, if the voids nucleate when $\sigma_{rr} = \sqrt{3k}$, the stresses can relax and the yield criterion can be satisfied. The solution, however, does indicate that voids nucleate below the surface and that the depth and the number of passes depend on the friction coefficient and the applied normal load.

In the analysis of void nucleation, it was assumed that the hard particles are rigid. However, in real materials the particles have some elasticity, which would result in values of σ_{rr} smaller than the ones calculated by the above procedure. An exact analysis of the interfacial stress for elastic inclusions has been done for an elastic solid which may be adopted for an elasto-plastic solid[16, 17].

The criterion for void nucleation from large particles may be a combination of a local shear strain and a local interfacial tensile stress criterion. In the above analysis, it was assumed that the local stress criterion was sufficient. This assumption, however, may be a good approximation for equiaxial particles, but not for elongated particles. In the analysis of elongated particles, the local strain concentrations are large and void nucleation generally occurs by particle fracture. Therefore, as the particles become more elongated, the local shear strain criterion for void nucleation may become the dominating criterion.

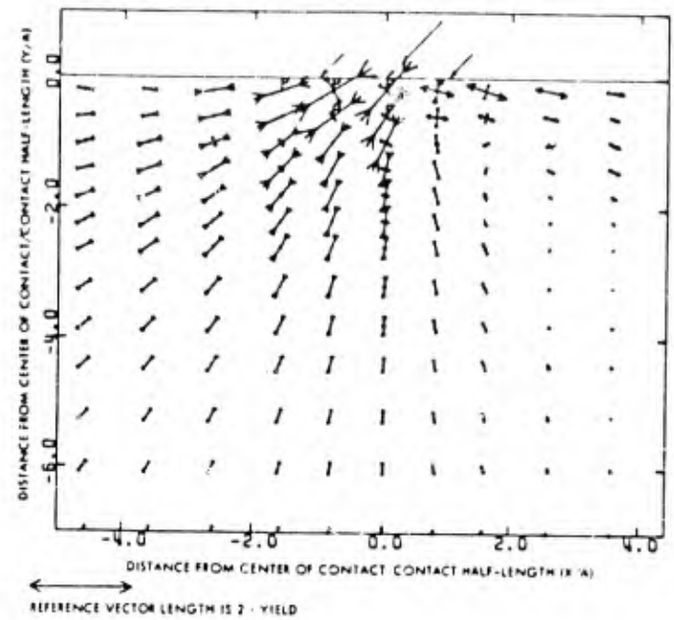
4.4 Crack Propagation Due to Surface Traction

In the preceeding subsection, crack nucleation in an elasto-plastic solid was discussed in great detail. It was shown that cracks can nucleate at subsurfaces when large plastic deformation occurs, although the applied load by the

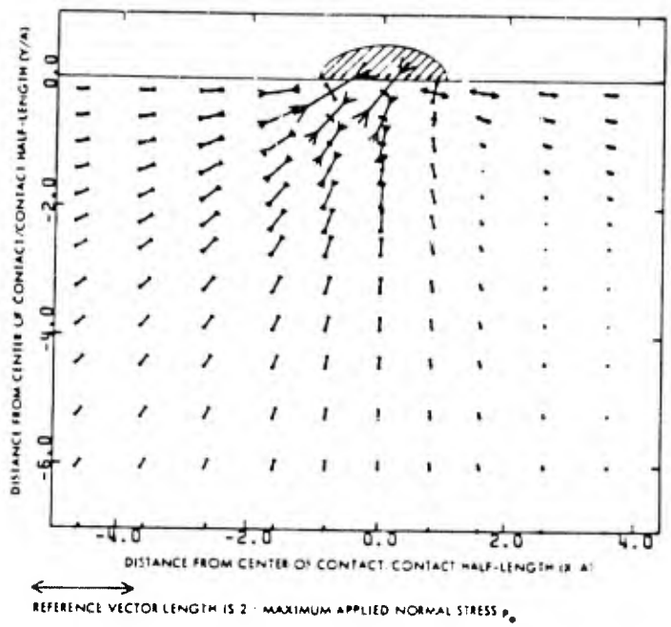
asperity, creates compressive and shear stresses. It was also briefly stated that cracks nucleate at the surface of brittle solids, especially under rolling contacts. These cracks propagate perpendicular to the surface into the subsurface since the maximum normal stress at the surface is parallel to the surface, as shown in Figure 4.25. These cracks eventually change direction due to the changing direction of the maximum normal stress and the nature of the stress field. Various cases of crack propagation were illustrated by Suh[18].

In this subsection the crack propagation in an elasto-plastic solid under sliding conditions will be analyzed. As shown earlier, cracks are nucleated within a narrow zone below the surface due to the compressive and tangential load applied by an asperity. These cracks may propagate due to the repeated loading by moving asperities at the surface. These subsurface cracks will propagate at different rates, depending on where the cracks are located. The crack that propagates the fastest will control the wear process. Although these cracks grow little by little and reach a steady state crack length, the analysis is easier to perform when cracks have grown to a finite size.

When stresses are applied at the surface of a solid with a subsurface crack as shown in Figure 4.26, part of the crack is in a compressive region while a very small section of the crack is in a tensile region. These cracks always experience combined loading. The crack tip in the tensile regime experiences a crack opening-closing mode of loading (i.e., Mode I) as well as shear loading, whereas the crack in the compression region only experiences compressive and shear loading. The material around the crack tip in the tensile region is commonly in elastic state due to the unloading of the material with the passage of the asperity.



(a)



(b)

FIGURE 4.25 THE STRESS FIELD UNDER AN ASPERITY CONTACT (COEFFICIENT OF FRICTION = 1). (LINES SHOW THE DIRECTIONS AND MAGNITUDES OF PRINCIPAL STRESSES, ARROWHEADS INDICATE SIGNS AND "p" INDICATES PLASTICITY.) (A) AN ELASTOPLASTIC SOLID (FINITE ELEMENT SOLUTION, APPLIED LOAD VECTORS NOT TO SCALE); (B) AN ELASTIC SOLID (SOLUTIONS FOR AN ELLIPTICALLY DISTRIBUTED LOAD, VECTORS NOT TO SCALE).

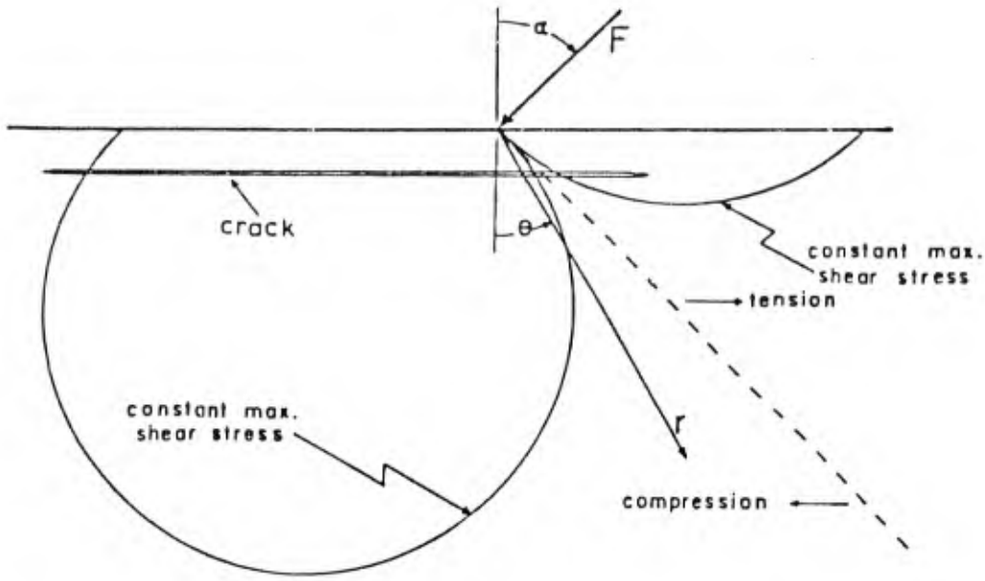


FIGURE 4.26 A POINT LOAD (LINE LOAD INTO THE PAPER, IN PLANE STRAIN) ON A HALFSPACE.

It has been well established in the fields of fracture mechanics and fatigue, that when cracks are loaded cyclically in Mode I, the crack propagation rate per cycle, dC/dN, correlates with the cyclic change in the stress intensity factor, ΔK_{II} as shown in Figure 4.27. The direction of the propagation in Mode I, correlates well with the direction of the maximum normal stress and also with the direction of minimum strain energy. When the change in the stress intensity factor per cycle is less than the threshold value, ΔK_{th}, cracks cannot propagate under Mode I loading.

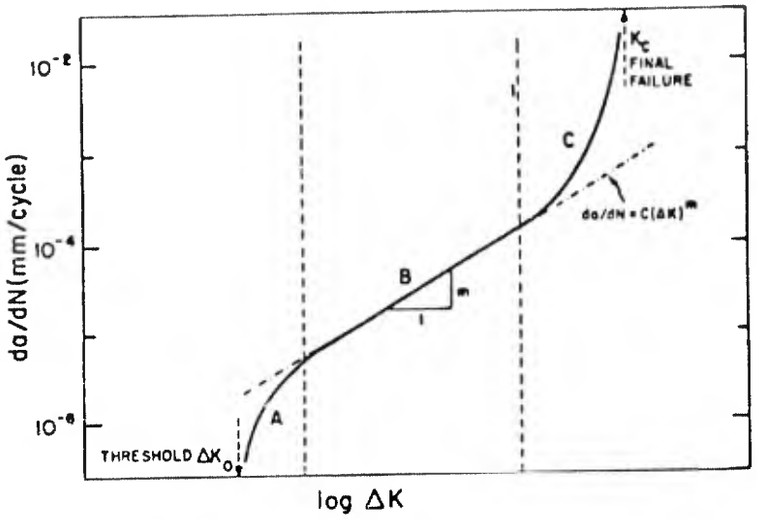


FIGURE 4.27 A TYPICAL PLOT OF THE CRACK EXTENSION PER CYCLE VS. THE LOGARITHM OF THE CHANGE IN THE STRESS INTENSITY FACTOR. THE PRIMARY MECHANISMS MAY BE DIVIDED INTO THREE REGIMES: REGIME A, NON-CONTINUUM MECHANISMS, SHOWING A LARGE INFLUENCE OF (I) MICROSTRUCTURE, (II) MEAN STRESS AND (III) ENVIRONMENT: REGIME B, CONTINUUM MECHANISMS (STRIATION GROWTH), SHOWING LITTLE INFLUENCE OF (I) MICROSTRUCTURE, (II) MEAN STRESS, (III) DILUTE ENVIRONMENT AND (IV) THICKNESS: REGIME C., "STATIC MODE" MECHANISMS (CLEAVAGE, INTERGRANULAR AND FIBROUS), SHOWING A LARGE INFLUENCE OF (I) MICROSTRUCTURE AND (II) THICKNESS BUT LITTLE INFLUENCE OF THE ENVIRONMENT.

Very little data are available on crack propagation under Mode II and Mode III loading. When there is combined loading of Mode I and Mode II (or III) type, there is very little known about the direction and the magnitude of crack propagation. It will be shown in this subsection that when the loading is both compressive and shearing, the crack propagation direction in an elasto-plastic solid is parallel to the τ_{max} direction.

4.4.1 Subsurface Crack Propagation in Linear Elastic Solids and Its Implications

For the case shown in Figure 4.23, Fleming and Suh analyzed the propagation of a subsurface crack parallel to the

surface using a linear elastic fracture mechanics (LEFM) model[19]. This treatment was based on the fact that a subsurface crack is closed in the plastic region in front of the asperity contact, while the trailing end of the crack behind the moving asperity contact is subjected to tensile stress in the elastic region. This stress intensity factor, due to an elliptically distributed load at the asperity contact, was calculated using an approximate method based on weighting factors. The stress intensity factor for Mode I, computed in this manner, was very small being in the neighborhood of Δk_{th}, while the stress intensity factor for Mode II was about an order of magnitude larger than that for Mode I. Essentially, similar results were obtained by Hills and Ashelby, Rosenfield and Keer, et al, although they also computed the stress intensity factor in the compressive zone[20, 21, 22].

Recently, Sin and Suh have caluculated the stress intensity factor using finite element method for both an elastic solid and an elasto-plastic solid[23]. The FEM results for the elastic case show that:

1) The maximum values of K_{II} occur when the crack tips lie right below the asperity contact.
2) The closer the crack is to the surface, the larger is the maximum value of K_{II}.
3) K_{II} increases with crack length.

All these values of stress intensity factors for horizontal subsurfaace cracks are quite small. This is not surprising, in view of the fact that the largest stress component, due to the surface traction exerted by the asperity contact, is the normal stress parallel to the surface. Therefore, the largest stress intensity factor for an elastic solid is expected to be associated with cracks perpendicular to the surface.

Linear elastic fracture mechanics were applied to the subsurface problems as an approximation to the real elasto-plastic solid. Whether the linear elastic fracture mechanics (LEFM) approaches are appropriate depends on the plastic zone size at the crack tip. If this size is too big or comparable with such a dimension as the distance from the crack tip to the free surface, LEFM is no longer valid and plasticity plays a significant role. In this case, the local stress and strain history become important in determining the fracture criterion.

The estimate of the plastic zone size for Mode II loading is given by

$$r_p = \frac{1}{2\pi}\left(\frac{\Delta K_{II}}{k}\right)^2 \qquad (4.41)$$

where ΔK_{II} is the stress intensity factor range in Mode II and k is the shear yield strength of the material. Tables 4.1 and 4.2 show the changes in stress intensity factor and estimated plastic zone sizes as a function of the depth of crack location for a small crack during one cycle of loading. Values for all parameters are the same as given before: k = 245 MPa = 25 kg/mm^2 and a = 10 μm. From these two tables, it can be seen that the plastic zone sizes estimated are comparable with the depth of crack and the stress intensity factor ranges calculated and are close to or less than the threshold intensity ΔK_{th}. Of course, there is no data on Mode II available. However, an approximate analysis indicates that the threshold intensity factor in Mode II is very close to that in Mode I[7].

There are two difficulties in accepting the linear elastic fracture mechanics approach to the subsurface crack problem. The cracks quite near the surface can have large values of ΔK_{II}, but they cannot propagate in a brittle manner due to the development of large plastic zones. On the other hand, cracks far below from the surface cannot grow since the stress intensity factor for these are much smaller than the threshold although the plastic zones are small and the plasticity restriction does not apply. Therefore, plasticity should come into the analysis of subsurface crack behavior under the moving asperity.

4.4.2 Crack Trajectory - Fracture Criteria in Mixed Mode

One of the major issues in studying crack propagation is how to predict the crack propagation direction. This is an appropriate topic to consider before the plasticity effects at the crack tip are further discussed.

Until recently, no suitable failure criterion could be found for the mixed mode fracture. Much of the attention has been given only to the problem of predicting the direction of crack extension when a body with cracks is simply loaded. There are two major criteria: maximum hoop stress and minimum strain energy density[24, 25]. The maximum hoop stress criterion states that crack growth will occur in a direction perpendicular to the maximum principal stress. On the other hand, the minimum strain-energy-density factor criterion postulates: (1) that the initial crack growth takes place in the direction along which the strain-energy-density factor possesses a stationary value, and (2) that crack initiation

d/a		μ : 0.25	0.5	1.0
0.3	Left	1.05	1.07	1.44
	Right	1.28	1.34	1.54
0.5	Left	1.08	1.11	1.15
	Right	1.35	1.40	1.49
1.0	Left	0.67	0.68	0.70
	Right	0.89	0.94	1.06

Note: a -- half length of the aperity contact
c -- length of the crack
d -- depth of the crack
μ -- friction coefficient

TABLE 4.1 STRESS INTENSITY FACTOR RANGE, K_{II} ($MNm^{3/2}$), AT LEFT AND RIGHT TIPS FOR A SMALL CRACK ($c = 1/4\alpha$).

d/a		μ : 0.25	0.5	1.0
0.3	Left	0.96	1.00	1.48
	Right	1.44	1.56	2.08
0.5	Left	0.76	0.80	0.88
	Right	1.20	1.28	1.48
1.0	Left	0.12	0.12	0.12
	Right	0.20	0.24	0.32

Note: Nomenclature defined in Table 4.1

TABLE 4.2 RATIO OF PLASTIC ZONE SIZE TO DEPTH OF CRACK AT LEFT AND RIGHT TIPS FOR A SMALL CRACK ($c = 1/4\alpha$).

occurs when the factor reaches a critical value. Comparison of the two criteria by several authors has shown that for tensile loading the differences between them are small. For compressive loading, however, not only do these two not agree, but neither criterion correlates well with physically observed behavior[26].

When these criteria are applied to the subsurface cracks in a compressive zone, they predict crack extension direction to be about 110° at the left (trailing) tip and about 70° at the right (leading) tip from the direction parallel to the surface, implying that crack extension occurs toward the surface at both tips. However, experimental results given in the next section show that the subsurface cracks grow parallel to the surface most of the time, before they become loose. McClintock has suggested, in the investigation of crack behavior in rail heads under rolling conditions, that cracks in a compressive field are most likely to grow in shear[27]. In fact, Forsyth has observed that fatigue cracks have two growth regimes[28]. In Stage I, cracks formed on the slip planes of the persistent slip bands, grow when they are most closely aligned with the maximum shear stress directions. In sliding wear, the slip planes tend to line up parallel to the surface maximum shear stress direction, and thus is likely to be the crack propagation direction.

In 2-dimensional deformation fields the stresses at the crack tip are expressed as[25]:

$$\sigma_{rr} = \frac{1}{2\sqrt{2\pi r}}\left[K_I \cos\frac{\theta}{2}\left(3 - \cos\theta\right) + K_{II}\sin\frac{\theta}{2}\left(3\cos\theta - 1\right)\right]$$

$$\sigma_{\theta\theta} = \frac{1}{\sqrt{2\pi r}}\cos\frac{\theta}{2}\left[K_I\cos^2\frac{\theta}{2} - \frac{3}{2}K_{II}\sin\theta\right] \qquad (4.42)$$

$$\sigma_{r\theta} = \frac{1}{2\sqrt{2\pi r}}\cos\frac{\theta}{2}\left[K_I\sin\theta + K_{II}\left(3\cos\theta - 1\right)\right]$$

and the maximum shear stress τ_{max} is given by

(4.43)

τ_{max} will have maximum values when the conditions of $\partial\tau_{max}/\partial\theta = 0$ and $\partial^2\tau_{max}/\partial\theta^2 > 0$ are satisfied. If we denote θ_m for the angle θ which satisfies the conditions and substitute into Eq. (4.42), then the stresses become $[\sigma_r]\theta_m$ $[\sigma\theta]_{\theta m}$, and $[\tau_r\theta]_{\theta m}$ respectively. Thus, using the Mohr's circle transformation, the maximum of τ_{max} and the angle between the θ_m-direction and this maximum shear direction can be determined. When the

distance r goes to zero, this direction ultimately becomes the direction of crack propagation.

If the above criterion is applied to the results obtained in the previous subsection for subsurface cracks, it predicts the angle to be between -5 and 5 degrees. These values are very small and therefore can practically be assumed zero, implying that cracks propagate in a plane coincident with the original cracks parallel to the surface.

4.4.3 The Crack Propagation Mechanism in Elasto-Plastic Solids

Linear elastic fracture mechanics are found to be useful in assessing the crack tip stress concentration and in determining the crack trajectory by mixed mode fracture criteria. However, the size restriction by plasticity considerations and the small stress intensity factors calculated (which is less than the threshold) suggest that the plastic fracture mechanics approach is required. Recently, Sin and Suh investigated this problem at MIT using the finite element method[23].

In this study ADINA (Automatic Dynamic Incremental Nonlinear Analysis, a finite element computer program for the static and dynamic displacement and stress analysis of solids, fluid-structure systems and structures) was used as in the elastic analysis case[29]. The material model used in this study was as follows: infinitesimal, material nonlinearity only[30].

The model used to calculate the elastic-plastic response under the moving load condition was the same as that for the elastic case. No dynamic effect was considered in the analysis. The material was assumed to be slightly work-hardening ($E_T = 10^{-4}$ E). The material properties used were as follows: isotropic, slightly work-hardening, E = 1.96 x 10^5 MPa = 2 x 10^4 kg/mm^2, E_T = 19.6 MPa = 2 kg/mm^2, ν = 0.28, $\sigma_y = \sqrt{3k}$ = 424 MPa = 43.3 kg/mm^2.

For the investigation of crack propagation only, a short crack (c = 1/4a) was used. Due to the prohibitively expensive computer cost, only a limited parameter study was conducted for the case of a = 10 μm, p_o= 4k, and μ = 0.25.

The problem was solved incrementally by moving the load step by step. For an accurate solution, the load increments per step should be sufficiently small. However, such a load step requires a large number of calculations that make the analysis very expensive. Therefore, larger load steps were

used with iteration to obtain efficient and accurate solutions. The use of iteration can introduce some difficulties. The convergence process may be slow, requiring a large number of iterations that can be expensive. Also, some iterative methods do not converge for certain types of problems or large load increments. For details of numerical computation, the reader is referred to Reference (23). Figure 4.28 shows the development of the plastic zone along with the moving load step by step when this element is used. At the beginning, the shape of the plastically deformed zone is more or less the same as for the case of no crack inside the zone, except right around the crack tips. Due to the presence of a crack, the stress field changes significantly when the load moves over the crack. It can also be noticed that there are some spots inside the plastic region where unloading has taken place. With the repeated loading-unloading, it is expected that the overall plastic zone should become smaller as in the case of no crack.

Examination of the displacement of the nodal points on the crack surfaces shows that the upper surface initially slides forward and then slides backward. This is consistent with the result by McClintock who investigated the behavior of a subsurface crack in a rolling contact problem[27]. In Table 4.3, the relative sliding displacements of crack tip nodel points are shown for two different depths of crack location. It shows that the relative displacement increases with sliding, then decreases, and finally changes sign. The following can also be observed from the table: (1) the crack closer to the surface has large values, and (2) the left tip usually has large values of relative sliding displacements.

In Figures 4.29 and 4.30, the crack tip shear strains are plotted along the distance from the crack tip for different stages of the loading position as the asperity contact moves from left to right. The distribution of shear strain increases until it reaches a maximum value and then decreases. After it attains a minimum, it increases again. At very near the tip, the shear strain changes from positive to negative, and then to positive again. For a crack located at d = 0.25a, the shear strain can reach more than 120% at a point of 0.00625 μm away from the tip.

As shown in the preceeding subsection, the subsurface cracks are likely to grow in the direction parallel to the surface, even though it was based on elastic solutions. Also the experimental findings of surface texture development due to sliding strongly suggest that the cracks propagate along the slip planes nearly parallel to the surface[31, 32].

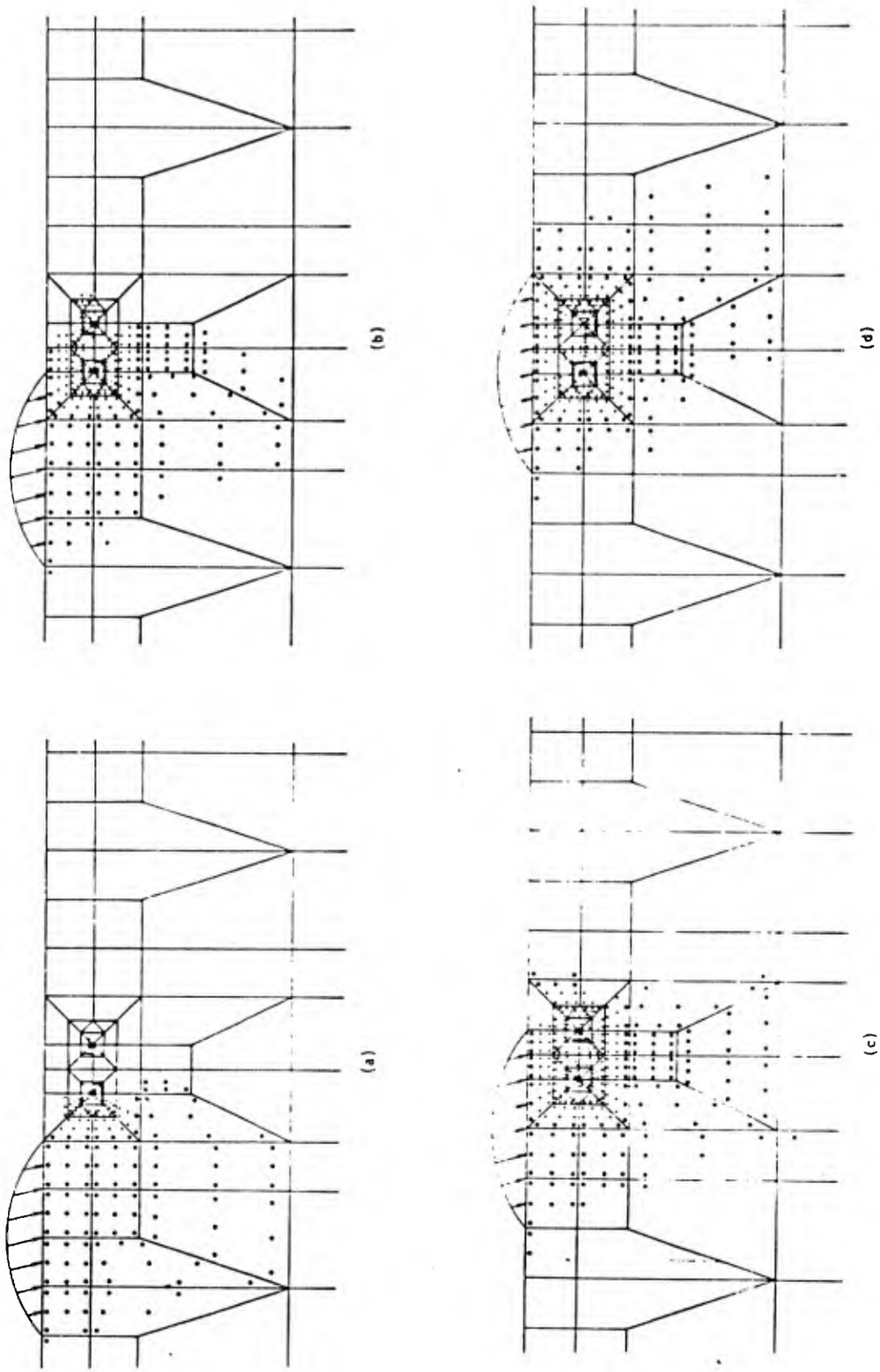


FIGURE 4.28

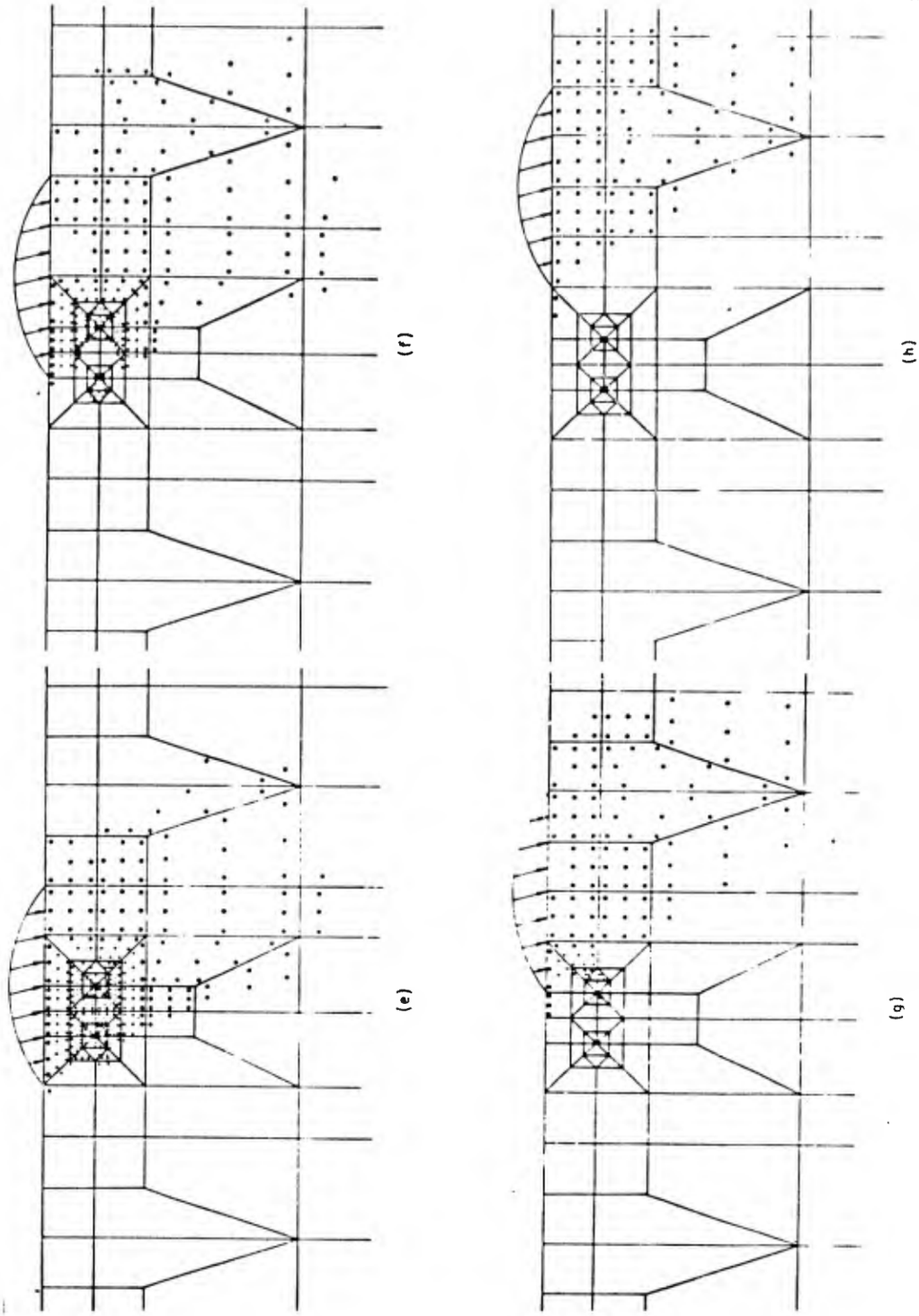


FIGURE 4.28 VARIATION OF PLASTICALLY DEFORMED ZONE AROUND A CRACK UNDER A MOVING ASPERITY: a = 10 μm, d = 0.5α, μ = 0.25, p_o = 4k = 980 MPa. DOTS INDICATE THE INTEGRATION POINTS.

TABLE 4.3 RELATIVE CRACK-TIP SLIDING DISPLACEMENT (μm)

(a) d = 0.5a

	a	b	c	d	e	f	g	h*
Right Tip	0.0009	0.0023	0.0033	0.0027	0.0015	-0.0009	-0.0008	-0.0001
Left Tip	0.0013	0.0030	0.0035	0.0028	0.0004	-0.0012	-0.0005	-0.00003

(b) d = 0.25a

	a	b	c	d
Right Tip	0.0003	0.0012	0.0077	0.0093
Left Tip	0.0002	0.0031	0.0141	0.0138

*Each step corresponds to the relative position shown in Fig. 4.28

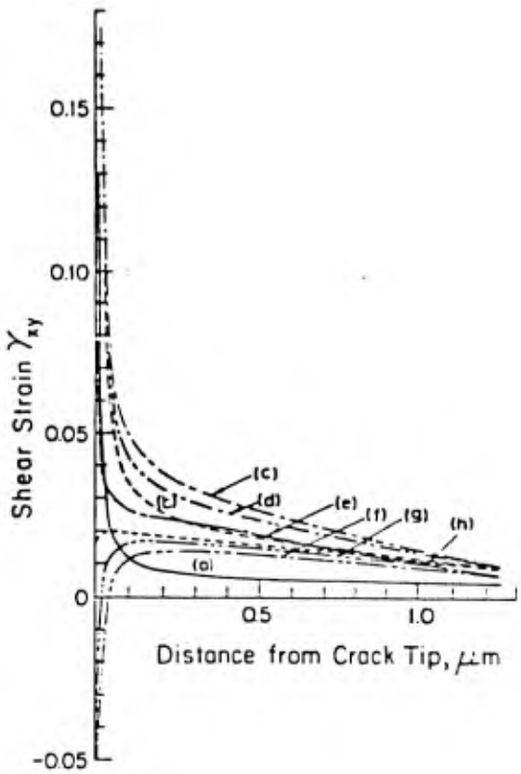


FIGURE 4.29 SHEAR STRAIN VERSUS DISTANCE FROM THE LEFT CRACK TIP WHEN PLASTIC ELEMENTS ARE USED. (A), (B) -- (H) CORRESPOND TO THE POSITION OF THE ASPERITY CONTACT RELATIVE TO THE CRACK TIP SHOWN IN FIG. 4.28

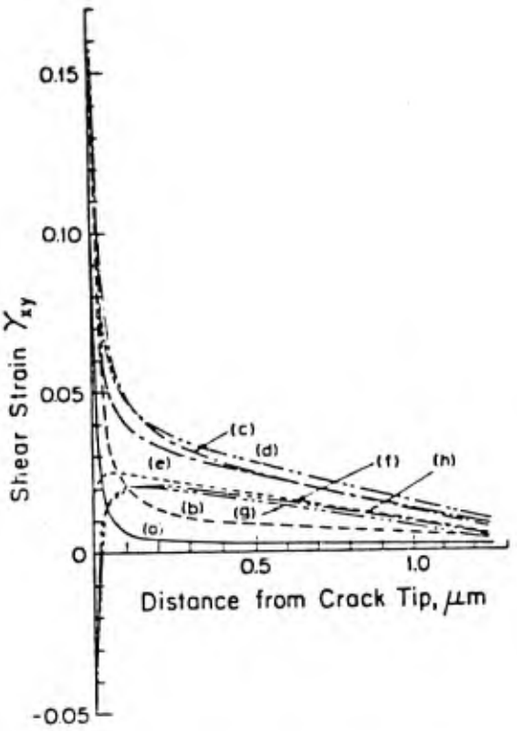


FIGURE 4.30 SHEAR STRAIN VERSUS DISTANCE FROM THE RIGHT CRACK TIP WHEN PLASTIC ELEMENTS ARE USED. (A), (B) -- (H) CORRESPOND TO THE POSITION OF THE ASPERITY CONTACT RELATIVE TO THE CRACK TIP SHOWN IN FIG. 4.28

The crack -opening displacement concept may be applied to the cracks in Mode II to determine the crack growth rate per cycle of loading. In this mode the relative sliding displacement occurs by means of slip due to crack tip deformation. If the maximum of this displacement is employed as a crack-tip sliding displacement (CTSD), denoted by ΔS, then the crack growth length, ΔC, may be expressed as

$$\Delta C = \Delta S - \Delta C_W \tag{4.44}$$

where ΔC_W is the length of rewelding. As for ΔC_W, there is no satisfactory condition. According to Kikukawa et al, rewelding seems to be affected by environmental conditions only in Mode I[33]. In Mode II, they found that the length of rewelding, and therefore the crack growth length, does not differ discernably between the air and vacuum tests. Moreover, the ratio of $\Delta C/\Delta S$ was found to be nearly 0.16, which is fairly small in comparison with the ratio of $\Delta C/COD$ (0.55) for Mode I in their study.

As implied by Kikukawa et al's study, a crack can grow to the point where fracture really occurs, i.e., the fracture conditions are satisfied. These conditions may be addressed in terms of fracture strain, γ_f. If the material element at a distance r_f from the crack tip attains the fracture strain, actual fracture will occur up to this point. The strain γ_{xy} along the distance r from the crack tip can be written as

$$\gamma_{xy} = K_\varepsilon / r^m \tag{4.45}$$

where K_ε is the strain intensity factor and m is the parameter which depends on the strain hardening exponent. For a nonhardening material and for the HRR field, m is unity. Therefore, r_f may be obtained by substituting γ_f for γ_{xy} in Eq. 4.45.

$$r_f = (K_\varepsilon / \gamma_f)^{-m} \tag{4.46}$$

Figure 4.31 shows the shear strain distribution near the crack tip of cracks located at d/a = 0.25 and d/a = 0.5. If the contact length is 20 μm and if the fracture strain γ_f is assumed to be 0.4, the material within 0.0030 μm of the crack tip of the crack located at d = 5 μm will fracture, according to Figure 4.31. When the crack is located at d = 2.5 μm, the material within 0.02 μm of the crack tip will fracture.

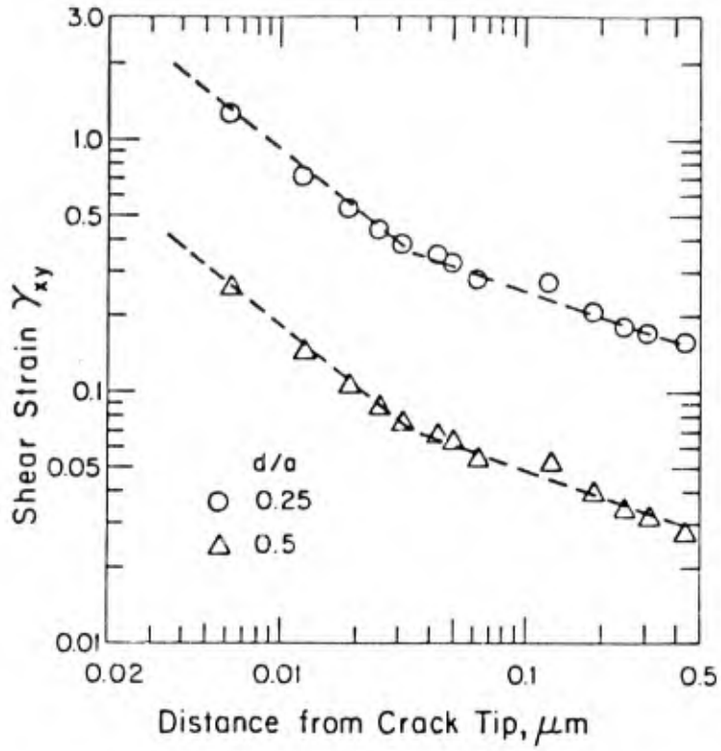


FIGURE 4.31 SHEAR STRAIN AS A FUNCTION OF DISTANCE FROM THE LEFT TIP FOR DIFFERENT DEPTHS OF CRACK LOCATION

4.4.4 Experimental Results

Crack propagation at the subsurface has been documented by Suh, et al[34]. Figures 4.32 and 4.33 show a typical micrograph of subsurface cracks in OFHC copper and iron solid solution, respectively. Figure 4.34 shows the early stages of crack propagation in annealed Fe-1.3%Mo. Cracks are not exactly parallel to the surface. This could be due to the loading condition, anistrophy of materials, and unique texture around second phases.

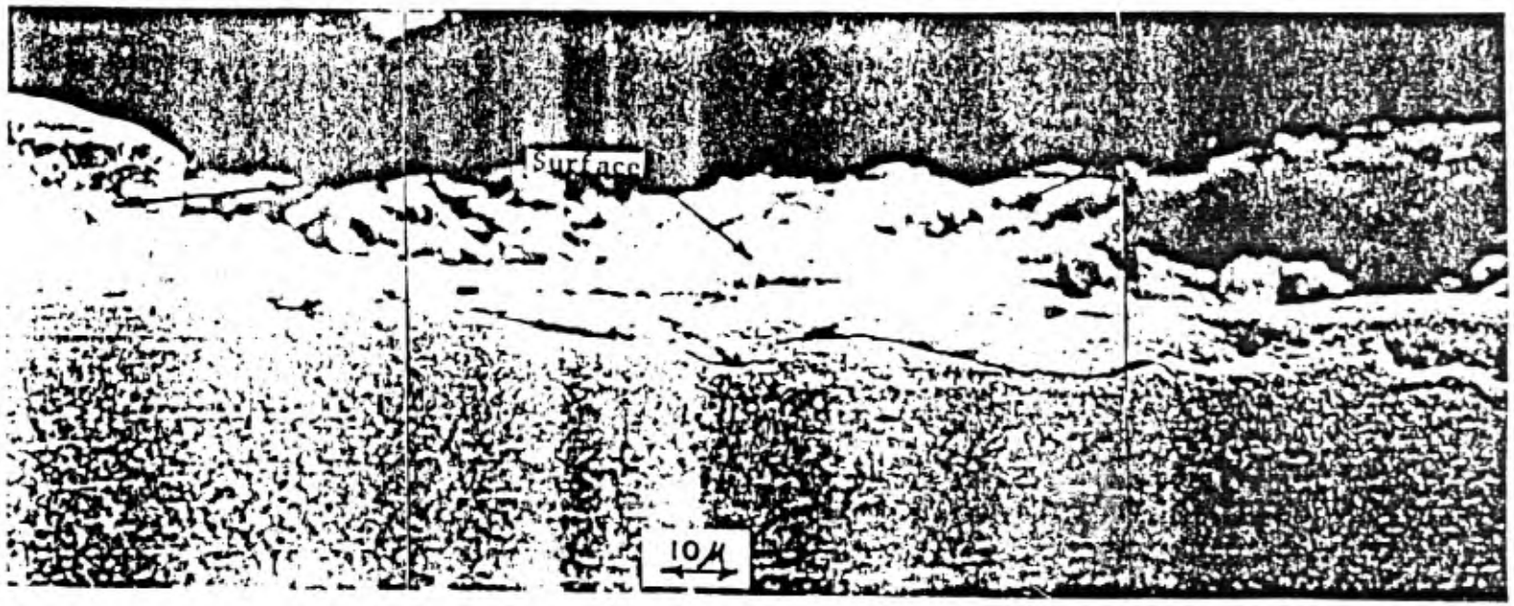


FIGURE 4.32 VOID AND CRACK FORMATION IN ANNEALED OFHC COPPER

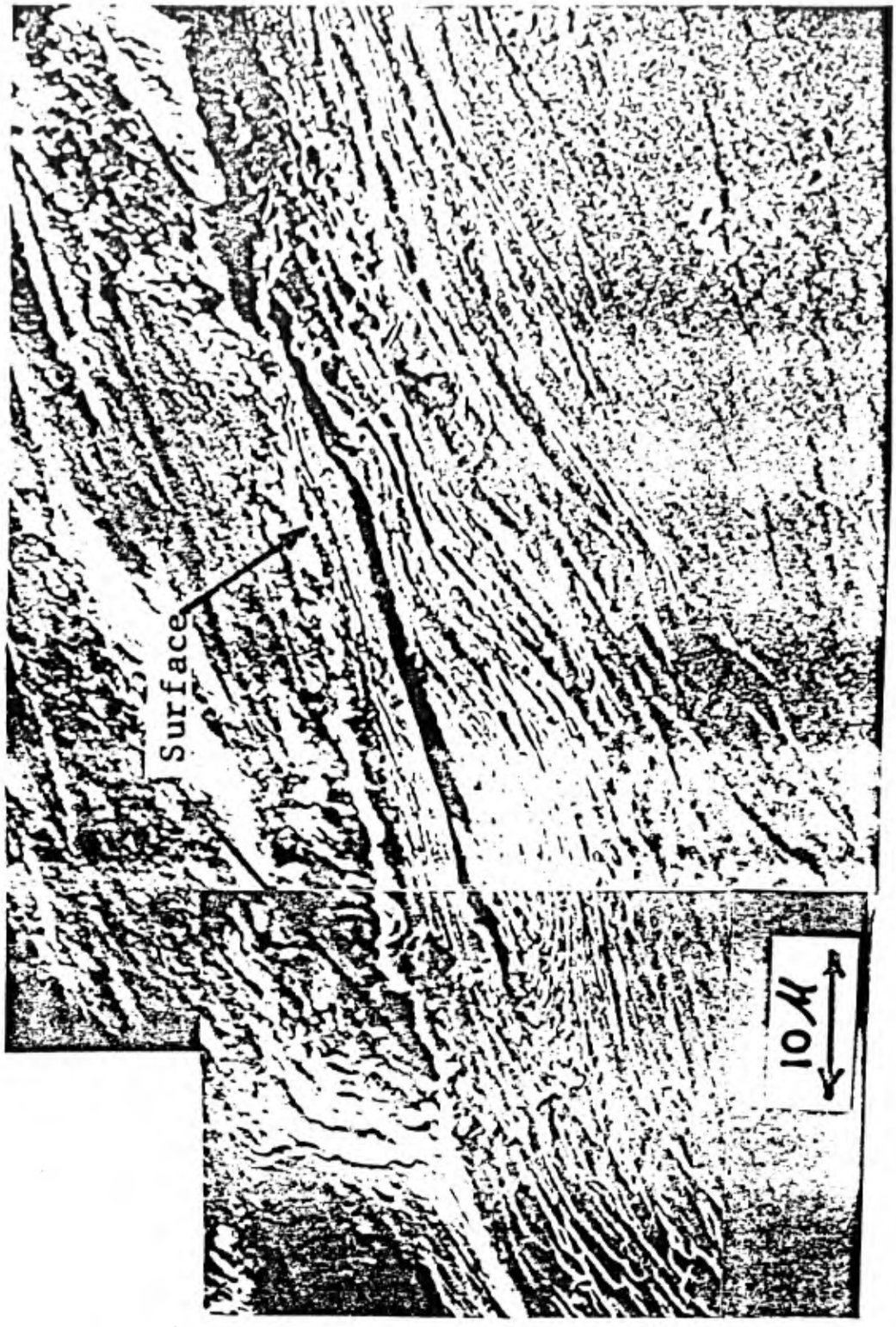


FIGURE 4.33 SUBSURFACE DEFORMATION AND CRACK FORMATION IN IRON SOLID SOLUTION

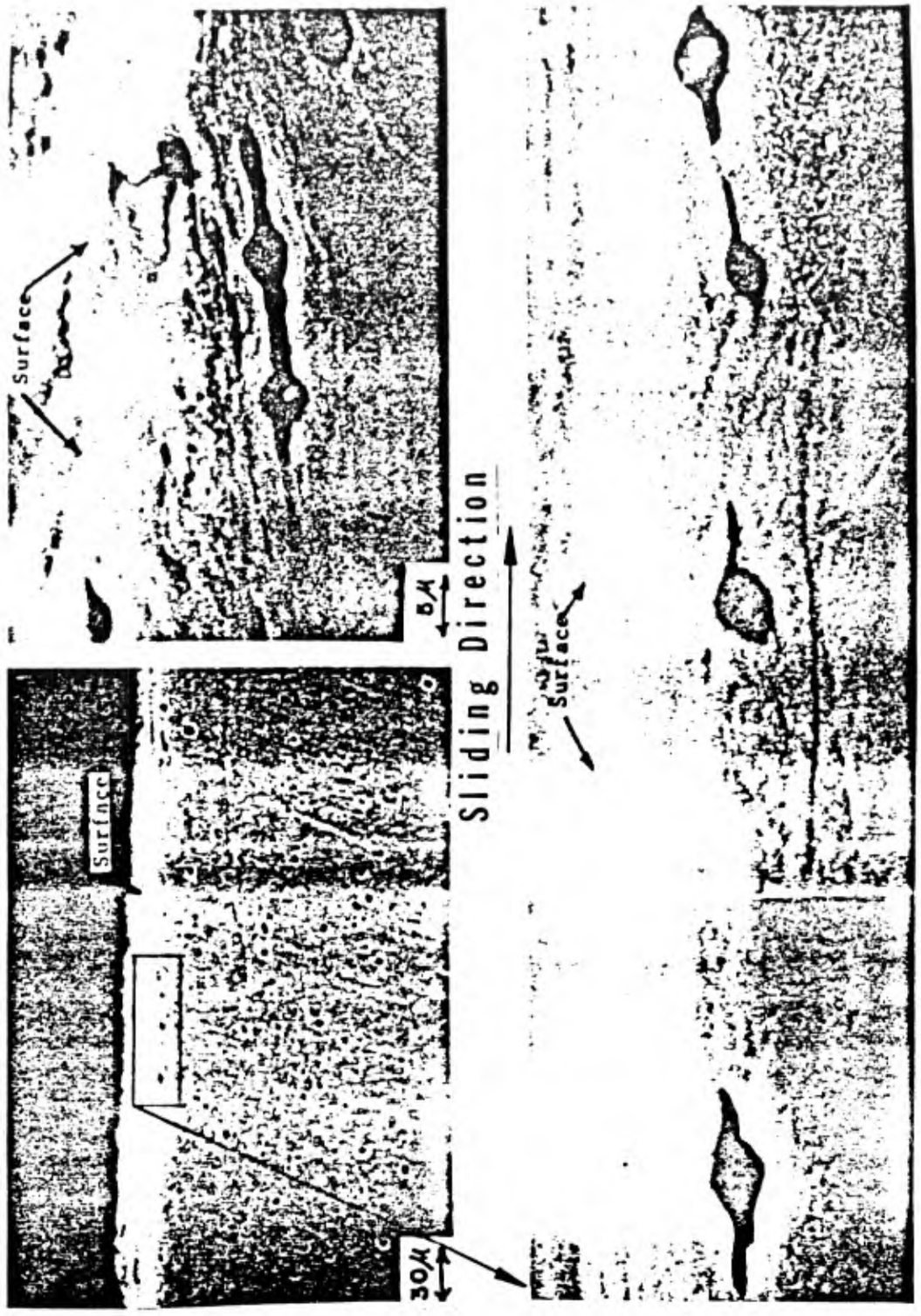


FIGURE 4.34 . VOID FORMATION AROUND INCLUSION AND CRACK PROPAGATION FROM THESE VOIDS NEAR THE SURFACE IN ANNEALED Fe-1.3% Mo

Cracks are prevalent in two phase metals. In high purity single phase metals cracks are not a common occurrence and consequently the wear rate is low. However, cracks have been observed even in high purity metals. Whether these cracks are due to dislocation interactions or to small impurities we do not know. Further work is needed to understand crack nucleation mechanism in single phase metals.

Cracks have also been observed in polymers, composities, and ceramics. Clerico showed that chopped glass fiber reinforced polymers also develop subsurface cracks[35]. Swain also reported formation of subsurface cracks in alumina[36]. All these cracks lead to failure of the surface layer.

4.5 Concluding Remarks

In this section it was shown that materials respond to the forces exerted at the asperity contacts by deformation, crack nucleation, and crack propagation. It is shown that in the case of elasto-plastic solid, the plastic strain accumulates under the cyclic loading by moving asperities. With the plastic strain accumulation, the stress at the particle/matrix interface grows. When the stress exceeds the adhesive strength between the particle and the matrix and when the inclusion is larger than about 500°A, crack nucleation occurs. Once these cracks are present they can propagate which seems to be caused by large plastic deformation at the crack tip along the direction parallel to the surface. Depending on the material and the loading condition, either crack nucleation or propagation can be the wear rate controlling process.

Although a great deal of emphasis is given to elasto-plastic solids in sliding situations, cracks can also develop in brittle solids. In this case, cracks are likely to nucleate at and perpendicular to the surface. This mode of surface damage is not prevalent under sliding contact.

4.6 References - Section 4

1. Merwin, J.E. and K.L. Johnson, "An Analysis of Plastic Deformation in Rolling Contact," Proc. Institution Mech. Engrs. 177 (1963) 697-690.

2. Johnson, K.L. and J.A. Jefferis, "Plastic Flow and Residual Stresses in Rolling and Sliding Contact," Proc of Symposium on Fatigue in Rolling Contact, Inst. Mech. Engrs., (1963) 54-65.

3. Jahanmir, S. and N.P. Suh, "Mechanics of Subsurface Void Nucleation in Delamination Wear," Wear, 44 (1977) 17-38; also Jahanmir, S., Ph.D. Thesis, MIT, 1976.

4. Hertz, H., Gesammette Werke, Vol. 1., Leipsig, Germany, 1895.

5. Smith, J.O. and C.K. Liu, "Stresses Due to Tangential and Normal Loads on an Elastic Contact with Application to Some Contact Problems," Journal of Applied Mechanics, Trans. A.S.M.E., 20 (1953) 157-166.

6. Gill, S., "A Process for the Step by Step Integration of Differential Equations in an Automatic Digital Computing Machine," Proc. Camb. Phil. Soc., 47 (1951) 90-108 and IBM, "Scientific Subroutine Package," Subroutine RKGS, 1975.

7. Sin, H.C., "Surface Traction and Crack Propagation in Delamination Wear," Ph.D. Thesis, MIT, 1981.

8. Anard, S.C., "Numerical Investigation of Stresses in the Inelastic Range in a Rolling Contact," Proc. Second Int. Conf. on Vehicle Structural Mechanics, Society of Automotive Engineers, 1977, pp. 121-127.

9. Augustsson, G., "Strain Field Near the Surface Due to Surface Traction," S.M. Thesis, MIT, 1974.

10. Suh, N.P., "An Overview of the Delamination Theory of Wear," Wear, 44 (1977) 1-16.

11. Argon, A.S., "Formation of Cavities From Nondeformable Second-Phase Particles in Low Temperature Ductile Fracture," J. Eng. Mat. Tech., Trans. ASME, (1976) 60-68.

12. Hirth, J.P. and D.A. Rigney, "Crystal Plasticity and the Delamination Theory of Wear," Wear, 39 (1976) 133-141.

13. McClintock, F.A., and A.S. Argon, "Mechanical Behavior of Materials," Addison-Wesley, Reading, MA, 1966.

14. Rhee, S.S. and F.A. McClintock, "On the Effects of Strain Hardening on Strain Concentrations," Proc. 4th Natl. Cong. Appl. Mech., ASME, (1962) pp. 10002.

15. Johanmir, S. and N.P. Suh, "Mechanics of Subsurface Void Nucleation in Delamination Wear," Wear, 44 (1977) 17-38.

16. Su, K.Y., "Void Nucleation in Particulate Filled Polymeric Materials and Its Implications on Friction and Wear Properties," Ph.D. Thesis, MIT, 1980.

17. Eshelby, J.D., "The Determination of the Elastic Field of an Ellipsoidal Inclusion, and Related Problems," Proc. Roy. Soc. A241 (1957) pp. 376-396.

18. Suh, N.P., "Wear Mechanisms: An Assessment of the State of Knowledge," Fundamentals of Tribology, (Ed. N.P. Suh and N. Saka), MIT Press, 1980.

19. Fleming, J.F. and N.P. Suh, "Mechanics of Crack Propagation in Delamination Wear," Wear, 44 (1977) 39-56.

20. Hills, D.A. and D.W. Ashelby, "On the Application of Fracture Mechanics to Wear," Wear, 54 (1979) 321-330.

21. Rosenfield, A.R., "A Fracture Mechanics Approach to Wear," Wear, 61 (1980) 125-132.

22. Keer, L.M., M.D. Bryant and G.K. Haritos, "Subsurface Cracking and Delamination," Solid Contact and Lubrication, A.S.M.E. Pulbication, A.M.D., (39) (1980) 79-95.

23. Sin, H.C. and N.P. suh, to be published.

24. Erdogan, F. and G.C. Sih, "On the Crack Extension in Plates Under Plane Loading and Transverse Shear," J. Basic Engr. 85 (1963) 519-527.

25. Sih, G.C., "Strain-Energy-Density Factor Applied to Mixed Mode Crack Problem," Int. J. Fracture, 10 (1974) 305-321.

26. Swedlow, J.L., "Criteria for Growth of the Angled Crack," Cracks and Fracture, ASTM STP 601, 1976, pp. 506-521.

27. McClintock, F.A., "Plastic Flow Around a Crack Under Friction and Combined Stress," Fracture 1977 (Eds. D.M.R. Taplin), Vol. 4, pp. 49-64, Pengamon, Oxford 1978.

28. Forsyth, P.J.E., "A Two Stage Process of Fatigue Crack Growth," Proc. Crack Prop. Symposium, Cranfield, Vol. 1, 1961, pp. 76-94.

29. Bathe, K.J., "ADINA - A Finite Element Program for Automatic Dynamic Incremental Nonlinear Analysis," Report 82448-1, Acoustics and Vibration Laboratory, MIT, 1975.

30. Bathe, K.J., "Static and Dynamic Geometric and Material Non-linear Analysis Using ADINA," Report 82448-2, Acoustics and Vibration Laboratory, MIT 1976.

31. Wheeler, D.R. and D.J. Buckley, "Texturing in Metals as a Result of Sliding," Wear, 33 (1975) 65-74.

32. Krause, H. and A.J. Demirci, "Texture Changes in the Running Surface of F.C.C Metals as the Result of Frictional Stress," Wear, 61 (1980) 325-332.

33. Kikukawa, M., M. Jono, and N. Adachi, "Direct Observation and Mechanism of Fatigue Crack Propagation," Fatigue Mechanisms, ASTM STP 675, 1979, pp. 234-253.

34. Suh, N.P., and Co-workers, "The Delamination Theory of Wear," Elsevier, 1977.

35. Clerico, M., "Sliding Wear Mechanisms of Polymers," Fundamentals of Tribology, (Eds. N.P. Suh and N. Saka), MIT Press, 1980.

36. Swain, M.V., "Microscopic Observations of Abrasive Wear of Polycrystalline Allumina," Wear, 35 (1975) 185-189.

5. SLIDING WEAR

5.1 Introduction

In Section 1, various wear processes have been briefly reviewed. It was stated that there are two classes of wear: one controlled by the mechanical behavior of materials and the other by the chemical behavior. One of these mechanical behavior controlled wear processes is sliding wear, which is often called inappropriately, adhesive wear in the literature. The purpose of this section is to examine various ways wear particles can be generated when two surfaces slide against each other. In this section, the discussion will be confined to those cases where the interfacial temperature rise is not high enough to cause wear processes controlled by the chemical behavior of materials.

Wear particles in elasto-plastic solids are generated during sliding by the following three mechanisms:

1) Asperity deformation -- Orginally existing surface asperities and the asperities generated due to the delamination wear can be removed upon deformation. They may be removed in a single asperity interaction or in multiple, repeated interactions.

2) Plowing — Surfaces can be plowed by wear particles generated by asperity removal or by delamination. Hard particles from the environment can also plow the surface. Surfaces can also be plowed by hard asperities of the counterface. The consequence of plowing can result in the generation of chips by a cutting action or the formation of ridges which deform and get removed due to subsequent loadings.

3) Delamination -- Large wear particles are removed by the process of plastic deformation of the surface layer, and subsurface crack nucleation and propagation. This process of wear particle generation will be discussed in great detail.

Wear particles in brittle solids are generated by different mechanisms from those of elasto-plastic solids. They are normally produced by cracks running perpendicular to the surface starting from minute surface defects or subsurface defects very near the surface. These cracks then change direction and small particles are created.

Normally brittle solids, such as aluminum oxide and even glass, sometimes act as elastic-plastic solids when extremely small chips are cut from the surface. This size effect has been observed in many different situations as in grinding, sliding, and hydrostatic deformation of materials.

In Section 2 it was shown that the surface is highly heterogeneous if not contaminated. This fact was supported from experimental observations, and through quantum mechanical and thermodynamic reasoning. Therefore, it was emphasized that ashesion is normally difficult to achieve except when fresh new surfaces are exposed during asperity contacts.

In Section 3, the mechanism of crack nucleation was examined in depth, whereas in Section 4 crack propagation was investigated from a continuum mechanics point of view.

The purpose of this section is to describe the delamination theory of wear both for metals and polymers. This section will also present the means of reducing wear caused by sliding.

5.2 The Delamination Theory of Wear

5.2.1 Description of the Theory

The delamination theory of wear was introduced in 1972 in order to explain the wear of metals and other solid materials[1]. The theory has since been supported by a large number of experimental results*[2]. The significance and implications of the mechanisms postulated by the delamination theory of wear can best be understood by considering the following questions:

1) Where does all the energy supplied by the external agent go?

2) Why and how does the coefficient of friction affect the wear rate?

3) Why do some hard metals wear faster than soft metals?

4) Why do most wear particles have an aspect ratio greatly different from unity?

5) Why does seizure occur?

6) How does the microstructure of metals affect the wear rate?

7) How do initial surface roughness and waviness influence the wear phenomenon?

The delamination theory of wear describes the following sequential (or independent if there are preexisting subsurface cracks) events which lead to loose wear sheet formation[1].

(1) When two surfaces come into contact, normal and tangential loads are transmitted through the contact points. Asperities of the softer surface are easily deformed and fractured by the repeated loading action, forming small wear particles. Hard asperities are also removed but at slower rates. A relatively smooth surface is initially generated, either when these asperities are deformed or when they are removed.

---

* The original paper is one of the most frequently cited papers in applied science according to the Institute for Scientific Information[3].

(2) The surface traction exerted by the harder asperities at the contact points, induces incremental plastic deformation per cycle of loading, accumulating with repeated loading. The increment of permanent deformation remaining after given cyclic loading is small compared with the total plastic deformation that occurs in that cycle due to the reversal of shear deformation.

(3) As the subsurface deformation continues, cracks are nucleated below the surface. Crack nucleation very near the surface cannot occur due to the triaxial state of compressive loading which exists just below the contact region.

(4) Once cracks are present (either by crack nucleation or from preexisting voids and cracks), further loading and deformation causes the cracks to extend and propagate, eventually joining with neighboring cracks. These cracks tend to propagate parallel to the surface at a depth governed by material properties and the state of loading. When the cracks cannot propagate because of the small stress concentration at the crack tip due to extremely small surface traction at the asperity contact, crack nucleation is the rate controlling mechanism.

(5) When the cracks finally shear to the surface, long and thin wear sheets delaminate. The thickness of the wear sheet is controlled by the location of subsurface crack growth, which is controlled by the normal and tangential loads at the surface.

A series of experimental studies conducted at M.I.T. have substantiated this theory, showing that the delamination process initiates when the subsurface plastic deformation causes the nucleation of voids, Figure 5.1[2]. With further deformation, these voids elongate and link up to form long cracks in a direction nearly parallel to the wear surface, Figure 5.2. At a critical length, these cracks shear to the surface, yielding a wear particle in the form of a long thin sheet as shown in Figure 5.3. The top surface of the wear sheet is generally smooth, while the fractured surface is rough sometimes showing dimples, Figure 5.4, and in some cases, shear failure surfaces. These experimental observations have been explained in terms of the theoretical models presented in the preceeding chapters.

5.2.2 A Model for Delamination Wear

A wear equation based on the delamination theory can readily be developed. For this purpose, consider a subsurface crack lying below the surface as shown in Figure 5.5. An

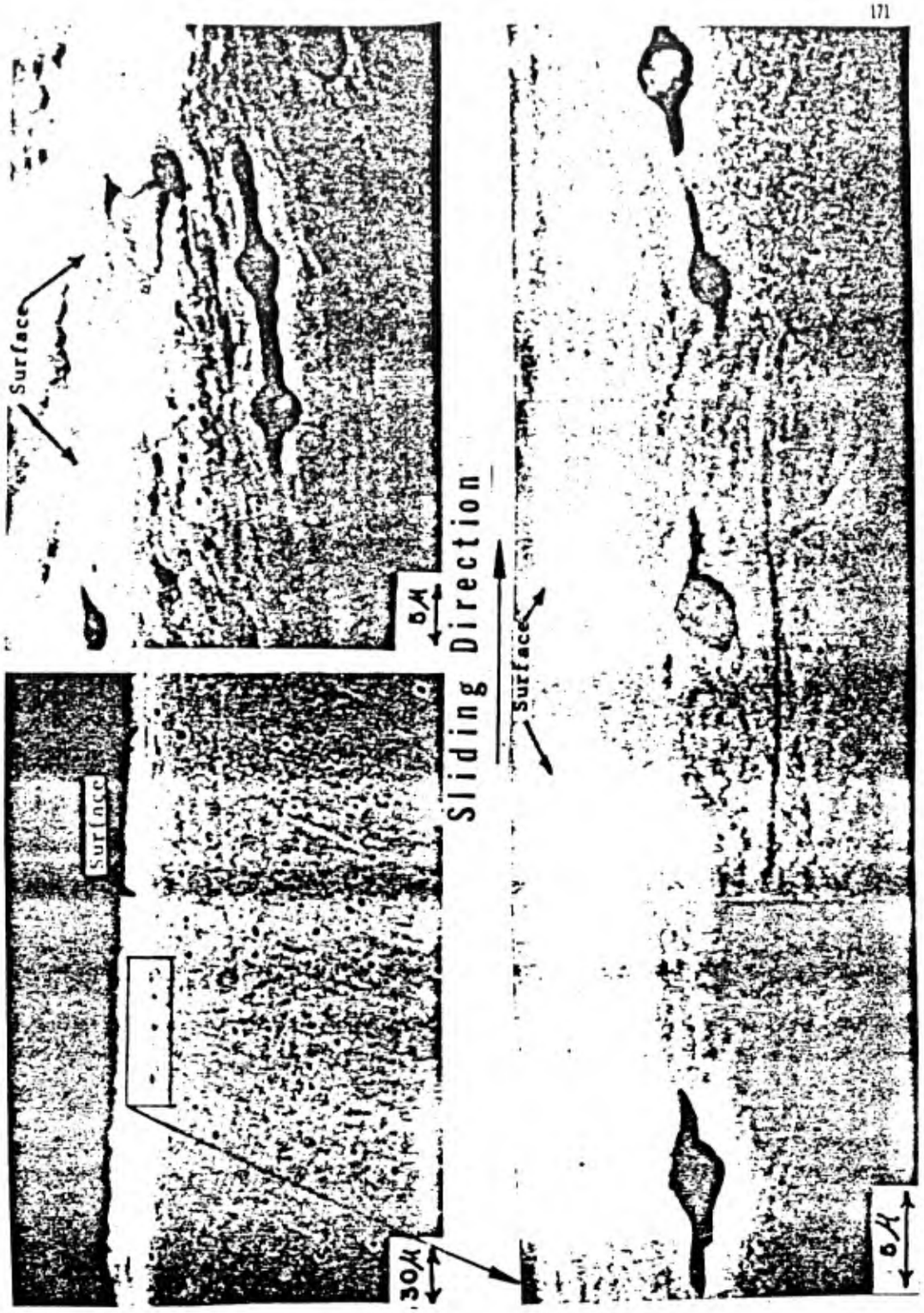


FIGURE 5.1 VOID FORMATION AROUND INCLUSIONS AND CRACK PROPAGATION FROM THESE VOIDS NEAR THE SURFACE IN ANNEALED Fe-1.3%Mo

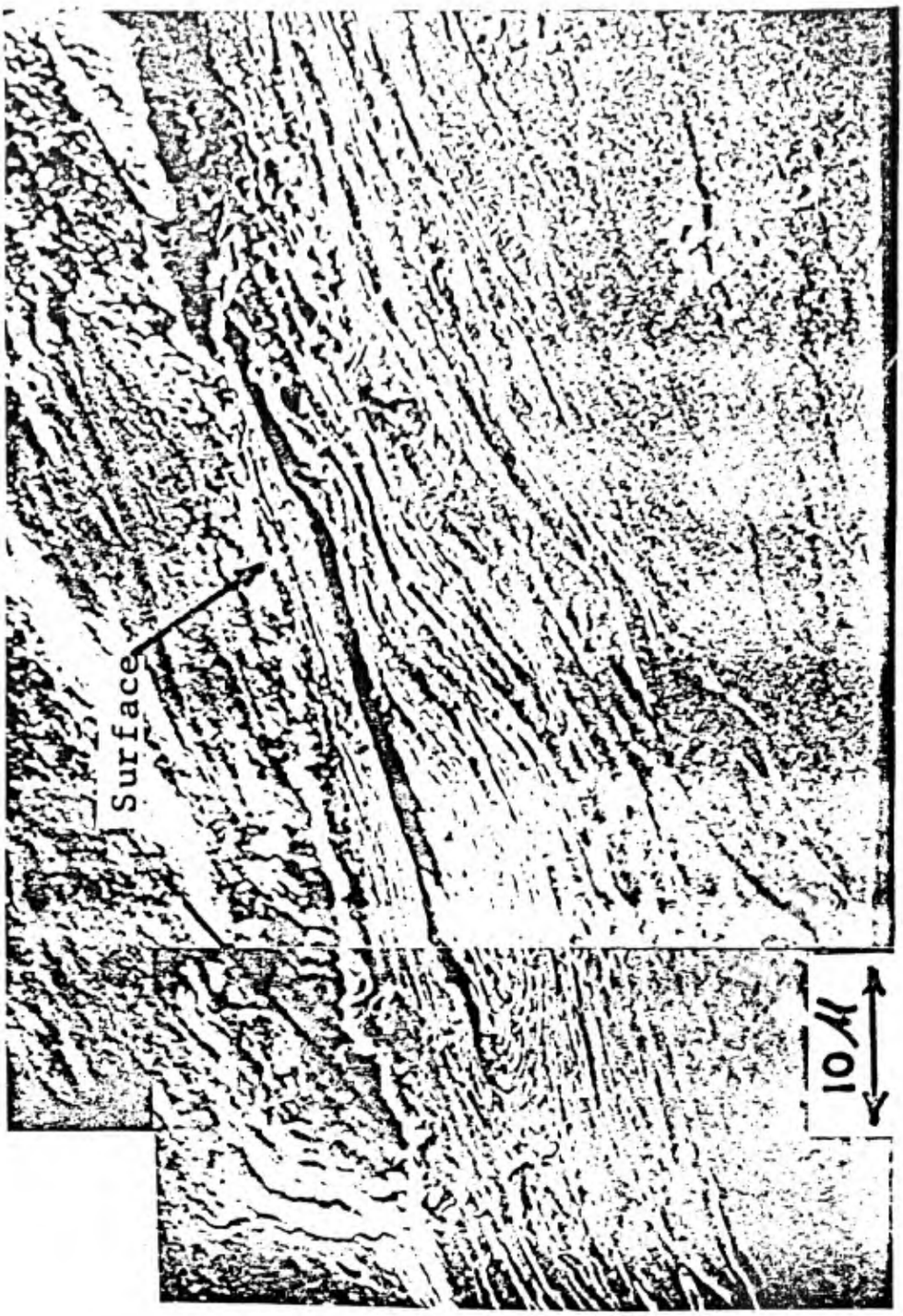


FIGURE 5.2 SUBSURFACE DEFORMATION AND CRACK FORMATION IN IRON SOLID SOLUTION

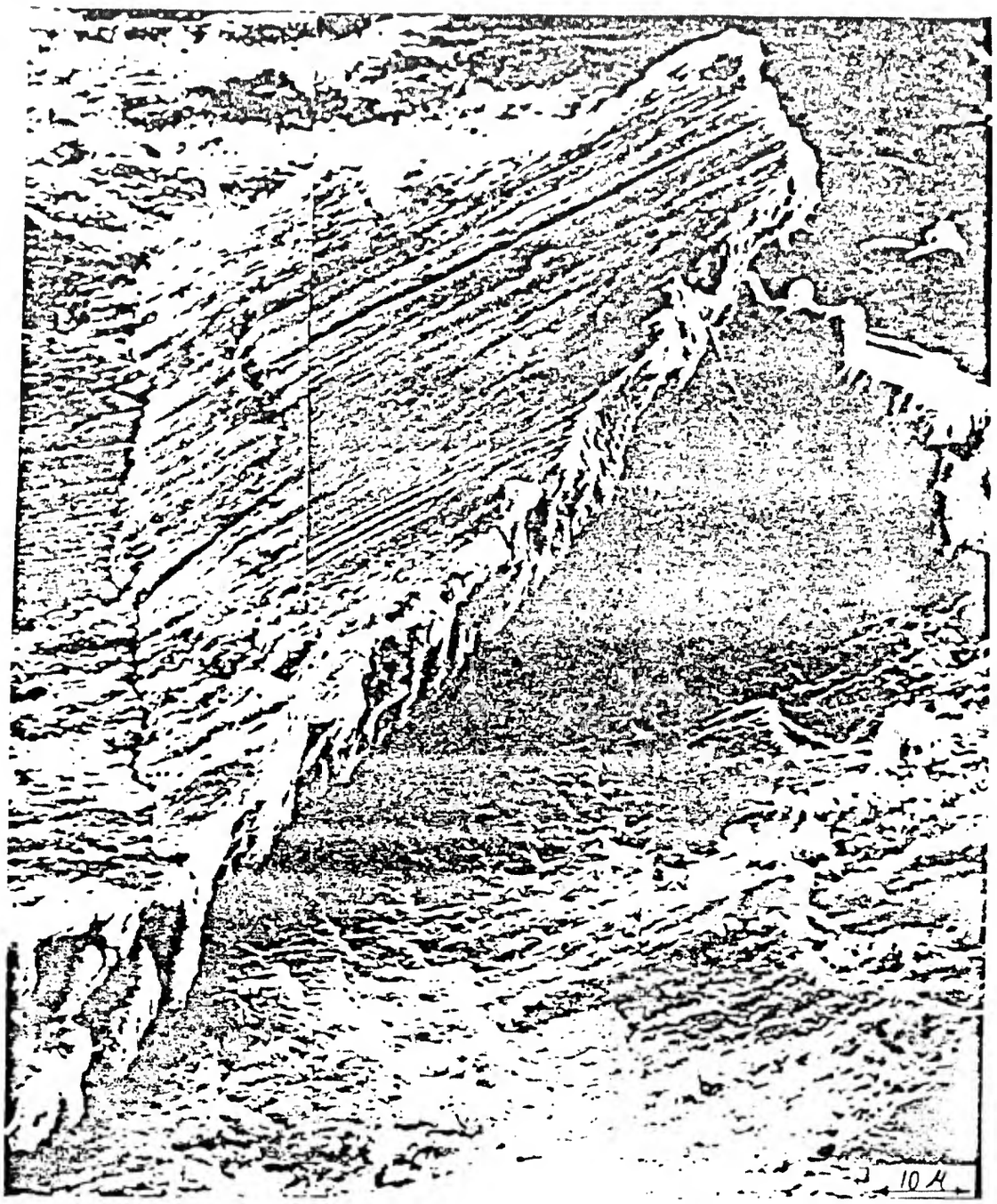


FIGURE 5.3 WEAR SHEET FORMATION IN IRON SOLID SOLUTION

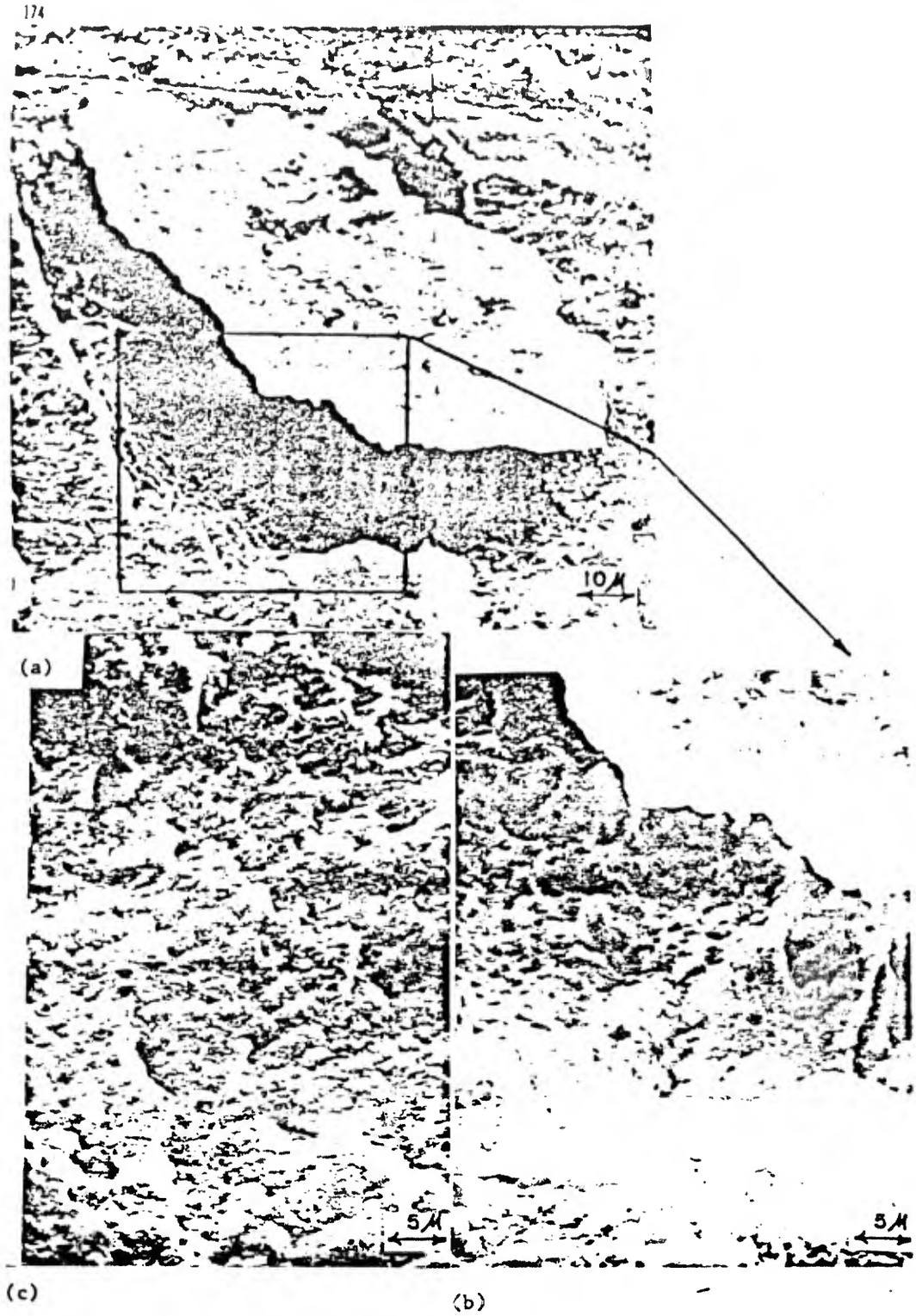


FIGURE 5.4 WORN SURFACE OF PURE IRON: (A) WEAR SHEET FORMATION; (B) SHEAR DIMPLES BENEATH THE WEAR SHEET IN (A); (C) DIMPLED APPEARANCE OF A WEAR CRATER

asperity is moving over the surface from left to right. The wear rate will be dictated by the crack propagation at both ends, L and R. The crack propagation for the ith cycle, ΔC_i, may be expressed as:

$$\Delta C_i = f\ (\mu,\ d,\ C,\ \text{material properties}) \tag{5.1}$$

for both ends. If N is the total number of asperity passes required for removal of one layer, the volume V_1 for one crack of width, w, lying at a depth d is obtained as

$$V_1 = w \cdot d \cdot \sum_{1}^{N} (\Delta C_{L_i} + \Delta C_{R_i}) \tag{5.2}$$

Therefore, the total volume, V, for one layer may be given by

$$V = N_c \cdot N_w \cdot w \cdot d \cdot \sum_{1}^{N} (\Delta C_{L_i} + \Delta C_{R_i}) \tag{5.3}$$

where N_c is the number of cracks along the sliding direction and N_w is the number of wear sheets in the direction of contact width. Since $N_w \times w$ is in the order of the contact width L_w, the volume V becomes

$$V = N_c \cdot L_w \cdot d \cdot \sum_{1}^{N} (\Delta C_{L_i} + \Delta C_{R_i}) \tag{5.4}$$

Let λ be the asperity contact spacing, ℓ_c the crack spacing, D the diameter of a specimen, and L the contact length as shown in Figure 5.6. Using this model the number N_c and N can be determined for a specimen and a slider, respectively.

Specimen:

Since the number of asperities per length of contact is L/λ, the number N is determined as

$$N = \frac{S}{\pi D} \frac{L}{\lambda} \tag{5.5}$$

where S is the sliding distance required for removal of one layer. Also, N_c is given by

$$N_c = \pi D/\ell_c. \tag{5.6}$$

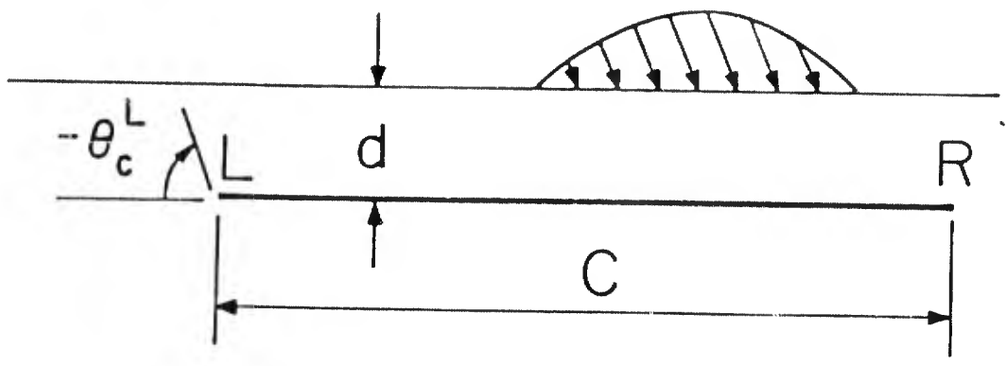


FIGURE 5.5 A SUBSURFACE CRACK UNDER A MOVING ASPERITY

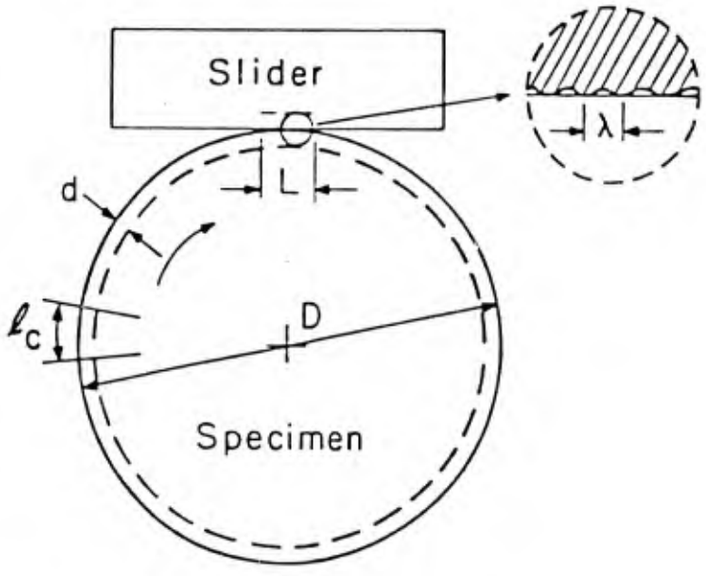


FIGURE 5.6 A MODEL OF WEARING SPECIMEN AND SLIDER

The wear rate is, by assuming $L_w = L$, finally obtained as

$$\frac{V}{S} = \frac{L^2 \cdot d \cdot (\Delta \bar{C}_L + \Delta \bar{C}_R}{\lambda \cdot \ell_c} \tag{5.7}$$

where $\Delta \bar{C}_L$ and $\Delta \bar{C}_R$ are the average crack propagation rates during N cycles.

Slider:

The numbers N and N_c are given by

$$N = S/\lambda \tag{5.8}$$

and

$$N_c = L/\ell_c \tag{5.9}$$

respectively. Therefore, the wear rate for a slider is obtained as

$$\frac{V}{S} = \frac{L^2 \cdot d \cdot (\Delta \bar{C}_L + \Delta \bar{C}_R)}{\lambda \cdot \ell_c} \tag{5.10}$$

which is the same expression as given in Eq. 5.7 for a specimen. The wear coefficient K can be obtained by using Archard's equation as

$$K = \frac{3H \cdot L^2 \cdot d \cdot (\Delta \bar{C}_L + \Delta \bar{C}_R)}{W \cdot \lambda \cdot \ell_c} \tag{5.11}$$

where H is the hardness of the material and W is the applied load.

Eqs. 5.7 and 5.11 show that the wear rate is directly proportional to the depth of crack and the average crack growth. The amount of crack growth, ΔC, depends on the depth, d. The results in Section 4 indicate that ΔC increase with decreasing d for a given friction coefficient of 0.25. For higher μ, the same tendency is expected.

ΔC depends on d because of shear strain concentration at crack tips. Under the moving load condition, it is highly likely that the plastic deformation is much more severe at the crack tips near the surface due to sliding between crack

surfaces even for low friction. However, cracks are only observed at a finite distance below the surface, because they cannot nucleate very near the surface because of high triaxial state of compressive stress, as shown in Section 3. The actual location of the crack will also be determined by the location of second phase particles around which crack nucleation occurs preferentially.

For a longer crack, a much larger shear concentration is expected due to sliding of crack surfaces with respect to each other, which is affected by the shear traction between the crack surfaces. Therefore, the crack propagation rate will increase with increasing total crack length.

When the total crack length reaches a critical size for unstable crack propagation leading to fracture, the direction of crack propagation will change toward the surface, eventually forming a wear sheet. Fracture criteria for the combined mode of loading considered in Section 4, may provide a qualitative explanation for this. Let us consider a subsurface crack under a moving asperity shown in Figure 5.5. If the left tip, L, is in the tensile zone behind the load, then the maximum hoop stress and the minimum strain-energy-density factor criteria predict the crack growth of about -70^{o} at the left tip. When the critical condition is reached with the tensile component, the crack will be fractured to the surface. Micrographs of wear sheets about to delaminate show that subsurface cracks always reach the surface at the trailing edge of the crack[3].

Material properties such as strain-hardening property are the important factors affecting crack propagation, ΔC. Also, the anisotophy of the properties are expected to affect ΔC. Due to repeated loading effects of cyclic hardening, cyclic softening, and the Bauschinger effects are also important. However, it seems impossible at the present time to incorporate all these factors in the analysis because of the extreme complexity. The fracture strain is the simplest parameter that can be used in determining ΔC.

Experimentally, it is well known that the wear behavior of materials strongly depends on the coefficient of friction[4, 5]. A high shear strain distribution is expected ahead of extending cracks when the coefficient of friction is high. Therefore, the crack propagation rate is larger for higher friction coefficients. The exact relation between friction and wear cannot be given because of limited data. However, the elastic solutions obtained in Section 4, although they have already been shown to be inadequate, may be used to discuss the functional relation qualitatively.

In Figures 5.7 and 5.8, the maximum Mode II stress intensity factor for a given d are plotted as a function of the coefficient friction, μ,. It can be seen that the relation between K_{II} and μ is approximately linear for a given d. If we assume that Paris' crack growth law is still valid in this case, then the crack growth, ΔC, may be expressed as

$$\Delta C \sim (a\mu)^n \tag{5.12}$$

where a is a constant. Since the wear volume is determined by the amount of the total ΔC accumulated, wear rates are a power function of friction. The exponent, n, is a constant which varies from 2 to 4 depending on material for ΔK values in stage II. This should give a rough idea about a functional relation between friction and wear, although the foregoing discussion is very approximate.

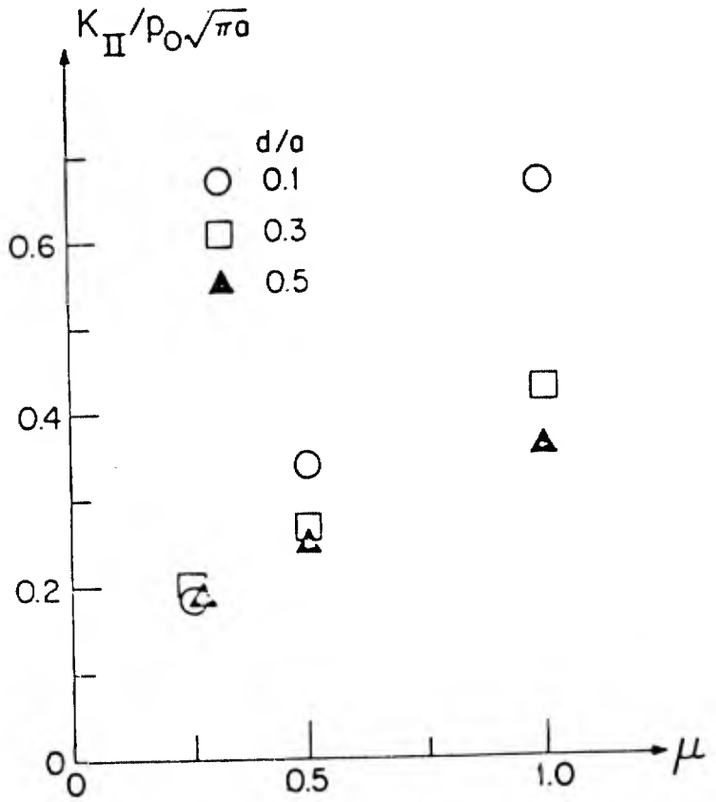


FIGURE 5.7 NORMALIZED ΔK_{II} AS A FUNCTION OF μ: c = 1/4a; LEFT TIP

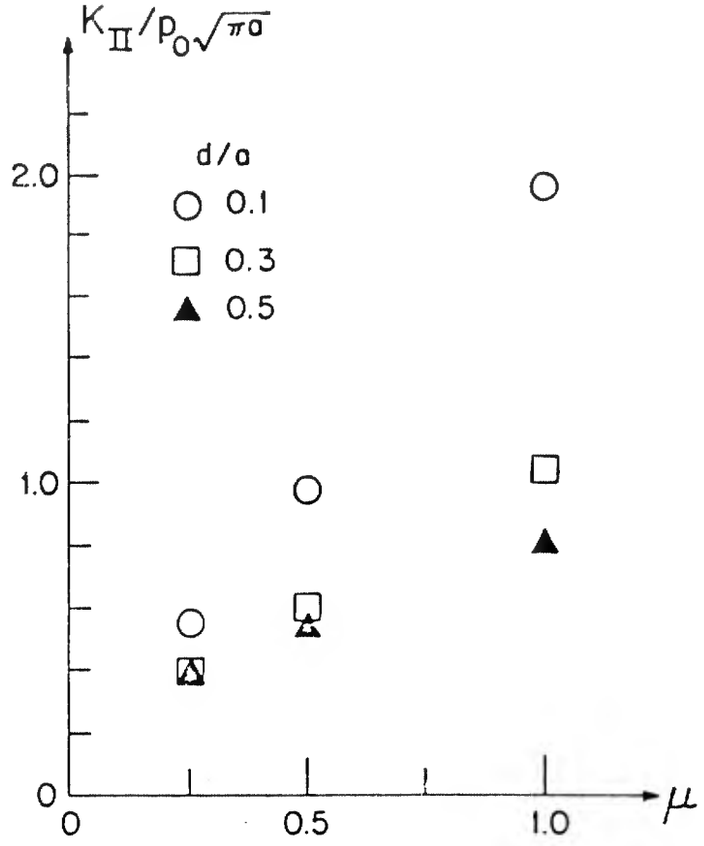


FIGURE 5.8 NORMALIZED ΔK_{II} AS A FUNCTION OF μ: c = 3a

5.2.3 Numerical Examples: Wear Coefficient Prediction

The wear model for delamination wear, i.e., Eqs. (5.7) and (5.11), may be used to predict wear coefficients for typical sliding wear situations. Two examples are given here.

Example 1:

The crack tip sliding displacement for d = 5 μm is 0.0035 μm (Table 4.3). If $\Delta\bar{C}_L = \Delta\bar{C}_R = 0.0035$ μm, and the apparent contact length L = 2 mm and both the asperity contact spacing, λ, and the crack spacing, ℓ_c, are assumed to be 100 μm, then

$$\frac{V}{S} = \frac{(2 \times 10^{-3})^2 (5 \times 10^{-6}) (0.007 \times 10^{-6})}{(100 \times 10^{-6}) (100 \times 10^{-6})} =$$

$$= 1.4 \times 10^{-11} \text{ m}^3/\text{m}$$

and for H = 100 kg/mm^2 and W = 1 kg,

$$K = \frac{3 \times 100 \times 10^6 \times 1.4 \times 10^{-11}}{1} = 4.2 \times 10^{-3}$$

Example 2:

For d = 2.5 μm, by taking $\Delta\bar{C}_L = \Delta\bar{C}_L = 0.014$ μm again from Table 4.3,

$$\frac{V}{S} = \frac{(2 \times 10^{-3})^2 (2.5 \times 10^{-6}) (0.028 \times 10^{-6})}{(100 \times 10^{-6}) (100 \times 10^{-6})}$$

$$= 2.8 \times 10^{-11} \text{ m}^3/\text{m}$$

and

$$K = 8.2 \times 10^{-3}$$

According to the experimental results, K is between 10-2 and 10^{-4} in most cases, which indicates that the model can predict wear behavior fairly well.

5.3 Microstructural Effects in Delamination Wear

The microstructure of metals affect the wear behavior of metals a great deal. Hardness and toughness are generally affected by the microstructure which in turn affect the wear properties. It has been well established that the hardness and the topography of the surface affect the number of asperity contacts and the size of each individual contact, in addition to the resistance to deformation of the surface layer[6, 7]. As we have seen in Section 4, the toughness of metals is closely related to the crack propagation rate, which also effects the rate of wear sheet formation.

Several different aspects of the microstructure of metals affect the wear process: grain size; the properties, volume fraction, size, and distribution of second phase particles; and texture of the surface layer. Each of these aspects will be discussed in this subsection.

From the friction and wear point of view the ideal material that has a low coefficent of friction and a low wear rate is a hard material in which cracks cannot be nucleated and chemically inert. Such a material then must be single-phase without any impurities and tough. However, many engineering materials cannot be made both tough and hard. Hard materials generally have low toughness. Therefore, the choice of materials for tribological application must be made through compromise.

The wear of single phase polycrystalline metals was investigated at MIT[8 - 10]. OFHC copper, copper-chromium (0.58 and 0.81 at % Cr), copper-silicon (2.3 and 8.6 at % Si), and copper-tin (1.4, 3.4 and 5.7 at % Sn) alloys were tested. Table 5.1 gives the heat-treatment conditions, grain size, and chemical composition of these materials. Figure 5.9 shows typical microstructures of these materials. Figure 5.10 shows the Vickers hardness of the solid solutions as functions of the solute content. It can be seen that different elements produce different hardening for the same atomic content of solute. The effectiveness of tin increasing the hardness of copper should be noted.

Alloy	Composition —		Recrystallization treatment		grain size
	Wt.%	At.%	Temperature(C)	Time (min)	(μm)
OFCH Cu	--	--	360	250	28
Cu-Sn	2.5	1.4	790	120	35
Cu-Sn	6.0	3.4	790	60	40
Cu-Sn	10.0	5.7	790	30	30
Cu-Si	1.0	2.3	850	300	75
Cu-Si	4.0	8.6	850	300	90
Cu-Cr	0.47	0.58	1070	5	450
Cu-Cr	0.66	0.81	1070	5	485

*The purity of the materials used was: OFHC 99.98%; Sn 99.89%; Si 99.9999%.

TABLE 5.1 EXPERIMENTAL MATERIALS: COPPER SOLID SOLUTIONS*

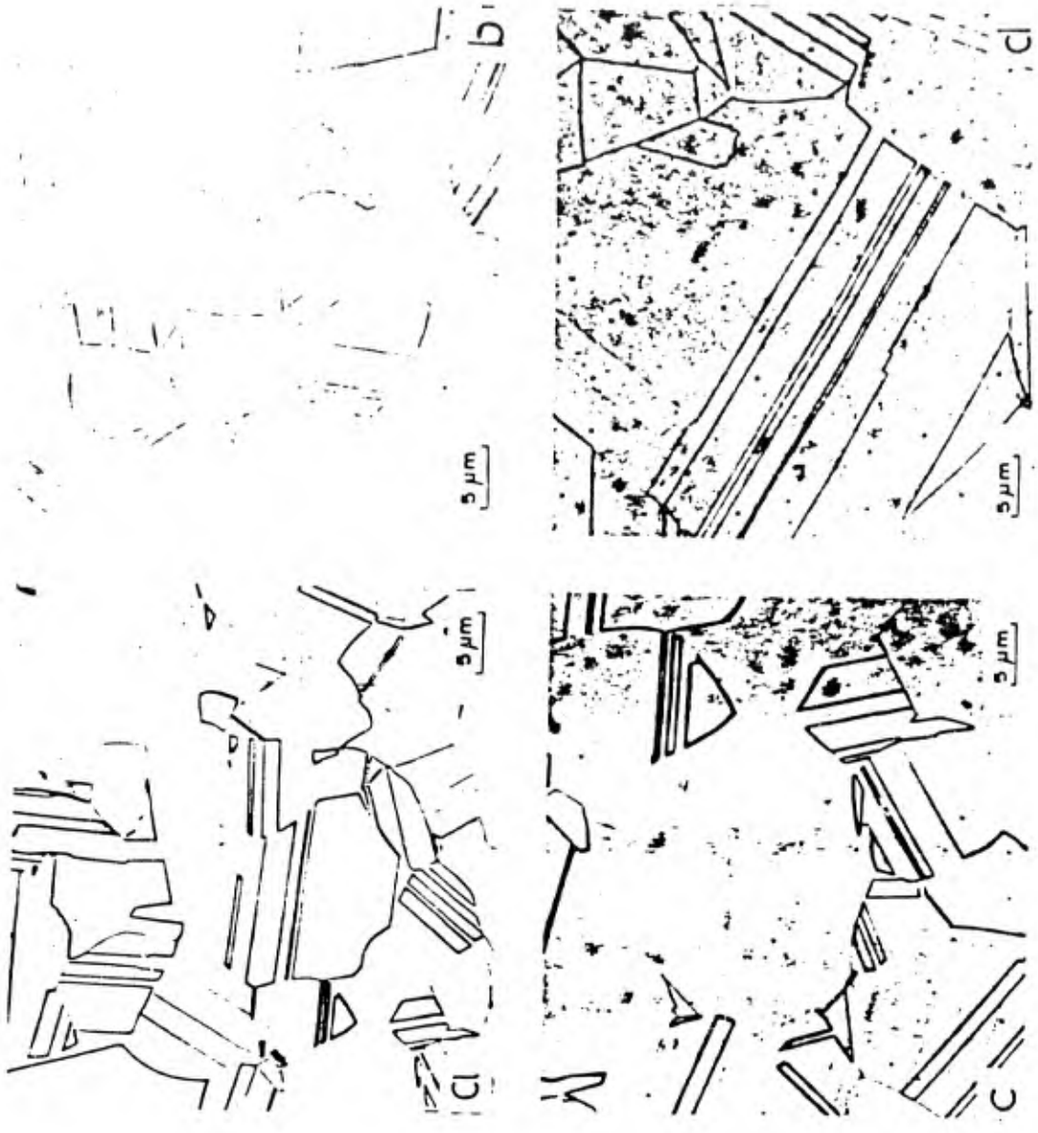


FIGURE 5.9 OPTICAL MICROGRAPHS SHOWING THE REPRESENTATIVE MICROSTRUCTURES OF SOLID SOLUTIONS: (A) OFHC COPPER, (B) Cu-5.7 at % Sn, (C) Cu-8.6 at % Si AND (D) Cu-0.81 at % Cr.

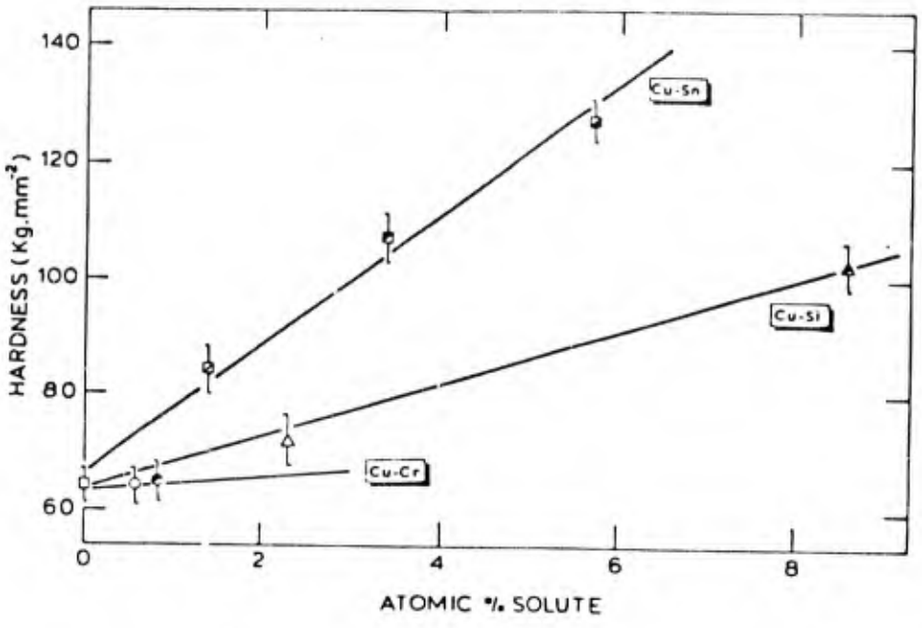


FIGURE 5.10 THE VICKERS MICROHARDNESS UNDER A 200 g LOAD VERSUS THE ATOMIC PERCENT OF SOLUTE. EACH POINT REPRESENTS AN AVERAGE OF FIVE MEASUREMENTS AND THE BARS REPRESENT THE STANDARD DEVIATION

An immediate consequence of the alloying was the change in the friction coefficient as shown in Figure 5.11. The harder Cu-Sn solution seems to have lower coefficients of friction than the Cu-Si solutions. Nevertheless, the difference in friction coefficient is not large enough to account for the large difference seen in wear rates which are shown in Figure 5.12. Figure 5.13 are SEM micrographs of some of the materials tested. The sliding direction in these micrographs is from the right to the left. The micrographs clearly show that delamination wear sheets have formed. It is also interesting to note that the top surface of the particles is very smooth. Figure 5.14 contains SEM micrographs of the subsurface of the specimens shown in Figure 5.13. The sliding direction is again from right to left. These micrographs indicate that cracks tend to propagate parallel to the surface and then extend to the surface. Although the angle of the crack, when the crack begins to propagate to the surface, is

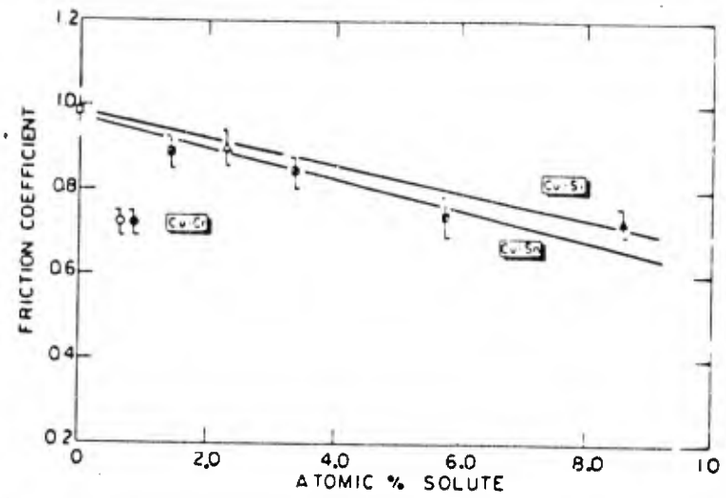


FIGURE 5.11 THE FRICTION COEFFICIENT AS A FUNCTION OF THE ATOMIC CONTENT OF SOLUTE. THE FRICTION COEFFICIENT WAS CALCULATED USING THE STEADY STATE TANGENTIAL FORCE. THE SLIDING SPEED WAS 2 m min^{-1}

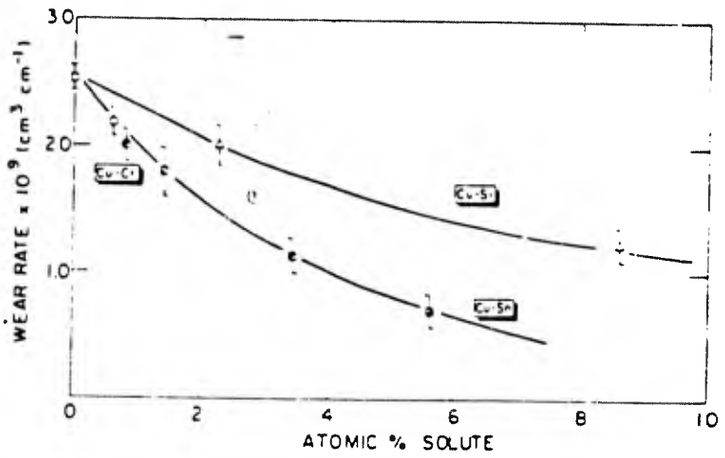


FIGURE 5.12 THE WEAR RATE AS A FUNCTION OF THE ATOMIC PERCENT OF SOLUTE. THE NORMAL LOAD WAS 2 kg AND THE DURATION OF THE TESTS WERE 100 MIN

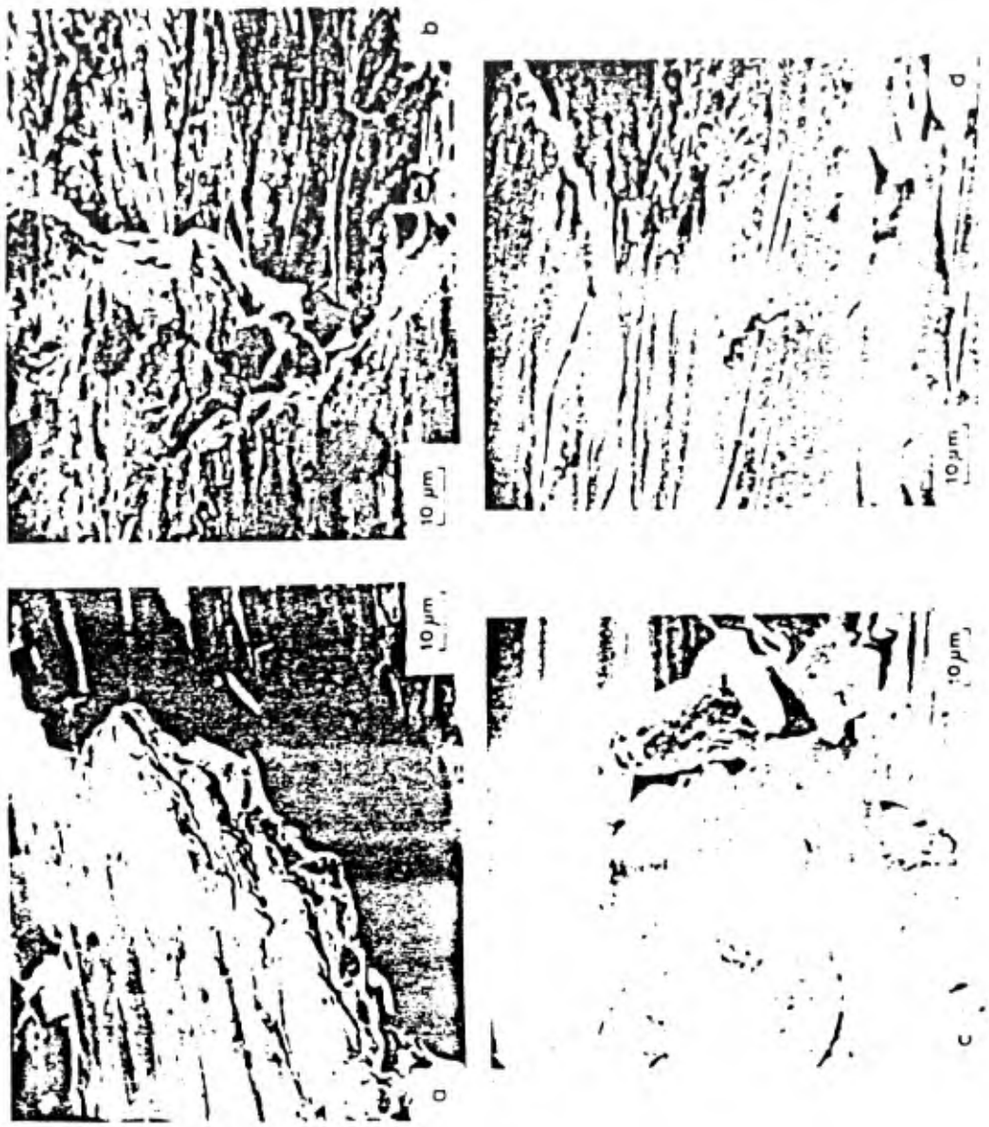


FIGURE 5.13 SCANNING ELECTRON MICROGRAPHS OF WEAR TRACKS (A) OFHC COPPER, (B) Cu-5.7 AT % Sn, (C) Cu-8.6 AT % Si AND Cu-0.81 AT % Cr. THE NORMAL LOAD WAS 2 kg AT A SLIDING SPEED OF 2 m min^{-1} AND THE SLIDING DISTANCE WAS 200 m

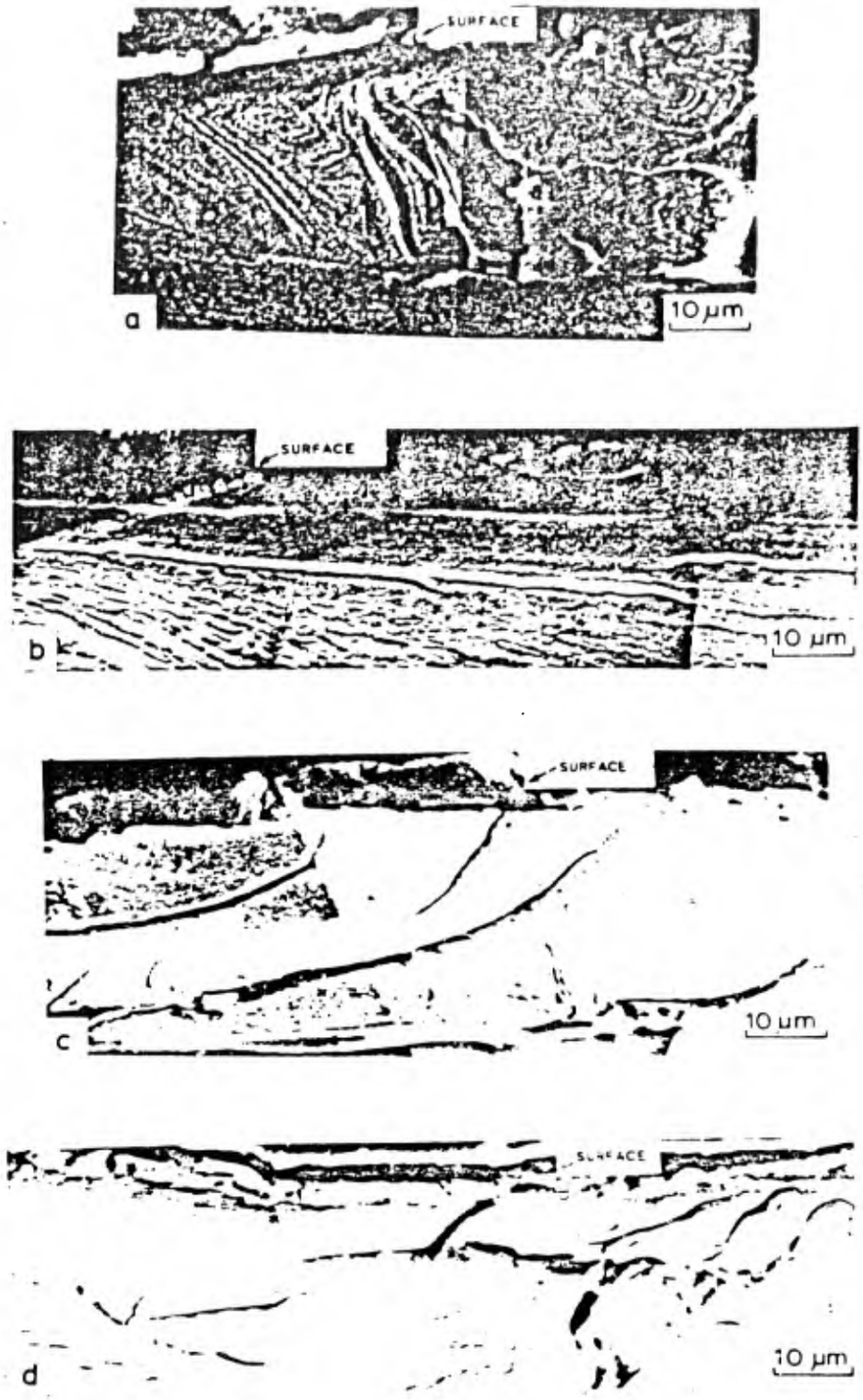


FIGURE 5.14 SCANNING ELECTRON MICROGRAPHS OF THE SUBSURFACE. THE MATERIALS AND TEST CONDITIONS ARE THE SAME AS IN THE PREVIOUS FIGURE

not always 70°, but it is interesting to note that most cracks change their direction as they approach the surface, more toward the direction perpendicular to the maximum normal stress direction. The depth at which cracks propagate seems to be different for specimens with different compositions. This difference is more striking between the micrographs of OFHC copper and Cu-5.7 a/o Sn solid solution; OFHC copper shows cracks propagating at a depth of the order of 50 μm, and the latter at a depth of the order of 15 μm.

The microscopic observations of the worn specimens show four facts: (a) that wear sheets are formed by cracking; (b) that there was extensive plastic deformation; (c) that there are subsurface cracks running parallel to the surface; and (d) that the cracks at the trailing end turn toward the surface. All of these observations are in accordance to the delamination theory of wear. These micrographs do not however, indicate how the cracks are initiated.

The effects of having a two phase structure on tribological behavior were investigated at MIT, using precipitation-hardened copper-chromium alloys (Cu-0.58 a/o Cr and Cu-0.81 a/o Cr) aged for different periods of time at 500° C[9]. The characteristics of the materials are given in Table 5.2. Figure 5.15 shows the variation in the hardness of the copper-chromium super-saturated solid solutions as a function of the aging time. The hardness of these materials initially increased with aging time and then decreased. The maximum value was reached after about 100 minutes of aging. The hardness is 65 kg/mm^2 for the solid solutions and 140 kg/mm^2 for the peak aged alloys. The aging time for the maximum hardness is different for the two alloys, whereas the maximum hardness is about the same.

Experimental materials

Parameter	Alloy	Aging time (min)		
		100	1 000	10 000
Volume fraction, $V_v \times 10^3$	Cu-0.58 Cr	5.19	5.25	5.31
	Cu-0.81 Cr	6.96	6.97	7.09
Mean free path λ (μm)	Cu-0.58 Cr	68.9	70.49	71.8
	Cu-0.81 Cr	51.84	53.02	53.00
Particle size d (μm)	Cu-0.58 Cr	0.54	0.55	0.58
	Cu-0.81 Cr	0.55	0.56	0.58

TABLE 5.2 EXPERIMENTAL MATERIALS

Figure 5.16 shows the friction coefficient and the wear rate as functions of the aging time. The friction coefficient is fairly constant for all treatment times for both alloys. The increase in hardness, resulting from the aging treatment, does rot seem to affect the friction coefficient, probably because the hardness of the slider was much greater than the specimens. The wear rate initially decreases by a factor of three for both Cu-Cr alloys and then increases approximately linearly; the slope seems to be the same for both alloys. The minimum wear rate does not correspond to the maximum hardness. (In these figures the peak hardness is indicated by arrows A and B for Cu-0.58 Cr and Cu-0.81 Cr, respectively). Figure 5.16 also gives a plot of the wear coefficient as a function of aging time, which shows that it increases rapidly after 5 minutes of aging and then levels off asymptomaticly to a constant value.

As the microstructural differences between the two Cu-Cr alloys are characterized by the volume fraction and the mean free path of particles, the wear resistance (the inverse of wear coefficient) is plotted as a function of the volume

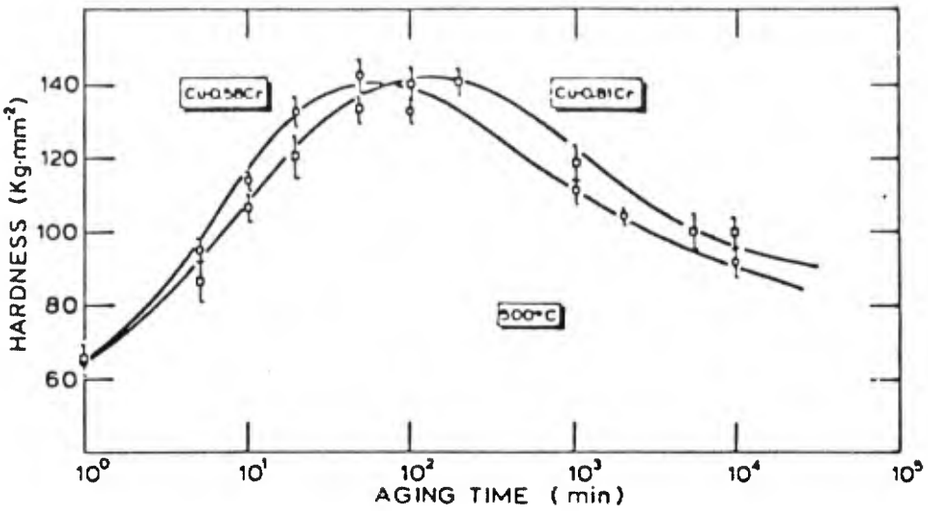


FIGURE 5.15 THE VICKERS MICROHARDNESS UNDER A 200 g NORMAL LOAD AS A FUNCTION OF AGING TIME. THE SPECIMENS WERE SUBJECTED TO AN AGING TREATMENT AT 500°C AND WERE WATER QUENCHED AT THE END OF THE TREATMENT

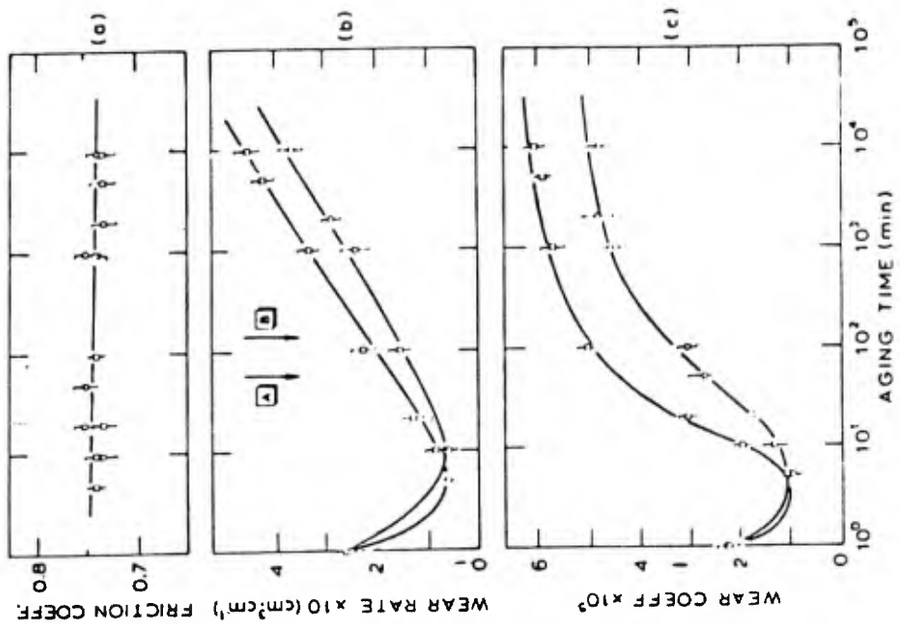


FIGURE 5.16 FRICTION AND WEAR PROPERTIES OF PRECIPITATION-HARDENED Cu-Cr ALLOYS AS A FUNCTION OF THE AGING TIME: (A) FRICTION COEFFICIENT, (B) WEAR RATE, AND (C) WEAR COEFFICIENT. THE NORMAL LOAD WAS 2 kg AND THE DURATION OF THE TESTS WERE 100 MIN AT A SLIDING SPEED OF 200 cm min^{-1}

fraction and the inverse of mean free path, as shown in Figure 5.17. The wear resistance decreases with an increase in the volume fraction and with the inverse of the mean free path of particles for the overaged alloys. These results are expected since the crack nucleation cites increases with the number of hard particles.

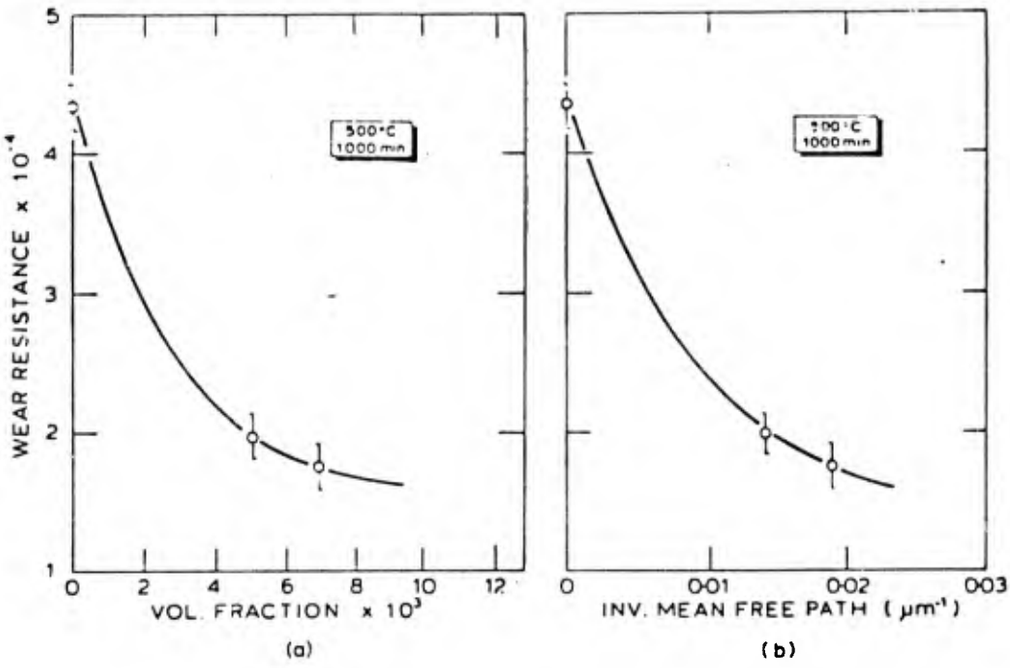


FIGURE 5.17 THE WEAR RESISTANCE (RECIPROCAL OF THE WEAR COEFFICIENT) VERSUS (A) THE VOLUME FRACTION AND (B) THE MEAN FREE PATH OF Cu-Cr ALLOYS FOR 1000 MIN OF TREATMENT

Figure 5.18 shows the micrographs of wear tracks of the precipitation-hardened alloys aged for 5 and 10,000 minutes. The sliding direction is from left to right. It can be seen that the surface details are similar to those shown earlier. The subsurface features for the same alloys are shown in Figure 5.19 where some second-phase particles can be seen in the overaged alloys. However, it is interesting to note that in Figure 5.19(a), subsurface cracks of the specimen aged for 5 minutes are very close to the surface, while for the specimen aged for 10,000 minutes, cracks are formed at a large depth. As an example of the morphology of wear particles, scanning electron micrographs of particles collected from the Cu-0.81 a/o Cr alloy aged for 10,000 minutes are shown in Figure 5.20. These particles are in the form of sheets and some have lamellar structure, possibly created by the presence of a large number of cracks above the crack that propagated the fastest.

These results with two phase metals show that the wear rate is also affected by the coherency of particles to the matrix. In the early stages of precipitation, the particles are coherent and therefore, the stress required to separate the particle from the matrix is large. Therefore, crack nucleation requires large amounts of subsurface deformation in order to develop sufficient interfacial stress between the matrix and the particle. Since the increased hardness decreases the deformation rate, the wear coefficient is decreased. The steep increase in wear coefficient after reaching a minimum value is due to the loss of coherency requiring less stress for decohesion of a particle from the matrix and is also due to the increased particle size. The deformation rate still decreases, but the deformation required for nucleation decreases even faster. Further, as the interparticle spacing decreases during the early stages of aging, cracks have to propagate smaller distances in order to join with neighboring cracks. This explains why the wear coefficient of the Cu-0.81% Cr alloy should be greater than that of the Cu-0.58% Cr alloy because the former has a larger volume fraction of second phase particles and possibly has a smaller interparticle spacing.

After the maximum hardness is reached, the volume fraction of particles does not increase with aging time anymore and coarsening of particles occurs, which increases the interparticle spacing. Crack nucleation at this point, tends to be relatively easy and the overall wear rate is controlled by the crack propagation rate. In this case, the crack growth and the interparticle spacing must be considered. Since the hardness does not change much for the overaged alloys, the number of asperities in contact must be nearly the

FIGURE 5.18 SCANNING ELECTRON MICROGRAPHS OF WEAR TRACKS OF THE PRECIPITATION-HARDENED Cu-0.81 AT % Cr ALLOY FOR AN AGING TIME OF (A) 5 MIN AND (B) 10000 MIN. THE SLIDING DIRECTION IS FROM LEFT TO RIGHT

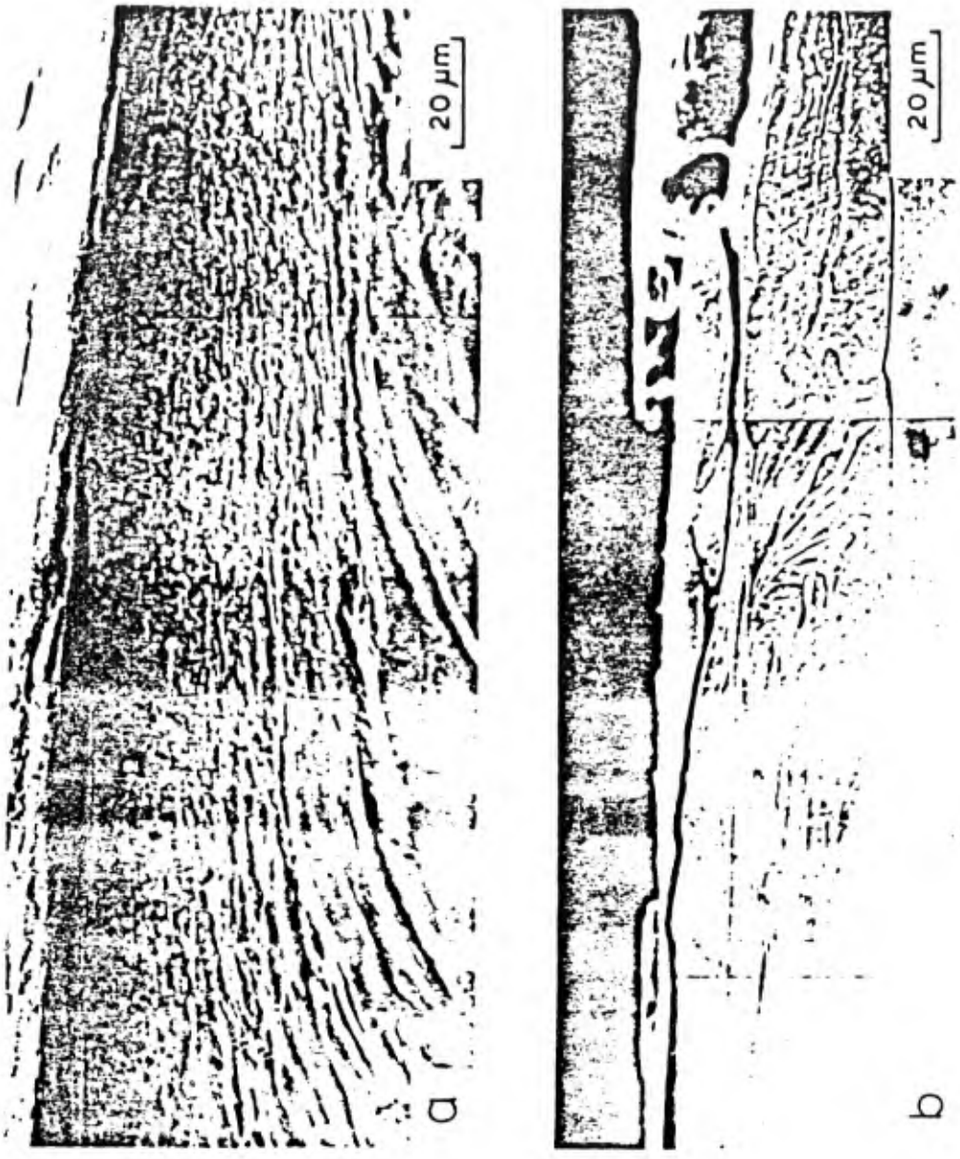


FIGURE 5.19 SCANNING ELECTRON MICROGRAPH OF THE SUBSURFACE OF PRECIPITATION-HARDENED Cu-0.81 AT % Cr ALLOY AGED FOR (A) 5 MIN AND (B) 10000 MIN

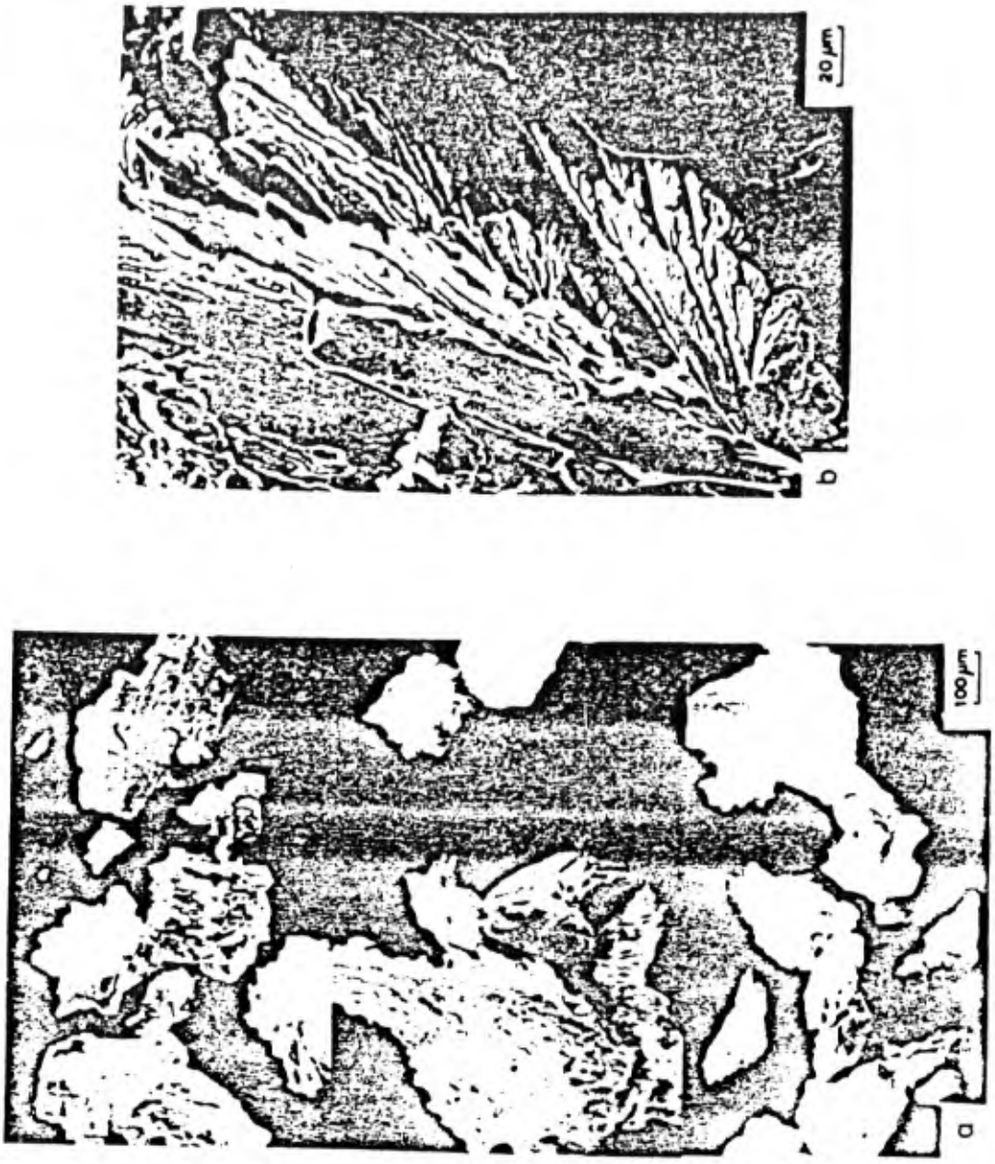


FIGURE 5.20 SCANNING ELECTRON MICROGRAPHS OF WEAR PARTICLES COLLECTED FROM WEAR TESTS ON THE SPECIMEN OF Cu-0.81 AT % Cr AGED FOR 10000 MIN

same and the number of asperities passing by will be about the same. Also since the matrices of both alloys are exactly the same, the crack growth rate must be the same to a first order approximation. Assuming that the material is delaminated in successive layers, the crack growth rate is calculated as the ratio of the mean free path to the number of cycles required to remove one layer. Such a calculation leads to a range of values between 4×10^{-4} and 6×10^{-4} μm/cycle for both alloys, and the wear rate depends basically on the mean free path of the interparticle spacing. Thus, when the overall wear process is controlled by the crack growth rate, the wear coefficient tends to level off toward asymptomatic values, since as the aging treatment is continued, the mean free path becomes roughly constant. This explains the leveling-off of the wear coefficient when the transition from crack nucleation rate controlled wear to crack propagation controlled wear occurs. The effect of the mean free path is also shown by the difference in the wear rate of Cu-0.81 a/o Cr and Cu-0.58 a/o Cr. As a higher volume fraction implies a smaller mean free path for the same particle size, the curve for Cu-0.81% Cr should be higher than that for Cu-0.58% Cr.

5.4 Sliding Wear of Polymeric Materials

5.4.1 Wear of Simple Phase Polymers

The basic thoughts involved in the delamination theory of wear can be applied to the wear of polymeric materials which have been used increasingly in bearing applications because of their low cost, excellent chemical resistance, and good wear properties.

Deformation properties of polymers have a profound influence on wear. Below the glass transition temperature, most amorphous polymers are relatively brittle. The sliding wear of glassy polymers is caused by crazing and brittle fracture when there is no thermal softening of the surface layer[11 - 15]. Semicrystalline polymers, which are usually used above the glass transition temperature and relatively ductile, show different wear behavior from the glassy polymers. Thick lump type wear debris or thin films are generated by plastic deformation of the surface layer in response to the applied normal and shear stresses[16 - 23, 19, 24 - 32]. Thin films are normally found on the counter surface face. The mechanism of the thin film transfer is not clear. The wear behavior depends upon the polymer properties and the sliding speed[19]. Thin films are formed only with highly linear crystalline polymers [such as polytetrafluoroethylene (PTFE), high density polyethylene (HDPE), polyoxymethylene (POM)], at slow sliding speed. The

thickness of the film is of the order of 50 to 200 A and the molecular chains are highly oriented in the sliding direction[25].

Many different explanations have been advanced to explain the mechanism of the thin film transfer. Steijn explained the thin film transfer phenomenon as resulting from the formation of adhesion junctions and the drawing of thin fims across the sliding surface[27]. Tanaka et al, argued that the thin film transfer of PTFE is due to easy destruction of the special banded structure, but not due to the drawing of molecular chains[29]. Makinson and Tabor claimed that the easy slipping of crystalline slices causes the thin film transfer[19]. According to Briscoe et al, the behavior of the thin film transfer at a low sliding speed, appears to be connected with a smooth molecular profile and not with the crystallinity or band structure of PTFE[25]. In the case of the crystalline polymers, that transfer thick lumpy films rather than thin films, little orientation was found in the film although equally strong adhesion junction might have been present. It is more probable that the reorientation occurs prior to the thin film transfer[16]. Then the thin film is formed by shearing of the oriented layer, which is followed by molecular orientation possibly due to the elongation of molecular chain during the shearing process.

A careful examination of worn surfaces reveals very large plastic deformation of the surface and subsurface layer, as discussed in Section 4. The deformation and the deformation gradient are largest at the surface and decays rapidly away from the surface. In crystalline materials such as metal crystals, grains deform in such a manner that the slip planes align nearly parallel to the surface. Therefore, it is expected that the surface layer will shear easily and fracture parallel to the surface due to the alignment of the slip planes with the surface. In crystalline polymers with no bulky side pendant groups, the crystalline region of the polymer is expected to align parallel to the surface. Consequently, the crystalline platelets are expected to shear off parallel to the surface and elongate when the surface is plowed by the asperities of the opposing surfaces. The peeling process may be facilitated if there are preexisting cracks or if cracks are nucleated during the deformation process.

In addition to the deformation of the surface layer, the deformation and fracture resulting from the bulk deformation of the surface due, to the Hertzian contact load in glassy polymers, must be considered in establishing the criteria for wear. When only a normal load is applied at a contact between

a sphere and a semi-infinite solid, the largest shear stress in the body occurs below the surface. The depth depends on the poisson ratio of the material and is 0.47 times the radius of the contact area (a) when poisson ratio is 0.330. When normal and tangential loads are present at the asperity contact, the maximum shear stress is about 0.343 Po at y = 0.4a where Po is the normal stress at the center of contact area and y is the coordinate axis perpendicular to the contact area[31, 32]. In glassy polmers, fracture can be initiated from these locations of maximum Hertzian shear stress. Therefore, in order to prevent wear, it will be necessary to insure that the applied load does not exert a stress greater than a critical value at the point where the Hertzian shear stress is maximum.

Based on the foregoing arguments, a wear model for polymers can be postulated using the yield criteria, Figure 5.21. The wear model divides the wear conditions into two regimes: the thin film transfer case (Regime I) and the thick film or lump transfer case (Regime II). In Regime I, the relationship between the applied shear stress and the material shear yield strength are so that the thin film transfer occurs when

$$\tau_s \geq k_s$$

and (5.13)

$$\tau_{max} < k_b$$

where τ_s is the tangential component of the surface traction due to various mechanisms discussed in Section 3, τ_{max} is the maximum Hertzian shear stress, k_s is the shear yield strength of the surface layer, and k_b is the bulk shear strength. In this regime thin film transfer occurs. For example, since PTFE and HDPE have much lower shear yield strength at the surface than in the bulk, they can plastically deform only at the surface layer under typical sliding conditions and generate thin films by the mechanism discussed earlier. In Regime II, both the shear stress at the surface and the maximum shear stress can exceed the material strength. Therefore, bulky wear particles will be generated when the following two conditions are satisfied:

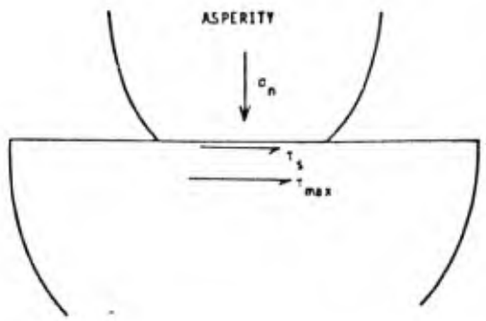
$$\tau_s \geq k_s$$

and (5.14)

$$\tau_{max} > k_b$$

As a result, plastic deformation may occur in the maximum shear region below the surface, generating lumpy wear particles.

The foregoing hypothesis has been checked by taking a polymer which belongs to Regime I and treating the surface so as to insure that the applied shear stress at the surface, τ_s, is less than the shear strength of the surface, k_s. This can be done by crosslinking the surface of polymers through helium plasma treatment.



Regime I : $\tau_s \simeq k_s$, $\tau_{max} < k_b$ — Thin Film Transfer

Regime II : $\tau_s \simeq k_s$, $\tau_{max} > k_b$ — Lumpy Thick Film Transfer

where σ_n = Normal Stress
τ_s = Shear Stress at the Surface
τ_{max} = Maximum Shear Stress in the Bulk
k_s = Surface Shear Yield Strength
k_b = Bulk Shear Yield Strength

FIGURE 5.21 CONDITIONS FOR WEAR OF POLYMERS BY THIN FILM TRANSFER IN CRYSTALLINE POLYMERS & BY LUMPY THICK FILM TRANSFER

5.4.2 Wear of Polymeric Composites

Many composites are made of polymers reinforced with fibers. Many bearings and gears made of phenolics, nylon, polyurethane, teflon (PTFE), and acetal have been reinforced glass fibers and graphite fibers for improved wear resistance and dimensional stability. These composites with chopped fibers behave in such a manner that their tribological behavior can be predicted and improved based on the delamination theory of wear. Two of these studies will be briefly reviewed here.

Clerico tested nylon reinforced with randomly oriented chopped glass fibers[33]. She found that these composites delaminate by plastic deformation and subsurface crack nucleation and propagation. The wear particles collected were thin sheets.

Sung and Suh investigated the effect of fiber orientation on wear based on the delmaination theory of wear[34]. Three types of fiber reinforced plastics were studied. One was a unidirectionally oriented composite with graphite fibers and an epoxy matrix, fabricated from prepegs. The second composite was composed of unidirectionally oriented Kevlar-49 fibers in epoxy. The third composite material was a commercial grade bearing material (RT Duroid 5813), a polytetrafluoroethylene (PTFE) based material. This composite contained about 16 wt.% of planar oriented micro-glass fibers with two-thirds of the fibers being preferentially oriented in one direction. The ratio of fibers oriented in three orthogonal directions are approximately 2:1:0. The composition and the properties of these composites are given in Table 5.3.

Properties of composites	
Unidirectional graphite fiber (Thornel 300)-epoxy (SP-228)	
Specific gravity	1.58
Fiber volume	60%
Longitudinal tensile strength	200 klbf in^{-2}
Longitudinal tensile modulus	18 x 10 lbf_2 in^{-2}
Compression strength	145 klbf in^{-2}
Transverse tensile strength	10 klbf in^{-2}
Transverse tensile modulus	1.2 x 10 lbf in^{-2}
Interlaminar shear strength	16 klbf in^{-2}
Unidirectional Kevlar-49 (DuPont)-epoxy	
Specific gravity	1.33
Fiber volume	65%
Longitudinal tensile strength	223 klbf in^{-2}
Longitudinal tensile modulus	17 x 10 lbf in^{-2}
Interlaminar shear strength	9 klbf in^{-2}
Bidirectional glass microfiber-MoS_2-PTFE (RT Duroid[a])	
Specific gravity	2.42
Glass fiber	16 wt.%
MoS_2	15 wt.%
Compression modulus[b]	158/183/130 klbf in^{-2}
Compression strength[b]	3.6/4.4/8.2 klbf in^{-2}

a Registered trademark of Rogers Corporation
b Three values are for the three orthogonal directions.

TABLE 5.3 PROPERTIES OF COMPOSITES

The experimental results show that the wear rate is the least when the fibers are placed perpendicular to the surface as shown in Figure 5.22 for the graphite fiber-epoxy composite. In this case, the friction was also the least as shown in Figure 5.23. Similar wear results were obtained with the Kevlar-epoxy composite, Figure 5.24, although the order of frictional force was changed. With the biaxially oriented glass micro-fiber - MoS_2 - PTFE composites the wear rate was again found to be the least when the largest fraction of fibers was oriented normal to the sliding plane Figure 5.25. These results indicate that when the fibers are normal to the surface, plastic deformation and crack propagation are both minimized, resulting in low wear rates.

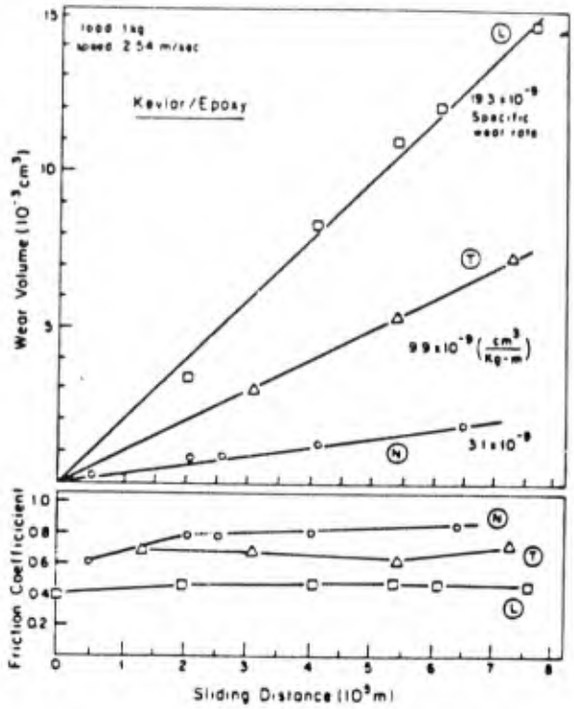


FIGURE 5.22 FRICTION COEFFICIENTS & WEAR VOLUME AS A FUNCTION OF SLIDING DISTANCE IN UNIAXIAL GRAPHITE FIBER-EPOXY COMPOSITE. SLIDING AGAINST 52100 STEEL, WITH FIBER ORIENTATIONS NORMAL LONGITUDINAL & TRANSFER TO THE SLIDING DIRECTION.

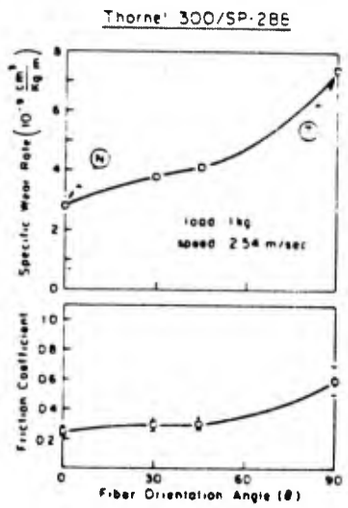


FIGURE 5.23 FRICTION COEFFICIENTS & SPECIFIC WEAR RATE AS A FUNCTION OF SLIDING DISTANCE IN UNIAXIAL GRAPHITE FIBER-EPOXY COMPOSITE SLIDING AGAINST 52100 STEEL, WITH VARYING FIBER ORIENTATIONS RANGING FROM NORMAL (θ=0) TO TRANSVERSE (θ=90°) TO THE SLIDING DIRECTION

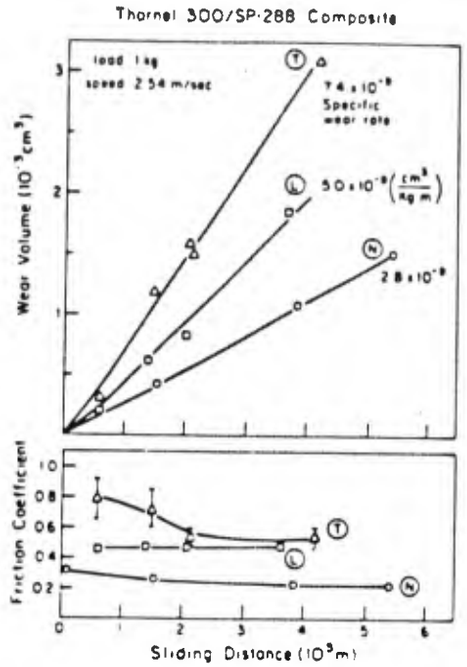


FIGURE 5.24 FRICTION COEFFICIENTS & WEAR VOLUME AS A FUNCTION OF SLIDING DISTANCE IN UNIAXIAL KEVLAR 49-EPOXY COMPOSITE SLIDING AGAINST 52100 STEEL WITH FIBER ORIENTATIONS NORMAL (N), LONGITUDINAL (L), AND TRANSVERSE (T), TO THE SLIDING DIRECTION

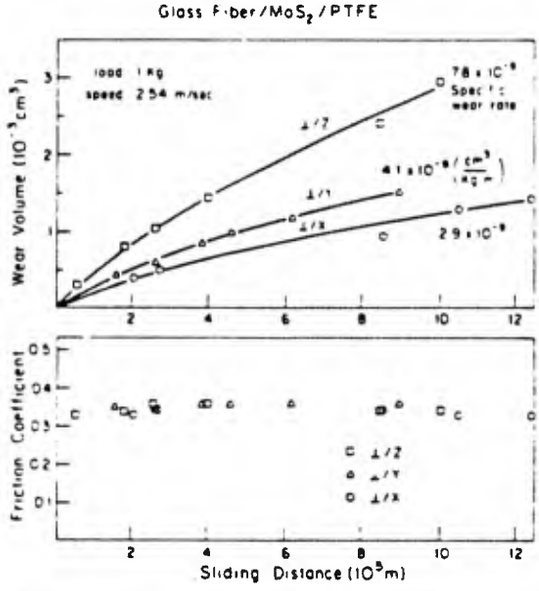


FIGURE 5.25 FRICTION COEFFICIENTS & WEAR VOLUME AS A FUNCTION OF SLIDING DISTANCE IN BIAXIALLY ORIENTED GLASS MICROFIBER MoS_2-PTFE COMPOSITE, SLIDING AGAINST 52100 STEEL WITH SLIDING PLANES NORMAL TO THREE ORTHOGONAL DIRECTIONS, X, Y, AND Z

5.5 Wear Minimization

A number of techniques have been devised to minimize the sliding wear of materials based on the ideas suggested by the delamination theory of wear. They are listed below without much elaboration. The details can be found in the references given.

1) Coating of the metal surface with a thin layer of soft metals. The thickness of the layer, which is very critical, is about 0.1 μm for nickel and 10 μm for gold. Such a thin layer prevents plastic deformation and thus, reduces wear[35].

2) Cross-linking of crystalline polymers to prevent the film transfer process by creating a strong layer on the surface so that the frictional force is always less than the strength of the surface layer[36].

3) Ion implantation of iron with titanium ion and nitrogen ion. This process creates a hard layer on the surface which prevents plowing, thus lowering the frictional force and the wear rate[37].

5.6 Concluding Remarks

In this section it was shown that elasto-plastic materials, be it a polymer or metal, wears by the delamination process. The delamination theory of wear explains the sliding wear process in detail and correctly. Wear of sliders is caused by plastic deformation, crack nucleation, and crack propagation.

Wear particles are also generated by asperity deformation and plowing, which accelerate the delamination process. However, in many cases the ultimate failure of the sliding surfaces is by the mechanisms postulated by the delamination theory of wear. This is because the dominant mechanism that controls the wear rate is the delamination mechanisms.

Sliding wear under high loads and speeds can also be caused by other mechanisms because of the high interfacial temperatures. This is a subject by itself.

5.7 References - Section 5

1. Suh, N.P., "The Delamination Theory of Wear," Wear, 25 (1973) 111-124.

2. Jahanmir, S., N.P. Suh, and E.P. Abrahamson, II, "Microscopic Observations of the Wear Sheet Formation by Delamination," Wear, 28 (1974) 235-249.

3. Jahanmir, S., "A Fundamental Study of the Delamination Theory of Wear," Ph.D. Thesis, Department of Mechanical Engineering, MIT, 1977.

4. Tohaki, M., "Microstructural Aspects of Friction," S.M. Thesis, Department of Mechanical Engineering, MIT, 1979.

5. Rabinowicz, E., "The Dependence of the Abrasive Wear Coefficient on the Surface Energy of Adhesion," Wear of Materials - 1977, ASME, New York, 1977, 36-40.

6. Gupta, P.K. and N.H. Cook, "Statistical Analysis of Mechanical Interaction of Rough Surfaces," Journal of Lubrication Technology, 94 (1972) 19-26.

7. Whitehouse, D.J., "The Effects of Surface Topography on Wear," Fundamentals of Tribology, (Eds. N.P. Suh and N. Saka), MIT Press, 1980, 17-52.

8. Pamies-Teixeira, J.J., N. Saka, and N.P.. Suh "Wear of Copper-Based Solid Solutions," Wear, 44 (1977) 65-75.

9. Saka, N., J.J. Pamies-Teixeira, and N.P. Suh, "Wear of Two-Phase Metals," Wear, 44 (1977) 77-86.

10. Saka, N., "Effect of Microstructure on Friction and Wear of Metals," Fundamentals of Tribology, (Ed. N.P. Suh and N. Saka), MIT Press, 1980, 135-170.

11. Puttick, K.E., L.S.A. Smith, and L.E. Miller, "Stress Field Round Indentation in Poly (Methyl Methacrylate)," J. Phys. D.; Appl. Phsy., 10 (1977) 617-632.

12. Billinghurst, P.R., C.A. Brookes, and D. Tabor, "The Sliding Processes as a Fracture-Inducing Mechanism," Proceedings of the Conference on the Physical Basis of Yield and Fracture, Oxford 1966, 253-258.

13. Matsushige, K., S.V. Radcliffe, and E. Baer, "The MechanicalBehavior of Polystyrene Under Pressure," J. Mat. Sci., 10 (1975) 833-845.

14. Van Den Boogaart, A., "Crazing and Characterization of Brittle Fracture in Polymers," Proceedings of the Conference on the Physical Basis of Yield and Fracture, Oxford (1966), 167-175.

15. Peterson, T.L., D.B. Ast, and E.J. Kramer, "Holographic Interferometry of Crazes in Polycarbonate," J. Appl. Phys. 45 (1974) 4220-4228.

16. Pooley, C.M. and D. Tabor, "Friction and Molecular Structure: The Behavior of Some Thermoplastics," Royal Soc. London, 329A (1972) 251-274.

17. Bowers, R.C., W.C. Clinton, and W.A. Zisman, "Frictional Properties of Plastics," Modern Plastics, Feb. (1954), 131-225.

18. Tanaka, K. and Y. Uchiyana, "Friction, Wear, and Surface Melting of Crystalline Polymers," Wear of Materials, ED. S.K. Rhee et. al., ASME (1977), 499-530.

19. Makinson, R.K. and D. Tabor, "The Friction and Transfer of Polytetrafluoroethylene," Proc. Roy. Soc. London, A281 (1964) 49-61.

20. West, C.H. and J.M. Senior, "Frictional Properties of Polyethylene," Wear, 19 (1972) 37-52.

21. James, D.I., "Surface Damage Caused by Polyvinyl Chloride Sliding on Steel," Wear, 2 (1958/59) 183-194.

22. Jain, V.K. and S. Bahadur, "Material Transfer in Polymer-Polymer Sliding," Wear of Materials, ASME (1977), 487-493.

23. Warren, J.H. and N.S. Eiss, Jr., "Depth of Penetration as a Predictor of the Wear, Polymers on Hard, Rough Surfaces," Wear of Materials, ASME (1977), 494-500.

24. Kar, M.K. and S. Bahadur, "Macromechanism of Wear at Polymer-Metal Sliding Interface," Wear of Materials, ASME (1977), 501-509.

25. Briscoe, B.J., C.M. Pooley and D. Tabor, "Friction and Transfer of Some Polymers in Unlubricated Sliding," Advances in Polymer Friction and Wear, Polymer Science and Technology Vol. 5A, Ed, H. Lee, Plenum Press (1974), 191-204.

26. McLaren, K.G. and D. Tabor, "The Friction and Deformation Properties of Irradiated Polytetrafluoroethylene (PTFE)," War, 8.(1965) 3-7.

27. Steijn, R.P., "The Sliding Surface of Polytetrafluoroethylene, on Investigation with Electron Microscope," Wear, 8 (1968) 193-212.

28. Sviridyonok, A.I., V.A. Bely, V.A. Smurigov, and V.G. Savkin, "Study of Transfer in Frictional Interaction of Polymers," Wear, 25 (1973) 301-308.

29. Tanaka, K., Y. Uchiyaman, and S. Toyooka, "The Mechanism of Wear of Polytetrafluoroethylene," Wear, 23 (1977) 153-172.

30. Timoshenko, S.P. and J.N. Goodier, "Theory of Elasticity," McGraw-Hill (1970), 409-414.

31. Poritsky, H., "Stresses and Deflections of Cylindrical Bodies in Contact with Application to Contact or Years and Locomotive Wheels," J. Appl. Mech., June (1950), 191-201.

32. Hamilton, G.M. and L.E. Goodman, "The Stress Field Created by a Circular Sliding Contact." J. Appl. Mech., June (1966), 371-376.

33. Clerico, M., "Sliding Wear of Polymeric Composites," Wear, 53 (1979) 279-301.

34. Sung, N.H. and N.P. Suh, "Effect of Fiber Orientation on Friction and Wear of Fiber Reinforced Polymeric Composites," Wear, 53 (1979) 129-141.

35. Saka, N., H.C. Sin, and N.P. Suh, "Prevention of Spline Wear by Soft Metallic Coatings," Final Report to DARPA and ONR, Contract No. N00014-76-C-0068, July, 1980.

36. Youn, J. and N.P. Suh, "Tribological Characteristics of Surface Treated Polymers," SPE ANTEC, (1981) 20-23.

37. Shepard, S.R. and N.P. Suh, "The Effects of Ion Implantation on Friction and Wear of Metals," ASLE/ASME paper to appear in Transactions, 1981.

6. CONCLUSIONS

Surface interactions are complex phenomena, requiring input from materials science, continuum mechanics, thermodynamics, and quantum mechanics. The surface is always highly contaminated because of the high free energy of the surface. The atoms of the contaminants are well mixed with the substrate atoms, again to lower the free energy.

Therefore, adhesion cannot occur readily under quasi-static conditions at low temperatures.

Friction is caused by plowing, adhesion, and asperity deformation. Therefore, it depends strongly on environment. Through proper design of the interface and materials, friction can be controlled. It was shown that friction must be considered in the Friction Space, since it is not an invariant in the temporal frame of tribological applications.

Finally, the response of materials are shown to be plastic deformation, crack nucleation, and crack propagation. Based on detailed considerations of each one of these processes, the delamination theory of wear was introduced. It is shown that although there are many different wear particle generating mechanisms, the delamination process is the rate determining process.

7. ACKNOWLEDGEMENTS

I am grateful to the financial support provided by the Defense Advanced Research Projects Agency and the Office of Naval Research over the years to conduct research on the Delamination Theory of Wear. Much of the work reported here was done through the partnership between my students, colleagues and myself. I am particularly indebted to Drs. S. Jahanmir, N. Saka, and H.C. Sin for their contributions to our research efforts. Thanks are also due Mrs. Margaret McDonald for her patience in typing the manuscript.

MATERIALS IN TRIBOTECHNICAL APPLICATIONS

A.W.J. de Gee
Metaalinstituut TNO, Apeldoorn, The Netherlands

1. INTRODUCTION

When set to the task of discussing the role of materials in tribology, one is immediately faced with the need to decide whether to follow a "theoretical model approach" or a "systems approach"[1, 2]. The first approach assumes a consistent behavior of materials under a variety of application conditions, the second allows for appreciable deviations in behavior from one application to the other.

Without rejecting the "model approach" all together (see below), in the present case the "systems approach" is chosen, if only because of the fact that the wear rate of a particular, well defined material may differ by several decades from one application (i.e. set of conditions) to the other. In the following, two examples are given of cases in which relatively minor changes in conditions result in a considerable - and a priori unpredictable change in tribological behavior.

The first example of the essential unpredictability of tribological processes is shown in Figure 1. This figure shows the pronounced influence of a change in dimensions of contacting bodies on their wear rates[3]. In the present case this is caused by a pronounced difference in friction induced surface temperature, as a result of the change in dimensions. Actually, the wear process changes radically from one geometry to the other. If a pin with R = 2 mm is combined with a ring with R = 40 mm, the wear mechanism is that of adhesive wear. If, on the other hand, a pin with R = 4 mm is combined with a ring with R = 20 mm, the wear mechanism is that of oxidative wear.

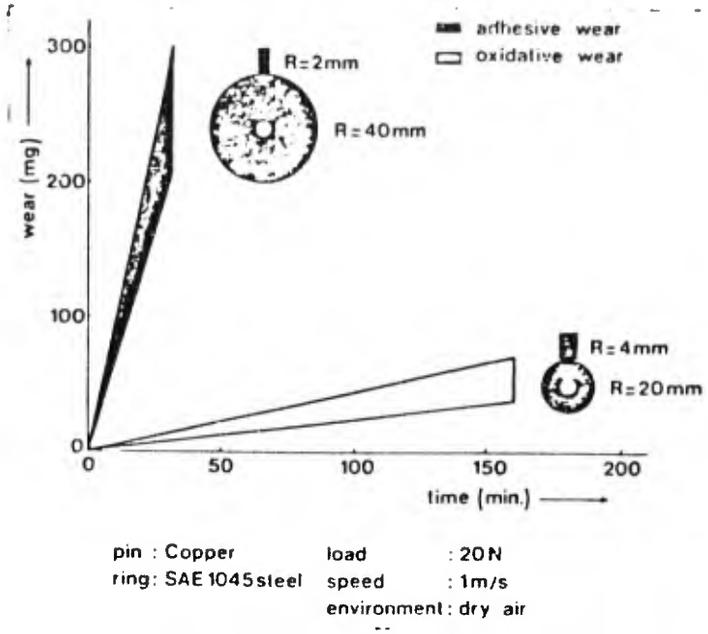


FIGURE 1 THE INFLUENCE OF A CHANGE IN DIMENSIONS OF CONTACTING BODIES ON THEIR WEAR RATES

The second example concerns the influence of the chemical composition of the bearing material and that of the lubricant cover gas on the wear of Cu 6 Sn alloys, Figure 2, used in combination with shafts of, respectively, carburized steel AISI 5120 and through hardened steel AISI D3[4]. Copper-tin alloys usually contain small quantities of phosphorous (up to 0.1%) as a result of the fact that the melt is deoxidized with "copper-phosphorus" (copper with approximately 10% Cu_3P). If oxygen from the air penetrates into the melt, phosphorous is oxidated preferentially, thereby protecting the metals against oxidation. The phosphorous oxide, thus formed, can be found in the slag. If desired, phosphorus-free alloys can be obtained by melting together pure metals under vacuum (no oxidation of the metals), by exposing a phosphorous containing melt for some time to the air, thereby removing the surplus of phosphorous by oxidation or by depositing the bearing material on a steel substrate by a process of metal spraying under controlled oxidation conditions. In the first instance the resulting alloy contains neither phosphorous nor tin oxide. In the latter two cases however, a very small amount of tin oxide is formed, which - after solidification - is present in the alloy in the form of finely divided crystals with a diameter from 0.1 to 0.5 μm.

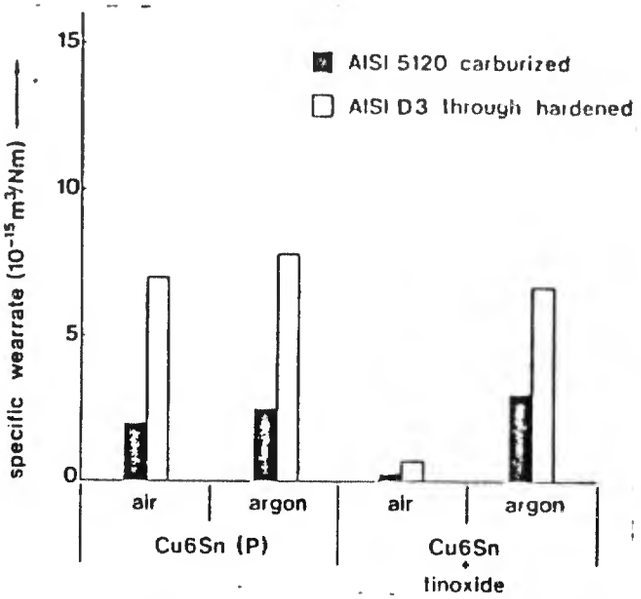


FIGURE 2 THE INFLUENCE OF THE CHEMICAL COMPOSITION OF BEARING MATERIAL AND LUBRICANT COVER GAS ON THE WEAR OF Cu 6 Sn ALLOYS

Laboratory tests with a tribometer as well as full scale bearing tests showed that these tin oxide containing materials behave exceptionally well, provided that the environment contains oxygen[5]. This effect is shown in Figure 2. The explanation of the pronounced differences in behavior is that tin oxide catalyses the formation of surface active, polymolecular products ("friction polymers") in the contact zone. Such polymer formation is possible only by virtue of the fact that the lubricant contains suitable "building stones." In the present case, such "building stones" are free radicals, which are formed in the lubricant by reaction with oxygen from the cover gas. Figure 2 also shows that the nature of the steel also has a pronounced effect on the wear of the Cu Sn alloy (although the surface roughness of the shafts was the same in both cases). Still the nature of the shaft does not affect the phosphorous/tin oxide effect.

There would not have been the remotest possibility of predicting a priori the above effects on the basis of theoretical models.

Although a strong case for a systems approach can be built upon the above evidence, this does not mean that theoretical modeling would not be worthwhile and - in fact - a necessary activity in tribology. Actually, it is only by virtue of such models that effects as discussed above can be explained. Also, theoretical models may show how far observations can be extrapolated to "related systems" (for instance: if the mechanism by which lauric acid functions in boundary lubrication

is understood, see Figure 3, one can safely predict that stearic acid will have similar effect, although it will still be hard to predict the minimum concentration at which such effect manifests itself for the first time). An obvious area in which the principle of "extrapolation by virtue of theoretical modeling" is applied widely is that of improvement of the properties of a particular class of materials, as for instance deformation - martensite forming manganese steels[5].

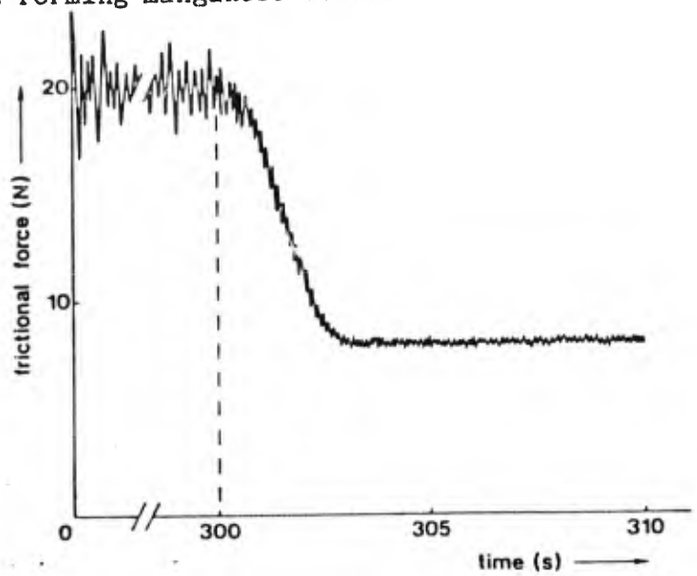


FIGURE 3 EFFECT OF ADDITION OF LAURIC ACID ON THE FRICTION OF BRONZE AGAINST STEEL

LUBRICATION CONDITION: BOUNDARY
300 s AFTER BEGINNING OF THE TEST 1 cm^3 of LAURIC ACID ($C_{11}H_{23}COOH$) IS ADDED TO THE LUBRICANT BATH (1000 cm^3 SAE 10 BASE OIL).

Having chosen for a systems - (application directed) approach for a discussion of "materials in tribology", practically meaningful application areas are to be defined.

In the present case a choice has been made for the following areas:

- lubricated plain bearings
- rolling element bearings, gears, cams and cam followers
- dry bearings
- earth moving, mining and dredging equipment.

The above four application areas cover a considerable percentage (80% at a rough estimate) of the technical areas where tribology is important. In addition, the tribological aspects of surface treatments and coatings are covered

separately, as such treatments and coatings may be applied in widely different application areas.

In order to promote a unified approach towards the behavior of materials in the above application areas, a brief survey of wear mechanisms and their interrelation precedes the detailed discussion of the separate areas.

1.1 Wear Processes

Wear is usually defined as the undesirable, progressive loss of substance from the operating surface of a body, occurring as a result of relative motion at the surface. It can be due to the removal of material as a result of frictional heating (wear by melting or evaporation) or to the continuous removal of reaction products (e.g. oxides) from the surface (chemical, corrosive or oxidative wear). In all other cases, wear is the result of local overstressing of one or both of the contacting bodies, due to normal and tangential (friction) forces. This may be (but is not necessarily) accompanied by transfer of material from one body to the other. When transfer occurs, the process is identified as one of adhesive wear; otherwise the term delamination wear is applicable. Finally, if the action of an abrasive is involved, the designations abrasion or abrasive wear are used.

If tangential forces are low, as is usual in rolling element bearings, the contacting surfaces may still suffer from surface fatigue, leading to wear as a result of cyclic variations in the normal force. If cyclic loading in a direction perpendicular to the surfaces is due to impact, the term impact wear is used. Finally, in erosion a solid surface suffers from attack by solid particles carried in a fluid. Depending upon the angle of attack, either abrasion or impact wear may predominate.

Figure 4 gives a survey of the different wear processes that may occur under conditions of sliding motion between two contacting bodies and shows how these relate to friction.

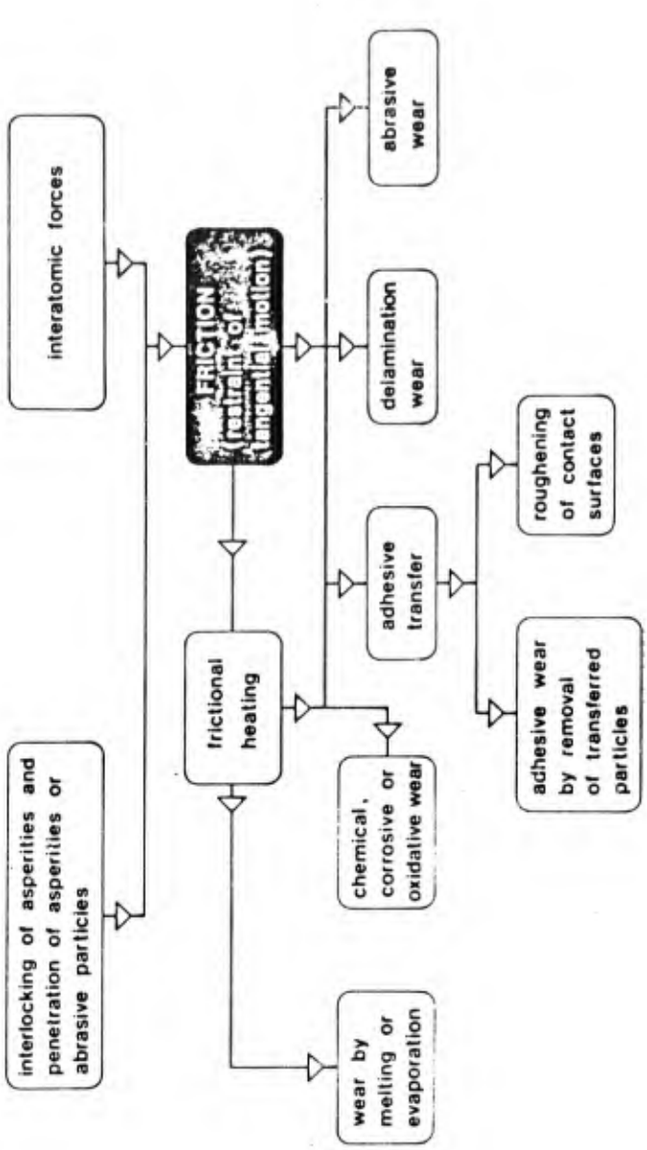


FIGURE 4 SURVEY OF WEAR PROCESSES THAT MAY OCCUR UNDER CONDITIONS OF SLIDING MOTION BETWEEN TWO CONTACTING BODIES AND THEIR RELATION TO FRICTION

A consequence of the fact that, usually, wear is caused by overstressing of the surface zone of materials due to normal and tangential forces, is that friction (as resulting from adhesive forces and plowing) and the dynamic strength properties of the materials, that are subjected to friction, are equally important in determining the wear rate. Although from this it seems reasonable to adhere to the statement "no wear without friction"

it is also clear that in general there will be no unambiguous relation between friction and wear. This is because of the fact that - from one material combination to the other - a difference in friction may or may not be balanced by a difference in dynamic strength properties.

2. MATERIALS FOR LUBRICATED PLAIN BEARINGS

2.1 Non-Tribotechnical Versus Tribotechnical Quality Criteria

In selecting materials for use in lubricated plain bearings, one usually applies non-tribological as well as tribological selection criteria. The first group comprises hardness (and the related properties compressive strength and deformability), the thermal properties heat conductivity, specific heat, maximum admissible surface temperature and coefficient of thermal expansion, the electrical properties (i.e. in particular the resistance against electrical discharge pitting) and, last but not least, the price as related to availability, machineability, etc. A general property of these non-tribological criteria is that they are - at least in a first approximation - system-independent. A non-tribological criterion which does depend strongly on the systems properties (i.e. in particular the environment and the temperature) is the resistance of the bearing material against corrosive action of the lubricant (particularly important if an "unconventional" liquid as, for instance, water is used).

In contrast to the above, quality criteria of a typical tribological nature are always heavily system-dependent. This means, among other things, that in practice one can never characterize a single bearing material as such. Instead one should always characterize the combination of bearing material, lubricant and journal material. This applies in particular if unconventional (nonferrous) journal materials and/or unconventional lubricants are to be used.

International cooperative work under the auspices of the International Standards Organization, ISO, has led to the definition of four main categories of tribological selection criteria, namely:

- behavior under conditions of boundary lubrication
- behavior under conditions of surface fatigue
- behavior under conditions of (fluid) erosion
- behavior under conditions of cavitation erosion.

2.2 Tribological Quality Criteria

Figure 5 gives a survey of the criteria which, together, determine the tribological behavior of bearing materials and their interrelation. In the following paragraph, Figure 5 is discussed in detail.

When considering behavior under conditions of boundary lubrication (A), one distinguishes between behavior under normal conditions of lubricant supply, (very) low speed and low thermal loading (A1), and the behavior under conditions of interrupted lubricant supply, nominal speed and high thermal loading (A2). The first set of conditions is encountered in plain bearings which normally function under conditions of hydrodynamic lubrication, but which are run at speeds far below nominal speed (as, for instance, during starting and stopping). Under such conditions, the development of frictional heat is negligible and the bearing runs under psuedo-isothermal conditions. The latter situation occurs in bearings in which the supply of lubricant stagnates while the bearing runs at nominal speed. Under such conditions, there is considerable development of frictional heat and the temperature of the bearing rises rapidly. In both situations, A1 and A2, the lubricant may or may not be contaminated with abrasive particles ("dust"). This leads to situations A1-1, A1-2, A2-1, and A2-2. In all these cases, the behavior of the combination of bearing material - lubricant - journal material can be characterized with three parameters, namely the coefficient of friction (a), the process roughness or, better, the contact parameter (b), and the wear rate (c).

A major difference between situations A1 and A2 is that in situation A1 the system rapidly attains equilibrium, the characteristic parameters coefficient of friction, process roughness and wear rate attaining constant values. In situation A2 on the contrary drastic transition effects are bound to occur. In this case it is not so much the high values of the characteristic parameters after reaching equilibrium which are important, but much more the time which elapses before a transition in one or more of the characteristic parameters occurs. Accordingly, problems associated with stagnation of the lubricant supply at nominal speed are usually solved by application of a suitable monitoring system, for instance based on continuous measurement of the bearing temperature. On the basis of this philosophy a better material-lubricant combination means a combination with a longer delay time.

Further particulars on determination and use of coefficient of friction, process surface roughness (or contact parameter) and wear rate can be found in References 7, 8, and 9.

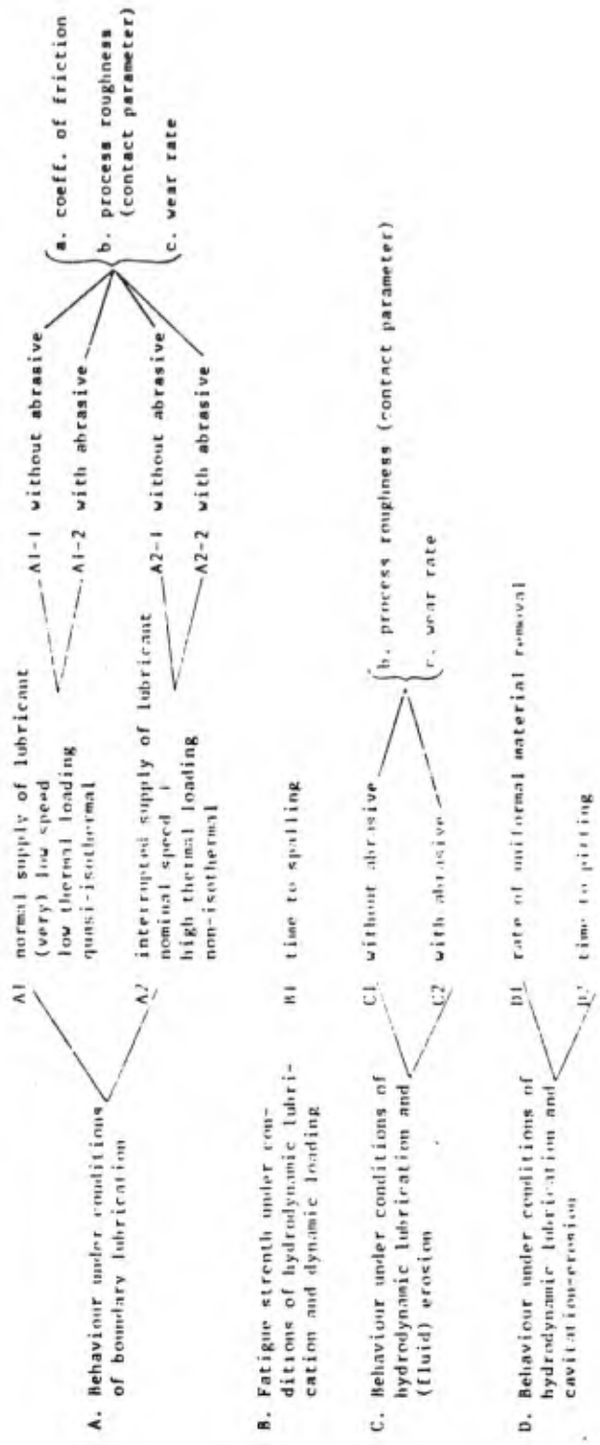


FIGURE 5 SURVEY OF THE CRITERIA WHICH DETERMINE THE TRIBOLOGICAL BEHAVIOR OF BEARING MATERIALS AND THEIR INTERRELATION

2.3 Typical Journal Bearing Materials and Their Performance Characteristics

Materials, particularly suited for application in plain bearings, should be able to carry the load on the bearing (no gross plastic deformation), should have an adequate fatigue life, should be readily deformable to accommodate misalignments, and should be able to embed dust particles. Some of these properties are conflicting. This applies for instance for fatigue life versus embedability. The first generally requires application of the bearing material in thin layers (the normal forces are then effectively carried by the hard steel substrate) while embedability, in particular of coarser particles, requires a considerable layer thickness.

According to the above, material combinations which are suitable for use in plain bearings, which occasionally have to function under conditions of solid-to-solid contact, preferably should possess low values of coefficient of friction, process surface roughness, and wear rate as well as adequate (fatigue) strength properties.

Metallic bearing materials, which more or less fulfill these requirements, usually contain elements which either do not form solid solutions with iron, the main constituent of most journals, or which form brittle intermetallic compounds with iron. The first fact is connected with the widely used, though controversial, rule which states that the tendency towards adhesion is the lowest for pairs of metals with almost zero mutual solubility, i.e. for metals which are the least compatible in a metallurgical sense. This point of view, which has been defended most strongly by Rabinowicz, has led to the formation of "compatibility chart" for various metal combinations, derived from binary diagrams of the respective elements[10].

In weighing the arguments of supporters and opponents of the "solubility rule," it should first be pointed out that strong forces between atoms of different types frequently lead to the formation of intermetallic compounds, thereby reducing rather than increasing solid solubility. In that case there will be no correlation between solid solubility and tendency towards adhesion. Further, it should be noted that the evidence which led to the design of the compatibility chart, without exception, concerns technical, that is, contaminated systems, in which oxide films and - in some cases - lubricants were present. This implies that, under the relevant conditions, contact between the virginal metal sufaces was largely inhibited by interfacial layers. This seems to be a crucial

factor because experiments in a vacuum showed beyond doubt that very clean surfaces of virtually insoluble metals such as silver and iron easily form strong adhesive junctions[11]. The above leads to the conclusion that the solubility rule is of practical consequence only for (lightly) contaminated surfaces, for which the interatomic forces are reduced to an appreciable extent.

The formation of brittle, intermetallic compounds in the friction interface can be a beneficial factor because, if such compounds are formed, the adhesive junctions tend to fracture in the original interface, with the result that adhesive material transfer from one sliding component to the other and - ultimately - adhesive wear does not occur. Obviously, a necessary requirement for brittle compound formation is that a certain amount of interfacial diffusion occurs. In practical systems one can hardly imagine a situation in which instantaneous diffusion will take place to any measurable extent during a single contact cycle. In such cases a certain amount of preliminary transfer will take place. Upon subsequent contacts frictional heating will facilitate interfacial diffusion with the result that brittle interfacial compounds can be formed, the transferred material remains attached for a small number of contact cycles only, and the wear particle, which forms ultimately, remains very thin.

Good examples of elements which do not form (solid) solutions with iron are lead and silver. The first element is a well-known component in lead based white metals and in leaded bronzes; the second metal can be used as such in the form of thin plating. Examples of brittle compound forming elements are tin and antimony. Again these are used frequently as main alloying components in bearing materials, most notably in white metals or babbitts, of which the lead babbitt Pb15Sb10Sn and the tin babbitt Sn12Sb are important representatives.

In many bearing materials, the alloying element with the favorable tribological properties, is present in the form of a separate, soft phase. This is, for example, the case in copper-lead, bronze-lead, and aluminum-tin bearings. In the first two cases, lead forms the second, soft phase; in aluminum-tin alloys this is the case for tin. Under load the softer phase is partly extruded and the interface becomes covered with a thin layer of lead or tin, respectively.

Obviously, in bearing materials, the ability of a metal surface to form suitable boundary lubricant films by reaction with the lubricant is an equally important feature. An example of this has already been given in the introduction (friction polymer formation, catalysed by finely divided tin oxide).

Similar, although not always equally strong effects are frequently found with other materials.

As stated above, plain bearings may also suffer from wear due to the presence of hard abrasive particles in the lubricant. This wear takes the form of (usually very mild) erosion if the effective width, d, of the abrasive particles is smaller than the minimum thickness of the lubricant film, h_{min}. However, if $d > h_{min}$ and in particular under conditions of boundary lubrication, more or less severe abrasive wear occurs. Under such conditions the power of the bearing surface to embed foreign particles (i.e. its embeddability), the surface hardness in the fully work-hardened condition and the elasticity of the journal are important criteria. Further, surface active components in the lubricant may play an important role in reducing the cutting action of angular abrasive.

Although the above properties (i.e. embedability, hardness, elasticity, and "lubricity") may be determined separately, Working Group 6 of Subgroup 2 of ISO/TC123 has recommended to adopt a "systems approach", i.e. to study the reaction of the system bearing material - lubricant - journal material to introduction of abrasive particles in the lubricant[7]. In principle this can be done in a laboratory apparatus, for instance of the "pin-on-ring" type, as well as in a real journal bearing. In both cases, the main criterion is the degree of damage (scratch formation) to the journal surface. This can be characterized adequately by means of Talysurf tracings as shown in Figure 6 for four classes of materials I - IV. In a specific test it was found that the white metals Pb Sb15Sn 10 and SnSb12 fall in class I, characterized by the fact that the surface condition of the steel mating surface remains virtually unaffected[12]. With Ag and CuSn6Pb10, class II, some local roughening of the surface of the ring occurs. Next, CuSn6Pb10 (P) and CuPb20 form class III, with appreciable roughening of the ring surface and finally A1Sn20 and CuSn8 fall in class IV, characterized by severe scratching of the contact surfaces.

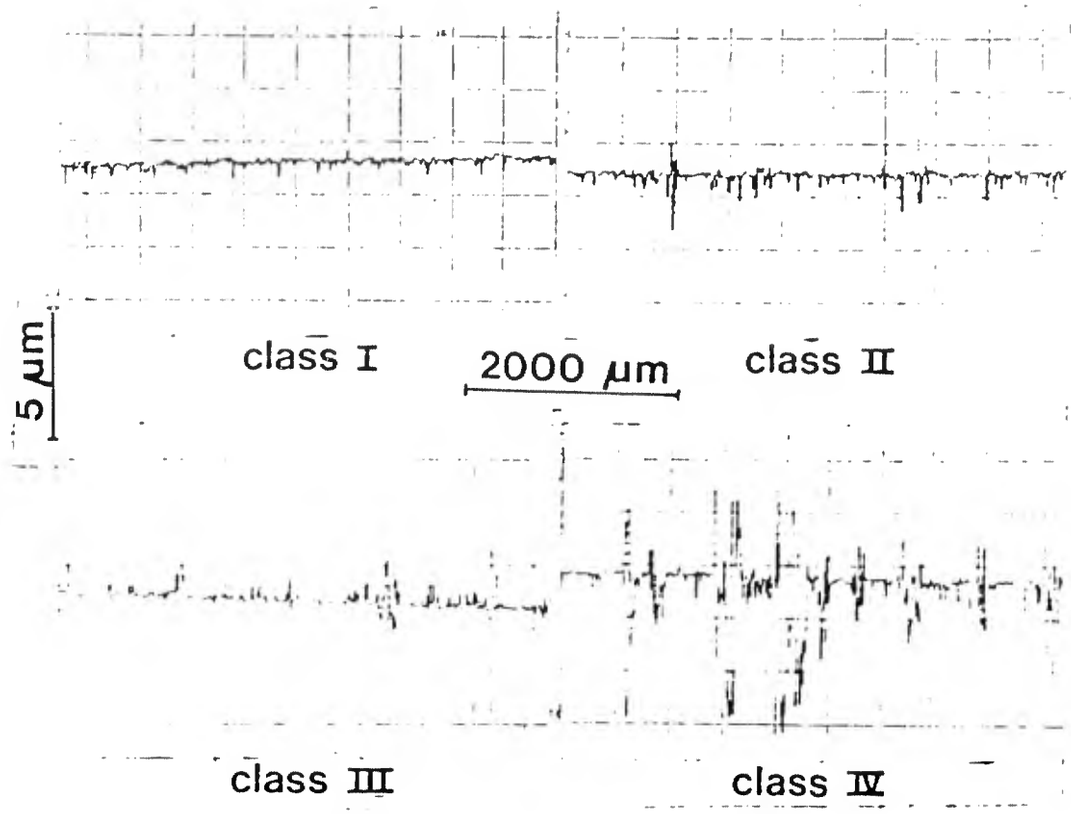


FIGURE 6 TALYSURF STYLUS TRACING OF STEEL SHAFTS, RUN AGAINST DIFFERENT BEARING MATERAILS UNDER CONDITIONS OF BOUNDARY LUBRICATION IN THE PRESENCE OF ABRASIVE DUST

As the mating surface, steel SAE 1045, and the lubricant were kept constant, differences in behavior must be due to differences in embeddability. The results show that this embeddability apparently does not correlate in a simple way with hardness. In fact the hardness of the A1Sn20 alloy and that of the white metals were the same, while the embeddability apparently differed considerably. Another striking feature is the significant difference in behavior between the phosphorus-free and the phosphorus containing, CuSn6Pb10 alloy. In a lubricant without abrasive, this difference is explained in terms of friction polymer formation. The present results permit the conclusion that such polymers also considerably mitigate the scratching action of hard particles in the lubricant.

If only because of the fact that for the greater majority of criteria, shown in Figure 5, generally accepted test methods

have not yet been developed, detailed information on the behavior of well-known bearing materials as discussed above is not yet available. Still, in Table 1 an attempt is made to summarize the available information on the functional behavior of such materials under conditions of boundary lubrication with or without abrasive. Although in this case, a draft ISO standard, describing a function-oriented test method, is available, and cooperative "round robin tests" have been performed by a number of laboratories, even in this case quantitative information cannot be given because of the pronounced system-dependent character of the characteristic parameters (a change in lubricant may drastically change the wear rates and even cause a change in the ranking of different materials)[7, 9]. In view of this, the information summarized in Table 1 is of qualitative nature and can only be used to assist in the choice of a potentially better material, if a particular material has been found to fail under application conditions. It should also be clear that the qualifications given in Table 1 refer to commercially available materials used in combination with a steel shaft and lubricated with a mineral oil with a low additive package. It should be noted that the possibility that special metallurgical or surface treatments may result in a much better quality than is specified in Table 1. The phosphorous-free tin-bronzes that were discussed in Section 1 form a clear example of this. Actually, the behavior of these materials under conditions of boundary lubrication is so superior that, in the second column of Table 1, they would easily score "+++"; however, they are not commercially available.

3. MATERIALS FOR ROLLING ELEMENT BEARINGS, GEARS, CAMS AND TAPPETS

3.1 Similar Characteristics

Rolling element bearings, gears, cams and tappets possess a number of similar characteristics as shown schematically in Figure 7. They are:

- Counterformal two body contact situation, leading to relatively high (Hertzian) contact pressures.
- Surfaces usually made from hardened steel.
- Elastohydrodynamic lubrication, characterized by the fact that the lubricant film formation is caused by hydrodynamic pressure built-up in the lubricant as well as by elastic deformation of the contact surfaces. As a result of such deformation, the thickness of the lubricant film is considerably larger than follows from "classical" hydrodynamic theories, which assume undeformable surfaces.

Material	behaviour under conditions of boundary lubrication	
	without abrasive particles [1]	with abrasive particles [2]
Pb-Babbitt (10 µm layer)	++	+
Pb-Babbitt (100 µm layer)	++	+++
Sn-Babbitt (10 µm layer)	+++	+
Sn-Babbitt (100 µm layer)	+++	+++
Silver (10 µm layer)	++	±
Silver (100 µm layer)	++	+
Aluminium-Tin	+	+
Copper-Lead	++	++
Sn-Bronze [3]	+	±
Sn Pb - Bronze [3]	+	++
Al-Bronze [4]	+	+

Notes

1) Rough estimate, based on data from different sources, concerning coefficient of friction, wearrate and process surface roughness (contact parameter); c.f. Fig. 5

2) Rough estimate, based on data from different sources, concerning damage to the steel shaft (wear and scratching)

3) See also the note in the text on phosphorus-free tin bronzes

4) Can only be used in combination with a hardened, for instance carburized, shaft.

Explanation of signs

+++ excellent
++ good
+ fair
} in most technical applications

± unsatisfactory for a number of applications.

TABLE 1 FUNCTIONAL BEHAVIOR OF SOME WELL-KNOWN BEARING MATERIALS, OPERATING UNDER CONDITIONS OF BOUNDARY LUBRICATION

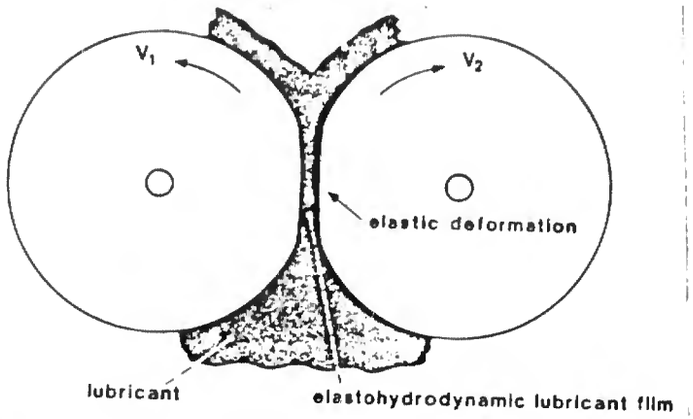


FIGURE 7 SIMILAR CHARACTERISTICS OF ROLLING ELEMENT BEARINGS, GEARS, CAMS, AND TAPPETS

In spite of the above similarities, there are also pronounced differences in tribo-technical behavior between rolling element bearings and rolling cams on the one hand and sliding cams on the other hand, while gears occupy an intermediate position. The main difference in this respect is that, in rolling element bearings and in rolling cams, the percentage of slip (sliding motion) usually is of the order of 1%. The same is true for gear teeth at the rolling circle. However, at top and bottom of gear teeth, up to 40% slip may occur, while sliding cams function with a slip percentage of 100%.

This difference in type of motion has important consequences for the tribological behavior of the materials, in particular because coefficients of friction in pure rolling are on the order of f = 0.001 - 0.01, while under conditions of sliding motion, in the presence of a lubricant, coefficients of friction are on the order of f = 0.01 - 0.1. As frictional energy is dissipated, mainly in the form of heat, this difference in friction results in the fact that the contact process in rolling element bearings and rolling cams nearly always proceeds at ambient temperature, while in sliding cams relatively high "friction induced temperatures" can occur. This difference in "thermal loading" has important consequences for the formation of an elastohydrodynamic lubricant film and - in connection with that - for the occurrence of different wear mechanisms. In rolling element bearings and rolling cams, functioning under "ideal" conditions (no misalignment, viscosity of the lubricant carefully chosen, no abrasive dust in the lubricant), one usually finds that the life of the components is limited by the occurrence of spalling (surface fatigue), caused by the cyclic variations in normal force. The same holds for

gear teeth at the rolling circle. However, at top and bottom of gear teeth and in sliding cams, the relatively high friction induced temperatures frequently interfere with lubricant film formation, as a result of which corrosive, abrasive, or adhesive wear may occur in an early stage of use.

3.2 Surface Fatigue/Spalling

In general, the resistance against surface fatigue of components of rolling element bearings and rolling cams increases with increasing hardness. However, as fatigue cracks are usually intitiated at sites in the material where high local stresses occur, local defects and impurity inclusions play a role of major significance. This aspect of fatigue induced spalling of rolling elements has been covered extensively by Scott et al[13, 14, 15]. As an example, Figure 8 shows the relation between lifetime under conditions of rolling contact and the (calculated) minimum film thickness/roughness ratio, h/R, for two cylinders (initial line contact), made of maraging steel ASTM A-538, grade C, lubricated with SAE 10W base oil. Two different manufacturing methods of the steel were applied, i.e. vacuum melting and electroslag remelting[16]. In the present case the minimum film thickness, h, was calculated with the equation of Dowson and Higginson for inital line contacts[17]:

$$h = \frac{1.6\alpha^{0.6} (\eta_o \bar{u})^{0.7} (E')^{0.03} R^{0.43}}{F_N^{0.13}} \qquad (1)$$

in which α is the pressure exponent of the viscosity, defined by:

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

η_o = dynamic viscosity of the lubricant at p = 1 atm.
$u = \frac{1}{2}(v_1 + v_2)$, in which v_1 and v_2 are the surface speeds of the cylinders

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

$R = \frac{R_1 R_2}{R_1 + R_2}$, in which R_1 and R_2 are the radii of curvature of the cylinders.

In the example given in Figure 8, R_1 = 74 mm and R_2 = 77 mm. The rotation frequency of the cylinders was 50 s^{-1}, so that v_1 = 23.24 m/s and v_2 = 24.18 m/s. The percentage of slip (sliding motion) thus amounted to:

$$\frac{v_1 - v_2}{\frac{1}{2}(v_1 + v_2)} \cdot 100\% = 4\%$$

The composite roughness parameter R was calculated with:

$$R = (R_{m_1}^2 + R_{m_2}^2)^{\frac{1}{2}}$$

in which R_{m1} and R_{m2} are the standard deviations in the distribution of roughness heights (r.m.s values) of cylinder 1 and cylinder 2, respectively.

In calculating h with Equation (1), it was assumed that during a test the viscosity of the oil was constant, i.e. that the temperature did not change significantly as a result of friction. As the percentage of slip was as low as 4%, this seemed a reasonable assumption. Variations in h/R were realized by changing the oil bath temperature, T, by which the value of η_o could be varied between $9.5 \cdot 10^{-3}$ Ns/m^2 (at T = 50°C) and $4.0 \cdot 10^{-3}$ Ns/m^2 (at T = 90°C).

The R-values were calculated for the new, as machined surfaces. Thus, the h/R values in Figure 8 refer to the lubrication condition directly upon the application of normal force. During running-in, the highest roughness peaks on the ring surfaces wear away, with the result that, upon running-in, h/R increases appreciably.

At each h/R ratio that was applied (i.e 0.7, 1.0, 1.2 and 2.5), ten pairs of vacuum melted rings and ten pairs of electroslag remelted rings were tested. The "L_{50} -lifetimes," given in Figure 8, were defined as the time which elapsed until five ring combinations in each group of ten, were damaged as a result of spalling fatigue. Figure 9 shows typical spalling damage which, in the present case, occurred suddenly after $7 \cdot 10^6$ revolutions.

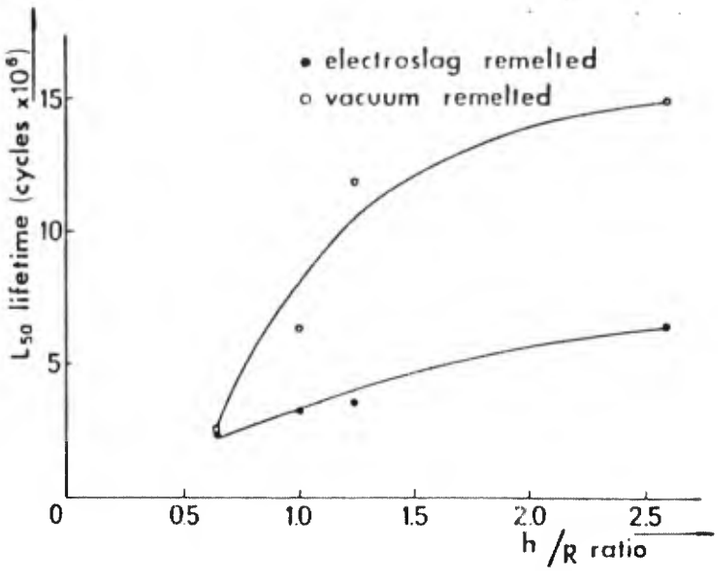


FIGURE 8 THE RELATION BETWEEN L_{50} LIFETIME AND (CALCULATED) MINIMUM FILM THICKNESS/ROUGHNESS RATIO $\frac{h}{R}$ FOR CONTACTING CYLINDERS (INITIAL LINE CONTACT OF MARAGING STEEL ASTM A-538, LUBRICATED WITH SAE 10 W BASE OIL)

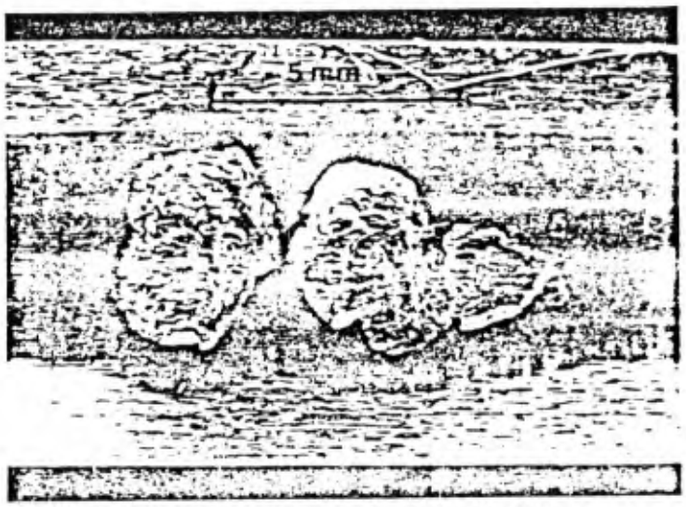


FIGURE 9 SPALLING DAMAGE

Figure 8 shows that for both steel batches the lifetime decreases with decreasing h/R ratio. This is due to the fact that the lower the value of h/R, the more roughness peaks penetrate the elastohydrodynamic lubricant film, in particular during the running-in period. During such penetration, the surfaces are locally damaged which induces the initiation of fatigue cracks from the surface into the bulk of the material. Figure 8 also shows that at h/R > 0.6 the vacuum melted steel has a significantly longer lifetime than the electroslag

remelted steel. The reason for this difference is that the vacuum melted batch contains less inclusions than the electroslag remelted batch.

At h/R > 2.5, crack initiation at the surface becomes a rare event. In fact fatigue failure in this lubrication regime is nearly exclusively due to crack initiation below the surface, in which stress concentrations around inclusions play an important role.

Metallographic examination of bodies, subjected to high cyclic loading, frequently shows the existence of "white etching zones" which frequently turn out to be prone to crack formation. Typical white etching zones, from which cracks originate, are shown in Figure 10. Metallographic work showed that the white etching zones in maraging steel consist of a soft modification of the material (hardness 300 HV instead of 560 HV), but with the same chemical composition as the surrounding matrix. It was also found that precipitation hardening as 480 °C completely restores the original structure and hardness of the material.

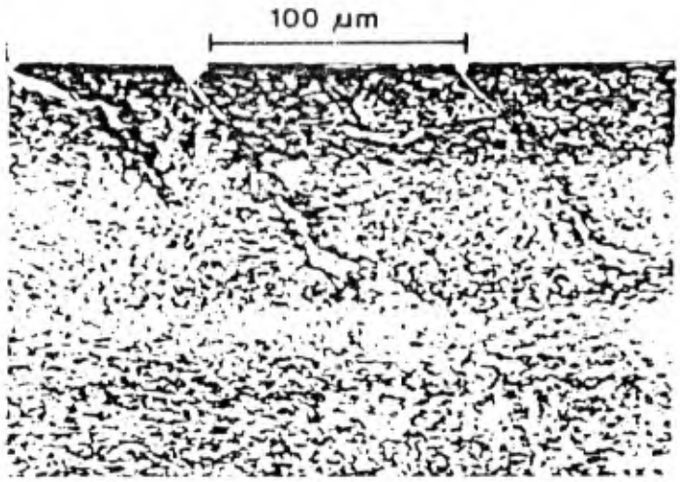


FIGURE 10 WHITE ETCHING ZONES FROM WHICH CRACKS ORIGINATE

As stated above, the pronounced decrease in lifetime at decreasing h/R ratio, is due to minor damage to the surfaces during running-in under conditions of asperity contact. A similar unfavorable effect can be due to corrosive attack of the steel surfaces. In this respect, it can be particularly dangerous if the lubricant contains water. Through a mechanism of preferential adsorption of water on the steel surface, followed by capillary condensation in tiny cracks formed during machining of the surfaces and local expansion, dangerous cracks can be formed on the surface which can lead to premature failure of the components, "infant-mortality."

Because of the pronounced influence of inclusions on spalling fatigue, very pronounced differences in lifetime can exist between different batches of steel with the same nominal chemical composition. An example of such a batch effect, dating back to the late sixties, is shown in Figure 11[18]. This shows the L_{10} lifetimes of the inner rings of 23 different groups of geometrically identical ball bearings, which were manufactured in an identical way, from different batches of vacuum degased ball bearing steel AISI 52100. The different batches were produced by seven different steel manufacturers. The L_{10} lifetime is defined as the time which elapses until 10% of the bearings in each group has been damaged by spalling fatigue. It can be seen that from one manufacturer to the other, pronounced differences occur. In particular, the results obtained with the materials from manufacturer E illustrate quite clearly that apparently identical materials may behave quite differently under conditions where surface fatigue predominates as a wear mechanism. It should be noted that since 1970 the situation has improved considerably. This is mainly due to the fact that steel to be used for rolling element bearings is now produced very carefully, thus keeping the impurity content and in particular, the distribution of the impurities in the material constant.

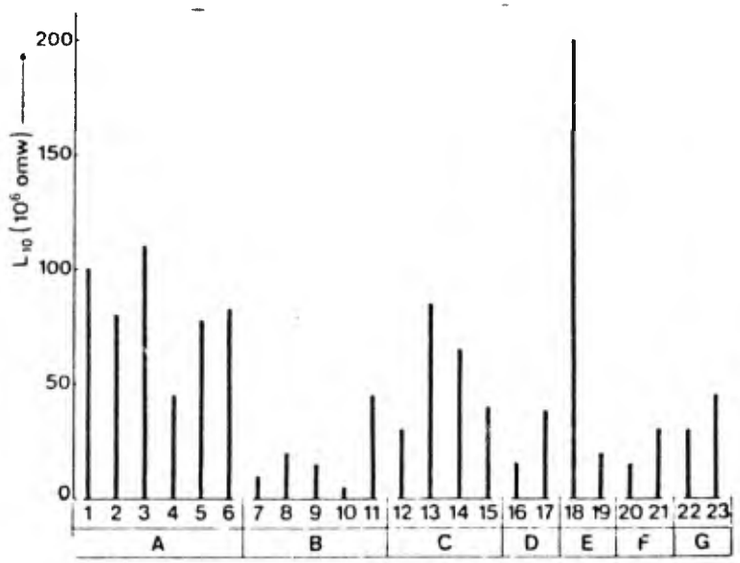


FIGURE 11 THE EFFECT OF STEEL QUALITY (BATCH EFFORT) ON THE ENDURANCE (L_{10} LIFETIME) OF BALL BEARINGS (FROM REF. 17)

1, 2, 3 .... 23: batch numbers
A, B, ... G: steel manufacturers

3.3 The Behavior of Real Rolling Element Bearings

Provided that rolling element bearings are assembled in accordance with the directions of the manufacturer, that a lubricant of adequate viscosity is used and that the lubricant does not contain water or dirt, the lifetime of the bearing is limited exclusively by subsurface (spalling) fatigue. Under such conditions the well-known Lundberg-Palmgren equation applies:

$$L_{10} = \left(\frac{C}{P}\right)^k \cdot 10^6 \qquad (2)$$

C is called the basic dynamic load capacity, defined as the force on the bearing under which 90% of the bearings has a lifetime of more than 1 million revolutions. P is the equivalent dynamic bearing load, which can be found by adding the radial force, Fr, and the axial force, Fa, according to:

$$P = XF_r + YF_a \qquad (3)$$

in which X and Y are constant factors for a particular type of bearing. The exponent k in Equation (2) is equal to about 3.

For each desired value of the L10-lifetime and for a known value of P, Equation (2) yields a value for C, on the basis of which, for each type of rolling element bearing, the necessary dimensions (diameters) of the bearing can be found in catalogues provided by the various rolling element bearing manufacturers.

The Lundberg-Palmgren Equation yields values for the L_{10}-lifetime, i.e. the time which elapses until 10% of the bearings in a larger group have suffered from spalling fatigue. Obviously, there may be cases in which one desires to obtain information on the time to lower failure percentages (e.g. L2-values). To that purpose, the L_{10} -values found, according to Equation (2), can be multiplied with the appropriate a_x factors given in Table 2, for failure percentages of 5, 4, 3, 2, and 1, respectively.

x	a_x
10	1.00
5	0.62
4	0.53
3	0.44
2	0.33
1	0.21

TABLE 2 CALCULATION OF TIME TO FAILURE PERCENTAGES $\leq$ 10%, ACCORDING TO: $L_x = a_x \cdot L_{10}$

It goes without saying that the Lundberg-Palmgren Equation cannot yield a prediction for the lifetime of one particular bearing. It only predicts the "average endurance" of a large amount of seemingly identical bearings. Even then one should be careful, because the Lundberg-Palmgen equation only applies under ideal conditions. Faulty assembly, abrasive dust in the lubricant, etc. can cause disasterous and essentially unpredictable reductions in lifetime.

3.4 Counterformal Contacts under Conditions of Sliding Friction

If counterformal surfaces are in contact under conditions of sliding friction (as in sliding cams and tappets), and elastohydrodynamic lubricant film can still be built up, provided that the conditions are suitable. However, because an appreciable amount of frictional heat is generated under conditions of sliding contact, the system is this case is certainly not isothermal by nature.

It has been shown that the lubrication condition of sliding counterformal steel contacts can be described in the form of a transition diagram, which defines the lubrication condition as a function of normal force F_N, sliding speed v and oil bath temperature, T[19, 20, 21, 22, 23]. Such diagrams apply equally well to ball-against-ball, ball-against-cylinder, crossed cylinder contacts, and also, they apply to virginal as well as to run-in surfaces. A cross section at constant T, for a completely oil submerged point contact, shows three regions: Figure 12; region I, in which friction is low and wear is nil or very low, region III, where severe wear and scuffing occur and an intermediate region II, characterized by a (temporarily) high friction but mild wear. It is assumed that in the three regions the following lubrication mechanisms apply:

region I: (partial) elastohydrodynamic lubrication

region II: boundary lubrication

region III: unlubricated contact (although the specimens are still fully submerged in the lubricant).

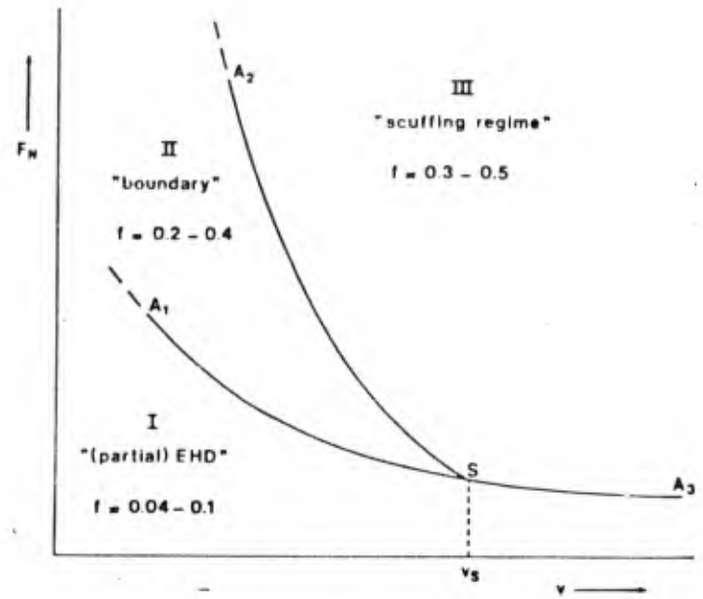


FIGURE 12 CROSS SECTION AT CONSTANT TEMPERATURE OF A TRANSITION DIAGRAM. APPLICABLE TO COMPLETELY OIL SUBMERGED SLIDING POINT CONTACTS

The lower curve A1-S-A3 has been called the "primary transition curve". It is believed to be continuous, point S merely being the intersection between curves A1-S-A3 and A2-S. Early work on the subject as well as recent research show that the location of the curve A1-S-A3 depends on viscosity. This is the main reason for assuming that in region I a thin elastohydrodynamic lubricant film keeps the surfaces apart[24, 25, 26]. The pronounced decrease in load-carrying capacity at increa sing speed of sliding is supposed to result from the effect of frictional heating in the contact zone, causing an appreciable decrease in effective lubricant viscosity[24]. Upon running in regions II or III, reaction (oxide) layer formation may cause a secondary transition towards a stable regime with very little wear. This regime is characterized by a coefficient of friction, f = 0.1 and a specific wear rate $< 0.1 \cdot 10^{-6}$ mm^3/Nm. If this regime is reached after functioning for a certain period of time in region II, the component normally is still serviceable. In fact, region II is probably identical to the region of "incipient scuffing," which is sometimes found in actual machine components[27]. On the contrary, when functioning

in region III, the component usually becomes unserviceable long before protective films have formed. Figure 13 shows characteristic friction force - time records, obtained at testing ball-against-cylinder contacts, Figure 14, at v = 0.5 m/s in marine diesel engine oil of 60°C under air cover, and at three different values of normal force, F_N.

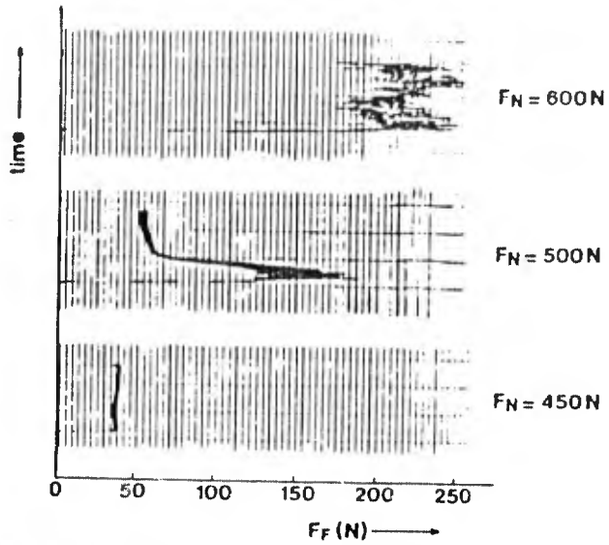


FIGURE 13 CHARACTERISTIC FRICTION FORCE - TIME RECORDS FOR BALL-AGAINST-CYLINDER CONTACT, LUBRICATED WITH MARINE DIESEL ENGINE OIL OF 60°C UNDER AIR COVER

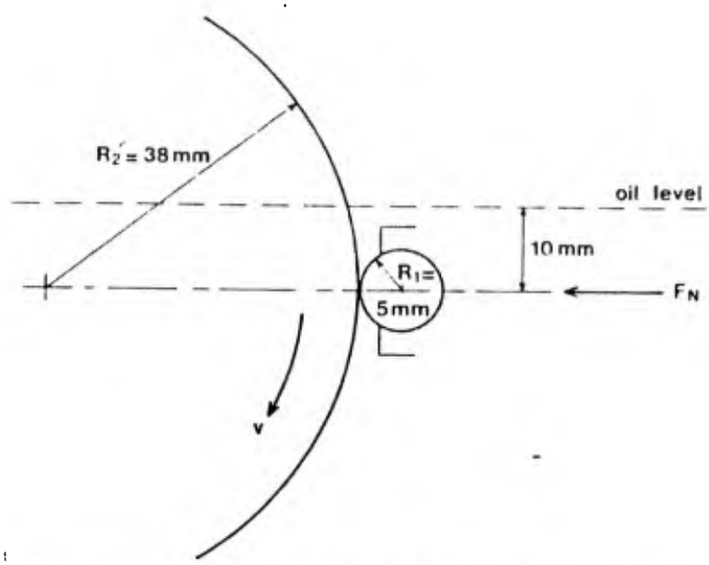


FIGURE 14 BALL-AGAINST-CYLINDER GEOMETRY

At normal forces F_N = 450 N, 500 N, and 600N, the system runs in, respectively, regions I, II, and III. Figure 13 clearly shows that in region II the process is regenerative by nature, friction decreasing rapidly to its f = 0.1 level. In this particular example, this took about 25 seconds. Figure 15 shows characteristic parts of the worn surfaces of the test specimens after termination of the given experiments. From the widths of the wear scars formed on the ball surfaces it can be seen that, in lubrication regime I, the wear rate of the pin is very small indeed. In lubrication regime II the wear rate is much higher, athough it still can be considered acceptable for most practical applications, mainly in view of the fact that the occurrence of wear is limited to the brief high friction period. In lubrication regime III, however, the wear rate is extremely high, which is completely unacceptable in practice. The corresponding ring surfaces, after termination of the tests, are also shown in Figure 15. It can be seen that, at F_N = 450 N (region I), the ring surface remains virtually undamaged. Actually, in most cases it is difficult to trace the wear track. At F_N = 500 N (region II), however, the wear track is clearly visible with the naked eye. Microscopic observation reveals that, in this region, adhesive material transfer occurs at a very minor scale. Finally, at F_N = 600 N (region III) severe adhesive wear effects are found upon inspection of the ring surface. This condition is similar to that of gear teeth which suffered from severe scuffing.

As stated earlier, the primary transition curve A1-S-A3, Figure 12, is assoicated with collapse of a partial EHD lubricant film. The location of the secondary transition curve A2-S, however, is probably associated with a metallurgical transformation in the steel, occurring within a narrow range of contact temperatures of the order of 500 - 600°C .

Elastohydrodynamic action being responsible for the partial separation of the surfaces in region I, one would expect the load-carrying capacity of the EHD film, the curve A1-S-A3, to decrease with decreasing speed in the very low speed region (effect not shown in Figure 12). Recent research showed that this curve does indeed show a maximum, occurring in the speed range of v = 0.001 - 0.003 m/s.

It goes without saying that the actual location of the transition curves in the transition diagram depends on a great number of factors, of which the most important are:

- the viscosity-temperature behavior of the lubricant
- the radii of curvature in the contact area and the local Hertzian pressure which results from these radii

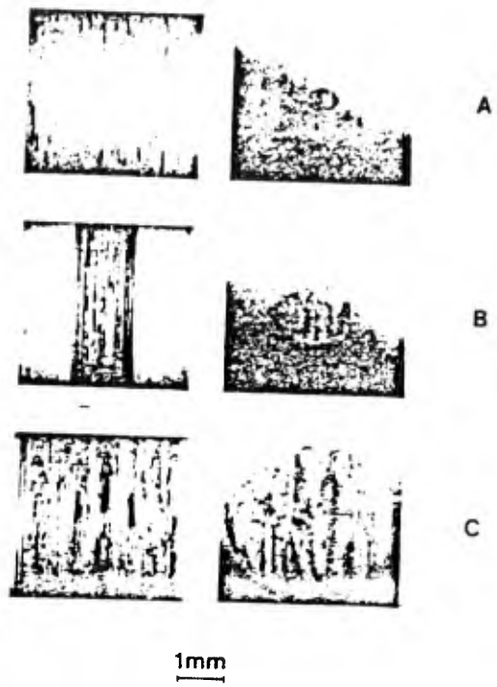


FIGURE 15 CHARACTERISTIC PARTS OF THE WORN SURFACES OF TEST SPECIMENS, PRODUCED IN, RESPECTIVELY, REGIONS I, II, AND III OF THE TRANSITION DIAGRAM (FIGURE 12)

- the roughness of the contacting surfaces, the chemical composition, and the degree of aeration of the lubricant
- the chemical composition and structure of the steel.

As an example, Figure 16 shows the effect on the transition diagram of adding the lubricant additive zinodialkyldithiophosphate (ZDP) to a mineral base oil. It can be seen that addition of ZDP results in a shift of the total diagram to higher F_N-values, which means that the load-carrying capacity of the system increases over the whole range of experimental conditions.

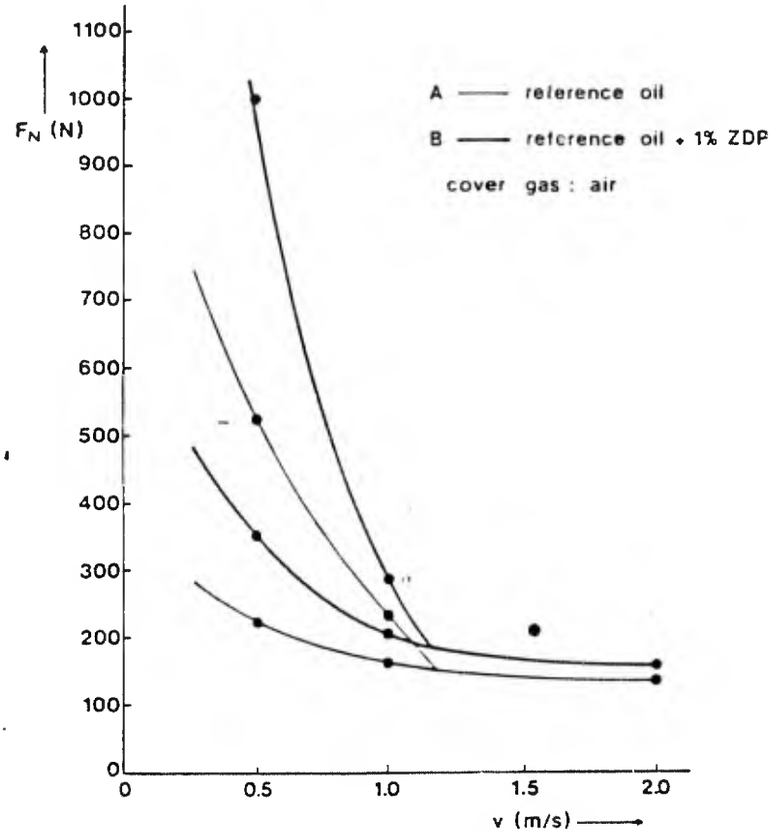


FIGURE 16 THE EFFECT OF ZINCDIALKYLDITHIOPHOSPHATE (ZDP) ON THE TRANSITION DIAGRAM FOR OIL SUBMERGED SLIDING POINT CONTACTS

Results obtained with different types of steel in mineral baseoil without additives under air cover, are shown in Figure 17. Three types of steel were used; steel AISI E 52100 with 1.5 wt.% chromium, steel AISI A2 with 5.6 wt.% chromium, and steel AISI D2 with 12 wt.% chromium. Each steel was subjected to three different heat treatments, which resulted in percentages of retained austenite (R.A.) of 2-4%, 17-21%, and 55% respectively.

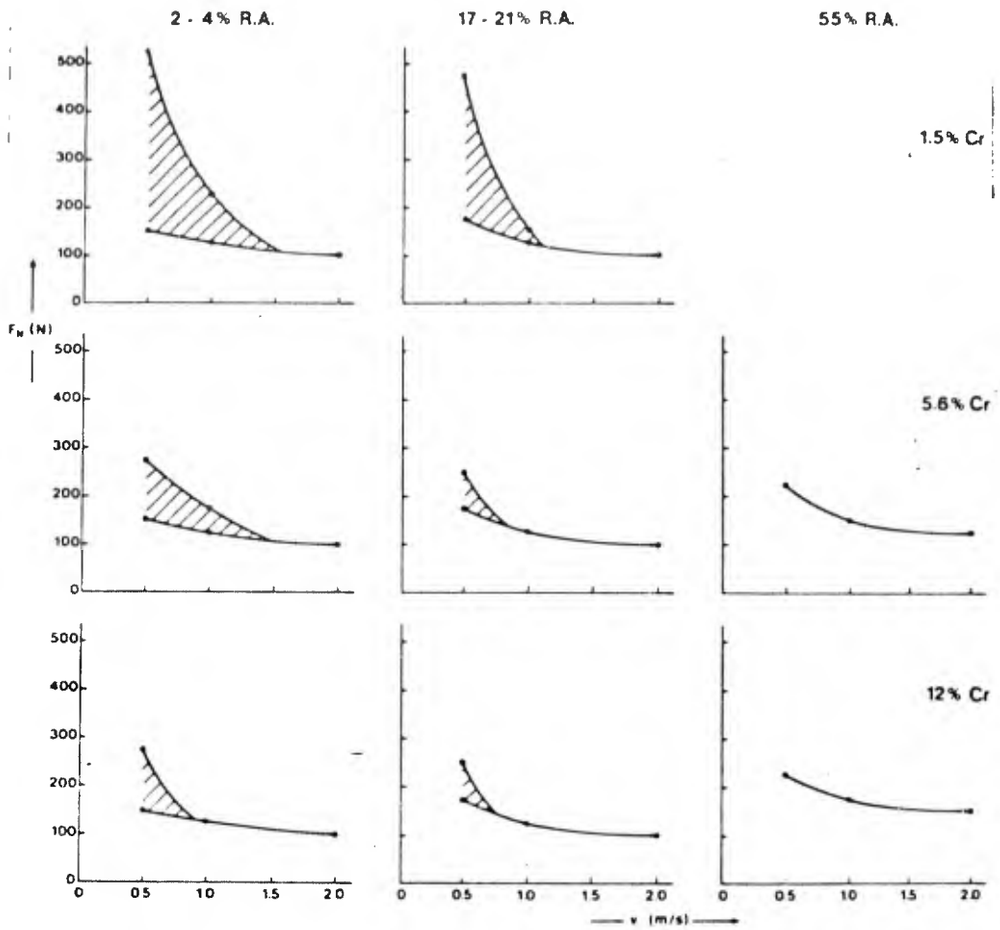


FIGURE 17 TRANSITION DIAGRAMS FOR STEELS WITH DIFFERENT PERCENTAGES OF CHROMIUM AND RETAINED AUSTENITE (R.A.)

Figure 17 shows that the chromium content of the steel has little or no effect on the load-carrying capacity of the partial EHD film (i.e. the location of the primary transition curves in the different diagrams). On the other hand, the chromium content is found to influence quite strongly the second primary transition (i.e. from the region of boundary lubrication or incipient scuffing to the scuffing region). In fact, at v = 0.5 m/s, the load-carrying capacity decreases from about 500 N for steel E 52100 to about 275 N for steels A2 and D2. The differences in heat treatment, as manifest in the R.A. percentages, are found to be of a dualistic nature. First, the load-carrying capacity of the partial EHD film shows a tendency to increase with increasing R.A. percentage. Secondly, the load-carrying capacity of the boundary film tends to decrease with increasing R.A. percentage. As a result of these effects,

the importance of regime II decreases appreciably with increasing percentage of alloying elements as well as with increasing percentage of retained austenite. This is borne out quite clearly by the diagrams shown in Figure 17.

Although a first primary transition (collapse of the EHD lubricant film) has been observed in many different systems, severe wear and scuffing in region III only occurs if untreated steel surfaces are used. Surface treatments (most notably salt bath nitriding) may result in a relatively mild wear process, even in region III. In such a case, wear remains mild until the protecting surface layer has been worn away. This aspect of steel-steel interaction is discussed in Section 6.

4. MATERIALS FOR DRY BEARINGS

4.1 General Aspects

Dry running, unlubricated bearings should not be used without special reason; if at all possible (and economically feasible), lubrication should be applied. The main reason for this is that with a few exceptions, only lubricated bearings, and obviously bearings in which the journal is kept in position by magnetic forces, can function completely without wear. Probably the only exceptions to this rule are lightly loaded rolling element bearings (although without lubrication it will be hard to avoid some cage wear), lightly loaded plastic bearings (as can be found, for instance, in office equipment), and bearings operating in vacuum and lubricated with molybdenum-disulphide. In such systems, the no-wear running period may sometimes surpass the desired lifetime. In lightly loaded plastic bearings in which the effects of contact with the mating component, usually a steel journal, are limited to elastic deformation of the surface asperities, the no-wear period is limited by the occurrence of micro-fatigue effects, as have been described by Bayer et al [28].

Also in dry running rolling element bearings the no-wear period is limited by fatigue. However, in contrast to the situation with plastic bearings, the damage after the first occurrence of fatigue effects in rolling element bearings usually is of such intensity that further use in the "wear period" is usually impossible. The benavior of bearings, lubricated with molybdenum-disulphide, is discussed in detail in subsection 4.4.

Reasons for choosing a dry running, unlubricated bearing, even in cases where a certain amount of wear may be expected, can be listed as follows:

1. The bearing has to function at high or low temperature, realistic boundaries being +250°C and -50°C.

2. The bearing has to function in vacuum.
   As far as this point is concerned it should be realized that nowadays lubricants with extremely low vapor pressure (10 - 50 Pa at 200°C) are commercially available.

3. The environment in which the bearing has to function must be kept extremely pure, so that leakage of lubricant cannot be allowed.
   Although, in principle, very efficient lubricant seals are available, dry running bearings are, for this reason, frequently applied in the textile industry, domestic appliances, etc.

4. The bearing has to function in a process fluid with nonlubricating (perhaps even degreasing) properties.
   Notwithstanding the presence of a fluid, one can, in these cases, often consider the bearing as "running dry."

5. The price of the equipment in which the bearing is to be used and/or the possibilities for maintenance are such that application of a lubricated system is virtually impossible.
   Again, this situation is found frequently in the field of domestic appliances; still in such equipment the grease lubricated spiral groove bearing is slowly gaining ground[29].

4.2 Materials

Materials for dry running, unlubricated bearings are frequently built-up on the basis of the following components which in some cases, may also be used as such:

- plastics (thermoplastic polymers and thermosetting resins)
- carbon and graphite
- metals with food running properties against iron such as silver, antimony and lead; see subsection 2.3
- molybdenum-disulphide MoS_2.

Common examples of materials which belong to one or more of these categories are described by Lancaster in the "Tribology Handbook"[30]. Data, taken from this reference, which are very useful in deciding which material to apply in a specific

practical application, are reproduced here in the form of Tables 3 and 4.

Table 3 contains information on plastic based materials. Although important differences in properties may occur from one material to the other, they have certain features in common:

- a relatively low maximum allowable surface temperature
- a relatively high coefficient of thermal expansion
- a very low heat conductivity.

Material		$\bar{p}_{max}$ MN/m^2	T_{max} °C	$\frac{\Delta L}{L}$ 10^{-6}/°C	λ w/m °C
thermoplastics	nylon, acetal	10	100	100	0.24
thermoplastics + fillers	fillers: - MoS_2 - PTFE - glass - graphite, etc.	15-20	150	60-100	0.24
PTFE + fillers	fillers: - glass - bronze - mica - carbon - metals	2-7	250	60-100	0.25-0.5
thermosets + fillers	phenolics, epoxies + asbestos, textiles, PTFE	30-50	175	10-80	0.4

$\bar{p}_{max}$ = maximum value static projected bearing pressure
T_{max} = maximum service temperature
$\Delta L/L$ = coefficient of thermal expansion
λ = heat conductivity

TABLE 3 SOME CHARACTERISTIC PROPERTIES OF PLASTIC BASED DRY BEARING MATERIALS (AFTER LANCASTER, REF. 29)

These facts are to be taken into account quite seriously in each practical design situation.

The characteristic properties of carbon or metal based materials are summarized in Table 4. Here we see that in contrast to the situation with plastic based bearing materials (summarized in the lower row of Table 4), these materials are characterized by a relatively high maximum allowable surface temperature, a generally quite modest thermal expansion and a

relatively excellent heat conductivity. From this, one rapidly gathers that these materials are particularly suited for high temperature applications.

Material		$\bar{P}_{max}$ MN/m^2	T_{max} °C	$\frac{\Delta L}{L}$ 10^{-6}/°C	λ w/m °C
carbon-graphite	may contain resin	1-3	500	1.5-4	10-50
carbon-metal	metal: Cu, Ag, Sn, Pb, Sb	3-5	350	4-5	15-30
metal-solid lubricant	bronze-graphite-MoS_2 Ag-PTFE	30-70	250-500	10-20	50-100
"plastics" (summary of Table 3)	frequently with fillers	2-50	100-250	10-100	0.24-0.5

$\bar{P}_{max}$, T_{max}, $\Delta L/L$ and λ: see legend Table 3

TABLE 4 SOME CHARACTERISTIC PROPERTIES OF CARBON OR METAL BASED DRY BEARING MATERIALS (AFTER LANCASTER, REF. 29)

The importance of maximum allowable service temperature and heat conductivity is clearly illustrated by Figure 18, which describes experience acquired in designing a refrigerating plant. Figure 18 shows the behavior of three different bearing materials, each based on a combination of polytetraflourethylene (PTFE) and bronze, used in combination with journals of stainless steel and copper, respectively, and used at an environmental temperature of -20°C. Bearing materials as well a journal materials show pronounced differences in heat conductivity. Bearing materials A and B possess a PTFE matrix, filled with bronze powder, while bearing material C possesses a bronze matrix, filled with PFTE. As a consequence of their structure, materials A and B have a relatively low heat conductivity, 4 to 10 mW/cm K, respectively. In material C, the fact that bronze forms the matrix guarantees a relatively high heat conductivity of 200 mW/cm K.

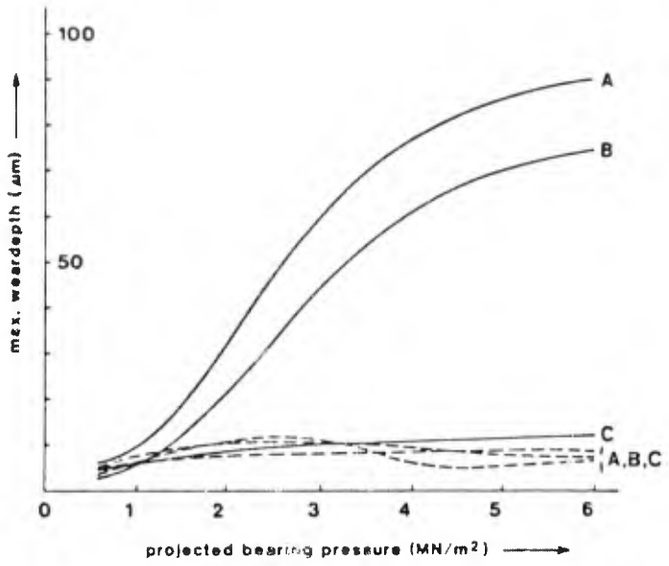


FIGURE 18 WEAR DEPTH AS A FUNCTION OF PROJECTED BEARING PRESSURE FOR THREE PTFE-BRONZE COMPOSITES, A, B AND C, USED IN COMBINATION WITH JOURNALS OF, RESPECTIVELY, STAINLESS STEEL (FULL LINE) AND COPPER (DASHED LINES)

Figure 18 shows that if materials A and B are used in combination with stainless steel which has a rather poor heat conductivity, a pronounced increase in wear depth occurs in the range of bearing pressures between 1 and 3 MN/m^2. This is caused by a pronounced increase in surface temperature of the bearing as a result of dissipation of frictional heat. Clearly, the heat conductivity of material C is so high that this phenomenon is not found if that material is used. When using copper as journal material, a material with a relatively high heat conductivity, none of the materials A, B, or C show the pronounced increase in wear depth that is found with materials A and B in combination with a stainless steel journal. This example clearly illustrates the important role of heat conductivity in the tribo-technical behavior of bearing materials based on plastic. The total energy that dissipates per unit of projecting bearing surface, as a result of friction (Q'), is given by

$$Q' = f.\ p.\ v \qquad (4)$$

in which f is the coefficient of friction, p is the projected bearing pressure, and v is the speed of sliding. A "maximum allowable pv-value" $(pv)_{max}$ is sometimes specified for a particular bearing material. If only because of the fact that f

may depend quite strongly on v as well as p, this can be misleading, in particular in the case of plastic based materials. Instead, the use of p against v curves is recommended[30]. For a number of plastic based materials such curves are given in Figure 19.

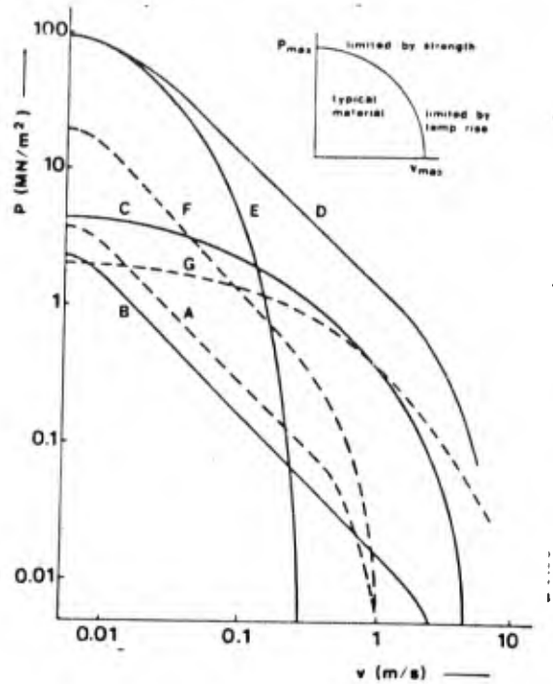
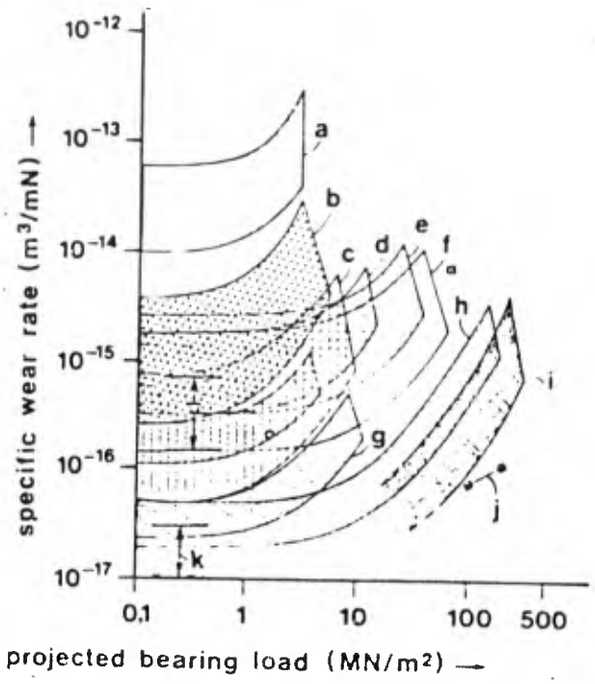


FIGURE 19 P-V CURVES FOR A NUMBER OF PLASTIC BASED BEARING MATERIALS (AFTER LANCASTER, REF. 29)

Clearly, information as given in Figure 19 is based on some form of general concensus as to what wear rates are acceptable for practical application. More precise information on the wear rates to be expected when using dry running bearings is hard to give because of the pronounced systems dependency of such data. As a first approximation, information on the specific wear rates of characteristic dry bearing materials can be found in Reference 31. Figure 20, taken from this reference, relates the specific wear rate, expressed in mm^3 material removed per meter sliding distance, per N normal load, for a number of materials to the projected bearing pressure p. This figure relates to situations in which either the heat production (low sliding speed) or the possibilities for heat conduction are such that "high" temperatures in the bearing are avoided. Figure 20 shows that, even within one "family" of materials, the specific wear rate may easily differ by a factor of 5-10. Therefore application oriented tribo-metry will generally be necessary if one wishes to obtain more precise information[5].



A : PTFE (unfilled)

B : carbon-graphite (including metal and thermoset impregnated types)

C : filled PTFE's

D : filled nylons and acetals

E : reinforced phenolic resin laminates with solid lubricant fillers

F : filled polyimides and p-oxybenzoyl polymers

G : Epoxy-bonded bronze / Pb filled PTFE

H : PTFE / Pb in porous bronze on a metal backing (available wear depth = 0.04 mm)

I : Woven PTFE / glass fibre / metal back (available wear depth = 0.25 mm)

J : PTFE fibre / high strength CF filament wound bearings

K : High modulus CF reinforced thermoset (pin-on-disc results).

FIGURE 20 SPECIFIC WEAR RATE AS A FUNCTION OF PROJECTED BEARING LOAD FOR A NUMBER OF "DRY BEARING MATERIALS" (AFTER CREASE, REF.30)

4.3 The Effect of Surface Roughness and Material Transfer on the Wear of Plastics

Apart from the pronounced effect of friction induced surface temperatures, the roughness of the mating journal surface and possible changes in that roughness, exert a pronounced influence on the tribological behavior of plastic bearing materials. Figure 21, taken from Reference 32, shows this for six different plastics. From the upper part of Figure 21, it can be seen that for all materials, the coefficient of friction, f, goes through a minimum, the location of which

varies with the nature of the material. The reason for this effect is that at low values of journal roughness, adhesive forces in the interface predominate (the materials "glue together"), while at higher roughness values the machining forces increase considerably, due to the chiselling action of the asperities of the steel journal surface. The lower part of Figure 21 shows that although for all materials, wear increases with increasing journal roughness, the intensity of the effect differs considerably from one material to the other. The wear of PTFE depends very strongly and the wear of nylon (PA-66) only very slightly on roughness.

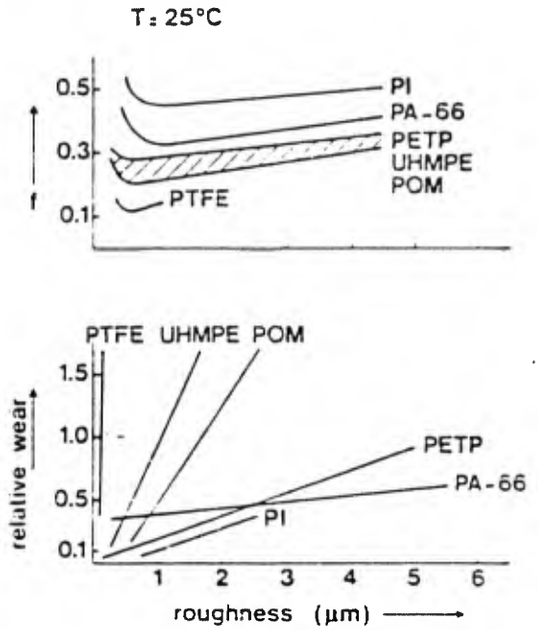
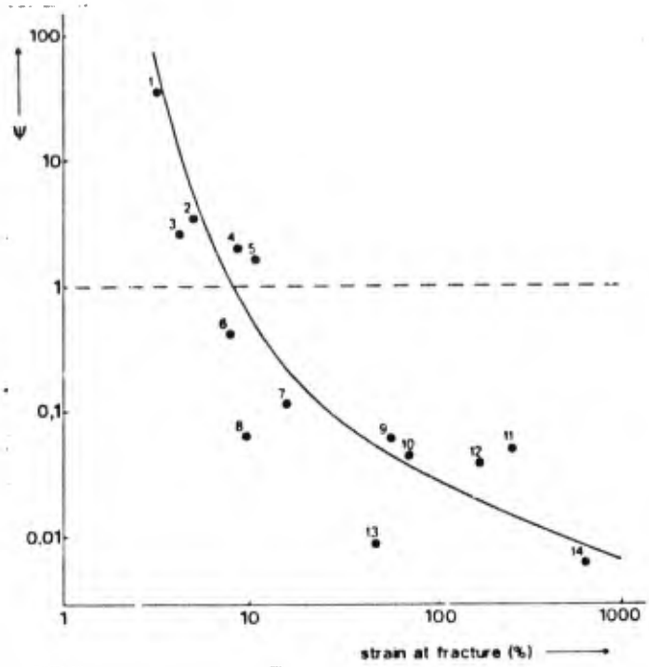


FIGURE 21 FRICTION AND WEAR OF A NUMBER OF UNFILLED PLASTICS AS A FUNCTION OF THE ROUGHNESS OF THE MATING SURFACE (AFTER ERHARD AND STRICKLE, REF. 31)

A complicating factor in discussing the effect of roughness of the mating surface on wear of plastic bearing materials is that the material of the bearing may transfer to the steel journal, thus masking the original roughness pattern. This effect has been studied and discussed comprehensively by Lancaster[33]. Figure 22, taken from this reference, shows 14 different plastic bearing materials in sliding contact with steel, the ratio, (λ), between the wear rates, characteristic for a wear process with transfer and the wear rates, characteristic for a wear process without transfer, as a function of the strain at fracture. Figure 22 shows that the ratio, Ψ, decreases appreciably with increasing ductility (increasing strain at fracture). For the relatively ductile plastics, 6-14, the formation of a transferred layer on the



1 : epoxy-828
2 : polystyrene
3 : polyvinylidenechloride
4 : polyether-17449
5 : polymethylmetacrylate (perspex)
6 : polytetrafluorchloroethylene
7 : acrylonitril-butadienestyrene colepolymer
8 : polycarbonate
9 : nylon 11
10 : nylon 6.6
11 : polytetrafluorethylene (teflon)
12 : polypropylene
13 : polyacetale
14 : polyethylene
R_a-value shaft : 0.08 μm
ambient temperature : 25°C.

FIGURE 22 RATIO Ψ BETWEEN THE WEAR RATES, CHARACTERISTIC FOR A WEAR PROCESS WITH TRANSFER AND THE WEAR RATES, CHARACTERISTIC FOR A PROCESS WITHOUT TRANSFER FOR 14 DIFFERENT BEARING MATERIALS IN SLIDING CONTACT WITH STEEL AS A FUNCTION OF THE DUCTILITY (STRAIN AT FRACTURE) (AFTER LANCASTER, REF. 32)

steel surface is favorable, because this reduces the roughness of the mating surface and thereby the micro-machining action of the asperities. For the relatively brittle plastics, 1-5, a transferred layer with good adherence to the journal surface is not formed with the result that, in this case, it is favorable to prevent transfer.

In the case of Lancaster's experiments, the transfer situation was realized by traversing several times over the same track in the unlubricated condition. A wear process without plastic transfer was realized by introducing a suitable lubricant, preventing hydrodynamic effects by using a suitably low viscosity.

In adapting the information of Figure 22 to practical tribological situations, it should be kept in mind that, with plastics, a decrease in temperature can result in a sudden drastic decrease in ductility at the so-called "embrittlement temperature." Figure 22 shows that as a result, the wear behavior of such a plastic can change drastically.

4.4 The Effect of Fillers

It is, again, Lancaster who described conprehensively the effect of fillers on the behavior of plastic bearing materials[33].

Addition of molybdenum-disulphide, MoS_2, or polytetrafluroethylene, PTFE, in either elastomers or resins, usually leads to transfer of the filler to the mating metal surface. This results in an appreciable decrease in coefficient of friction and thus, in reduction of frictional heat. The ultimate effect on wear cannot easily be predicted because the reduction in friction is usually counterbalanced by a reduction in mechanical strength of the plastic. According to Lancaster, the strength reducing effect of PTFE is less pronounced than that of MoS_2.

Another well-known addition to plastic bearing materials consists of carbonfiber. Such fiber gives additional mechanical strength to the plastic, may be instrumental in conducting the frictional heat from the surface, and may also favorably reduce the roughness of the mating steel surface. If the fibers possess a graphite-like structure, a transferred film can be formed, in the same way as described above for MoS and PTFE. The formation of such transferred films can lead to a considerable reduction in wear rate. An example of this is shown in Figure 23, again taken from Reference 33. It can be seen that the initially relatively high wear rate decreases appreciably as a result of the gradual formation of a

transferred graphite film on the mating steel surface. Addition of water results in removal of the transferred film and in a sudden, drastic increase in wear rate.

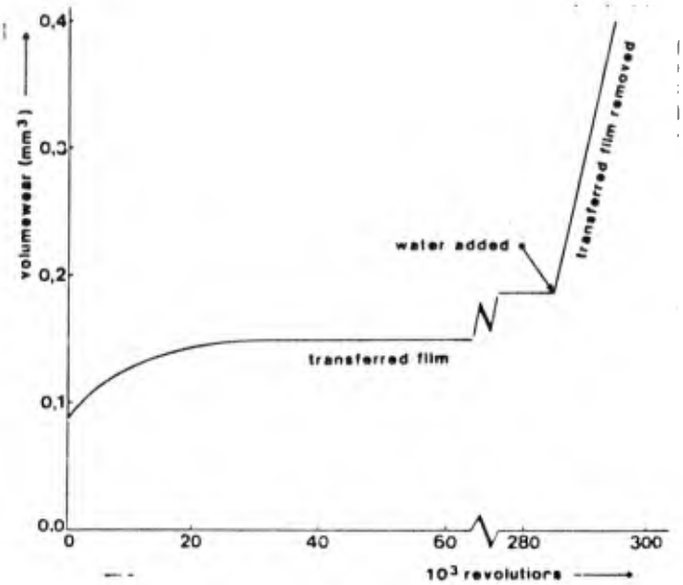


FIGURE 23 EFFECT OF THE FORMATION OF A TRANSFERRED FILM OF GRAPHITE ON THE WEAR OF EPOXYRESIN, FILLED WITH NONPOLISHING CARBON FIBRE (AFTER LANCASTER, REF. 32)

Carbonfibers with a cubic structure (non-graphitic) do not form a transferred film on the mating steel surface. They have, however, a polishing action, as a result of which the roughness of the mating surface decreases gradually with time. This also has a favorable effect on wear of the filled plastic, as shown in Figure 24, again taken from Reference 33. Figure 24 shows the volume wear of PTFE, filled with, respectively, non-polishing carbon fibers (A) and polishing carbon fibers (B), as a function of the number of revolutions and for three different initial roughnesses of the mating steel surface. It turns out that, when using the nonabrasive fiber, the wear rate increases considerably with increasing surface roughness of the steel. This is in accordance with the information from the lower part of Figure 21. However, if the plastic filled with polishing fibers is used, the volume of wear is found to be nearly independent of the initial surface roughness of the mating steel surface. Actually, the differences which are still found are due to the fact that it takes some time for the roughness of the steel surface to attain its equilibrium value. From the slope of curve 1 in part A and curves 1, 2, and 3 in part B, it can be concluded that the c.l.a.l value of the polished steel surface is approximately 0.04 μm.

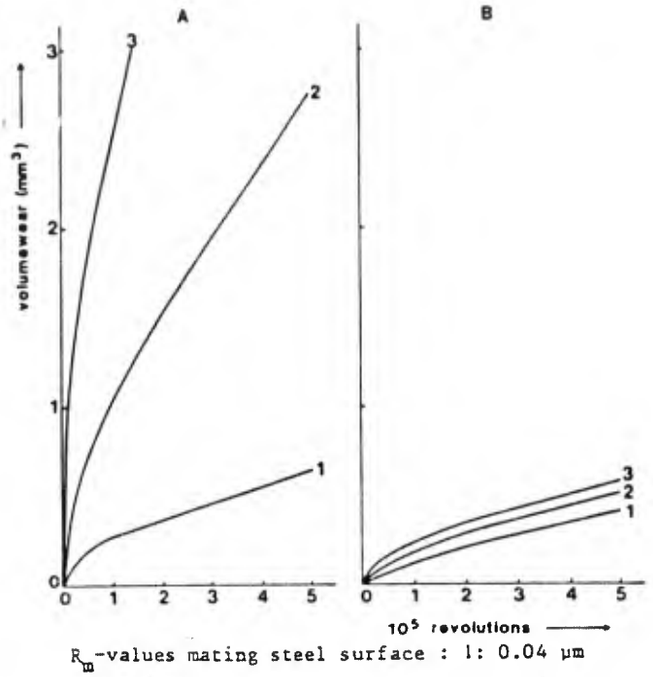


FIGURE 24 WEAR OF PTFE FILLED WITH, RESPECTIVELY, NONPOLISHING (A) AND POLISHING (B) CARBON FIBERS AS A FUNCTION OF THE NUMBER OF REVOLUTIONS OF THE JOURNAL FOR THREE DIFFERENT INITIAL ROUGHNESSES OF THE MATING STEEL SURFACE

The favorable effect of polishing is even more pronounced in the case of water lubricated epoxyresin-carbonfiber composites, because water in any case prevents the formation of transferred films, Figure 23.

Figure 25 shows the relationship between the wear rate and the initial surface roughness of the mating steel surface for the following three cases:

a) composites with polishing carbonfiber in contact with stainless steel
b) composites with non-polishing carbonfiber in contact with stainless steel
c) composites with polishing carbonfiber in contact with hard chromium.

Figure 25 shows that the polishing action in case (a) leads to a situation in which the specific wear rate is virtually independent of the initial roughness of the mating steel surface. Clearly, a similar effect is not found if a composite

with non-polishing carbonfiber is used, case (b). The same is true in case (c), because, in that situation, the polishing fiber is not able to polish the hard chromium, as a result of which the effect of the initial roughness of the steel surface persists throughout the entire lifetime of the component.

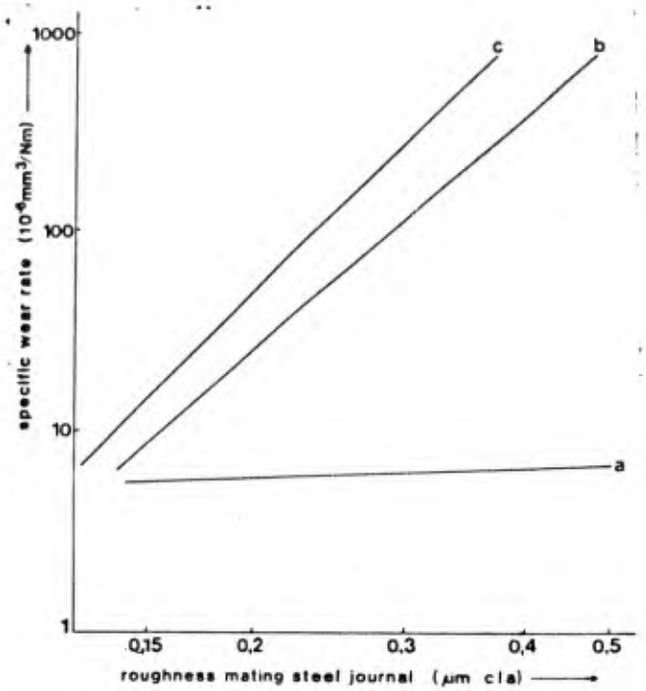


FIGURE 25 RELATION BETWEEN THE WEAR RATE AND THE INITIAL SURFACE ROUGHNESS OF THE MATING STEEL SURFACE FOR THE FOLLOWING CASES:

A. COMPOSITES WITH POLISHING CARBON FIBER IN CONTACT WITH STAINLESS STEEL
B. COMPOSITES WITH NONPOLISHING CARBON FIBER IN CONTACT WITH STAINLESS STEEL
C. COMPOSITES WITH POLISHING CARBON FIBER IN CONTACT WITH HARD CHROMIUM

(AFTER LANCASTER, REF. 32)

It is not accidental that in tests, the results of which are shown in Figure 25, stainless steel or chromium were used. If ordinary steel had been used, corrosion might have led to an appreciable change in surface roughness. In practice such corrosive attack of the mating steel surface usually leads to a reduction in roughness and thus, to a reduction in wear of the plastic bearing material. However, in some cases, depending on the corroding agent, the roughness may increase, with an accompanying unfavorable increase in wear rate.

Lancaster studied the wear of epoxyresin-non-polishing carbonfiber composites in different environments. The results are shown in Figure 26. The lowest wear rate is found in dry condition because a transferred graphite film is formed. On the other hand, the highest wear rate is found in water, because water prevents the formation of a transferred film, while a decrease in roughness of the mating surface by either polishing or chemical etching (corrosion) does not occur. When using seawater, some "chemical polishing" occurs, because the stainless steel which is used as mating surface, is not completely stable in seawater. This results in a considerable reduction in wear rate of the plastic. This effect is enhanced considerably by using solutions of coppersulphate or ironchloride, which produce a stronger etching effect on stainless steel. The organic liquids, "mineral oil" and "diester", have no corrosive action with respect to stainless steel. In this case, the considerable reduction in wear rate as compared to that in water, is to be ascribed to the formation on the steel surface of a transferred film consisting of a mixture of the resin and the lubricant.

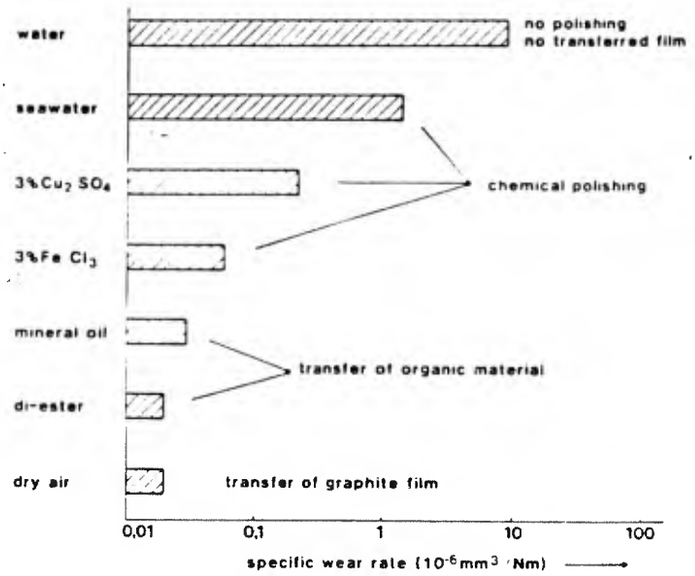


FIGURE 26 WEAR RATE OF EPOXYRESIN NONPOLISHING CARBON FIBER COMPOSITES IN DIFFERENT ENVIRONMENTS (AFTER LANCASTER, REF. 32)

It has already been mentioned that corrosive attack of the metal surface can lead to an increase in surface roughness and, thereby, to an increase in wear of the plastic bearing material. The same holds for fillers with a pronounced action. If the individual particles of the filler are relatively big, the

roughness of the mating steel surface may increase rather then decrease as a result of contact with the bearing material, resulting in an increase in wear rate of the plastic. This may happen if glass is used as a filler, in particular if the mating surface is relatively soft (aluminum, bronze, titanium, etc.).

From the above, it will be clear that the tribological behavior of plasti-metal combinations is rather complicated. In fact function-oriented tribo-metric testing sill frequently be necessary if one wants to find the best material for a particular application[5].

4.5 Plastic-Plastic Combinations

In the above subsection, the behavior of plastic-metal combinations was discussed. Thereby the important effect of frictional heating of the plastic in the contact zone was emphasized. This is even more important in the case of plastic-plastic combinations, which generally have a very poor heat conductivity. Thus, they can only be used in very lightly loaded and/or very slowly moving mechanisms. Published information on the behavior of plastic-plastic combinations is very scarce. From research performed at TNO, the results of which as yet are unpublished, the following general rules can be derived:

- The greater the difference in chemical composition and structure of the contacting surfaces, the lesser the wear will be. This is another illustration of the "compatibility rule" which was discussed for metal-metal combinations in subsection 2.3.
- The presence of fillers with polishing effects in one of the contacting surfaces, invariably leads to a considerable increase in wear rate of the mating surface. In this case, plastic-plastic combination, this also holds for polishing carbonfibers.
- Contrary to the situation with plastic-metal combinations, introduction of a liquid will always lead to a decrease in wear rate because, in this case, a decrease in adhesion and material transfer will always be favorable.

4.6 Solid Lubricants

As stated in the introduction of this chapter solid lubricants (i.e. in particular molybdenum-disulphide, MoS_2, and graphite) frequently form an important component in materials for dry bearing applications. As such, dry lubricant films, formed by vigorous rubbing-in of suitably pretreated metal surfaces may already result in a dry running bearing with

favorable performance characteristics. Further resin or waterglass bound films find widely spread applications and finally, graphite of MoS^2 may be added to lubricating grease.

An important argument for applying a solid lubricant is the very good thermal stability which can be realized with some systems. Actually, dry MoS^2 powder in neutral or reducing environments may function adequately up to temperatures of 800°C while a lubricant as calciumfluoride with ceramic binder may function adequately up to 1200°C.

If the solid lubricant is not replenished during use, the applicability is limited by the endurance of the initially applied lubricant film. The following is based on Reference 34, updated if necessary.

4.7 Main Properties of MoS_2 and Graphite

Molybdenum-disulphide as well as graphite have a hexagonal crystal structure, as shown for graphite in Figure 27 and for MoS_2 in Figure 28. In both cases the hexagonal structure leads to the formation of platelike crystals with a thickness to length ratio on the order of 1 : 1000. In pure MoS_2, the attractive forces between neighboring sulphur planes are very low, with the result that shear is easy. On the other hand, the load-carrying capacity in a direction normal to the hexagonal plane is very high. This explains the extremely favorable friction properties of MoS_2. Unfortunately, impurities of different kinds in particular oxygen or water hinder the free movement of the hexagonal movement crystals. This is entirely different with graphite, which shows low friction only if an adequate amount of "impurity molecules" is adsorbed on the carbon lattice. In this respect water vapor is found to be extremely favorable. Obviously, this basic difference in response to "impurities" has important consequences for the practical application of MoS_2 and graphite. For example, in a vacuum molybdenum-disulphide will show an excellent performance, while graphite will fail utterly as a lubricant. On the other hand, in very humid environments, (relative humidity > 80%) and under conditions of frictional contact, molybdenum-disulphide will hydrolyze spontaneously, leading to the formation of molybdenum-disulphide, molybdenum-trioxide, sulphur-dioxide, and free sulphur. Under such conditions, graphite functions fully adequately. The above hydrolysis of MoS_2, is one of the most convincing examples of "tribo-chemical reactions", as in vitro, it is impossible to decompose MoS_2, even by boiling with a concentrated solution of hydrochloric acid.

As far as the load-carrying capacity of the film is concerned, MoS_2 is undoubtedly highly superior to graphite. In

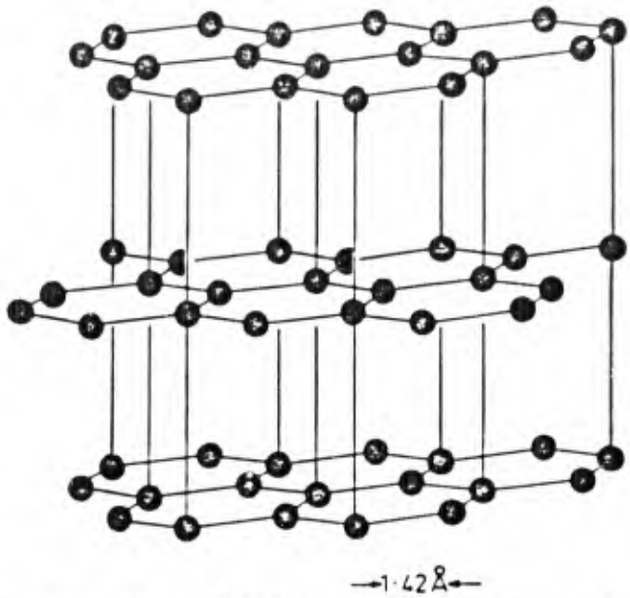


FIGURE 27 CRYSTAL STRUCTURE OF GRAPHITE

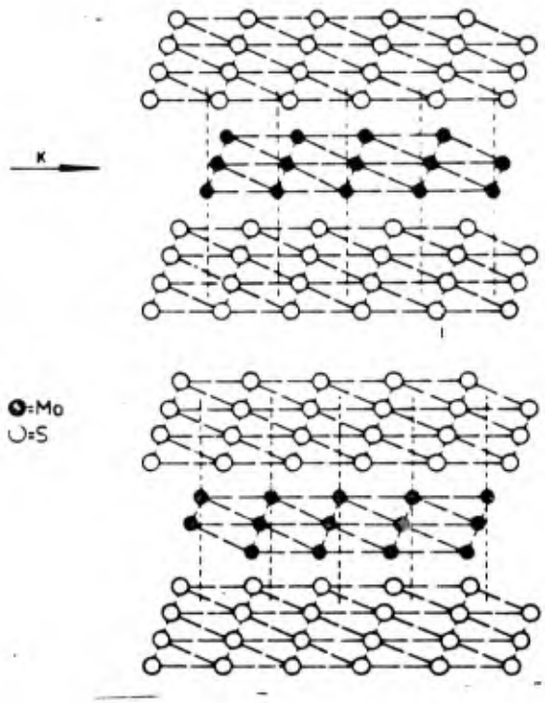


FIGURE 28 CRYSTAL STRUCTURE OF MOLYBDENUM-DISULPHIDE

fact, an MoS_2 film, formed from dry powder will carry any load which does not lead to gross plastic deformation of the substrate. The same is by no means true for graphite. In MoS_2 lubricated systems, one will usually find a decrease in coefficient of friction with increasing normal force. For a lubricant film formed from dry MoS_2 powder, this behavior is illustrated in Figure 29. It can be seen that the relative humidity of the atmosphere is an important parameter in the f/F_N relation. This is probably due to the fact that the friction behavior of MoS_2 improves with decreasing humidity. With increasing normal force, the heat which is generated in the friction zone, increases leading to an increase in surface temperature of the lubricant film. At a constant relative humidity of the surrounding atmosphere, this leads to a decrease in coefficient of friction with increasing normal force. Figure 29 shows that very low coefficients of friction can be obtained with MoS_2, of the order of $f = 0.03$. Even under optimum conditions, the coefficients of friction, measured with graphite as a lubricant, are higher of the order of $f = 0.1$. As is to be expected, on the basis of the friction mechanism, described above, an increase in normal force will, in this case, lead to a usually moderate increase in f.

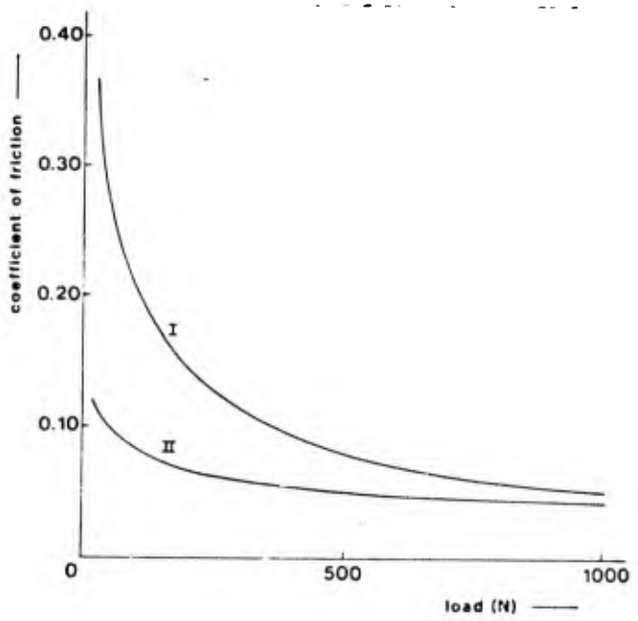


FIGURE 29 RELATION BETWEEN COEFFICIENT OF FRICTION AND NORMAL FORCE FOR MOLYBDENUM-DISULPHIDE POWDER FILMS IN DIFFERENT ENVIRONMENTS

I : AIR OF 25°C AND 50% RELATIVE HUMIDITY
II: AIR OF 25°C AND 0% RELATIVE HUMIDITY

If the solid lubricant is supplied continuously from the matrix of the bearing material (as, for instance, in the case of plastic - MoS_2 composites), the endurance of the lubricant film does not come into play. However, in cases where one relies on the performance of a rubbed-in and burnished film on a steel substrate, the endurance usually is the determining factor in a practical application.

As far as molybdenum-disulphide is concerned, the best performance is obtained if the contact surfaces are made of a steel with HV = 200-300 and a steel with HV = 700-900, respectively. The MoS_2 powder is rubbed on the harder surface, after pretreatment by wet blasting with quartz or Al_2O_3 to a final roughness of approximately 0.1-0.3 µm c.l.a. An even better performance/endurance is found if the substrate is finished by phosphating. If solid lubricant films are formed by mechanical rubbing-in, which is a usual and dependable application procedure, the layer thickness in the fully compressed state depends on the rubbing-in time. Thus, the time necessary to form a layer with a compressed thickness of 1 m will vary considerably from one powder to the other. For MoS_2, in two modifications (i.e. average particle sizes of 30 µm and 3 µm, respectively), the related (also hexagonal) compound tungsten-disulphide WS_2 and for graphite, this rubbing-in time is given in Table 5. It can be seen that the rubbing-in time indeed varies considerably from one powder to the other, a fact which should be taken into serious account if such powders are to be applied in practice. Table 5 also shows the endurances measured under a normal force of 200 N and at a speed of sliding of 0.25 m/s in, respectively, dry air and nitrogen with approximately 100 ppm oxygen. It runs out that the endurances of 1 µm thick layers of, respectively, coarse and fine MoS_2, is the same (i.e. 18 hours in dry air and more than 100 hours in nitrogen). Thus , the difference between coarse and fine powdered MoS_2, which is frequently claimed on the basis of practical experience, is entirely due to differences in rubbin-in time. Table 5 also shows the relative inferiority as far as the endurance of rubbed-in films is concerned of WS_2 and graphite. In accordance with the friction mechanism described above, the endurance of a graphite film in nitrogen is virtually nil, while under such conditions, WS_2 still performs moderately well. Because of the ease of layer formation during rubbing-in, WS_2 excels at relatively short rubbing-in times. This is shown in Table 6 which lists the endurance in dry air for layers formed during 1 minute rubbing-in procedures. This table shows that at such short rubbing-in times, coarse MoS_2 particles do not function at all.

POWDER	PARTICLE SIZE	RUBBING-IN[1] TIME	ENDURANCE [2] (h)	
	(μm)	(min)	DRY AIR	NITROGEN WITH 100 ppm OXYGEN
MoS_2	10-50	200	18[3]	> 100
MoS_2	1- 5	20	18[3]	< 100
WS_2	1-25	1	2[3]	22
GRAPHITE	1-10	20	4[4]	< 1

TABLE 5 ENDURANCE OF SOLID LUBRICANT FILMS, FORMED BY RUBBING-IN WET-BLASTED STEEL SUBSTRATES WITH DRY POWDER TO A LAYER THICKNESS (COMPRESSED) OF 1 μm

1. TIME NECESSARY TO FORM A LAYER WITH A THICKNESS OF 1 μm (IN THE FULLY COMPRESSED STATE)

2. TEST CONDITIONS: NORMAL FORCE: 200 N
   SLIDING SPEED: 0.25 m/s

3. DRY AIR

4. AIR WITH 60% REL. HUM.

RUBBING-IN PROCEDURE; NEOPRENE SPONGE UNDER 14 N FORCE

particle size (μm)	powder	endurance in air[1] (h)
10-50	MoS_2	< 0.1[2]
1- 5	MoS_2	1[2]
1-25	WS_2	2[2]
1-10	graphite	< 0.1[3]

TABLE 6 ENDURANCE OF SOLID LUBRICANT FILMS, FORMED BY RUBBING-IN WET-BLASTED STEEL SUBSTRATES WITH DRY POWDER DURING 1 MINUTE

1) TEST CONDITIONS: SEE LEGEND TABLE 5

2) DRY AIR

3) AIR WITH 60% REL. HUM.

The unfavorable effect of oxygen on the endurance of MoS_2 is once more shown in Figure 30. In a morphological sense, the intake of oxygen manifests itself by the formation of a membraneous like layer with a metallic luster. Such layers, which are to be considered as intermediate degradation products of MoS_2 powder films, suffer from dynamic blister formation. Figure 31 shows a magnified picture of a part of a blistered MoS_2 film. The blisters which can be seen have formed directly after the passage of the slider. Upon subsequent slider passage, they would have been flattened only to form again after repeated slider passage. Figure 31 clearly shows a crack formed in the central blister in a direction perpendicular to the direction of sliding. Such cracks ultimately lead to descaling of relatively large parts of the MoS_2 membrane.

5. MATERIALS FOR MINING, EARTH MOVING AND DREDGING EQUIPMENT

In mining, earth moving, and dredging equipment three wear processes predominate; abrasive wear, erosion, and impact wear.

5.1 Abrasive Wear

Situations of pure abrasive wearare frequently encountered in cases where a component moves through a loosely packed soil or where a load of abrasive material (for instance sand) is moved with respect to a guiding surface.

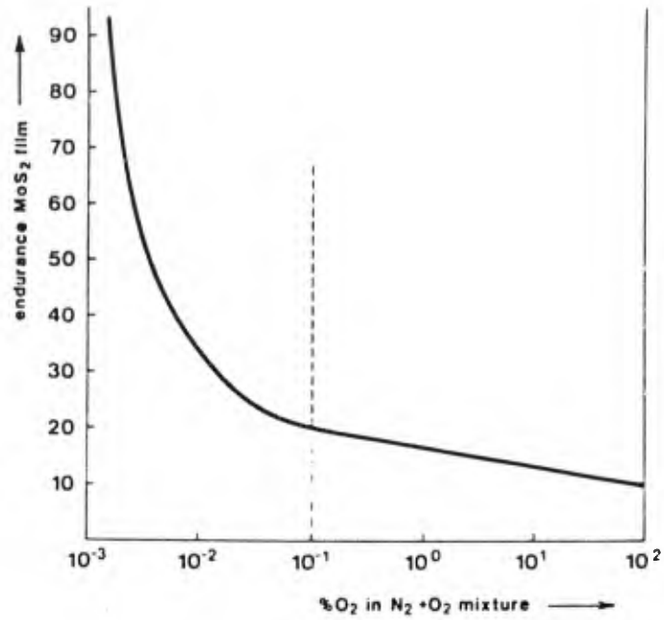


FIGURE 30 THE EFFECT OF OXYGEN ON THE ENDURANCE OF MoS_2 POWDER FILMS

FIGURE 31 BLISTERS IN AN MoS_2 POWDER FILM

Although, from a systems point of view, the abrasive wear mechanism is relatively simple or in any case well defined, it has taken a considerable period of time before a clear insight in the material factors which control the tribological behavior of materials under conditions of abrasion could be developed. Quite recently, Torrance developed an expression relating the abrasion resistance of pure metals or alloys to their hardness and Young's moduli[35]. His expression seems to agree well with a wide range of published experimental results. The wear resistance under conditions of pure abrasion is usually measured in tests in which a small metal cylinder or pin is weighed and then pressed under a standard load against a moving abrasive surface, generally a standard grade of abrasive paper. The abrasive is slowly traversed beneath the pin in such a way that the pin is always rubbed by fresh abrasive. The pin may describe a spiral on the abrasive disc or follow a helix on an abrasive drum. The test is stopped when the pin has traversed a standard distance. It is then reweighed and the volume, v_1, lost during the test, is calculated from the weight loss. The test is then repeated using a reference material and the relative wear resistance, β_i, of the metal is expressed as $\beta = v_r/v_i$; where v_r = the volume loss of the reference material. Use of a reference material is made, because differences in quality of the abrasive (paper) may result in appreciable differences in the absolute wear amounts of the test materials. If β_i is plotted against the Vickers hardness for a series of metals, a remarkable fact emerges; all the pure metals and annealed alloys lie on a straight line passing through the origin, but alloys, hardened by heat treatment, lie beneath this line on straight lines of lower gradient. This is shown in Figure 32, from Reference 36.

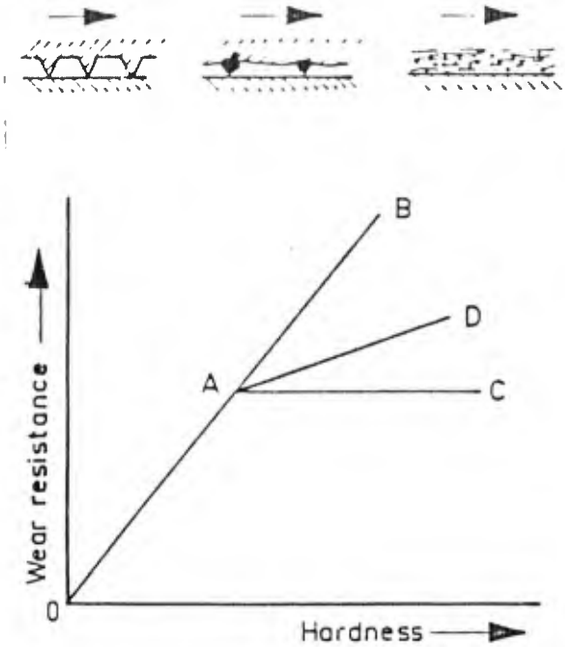


FIGURE 32 RELATION BETWEEN WEAR RESISTANCE AND HARDNESS FOR PURE METALS (LINE O-A-B), HEAT TREATED STEELS (LINE A-D) AND COLD WORKED MATERIALS (A-C)(AFTER KRUSCHOV, REF.35)

The anomalous behavior of heat treated alloys has been explained partly by Richardson, who confirmed that hardening a metal by cold working, has no effect on its abrasion resistance[37]. He also observed that the surface hardness of an abraded metal is considerably higher than its bulk hardness, because of the intense plastic strain which is induced by abrasion. He therefore suggested that the abrasion resistance of a metal correlates better with its hardness in a heavily worked condition than with its bulk hardness. Although experiments showed this to be the case, heat treated alloys were still found to fall on lines of lower gradient than pure metals.

A second approach was used by Buttery and Archard, who attempted to simulate the action of an abrasive grit by plowing scratches in the surfaces of various metals with a Vickers indenter[38]. By careful measurements made on Talysurf traces taken across the scratches, they tried to find how much of the scratch volume was pushed into sidewalls and how much was removed as wear debris. On the basis of these results, they suggested that an abrasive grit removes metal more effectively from a heat treated alloy than from a pure metal, in this way explaining in a semi-quantitative way, the lower than expected abrasion resistance of heat treated materials.

The above results were criticized by Johnson, who pointed out that although there is a tendency for metal, displaced by an indenter, to be ploughed plastically into sidewalls, it is possible for some of the displacement to be accommodated by a small elastic distortion of metal adjacent to the plastic zone[39]. Torrance took up this argument and developed the following relation for the total volume of material, removed by an abrasive[35]:

$$v = \frac{b\ Fs}{H\ \tan^2\theta}\left(1 + \frac{KH}{E}\right) \qquad (5)$$

in which b = the fraction of the groove volume removed as wear debris; F = the force applied to the test specimen; s = the length of the grooves ploughed by the abrasive particles; H = the Vickers indentation hardness of heavily worked material (i.e according to Richardson); θ = the semi-apical angle of the (supposedly conical) particles; E = the Young's modulus of the surface that undergoes abrasive wear; and K = a factor which in a first approximation can be considered as a constant for a wide variety of metals and θ-values. Since the wear resistance, β_i, is defined as:

$$\beta_i = v_r/v_i$$

Equation (5) gives:

$$\beta_i = \frac{H_i}{H_r}\,\frac{1 + KH_r/E_r}{1 + KH_i/E_i} \qquad (6)$$

in which the index, i, refers to the material under test and the index, r, to the reference material.

Although Torrance's equation is based on many assumptions and approximations, it correlates surprisingly well with experimental results, as is shown in Reference 35. In this reference the measured values of the wear resistance, β_i, are correlated with the values of β_i calculated with Eq. (6) for a large number of tests performed by Richardson, Moore and Mutton, and Watson[37, 40, 41]. It is shown that, despite the many simplifying assumptions, Eq. (6) gives a much better prediction of the relative abrasion resistance than was hitherto possible.

As stated in subsection 1.1, friction and wear do not usually correlate if different materials are compared in a single test. This applies equally well for abrasion as for other tribological situations. In abrasive wear, as occurring

in the cutting of tightly packed soil in a dredging operation, one particular construction material may come in contact with soils of different composition, surface structure, etc. In that case, the chances of finding workable correlations between friction and wear are much better. Figure 33, taken from Reference 42, shows that this is indeed the case. Correlations, as shown in Figure 33, open the possibility of monitoring cutter wear by means of power loss measurements. It was found that application of this principle in dredging practice actually resulted in improved efficiency in the use of cutter teeth.

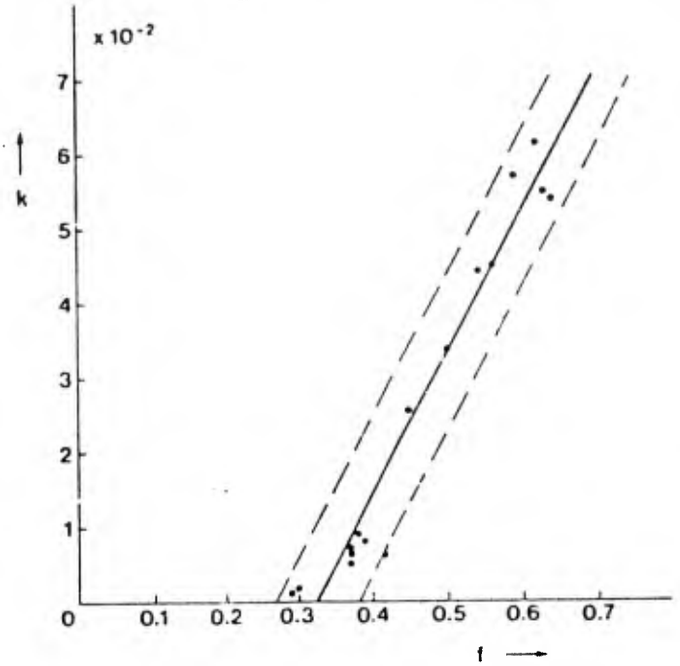


FIGURE 33 CORRELATION BETWEEN WEAR COEFFICIENT k AND COEFFICIENT OF FRICTION f, WHERE k = (VOLUME OF WEAR) (HARDNESS) / (LOAD) (SLIDING DISTANCE). ——, LINEAR REGRESSION CURVE, ----, CONTROL LIMITS AT TWO STANDARD DEVIATIONS TO EITHER SIDE OF THE REGRESSION CURVE. 95% OF THE DATA ARE EXPECTED TO LIE WITHIN THESE CONTROL LIMITS.

5.2 Impact Wear

Wear Phenomena associated with impact are described comprehensively by Engel[43]. In mining or earth moving operations, impact wear can be due to collisions between metal components and hard rocks. Under those conditions, the resistance of the material against impact induced surface fatigue appears to be of crucial importance. This necessarily implies that under rock-moving conditions, the materials should be hard and ductile, a state of affairs which is, in fact, desirable in the entire field of wear, but which is not obtained easily. In the application area under discussion, it is in

particular the manganese steels which have emerged as wear resistant materials with unique properties. In fact, the favorable wear behavior of these materials, the "Hadfield steels", is due to the fact that they are open to the formation of deformation-martensite as a result of mechanical work. Recently, the situation has been summed up by Zum Gahr et al[6]. The authors show beyond a doubt that there may be pronounced differences between steels from a single "family", only by virtue of the fact that some can form deformation martensite and others cannot. Actually this is a fine example of a case in which, within a particular class of materials, a theoretical model based on metallurgical knowledge may adequately predict the field behavior of new alloys.

5.3 Erosion

Erosion is loss of material from a surface due to contact with particles which are carried in a fluid. Practical cases in which the lifetime of components is limited by erosion are encountered in application areas which range from dredging operations (cold; seawater) to gas turbine applications (hot; corrosive gasses).

In erosion the mechanisms of abrasion and impact wear usually occur simultaneously, the relative importance of each of these mechanisms depending on the predominant angle of impact, α, under which the eroding particles hit the wearing surface. Obviously this complicates the picture quite considerably, meaning that a systems approach is even more appropriate in the case of erosion than in the case of pure abrasion or pure impact wear. Another factor of considerable influence is formed by the presence of the medium which carries the particles, which may be either gas or liquid and which may exert an independent corrosive action on the wearing surface. Other important parameters determining the behavior of materials under conditions of erosion are the particle concentration, the nature and distribution of the eroding particles, the speed, and the temperature.

For the case of aluminumoxide particles in an air stream attacking aluminum and plain carbon steel surfaces, it has been attempted successfully to set up theoretical models for the relation between erosion and angle of attack, α [44, 45]. Figure 34 summarizes the results of this work in terms of erosion versus α curves for extremely ductile (1), extremely brittle (2) and "ordinary" materials (3). From Figure 34 it can be seen that when working with ordinary engineering materials, as for instance steels, one finds a maximum in the erosion curve, which usually falls in the range $20^{\circ} < \alpha < 40^{\circ}$. The precise location of the maximum, however, depends quite strongly on the

properties of the wearing surface and on the precise experimental conditions, which limits the predictive power of the theoretical model quite severely. In fact it is not likely that a general erosion-theory, which covers the effects of all influencing factors (including corrosion), will ever be developed. For that reason, for the majority of new technical applications, one has to rely on simulative testing in selecting suitable materials[46].

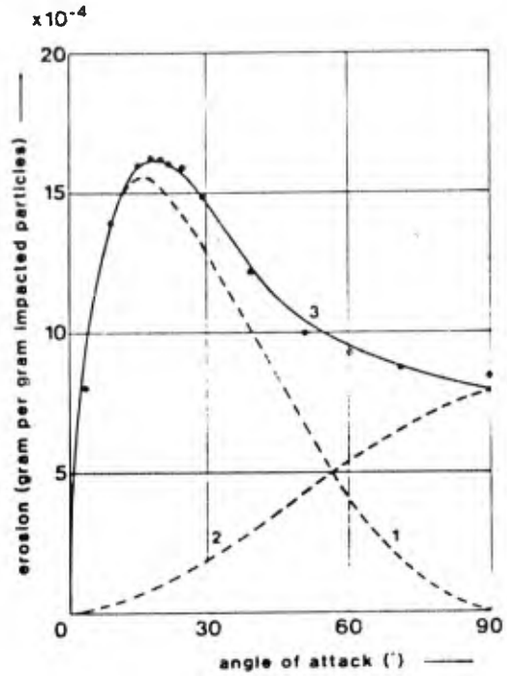


FIGURE 34 EROSION VERSUS ANGLE OF ATTACK CURVES FOR DIFFERENT TYPES OF ERODING MATERIAL (AFTER NEILSON & GILCHRIST, REF. 44)

1: IDEAL DUCTILE MATERIAL (THEORETICAL CASE)
2: EXTREMELY BRITTLE MATERIAL (FOR EXAMPLE, GLASS)
3: ORDINARY ENGINEERING MATERIAL (IN THIS CASE ALUMINIUM, ERODED BY 210 μm ALUMINIUMOXIDE PARTICLES IN AN AIR STREAM AT V = 125 m/s.

A typical result, obtained in a current research program, performed at the TNO Laboratory, in which different materials are compared with respect to their behavior under conditions of erosion in particle-seawater slurries, is shown in Figure 35. This figure shows the relative wear rates, β_i, for "rubber" and "polybutene", calculated by relating their volume losses to that of plain carbon steel C22 (Vickers hardness 150). Actually, if the volume loss of the reference steel exceeded that of the test material, a factor β_{i+} was calculated according to:

$$\beta_{i+} = \frac{v_{steel}}{v_{test\ mat.}}$$

If the volume loss of the test material exceeded that of the reference material, a factor β_{i-}, was calculated according to:

$$\beta_{i-} = \frac{v_{test\ mat.}}{v_{steel}}$$

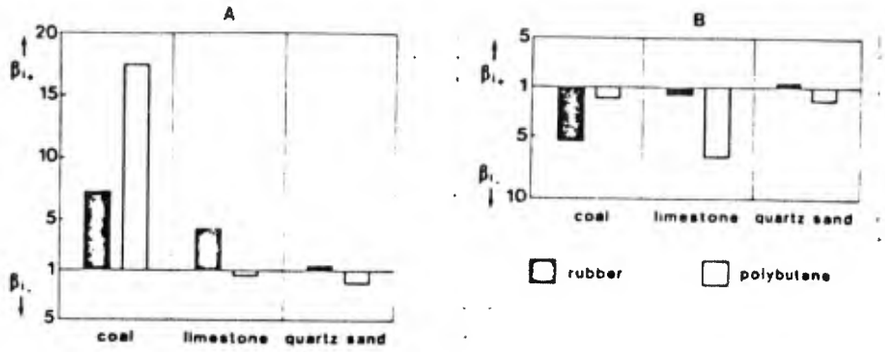


FIGURE 35 RESULTS OBTAINED IN TESTING "RUBBER" AND "POLYBUTENE" IN, RESPECTIVELY, COAL-SEAWATER, LIMESTONE-SEAWATER AND QUARTZSAND-SEAWATER SLURRIES.

THE PICTURE CONTAINS RELATIVE WEAR RATES, OBTAINED BY COMPARING THE WEAR OF THE TEST MATERIALS WITH THE WEAR OF PLAIN CARBON STEEL C22 (150 HV).

A: RESULTS OBTAINED UNDER CONDITIONS OF ELECTROCHEMICAL CORROSION

B: RESULTS OBTAINED UNDER CONDITIONS WHERE ELECTRO-CHEMICAL CORROSION IS INHIBITED EFFECTIVELY.

In the present tests, the abrasive agents were ground coal, limestone, and quartz sand, each with an average particle diameter about 600 μm. The other test parameters were: concentration of abrasive, 19 vol.%; angle of attack, α: 80°; slurry speed, 4 m/s; and exposition time, 6 h. Part A of Figure 35 shows results obtained under conditions in which (for steel

C22) electrochemical corrosion could occur; pieces of other metal attached to the test specimen. Part B of Figure 35 shows results obtained with electrically isolated test specimens. It can be seen that in quartz sand/seawater mixtures, the results obtained under "corrosion conditions" (part A of Figure 34), are the same as those obtained under "noncorrosion conditions" (part B of Figure 34), with rubber performing slightly better and polybutene perfoming worse than steel C22 by a factor of approximately 2. Apparently in quartz sand/seawater slurries, electro-chemical corrosion effects do not contribute significantly to the process of material removal. Figure 34 also shows that this is quite different in coal/seawater and limestone/seawater mixtures where the results obtained under "corrosion conditions" differ considerably from those obtained under "noncorrosion" conditions.

Figure 35 only gives an impression of the situation at particular values of angle of attack α, particle size (distribution), and abrasive concentration. Variations in the values of each of these parameters considerably influence the results.

Results as shown in Figure 35 can be applied in practice by first performing tests under real operating conditions in prototype equipment made from the reference material (in this case steel C22), and then applying materials with high β_{i+} values at locations which are found to suffer from excessive wear. Such selective application of better, but generally also more expensive, materials may contribute significantly to the overall performance of the installation, paying due regards to economic considerations.

In the present case, it would not be wise to apply rubber or polybutene coatings for use in coal or limestone slurries if, in practice, electrochemical corrosion can effectively be suppressed (in quartz sand, however, application of rubber might improve the situation somewhat). If, however, electrochemical corrosion also plays an important part in the practical application, the use of rubber or polybutene (in coal-seawater slurries) or rubber (in limestone-seawater slurries) will considerably improve the situation.

6. THE TRIBOLOGICAL ASPECTS OF SURFACE TREATMENTS AND COATINGS

Thermal and thermochemical treatments of surfaces and coating techniques become increasingly important because of two reasons:

- the demands on machines and equipment in relation to exposition to hostile environments and high temperatures increase,
- the need for cheaper and more readily available construction materials becomes more pressing.

Actually when applying a high quality surface treatment or a coating to a relatively simple substrate material, it may be possible to reconcile the above requirements effectively.

By now, a large number of coatings and surface treatments are available, which have as a rule, been developed and exploited independently. Thus it is difficult, if not entirely impossible, to compare the features of the different processes on a really objective basis. Still, this section tries to provide a rough guide for process selection, which should be followed by function-oriented tribometry and careful weighing of the tribological as well as the economical factors, if the best possible solution for a given practical application is to be found[5].

Much of the information contained in this section is derived from a multi-client study, performed by the International Research and Development Corporation, Ltd[47]. However, experiences gained at TNO and information from other literature sources are included as well.

6.1 Selection Criteria

Although the "wear resistance" undoubtedly is the most important criterion in selecting a suitable surface treatment or surface coating, other properties will be more or less important, depending on the details of the practical application. Actually, in each application, each selection criterion should be allotted a "weighing factor." In addition to "wear resistance," the following criteria may come into play:

- layer thickness
- workpiece temperature during treatment
- porosity
- hardness
- ductility
- adherence
- price
- corrosion resistance
- "fatigue strength"
- maximum service temperature

The layer thickness should be adapted to the requirements of the practical application. As an example, in the case of precision

measuring equipment, one will generally apply highly wear resistant treatments or coatings of relatively small thickness, because the admissible amount of wear usually is very small. On the other hand, in equipment for the dredging industry one will usually apply surface welded coatings of considerable thickness (up to 20 mm). A quantitative example of the interaction wear rate/layer thickness/type of application is given in the following.

Table 7 shows data relevant to surfaces treated by gasnitriding and hard chromizing, both processes belonging to the family of "thermochemical treatments." Under the present conditions of load, speed, lubrication, and temperature, the wear rate of the gasnitrized surfaces was 0.2 μm per day and that of the surfaces treated with chromium 0.02 μm per day (Table 7, column 3). The wear rate of the substrate material was 50 μm per day in both cases. The effective layer thickness of the gasnitrided surfaces was 300 μm and that of the surfaces treated with chromium, 20 μm. From the data, it follows that at a maximum allowable depth, Δh_{max} of 300 μm, the liftime of the gasnitrided surface, t_{max}, is about four years. On the other hand at a maximum allowable wear depth, Δh_{max} of 20 μm, t_{max} reduces to only three months (0.27 year). In the case of surfaces treated with the hard-chromizing process, the maximum allowable wear depth is irrelevant, because in both cases the lifetime of the layer is 2.7 years, being the time necessary to remove the chromium diffusion layer by wear.

From this example we see that if the requirements are:

- lifetime : 4 years
- maximum wear depth : 300 μm

gasnitriding is the appropriate process.

If, on the other hand, the requirements are:

- lifetime : 2 years
- maximum wear depth : 20 μm

the hard-chromizing process is to be chosen.

process	layer thickness (μm)	wearrate (μm/day)		t_{max} (year)	
		layer	substrate	Δh_{max} = 300 μm	Δh_{max} = 20 μm
gasnitriding	300	0.2	50	4.1	0.27
chromizing	20	0.02	50	2.7	2.7

TABLE 7 LIFETIME t_{max} OF, RESPECTIVELY, GASNITRIDED AND CHROMIZED SURFACES FOR TWO VALUES OF THE MAXIMUM ALLOWABLE WEAR DEPTH Δh_{max}

Although in reality, the situation may be much more involved than is suggested in Table 7, if only because more processes are included in the comparison, the example shows that data on the wear rate alone are not sufficient to determine which process is most suitable for a particular application. The workpiece temperature during performance of the surface treatment or the application of the surface coating is important because a relatively high temperature may lead to severe distortions of the workpiece. In general, there will be a tendency to apply the high temperature treatments only in cases where a certain degree of distortion is acceptable. If this is not the case, one has to reckon with additional costs of machining of the surface treated layer or coating and also with possible undesirable local changes in layer thickness as a result of machining. The porosity of a coating can influence its mechanical strength and thereby its wear resistance. Also pores, which connect the coating surface with the substrate surface, may transmit a corrosive environment which may lead to interface corrosion. Obviously hardness and ductility are mutually related properties. As stated previously for most applications, one wishes to obtain a high hardness and a high ductility. In practice this usually means that an acceptable compromise has to be found. The importance of the criteria "adherence" and "price" is self-evident. Corrosion resistance, "fatigue strength" and maximum service temperature are typical system-dependent properties. It depends much on the type of application, i.e. the composition of the environment, the temperature, the loading pattern, etc., whether special requirements as to these criteria are to be met.

From the above, it will be clear that a surface treatment or surface coating which behaves excellently under one particular set of conditions may fail utterly under other conditions. Thus commercial claims regarding "universally

applicable coatings" should be treated with a healthy scepticism.

6.2 A Survey of Surface Treatment and Surface Coating Techniques

The (commercially) available surface treatment and surface coating techniques which are available to the designer of wear resistant surfaces are summarized in Figures 36 and 37 respectively. It can be seen that one may choose from a great amount of tehnical possibilities. Actually, the situation is even more involved because in certain cases many competitive commercial processes are on the market which may or may not lead to essential differences in quality. Obviously, the scope of this presentation does not permit a detailed discussion of each of the individual processes mentioned in Figures 36 and 37. Instead Table 8 provides for each of the main categories information on the following criteria:

- maximum layer thickness, that can be obtained within reasonable limits of economic procedure
- maximum temperature of the workpiece during treatment
- maximum hardness that can be obtained.

Further, Table 8 lists for each of the main categories some typical examples of materials suitable for treatment or coating materials and also some typical applications.

If ranges are given for the maximum values of layer thickness and workpiece temperature or hardness, this indicates that different processes, which belong to one and the same category, yield different values for these parameters. For instance in the category "thermal hardening", the process of "flame hardening" yields a hardening depth up to 5000 µm; laser hardening, on the other hand, produces a maximum hardening depth of some 500 µm. In some cases the differences thus found are even more extreme. For instance, the case within the group or "thermochemical treatments" is cited. Here the "chromizing" treatment yields a maximum layer thickness of 20 µm, while "carburizing" may easily yield layers with a maximum thickness of 4000 µm. It should finally be noted that the examples as to coating materials and characteristic applications are by no means exhaustive. In fact the more recently developed processes such as, certain vapor deposition methods and electrochemical treatments, followed by thermal diffusion may find wider applications in the near future[48, 49].

Table 8 shows that, in principle, a confusingly large amount of different surfaces can be obtained, which makes the ultimate choice for a given practical application difficult.

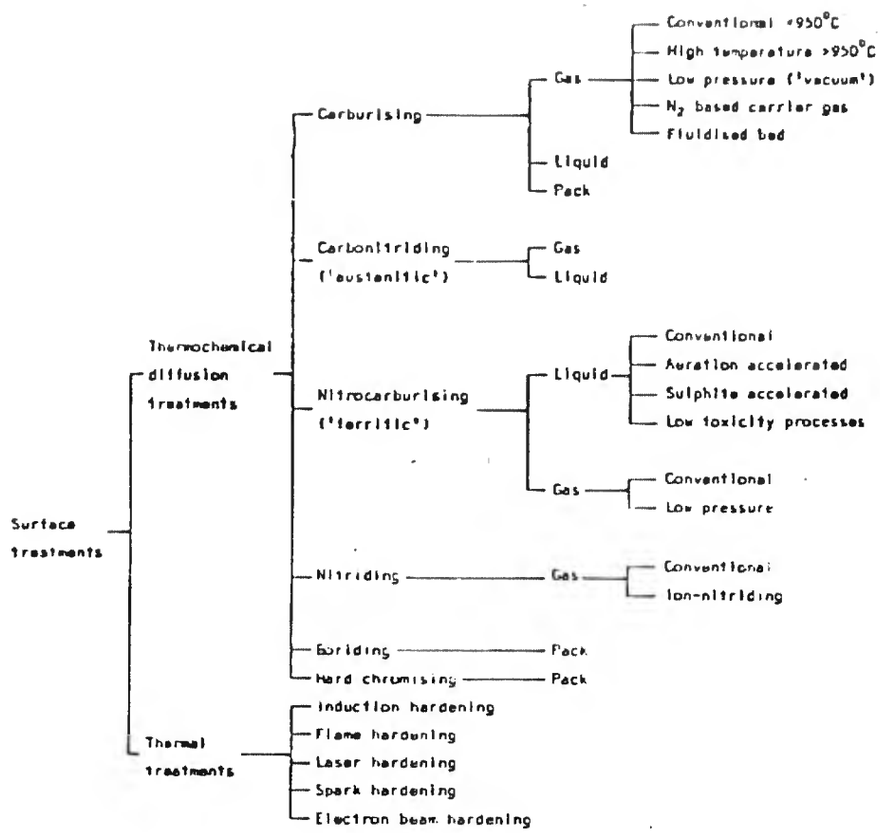


FIGURE 36 SURVEY OF SURFACE TREATMENT TECHNIQUES

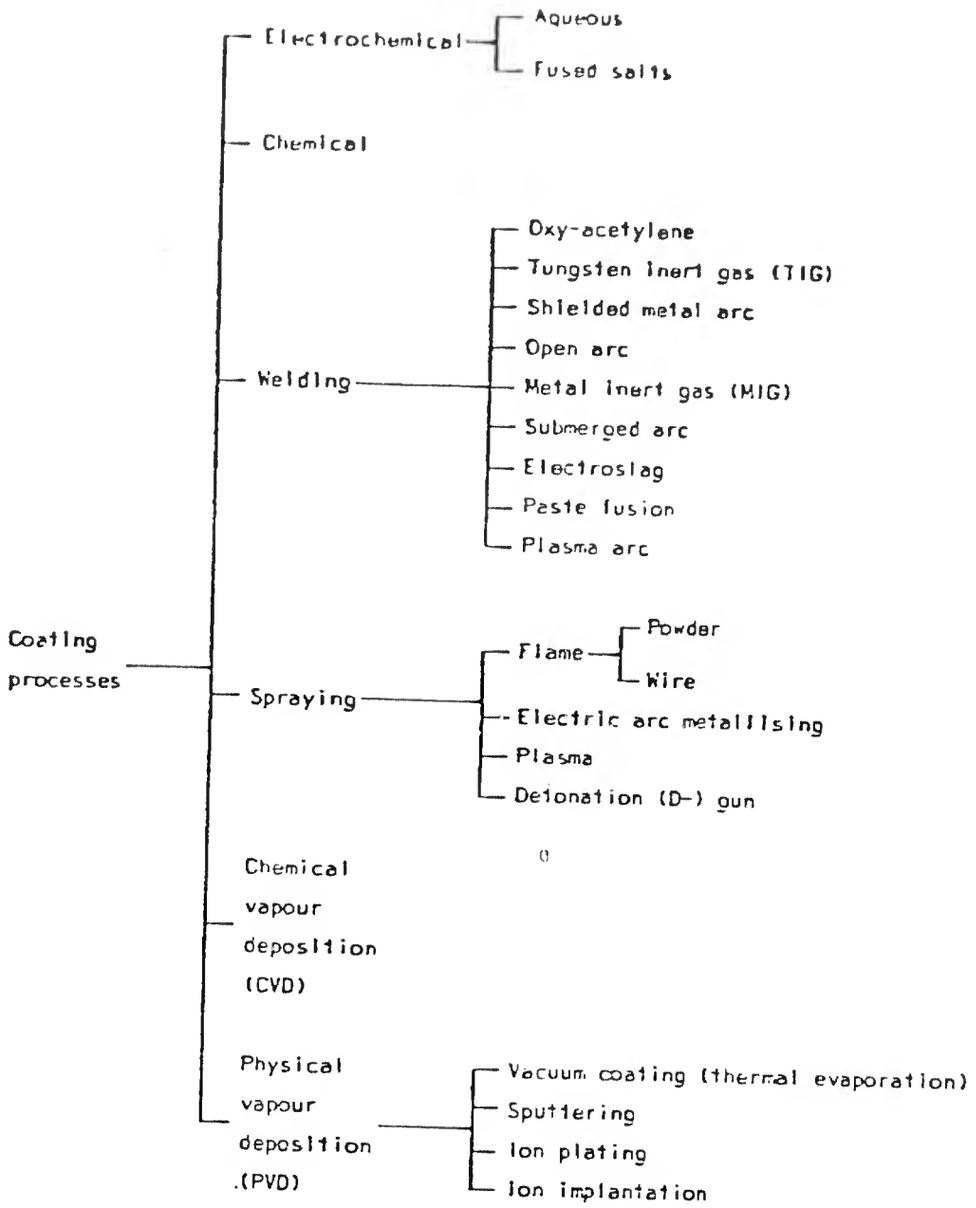


FIGURE 37 SURVEY OF SURFACE COATING TECHNIQUES

TABLE 8 A SURVEY OF SURFACE TREATMENT AND SURFACE COATING TECHNIQUES

process	maximum thickness coating (μm)	maximum temperature workpiece (°C)	some typical examples of coating materials	maximum hardness	some characteristic applications
thermal hardening	500 - 6000	100 - 1000	Steel (0,3% - 0,5% C)	500 - 900	automobile engine components
thermochemical treatments	20 - 4000	400 - 1000	Steel + C, N, B, Cr or V	500 - 2000	gears, roller bearing races, crankshafts
electrochemical treatments	50 - 500	50	Ag + Rh	150	rotary switches
			Cr	1000	cylinder liners
			Cr_3C_2 / Co	2500/500	diesel engine piston rings
			Al_2O_3 on Al (anodizing)	(1000)	plastic injection moulds
chemical treatments	5 - 50	50 - 200	NiP or NiB	950	casting dies, thread guides
			$MnPO_4$ / $Mn_3(PO_4)_2$ (phosphating)	380	gears, tappets, cam shafts
welding	20000	1400	12% Mn steel	600	earth and rock engaging equipment; dredging cutter heads
			martensitic steels	900	
			Co-alloys	700	
			WC/Ni	2000/700	
spraying	1000	100 - 1100	13% Cr steel + oxides	350	shafts, valve seats, dies
			Mo + oxides	400	
			NiAl	350	
			Co-alloys	700	
			WC/Co or Cr_3C_2/NiCr	2000/1000	
			Al_2O_3 or Cr_2O_3	2500	
chemical vapour deposition (CVD)	10	1000	TiC, TiN, W_2C, WC, Al_2O_3, etc.	3000	metal cutting tools wire drawing dies
physical vapour deposition (PVD)	5 - 150	20	TiC, TiN, W_2C, WC, B_4C, Si_3N_4, etc.	4000	metal cutting tools compressor blades
			Al, Pb, In, W	100 - 1000	ball bearings, shafts
			C, N	1500	wire drawing dies

The issue is further complicated by the fact that, in a particular application, surfaces which have been treated by different processes may well be combined. That this can be very beneficial is illustrated in Figure 38. This shows that the total amount of volume wear, produced in 20 hours testing under the conditions specified, differs significantly for combinations of nitrocaburized (NC) and carburized (C) surfaces, with the latter combination producing about twice the amount of wear as the first combination. Remarkably enough, the volume wear of both surfaces, but in particular that of the carburized surfaces, but in particular that of the carburized surfaces, is reduced considerably if combinations or carburized and nitrocarburized surfaces are made. In fact Figure 38 shows that if in particular the shaft, item B in Figure 38, is to be protected, the combination carburized shaft --- nitrocarburized mating surface yields an approximately 25 times better performance than a combination carburized shaft - carburized mating surface. In a way this is still another example of the "compatability effect" that was also described in subsection 2.3 for combinations of metals and in subsection 4.5 for combinations of plastics.

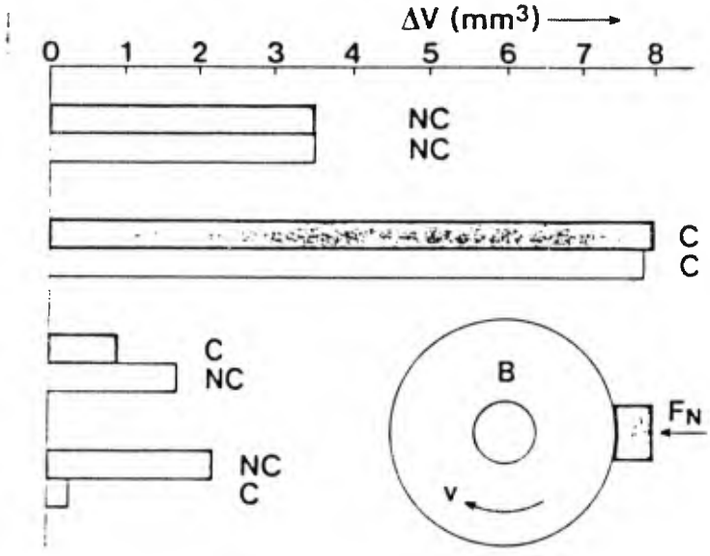


FIGURE 38 VOLUME WEAR MEASURED WITH A PIN (A) AND RING (B) TEST RIG, AN ILLUSTRATION OF THE "COMPATIBILITY EFFECT"

NC: NITROCARBURIZED LAYER
C : CARBURIZED LAYER
NORMAL FORCE: 1000 N, SPEED OF SLIDING: 0.2 m/s, EXPERIMENT
DURATION: 20 HOURS, ENVIRONMENT: DRY AIR OF 200°C

6.3 Thermochemical Processes

In Table 9 the main properties and functional characteristics of the thermochemical treatments carburizing, carbonitriding, nitrocarburizing, nitriding, boriding, and chromizing are summarized. For comparison, the same properties are given for laserhardening (category "thermohardening"), hard chromium plated surfaces (category "electrochemical processes") and WC/Co coatings (category "spraying"). Typical performance characteristics of the various treatments and coatings are the maximum surface temperature, the fatique strength ,the wear resistance under conditions of abrasive and adhesive wear, and the corrosion resistance. The maximum service temperature of a component which has been surface treated or coated depends on a number of different factors, some of which are system oriented. These are:

- Strength/temperature relation of the substrate. As a result of softening of the substrate the component as a whole may become inoperable, even if the coating itself still functions adequately.
- Diffusion into the substrate material of one or more of the elements which give the surface its special properties, resulting in a disappearance of the coating into the substrate.
- Reaction with the surrounding atmosphere (usually oxidation).
- Decrease in hardness of the treated surface or the surface coating with increasing temperature.

Depending on the type of treatment and the type of substrate material, one or more of the above factors will predominate and control the maximum service temperature. In column 5 of Table 9 approximate values of this maximum servie temperature are given, under the assumption that the first of the above mentioned factors (strength-temperature relation of the substrate) does not interfere with the function of the component. In considering the practical consequences of the maximum service temperature, it should be borne in mind that the temperature in the friction interface can be considerably higher than the environmental temperature, because of frictional heating effects. Thus in deciding whether a particular treatment or coating can be applied the real surface temperature should be estimated.

Because of the very strongly system oriented character of the properties "fatigue strength", "wear resistance," and "corrosion resistance," it is not possible to give for these properties quantitative data (although, in treatment selection for a well-defined practical application, function oriented

process	thickness coating (µm)	maximum temperature workpiece (°C)	maximum hardness HV	"fatigue strength"	"wear resistance" abrasive	"wear resistance" adhesive	"corrosion resistance"	maximum service temperature (°C)
laser hardening	500	100	500 - 900	++	+	+	±	200
carburizing	250 - 4000	850 - 950	700 - 900	+++	+	+	±	200
carbonitriding	50 - 750	750 - 900	600 - 850	++	+	++	±	200
nitrocarburizing	20*, 1000**	570	500 - 650*	++	+	++(+)	±	200
nitriding	400 - 600	500 - 525	800 - 1050	+	++	+++	+	500
boriding	25 - 100	850 - 1000	1000 - 2000	+	+++	+(+)	+	500
chromising	10 - 20		1500 - 2000	+	+++	+++	+++	600
chromium plating	2 - 250	50	1000	-	++	+++	+++	500
WC/Co plasma spraying	500	250	2000/1000	±	++	++	+	500

*thickness and hardness of compound layer on mild steel
**total depth of diffusion zone

TABLE 9 IMPORTANT PROPERTIES OF SOME SURFACE TREATED OR COATED SURFACES

tribometry may produce such data). As far as the "fatigue strength" is concerned, Table 9 shows that the thermochemical treatments as well as laser hardening (and "thermohardening" in general) lead to an increase in fatigue strength of the surface. This is because during the formation of martensite in the surface zone (as occurs during thermal hardening) as well as by diffusion of carbon, nitrogen, borium, or chromium in the surface, compressive stresses are built up in the surface zone, which counteract the formation of microcracks. Table 9 shows that, in particular, carburizing leads to an appreciable increase in "fatigue strength" of the surface. As far as this point is concerned two reservations should be made:

- If the surface treated components are finished by grinding, which is usually the case after carburizing, carbonitriding, and nitriding, the grinding process may, but not necessarily does, cause tensile stresses in the surface which may nullify the favorable effects of the surface treatment itself.

- If possible, a sharp transition between surface treated zone and substrate material should be avoided, because this may lead to spalling phenomena. This is in particular the case if the surface treated component is to be applied under conditions of cyclic loading, as in rolling element bearings. In any case it should be avoided that the location of the zone with maximum stress coincides with the transition area.

Table 9 shows that plasma sprayed WC/Co layers usually do not influence the fatigue strength of the component. Hard chromium plating, on the other hand, may lead to a significant decrease in fatigue strength because cracks in the chromium layer may initiate cracks in the substrate. This can be remedied by applying an intermediate layer, acting as crack barrier (for instance a thin nickel-chromium layer). However, this leads to a considerable increase in costs.

As far as the "wear resistance" is concerned one should distinguish between the resistance against "abrasive wear" and that against "adhesive wear". If abrasive wear predominates, hardness and elastic modulus are the predominating factors. As far as adhesive wear is concerned, diminishing friction is of equal importance. In this respect the nitrocarburizing, nitriding and chromizing processes distinguish themselves favorably from the other processes. Table 9 also shows that, as far as "wear resistance" is concerned, hard chromium plating and WC/Co plasma spraying in general are a match for the thermochemical process. Table 9 finally shows that the corrosion resistance of the treated surfaces is generally

similar or somewhat better than that of the untreated substrate. Conforming to expectations, chromium containing surfaces, obtained by chromizing or chromium plating of surfaces, generally show an excellent corrosion resistance because of the intrinsic favorable properties of chromium. The corrosion resistance of WC/Co sprayed layers depends on their porosity. If porosity is very low (c.f. 50) or if a porous layer is sealed with a resin, a very good resistance against corrosion can be obtained.

6.4 CVD and PVD Processes

In the seventies chemical vapor deposition (CVD) and physical vapor deposition (PVD) have rapidly gained a foothold in the world of surface coating techniques. In chemical vapor depsoition a very dense, hard, and well-adhering coating is formed on the surface of a substrate, heated to a temperature of approximately 1000°C, by reaction of gaseous compounds[48]. The most frequently applied materials are listed in column 4 of Table 7; suitable substrate materials are steels, cermets, and carbides. The layer thickness is limited to approximately 10 µm, because unacceptable stresses build up in thicker layers. The equipment necessary for performing the CVD process consists in principle of a vaporizer in which reactive gasses are added to a carrier gas, a container in which the substrate to be coated is placed and in which the chemical reactions occur and a heat supply system. The primary reaction partners are usually liquid. The carrier gas, usually hydrogen, is led through the liquids during which it becomes saturated with the relevant vapors. Heating takes place by induction or radiation.

A typical example is the formation of titaniumcarbide via the reaction:

$$TiCl4 + CH4 \rightarrow TiC + 4HCL$$

in which hydrogen is used as a crrier gas. At atmospheric pressure, dense TiC layers are formed on the substrate at temperatures in the range of 900°- 1200°C, at a rate of layer formation of approximately 5 µm/hour. The deposition temperature can be lowered to 700°C by working under reduced pressure of approximately 0.1 kPa. The structure of the TiC coating tends to be columnar and this leads to easy crack propagation and consequent lowering of transverse rupture strength of the substrate, a decrease of 40% having been measured with cemented carbides. Attempts to reduce this effect, by the use of catalysts which increase the number of nucleation sites on the substrate, have been successful[52]. Up till now, the use of CVD coated surfaces has virtually been restricted to tool tips. However, in the late seventies a

number of interesting tribological applications became known[53]. These include rolling element bearings, sliding gyroscope bearings, and measuring equipment.

A typical advantage of the CVD process is its excellent "throwing power", by which an irregularly shaped component can be covered on all sides with a layer of uniform thickness.

Physical vapor deposition (PVD) processes are always performed at low pressure, the vapor to be deposited being formed by thermal evaporation or electrical emission ("sputtering") of a suitable material, in some cases in the presence of a chemically active gas. The desired coating forms by condensation at the substrate surface. All PVD processes have in common that one needs relatively expensive equipment for their performance which is also energy intensive. On the other hand, the coating quality of some PVD processes is superior. In particular this holds for the newer process of ion-implantation[54]. In this process ions are accelerated by voltages of 100 - 150 kV and are focused on the substrate to be coated. By the high velocity with which the ions hit the substrate, they can penetrate to a depth of approximately 1 μm. This yields a very good adhesion. A typical example is dies for wire drawing made of WC, the endurance of which was increased by a factor of 10 by ion implantation of carbon.

7. REFERENCES

1. Salomon, G., "Recent Advances in the Application of Systems Thinking in Tribology," lecture presented at the eighth meeting of the International Research Group on Wear of Engineering Materials, Saint-Ouen (Paris), May 1981; to be published (text can be obtained from the Technical Secretariat, c/o TNO, P.O. Box 541, 7300 AM Apledoorn, The Netherlands).

2. Czichos, H., "Tribology, a Systems approach to the Science and Technology of Friction, Lubrication and Wear," Elsevier, Amsterdam, 1978.

3. Begelinger, A. and de Gee, A.W.J., "Synopsis of Results from an International Co-operative Wear Programme," Lubrication Engineering 26 (1970), 56-63.

4. Vaessen, G.H.G. and de Gee, A.W.J., "Boundary Lubrication of Bronzes - Metallurgical Effects," Transactions ASLE 16 (1973) 203-207.

5. De Gee, A.W.J., "Selection of Materials for Tribotechnical Applications - The role of Tribometry," Tribology International, August 1978, 233-239.

6. Bauschke, H.M., Hornbogen, E, and Zum Gahr, K.H., "Abrasive Wear of Austenitic Steels," Aeitschr. fur Metallkunde 72 (1981), 1-13 (in German).

7. ISO/DP7148 "Testing of the Tribological Behavior of Bearing Materials for Oil Lubricated Applications; Part I: Coefficient of Friction and Wear Rate under Conditions of Boundary Lubrication and Ample Supply of Lubricant to the Friction Couple," Copies can be obtained from: H. Tepper, Deutsches Institut fur Normung e.V., Kamekestrasse 8, D-5000 Koln 1, Germany.

8. De Gee, A.W.J., "Selection of Materials for Lubricated Journal Bearings," Wear 36 (1976), 33-61 and Wear 42 (1977), 251-261.

9. Habig, K.H., Borszeit, E., and de Gee, A.W.J., "Friction and Wear Tests on Metallic Bearing Materials for Oil Lubricated Bearings," to be published (Wear 1981).

10. Rabinowicz, E., "The Influence of Compatability on Different Tribological Phenomena," Transactions ASLE 14 (1971), 206-212.

11. Buckley, D.H., "Fretting in Aircraft Systems," AGARD Conference Proceeedings Nr. 161 (1975), 13/1-13/15.

12. Begelinger, A. and de Gee, A.W.J., "Abrasive Wear of Bearing Materials," Proceedings 2nd Eurotrib Conference, Dusseldorf, October 1977, Band I.

13. Scott, D., "The Effect of Steel Making, Vacuum Melting and Casting Techniques on the Life of Rolling Bearings," Vacuum 19 (1969), 167-169.

14. Scott, D., and McCullagh, P.H., "The Role of Nitrogen Content on the Rolling Contact Fatigue Performance of EN31 -Ball Bearing Steels," Wear 25 (1973), 339-344.

15. Scott, D., "New Materials for Rolling Mechanisms," Wear 43 (1977), 71-87.

16. Vaessesn, G.H.G. and de Gee, A.W.J., "Rolling Contact Fatigue of Maraging Steels of Different Production History, Influence of Film Thickness/Roughness Ratio," Paper C7, Second Symposium on Elastohydrodynamic Lubrication, I. Mech. E., Leeds, 1972.

17. Dowson, D. and Higginson, G.R., "Elastohydrodynamic Lubrication," International Series in Material Science and Technology, Vol. 23, Pergamon Press, 2nd Edition, 1977.

18. McCool, J.I., "Load Ratings and Fatigue Life Prediction for Ball and Roller Bearings," Transactions ASME, Journal of Lubrication Technology, January 1970, 16-22.

19. Salomon, G., "Failure Criteria in Thin Film Lubrication," Wear 36 (1976), 1-6.

20. Czichos, H., "Failure Criteria in Thin Film Lubrication; Investigations of Different Stages of Film Failure," Wear 36 (1976), 13-17.

21. Begelinger, A. and de Gee, A.W.J., "On the Mechanism of Lubricant Film Failure in Sliding Concetrated Steel Contacts," Transactions ASME, Journal of Lubrication Technology 98 (1976), 575-579.

22. Begelinger, A., de Gee, A.W.J. and Salomon, G., "Failure of Thin Film Lubrication - Function-Oriented Characterization of Additives and Steels," Transactions ASLE 23 (1980), 23-34.

23. Begelinger, A. and de Gee, A.W.J., "Failure of Thin Film Lubrication - The Effect of Running-In on the Load Carrying Capacity of Thin Film Lubricated Concentrated Contacts," Transactions ASME, Journal of Lubrication Technology 103 (1981), 203-211.

24. Begelinger, A. and de Gee, A.W.J., Thin Film Lubrication of Sliding Point Contacts of AISI 52100 Steel," Wear 28 (1974), 103-114.

25. Czichos, H., "Failure Criteria in Thin Film Lubrication; the Concept of a Failure Surface," Tribology International 7 (1974), 14-20.

26. Begelinger, A. and de Gee, A.W.J., "Failure of Thin Film Lubrication - A Detailed Study of the Lubricant Film Breakdown Mechanism," to be published in Wear.

27. Fowle, T.I., "Gear lubrication: Relating Theory to Practice," Lubrication Engineering 32 (1976), 17-34.

28. De Gee, A.W.J. and Vaessen, G.H.G., "A Note of Bayer and Ku's Model for Zero Wear," Wear 18 (1971), 492-496.

29. Muyderman, E.A., "Grease Lubricated Spiral Groove Bearings," Tribology International, June 1979, 131-137.

30. Lancaster, J.K., "Dry Rubbing Bearings," in Tribology Handbook, ed. M.J. Neale, Butterworth, London, 1973.

31. Crease, A.B., "The Wear Performance of Rubbing Bearings - Improved Data for Design," Paper 10, Proceedings 3rd Leeds-Lyon Symposium, Leeds, I. Mech. E., London, 1976.

32. Erhard, G. and Strickle, E., "Gleitelemente aus thermoplastischen Kunststoffen," Kunststoffe 62 (1972), 1-9 (in German).

33. Lancaster, J.K., "Dry bearings: A Survey of Materials and Factors Effecting Their Performance," Tribology International 6 (1973), 219-251.

34. Salomon, G., Begelinger, A., van Bloois, F.I., and de gee, A.W.J., "Characterization and Tribological Properties of MoS2 Powders and Related Chalcogenides," Transactions ASLE 13 (1970), 134-147.

35. Torrance, A.A., "The Correlation of Abrasive Wear Tests," Wear 63 (1980), 359-370.

36. Krushcov, M.M., "Resistance of Metals to Wear by Abrasion as Related to Hardness," Proc. Conf. on Lubrication and Wear, 1957, I. Mech. Eng., London, 46-55.

37. Richardson, R.C.D., "The Maximum Hardness of Strained Surfaces and the Abrasive Wear of Metals and Alloys," Wear 10 (1967), 353-382.

38. Buttery, T.C. and Archard, J.F., "Grinding and Abrasive Wear", Proc. Inst. Mech. Eng., London, 185 (1970 - '71), 537-542.

39. Johnson, K.L., Discussion of Ref. 37, Proc. Inst. Mech. Eng., London, 185 (1970 - '71), D 205.

40. Moore, M.A., "The Relationship Between the Abrasive Wear Resistance, Hardness, and Microstructure of Ferritic Materials", Wear 28 (1974), 59-68.

41. Mutton, P.J. and Watson, J.D., "Some Effects of Microstructure Abrasion Resistance of Metals," Wear 48 (1978), 385- 398.

42. De Gee, A.W.J., "A Note on the Relation Between Friction and Wear," Wear 65 (1981), 397-398.

43. Engel, P.A., "Impact of Materials," Elsevier, Amsterdam 1976.

44. Bitter, J.G., "A Study of Erosion Phenomena," Wear 6 (1963), 169-190.

45. Neilson, J.H. and Gilchrist, A., "Erosion by a Stream of Solid Particles," Wear 11, 1968, 111-122.

46. De Bree, S.E.M., Begelinger, A. and de Gee, A.W.J., "A Study of the Wear Behavior of Materials for Dredge Parts in Water-Sand Mixtures," Proc. Third International Symposium on Dredging Technology, BHRA, Bordeaux, March 1980, 299-314.

47. Arthur, G., Birch, D., Michie, G.M., Moorhouse, P., and Wells, T.C., "Wear Resistant Surfaces - A Guide to Their Production, Properties and Selection," Report International Research and Development Co. Ltd., England, 1979.

48. Powel, C.F., Oxley, J.H., and Blocker, J.M., "Vapor Deposition," John Wiley and Sons, New York, 1966.

49. Caubet, J.J., and Gregory, J.C., "Thermal and Chemico-Thermal Treatments of Non-Ferrous Materials to Reduce Wear," Tribology International 4 (1971), 8-14.

50. Habig, K.H., "Verschleiss und Harte von Werkstoffen" ("Wear and Hardness of Materials"), Carl Hanser Verlag, Munich-Vienna, 1980.

51. Van Nederveen, H.B., Verburgh, M.B., and Houben, J.M., "The Densification of Plasma Sprayed Coatings by Subsequent Hot Isostatic Pressing," paper 51, Proc. 9th International Spraying Conference, The Hague, May 1980.

52. Hintermann, H.E., Gass, H., and Lindstrom, J.N., "Nucleation and Catalysed Growth of TiC, Produced by CVD on Cemented Carbide," Third International Conference on CVD, 1972, 352-360.

53. Hintermann, H.E., Menoud, C., and Maillat, M., "Friction and Wear Characteristics of SiC, TiC and TiN Hard Coatings Rubbing Against Each Other," Lecture Presented at the 8th Meeting of the IRG on Wear of Engineering Materials, Saint-Ouen, May 1981, to be published in Wear.

54. Hartley, M.E.W., "Ion Implantation and Surface Modification in Tribology," Wear 34 (1975), 427-438.

SURFACES

M. J. EDMONDS
BRIGHTON POLYTECHNIC

1. INTRODUCTION

Surfaces are at the heart of the multidisciplinary subject of Tribology. Without an accurate assessment and characterization of surfaces, tribological research cannot progress. In this chapter the author has outlined some of the important aspects concerned with surface technology - ranging from the composition of a surface, specification, terminology, characterization, selection and production, to some of the advances that have taken place in recent years.

The discussion closes with a short account of possible future trends in surface technology and the likelihood of in-process and on-line measurement within the next decade.

1.1 Brief Historical Review

Interest in surfaces and the effects of movement of one surface relative to another, almost certainly stretches back in history to the stone age, Figure 1. Although this may perhaps be disputed by some scholars on the grounds of lack of positive archaeological evidence, it is difficult to visualize primitive man manipulating heavy objects other than by rolling them on logs, thus creating the earliest type of plain bearing. We can also appreciate him rubbing a piece of wood or bone perhaps against a stone, in order to sharpen an edge during the manufacture of tools or weapons. As he progressed, we can imagine him using water or animal fats to lubricate the surfaces to make them easier to operate.

FIGURE 1 "TRIBOLOGY? THAT'S THE STUDY OF TRIBES ISN'T IT!!!

We have to wait many centuries though before any positive documentary evidence is available to support ideas and practices. Indeed, the earliest known detail is that credited to Leonardo da Vinci (1452 - 1519) who, as far as we are aware, was the first person to develop and record ideas of the basic concepts of friction[1].

Interest in surfaces today, although far removed from primitive man, results from the evolution of the rolling bearing, or to use the modern word which embraces all such matters - TRIBOLOGY.

It is really to the study of the rolling bearing that we have to seek for our origins in the subject. Archaeologists believe that the ancient Egyptians and Greeks made considerable use of the principle of rolling bearings. A piece of a bearing believed to have been devised by the Greek Diades (330 BC) is still in existence and was used as part of a battering ram. Fragments of materials found in Lake Nemi - Italy, in 1928, again support the idea that rolling bearings were used more than 2,000 years ago. The period of the Industrial Revolution in Britain and Europe inevitably made a major impact into the field of rolling contact. A notable exhibit in the Norwich Museum of an iron ball thrust bearing, with many design characteristics of modern bearings, made its appearance about 1780.

In 1787 one of the earliest British patents, number 1580, was awarded to Mr. John Garrett of Gloucester for the

arrangement of various types of rolling elements to form bearings. Several interesting developments occurred in the years that followed; and in the middle part of the eighteenth century, with the advent of the bicycle, patents for a variety of rolling applications became abundant.

In 1881, Heinrich Hertz published in Germany, a treatise destined to become a classic, on the deformation of curved elastic bodies in contact[2]. This work provided the first mathematical background to analysis involving surfaces. Apart from a few papers presented to learned societies, there appeared to be very little literature available directly concerned with Surface Technology until about 1930 when a paper by R. E. W. Harrison appeared, which was concerned with a survey of Quality Standards[3]. Shortly after, a paper by Firestone, Durbin, and Abbott was published concerning tests for smoothness of machined finishes[4]. The instrument they used to provide surface profile measurements was an optically amplified stylus system in which a lever arrangement was used to record on photographic paper. A year later in 1933, Abbott and Firestone published a paper which may be regarded as the forerunner for all the subsequent surface analysis work that has taken place[5]. By accepting the fact that no surface is truly flat but made up of an irregular array of hills and valleys, they suggested that if an imaginary mean plane is drawn through the surface profile irregularities then it is possible, by drawing parallel planes on either side of the mean plane, to produce a curve of statistical features of the surface texture.

Figure 2 illustrates a typical surface profile showing the depth versus bearing area relation. The curve is now known as the "Abbott bearing area curve," and demonstrates the fact that when two surfaces are interfaced, the real area of contact supporting the surfaces is only a small percentage of the nominal area. This concept will be discussed in greater detail later, since it is an important aspect of surface technology.

To a large extent, until about 1940, much of the analytical interest in surfaces seems to have remained within the disciplines of physics, chemistry and metallurgy. Engineers at this time were more concerned with methods of production of surfaces, and the design of machines that produced them.

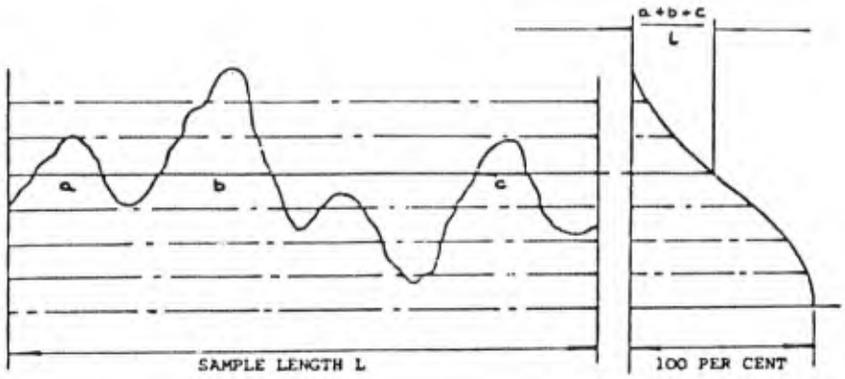


FIGURE 2 THE BEARING AREA CURVE

However, as technology advanced, it became imperative that engineers cross the disciplinary boundaries and gain a greater fundamental understanding of the surfaces and materials they were dealing with.

The period of the second world war produced the next significant advances in the technology. These advances resulted from the fact that aircraft engines had to be developed at an ever increasing rate, be kept running, and respective engine mechanical failures reduced to a minimum. Led by Dr. R. E. Reason, a team of engineers in Leicester set out on a research and development program to develop an instrument capable of quantitative assessment of surface texture, and suitable for use alongside production engineering workshops[6]. Hitherto, all previous assessment attempts were mainly confined to laboratories. Standards of manufacture were generally accomplished by a "go and no go" gauge arrangement. If a component looked good, felt good (tactile testing), and worked, then the process of its manufacture was repeated. Such a system proved extremely difficult, particularly when it became necessary to subcontract work out as the need arose to increase component production rate. Some means had to be established to indicate the type of surface finish specified by the designer, in order to supplement tolerance specifications. This led to the "centre-line-average" technique, or Ra value as it is known today, based on the assessment instruments developed by R. E. Reason and his team.

This brings us up to the early 1950's. One other major advance of special significance connected with surface technology must be noted, and that is the work of Bowden and Tabor, the principal tribophysicists of the day[7]. In 1950 they published their classical work, The Friction and Lubrication of Solids, Part I. Many important topics, vital to the study of Tribology are covered in their text, ranging from area of

contact between solids, friction, and adhesion to the influence of liquid films.

During the past thirty years, many advances have taken place in the study of Surface Technology, and the remainder of this chapter will be concerned with outlining various features that have emerged. To quote Dr. Tom Thomas; "the number of papers published on matters relating to Surface Technology is now doubling each year." This illustrates the wide range of interests associated with the topic, and its place in modern technological systems. The most comprehensive survey on Surface Technology is, to the best of the author's knowledge, that of Thomas and King[8]. This book contains 651 references together with abstracts set in chronological order from 1921 to 1976, and it provides the most useful literature survey on the topic so far in print.

2. COMPOSITION

To the layperson, a surface probably represents simply the outside face of a body, but it is in fact much more complicated than that. A surface must be recognized as an amalgam consisting of several zones which grow organically out of its parent material having both chemical and physical characteristics dependent on the composition of the material itself and the environment in which it is situated.

The base of the surface amalgam consists of a zone of work-hardened material on top of which is a microcrystalline structure. This is termed the BIELBY layer, and for surfaces which have been derived by means of a machining operation, is formed as a result of the melting and flow of the surface molecular layers. The layers form due to hardening and subsequent quenching as they cool against the underlying material. Figure 3 shows a cross-section of this very complex structure. The basic structure is further complicated by the addition of chemical reactions with the atmosphere, contaminants of dust particles, and molecular films absorbed from the environment. In Figure 4, orders of magnitude of depth are given for the various structural zones, plotted on a log scale. very great and the surface will yield plastically. Penetration

From values such as these, it is clear methods of production not only influence the surface finish, but must also have an effect on the subsurface. A very interesting paper by Wallace shows the condition of subsurface material after different finishing processes, and appears to substantiate the figures quoted by Halling and Barwell[10, 9, 20].

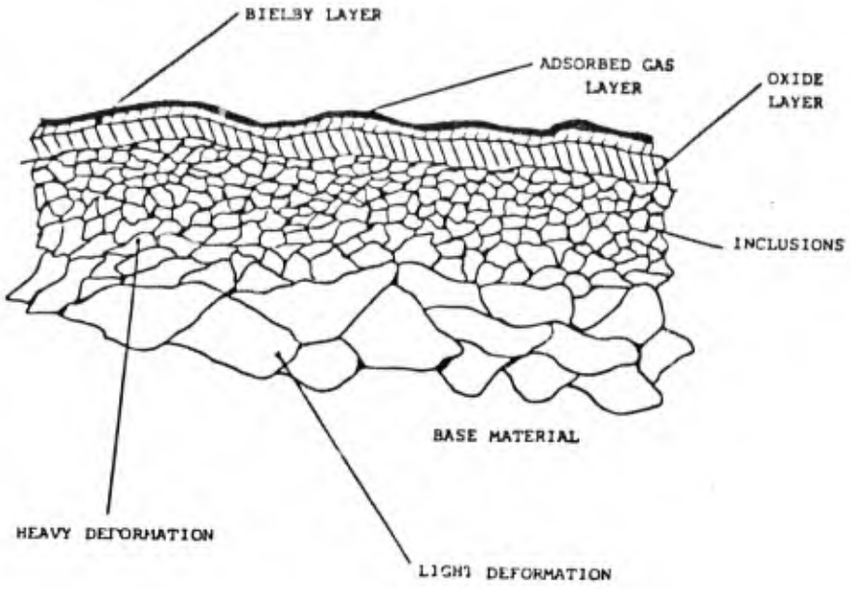


FIGURE 3 SURFACE COMPOSITE

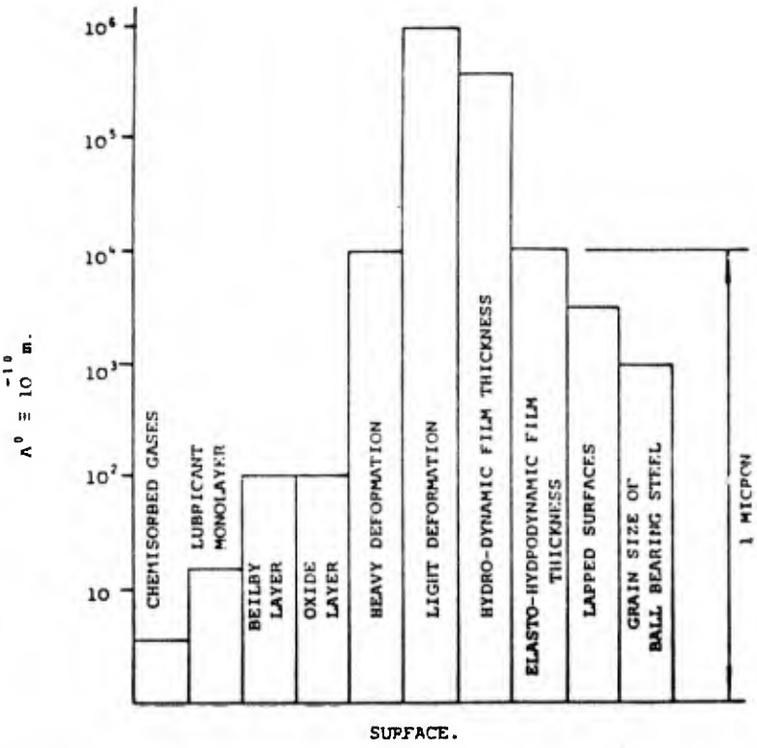


FIGURE 4 TYPICAL ORDERS OF MAGNITUDE OF SURFACE FEATURES

3. HARDNESS

The ability of a surface to resist bulk surface deformation in the plastic mode is referred to as hardness. This is usually measured as a result of penetration of the body by a diamond shape in the form of an inverted pyramid. As the diamond impinges on the surface, stress at the point of entry will be will continue until the diamond has penetrated sufficiently to have interacted with sufficient material to produce an opposing force which will bring it to rest. It is usual to quote, as a hardness figure, the value of the applied load divided by the area of the resulting permanent impression.

Values of hardness have been found to correspond closely with a value of three times the yield stress in compression. Likewise, when two surfaces are brought together, the highest peaks (asperities) on the one surface will come into contact with asperities on the other surface and the intensity of load initially will be very great. It may be expected that the asperities will first yield elastically, but in most metals the elastic component is almost negligible, and the plastic mode appears to predominate [11]. The scale of action will, however, be much smaller than that of the conventional hardness test.

The hardness, expressed as a stress, is normally referred to as the flow stress of the material. If work- hardening is ignored, and the material assumed to be uniform, then it can be shown that the true area of contact will be reached when the applied load equals the product of the true area of contact and the flow stress[12].

4. SURFACE TEXTURE MEASURMENT

Many methods for surface texture measurement both contacting and noncontacting, have been devised over the past thirty years[8]. Generally speaking, non-contacting methods involve optical systems, and contacting methods utilize a fine pointed stylus. A very comprehensive survey, which includes comparisons of all systems developed up to 1942, was published by Schiesinger and more recently by Schneider[13, 17]. The major advantage with optical techniques is the fact that no surface damage is incurred, whereas traversing a stylus over a surface often results in a very fine scratch. Despite this situation, the most widely used instrumentation employs a stylus. This utilization is not only a function of convenience, but principally because more quantitative detail can be obtained. Nevertheless, much qualitative information can be derived from optical systems, particularly electron microscopes. Table 1, below, gives some idea of the resolution obtainable from a variety of techniques.

Technique	Resolution (μm)	
	Lateral	Vertical
Profilometric Instruments	1.3-2.5	0.005-0.25
Electron Microscope	0.005	0.0025
Optical Microscope	0.25-0.35	0.18-0.35
Interference Microscope	0.25	0.25
Oblique Sectioning	0.25	0.025

TABLE 1 SURFACE RESOLUTION

4.1 Stylus Systems

The early stylus instruments used a phonograph needle; however, modern systems enploy a truncated diamond pyramid of tip radius approximately 2.5 μm[15, 16]. The angle between the faces is 112°, and the dimensions of the rectangular flat about 3 μm x 8 μm. The shorter edge is arranged to be parallel with the direction of motion. The applied load is of the order of 0.7 mN. Variations of styli are available for special applications, down to a tip radius of 0.1 μm.

Vertical movement of the stylus is recorded electrically via an L.V.D.T. (linear variable differential transformer), with the displacement amplified. The electrical signal from a displacement-sensitive transducer can provide three kinds of information: (1) an analogue record, i.e. a profile graph on a chart recorder; (2) a digital record of closely spaced ordinates of the profile graph, which can be recorded on punched or magnetic tape for subsequent processing in a computer; or (3) meter indications of numerical values which are descriptive of the profile.

It should be pointed out that the graphical description of a profile should not be taken too literally, because magnification factors on the instruments yield up to 10,000 times vertical displacement relative to the horizontal displacement. What appears to be a close spike-like profile in the compressed state would yield, if plotted on a 1:1 ratio, a gentle undulating surface, characteristic of moorland rather than say a profile view of the Alps. By virtue of this, penetration into the valleys by the stylus is not such a

critical factor as might be assumed. Although, if measurements of abrasive surfaces such as the assessment of grit in grinding wheels are considered, then the problem could be more acute. If such an assessment as this is desired, it would seem more practical to approach the problem a different way utilizing the electron microscope.

4.2 Electron Microscope

As mentioned earlier, the S.E.M. (scanning electron microscope) does offer an alternative means of surface geometry assessment. Magnifications up to 5,000 x are easy to relate to ordinary hand-held magnifiers. If it is desired to simply look at a surface at a reasonable magnification, using 'Y' modulation can be useful. This means operating the S.E.M. such that an isometric picture is built-up, by scanning the surface in a similar fashion to tracing a profile using a stylus system. Figure 5 illustrates the view of a medium ground surface using this technique, together with its magnified S.E.M. image, the angle being set at 45 °.

5. SPECIFICATION

No surface exists in nature, or man-made, which can be described as being truly planer. The reason for this is that all matter is made up of atoms, and for such a state to exist the center of all atoms, or ions in the surface layer, would have to lie in the same plane. No element appears to be able to conform to this requirement. The nearest element which might qualify is very carefully cleaved mica. In the vast majority of cases, the surface of a solid has irregularities, although in some cases only atomic in scale, but frequently very much larger. For machined surfaces, we can take it for granted that whatever method of preparation has been applied, the surface will always have irregularities of some form or another. The measurement and adequate assessment of surfaces is of vital importance to all engineering and science systems. Friction, wear, contact deformation, reflection, heat transfer, electrical and thermal conductance, tolerances and fits are all influenced by and depend on a comprehensive characterization of the micro- and macro-geometrical texture. Many papers have now been published on each of the above mentioned topic areas; subjects ranging from the assessment of the surface of ships hulls to the surface of lip seals[18, 19]. As already mentioned, Thomas and Kings book offers a wide range of references together with an excellent literature review; and readers are advised to use this as their starting point for any particular area of interest.

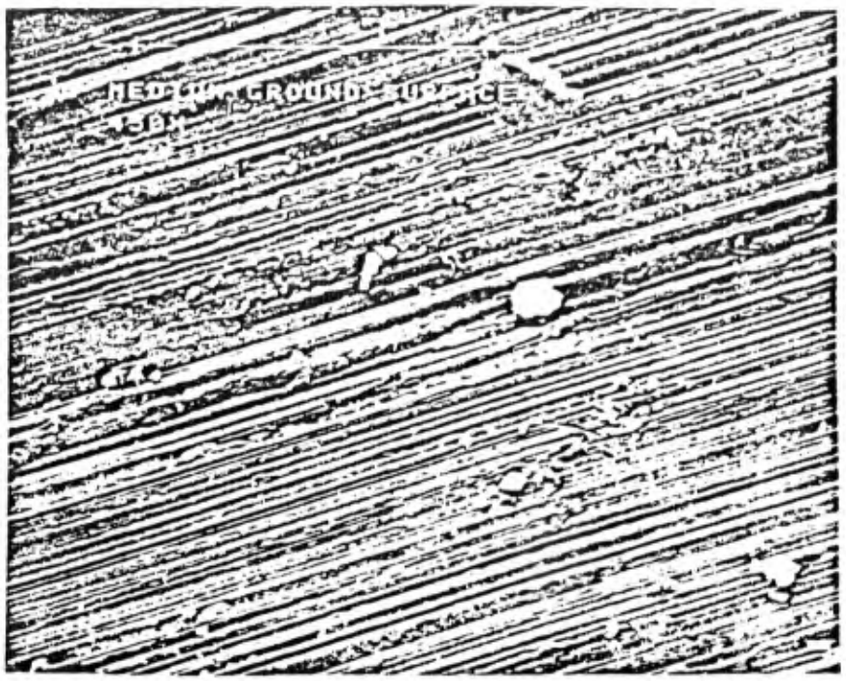


FIGURE 5(A) S.E.M. SCAN OF A MEDIUM GROUND SURFACE

FIGURE 5(B) S.E.M. SCAN OF 'Y' MODULATION

5.1 Terminology

The geometric texture of machined surfaces is controlled by characteristics of the finishing process by which the surface is produced. On a microscopic scale, irregularities are evident, and roughness is formed by fluctuations in the surface of short wavelengths (termed microroughness). The form is characterized by hills (known as asperities) and valleys of varying amplitudes and spacing, see Figure 6. Orders of magnitude of measurement are either microinches or micrometers (micron), (1 μm = 1 x 10^{-6}m). In addition, on many surfaces a larger wavelength of roughness can also be observed called waviness, which is referred to as macroroughness. Normally this is measured in mm. Other features which also exist in a surface in the form of undulations, and exhibit much larger wave-lengths, are errors of form. Again, these may be measured in mm. Such features as these are generally caused by vibrations of the machine, work piece, or the cutting tool.

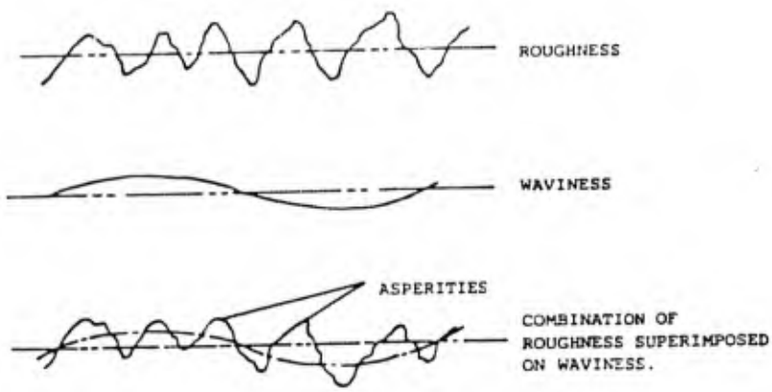


FIGURE 6 COMPONENTS OF A SURFACE

The distribution of the asperities over a surface can be either directional or homogeneous in all directions, depending on the method of preparation. Surfaces produced by shaping, milling, or planning for example have a definite pattern or lay. But those produced by lapping, electropolishing, etc., show an isotropic distribution in all directions.

5.2 Symbols Used in Surface Analysis (ANSI B461 - 1978)

5.2.1 Center-Line Average

The universally recognized parameter of roughness has the symbol Ra. Formerly, this was called the center-line average, C.L.A. (British), and the Arithmetic Average, A.A., (America). It is the arithmetic mean of the departure, "y", of the profile

from the theoretical mean line, and is normallly determined as the mean result of several consecutive sampling lengths, "L", see Figure 7.

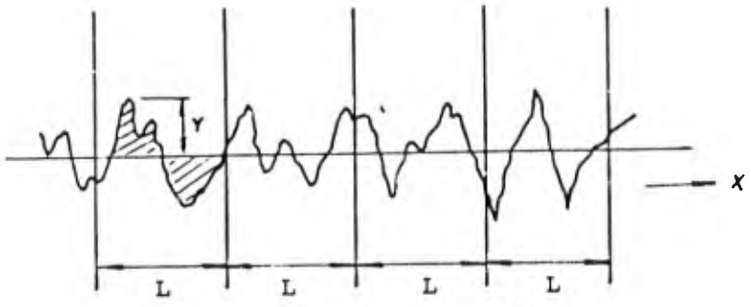


FIGURE 7 CENTER-LINE AVERAGE

5.2.2 Root Mean Square

The Root Mean Square value, R.M.S., is no longer used in the context of surface texture measurement, but is used in statistical analysis. It was formerly given the symbol Rq

where $Rq = \sqrt{\frac{1}{L} S_o^{L} Y^2 (x)dx}$

In statistical analysis the symbol used is "σ".

5.2.3 Ten Point Height

Rz is the ten point height, and is the average distance between the five highest peaks and the five deepest valleys within the sampling length and measured perpendicular to it, see Figure 8.

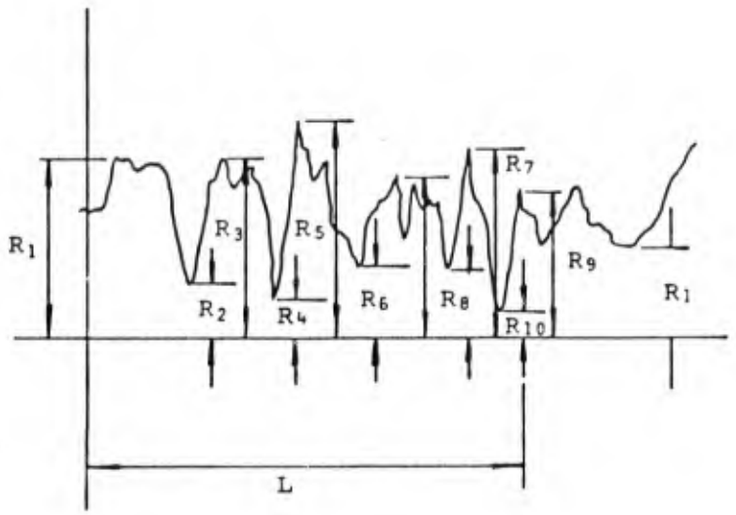


FIGURE 8 TEN-POINT HEIGHT

5.2.4 R_t, R_{max} and R_{tm}

R_t is the maximum peak to valley height within the assessment length. R_{max} is the maximum peak to valley height within a sampling length L. Since the value can be affected by a spurious scratch or a particle of dust on the surface, it is more usual to use the average, "R_{tm}", of five consecutive sampling lengths, see Figure 9.

$$\text{where } R_{tm} = \frac{R_{max_1} + R_{max_2} + R_{max_3} + R_{max_4} + R_{max_5}}{5}$$

$$R_{tm} = \frac{1}{5} \sum_{i=1}^{i=5} R_{max_1}$$

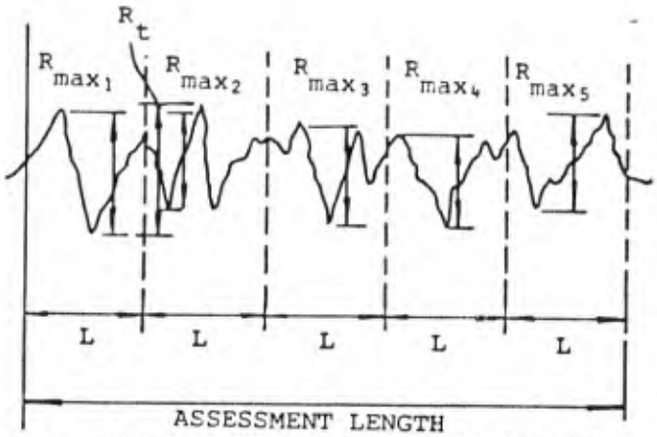


FIGURE 9 PEAK TO VALLEY HEIGHT

5.2.5 R_p and R_{pm}

R_p is the maximum profile height from the mean line within the sampling length. R_{pm} is the mean value of R_p determined over 5 sampling lengths, see Figure 10.

$$\text{where } R_{pm} = \frac{R_{p_1} + R_{p_2} + R_{p_3} + R_{p_4} + R_{p_5}}{5} = \frac{1}{5} \sum_{i=1}^{i=5} R_p$$

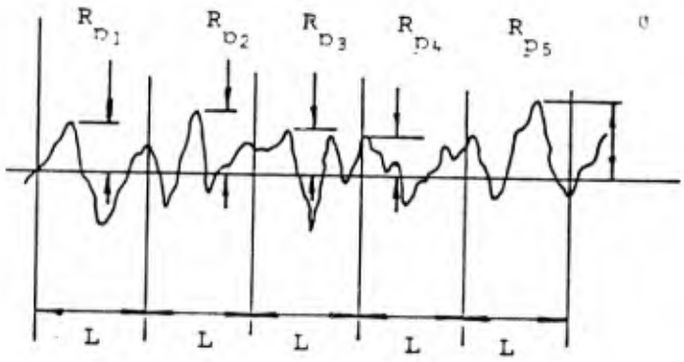


FIGURE 10 MAXIMUM PROFILE HEIGHT

5.2.6 P_c

Peak count is the number of peak/valley pairs per inch projecting through a band width, b, see Figure 11.

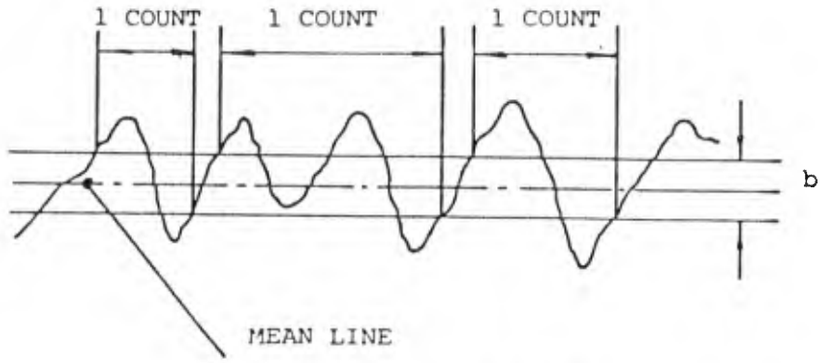


FIGURE 11 PEAK COUNT

5.2.7 Average Wavelength

Because no single parameter can adequately represent the complex patterns found on a surface, other parameters were introduced, but all are derived principally from amplitude, and as such are relatively unaffected by the spacing of irregularities. Universal acceptance of these symbols and their meaning has not, as far as the author is aware, been totally agreed upon. In Europe, the symbol, λa is used which is not derived from amplitude but from spacing. This is termed the average wavelength (λa) and is now used in conjunction with Ra. Thus, the average roughness and the openness or closeness of the surface texture can be specified.

Some typical values of Ra and λa are shown in Table 2.

Process	Ra (µm)	λa (µm)
Turning	1.7	100
Fine turning	0.3	37
Grinding (flat)	0.95	77
Circumferential Grinding	0.55	33
Single tooth milling	2.2	300
Multiple tooth milling	5.1	1300
Fine cut planning	5.1	440

TABLE 2 Ra VERSUS λa VALUES

From the table we can see how two surfaces prepared by different machine processes can produce a similar value of Ra, and yet have a value of λa grossly different. Clearly it is necessary to specify both parameters and not simply rely on just the Ra value alone.

5.3 Cut-Off

The question of "cut-off" is another feature which should be added to the specification of a surface. Since the surface texture can include roughness of different crest spacings which may be superimposed on one another, it is necessary to choose a sampling length, L, sufficiently long enough to enable the various crest spacings to be accommmodated. Electrical integrating instruments have now been devised such that the meter cut-off values are made equal to the sampling lengths. When the meter is switched to the desired meter cut-off, the pick-up used to follow the profile of a surface gives the average results from several consecutive sampling lengths.

Table 3 shows a selection of suitable cut-offs which are normally used for a variety of common manufacturing processes.

Manufacturing Process	Typical roughness height (Ra) μm	Range of Peak Spacings (mm)				
		0.08	0.25	0.8	2.5	8.0
Super Finishing	0.05 - 0.2	•	•	•		
Lapping	0.05 - 0.4	•	•	•		
Honing	0.1 - 0.8		•	•		
Grinding	0.1 - 1.6		•	•	•	
Diamond turning	0.1 - 0.4		•	•		
Turning	0.4 - 6.3			•	•	
Boring	0.4 - 6.3			•	•	•
Broaching	0.8 - 3.2			•	•	
Milling	0.8 - 6.3			•	•	•
Shaping	1.6 - 12.5			•	•	•

TABLE 3 SUITABLE CUT-OFF VALUES

Various limiting features are apparent when selecting the appropriate cut-off value such as the finite dimensions of the stylus tip, the inherent noise level of the instrument, and the

type of skid nose piece which may be added to the instrument. It is therefore advisable to seek guidance from the appropriate instrument manuals before proceeding to take surface measurements.

5.4 Surface Designation

The cost of machining to achieve set tolerances is largely dependent on the skill of the operator, the quality and type of machine tool, and the material. If we add to this just the basic surface texture requirement, Ra, then clearly the cost of production must increase. A graph showing relative costs plotted against surface texture is shown in Figure 12. The cost of failure of a component which can in turn affect other components, and which can lead to catastrophic failure, must be viewed relative to the cost of producing the component in the first place. It is therefore necessary to first decide the need, then to assess the value, and finally to communicate very specifically the necessary information to all those people involved in the manufacture of any particular component or system. In order to do this, the design engineer must suppply very specific information relating to the surface finish desired. Ideally, a detailed engineering drawing should contain information on the maximum and minimum Ra value, the maximum waviness height, the appropriate cut-off, the lay, and maximum roughness spacing.

At first sight, the above detail may appear to present an appreciable burden to both the detail draughtsman and the machine operator. However, the problem can be overcome reasonably well using the existing technique of a tick, "√", to indicate where machining has to be performed. Based on ANSI Y 1436 - 1978, Reference 21, which is almost in agreement with I.S.O. 1320, the following format, Figure 13, has been agreed upon.

In practice, not all areas on a machine component need to be treated with such rigor. Only those parts which require precision machining and close tolerances such as slideways, shaft fittings, bearings, etc., or areas where contacting surface are concerned, need the full specification. Other than this, the Ra value is generally sufficient. But how does the machine operator interpret the values? By far and away the best method is simply experience, together with visual examples and the old and surprisingly sensitive fingernail test. With the newer instruments coming onto the market, sample specimens can be taken and assessed fairly rapidly. Unfortunately, this often involves removing the component from the machine tool, and it is particularly time consuming and difficult to reset, should the component not meet the requirements.

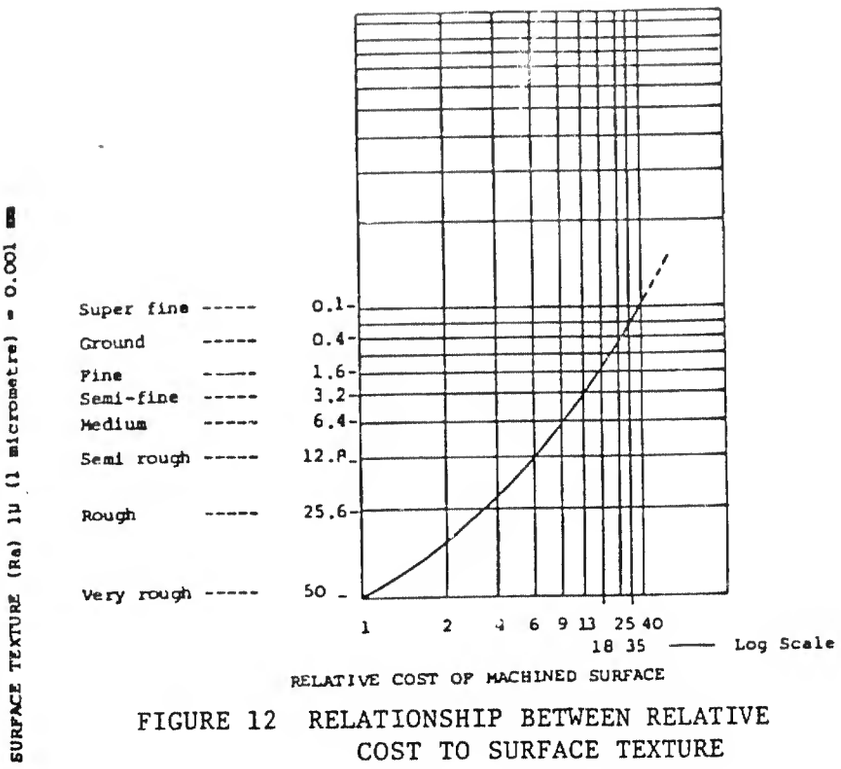


FIGURE 12 RELATIONSHIP BETWEEN RELATIVE COST TO SURFACE TEXTURE

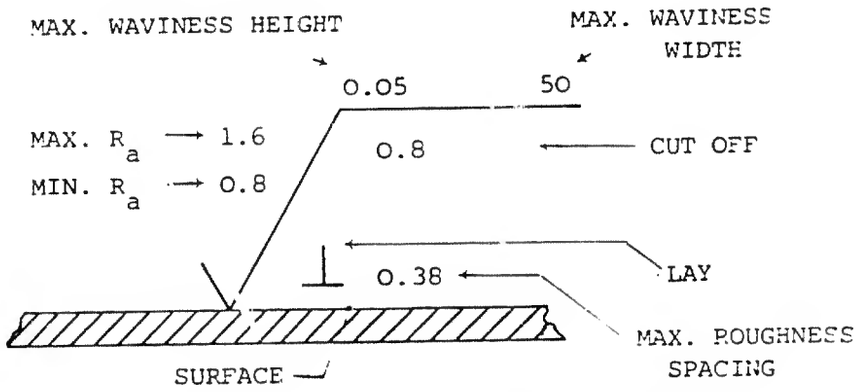


FIGURE 13 SURFACE TEXTURE SYMBOLS

ALL VALUES ARE IN mm EXCEPT RA = µm
✓SURFACE MAY BE PRODUCED BY ANY METHOD
✓MATERIAL REMOVAL BY MACHINING IS REQUIRED:
(I.E., MATERIAL MUST BE PROVIDED FOR THAT PURPOSE)
.025m ✓MATERIAL REMOVAL ALLOWANCE (mm)
✓MATERIAL REMOVAL PROHIBITED

5.5 Replication

In an effort to overcome the problem mentioned above, considerable work has now been done on surface replication[22, 23]. This is a technique which permits a plastic replica to be made of the area of interest which can be assessed away from the machine. Depending on the type of surface, very good correlations can be obtained between the original surface and the replica. In some cases a 95% accuracy has been claimed. The major problem is the greater the adhesion of the replica to the surface the more difficult it is to remove it. However, the better the adhesion, the more faithful is the reproduction. A compromise, as in most cases in engineering, has so far to be accepted.

6. CHARACTERIZATION

6.1 The Statistical Parameters

The principal production parameters for describing machine surfaces have been outlined in the preceeding subsections. In 1968 a new technique was introduced by Peklinik by applying spectral analysis to the study of surface topography[24]. This technique has probably produced the most significant advance in the subject in the past decade. Based on the mathematical treatment of natural phenonmena, namely the motion of ocean waves, mathematician Lonquet-Higgins produced a mathematical treatise which laid the foundations of what today has become one of the most powerful tools in the study of signal processing[25]. In recognizing that machined surfaces are made up of random and periodic components analogous to noise and vibration signals, surface scientists have developed and utilized the technique of spectral analysis with a considerable degree of success in recent years[26, 27, 28, 29, 30, 31]. The most significant advance, and the one which has created the most valuable contribution, is the Auto-correlation function, together with its Fourier-transfrom, the power spectral density function. These mathematical functions aim to separate the random and periodic components that occur in many surfaces generated by machining processes. Separation of these components is possible by analyzing the two- dimensional profile of a surface.

6.2 The Auto-Correlation Function

The analysis required to obtain the Auto-correlation function of a surface consists of making comparisons between a given length of profile and the same profile displaced in the length domain by increasing amounts. At each stage, ordinates in the overlapping portions of each profile are multiplied and

their sum is reduced to an average value. Figure 14 illustrates the situation.

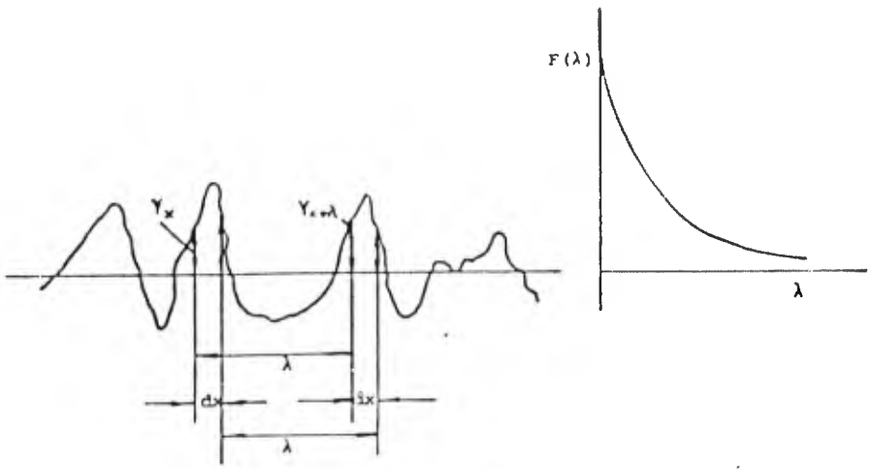


FIGURE 14 THE AUTO-CORRELATION FUNCTION

Consider the points Yx and Yx+λ on the profile separated by a distance λ. The product of their heights is Yx, x+λ. If the length λ is moved a distance dx along the profile then its ends define a new set of points Yx + dx, Yx + dx + λ, which will yield another height product. By repeating this procedure for the whole length of the profile, the following parameter can be defined:

$$\sigma^2_{YY},\lambda = \frac{1}{L - \lambda} \int^{L-\lambda} Y_x,\ Y_x + \lambda \qquad (1)$$

where σ is the R.M.S. value.

If λ is small in comparison with the average dimensions of the peaks then the products Yx, Yx + λ will, in the main, be mostly finite and positive; thus, Equation (1) will be finite and positive. If λ is large in comparison with the average dimensions of the peaks, the products are also likely to be of either sign, and their sum will therefore be small in magnitude. By plotting σ^2_{yy}, λ against λ, the function decays from a finite positive value to nearly zero, and the rate of decay can then be used as a measure of the spacing of the surface texture. This function is called Auto-covariance, which for zero delay is the variance, i.e. σ^2_{yy}, o = σ^2.

If both sides of the equation are normalized by dividing by σ^2, the Auto-correlation coefficient is thus obtained from:

$$F\lambda = \frac{1}{\sigma^2 (L-\lambda)} \int_{o}^{L-\lambda} Y_x, Y_x + \lambda^{dx} \qquad (2)$$

Since this function has the properties associated with the correlation coefficient of discrete statistics, i.e. its numerical value can vary between ± 1, the variation with delay length λ is the Auto-correlation function.

An Auto-correlation function of 1.0 indicates complete agreement, whilst - 1.0 indicates complete disagreement. A value of zero shows that no correlation exists.

Figure 15 illustrates two typical surface profiles that might be produced by differenct processes, together with their corresponding Auto-correlation functions. The shape of Figure 15(A) is somewhat periodic in form and typical of a surface that might be produced by end-milling or by shaping, operating across the lay. The plot of the Auto-correlation function follows a similar periodic form. The general decay indicates a decrease of correlation as the sampling interval, ℓ, increases, and the oscillation component indicates the inherent periodicity of the profile. Figure 15(C), on the other hand, is random and typical of a ground surface. The rate of decay of the Auto-correlation function, Figure 15(D), is fairly rapid, and some periodicity can also be seen. Again, this illustrates that the profile has been obtained operating across the lay. The initial decay is characteristic of the surface component, and the mean distance between the points where the function becomes zero is an indication of the half-wavelength of the surface waviness. At this point, the surface ordinates in the construction of the curve are statistically independent, and the function may be approximated by:

$$F(\lambda) = e^{-x/\alpha 1} \qquad (3)$$

where $\alpha 1$ is the correlation distance.

Peklenik suggested that most engineering surfaces can be categorized by one of five Auto-correlation functions, which range from linear to sinusoidal functions to the preceeding exponential function, Equation (3)[24].

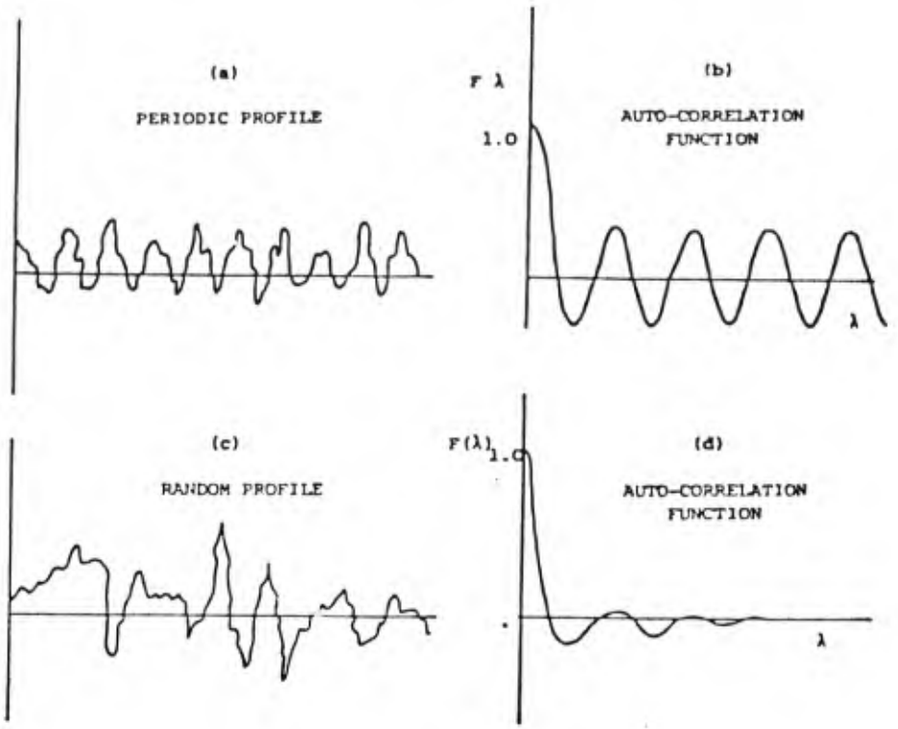


FIGURE 15 TYPICAL SURFACE PROFILES AND THEIR RESULTING AUTO-CORRELATION FUNCTIONS

6.3 The Power Spectral Density Function

As previously mentioned, the Fourier transform of the Auto-correlation function is termed the power spectral density of the discrete wavelengths contained in the profile variation, that is to say, it reveals the dominant frequencies in the profile[32, 33]. Large wavelength periodic components appear as spikes in the P.S.D. curve and, generally speaking, for a random signal with a wide spectral density distribution, the Auto-correlation function exhibits a high rate of decay. The P.S.D., $P(\omega)$, and Auto-correlation function are related by:

$$P(\omega) = 4 \int_0^{L-\lambda} \sigma^2_{YY,\ \lambda} \cos 2\pi \lambda f\ \lambda d \ \ldots \qquad (4)$$

The curve of $P(\omega)$, Figure 16, is the power spectrum due to the fact that its instantaneous value at a given frequency is proportional to the mean square of the signal at a given frequency, and hence its power. The area under the power spectrum curve is the total power present at all the

frequencies, i.e. the mean square value of the whole signal, or the square of the r.m.s. roughness.

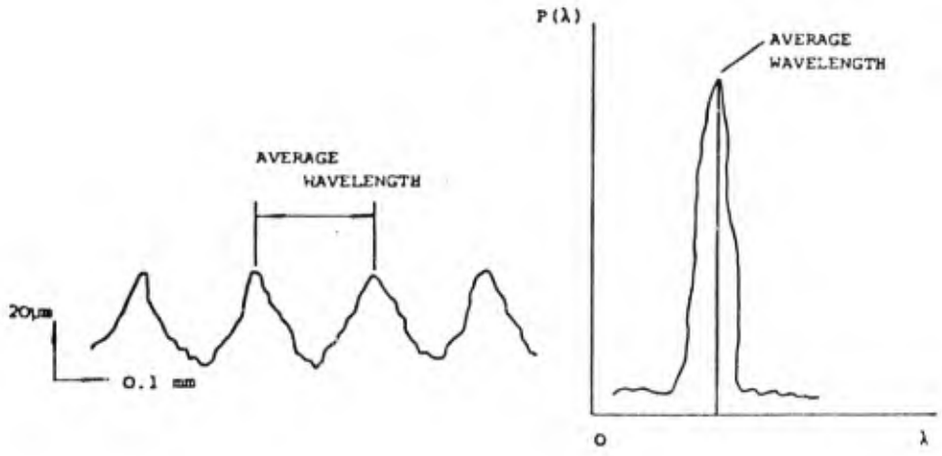


FIGURE 16 THE POWER SPECTRAL DENSITY FUNCTION

Any surface with a periodic component will have well defined peaks in its surface profile, which will also show up in the power spectrum curve. Depending on the number of ordinates chosen for a given length of profile, the P.S.D. curve will reveal not only the predominant frequency but also, in order of magnitude, any underlying frequencies or harmonics. This is particularly so for machined surfaces typically end-milled, turned, etc. However, for surfaces which have been produced by grit-blasting, where an appreciable periodic component is not apparent, i.e. when the surface is random rather than deterministic and the height distribution is Gaussian, the frequencies may be suppressed.

Generally, for most surface specifications, the Auto-correlation length can only be determined from the Auto-correlation function. Also, any periodic variation in the profile appears as a similar periodic variation in the curve of the auto-correlation function versus ordinate separation. In the P.S.D. function, any periodic function in the underlying surface appears as discontinuities. Nayak and Nara proposed that waviness and roughness should be discussed in terms of bandwidths of the spectrum, rather than of fixed wave lengths[34,35]. They proposed that the previous deterministic view of small sinusoids, namely "roughness" imposed on large sinusoids of fixed wavelengths, i.e. waviness or errors of form, should be

replaced by a picture of overlapping bands of wave lengths from a continuous spectrum. From this it follows that the size, shape, and height of the surface texture also forms a continuous distribution depending on the spectrum. Thomas and Sayles adopted this approach and proposed a theory to predict the total roughness of a specimen of any size from a standard profilometer measurement[36]. Their theory suggests that any two isotropic surfaces produced by ordinary engineering processes, no matter by what means, will have identical power spectra if their measured roughness is the same. This is because, as they suggest, no profilometer analysis would be able to distinguish between a surface roduced by blasting with very small beads and one lapped with coarse particles of abrasive materials to the same finish. To a certain extent, this is in agreement with the statistical theory of surface generation to Williamson[36].

A typical example of an engineering surface is shown in Figure 17.

6.4 The Height Distribution of Surface Texture

The Abbott and Firestone bearing area curve has already been mentioned. In statistical terms, this is the cumulative distribution of the ordinate distribution curve, and can be written as:

$$F(z) = \int_{-\infty}^{\infty} \psi \ (z) \ dz. \qquad (5)$$

Where z refers to the heights of the ordinates in the profile measured from the center-line, and $\psi(z)$ is the probability density function of the distribution of these heights. In practice, the derivation is achieved by taking measurements of z_1, z_2, etc. at some discreet interval and summing the number of ordinates at any given height level. Computer programs are now available for performing this task, and a typical curve for a grit blasted surface is shown in Figure 18. The straight portion of the graph indicates a Gaussian distribution, and is characteristic of a surface whose texture is uniformly distributed. For a more detailed discussion of this, readers are referred to Williamson's paper, Reference 28, or Halling, Reference 9.

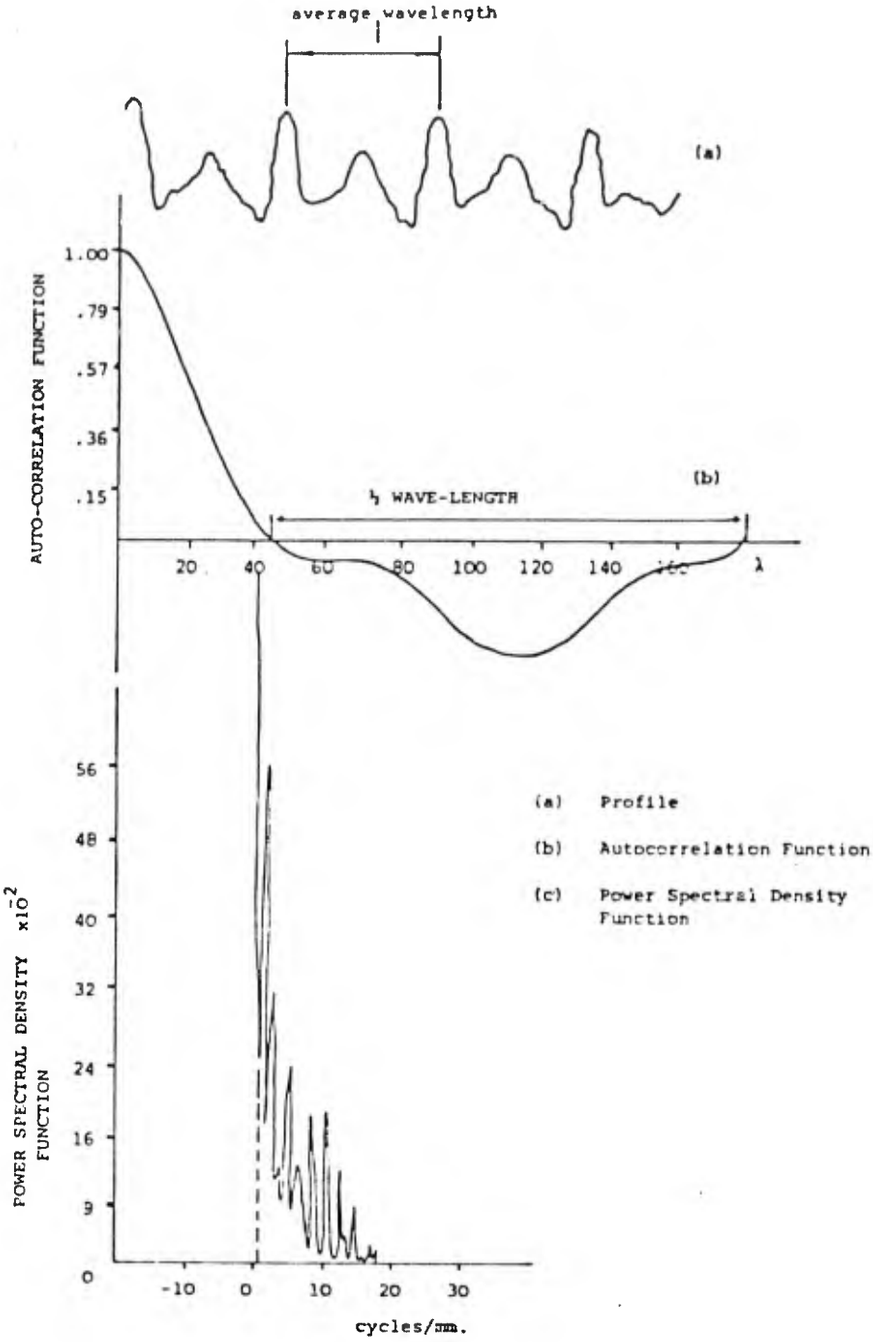


FIGURE 17 TYPICAL CHARACTERISTIC CURVES FOR AN END-MILLED SURFACE

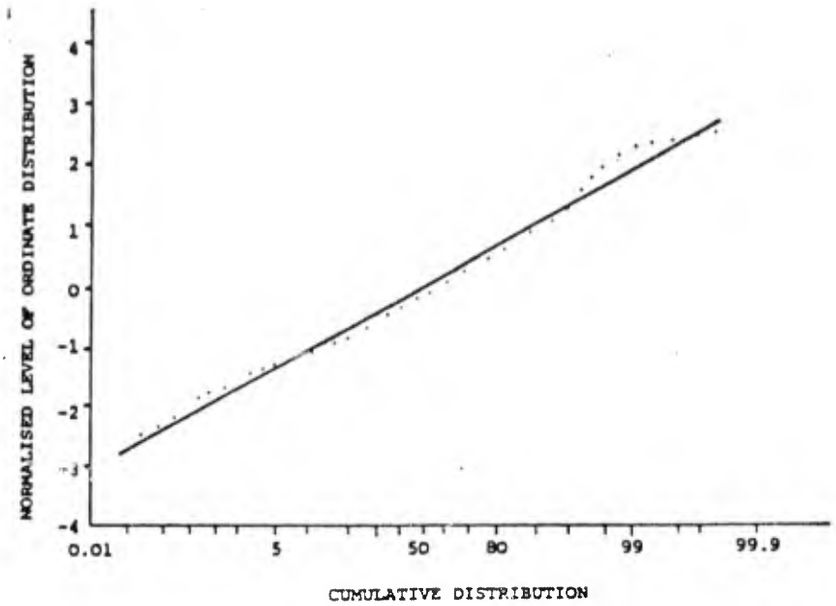


FIGURE 18 TYPICAL ORDINATE DISTRIBUTION CURVE FOR A GRIT-BLASTED SURFACE

Although several common surface preparations produce near-Gaussian distributions, many do not. It is therefore necessary to define some other statistical parameters for measuring the various forms.

6.5 Moments, Skewness and Kurtoris

6.5.1 Moments

The n^{th} moment of the distribution curve $\psi(z)$, about the mean is defined by:

$$M_n = \int_{-\infty}^{\infty} z^n \psi(z)\, dz. \qquad (6)$$

Thus the R.M.S. value $\equiv \sigma \equiv \left[\int_{-\infty}^{\infty} z^2 \psi(z)\, dz\right]^{1/2}$

$$= \left[\text{2nd moment of } \psi(z)\right]^{1/2}$$

also the C.L.A. $\equiv 2 \int_{0}^{\infty} z\psi(z)\, dz.$

$$= 2 \times \{\text{the } 1^{st} \text{moment of } 1/2\ \psi(z)\}.$$

Clearly the 1st moment of the whole of $\Psi(z)$ about the mean is zero, from which the center line of the distribution can be obtained.

For a Gaussian distribution curve the nth moment is given by:

$$M_n = \frac{1}{\sigma (\pi)^{1/2}} \int_{\infty}^{\infty} z^n e^{-z^2/2\sigma} dz \qquad (6)$$

If n is odd the term vanishes, which it would for a symmetrical curve, and if n is even then

$$M_n = \frac{n!}{2^{n/2} (n/2)!} \sigma^n \qquad (7)$$

Thus the 2nd moment becomes just σ^2 which is the variance.

6.5.2 Skewness

Skewness is a measure of the departure of a distribution curve from symmetry and is defined by:

$$S = \frac{\int_{-\infty}^{\infty} z^3 (\Psi) (z) dz}{\sigma^3} \qquad (8)$$

$$= \frac{3^{rd} \text{ moment of } \Psi(z)}{\sigma^3}$$

If the curve is a Gaussian distribution symmetrical about the axis then the skewness would be zero.

If the curve is either positive or negative then the shape of the skew would follow a shape similar to that shown in Figure 19.

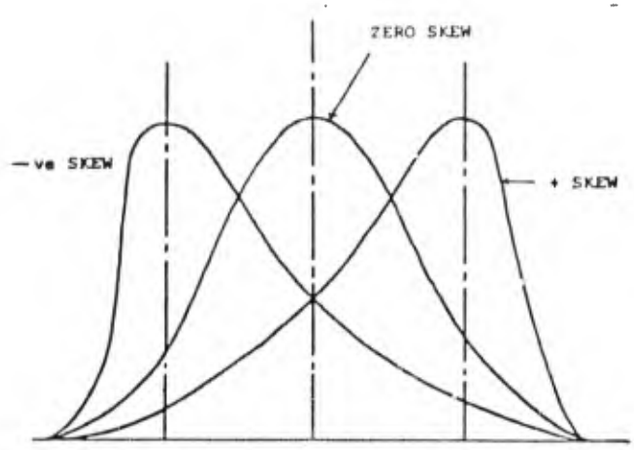


FIGURE 19 SKEWNESS

6.5.3 Kurtosis

This indicates the spikiness in the surface texture and is measured by the hump on the distribution curve. It is defined by:

$$k = \int_{-\infty}^{\infty} \frac{z^4 \psi(z)\, dz}{\sigma^4}$$

$$= \frac{4^{th} \text{ moment of } \psi(z)}{\sigma^{-4}} \qquad (9)$$

For a Gaussian distribution k = 3, obtained by substitution in Equation (7), i.e.

$$k = \frac{1}{\sigma^4} \frac{4 \times 3 \times 2}{2 \times 2 \times 2} \sigma^4 = 3$$

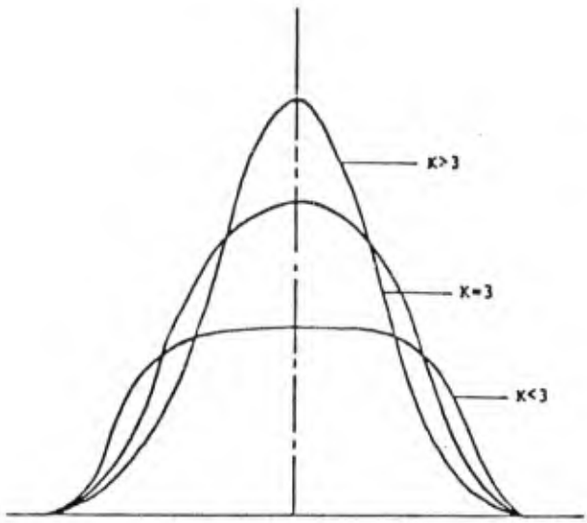


FIGURE 20 KURTOSIS

If k is less than 3 the curve is called platykurtic, and if k is greater than 3 the curve is called lepto-kurtic.

7. TOPOGRAPHY, AND REAL AREA OF CONTACT

So far the discussion has centered around the profile of a surface and the analysis concerned with two- dimensional surface topography. But topography, as the name implies, is three-dimensional; it has length, breadth, and depth.

The first successful attempt at describing a surface three dimensionally by means of a contacting method, namely a stylus, and utilizing multiple parallel tracing can be attributed to Williamson in 1968[28]. Since then, other workers have reported similar systems depending on the application to which their interests are directed, i.e. static, rolling, or sliding contact[38, 39, 40, 41, 42]. The major problems and limitations imposed on these systems were in creating an arbitrary flat datum, and the size of an area compatible with the amount of information that could be analyzed due to the vast data storage and data handling problems that such systems incur. An additional problem was the time element to physically record the vast amount of information. Advances in computer technology during recent years have greatly assisted in this problem, and future workers will be able to consider analyzing surface areas larger than have been possible hitherto.

Figure 21 shows a 2.5 mm^2 isometric view of a rough turned surface of roughness Ra = 20.64 μm. The horizontal axis is a 100 x 100 data point array, and the vertical axis magnified to 255 data points, with the peak to valley height being 54 μm.

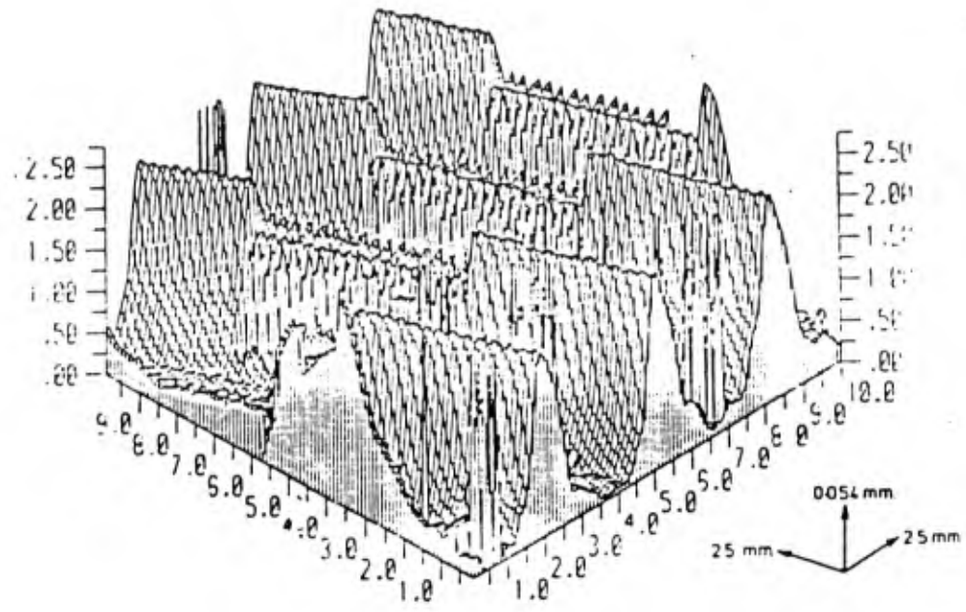


FIGURE 21 ISOMETRIC VIEW OF A ROUGH TURNED SURFACE

The machining marks are clearly visible indicating the undulating nature of the surface topography. Throughout the central portion, the asperities appear suppressed across the ridges, which suggests that possibly some flaw may have existed in the machine tool, or alternatively that over that portion of the surface the area had been damaged. Subsequent optical examination of the actual surface did in fact reveal a damaged portion - a small indentaion about 1 mm wide and about 10 μm deep. Other areas of the surface appeared to be satisfactory, which indicated that the cutting tool was not the cause. The important feature revealed by this form of analysis is that the technique is a useful means of detecting and quantifying detail of small orders of magnitude. "Wear" is a typical example where

the application could be useful, particularly if the instrumentation has relocation facilities[37, 38].

Figure 22 shows a contour map of the surface with eight different contour levels, equally divided between 0 and 255 data points where 20.65 μm = 255 data points. Again, the undulating nature of the surface is clearly revealed. The darker areas, where it appears that contours coalesce, can also be seen as sharp "spikes" in the isometric view. The explanation for this is that, due to the very rough nature of the surface, the stylus response could not cope with the speed of travel as it scanned the surface and "bounced" in this area. Subsequent tests on other types of machine surfaces did not reveal spikes of this nature, which suggests that speed of traverse is a necessary consideration, and particularly so if analysis is undertaken on very rough surfaces.

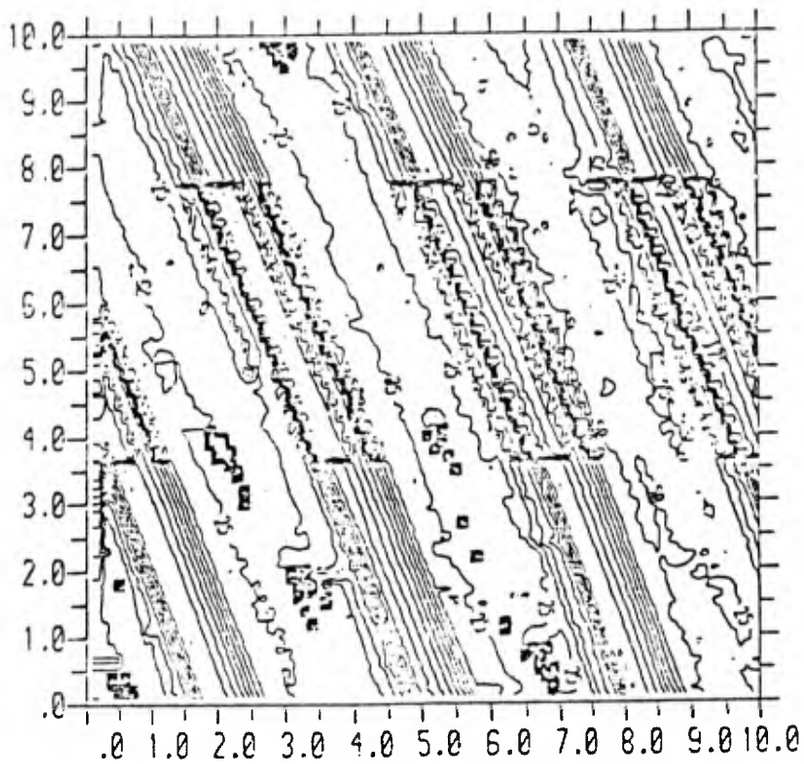


FIGURE 22 CONTOUR PLOT OF A ROUGH TURNED SURFACE

7.1 Real Area of Contact

When one solid is placed in contact with another solid the dominant asperities, which may be situated on the peaks of the waves of both solids initially coming into contact, suffer

deformation and carry on deforming until the load is finally supported by areas of discrete contacts. Any increase in loading, such as bolting or rivetting, causes further deformation of asperity contacts, and also may extend to the waviness in the surface. Most of the asperities deform plastically, but elastic deformation of asperities can occur at the edges of the microscopic contact areas where the stresses are relatively low. The induced stresses in the surface waves however, are always much smaller than those in the individual asperities, and generally speaking wave deformation is normally regarded as being elastic. Several other factors influence the pattern of contact, namely the magnitude and time dependence of the applied force, the stiffness and hardness of the superficial material, and the likelihood of contaminant layers or interfacial fluids.

In many engineering applications it is necessary to pass a heat flux, or transfer electrical energy from one solid to another, and it is the true area of contact that usually dictates the thermal and electrical resistance which occurs at the interface.

7.2 Deformation of Contacting Surfaces Under Load

Greenwood and Williamson proposed in their classic paper, "Contact of Nominally Flat Surfaces," that most surfaces do not undergo a transformation from elastic to plastic deformation as the load increases, but deform plastically even under the lightest loads[11]. By using a statistical technique to characterize a surface, they showed that engineering surfaces exhibit, to a very close approximation, a Gaussian distribution of surface heights. This led them to the belief that not only were the laws of friction dependent upon the topography of the surface, but more importantly they were able to postulate, for a Gaussian surface, a parameter which would indicate the mode of deformation.

This parameter is known as the plasticity index, β

$$\text{where } \beta = \frac{E'}{M}\left|\frac{\sigma'}{\rho_s}\right|^{1/2} \qquad (10)$$

$$\text{and } \frac{1}{E'} = \frac{1-\nu^2}{E_1} + \frac{1-\nu^2}{E_2} \qquad (11)$$

$$\text{and } \frac{1}{\beta} = \frac{1}{\rho_1} + \frac{1}{\rho_2} \qquad (12)$$

$$\text{and } \sigma' = (\sigma_1^2 + \sigma_2^2)^{1/2} \qquad (13)$$

The deformation was considered to be plastic for all but the lightest loads, when $\beta>1$, and elastic under all but the most extreme loading conditions, if $\beta<0.6$. In the limited range, $0.6<\beta<1$, an increase in load could change the contact from elastic to plastic. The majority of surfaces investigated by Greenwood gave mainly plastic deformation. The value of β was formed from calculations based upon the onset of plastic deformation, but for a single contact there is something like a 200 fold range of load (which implies approximately a 50 fold range of the deformation depth) between the onset and what Tabor has termed "full plasticity"[7]. Therefore, if severe plastic deformation is being considered, it seems likely that much higher values of β may be required. This view is shared by Whitehouse and Archard who proposed that the values of β under estimated the probability of plastic flow, because the curvature of the peaks increase with increasing asperity height[44]. Thus the highest peaks, which are those involoved in contact, have a smaller radius than the total population of peaks.

7.3 Sampling Interval

Very little experimental data is available relating sampling interval to mean asperity curvature, β, which is a necessary parameter in accurate determination of theoretical predictions. It has been suggested that for surfaces which possess a Gaussian distribution of surface heights, a three point analysis should be used to obtain the correct value of $\bar{\rho}$[43]. The value taken should be that value which corresponds to a sampling interval of 1/3 of that resulting from the main frequency of the surface structure, (i.e. 1/3 of the sampling interval corresponding to an Auto-correlation function of +0.1). Some surface structures may possess more than one principal frequency, in which case it would be necessary to take a value of 1/4 or even possibly 1/5 of the value corresponding to 0.1. However, a sampling interval of about 1/3 seems to be appropriate in most cases.

In general, the plasticity index is significant only if it is applied to the main long wavelength structure of the surface. Whitehouse and Archard argue that if values of β_s corresponding to the smallest scale structure is used, then the derivation of

the plasticity index becomes invalid because of the deformation of adjacent asperities[44].

7.4 Estimation of Real Contact Area

Archard has shown that when elastic deformation occurs the actual contact area and load are related by the equation

$$A_a = KW^m \qquad (14)$$

where K is a function of the assumed surface structure and the elastic constants of the material[45]. Analysis based on different mathematical models show, Figure 23, that m lies between 2/3 (for contact between a smooth sphere and a smooth flat) and 1 (the value approached by more complex models which more nearly simulate the roughness and waviness of real surfaces). As the complexity of the mathematical models increases, the number of contact areas similarly become more near directly proportional to the load with the size of each contact area becoming less dependent on load.

The theory developed by Greenwood and Williamson for elastic contact takes into account two material properties; the hardness and modulus of elasticity, and three topographic parameters; the mean radius of asperities, their surface density, and the spread of mean height[11]. The theory led to expressions of total real contact area, the number of micro-contacts, the load, and the (electrical) resistance between two contacting surfaces, in terms of separation of their mean planes. In agreement with experimental evidence, the theory indicates that the number of micro-contacts and the real area of contact depend only on the load and not on the nominal contact pressure. Further, it also indicates that the separation of the surfaces is not very sensitive to the pressure, the separation of similar surfaces being approximately equal to the center line average of roughness. The fact that the ratio of real contact area to load is nearly constant for elastic contact leads to the concept of an "elastic hardness" which can be used for predicting the real contact area for given loads, just as the conventional hardness is used when plastic deformation is assumed.

The authors assumed that the heights of the asperities were represented by a well defined continuous distribution function. By measuring several metal surfaces, they found that even though the entire height distribution might not be represented by any known function, the uppermost peaks formed a reasonable good

approximation to a Gaussian distribution. In other words, if the contact load is not very high, the Gaussian distribution of heights is certainly a good appproximation.

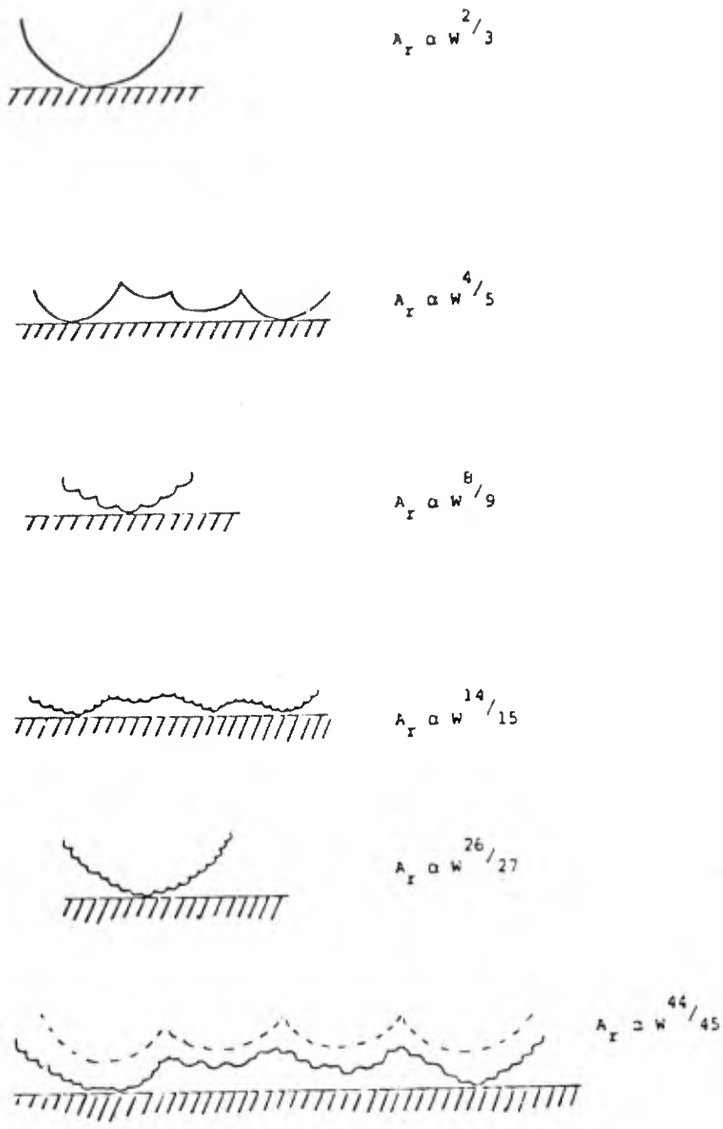


FIGURE 23 ELASTIC SURFACE MODELS

The next significant advance on the mechanism of surface contact was made by Tsukizoe and Hisakado[46, 47]. They considered the contact between two surfaces, one of which had a Gaussian distribution of surface heights. In order to predict the number of contacts and the average contact spot radius, they

had to make the following assumptions about the shape and distribution of surface asperities:

i) The distribution curve obtained from the profile curve had a normal distribution.
ii) The surface was isotropic.
iii) The surface contained a large number of asperities in the form of cones of equal base angle.
iv) Plastic deformation occurred at the contacts. Hence, no interference occurred between asperities, and the ratio of the real area of contact, A_r, to the nominal area of contact A_N, was given by:

$$\frac{A_r}{A_N} = \frac{P}{P_m} \tag{15}$$

or

$$\frac{A_r}{A_N} = \frac{W}{A_N P_m} \tag{16}$$

where P = applied nominal pressure

P_m = flow pressure of the softer material

W = applied normal load.

Based upon the above assumptions, Tsukizoe and Hisakado deduced expressions for the separation, the real area of contact, the number of contact points, the average radius of the contact points, and the distribution of the radii of the contact points. A similar analysis was recently presented by Jones[43], which is based upon the work of Tsukizoe, Hisakado and Thomas[46, 47, 48].

Tsukizoe and Hisakado obtained an expression for the mean plane separation, t = (u/ ∂, and the nominal and real areas of contact expressed in terms of the probability of contact for the normal distribution;

$$\Phi(t) = \left| 1 - \frac{2W}{A_{N/P_m}} \right| \int_{\circ}^{\infty} \phi\,(t)\, dt \tag{17}$$

where $\Phi(t) = \int_0^t \phi(t)\, dt$ (18)

They assumed that the surface was isotropic and that the distribution curve obtained from the profile curve of the surface was normal[46];

$$\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \quad (19)$$

where t is the normalized plane separation given by $t = u/\sigma$, u is the mean plane separation, and, σ, the r.m.s. roughness for the two surfaces, is obtained from

$$\sigma = (\sigma_1^2 + \sigma_2^2)^{1/2} \quad (20)$$

where σ_1 and σ_2 are the r.m.s. roughness of the profile curves for surfaces 1 and 2 respectively.

They also assumed plastic deformation and hence constant material flow pressure.

The real area was estimated from Equation (16) It was assumed that the flow pressure was almost load independent and there was no work-hardening of the conical asperities which have a small base angle.

Finally, the population of micro-contact bridges formed between mating surfaces may be expressed in terms of probability functions.

Figure 24 shows the distribution of sizes of micro-contacts as a function of dimensionless mean plane separation, t, (and hence of the imposed load) for contacts between nominally flat isotropic surfaces containing Gaussian distributions of surface heights[46, 47]. The data is presented in dimensionless form, micro-contact radii being normalized using a normalized mean absolute surface slope, ψ. The plotted function is of the form

$$f(a'_i) = \frac{\left[(t + a'_i)^2 - 1\right] \phi(t + a'_i)}{t\, \Phi(t)} \quad (21)$$

where $a'_i = \pi\Psi a_{i/2}$

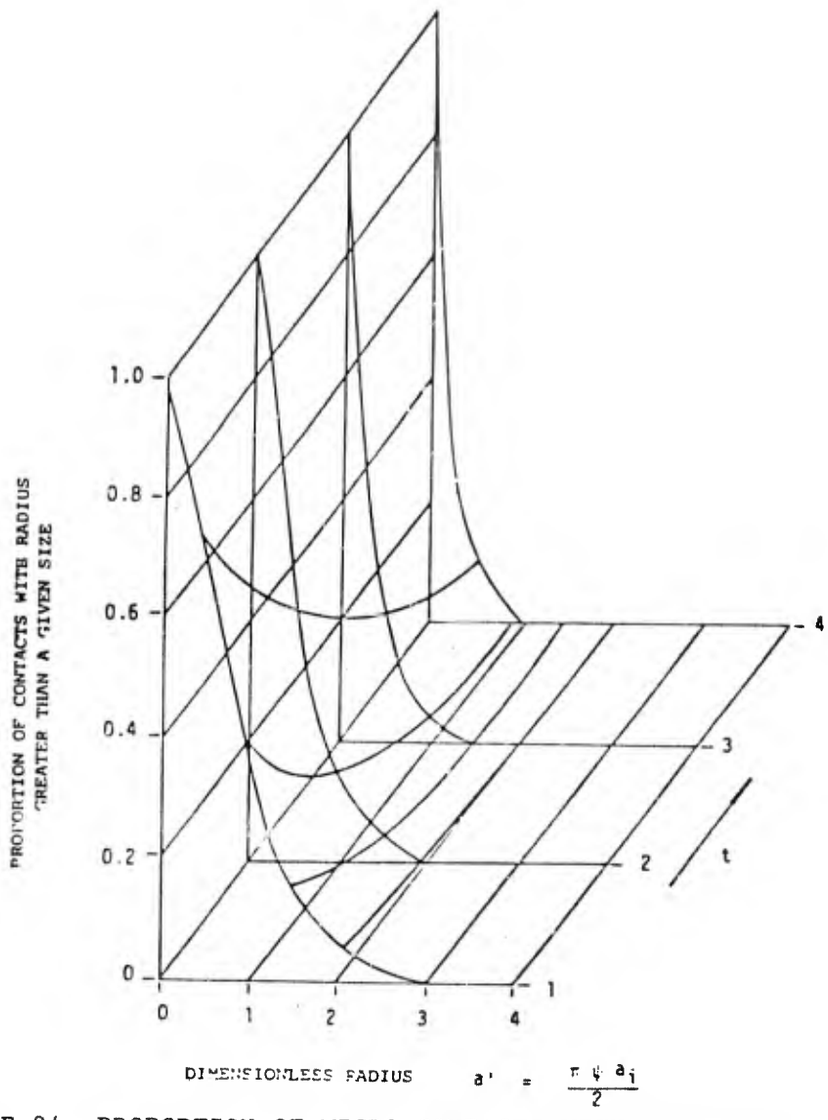


FIGURE 24 PROPORTION OF MICRO-CONTACTS WITH RADIUS GREATER THAN A GIVEN SIZE

Tsukizoe and Hisakado presented the theoretical relationships relating load with mean plane separation, mean contact spot size and number of contact spots, graphically in dimensionless form[46, 47]. From this it was shown that as the load increased, the average contact spot size increased slowly while the number of contacts increased quite rapidly. Thus, the

total area of contact would probably be increased by the increase in the number of contacts.

Tsukizoe and Hisakado tested their hypothesis experimentally using similar and dissimilar materials produced by different machining processes. They concluded that:

i) The relation between load and separation given by Equation (17) was reasonable for contact between normally distributed surfaces.
ii) The real area of contact under the applied load was independent of the surface roughness. The experimental results confirmed the assumption that deformation of the metal at the contact was plastic.
iii) The number of contact points under the appplied load was influenced by the degree of roughness and was proportional to $\tan^2/\sigma 2$.
iv) Over a wide load range ($10^{-4} < P/P_m < 10^{-2}$ MNm^{-2}) the mean contact radius increased by a factor of 1.6 while the number of contacts increased by a factor of 42.2. Therefore, the total contact area increased mainly as a result of the increase in the number of contacts.

In general, the theoretical predictions and experimental results compared quite favorably. For a dimensionless loading, $3 \times 10^{-3} < P/P_m < 2 \times 10^{-2}$, the discrepancy for the main plane separation and load was quite small. O'Callaghan and Probert assumed that when two surfaces are in contact, the contacting regions deform in an ideal plastic manner and the true contact between a rough surface and a flat surface can be regarded as that due to a number of small indentations[49]. Hence the true contact pressure equals the micro-indentation hardness:

$$\text{i.e.} \quad A_r = P/_M \qquad (22)$$

where M is the Vickers micro-hardness of the softer surface.

Thomas and Probert extended the theory of Tsukizoe and Hisakado and developed the following dimensionless relationship for number and mean size of contacts[48];

$$\frac{N_a}{\quad} = \frac{\pi t \ \phi(t)}{8} \qquad (23)$$

and $\bar{a}\psi = \frac{2}{Mt}$ (24)

The dimensionless mean plane separation, t, was linked to the real area of contact, A_r, through the normal probability integral[48].

$$A_r = \frac{1}{2} - \Phi(t) \quad (25)$$

Subsituting Equation (22) into Equation (25) yields

$$\frac{P}{M} = \frac{1}{2} - \Phi(t) \quad (26)$$

Thus the number of contacts spots per unit nominal area, N_a, and their mean radius a, can be related to the loading pressure, P. Subsequently the real area per unit area of contact may be obtained from expression

$$A_r = N_a \pi(\bar{a})^2 \quad (27)$$

The analysis required the measurement of r.m.s. roughness, σ mean asperity slope, $|\psi|$, and micro-hardness, M.

In general, it is extremely difficult to measure the total number of micro-contact spots in practice. Even at very high objective magnification ranges, the resolution of the lens limits the lower size of spots that can be measured. However, it is possible to measure size distribution of the larger micro-contact spots to a high degree of accuracy. In order to compare experimental data with theoretical predictions it is common to use the latter in cumulative form[43, 50].

Figure 25 shows the proportion of contact area supported by contacts of radius greater than a given size, plotted against t and a'. From Figures 24 and 25 it is clear that the disadvantage of not being able to measure the smaller micro-contact spots, is less pronounced for contact area than for the number of contact spots.

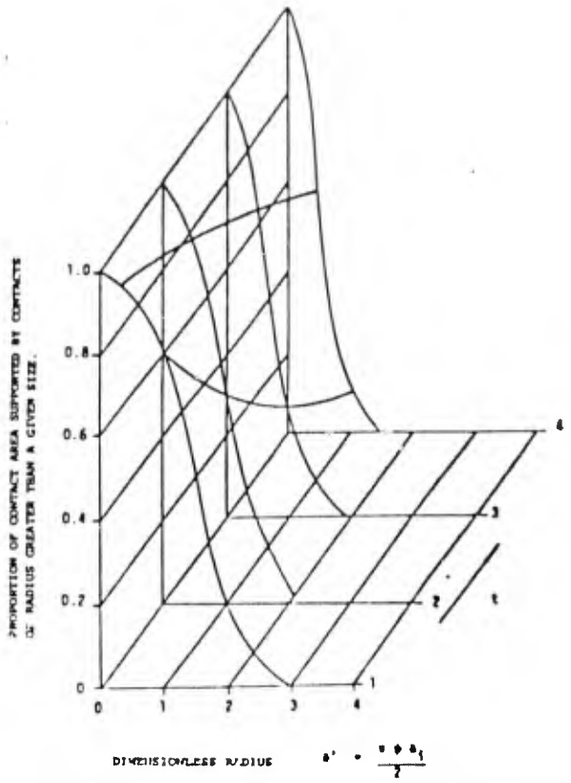


FIGURE 25 PROPORTION OF CONTACT AREA SUPPORTED BY CONTACTS OF RADIUS GREATER THAN A GIVEN SIZE

O'Callaghan and Probert showed that the analysis of Tsukizoe and Hisakado only applies when[49]:

$$2\,\Phi(t) = 1 - \Phi(t)/_{t} \qquad (28)$$

i.e. when P/M < 0.3.

When P/M > 0.3 the distance between adjacent asperities becomes too small for assuming that the deformation of the surface in contact was ideally plastic and no interference occurred between asperities.

By using the arithmetic mean radius, $\bar{a}$, in Equation (24), the real area of contact would be underestimated by approximately 23.5%. The correct parameter should therefore be the r.m.s. spot radius $\overline{\bar{a}}$ [43]. Equation (24) therefore becomes:

$$A_r = N_a\,\pi(\overline{\bar{a}}) \qquad (29)$$

where $\bar{a} = 1.43\ \bar{a}$ (30)

For soft, rough surfaces in contact with harder smooth surfaces, the contact areas are formed by deformation of asperities. The correct hardness for this mode of deformation would be approximated by the Mallock hardness. The Mallock hardness is defined as the flow pressure of a single asperity under load.

The deformation of single asperities under load has been studied by Thomas et al, References 51 and 52. Uppal and Probert have shown that the flow pressure of the single asperity varied with applied load[52]. Thomas found that the flow pressure for single asperity models was several times the Vickers hardness for deformation ratios greater than 0.8. The latter also found variation of the asperity flow pressure with varying asperity slopes. An important factor shown by Uppal and Probert was that when an idealized asperity was deformed, material from the apex was displaced to the shoulders[52]. For multi-asperity models this would result in asperity interaction and yet another parameter influencing the flow pressure of material in contact.

O'Callaghan and Probert stated that despite the existence of many theories to predict values for the parameters describing static contact, there is very little direct experimental data to corroborate the various estimates, and thus it is difficult to quantify the overall effect of the usual simplifying assumptions which are listed below[49]:

i) Large scale surface undulations are neglected.
ii) No interference between adjacent asperity bridges occur. In fact, contact bridge interaction ensues under much lower loads than has been previously suspected.
iii) Contact regions are circular. This applies only at very low loads, and is then only an approximation.
iv) The distribution of surface heights is Gaussian.
v) Asperities may be described as right circular cones.
vi) Contacting asperities deform in an ideal plastic manner. This implies that as loading proceeds all the material within intersections of the surfaces is nullified. It follows that the contact behavior at all loads may be predicted from material properties and geometries measured prior to loading. However, since matter is conserved, the material from the peaks must enter the valleys in a surface. The result of neglecting this displaced material is that

the calculated values of the mean plane separation under load are less than the true values.

7.5 Image Analysis

As previously mentioned, it is extremely difficult to measure contacts very accurately; but one method developed by the author, which has proved quite fruitful, utilizes a Quautimet Image Analyzer, Figure 26[41]. This instrument, which was originally designed to size and count flaws in a metal structure, consists of a microscope through which a specimen is viewed, and a videcon head which transfers the image to a viewing screen via a computer. The intensity of light reflected from the surface of the specimen can be adjusted at selected levels and, depending on the desired mode of operation, detects area fractures in the image, their chord length, their perimeter, and their number and size distribution.

The adopted technique consisted of pressing a hard (steel) machined surface against an optically flat soft (copper) surface under a known load, and upon separation, placing the impressed surface and the copper-surface, under the microscope. The imprint of the asperities are clearly visible even under low magnification. It is then relatively easy to count the number of asperities using the Image Analyzer by taking radial counts across the surface.

Figure 27 shows the spiral effect of a turned surface with the Analyzer set in the perimeter mode. The contact features are those produced by pressing a turned stainless steel surface against a flat copper surface under a load of 2.2 MN_m^{-2}. The contact features away from the center of the specimen, again of turned surface under a load of 18 MN_m^{-2}, are shown in Figure 28. Figure 29 presents contact features of a grit-blasted surface are shown under a load of 2.2 MN_m^{-2}. The analyzer automatically records the data, which can then be fed directly to a teletype printer for subsequent analysis or other data processing devices.

8. PRODUCTION OF SURFACES

A variety of techniques are available for the production of surfaces, several of which are listed in Table 4. Typical values of Ra that should be achieved by each technique are included (providing the tools are in good condition). The latter point, regarding the quality of the tool is particularly important since a blunt, worn, or damaged tool can result in severe degradation of the workpiece in the form of induced stresses and high temperature gradients, which can severely affect the surface integrity of the material as well as the

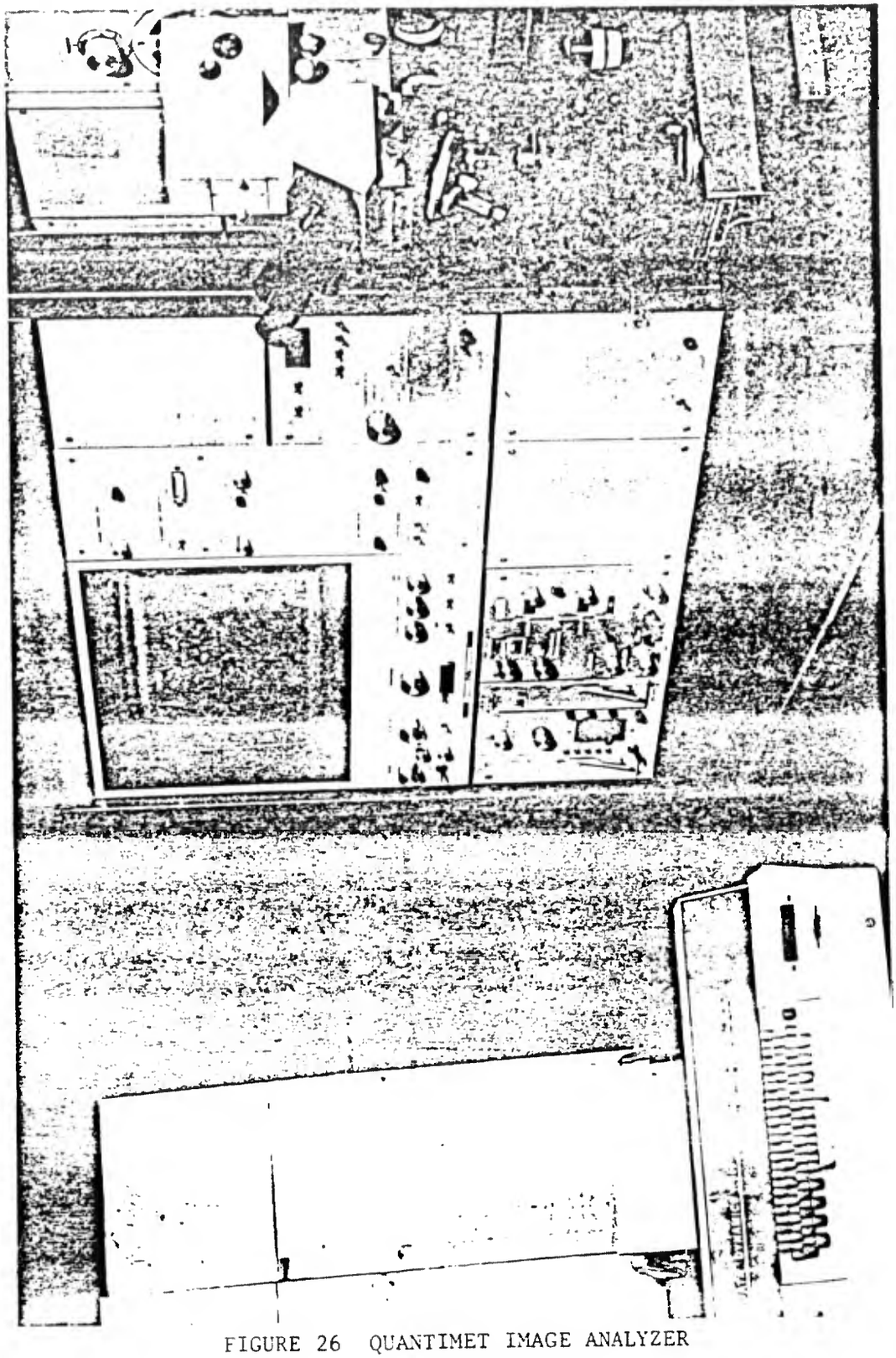


FIGURE 26 QUANTIMET IMAGE ANALYZER

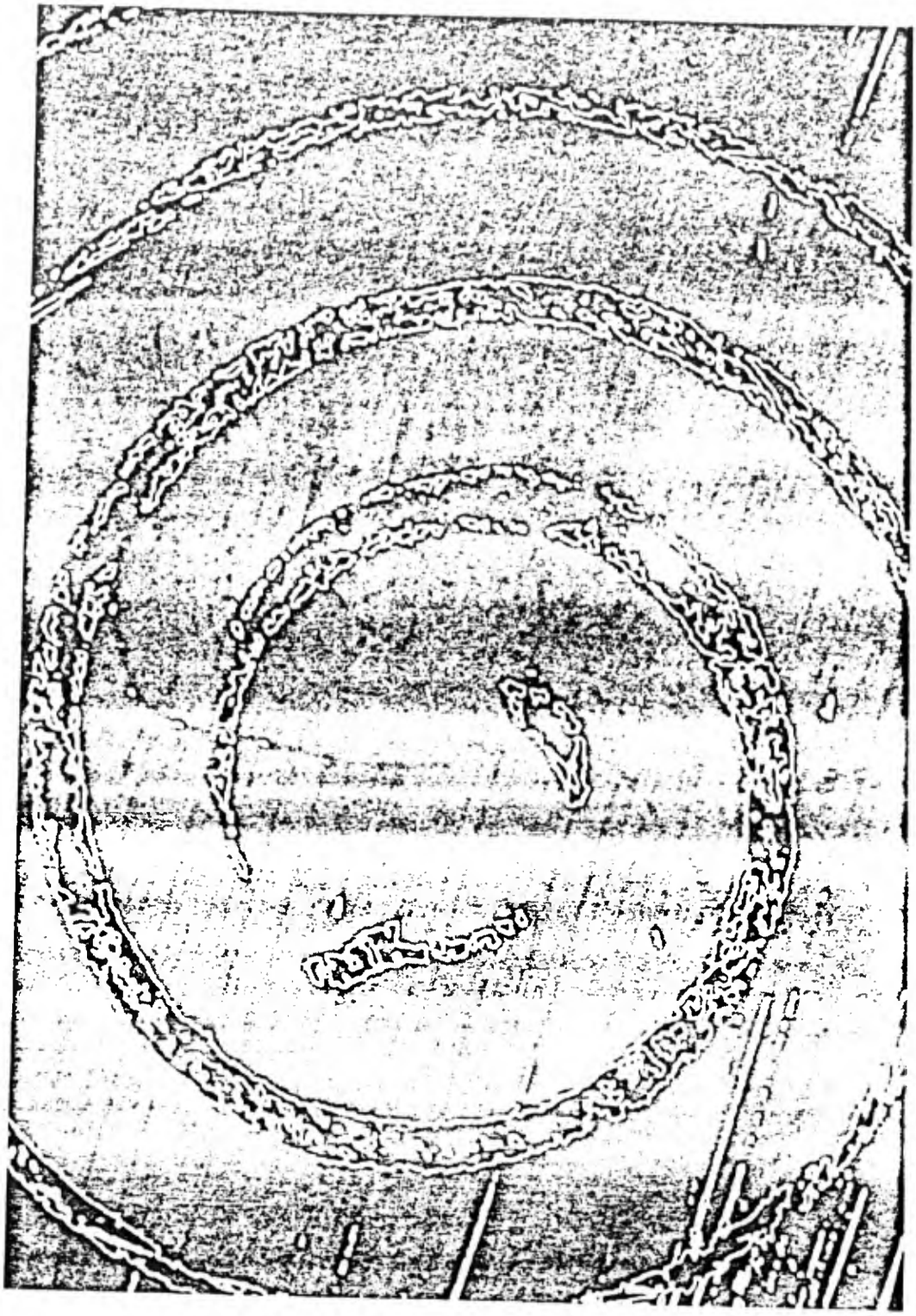


FIGURE 27 CONTACT SPOTS OF A TURNED SURFACE SHOWING SPIRAL EFFECT

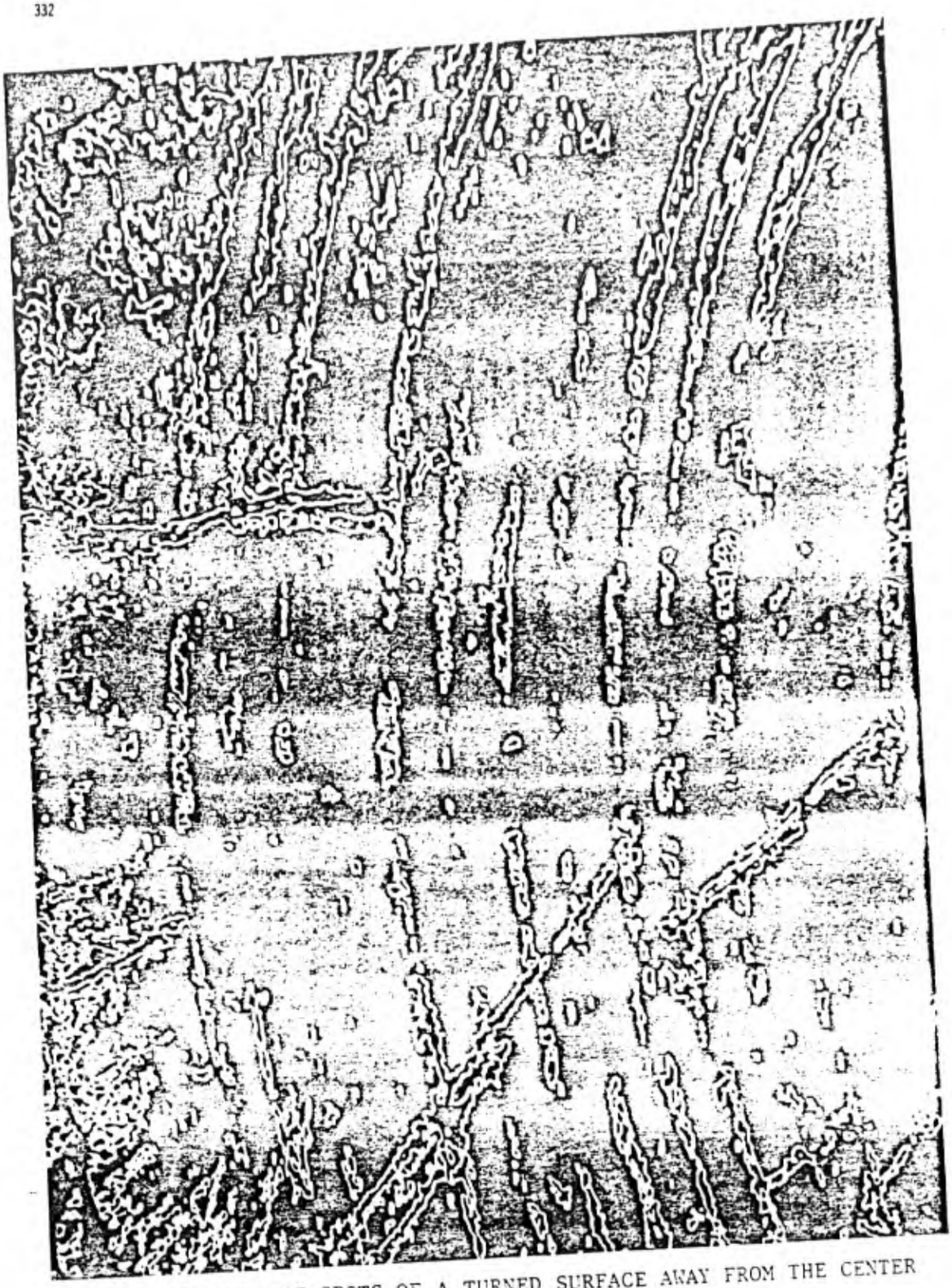


FIGURE 28 CONTACT SPOTS OF A TURNED SURFACE AWAY FROM THE CENTER OF THE SPECIMEN

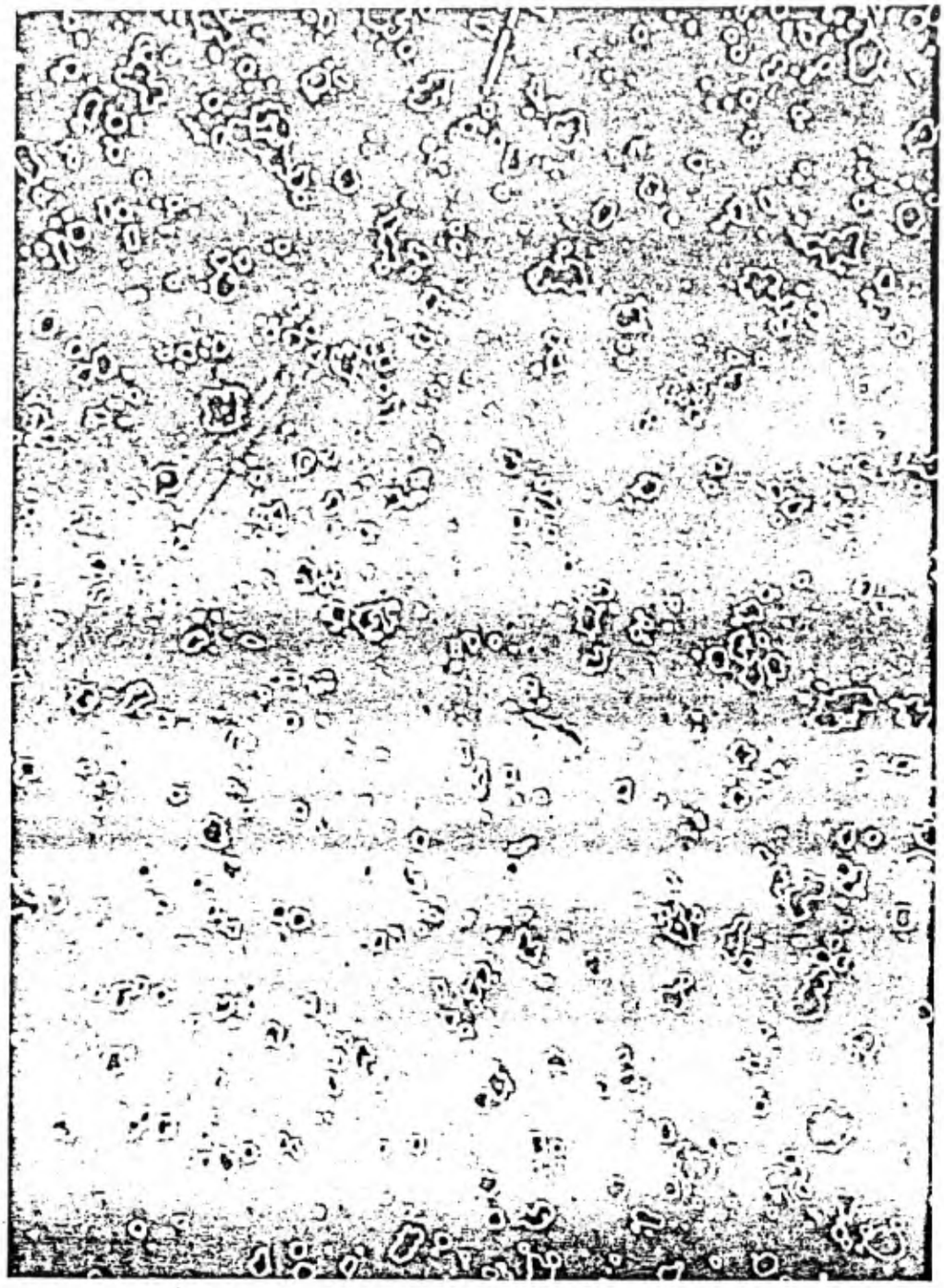


FIGURE 29 CONTACT SPOTS OF A GRIT-BLASTED SURFACE

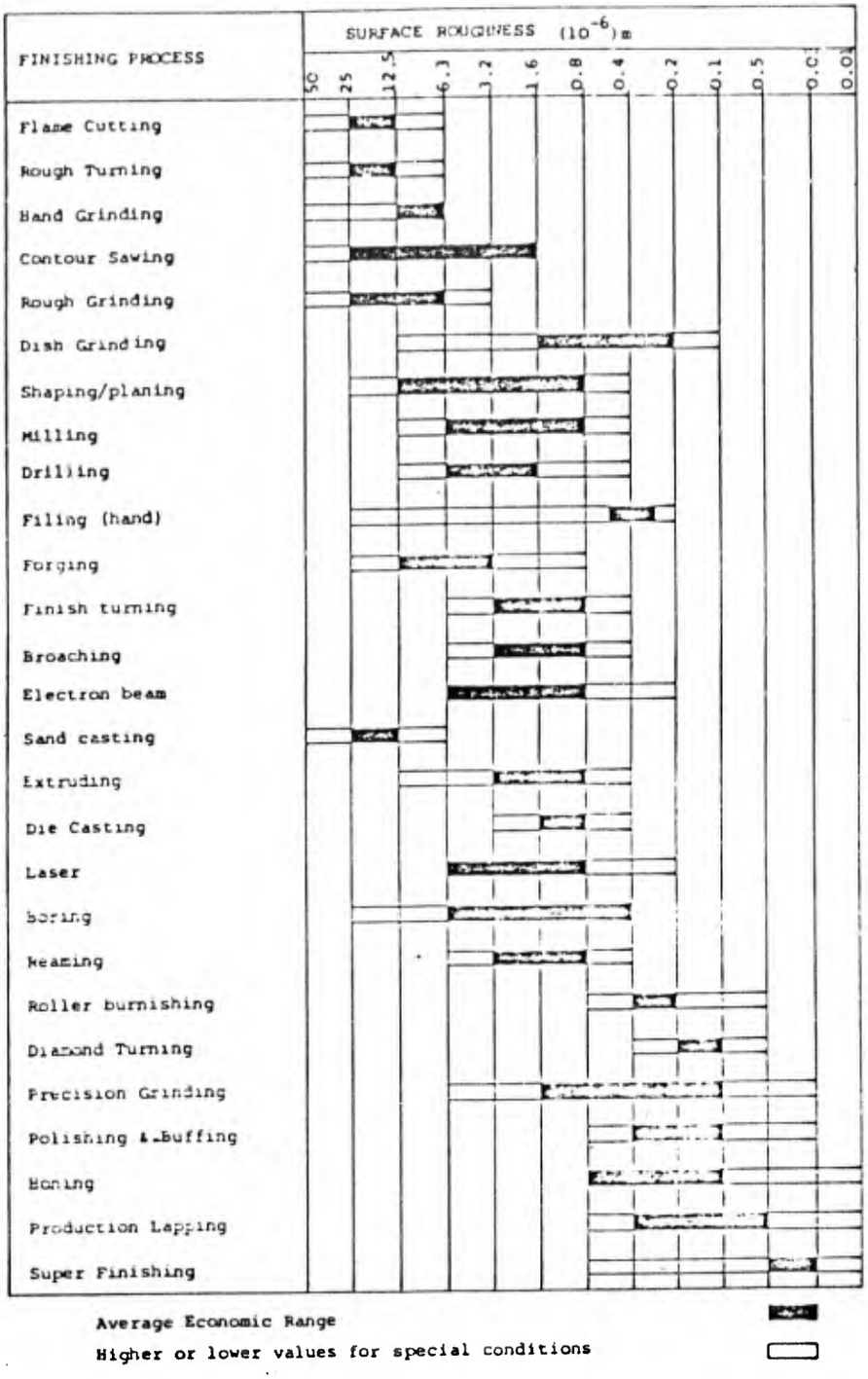


TABLE 4 SURFACE ROUGHNESS PRODUCED BY COMMON PRODUCTION METHODS

surface finish. The need therefore, to control the quality of surface texture, is of paramount importance and is normally a functional requirement of a production engineer. Control is generally instituted not so much in seeking superlative finishing of components, but rather to secure a surface texture of known type and roughness which from experience, has proven to be the most suitable to provide long life, fatigue resistance, maximum efficiency, and functional requirement at the lowest possible cost. The benefits derived are many and include such things as reduction in vibration, wear, and energy consumption.

8.1 Surface Integrity

Although surface finish is of vital importance in any tribological situation, it is of little value, if in creating the desired finish, the surface integrity has been altered to such an extent that the component fails as a result of the stresses it is likely to encounter. Considerable research has been carried out in recent years to establish working criteria on the effects of fatigue, micro-cracks and surface irregularities which are often likely initiation sites for failure. In addition, metallurgical and mechanical alterations, as a result of excessively high removal rates, can reduce the overall quality of the component.

Essentially, the term surface integrity is the study which embraces the interrelationships of metallurgy, machinability, and mechanical testing. For a very good and comprehensive review of all such matters, readers are referred to Section 18 of ASTM-E673 (1980); "Surface Analysis Definitions and Terminology," or alternatively, Surface Integrity Encyclopedia (Special unpublished collection of data and effects from specific material-process combinations), maintained for reference by the Machinability Data Center, Cincinnati, Ohio 45209.

8.2 Surface Roughness as a Result of Geometric and Kinematic Reproduction of Tool Point

A machined surface is created as a result of geometric and kinematic reproduction of the tool, and is generally dictated by both the kind of material and the tool point shape. The tool point shape is characterized by the following:

i) the nose radius,
ii) the leading cutting edge angle and the auxiliary cutting edge angle.

The factor which characterizes the tool point and workpiece kinematics, and which influences surface roughness, is the feed

per tool point. In the case of single-tool turning for example, this is equivalent to the feed per revolution. The resulting swarf material is either of the form of chips or spiral formation.

Considerable research has taken place in this area and various theoretical treatments have been developed, to try and correlate predictable surface finish relative to tool shape, speed of cut, feed, etc[53, 54]. Generally, the theoretical values calculated according to reproduction formula are smaller than the real mean total height of surface irregularities primarily due to the influence of physical and dynamic phenomena. Features such as built-up-edge formation, friction between the cut surface and the tool point, vibrations, and embedding of particles of material being machined, or of the tool point, all influence the effect of surface finish. Tables have been produced in order to aid the machine operator and a wealth of information is available relating cutting speeds and material composition[55]. In addition, some very useful nomograms have also been produced which offer an approximate selection of cutting conditions to achieve a given surface finish[56]. The presence of numerous variables, however, make it impossible to give any precise answers, and thus the skill and experience of the technical operators is still the best criteria.

9. SELECTION OF SURFACES

The selection of surfaces for production purposes is in many respects a matter of subjective judgement based on experience of previous similar systems. As previously mentioned, cost is one of the governing factors in the manufacturing process as shown by the exponential form of the graph in Figure 12. Considerable information has been gained from various surveys covering characteristics of most commonly employed metal forming processes. These include such features as:

i) Dimensional accuracy
ii) As produced surface finish
iii) Machine-finish allowance
iv) Maximum and minimum component size and section thickness
v) Minimum production total
vi) Ideal angle, form of undercut, holes, etc.
vii) Selection of alternative types of material.

Set out in tabular form, comprehensive detail of the above characteristics can be found in the Engineering Design Guides, Series No. 29; The Selection of Materials, by Waterman, Reference 57. These guides are based on information derived

from the Materials Optimizer of the Fulmer Research Institute. Other useful sources of information can be found in References 58, 59, and 60.

For the purpose of this short discussion, Tables 5 and 6 are presented by courtesy of the Editor of the Materials Optimizer, Fulmer Research Institute Ltd. These tables illustrate some of the practical features which must be taken into account, concerning the selection of surfaces.

10. CONCLUSION

This chapter has been written essentially as an overview of a very complex discipline within the field of Tribology. Many detailed areas have, by necessity, been omitted, but it is hoped that the author has been able to convey an appreciation of the subject, and that it will be helpful to those readers who may future, and in what way will surface measurement trends progress? The answer almost certainly lies in-process and on-line measurement, as recently discussed by Young, Vorburger and Teague[61]. Some of the techniques developed in recent years, using non contacting methods, and in particular optical systems, yield the most promising techniques. Laser optics are already now well established, and the data-handling capabilities of modern micro-computers, which can respond rapidly to change, particularly in an area scan, offer advantages to the surface technologist unheard of only a decade ago. Other areas which might be explored are non contacting capacitive transducers, again linked to the computer, and possibly air-jet transducers. Statistical analysis of surfaces will undoubtedly be performed operating in the double integral domain, and no doubt new parameters which more accurately describe a surface, will be developed. Techniques recently reported by Thwaite concerning optical Fourier transforms open the doors to exciting new developments[62].

The need for surface technologists to pursue their activities is paramount, because the development of automated quality control through in-process inspection, linked on-line with process computer systems, will form an integral part of manufacturing assemblies within the next decade.

11. NOTATION

a	radius of a micro-contact spot	m
a^1	dimensionless micro-contact spot radius	
$\bar{a}$	arithmetic micro-contact spot-radius	m

Type of contact		Examples	Type of wear	Remarks	Materials for use
Movement between two components in contact	Conformal (area contact)	Plain bearings, piston rings in cylinders, electrical contacts	Adhesive wear, surface fatigue, some abrasive wear and fretting	Surfaces usually need some running-in	Soft materials operating with hard materials
	Counterformal (concentrated contact)	Rolling bearings, gears, cams and tappets, electrical contacts	Surface fatigue, some adhesive wear and abrasive wear	Adhesive wear occurs only when there is a substantial amount of sliding	Hard with hard
	Conformal and counterformal	Plain bearings, piston rings in cylinders, rolling bearings, gears, cams and tappets, electrical contacts	Running-in	Involves mutual adjustment of the surfaces of the two components	Surface treatments and coatings or special surface finishes
	Particulate solids rubbing over a surface	Excavator buckets, sandblast apparatus, and other particulate materials handling plant	Abrasive wear	The type of abrasive wear experienced depends mainly on the contact pressure	Hard (sliding contact), elastic (bouncing contact)
Movement between a component and a mass of material	Particulate solids or liquids in a fluid over a surface	Equipment pumping abrasive slurries, turbine blades in wet steam aircraft leading edges in rain	Particle-impact erosion	Only becomes noticeable at high impact velocities and or high particle densities	Hard (low impact angles), elastic and ductile (high impact angles)
	Stream of fluid flowing over a surface	Ship propellers, pipes and ducts	Cavitation erosion	Requires low local pressure to initiate the process Corrosion can occur	High ultimate resilience characteristic

TABLE 5 CONDITIONS CAUSING WEAR (MATERIALS OPTIMIZER FULMER RESEARCH INSTITUTE 1974)

Material	Machinability index
Ball-bearing steel (BS 970 534A99)	30
Inconel (77.5% Ni-16.0% Cr-6.5% Fe)	35
Phosphor bronze (9.5% Cu-5.0% Sn-0.2%P)	40
Stainless steel (18% Cr-8% Ni)	45
Nickel steel (3.5% Ni.0.3% C)	50
Nickel-chrome steel (0.45%C)	50
Pearlite cast iron	50
Wrought iron	50
Monel (70% Ni-30% Cu)	55
Structural steel	60
Copper ½-hard rolled	60
Aluminium bronze (5% Al)	60
Chrome-molybdenum steel (BS 970 830M31)	65
Cast steel (0.35%C)	70
Chrome steel (free cutting)	70
Cast copper	70
Brass α	80
Cast iron (soft)	80
Free-cutting mild steel	100
Malleable cast iron (ferritic)	120
Free cutting α or β brass	200-400
Aluminium (half hard)	300-1500
Magnesium (6.5% Al)	500-2000

Machinability index is only a very rough guide to the relative ease or difficulty of machining different materials. The figures may be taken to represent ease of machining in terms of speed of metal removal, higher values denoting easier or faster machinability.

TABLE 6 MACHINABILITY INDEX OF SOME COMMON METALS
(MATERIALS OPTIMIZER)
(FULMER RESEARCH INSTITUTE 1974)

$\bar{a}$	RMS mean micro-contact spot radius	m
$e^{\alpha\lambda}$	decay of exponential function	
t	normalized mean plane separation	
y	profile height ordinate	m
y (x)	surface profile	mm
y (x + λ)	height of a surface profile at coordinate (x +λ)	mm
A_r	real area of contact	m^2
A_n	nominal area of contact	m^2
A_i	real area of contact per unit nominal contact area (= Ar/An)	
E	elastic modulus of a material	Nm^{-2}
E^1	effective elastic modulus of a contact	Nm^{-2}
F (λ)	Auto-correlation function	
M	Vickers micro-hardness	Nm^{-2}
N	number of micro-contacts	
N_a	number of micro-contacts per unit area	m^{-2}
P	applied pressure (W/A)	Nm^{-2}
P (W)	power spectral density	
Pm	flow pressure of a single asperity	Nm^{-2}
W	applied loading	N
W*	dimensionless loading factor ($= W/\sigma_s^2 M$)	

α^1	Correlation distance	m
β	plasticity index	
λ	sample interval	m
σ	RMS roughness	m
σ	RMS roughness of a surface diameter d.	m
ρ	radius of curvature	m
ν	Poissons ratio	
μ	microscopic	10^{-6}m
ψ	surface slope	radius
$\lvert\psi\rvert$	mean absolute surface slope	radius
ϕt	normal probability density function	

$$= \frac{e^{-t^2/2}}{\sqrt{2\pi}}$$

$\Phi(t)$ Normal probability integral

$$\int_0^t \phi\,(t)\,dt$$

12. REFERENCES

1. DaVinci, L., Codice Atlanticus, Milan, British Museum Library.

2. Hertz, H., The Contact of Elastic Bodies, Miscellaneous Papers (London Macmillan, 1896) 146 - 168, 173 - 183.

3. Harrison, R.E.W., "A Survey of Surface Quality Standards and Tolerance Costs Based on 1929-30 Precision Grinding Practice," Trans A.S.M.E. 53, 11 - 25 (1931).

4. Firestone, F.A., Durbin, F.M., and Abbott, E.J., "Test for Smoothness of Machined Surfaces," Metal Progress 21, 57-9 (1932).

5. Abbott, E.J., and Firestone, F.A., "Specifying Surface Quality," Mech. Engng., 55, 569-72 (1933).

6. Reason, R.E., "Surface Finish and Its Measurement," J. Inst. Prod. Engrs., 23, 347-72 October (1944).

7. Bowden, F.P., and Tabor, D., "The Friction and Lubrication of Solids, Part I." Book published by Oxford University Press U.K. (1950).

8. Thomas, T.R., and King, M., "Surface Topography in Engineering - A State-of-the-Art Review and Bibliography," Guy, N.G. (ed) (B.H.R.A.) Fluid Engineering, Beds., U.K.

9. Halling, J. (ed), "Principles of Tribology," Macmillan Press Ltd. (1978).

10. Wallace, D.A., "Surface Finish," S.A.E. Trans. (1940).

11. Greenwood, J.A., and Williamson, J.B.P., "Contact of Nominally Flat Surfaces," Proc. Roy. Soc. A295, 300-19 (1966).

12. Greenwood, J.A., "The Area of Contact Between Rough Surfaces and Flats," Trans. A.S.M.E., Ser (F)., J. Lubrication Technology, 89, 1, 81 - 91 (1967).

13. Schlesinger, G., "Surface Finish," Inst. of Prod. Engrs. Report, 231 (1942).

14. Bennet, H.E., and Porteous, J.O., "Relation Between Surface Roughness, and Specular Reflectance at Normal Incidence," J. Opt. Soc. AM., 51, 123-9 (1961).

15. Abbott, E.J., and Goldschmidt, E., "Surface Quality," Mech. Engng. 59, 813 - 825 (1937).

16. Barash, M.M., "Measuring the Finish of Rough Surfaces," Int. J. Mach. Tool Des. Res. Vol. 3, 97 - 100 (1963).

17. Schneider, E.J., "Surface Finish Testing - The Present State-of-the-Art and its Future," S.A.E. Paper 700144, 6 (1970).

18. Chaplin, P.D., "The Analysis of Hull Surface Roughness records," European Shipbuilding, 16, 40-7 (1967).

19. Thomas, T.R., Holmes, C.R., McAdas, H.T., and Bernard, J.C., "Surface Features Influencing the Effectiveness of Lip Seals: A Pattern - Recognition Approach," S.M.E. Paper 1075-128, 16 (1975).

20. Barwell, F.T., "Bearing Systems - Principles and Practices," Oxford University Press (1979).

21. Surface Analysis Definitions of Terminology Relating to. ASTM-E673 (1980).

22. Pearson, J. and Hopkins, M.R., "Plastic Replicas for Surface - Finish Measurement," J. Iron and Steel Inst., 67 - 70 (May 1948).

23. Young, A.P., and Clegg, B.H., "Replica Method for Examining Surface Profiles," Rev. Sci. Instrum. 30, 444-6 (June 1959).

24. Peklenik, J., "New Developments in Surface Characteristic and Measurements by Means of Random Process Analysis," Proc. Inst. Mech. Engrs. London Part 3K, 182, 108-26 (1967).

25. Lonquet-Higgins, M.S., "Statistical Properties of an Isotropic Surface Random Surface," Phil. Trans. Royal Soc., A250, 157 - 174 (1957).

26. Bendat, J.S., and Piersol, A.G., "Measurement and Analysis of Random Data," Wiley (1966).

27. Greenwood, J.A., "Surface Measurement, Experimental Methods in Tribology," Proc. Inst. Mech. Engrs. London, Part 3G vol. 182, 1 - 6 (1968).

28. Williamson, J.B.P., "Microtopography of Surfaces, In Properties and Metrology of Surfaces," Proc. Inst. Mech. Engrs. London, Park 3K, 182, 21 - 30 (1967).

29. Whitehouse, D.J., "Some Ultimate Units on the Measurement of Surfaces Using Stylus Techniques, Measurement and Control," 8, 147 - 151 (1975).

30. Whitehouse, D.J., "Modern Trends in the Measurement of Surfaces," Rev. M. Mec. 21, 19 - 28 (March 1975).

31. Thomas, T.R., "Recent Advances in the Measurement and Analysis of Surface Microgeometry," Wear (33) 2, 205-33 (July 1975).

32. Wilen, J.E., "Characterization of Cylinder Bore Surface Finish, A Review of Profile Analysis," Wear, Vol. 19, 143 - 162 (1972).

33. Radhakrishnan, V., "Statistical Behavior of Surface Profiles," Wear, Vol. 17, 259 - 267 (1971).

34. Nayak, P.R., "Random Process Model of Rough Surfaces," Trans. Amer. Soc. Mech. Eng. Vol. 93, No. 2, 398 - 407 (1971).

35. Nara, J., "About the Standardization and Spectral Measurement of Surface Waviness," Annals of International Committee for Prod. Eng., Research Vol. 9, No. 3, 687 - 693 (1971).

36. Williamson, J.B.P., Pullen, J., Hunt, R.T., "True Shape of Solid Surfaces," In Surface Mechanics Ed F.F. Ling Book Pub. ASME, New York, 24 - 35 (1969).

37. Williamson, J.B.P. and Hunt, R.T., "Relocation Profilometers," J. Phys. E. Sci. Instrum., 1, 7, 749-52 (July 1968).

38. Grieve, D.J., Kaliszer, H., and Rowe, G.W., "A Normal Wear Process Examined by Measurements of Surface Topography," ANN. C.I.R.P., 18, 585-92 (1970).

39. Peklenik, J., and Kubo, M., "A Basic Study of a Three Dimensional Assessment of the Surface Generated in a Manufacturing process," Ann. C.I.R.P., 16, 257-65 (1968).

40. Thomas, T.R., and Sayles, R.S., "Mapping a Small Surface," Journal of Physics. E. Sci, Inst., Vol. 9, 855 - 861 (1976).

41. Edmonds, M.J., "Effects of Surface Configurations on Thermal Energy Transfers Across Pressed Contacts," Ph.D. Thesis, Cranfield Inst. of Tech. (1978).

42. Williams, A., and Idrus, N., "Detection and Measurement of Damage to Surfaces After Static Contact Loading," Proc. Int. Conf. Metrology and Properties of Eng. Surfaces, Leicester Poly, UK., 281 - 291. (April 1979).

43. Jones, A.M., O'Callaghan, P.W. and Probert, S.D., "Prediction of Contact Parameters From Topography of Contacting Surfaces," Wear (31), 89 - 107 (1975).

44. Whitehouse, D.J., and Archard, J.F., "The Properties of Random Surfaces of Significance in Their Contact," Proc. Roy. Soc., A 316, 97 - 121 (1970).

45. Archard, J.F., "Elastic Deformation and the Laws of Friction," Proc. Roy. Soc. 243 A, 190 - 205 (1957).

46. Tsukizoe, T., and Hisakado, T., "On the Mechanism of Contact Between Metal Surfaces - The Penetrating Depth and Average Clearance (Pt. 1)," Trans. A.S M.E. Ser. D.J., Basic Engng. 873, 666 - 674 (1965).

47. Tsukizoe, T., and Hisakado, T., "On the Mechanism of Contact Between Metal Surfaces, Pt. II the Real Area of Contact, and the Number of Contact Points," Trans. A.S.M.E. Ser. F. 3 Lubr. Tech. 90, 1, (81 - 88) 1968.

48. Thomas, T.R., and Probert, S.D, "Establishment of Contact Parameters from Surface Profiles," Br. J. Appl. Phys., Ser. 3.3, 277 - 291 (1970).

49. O'Callaghan, P.W., and Probert, S.D., "Effects of Static Loading on Surface Parameters," Wear (24), 133 - 145 (1973).

50. Edmonds, M.J., Jones, A.M., O'Callaghan, P.W., and Probert, S.D., "Prediction and Measurement of Thermal Contact Resistance," Wear (50), 299 - 319 (1978).

51. Thomas, T.R., Uppal, A.H., and Probert, S.D., "Hardness of rough surfaces," Nature Physical Science, Vol. 229, No. 3, 86 - 87 (1971).

52. Uppal, A.J., and Probert, S.D., "Deformation of Single and Multiple Asperities on Metal Surfaces," Wear (20), 381 - 400 (1972).

53. Radford, J.D., and Richardson, D.B., "Production Engineering Technology," 3rd Edition, McMillan (1981).

54. Kaczmarck, J., "Principles of Machining - By Cutting Abrasion and Erosion," Peter Peregrinns Ltd. (1976).

55. Schubert, P.B. (Ed.), "Machinery's Handbook," 21st edition, Industrial Press. Inc. (1979).

56. Kachmarck, J., "Technological Possibilities of Obtaining Good Surface Finish," Czasopismo Technique 4 1961.

57. Waterman, N.A., "The Selection of Materials," Engineering Design Guides No. 29, O.U.P.

58. ASTM Standards ASTM-E673 Section 18 (1980).

59. Surface Integrity Encyclopedia (Special unpublished collection of data and effects from specific material - process combinations). Maintained for reference by the Machinability Data Center, Cincinnati, Ohio.

60. British Standards. BS. 4500. ISO Limits and Fits Part I: General, Tolerances and Deviations (1969), Also BS 1134, Part 1 and 2 (1972).

61. Young, R., Vorburger, T., and Teague, C., "In-Process and On-Line Measurement of Surface Finish," Annals of CIRP. Vol. 29, 1 (1980).

62. Thwaite, E.G., "The Direct Measurement of the Power Spectrum of Rough Surfaces by Optical Fourier Transformation," Wear, 57, 71 - 80 (1979).

LUBRICATION

W. O. Winer
Georgia Institute of Technology

1. INTRODUCTION

The objective of lubrication is to separate two surfaces of a mechanical system which are moving relative to one another so that the energy dissipation and the surface degradation are held at acceptably low levels consistent with the engineering design objectives for the mechanism. There are many possible ways of meeting these objectives including full film lubrication, boundary lubrication, solid film lubrication, and mixed film lubrication. In this chapter, the function and regimes of lubrication are discussed as well as several specific mechanisms. Emphasis in the material presented, is on introducing the concepts with physical understanding, and directing the reader to the literature available for in-depth study.

The types of mechanical components that require lubrication include slider bearings, rolling bearings, gears, cams, guideways, and others. These components have the common functional requirements of transmitting a force and/or guidance from one component to the other. They all involve the relative motion of one surface to another. The role of lubrication is to provide for minimum surface degradation and minimum energy dissipation at the interface between the two surfaces. The nature of this solution depends on the many defining characteristics and design requirements for the system. These considerations include the kinematics and dynamics of the system, the nature of the environment in which it must function, the composition of the bodies involved, and the design life criteria to be employed.

The objective of lubrication as stated above can lead to solutions which are not normallly though of as lubrication. For example, Halling suggests that the question should be "What is the best solution to the problem of carrying load across the interface with acceptable friction and wear?[1]" When asked in this fashion, we realize that the support of load with minimum friction and wear can be accomplished in some cases without the presence of a lubricant or lubrication. For example, in Figure 1 from Halling, eight different solutions to tribological problems are presented. At least three of these, (e, f, and h) would not normally be considered solutions involving lubrication. In those cases where the relative motion between the two surfaces has very small amplitude, the load could be carried by an elastomer, flexible strips, or a magnetic field, all of which are capable of accommodating the kinematic motion between the surfaces. The solutions shown in (a & g), namely dry contact and rolling elements, although not necessarily involving lubricants, would be considered to be traditional lubrication solutions to tribological problems. The three remaining examples, (b) chemical films, (c) laminar solid films, and (d) pressurized lubricant films are in the traditional sense lubrication solutions to the tribological problems. The chemical film solution (b) and the lamellar or solid film solution (c) can be thought of as similar mechanisms in the sense that they both involve solid boundary films on the substrate surfaces which permit the two solids to come together causing attrition and shear of the solid films between the surfaces. Within solution type (d) (pressure lubricant films) are included many classical types of lubrication where a pressurized film of a fluid (gas or liquid) is formed between the two surfaces causing them to be separated by virtue of the fact that the integrated value of the pressure over the surface is equal to the load being applied. The lubricant film in this case, is readily deformed and sheared with relatively low energy dissipation. Because, in this case, the surfaces are completely separated, there is little or no wear occurring on the surface. The pressurized lubricant film that exists can be formed in many ways. The film can be the result of an externally pressurized lubricant being injected between the two surfaces, in which case it is referred to as hydrostatic lubrication or it can be formed as a result of the relative motion of the two surfaces and the geometry of the film resulting in a self-acting bearing.

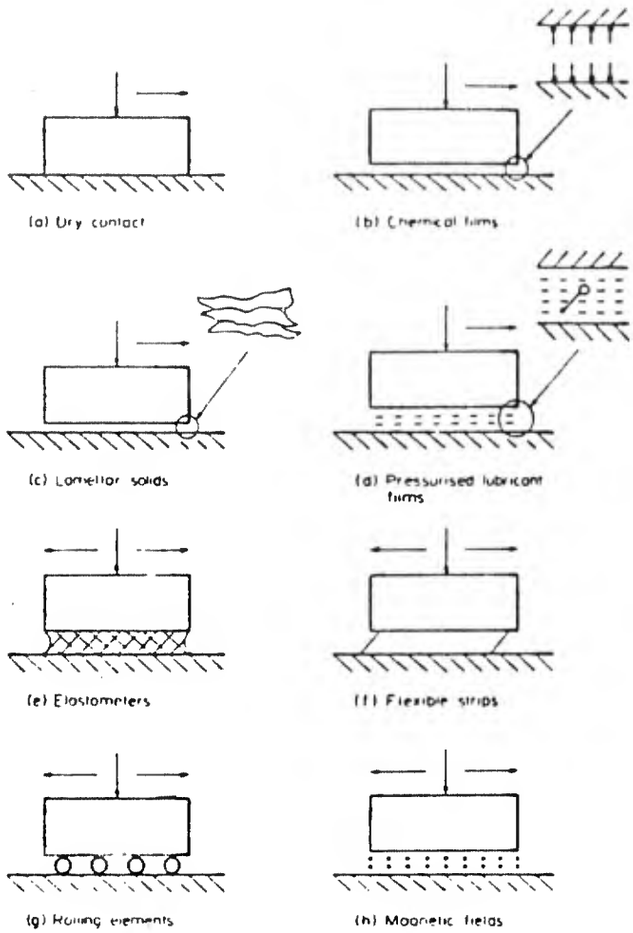


FIGURE 1 METHODS OF SOLUTION OF TRIBOLOGICAL PROBLEMS (HALLING, 1)

In many mechanical mechanisms to be lubricated, more than one type of lubrication can exist in the operating life of the device. For example, in self-acting fluid film lubrication, the relative motion of the two surfaces is necessary to generate the film pressure. However, in such mechanisms at some time during the operating cycle the kinematics may change, for example when the machine comes to rest, or the loads may change in such a fashion that the pressurized film is no longer generated. Under such conditions, the surfaces are permitted to come together resulting in the boundary ... lubrication or chemical film lubrication mechanism referred to in Figure 1 (b) or (c).

2. REGIMES OF LUBRICATION

Regimes of lubrication are normally associated with the dominate lubrication mechanism for the mechanical component being studied. In the present context, it is useful to divide the lubrication mechanisms into three broad regimes. These three regimes can be referred to as full film lubrication, boundary lubrication, and mixed film lubrication. In the literature it is not uncommon for other authors to divide the regimes of lubrication more finely than this. For example, to divide full film lubrication into hydrodynamic lubrication, hydrostatic lubrication, and elastohydrodynamic lubrication. The concept of regimes of lubrication is also used within these various categories to refer to the dominant mechanisms operating for particular contact. For example, as we will see in the regimes herein referred to as full film lubrication , there is a sub-category of thin film lubrication in which there are four possible dominant mechanisms, depending on the kinematics and material properties of the contact. These four sub-categories will also be referred to as regimes of thin film lubrication. However, for introductory purposes, it is helpful to think in terms of full film lubrication being that lubrication mechanism where the surfaces are completely separated by a film of lubricant, which may be a fluid, a gas, or a grease. Within this category, we find hydrodynamic lubrication, gas bearings, hydrostatic lubrication, and elastohydrodynamics.

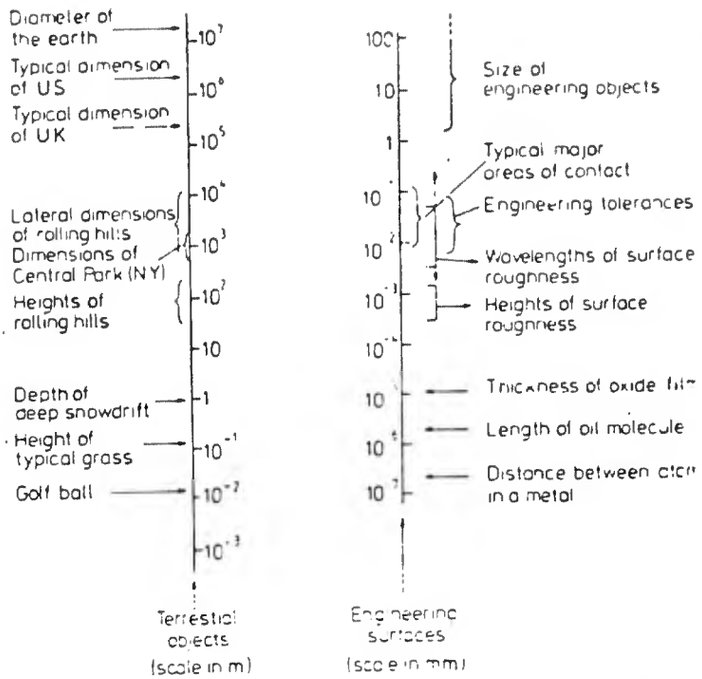
The second broad category referred to here is boundary lubrication. In this regime, lubrication is accomplished by the presence of surface layers on the solid which are sacrificially worn away as a result of the sliding motion between the two surfaces. No hydrodynamic action is present or necessary. In virtually all situations, there is some sort of an adsorbed layer formed on the surface which inhibits adhesion between the surfaces and is sheared during relative motion, hence acting as a lubricant. Within this category of boundary lubrication, we have the sub-categories referred to as EP (extreme pressure) lubrication, solid film lubrication, and unlubricated mechanisms such as those cases where the solid component is made out of a self-lubricating material.

The third general regime of interest is mixed film lubrication, which is, as the name implies, a combination of the full film type lubrication and boundary lubrication. This regime occurs when the hydrodynamic action generates a pressure which is insufficient to completely separate the two surfaces. This separation can be viewed in the context of the ratio of the film generated by the full film mechanisms to the combined surface roughness of the two surfaces. In virtually

all hydrodynamic lubrication analysis, it is assumed that the surfaces are completely smooth when a film thickness is predicted from the analysis. If that film thickness is not greater than the composite surface roughness of the two mating surfaces, then one would expect some surface contact to occur between the high spots of the two surfaces. At those contacts the lubrication mechanism must be of the boundary lubrication type and, therefore, the overall lubrication mechanism for the device is a mixed film type of lubrication.

The relative scales of the solid components, the lubricant films, and the surface roughnesses involved in these contacts are shown in Figures 2(a) and 2(b). Figure 2(a) is attributed to Archard in the Wear Control Handbook[2]. It shows on the logarithmic scales, the relative sizes of the typical lubricant molecules, and typical oxide film thicknesses, as well as surface roughnesses and engineering tolerances for typical engineering objects. For comparison, it also shows the relative sizes of objects that are readily understood by the reader. Figure 2(b), which is attributed to Halling, shows the relative sizes of the different lubricant layers[1]. Chemisorbed gas, lubricant monolayers, and oxide film are typical boundary lubricant films on the surface. They can be seen in relation to the surface roughness of a lapped engineering surface that is a very smooth engineering surface and elastohydrodynamic or normal full film lubricant films which are typical of the full film lubrication regime. In Figure 2(b), the relationship can also be seen between these two types of lubricant films and the solid surface structure of the surface to be lubricated. This surface consists of a rather ill-defined surface layer referred to as the Beilby layer, a heavily deformed region resulting from production of the surface, and the more lightly deformed material below that, before the solid substrate is reached. The relative scales for the two figures can be easily appreciated by noting that one Angstrom is 10^{-7} mm.

The ratio of the full film lubrication film thickness to the combined surface roughness has long been recognized as an important parameter in lubrication. The early work of Stribeck on journal bearings showed that for small values of ZN/P (Z being the viscosity, N the shaft rotational speed, and P the load on the bearing), hydrodynamic lubrication would be less likely to occur resulting in higher values of the coefficient of fiction and thus greater possibilities of wear occurring. As the parameter of ZN/P increased, the coefficient of friction decreased, went through a minimum,, and then slowly increased again as shown in Figure 3, from Bowden and Tabor[3]. The increase in coefficient of friction for low ZN/P, is the result of increasing solid surface



The Dimensions of Details of Engineering Surfaces are Shown on the Right Hand Side and Those of Terrestrial Surfaces on the Left-Hand Side. The Scaling Factor Between the Two Systems is One Hundred Million.

FIGURE 2(A) COMPARISON BETWEEN SCALES OF SIZE OF ENGINEERING & TERRESTRIAL OBJECTS (REF. 2)

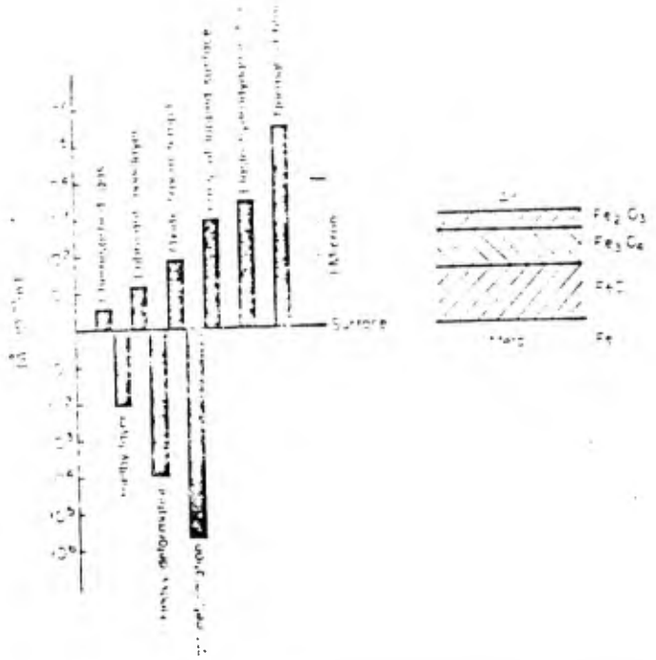


FIGURE 2(B) ORDER OF MAGNITUDE OF SURFACE FEATURES (REF. 1)

contact as the hydrodynamic load carrying ability of the film decreases with decreasing ZN/P. This coefficient of friction is the result of a mixed film mechanism combining the viscous shear of the hydrodynamic film and the solid contact shearing of the boundary lubrication films. The boundary lubrication films typically have higher resistance to shear resulting in higher coefficients of friction. It is also in this regime of low ZN/P, where wear will occur while at the higher ZN/P region where full fluid film lubrication dominates, little wear is expected.

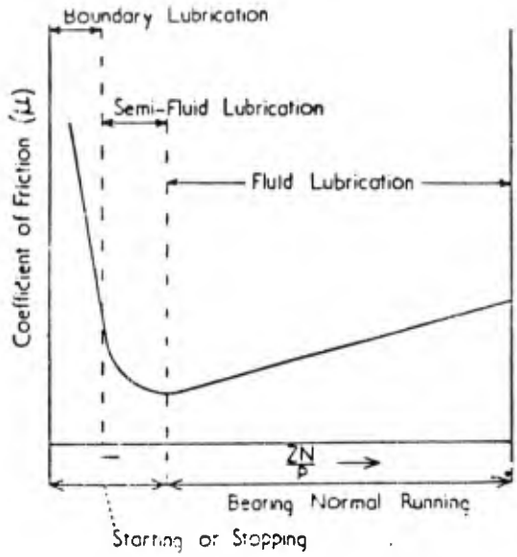


FIGURE 3 FRICTIONAL BEHAVIOR OF A JOURNAL AND BEARING AS THE VISCOSITY Z, NUMBER OF REVS. PER MINUTE N, AND NOMINAL PRESSURE P, ARE VARIED. UNDER CONDITIONS OF PURE FLUID LUBRICATION THE COEFFICIENT OF FRICTION μ IS LOW AND PROPORTIONAL TO AN/P. UNDER LESS FAVORABLE CONDITIONS THE LUBRICANT FILM BREAKS DOWN AND THERE IS A MARKED RISE IN FRICTION (REF. 3)

The fact that the increase in coefficient of friction for low ZN/P is the result of contact between the solids is clearly demonstrated by Czichos in Figure 4[4]. In this figure, the coefficient of friction and the electrical contact time between the surfaces are each plotted as a function of sliding velocity in the left hand part of Figure 4. In each case of a different temperature or different surface roughness, it can be seen that as the time during which no contact occurs is increasing, the coefficient of friction drops and reaches a minimum when the no-contact time reaches 100%. On the right-hand portion of Figure 4, the coefficient of friction and no-contact time fraction are plotted as functions of the

λ ratio which is the ratio of the hydrodynamically generated film thickness to the composite surface roughness. The composite surface roughness is defined as the square root of the sum of the squares of the two RMS surface roughnesses. When the results are rationalized in this fashion, it becomes clear that as the λ ratio increases, the coefficient of friction decreases and the amount of surface interaction decreases. When the λ ratio is less than one, the amount of surface interaction and the coefficient of friction increase and expected wear would also increase. In this particular case, λ ratios less than one appear to indicate that boundary lubrication is the dominant mode, λ ratios from about 1 to 2-1/2 represent mixed film lubrication and for λ ratios greater than about 2-1/2 the mode of lubrication seems to be completely full film lubrication. Consequently, from a design standpoint one can expect, for low λ ratios in the neighborhood of 1 or 2 or less, that the energy dissipated in the contact and the wear occurring in the contact will be greater than that occurring at higher λ ratios where the surfaces are completely separated by a hydrodynamic lubrication film.

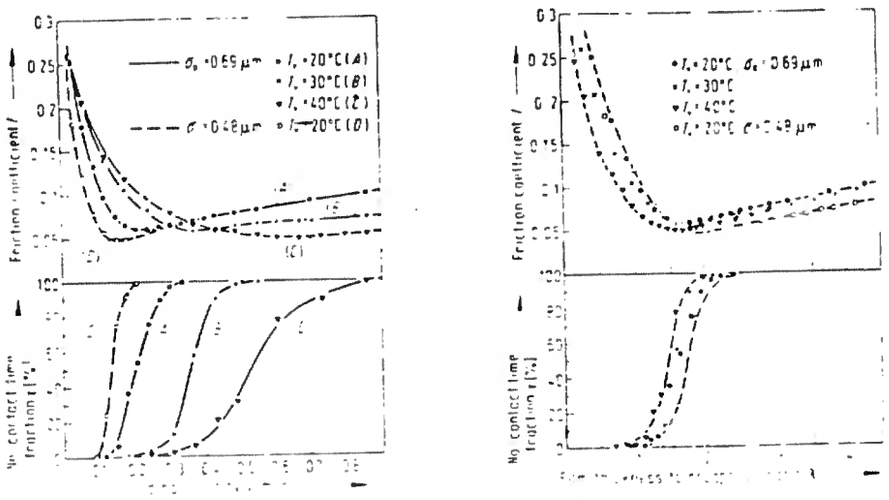


FIGURE 4 (REF. 4) STRIBECK'S CURVE AND THE NO-CONTACT TIME FRACTION FOR A ROUGH LINE CONTACT (F_N=9N; P_H=32.3 MN m^{-2}).

FRICTION AND NO-CONTACT TIME FRACTION VERSUS FILM THICKNESS-TO-ROUGHNESS RATIO (F_N=9N; P_H=32.3 MN m^{-2}).

The λ ratio in concentrated contacts such as cams, gears, and rolling element bearings also has an influence on the life

of the component. In these concentrated contacts if the surfaces are completely separated by a hydrodynamic lubricant film, the mode of failure would be expected to be one of fatigue failure of the surfaces. However, if the surfaces are interacting as a result of a low λ ratio, permitting the asperities to contact one another from the two surfaces, the life of the component will be reduced due to this interaction. Although the details of the relationship between life and lambda ratio are not completely agreed upon by all those active in the field, it can be schematically shown in Figure 5. As a general concept to qualitatively guide design, this is a useful idea; however in quantitative detail, it must be used with some caution. The standard life referred to here is the catalog life in the case of rolling element bearings, but the amount of increase or decrease expected is not clearly defined. In the case of the definition of the λ ratio, there is some uncertainty associated with the film thickness developed and also in the definition of the composite surface roughness. As discussed elsewhere in this text, the details of measuring the composite surface roughness are debatable. For example, cutoff length for sampling wavelength range to be considered and the relationship of these to the size of the lubricated contact, all influence the surface roughness measurement.

Recent work by Winer and Bair has shown, Figure 6, that for concentrated contacts the coefficient of friction (or traction) in the transition region between full film and boundary lubrication is more complicated than in the case of the conformal contacts such as journal bearings[5]. In the case of conformal contacts, the pressures are relatively low and the viscous drag causing the friction in the full film portion of the behavior is dominated by the viscosity of the lubricant. However, in concentrated contacts where the pressures can be quite high, the limiting shear stress behavior of the lubricant plays a role in determining the friction. This property will be discussed more fully in the following subsection on "Lubricants".

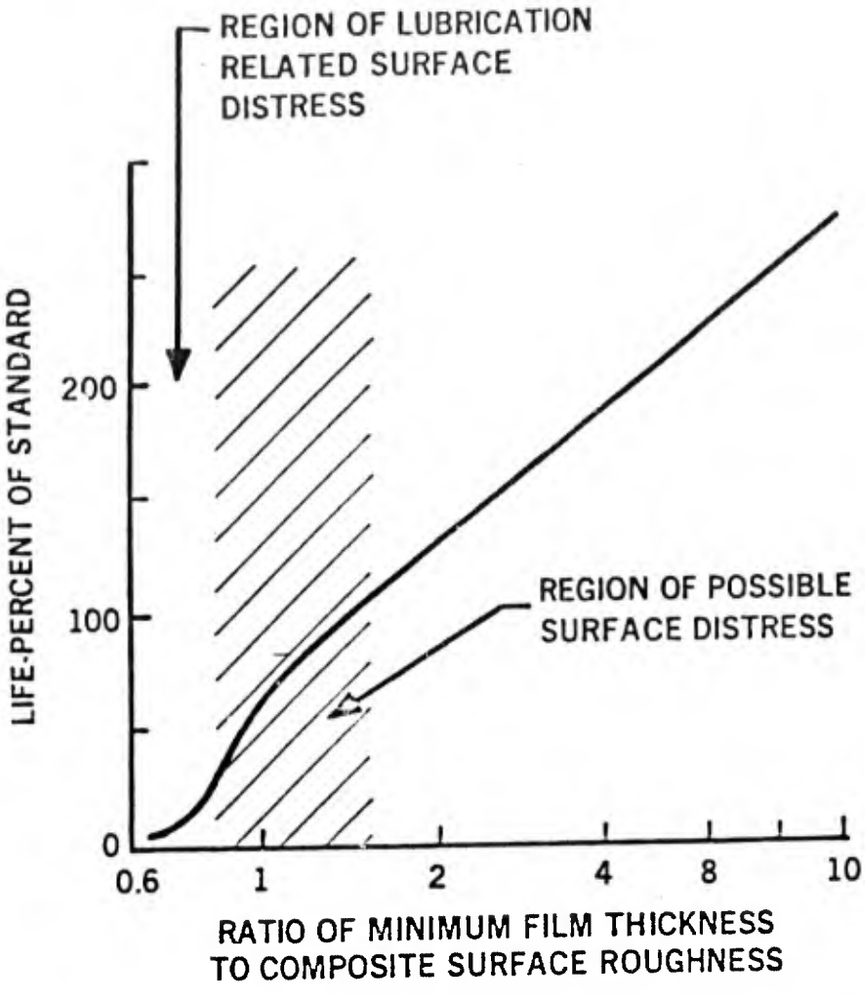


FIGURE 5 FATIGUE LIFE AS A FUNCTION OF THE MINIMUM LUBRICANT FILM THICKNESS/ROUGHNESS RATIO

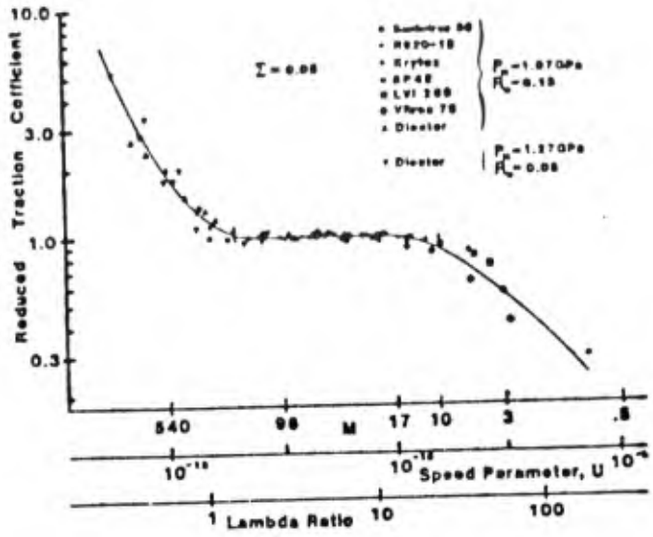


FIGURE 6 REDUCED TRACTION COEFFICIENT IN CONCENTRATED CONTACTS FOR A NUMBER OF LUBRICANTS AS A FUNCTION OF VARIOUS PARAMETERS (REF. 5)

3. BOUNDARY LUBRICATION

In the context of this discussion of lubrication, boundary lubrication includes all cases where the lubricant is a surface film or surface layer of a solid or near solid material which is sacrificially sheared between the two surfaces in order to maintain low friction and wear. In more complete discussions found elsewhere in the literature, a finer division is made in terminology and various types of what is herein referred to as boundary lubrication which may be discussed as separate categories under different names such as solid lubricants, E.P. lubricants, or oxide film lubrication. Boundary lubrication may exist as the only mode of lubrication between two surfaces, during part of the operating cycle, or over part of the geometry as in the mixed film lubrication mode.

The boundary lubrication films can be formed by several mechanisms. They may be mechanically put in place, such as burnishing on solid lubricants, for example, MoS_2 or graphite. Boundary lubricant films may be formed as a result of physical adsorption of materials in the environment onto the surface. They may be the result of chemical adsorption of materials in the immediate environment reacting with materials already on the surface, or they may even be the result of

self-lubricating materials as the substrate structural material when the component is made out of a composite material.

Boundary lubrication is provided by a sacrificial, low shear strength solid layer which adheres to the surface and is replenished, as it is worn away, by adsorption, reaction or insitu formation mechanisms in the temperature range required by the application. A highly simplified example which is attributed to Bowden and Tabor is shown in Figure 7 from Czichos[3, 4]. This is a hypothetical schematic of what might be expected of a system with a lubricating oil present and requiring boundary lubrication. If the operating conditions are such that no hydrodynamic film is developed as a result of the oil present, then the boundary lubrication characteristics of a nonpolar or nonreactive base oil would be rather poor resulting in a high coefficient of friction over the temperature range of interest as indicated by curve I. If a small percentage (a fraction of one percent) of a fatty acid which melts at some temperature T_m, is added to the base oil, the fatty acid will preferentially adsorb on the surface forming a low shear strength solid fatty acid layer which acts as a boundary lubricant at temperatures below its melting point. This fatty acid layer will result in low friction as the temperature is increased until the melting temperature is reached, at which point the fatty acid layer is desorbed leaving the surfaces again without a boundary lubricant and thus the coefficient of friction will rise to that which occurred with the base oil alone. If instead of the fatty acid one were to add an E.P. (extreme pressure) additive which chemically reacts with the surface of the solid at some temperture, T_r, then at lower temperatures there would be no solid boundary film on the surface and the friction would be high until the temperature of the surface reached the reaction temperature for the extreme pressure additive. At this temperature and above, the reaction which takes place results in the formation of a new material which is a solid layer on the surface and has a low shear strength resulting in low friction. Therefore, the E.P. additive acts as a good boundary lubricant at temperatures above its reaction temperature. Although not shown in this figure, the reaction products of the E.P. additive would of course melt or desorb at some higher temperature, again resulting in no lubrication and high friction. The E.P. additive is shown in curve III. Hypothetically one could formulate a lubricant consisting of both a small amount of the fatty acid and a small amount of the E.P. additive and resulting in low friction and good lubrication over the entire temperature range as shown by curve IV, the hypothetical composite lubricant behavior. In theory, with this composite lubricant the fatty acid would act

as the lubricant at low temperatures and would not desorb until temperatures above those at which the E.P. additive reacted to form another film on the surface. At the higher temperatures, the E.P. additive reaction products would function as the lubricant. At still higher temperatures, the reaction products of the E.P. additive would also be desorbed, as previously stated.

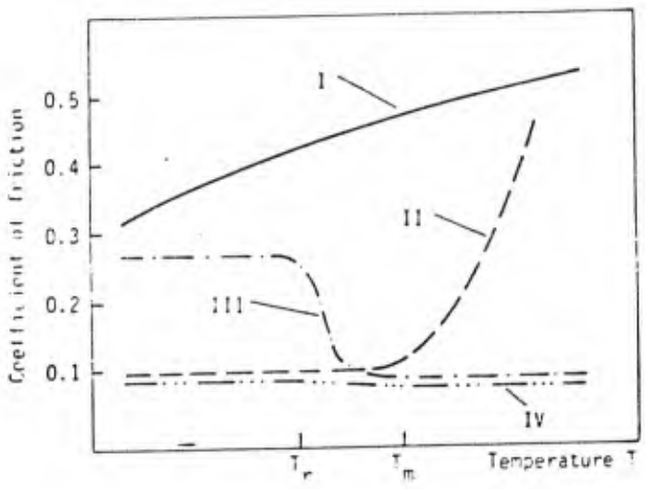


FIGURE 7 FRICTION BEHAVIOR OF BOUNDARY-LUBRICATION SYSTEMS (SCHEMATIC) (REF. 4)
I. NON-POLAR BASE OIL (B.O.)
II. B.O. PLUS FATTY ACID WHICH MELTS AT T_M
III. B.O. PLUS E.P. ADDITIVE WHICH REACTS AT T_R
IV. HYPOTHETICAL COMPOSITE

Although the E.P. additive characteristics described above are frequently referred to in the literature, it is generally recognized that such behavior is incorrectly named. The term "extreme pressure" (E.P.) was the result of early studies of the behavior of certain boundary lubrication additives to lubricating oils. In those studies it was recognized that these additives became active at high loads, that is high contact pressures, hence the name "extreme pressure additives" was given to them. However, it is now recognized that it is not the pressure that is important but

the temperature which resulted from the high pressure which actually causes or permits the reaction to take place. Therefore, these additives would more correctly be termed "extreme temperature additives."

In practice, of course, boundary lubrication is much more complicated than the simple description given above. Boundary lubrication is an area of considerable practical importance. However, it is an area in which relatively little fundamental research has been done and relatively little fundamental understanding exists, at least when compared to the existing understanding and efforts put into hydrodynamic or elastohydrodynamic lubrication. Boundary lubrication is very much an interdisciplinary field involving not only mechanical engineering but also very significant portions of chemistry and physical chemistry. The details of the reactions and behavior of boundary lubricants, which in general are proportionately small quantities added to bulk lubricants, are very complex. Additive interaction and additive competition for reactions with the surface play a significant role in the practical functioning of lubricants in the boundary lubrication regime. The limits of lubrication which are associated with gross surface disruption and sometimes referred to as scoring, scuffing, or severe wear, are also very much dependent upon the limits of boundary lubrication and are extensively studied but poorly understood.

For a more complete understanding of the state of the knowledge on boundary lubrication, the reader is directed to the vast literature on the subject. The books of Bowden and Tabor, Czhichos, Braithwaite, Cameron, and the Standard Handbook of Lubrication Engineering have extensive discussions of boundary lubrication[3, 4, 6, 7, 8]. The ASME publication, Boundary Lubrication: An Appraisal of the World Literature, published in 1969 and edited by Ling, Klaus, and Fein is a very excellent source of the review of the world literature through 1969[9]. A more recent review by Sakurai entitled, "The Role of Chemistry in the Lubrication of Concentrated Contacts," is also a very thorough review of the work of one of the leaders in the field of boundary lubrication[10].

4. MECHANISMS AND TYPES OF FULL FILM LUBRICATION

Full film lubrication exists when the two surfaces moving relative to one another are completely separated by a lubricant film. The lubricant might be either a liquid or a gas and the surface motion might be steady or unsteady. Whatever the material and the surface kinematics are, the operating conditions are such that the lubricant has a pressure developed in it which when integrated over the

surface area is equal to the load being applied to the two surfaces. The behavior of the film can be analyzed and described in terms of fundamental principles of mechanics. The fundamental principles involved are continuity (the conservation of mass) and the conservation of momentum (Newton's Second Law). These two fundamental concepts are combined with a number of appropriate simplifying assumptions into a single equation which is referred to as the Reynolds' Equation. Reynolds first developed and presented this equation governing hydrodynamic lubrication in 1886. If the system can be assumed to have constant properties, then the Reynolds' Equation with appropriate boundary conditions, is all that is required to analyze the behavior of the full film. However, if variable properties, as a result of temperature rise, are to be included, then one must also use the energy equation to predict the temperatures in the film and some appropriate constitutive equations describing the change of the properties with temperature. If property variations are permitted with pressure, as in the case of elastohydrodynamic lubrication or gas bearing lubrication, then additional constitutive equations relating the properties to temperature and/or pressure are required.

A detailed derivation of the Reynolds' Equation can be found in many texts in the literature. These derivations can be found at whatever level of sophistication suits the reader from the relatively elementary presentation of Halling, Trumpler, or Dowson and Higginson to the more advanced treatements of Gross, Constantinescu, Pinkus and Sternlicht, and Tipei[1, 11, 12, 13, 14, 15, 16].

Fundamentally, one analyzes the volume between the surfaces and requires continuity of flow, i.e., conservation of mass of the flow through this volume, and obtains the velocity profiles from the momentum equation. These velocity profiles are integrated across the film resulting in a single equation of the form shown in Equation (1) for the geometry shown in Figure 8.

$$\frac{\partial}{\partial x}\left(\frac{\delta h^3}{\mu}\frac{\partial p}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{\delta h^3}{\mu}\frac{\partial p}{\partial y}\right) = 6\left[U_1 - U_2\right]\frac{\partial(\delta h)}{\partial x} + 6\,\delta h\,\frac{\partial}{\partial x}\left[U_1 + U_2\right] + 12\,\frac{\partial(\delta h)}{\partial t} \qquad (1)$$

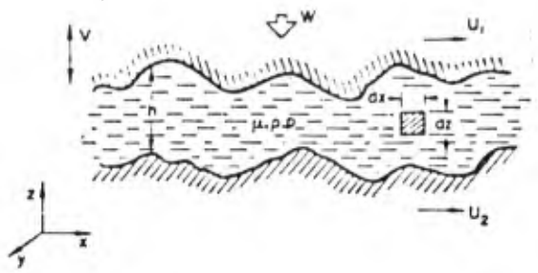


FIGURE 8 HYDRODYNAMIC LUBRICANT FILM GEOMETRY

The major assumptions that were used to develop Equation (1) are the neglect of gravitational and inertial force terms, the assumption of Newtonian lubricant behavior, the assumption that the film thickness is small compared to the lateral dimensions of the film, and the no-slip condition of the liquid at the liquid-solid or gas-solid boundaries of the film. Equation (1) is a rather complicated nonlinear partial differential equation in three independent variables, x and y being the lateral dimensions of the film, and t as time. U_1 and U_2 are the surface velocities of the two surfaces as shown in Figure 8; ρ, p and μ are the lubricant density, pressure, and viscosity, respectively, and are in general treated as functions of space and time; h is the lubricant film thickness, also a function of space and time. Physically, the left hand side of Equation (1) can be thought of as being related to the load being carried by the film because it is related to the pressure distribution which when integrated over the contact area will give the load being carried. The right hand side of Equation (1) can be thought of as related to the mechanisms generating the pressure and therefore generating the load being carried. The first term on the right hand side is the "geometric wedge" term which involves the product of the velocity difference between the two surfaces and the wedge shape or slope of the film resulting from the fact that the two surfaces are not parallel. This is the most common mechanism of generating the pressure in a hydrodynamic lubrication film. The second term on the right hand side is sometimes referred to as the stretching term. It involves the product of the film thickness and the change of surface velocity with distance along the surface. This is a much less common mechanism for generating pressure in the film than the first term.

The third term on the right hand side is referred to as the squeeze film term because it involves the time rate of change of film thickness between the two surfaces. That is, it represents the normal motion between the two films as the gap is closed. The squeeze film term is common in time varying load or time varying film thickness applications, such as in connecting rod bearings. It is a time unsteady phenomena.

If we further assume that the lubricant properties of density and viscosity are constant and the lower surface is not moving ($U_2 = 0$), then Equation (1) becomes Equation (2) which is a much more commonly recognized form of the Reynolds' Equation. All three basic terms described above in Equation (1) are still present on the right hand side of the equation.

$$\frac{\partial}{\partial x}\left(h^3 \frac{\partial p}{\partial x}\right) + \frac{\partial}{\partial y}\left(h^3 \frac{\partial p}{\partial y}\right) = 6\mu U_1 \frac{\partial h}{\partial x} + 6\mu h \frac{\partial U_1}{\partial x} + 12\bar{\mu} V \qquad (2)$$

In Equation (2) the new variable of μ represents the average viscosity in the film; U_1 represents the sliding velocity between the two surfaces; and V represents the relative velocity between the two surfaces in the direction perpendicular to the plane, that is, the squeeze film velocity. Equation (2) is a partial differential equation with the pressure as the dependent variable for which we are solving. The film thickness, surface velocities, and viscosity are assumed to be known parameters. The solution of this equation requires boundary conditions which will be discussed subsequently. The solution results in the pressure distribution being a function of x, y and t' as well as being inherently a function of the film thickness, surface velocities, and viscosity.

The surface motions associated with the wedge and squeeze film mechanisms of film generation are shown in Figure 9[1]. The wedge action is shown in Figure 9(a) and the squeeze film action is shown in Figure 9(b). Figure 9(c) combines the squeeze film and wedging action which exists in the case of lubricated rollers or in a journal bearing application.

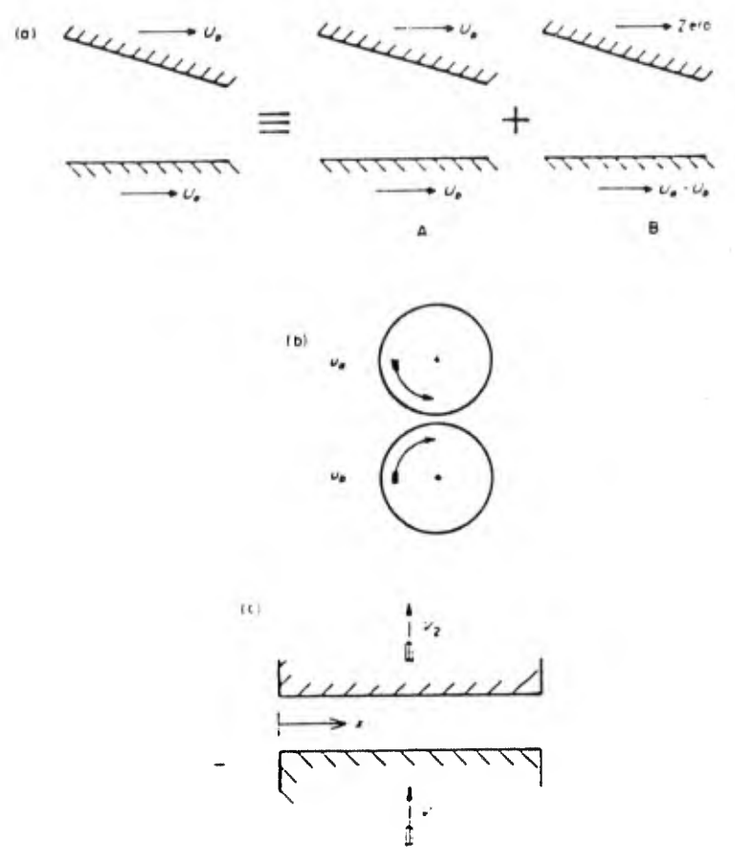


FIGURE 9 SCHEMATIC OF THE MECHANISMS OF HYDRODYNAMIC LUBRICATION (REF. 1)

The boundary conditions used in solving the Reynolds' Equation fall into three basic types: (a) the full Sommerfeld, (b) the half Sommerfeld, and (c) the Reynolds' boundary conditions. The choice of boundary conditions used depends on the particular system of interest and the applications of interest. All solutions to the Reynolds' Equation that exist in the literature and from which design curves have been made, have assumed one of these different types of boundary conditions. It is important to understand the difference in the boundary conditions, the applicability of the boundary conditions, and to make the correct choice when using design charts.

In slider bearings, the boundary conditions are less ambiguous because the pressure can be assumed to be the ambient pressure around the edge of the slider bearing pad.

However in journal bearings, the boundary conditions become more complicated and the three different choices are available and have been used. The full Sommerfeld solution was first used by Arnold Sommerfeld shortly after the turn of the century and consists of assuming that the pressure is equal to some ambient pressure, p_o, at $\theta = 0$, and 2π where θ represents the angular coordinate around the circumference of the bearing. The load vector passes through $\theta = 0$ and π with $\theta = \pi$ at the point of minimum film thickness. It is reasonable to expect that the pressure at the location chosen as 0 is equal to that at the same location which also represents 2π around the bearing. The difficulty with the full Sommerfeld boundary condition assumption is the pressure distribution that results from it when applying that boundary condition to the Reynolds' Equation.

Figure 10 shows the pressure distribution that results in the case of each of these three boundary conditions[1]. The full Sommerfeld solution is shown to result in a sub-ambient pressure distribution in the diverging region of the contact. The angular coordinate, θ, is selected in such a way that $\theta = 0$ represents the location of maximum film thickness, and $\theta = \pi$ represents the location of minumum film thickness in the journal bearing. Therefore, for $\theta = 0$ to π, the film is converging while from π to 2π the film is diverging in thickness. Depending on the operating conditions and the magnitude of p_o, the ambient pressure, the sub-ambient pressure in the diverging region could be large enough to result in cavitation or rupture of the film. If the film ruptures there will be no pressure in that film other than the ambient pressure, p_o. However, the sub-ambient pressure in the diverging region can exist if the p_o is sufficiently high that the pressure reduction in the diverging region does not result in a low enough pressure to cause cavitation in the film. In that case, the full Sommerfeld boundary condition and solution would be appropriate and the load would be the integral of the curve shown. However, in the vast majority of applications, the ambient pressure, p_o, is essentially atmospheric pressure and the pressures developed in the film are large compared to the ambient pressure. In this case, the sub-ambient pressure of the diverging region predicted by the Sommerfeld solution, would result in a near absolute zero pressure or even negative pressure (that is, tensile stress in the film) which, except in very special circumstances, the film could not withstand. The film in the diverging region would therefore rupture. In this much more common situation, the remaining two types of boundary conditions are more appropriate. The most elementary of the remaining two types of boundary conditions is the half Sommerfeld boundary condition, which simply takes the full Sommerfeld solution and

acknowledges that the sub-ambient pressure predicted in the diverging region will not occur, and therefore sets the pressure in the diverging region (i.e., from $\pi \leq \theta \leq 2$) to be the ambient pressure p_o. In this case only the positive portion of the pressure distribution in the converging region is integrated to obtain the load being carried by the film.

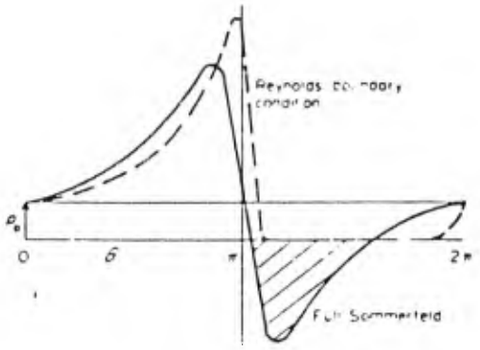


FIGURE 10 PRESSURE DISTRIBUTIONS IN JOURNAL BEARING RESULTING FROM CHOICE OF DIFFERENT BOUNDARY CONDITIONS (REF. 1)

The Reynolds' boundary condition is for a similar operating bearing to that used for the half Sommerfeld solution, but is somewhat more complicated and considered to be more correct for the fundamental behavior of the lubricant. The Reynolds' boundary condition states that the pressure must be equal to the ambient pressure at $\theta = 0$, but at the end of the positive portion of the pressure generated at some unknown location, θ_a, both the pressure must be equal to the ambient pressure and the pressure gradient, $dp/d\theta$, must be equal to 0. The difference between the results from the half Sommerfeld and the Reynolds' boundary conditions is in most cases small. The difference between the solutions obtained with the full Sommerfeld boundary conditions compared to the half Sommerfeld or Reynolds' conditions can be significant depending on the operating regime of the bearing.

The half Sommerfeld and Reynolds' boundary conditions should be used in most applications, particularly those where the ambient pressure of the bearing is atmospheric or near atmospheric pressure. However, if the bearing is operating in a pressurized environment, it would be appropriate to consider the full Sommerfeld solution and to examine the pressure generated in the diverging region relative to the ambient pressure. In the case of gas bearings, the full Sommerfeld

boundary condition is the appropriate one to use, but because the density is not constant the pressure distribution developed will not be antisymmetrical about the minimum film thickness as shown in Figure 10, for an incompressible liquid lubricant.

The designer employing hydrodynamic lubrication solutions would prefer to have as a solution to the Reynolds' Equation the minumum film thickness as a function of the viscosity, surface velocity, and time. The designer would likewise prefer to have the viscous drag, flow rates, and energy dissipation also as a function of the independent variables available, namely the viscosity, speeds, and load. This is referred to as the inverse hydrodynamic problem. However, what is obtained from the Reynolds' Equation is the pressure distribution which then must be integrated to find the load as a function of the film thickness as well as the other variables. These solutions are then plotted in dimensionless fashion to obtain design curves which can be used.

From the above discussion of the Reynolds' Equation, it becomes apparent that the only property of the liquid lubricant that is important in determining its ability to be a hydrodynamic lubricant is the viscosity of the lubricant. Therefore the lubricant can be any liquid material and, depending on the surface speeds, geometry, and load, any viscosity could be used to develop a hydrodynamic lubrication film. The viscosity of liquids is a strong function of temperature. Figures 11(a) and (b) show the viscosity of typical ASTM grade hydrocarbon lubricants and a wide variety of liquids, respectively, as a function of temperature. Additional discussion of lubricant viscosity behavior can be found in the following chapter on lubricants.

4.1 Hydrodynamic Lubrication With Liquids in Journal Bearings

Solutions of the Reynolds' Equation for journal bearings can be found for design purposes in several locations in the literature. The early work of Ramondi and Boyd is readily available in the Standard Handbook of Lubrication Engineering[8]. Other sources include those already mentioned; References 1, 7, 11 - 16. Additional more recent work with special emphasis can be found in the publications of Rohde, et al and Szeri[17, 18, 19]. In these publications, the results are presented as design charts relating the dimensionless groups of important variables for the operation of the journal bearing. The general journal bearing geometry considered is shown in Figure 12. The primary independent variables employed include the Sommerfeld number, S:

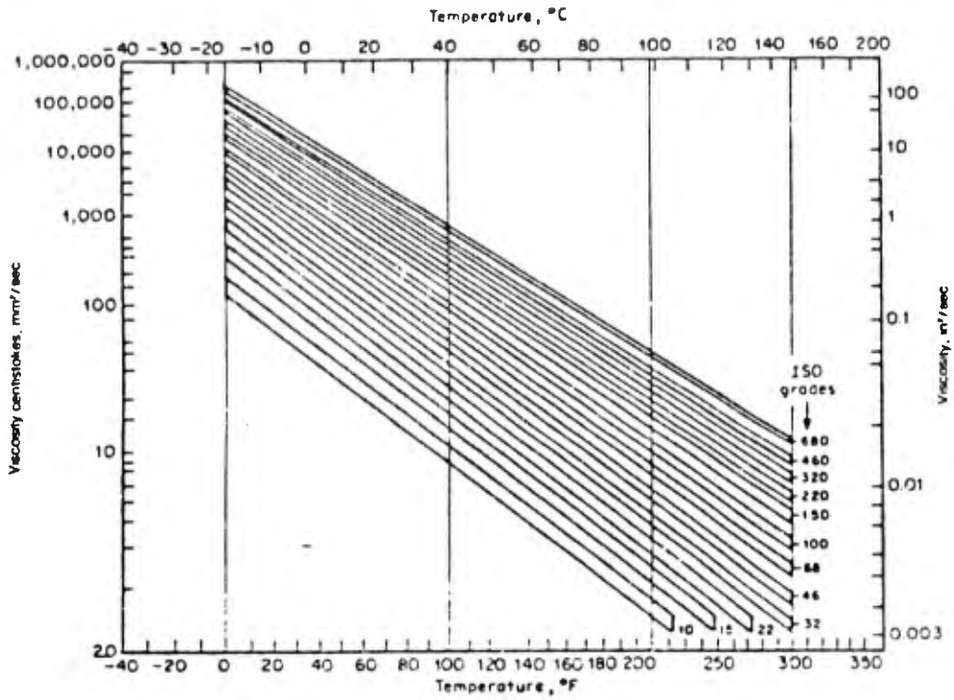


FIGURE 11(A) VISCOSITY-TEMPERATURE BEHAVIOR OF TYPICAL 100 VI LUBRICANT IN ISO VISCOSITY GRADES (REF. 2)

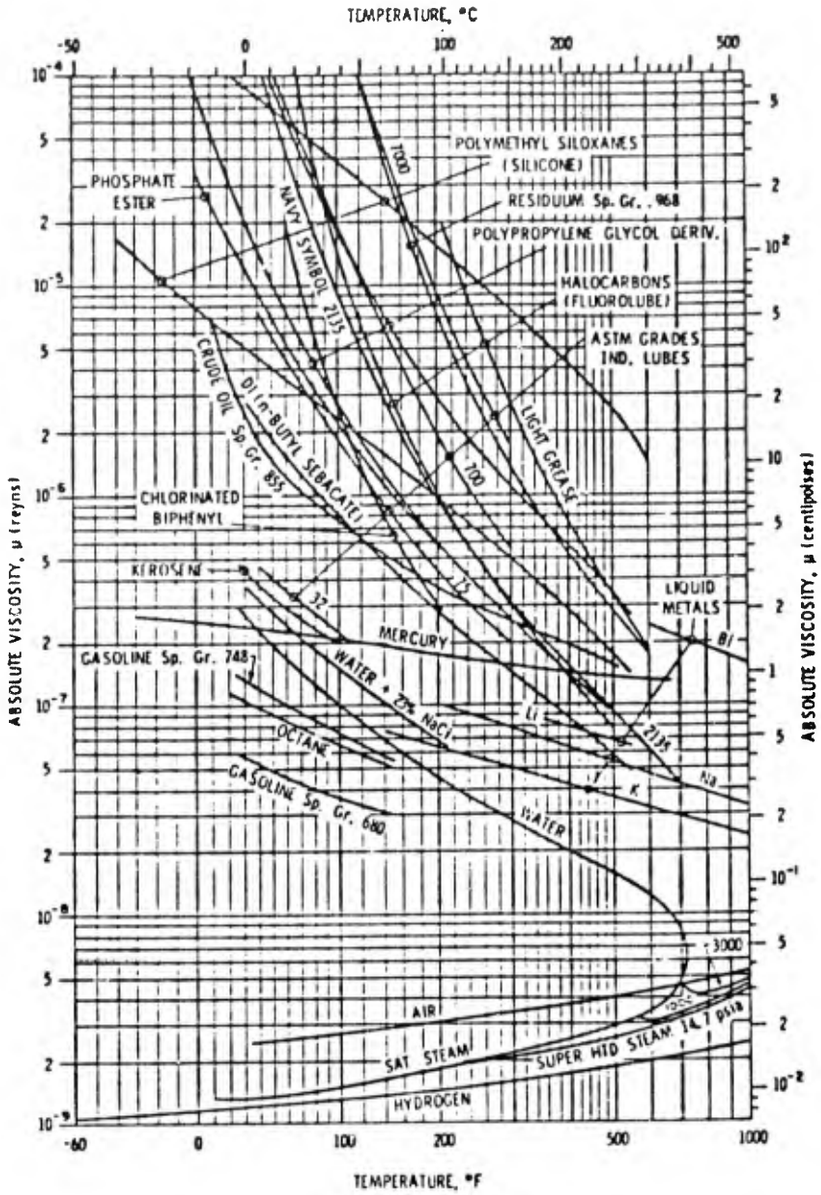


FIGURE 11(B) VISCOSITY-TEMPERATURE CURVES FOR A NUMBER OF FLUIDS (REF. 8)

$$S = \frac{\mu N}{\underline{P}} \left(\frac{r}{c}\right)^2$$

where μ = lubricant viscosity

N = shaft rotational speed

$\underline{P}$ = load divided by the projected area (length times diameter)

r = shaft radius

c = radial clearance between shaft and bearing

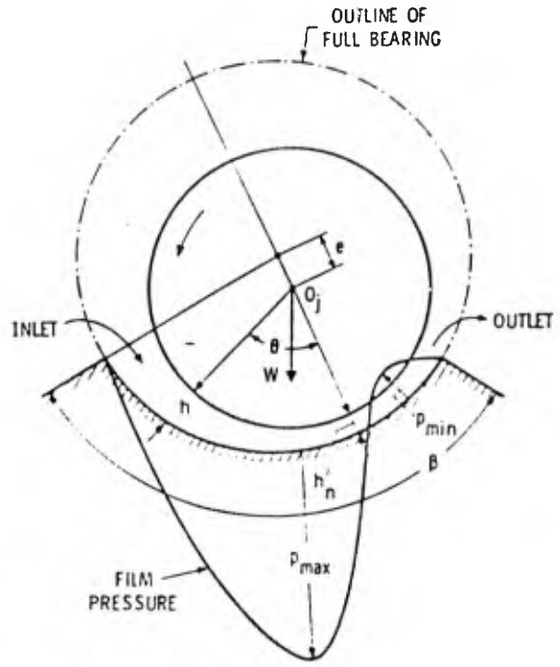


FIGURE 12 JOURNAL BEARING NOMENCLATURE (REF. 8)

The length-to-diameter ratio of the bearing is L/D' and the bearing angle is β. A bearing angle of 2π represents a full journal bearing surrounding the shaft and a bearing angle less than 2π represents a bearing shell which only surrounds part of the shaft.

In the design charts available, a number of dependent dimensionless groups are considered. These groups include: the dimensionless film thickness variable which is the minimum

film thickness divided by the bearing radial clearance; ho/C; the eccentricity, E, which is one minus the film thickness variable

$$\varepsilon = 1 - ho/C;$$

the temperature rise variable

$$\delta C_p \Delta t/P$$

where Δt is the average temperature rise of the lubricant and C the specific heat;

the total flow variable

$$Q/rcNL$$

representing the lubricant flow rate required to maintain the film, Q, the side flow variable

$$Q_s/Q$$

representing the portion of lubricant that flows out the end of the bearing; Q_s,

the friction variable

$$f\frac{r}{c}$$

which is a measure of the effective coefficient of friction, f, resulting from the viscous shearing action of the film; and the attitude angle of the bearing, ø, which is the angle between the load vector and the location of minimum film thickness.

The attitude angle with the eccentricity ratio gives a measure of the relative positions of the shaft center and the bearing center. In some solutions, the location and value of the maximum pressure and the zero pressure of the film are also given.

For the purposes of illustration and introducing the ideas involoved, the film thickness chart and the temperature rise chart for the full 360° journal bearing in a range of length-to-diameter ratios are presented in Figures 13 and 14. In each of these cases, it is seen that the abscissa is the Sommerfeld number or bearing characteristic number, which includes the viscosity, the rotational speed, load, and the

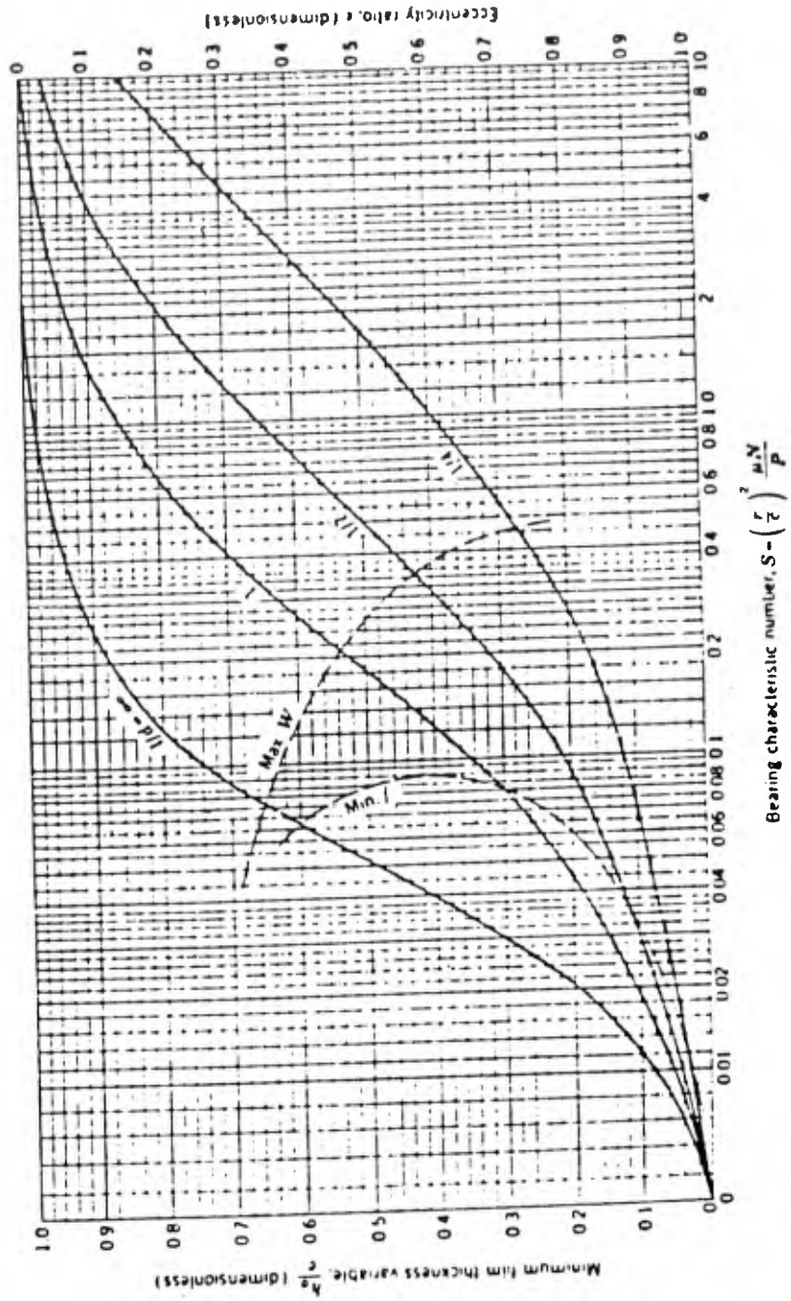


FIGURE 13 CHART FOR MINIMUM-FILM THICKNESS VARIABLE AND ECCENTRICITY RATIO. THE LEFT BOUNDARY OF THE ZONE BETWEEN THE MAX. W AND MIN. F CURVES DEFINE THE OPTIMUM h_o FOR THE MINIMUM FRICTION; THE RIGHT BOUNDARY IS THE OPTIMUM h_o FOR MAXIMUM LOAD (REF. 2). AS ADAPTED FROM RAIMONDI AND BOYD (REF. 8)

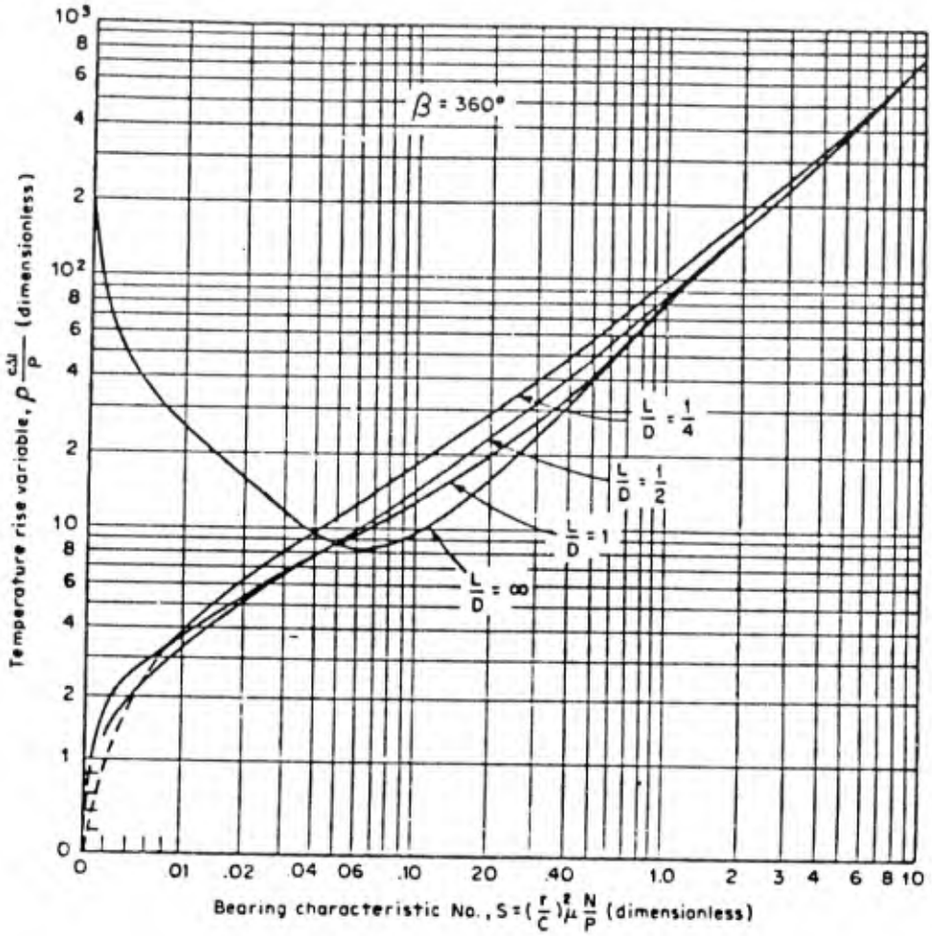


FIGURE 14 TEMPERATURE RISE VARIABLE FOR FULL JOURNAL BEARINGS (REF. 2) (ADAPTED FROM RAIMONDI AND BOYD REF. 8)

clearance ratio for the bearing. The dependent variable or ordinate on Figure 13 is the minimum film thickness ratio of the eccentricity ratio while on Figure 14 the ordinate is the dimensionless temperature rise variable. The temperature rise variable is based on the assumption that all of the energy dissipated by viscous shearing of the film is carried away by the flow of lubricant through the film and none of that energy is conducted to the solid surfaces. This is a conservative estimate in most applications. The temperature rise variable is important because the temperature rise in the film causes a reduction in the lubricant viscosity and the viscosity enters into the bearing characteristic number. Therefore, an iterative solution is required if one is to consider the influence of temperature rise on the bearing operation.

To illustrate the use of these charts consider the following operating conditions for a full journal bearing;

$r = 25.4$ mm

$c = 38$ μm

$L = 51$ mm

$W = 4480$ N

$N = 30$ rps

$\mu = 200$ mPa.sec (e.g. a typical SAE 20 lubricant at 58C)

The average pressure P in the bearing is

$$P = \frac{W}{2rL} = 1.73 \text{ MPa}$$

The Sommerfeld number, also known as the bearing characteristic number S, needed in Figure 13 is;

$$S = \left(\frac{r}{c}\right)^2 \frac{\mu N}{P} = \left(\frac{0.0254}{38 \times 10^{-6}}\right)^2 \frac{20 \times 10^{-3} \times 30}{1.73 \times 10^{-6}} = 0.155$$

From Figure 13 for S = 0.155 and L/d = 1,

$$\frac{h_o}{c} = 0.46$$

Therefore, the minimum film thickness is

$$h_o = 0.46 \times 38 \times 10^{-6} = 17.5\ \mu m$$

The accuracy of h_o using the above simple procedure depends on how accurately one can estimate the effective viscosity, μ, based on an estimated effective film temperature. Because of the strong dependence of viscosity on temperature, the influence of increased effective temperature on film thickness should be checked. Assuming the heat generated by viscous action is available for increasing the lubricant temperature and that the mean temperature of the lubricant leaving the sides of the bearing is equal to the average of the inlet and exit temperatures, the temperature rise of the lubricant is a function of the Sommerfeld number, S, and L/d. The dimensionless temperature rise, $\rho c \Delta t / P$, is presented in Figure 14.

The relationship between the effective film temperature (T_{eff}), used to determine the effective film viscosity, and the temperature rise (Δt) depends on the size of the bearing, the method of lubricant feed, and the thermal characteristics of the bearing. Small bearings ($D \leq 75$ mm = 3 in.) tend to be nonadiabatic and a recommended relationship between the temperature rise (Δt) and the effective film temperature (T_{eff}) is

$$T_{eff} = T_i + \frac{\Delta t}{2} \quad (D \leq 75 \text{ mm})$$

which is the relationship assumed in constructing Figure 14. However, larger bearings (typically $D > 75$ mm = 3 in.) fed with an adequate supply of lubricant tend to operate nearly adiabatically and therefore the recommended effective working temperature (T_{eff}) is taken as

$$T_{eff} = T_i + \Delta t \quad (D > 75 \text{ mm})$$

In both cases, T_i is the inlet temperature of the lubricant supplied to the bearings.

The effective temperature is used to determine the effective film viscosity from a viscosity temperature plot for the lubricant. This viscosity is then used to calculate the Sommerfeld number, S, to determine the thermally corrected minumum film thickness, h_o, from Figure 13.

The determination of the effective operating temperature requires an iterative approach. The approach suggested is to

plot the effective temperature of the bearing, as a function of lubricant viscosity, on the same graph as the viscosity temperature data for the lubricant. Where this bearing characteristic curve crosses the lubricant data curve will give the effective film temperature and effective viscosity for that bearing-lubricant combination.

The lubricant property, ρc, the volumetric specific heat, does not vary greatly with temperature within a given chemical class of lubricants. For hydrocarbon mineral oil lubricants, which are the most common, this property is 1.36 MPa/C, 1.36 x 10^6 N/m^2/C (107 lb/in.^{2}F).

For the above journal bearing example (e.g. D = 51 mm, L/D = 1, therefore use Figure 14 for the effective lubricant temperature)

$$S = \frac{\mu N}{P} \left(\frac{r}{C}\right)^2 = 7.74\ \mu$$

if the viscosity is in units of Pa.sec. By selecting values of viscosity, the above expression along with Figure 14, can be used to calculate the effective temperature as a function of viscosity. If the inlet temperature is 55C for this example, the values of this calculation are shown below:

μ Pa.sec	S	$\frac{\rho c/\Delta t}{P}$	$\frac{\Delta t}{2}$/C	T_{eff}/C
7×10^{-3}	0.054	8.7	5.5	60.5
14×10^{-3}	0.108	13.0	8.3	63.0
28×10^{-3}	0.217	21.0	13.4	68.0
56×10^{-3}	0.433	36.0	23.0	78.0
100×10^{-3}	0.774	65.0	41.0	96.0

If these are plotted on a viscosity temperature curve for the lubricant where the bearing curve crosses, the lubricant curve indicates the effective lubricant viscosity and temperature for that combination of bearing and lubricant. If

the lubricant were a typical SAE 10, the effective viscosity and temperature would be 14 mPa.sec and 62°C. Substituting this viscosity into the Sommerfeld Number for this example gives S = 0.11 and from Figure 13, $h_o/c = 0.37$ so the thermally corrected h_o is 14 μm for a reduction of 20% compared to the isothermal calculation. The above method assumes perfect axial alignment between the shaft and bearing. Once this average effective bearing characteristic is obtained, one can go to the other design charts that are available and determine the viscous torque on the shaft, the flow rate of lubricant required, the shaft center location relative to the bearing center location, as well as the location and magnitude of the maximum pressure and the location of the beginning of film rupture.

4.2 Slider Bearings

Dimensionless design charts obtained from solutions of the Reynolds' Equation are also available for slider geometry. They exist for both the fixed geometry slider pad and the tilting pad bearing. Complete sets of the design charts can be found in the literature, i.e., the solutions of Ramondi and Boyd [8]. For the purposes of introductory example, the fixed pad slider will be introduced in Figure 15. In this case, the dimensionless independent parameters are the bearing characteristic number,

$$K_f = \frac{1}{m^2}\frac{\mu U}{PB}$$

and the length to width ratio of the bearing represented by $\frac{L}{B}$.

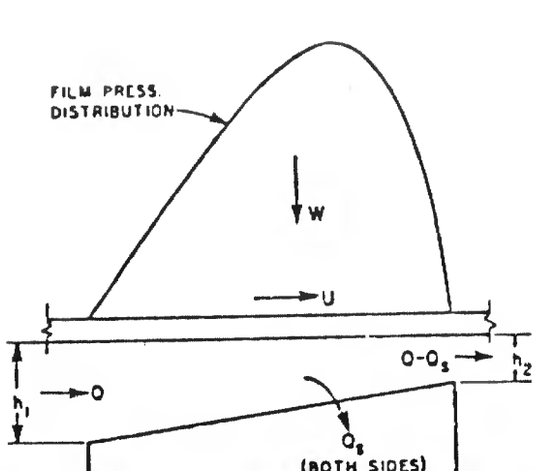


FIGURE 15 DIAGRAMMATIC SKETCH OF FIXED-PAD BEARING (REF. 2)

The bearing characteristic number, K_f, and the width to length ratio, L/B, are dimensionless where

B = length of pad in direction of motion, (m)

L = width of pad perpendicular to direction of motion, (m)

m = slope of pad surface = $(h_1 - h_2)/B$, (dimensionless)

h_1 = inlet film thickness, (m)

h_2 = outlet or minimum film thickness, (m)

U = velocity of sliding surface, (m/sec)

P = load per unit area = W/LB, (N/m^2)

W = load on pad, (N or lb)

μ = average lubricant viscosity, ($N.sec/m^2$)

As in the journal bearing case, slider bearing design charts are available for several independent dimensionless groups including the minimum film thickness variable and the temperature rise variable, as shown in Figures 16 and 17

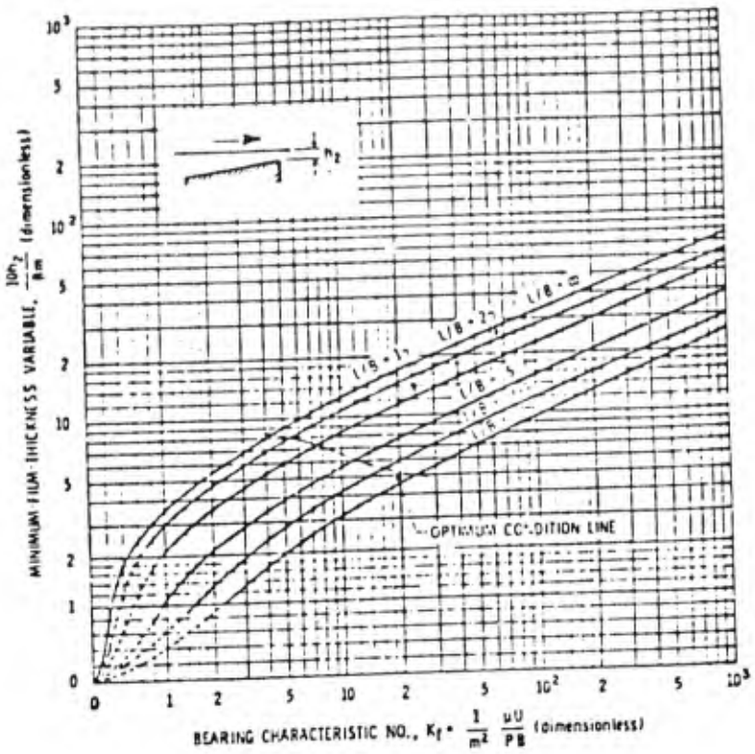


FIGURE 16 CHART FOR DETERMINING MINIMUM FILM THICKNESS (REF. 8)

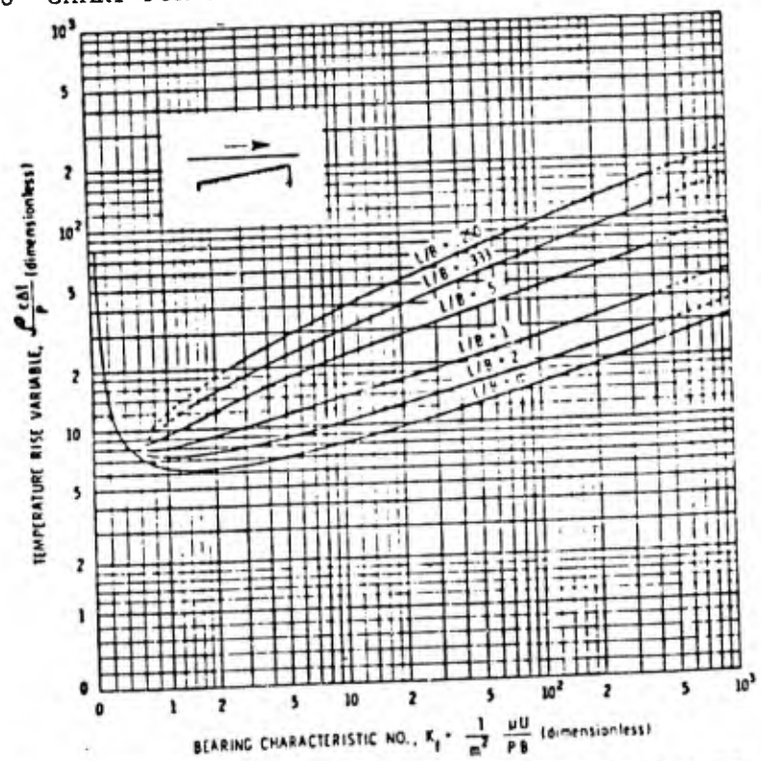


FIGURE 17 CHART FOR DETERMINING TEMPERATURE RISE (REF. 8)

respectively. In the reference literature, design charts for the dimensionless flow rate, side flow, and friction variable are also available.

The use of these charts can best be seen by the following example involving the analysis of a simple single slider with temperature dependent viscosity as the lubricant.

For example, consider the following case:

$W = 1780$ N

$U = 30.5$ m/sec

$B = 0.0254$ m

$L = 0.0254$ m

$m = 0.001 = (h_1 - h_2)/B$ (dimensionless)

and a typical mineral oil supplied at 70°C.

As in the preceeding example with the journal bearing, the bearing behavior is calculated for a range of viscosities. These calculated results are shown as follows:

$\mu / \frac{N \cdot sec}{m^2}$	K_f	$\frac{\rho c \Delta t}{P}$	$\frac{\Delta t}{2}$/C	T_{ave}/C
10^{2}	4.35	10	10	80
2×10^{-2}	8.70	11.8	12	82
4×10^{2}	17.40	14.0	14	84
8×10^{2}	34.8	17.0	17	87
10^{3}	43.5	18.0	18	88

If these were plotted on a viscosity-temperature curve for various SAE oils, the resulting operating points would be found:

SAE Grade	T_{ave}	K_f	$\frac{10\ hz}{m}$	h_2
30	81°C	5.66	6.5	16.5 μm
10	79°C	3.39	5.0	12.7 μm

4.3 Squeeze Films

As mentioned in the introductory portion of this section, squeeze films result from the time unsteady behavior of lubricating films. If the load is fluctuating or the surface velocities perpendicular to the plane of the film change, the lubricant must be squeezed out from between the films resulting in squeeze film action. During the time of the squeezing, a pressure is developed in the film which carries a transient load applied to the surface. This mechanism arises from the last term on the right hand of the Reynolds' Equation. Squeeze film behavior in its simplest form has been demonstrated for well over a hundred years and was analyzed and published prior to the Reynolds' Equation case.

In applications of journal bearings or slider bearings with fluctuating loads, such as connecting rod bearings and reciprocating devices, the load histroy is complex and the solution of the problem is complicated. This field is highly specialized and has relatively few persons contributing to it. Recent examples of some of these developments can be found in the references of Booker and co-worker[20, 21]

5. HYDROSTATIC LUBRICATION

Hydrostatic lubrication results from the introduction of a pressurized lubricant into the oil film. This is usually done with a recessed pocket in the oil film and the pressure distribution, resulting from this oil passing out through the oil film, carries the load. Unlike the hydrodynamic lubrication described above, no surface velocity is required to develop the film. The pressure is developed from an external source of a hydraulic pump. From the mathematical standpoint, this means that the right-hand side of the Reynolds' Equation is or can be zero and the boundary conditions are different in that a particular pressure is imposed upon the system. Hydrostatic bearings are particularly advantageous for carrying very high loads with

little or no velocity between the two surfaces. A typical hydrostatic bearing configuration is shown in Figures 18 and 19[1].

The hydrostatic bearing system consists of a reservoir and associated filtering system, a hydraulic pump, and a compensating element (or restrictor) between the pump and the pad recess. The lubricant may be fed to several pads from a single pipe through a manifold system with restrictors between the manifold and the individual pads.

As in the case of journal bearings and slider bearings, hydrostatic bearings are also a highly developed field in which design curves can be obtained in the literature[1, 8, 22].

The design of a hydrostatic bearing involves the behavior of a single pad and the interaction of several single pads and compensator systems. Both the relative size of the recess in the pad and the pressure drop across the compensator influence the performance of the bearing. The performance is measured in terms of flow rate, power consumption, and stiffness. The stiffness is the derivative of the load with respect to change in film thickness. The behavior of these quantities is given in terms of dimesionless pressure, flow, and power factors for a given geometry as illustrated in Figure 20 for a single circular pad with a circular recess. One can see from Figure 20 that the minimum power consumption, that is pump power required, will occur with a dimensionless pocket radius of about 0.5 in this particular case. In the literature cited, similar dimensionless design curves can be found for a variety of different pad and pocket geometries.

As illustrated in Figure 21 the pads can be arranged around a shaft to result in a hydrostatic journal bearing as well as the linear slider bearings implied above[1].

6. GAS BEARINGS

Gases can be viewed as fluids with very low viscosity and variable density. There is no reason why the previously discussed Reynolds' Equation cannot be applied to gases as lubricants. Because of the very low viscosity of the gases, the operating conditions suitable for gas film lubrication are different from those associated with liquid film lubrication. However, because of the very low viscosity, one would expect low friction, low energy dissipation, and low temperature rise. Gases afford a wide range of temperature for applications because of the high thermal stability of gases at elevated temperatures as well as low temperature capabilities beyond that normally thought of with most liquids. However,

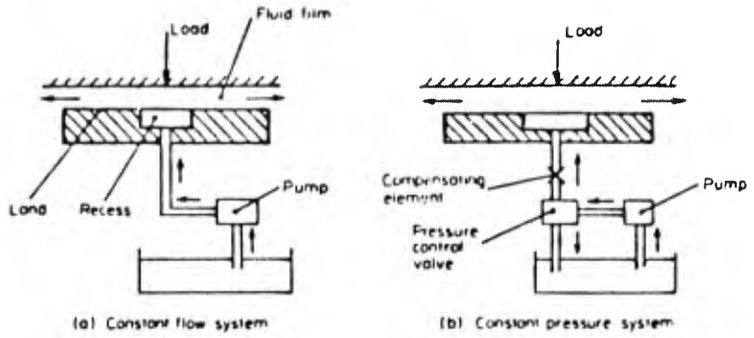


FIGURE 18 TYPICAL ARRANGEMENTS OF HYDROSTATIC BEARING SYSTEMS (REF. 1)

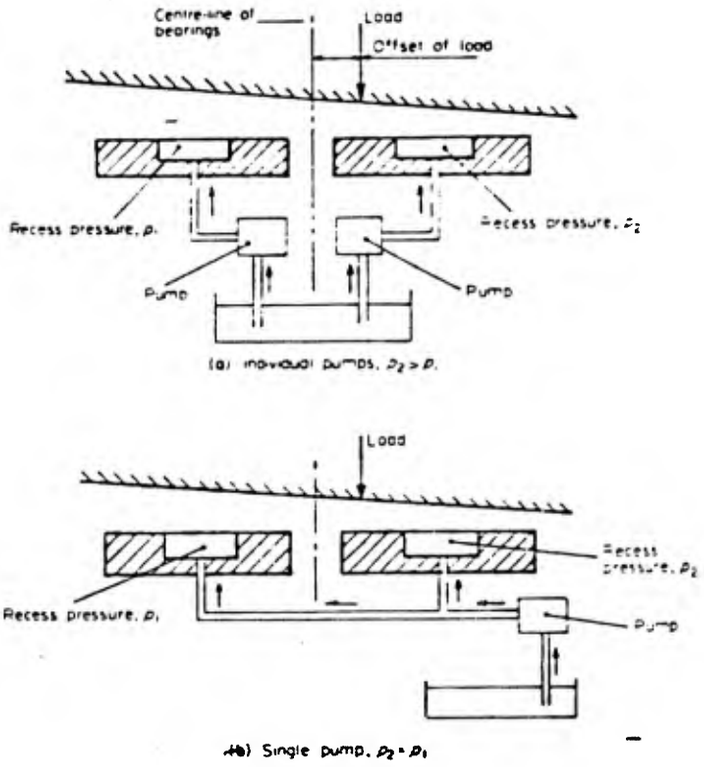
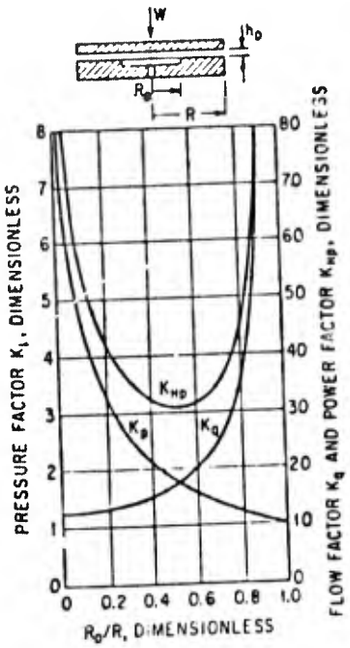


FIGURE 19 MULTIBEARING SYSTEMS (REF. 1)



$$p_o = H_p \frac{W}{\pi R^2}$$

$$Q = K_q \frac{W}{\pi R^2} \frac{h^3}{12\mu}$$

$$Hp = K_{Hp} \left[\frac{W}{\pi R^2}\right]^2 \frac{h^3}{12\mu}$$

FIGURE 20 PERFORMANCE FACTORS FOR CIRCULAR STEP BEARING (REF. 8)

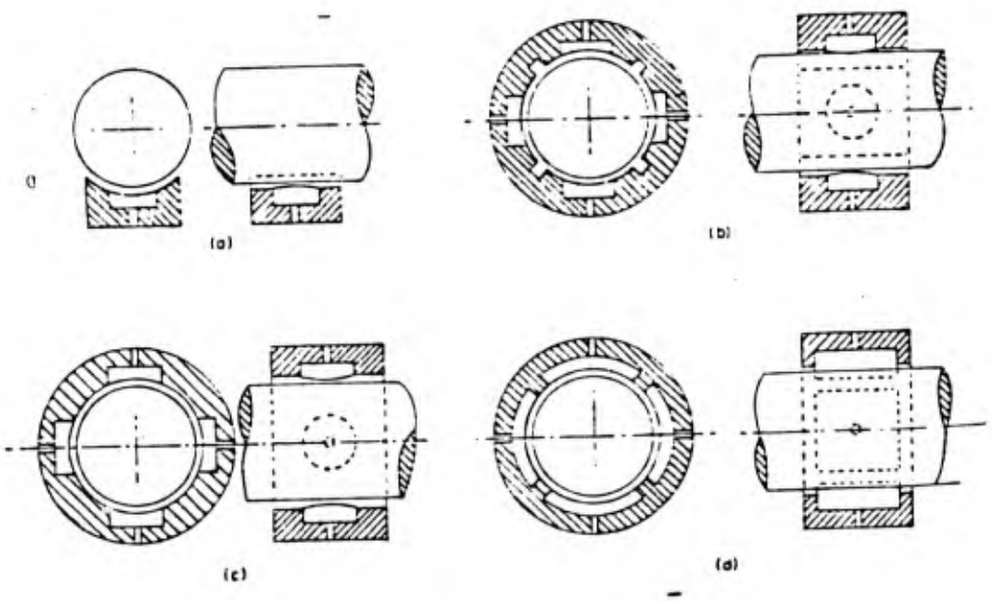


FIGURE 21 EVOLUTION OF HYDROSTATIC JOURNAL BEARINGS (REF. 1)

the disadvantages of gas film lubrication are also associated with low viscosity, which requires either thin films or high velocities and low loads to find suitable operating conditions. Also in the case of a gas as a lubricant, there is no boundary lubricant or extreme pressure lubricant present that can act as a backup lubricant when surface contact occurs. Because of the compressibility of the lubricant and the very low dissipation associated with low viscosity, gas bearings have more of a tendency to become dynamically unstable than liquid film bearings. The dynamic instabilities associated with gas bearings are for more important and more complicated than those associated with liquid bearings. Because of the very small film thickness at which gas bearings operate, the manufacturing of gas bearings is difficult requiring high precision in both the manufacturing techniques and the alignment of the bearing. Comparisons of gas, oil and water bearings, and their range of minimum film thickness operation and typical unit loading ranges are shown in Table 1. Recent developments in compliant gas film bearings have permitted the unit loading to be increased from that shown in Table 1 to as much as 50 psi. Design solutions for gas bearings are also available in the literature[13, 23].

Table 1.

Practical Comparison of Bearings Using Various Fluid Lubricants

Fluid	Viscosity, reyns	Typical min film thickness in bearing applications, in.	Typical unit load* P, psi
Oil	3×10^{-6}	0.002 - 0.004	200 - 500
Water	1×10^{-7}	0.0004 - 0.001	25 - 75
Air	0.3×10^{-8}	0.00005 - 0.0004	1 - 10

* Hydrodynamic bearings

The pressure distribution in a gas bearing is different from that in a liquid bearing because of the absence of cavitation in a diverging region. The pressure distribution in a typical journal bearing is shown in Figures 22 and 23[8].

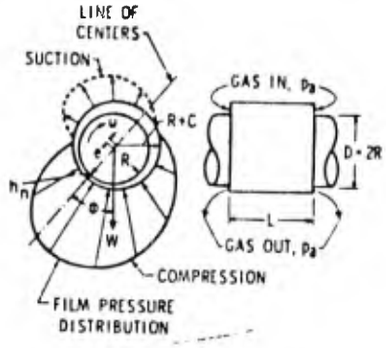


FIGURE 22 FULL JOURNAL BEARING OR PLAIN CYLINDRICAL BEARING (REF.8)

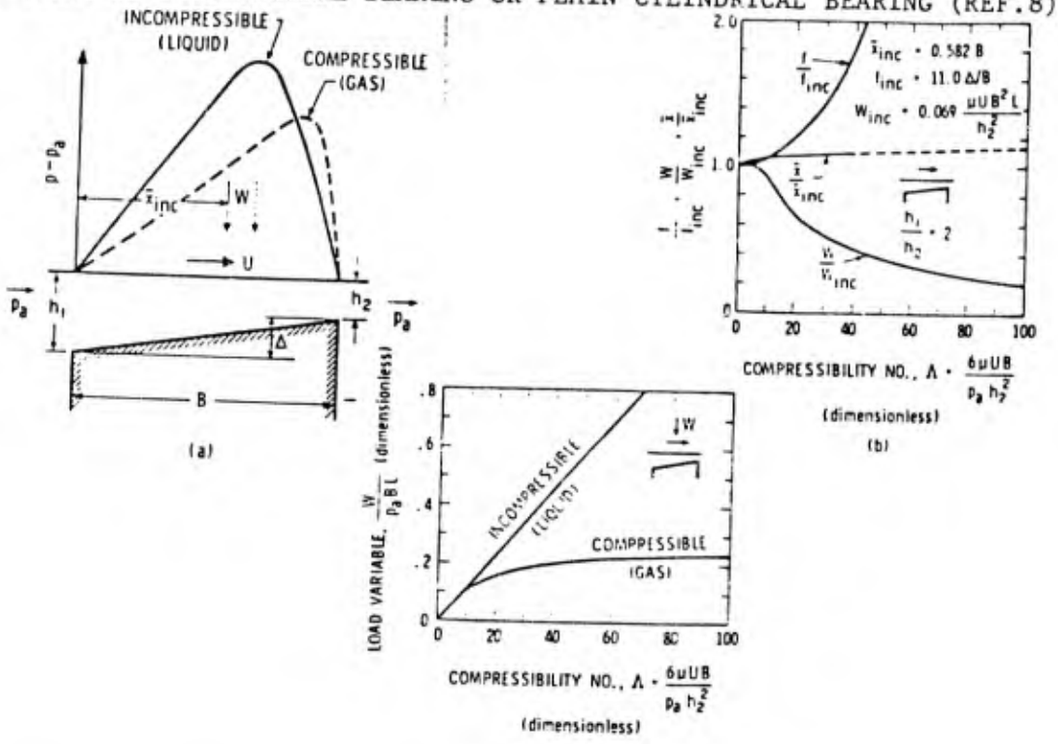


FIGURE 23 TYPICAL COMPRESSIBILITY EFFECTS IN HYDRODYNAMIC BEARINGS (SLIDER BEARING, L/B=1, h_1/h_2=2). (A) EFFECT ON PRESSURE DISTRIBUTION. (B) EFFECT ON COEFFICIENT OF FRICTION f= RATIO OF FRICTION FORCE PER POUND OF LOAD CARRIED = RECIPROCAL OF EFFICIENCY; EFFECT ON CENTER OF PRESSURE (OR PIVOT POSITION) $\bar{x}$; AND EFFECT ON LOAD CAPACITY W. (SUBSCRIPT "INC." DENOTES VALUE FOR LIQUID LUBRICATED BEARING.) NOTE THAT BOTH EFFICIENCY AND LOAD CAPACITY OF GAS BEARING DECREASE RELATIVE TO LIQUID BEARING BECAUSE OF COMPRESSIBILITY; CENTER OF PRESSURE MOVES TO REAR. ALSO NOTE THAT FOR SMALL Λ (Λ < 5, SAY) PERFORMANCE OF GAS AND LIQUID BEARINGS IS IDENTICAL. (C) EFFECT OF LOAD CAPACITY (REF. 8)

The importance of the boundary conditions discussed above in the case of liquid bearings is not as important as gas bearings. However, the density must be assumed to vary with the gas pressure in the bearing, although the viscosity is normally assumed to be constant because of the low energy dissipation and therefore the low temperature rise. In the case of gas bearings, the dimensionless groups are obtained for plotting the design relations for the important operating variables. The primary independent variables in this case are the length-diameter ratio and the compressibility number, Λ, which replaces the Sommerfeld number or the bearing characteristic number in the cases of the journal bearing or the slider bearing respectively. Figure 23 shows a comparison of incompressible liquid and compressible gas in a slider bearing. The difference in pressure distribution, shown in Figure 23(a), indicates that the peak pressure moves towards the exit position of the pad relative to that of the incompressible bearing. In Figure 23(b), the friction load ratio and the center of pressure for compressible versus incompressible bearings, are plotted as a function of compressibility number. The pivot position for a tilting pad bearing is only slightly influenced by the compressibility with the pivot position moving towards the rear of the pad compared to that of the incompressible case. However, the load of the compressible bearing decreases substantially with increasing compressibility number. This influence of the load is shown not only in Figure 23(b) but also in Figure 23(c), which shows that as a compressibility number increases, the load being carried asymptotically approaches a value related to the atmospheric pressure, Pa, for the gas bearing. Figure 24 shows a typical design curve for gas journal bearings in which the two independent parameters are the length-to-diameter ratio and the compressibility number. In this case, the eccentricity ratio is the additional parameter in plotting the results of a load variable versus compressibility number. For each eccentricity ratio, the load being carried approaches an asymptotic value for large values of the compressibility parameter.

The following example gives some indication of the magnitude of variables involved in a gas journal bearing. Consider the following operating conditions for a journal bearing operating on air at one atmosphere pressure and 50° C;

$D = L = 100$ mm

$W = 10^3$ N

$C = 5$ μm

$Pa = 100$ kPa

$W = 100$ rad/s.

$\mu = 2 \times 10^{-5}$ Pas

The load variable is

$$\frac{W}{2RLPa} = \frac{10^3}{0.1 \times 0.1 \times 10^5} = 1.0$$

and the compressibility number is

$$\frac{\Lambda}{6} = \frac{\mu W}{Pa} \left(\frac{R}{C}\right)^2 = \frac{2 \times 10^{-5} \times 100}{10^5} \left(\frac{5 \times 10^{-2}}{5 \times 10^{-6}}\right)^2 = 2$$

From Figure 24, for these conditions, the eccentricity is found to be about 0.53. Therefore, the minimum film thickness is $h_o = \varepsilon \times C = 0.53 \times 5 = 2.6$ μm.

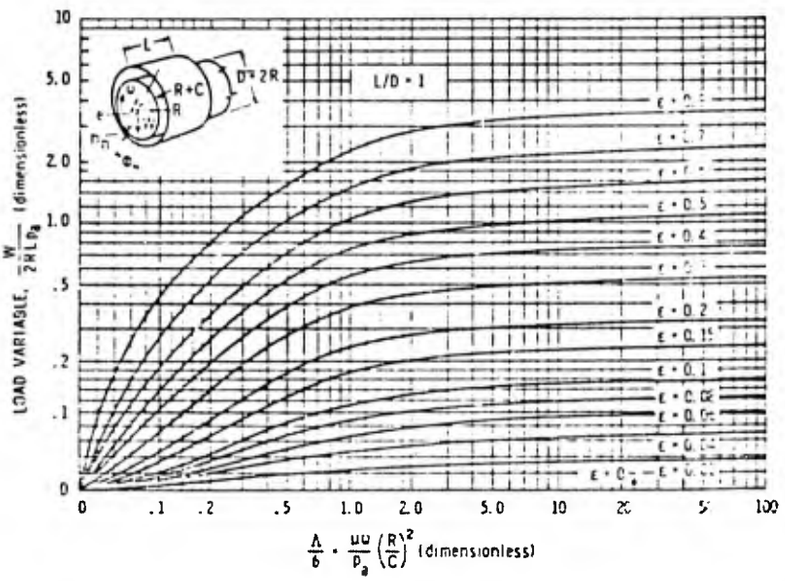


FIGURE 24 LOAD CAPACITY FOR FULL BEARING, L/D=1 (REF. 8)

7. ELASTOHYDRODYNAMIC LUBRICATION

Elastohydrodynamic lubrication occurs in concentrated contacts such as cams, gears, and rolling bearings. The

origin of the name comes from the fact that in these cases the load is concentrated over a very small region because the surfaces are nonconformal in nature. These concentrated loads result in sufficiently high local pressures that the surfaces deform elastically in amounts significant relative to the size of the generated hydrodynamic lubrication film thickness. In addition, because of the high pressure the lubricant properties, primarily the viscosity, will also change. Discussions of elastohydrodynamic lubrication can be found in the traditional lubrication texts such as Halling and Cameron[1, 7]. More thorough discussions of elastohydrodynamic lubrication will be found in the works of Dowson and Higginson and Hamrock and Dowson[12, 24].

The geometries associated with elastohydrodynamic lubrication are usually classified as either line contact or elliptical contacts. Sometimes elliptical contacts are also referred to as point contacts which are actually a special case of the elliptical contact. Line contact is that which occurs when two cylinders are loaded against one another as might be expected in the case of gears, cams, or cylindrical roller bearings. The fundamental geometry for line contact is shown in Figure 25[2]. In the highly loaded line contact case, the pressure distribution in the film is assumed to be Hertzian in nature even if the surfaces are separated by a thin film. The film-thickness distribution and the pressure distribution are also shown in Figure 25. The most commonly accepted expression for film thickness distribution in line contacts, is that attributed to Dowson and Higginson and shown in Equation (3) in which the dimensionless film thickness, h_{min}/R, is shown to be a power law function of a dimensionless material parameter, G, the dimensionless speed parameter, U, and a dimensionless load parameter, W.

$$\frac{h_{min}}{R} = 2.65 \frac{G^{0.54} U^{0.7}}{W^{0.13}}$$

where $G = \alpha E$ dimensionless material parameter

$U = \dfrac{\mu_o (u_1 + u_2)}{2ER}$ dimensionless speed parameter

$W = \dfrac{w}{ERL}$ dimensionless load parameter

α = pressure viscosity coefficient based on the piezo-viscous relation $\mu = \mu_o e^{\alpha p}$

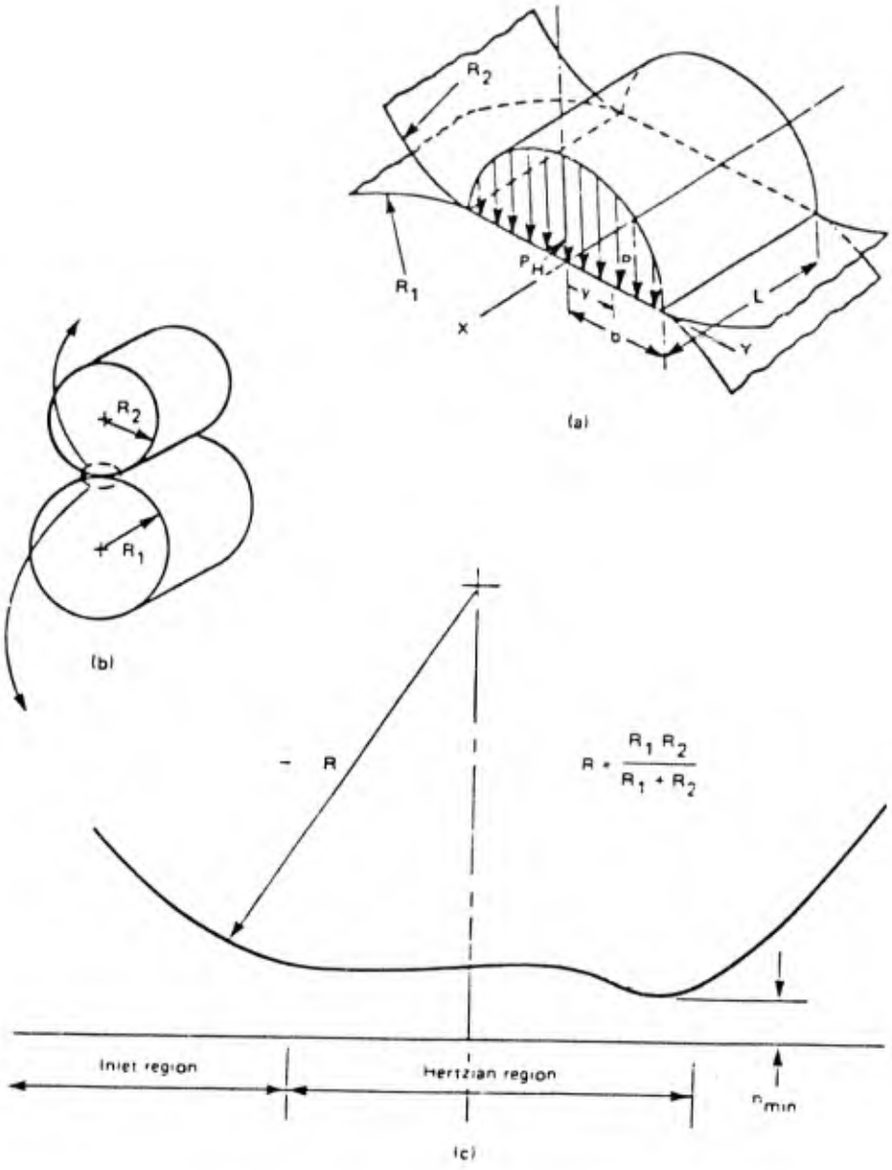


FIGURE 25 GEOMETRY OF A LINE CONTACT (REF. 2)

(A) HERTZIAN PRESSURE DISTRIBUTION
(B) CONTACTING CYLINDERS
(C) EHD FILM THICKNESS DISTRIBUTION

$$= \frac{1}{\mu}\frac{\partial \mu}{\partial P}\bigg|_{p=o} \quad (m^2/N) \quad (e.g., \alpha_{OT})$$

or

$$= \alpha^* = \left[\int_0^{p-\infty} \frac{\mu_o}{\mu(P)}\, dp\right]^{-1} \quad (m^2/N)$$

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

μ_o = lubricant viscosity (Pa.sec, N.sec/m) at inlet surface temperature

u_1, u_2 = surface velocities relative to contact region (m/sec)

$$R = \frac{R_1 R_2}{R_1 \pm R_2} \quad (m)$$

where the plus assumes external contact (both surfaces convex) and the minus internal contact (the surface with the larger radius of curvature is concave).

w = total load on the cylinder (N)

L = length of the cylinder (m)

One of the more interesting aspects of elastohydrodynamic film thickness is the rather small dependence of the film thickness on the load (e.g., to the -0.13 power). This expression shows that for a given system the film thickness is primarily a function of the speed parameter which is primarily dependent upon the viscosity and the average surface velocity.

In elastohydrodynamic lubrication, two physical lubricant properties are important in determining the film thickness. They are the viscosity in the inlet region and the pressure viscosity coefficient of the lubricant. Both the viscosity and the pressure viscosity coefficient are functions of temperature and are discussed in more detail in the following chapter entitled, "Lubricants."

The applicability of Equation (3) can be seen from an example of two steel rollers in external contact with a typical SAE 10 oil at about 40°C:

$L = 100$ mm

$R_1 = R_2 = 76$ mm

$w = 35.3$ kN

$E_1 = E_2 = 207$GPa

$\nu_1 = \nu_2 = 1/3$

$u_1 = u_2 = 12.7$ m/s

$\mu_o = 27.6$ mPas

$\alpha = 14.5$ GPa^{-1}

Therefore

$R = 38$ mm

$E = 233$ GPa

$G = 3379$

$U = 3.96 \times 10^{-11}$

$W = 3.987 \times 10^{-11}$

and

$$\frac{h_{min}}{R} = 1.58\ \mu m$$

At high surface speeds the lubricant viscous energy dissipation in the inlet region is sufficient to cause an increase in the temperature of the lubricant resulting in a decrease in the inlet viscosity. This phenomena is referred to as the thermal film thickness reduction and can be related to the energy dissipation in the inlet and the thermal conductivity of the lubricant. A thermal reduction factor, which is the ratio of the film thickness resulting when energy dissipation occurs in the inlet, divided by the isothermal film thickness as a function of a thermal reduction factor is attributed to Cheng and shown in Figure 26[2].

The thermal reduction factor Q_m is defined as follows:

$$Q_m = \frac{\mu_o u^2 \delta}{k_f} \text{ (dimensionless)}$$

where

μ_o = lubricant vicosity at atmospheric pressure and the inlet surface temperature $(\frac{N.sec}{m^2})$

$\mu = \frac{\mu_1 + \mu_2}{2}$ the average surface velocity (m/sec)

δ = lubricant viscosity temperature collection $\frac{1}{\mu}\frac{\Delta\mu}{\Delta T}(C^{-1})$

k_f = lubricant thermal conductivity $(\frac{N}{sec\ K}$ which is $\frac{W\ m}{m^2 K})$

In the above example, with a typical SAE 10 mineral oil for which the thermal conductivity is about 0.13 W/mK and the viscosity-temperature coefficient is about 0.045 C^{-1}, the Q_m = 1.6 and the thermal reduction is ϕ_T = 0.76. Therefore, the thermally corrected film thickness would be $h_{oT} = h\phi_T$ = 1.2 μm.

Concentrated contacts are not always elastohydrodynamic in the sense that both the elasticity and the viscosity variation of the lubricant with pressure is important. In the case of line contacts, there are three other possible regimes of operation. They are: 1) the case where neither elastic deformation nor viscosity change are important, which is referred to as the rigid isoviscous case, (R-I); 2) where the surfaces can be assumed to be rigid and the viscosity varies with pressure (R-V); and 3) where the elasticity of the surfaces results in significant deformation of the surface but the pressures are low enough that the viscosity can be treated as an isoviscous material (E-I). The equations governing the film thickness in these different regimes are quite different in their nature. The regimes can be divided in terms of two dimensionless variables, g_1 and g_2 which are shown in Figure 27[2]. The governing film thickness equations for the different regimes are presented as follows:

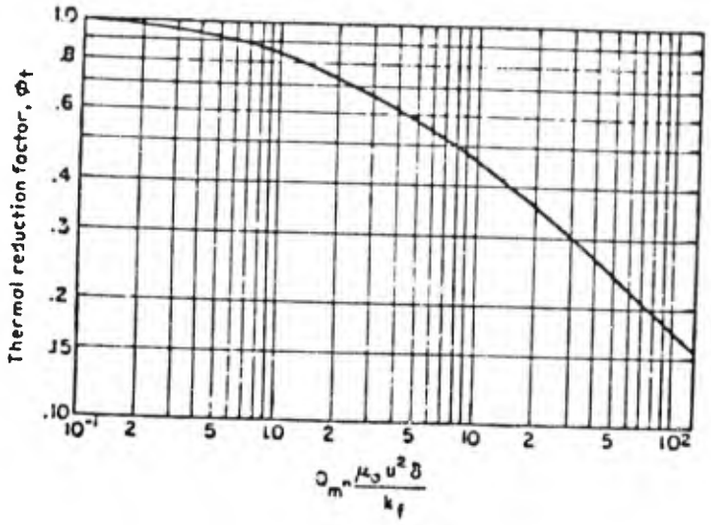


FIGURE 26 THERMAL REDUCTION FACTOR (REF. 2)

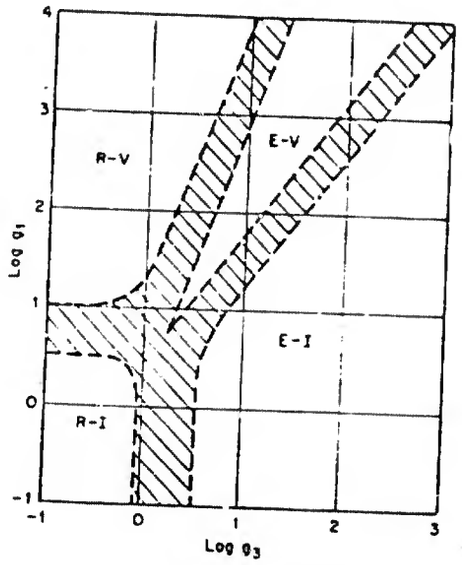


FIGURE 27 REGIMES OF LUBRICATION FOR LINE CONTACTS (REF. 2)

E-V: the elastic solid-variable viscosity lubricant regime which is assumed for the film thickness Equation (3). In terms of the variables used here Equation (3) becomes

$$h' = 2.65\, g_1^{0.54} g_3^{0.06}$$

E-I: elastic solid-isoviscous lubricant regime where the film thickness equation is

$$h' = 3.0\, g_3^{0}$$

R-V: rigid solid-variable lubricant regime where the film thickness equation is

$$h' = 1.66\, g_1^{2/3}$$

R-I: rigid solid-isoviscous lubricant regime where the film thickness equation is

$$h' = 4.9$$

In Figure 27 and the above film thickness equations, the following definitions have been used,

$$h' \equiv \frac{2h_{min}\, w}{\mu_o R(u_1 + u_2)} \quad \text{dimensionless minimum film thickness}$$

$$g_1 \equiv \left[\frac{2\alpha^2 w^3}{\mu_o R^2 (u_1 + u_2)}\right]^{1/2}$$

$$g_3 \equiv \left[\frac{2w^2}{\mu_o E R (u_1 + u_2)}\right]^{1/2}$$

$$W \equiv \frac{\text{load}}{\text{length of contact}}$$

Elliptical point contacts result when two curve surfaces with curvatures in both directions come in contact, such as a ball bearing or many gear configurations. The generalized geometry for elliptical contact is shown in Figure 28. The governing equations appropriate for the elliptical contact attributed to Hamrock and Dowson are presented as follows[24];

$$H_{c.F} = 2.69\ U^{0.67} G^{0.53} W^{-0.067} (1 - 0.61e^{-0.73k}) \qquad (4a)$$

$$H_{min.F} = 3.63\ U^{0.68} G^{0.49} W^{-0.073} (1 - e^{-0.68k}) \qquad (4b)$$

where

$H_{c.F}$ = $h_{c.F}/R_x$ (dimensionless)

$h_{c.F}$ = central film thickness for flooded contacts

$H_{min.F}$ = $h_{min.F}/R_x$ (dimensionless)

$h_{min.F}$ = minimum film thickness for flooded contacts

$k = \frac{a}{b}$ = elliptically parameter

$R_x = R_{x1}R_{x2}/(R_{x1} + R_{x2})$ (m)

$R_y = R_{y1}R_{y2}/(R_{y1} + R_{y2})$ (m)

a = semi-axis of contact ellipse perpendicular to the motion (m)

b = semi-axis of contact ellipse parallel to the motion (m)

$W = \frac{w}{ER_x^2}$ (dimensionless)

w = the total load (N)

$U = \frac{\mu_o(u_1 + u_2)}{2\ ER_x}$ (dimensionless)

$G = \alpha E$ (dimensionless)

$\frac{1}{\bar{E}} = \frac{1}{2}\left(\frac{1 - \nu_1^2}{E_1} + \frac{1 - \nu_2^2}{E_2}\right)$ (m^2/N)

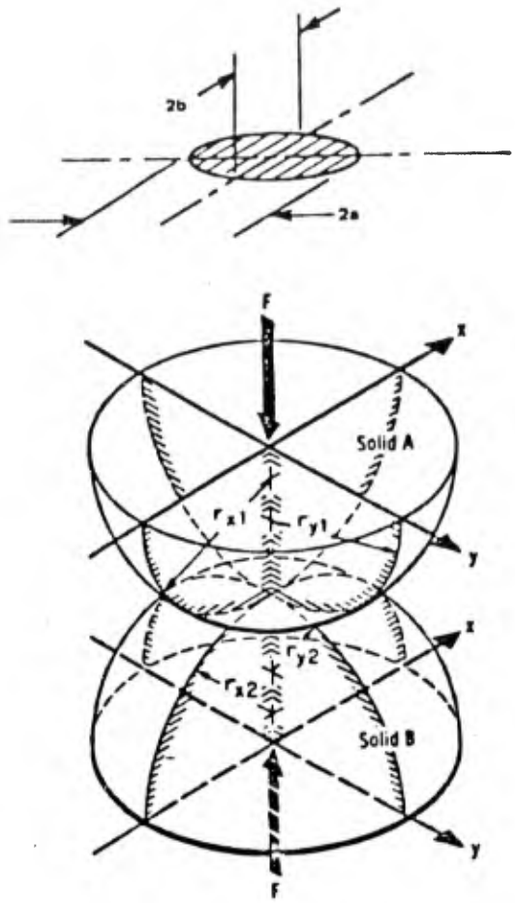


FIGURE 28 GEOMETRY OF CONTACTING ELASTIC SOLIDS (REF. 24)

E_1, E_2, ν_1, ν_2 = moduli of elasticity and Poission's ratio of body 1 and 2. The surface velocities, u_1 and u_2, lubricant inlet viscosity, μ_o, and pressure viscosity coefficient, α, are defined for line contacts.

The radii of curvature are assumed positive if the line surface is convex and negative if concave. The radius of curvature for a flat surface is infinite.

For example, consider the following elliptical contact. The principal radii of an EHD point contact between a 25.4 mm diameter steel roller with a 51 mm crown radius (surface 2) and a 76.2 mm diameter steel cylinder with no crown radius (surface 1) are

R_{x1} = 38 mm

R_{x2} = 12 mm

R_{y1} = ∞m

R_{y2} = 51 mm

The load and speeds are

w = 445 N

u_1 = 12.7 m/sec

u_2 = 12.7 m/sec

μ_o = 27.6 mPa.sec

α = 14.5 GPa^{-1}

$$R_x = \frac{R_{x1}\,R_{x2}}{R_{x1}+R_{x2}} = \frac{38 \times 12.7}{38 + 12.7} = 9.52 \text{ mm}$$

$$R_y = \frac{R_{y1}\,R_{y2}}{R_{y1}+R_{y2}} = 51 \text{ mm}$$

E = 233 GPa

G = αE = 3373

$$U = \frac{\mu_o(u_1 + u_2)}{2 \times E \times R_x} = \frac{27.6\times10^{-3}(12.7+12.7)}{2\times233\times10^{9}\times9.52\times10^{-3}} = 1.58 \times 10^{-10}$$

$$W = \frac{w}{ER_x^2} = \frac{445}{233\times10^{9}\times(9.52\times10^{-3})^2} = 2.11 \times 10^{-5}$$

$$k = \frac{a}{b} = 1.03\ (R_y/R_x)^{0.64}$$ [See section 3.3, Equation 16 of Reference (2)]

$$= 1.03\ (51 \times 9.52)^{0.64}$$

$$= 3.015$$

Then Equation (4a) gives

$$H_{c.F} = 2.69 \times (1.58 \times 10^{-10})^{0.67}\ (3373)^{0.53}$$

$$(2.11 \times 10^{-5})^{-0.067}$$

$$\times\ (1 - 0.61 \times e^{-0.73 \times 3.015})$$

$$= 1.037 \times 10^{-4}$$

$$h_{c.F} = 9.52 \times 10^{-3} \times 10.7 \times 10^{-4} = 0.98\ \mu m$$

and Equation (4b) gives

$$H_{min.f} = 3.63 \times (1.58 \times 10^{-10})^{0.68}\ (3373)^{0.49}$$

$$(2.11 \times 10^{-5})^{-0.073}$$

$$\times\ (1 - e^{-0.68 \times 3.015})$$

$$= 8.04 \times 10^{-5}$$

$$h_{min.F} = 8.04 \times 10^{-5} \times 9.52 \times 10^{-3}$$

$$= 0.76\ \mu m$$

As in the case of line contacts, there are different regimes of operation for the elliptical point contacts. These regimes have been developed by Hamrock and Dowson and are shown for the case of an ellipticity ratio of one in Figure 29[24]. The four different regimes involved are the same as those in the case of the line contact, however, in the case of point contact there is the added variable of the ellipticity ratio of the contact. Additional plots for other ellipticity ratios can be found in Reference 24. The appropriate

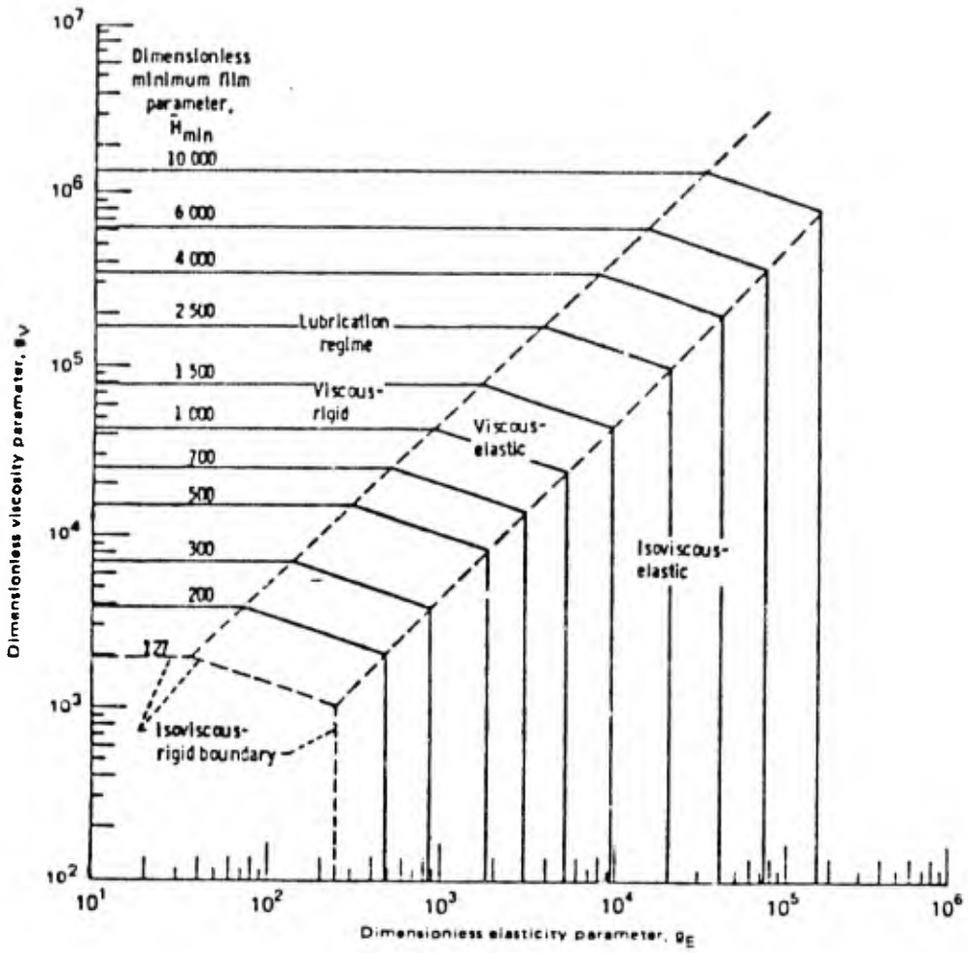


FIGURE 29 MAP OF LUBRICATION REGIMES WITH DIMENSIONLESS-FILM-PARAMETER CONTOURS ON LOG-LOG GRID OF DIMENSIONLESS VISCOSITY AND ELASTICITY PARAMETERS FOR ELLIPTICITY PARAMETERS OF 1 (k = 1) (REF. 24)

equations for the different regimes of the elliptic contact are given below:

IR: (Isoviscous-rigid regime)

$$\{H_{min}\}_{IR} = 128\beta\,\Phi^2\,[0.13\,\tan^{-1}(\tfrac{\beta}{2}) + 1.68]^2$$

where

$$\beta = \frac{R_y}{R_x} \simeq (\frac{k}{1.03})^{1/0.64}$$

$$\Phi = (1 + \frac{2}{3\beta})^{-1}$$

VR (Viscous-rigid regime):

$$\{H_{min}\}_{VR} = 1.66\,g_V^{2/3}\,(1 - e^{-0.68k})$$

IE (Isoviscous-elastic regime):

$$\{H_{min}\}_{IE} = 8.70\,g_E^{0/67}\,(1 - 0.85\,e^{-0.31k})$$

VE (Viscous-elastic regime):

$$\{H_{min}\}_{VE} = 3.45\,g_V^{0.49}\,g_E^{0.17}\,(1 - e^{-0.68k})$$

where all variables are defined as previously in this section except

$$\{\bar{H}_{min}\} = \bar{H}_{min}\,(-)_{\bar{U}}^{W} = \frac{4w^2 h_{min}}{(u_1 + u_2)^2\,\mu_o^2\,R_x^3}$$

where

$$g_v = \frac{GW^3}{U^2}$$

$$g_e = \frac{W^{8/3}}{U^2}$$

The above discussed methods of determining film thickness in elastohydrodynamic lubrication are not the only ones available in the literature. A very useful and simplified approach for determining the elastohydrodynamic film thickness in standard machine components has been developed and presented in the Mobil EHL Design Guide[25]. In this design guide, component materials are assumed to be steel. Certain standard geometries and surface roughnesses are assumed and the operating parameters are given in terms of quantities more familiar to the designers, such as rotational speed of the shaft and load on the shaft. The lubricant properties are lumped into a lubricant parameter which is the product of the viscosity and the pressure viscosity coefficient. When compared with the above equation for elastohydrodynamic lubrication, it is noted that these two parameters are in those equations to two different powers. However, because of the relatively limited range of variation of the pressure viscosity coefficient and the practically small difference between the power on the viscosity and the power on the pressure viscosity coefficient the lumping of these two parameters seems to be quite satisfactory for determining design information.

8. MIXED FILM LUBRICATION

Mixed film lubrication exists when the film thickness developed by the above hydrodynamic mechanisms is inadequate to separate the two surfaces being lubricated. In this case, some surface asperities interact with resulting boundary lubrication at those interactions. Wear, surface distress, and an increase in the energy dissipation compared to the complete hydrodynamic film can be expected. A clear understanding of the nature of mixed film lubrication is not available. All component run-in or wear-in is under mixed film lubrication. A better understanding of run-in is desirable because the difference between run-in and failure in new components is often unpredictable and certainly not well understood. Mixed film lubrication is an area of current active interest and one in which there is a great deal of descriptive information in the literature. A recent symposium held by ASME gives and indication of the current state-of-the-art in this field[26].

9. CONCLUSION

Lubrication is a complex and broad subject difficult to summarize in one brief review chapter. The various modes of lubrication have been introduced and the reader is directed to the appropriate literature for further study.

10. REFERENCES

1. Halling, J., Ed., "Principles of Tribology," MacMillan Press, Ltd., London (1975).

2. Peterson, M.B., and Winer, W.O., Eds., "Wear Control Handbook," ASME, New York (1980).

3. Bowden, F. P. and Tabor, D., "The Friction and Lubrication of Solids," Oxford University Press, New York (1954), and "Friction and Lubrication of Solids, II," Oxford University Press (1964).

4. Czichos, H., "Tribology," Elsevier Scientific Publishing Company, New York (1978).

5. Bair, S., and Winer, W. O., "Regimes of Traction in Concentrated Contact Lubrication," ASME Paper No. 81-LUB-16 (to be published in Trans. ASME, Journal of Lubrication Technology, 1982).

6. Braithwaite, E. R., "Solid Lubricants and Surfaces," Pergamon Press, New York (1964).

7. Cameron, A., "Basic Lubrication Theory," Longman, London (1966), and "Principles of Lubrication," J. Wiley and Sons, New York (1966).

8. O'Connor, J. J. and Boyd, J., "Standard Handbook of Lubrication Engineering," McGraw-Hill, New York (1968).

9. Ling, F. F., Klaus, E.E. and Fein, R. S., Ed., "Boundary Lubrication: An Appraisal of World Literature," ASME (1969).

10. Sakurai, T., "Role of Chemistry in the Lubrication of Concentrated Contacts," Trans. ASME, Journal of Lubrication Technology, 103, No. 4 (1981) 473 - 485.

11. Trumpler, P.R., "Design of Film Bearings," MacMillan, New York (1966).

12. Dowson, D., and Higginson, G. R., "Elasto-hydrodynamic Lubrication," Pergamon Press, New York (1966).

13. Gross, W. A., et. al, "Fluid Film Lubrication," John Wiley and Sons, New York (1980).

14. Constantinescu, V. N., "Gas Lubrication," ASME, New York (1969).

15. Pinkus, O., and Sternlicht, B., "Theory of Hydrodynamic Lubrication," McGraw-Hill, Inc., New York (1961).

16. Tipei, N., "Theory of Lubrication with Application to Liquid-Gas-Film Lubrication," Sanford University Press (1962).

17. Rohde, S. M., Maday, C. J., and Allaire, P. E., Eds., "Fundamentals of the Design of Fluid Film Bearings," American Society of Mechanical Engineers, New York City (1979).

18. Rohde, S. M., Wilcock, D. F., and Cheng, H. S., "Energy Conservation through Fluid Film Lubrication Technology: Frontiers in Research and Design," American Society of Mechanical Engineers, New York City (1979).

19. Szeri, A. Z., "Tribology," Hemisphere Publishing Company, New York (1980).

20. Booker, J. F., "Design of Dynamically Loaded Journal Bearings," (in Reference [17]) (1979).

21. Goenka, P.K., and Booker, J. R., "Effect of Surface Ellipticity on Dynamically Loaded Cylindrical Bearings and Joints," ASME Paper No. 81-LUB-1 (to be published in Trans. ASME, Journal of Lubrication Technology, 1982).

22. Rippel, H. C., "Cast Bronze Hydrostatic Bearings Design Manual," Cast Bronze Bearing Institute, Inc. (1964).

23. Grassam, N. S., and Powell, J. W., Eds., "Gas Lubricated Bearings," Butterworths, London (1964).

24. Hamrock, B. J., and Dowson, D., "Ball Bearing Lubrication: The Elastohydrodynamics of Elliptical Contacts," J. Wiley and Sons, (1981).

25. Mobil EHL Design Guide, Mobil Oil Co.

26. Rohde, S. M., and Cheng, H. S., "Surface Roughness Effects in Hydrodynamics and Mixed Lubrication," American Society of Mechanical Engineers, New York (1980).

11. ACKNOWLEDGEMENTS

Figure Number	Source (Chapter Reference)/Citation
1, 2(B)	Ref. 1, with permission of the Editor, J. Halling
3	Ref. 3, with permission of D. Tabor, F.R.S.
4, 7	Ref. 4, with permission of H. Czichos
9, 10	Ref. 1, with permission of the Editor, J. Halling
11(B), 12, 16, 17	Ref. 8, Standard Handbook of Lubrication Engineering, O'Connor and Boyd, Ed., 1968 McGraw-Hill Book Co., with permission
18, 19	Ref. 1, with permission of the Editor, J. Halling
20	With permission of the American Society of Lubrication Engineers, publishers of Lubrication Engineering and ASLE Transactions
21	Ref. 1, with permission of the Editor, J. Halling
22, 23, 24	Ref. 8, Standard Handbook of Lubrication Engineering, O'Connor and Boyd, Ed., 1968 McGraw-Hill Book Co., with permission