NASA SP-181

INTERDISCIPLINARY APPROACH TO FRICTION AND WEAR

A symposium held at
SOUTHWEST
RESEARCH INSTITUTE
San Antonio, Texas
November 28–30, 1967

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

NASA SP-181

INTERDISCIPLINARY APPROACH TO FRICTION AND WEAR

Proceedings of a NASA-sponsored symposium held
November 28–30, 1967, in San Antonio, Texas

Edited by P. M. Ku
Southwest Research Institute

Scientific and Technical Information Division
OFFICE OF TECHNOLOGY UTILIZATION 1968
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
Washington, D.C.

For sale by the Superintendent of Documents
U.S. Government Printing Office, Washington, D.C. 20402
Price $2.25 (Paper cover)
Library of Congress Catalog Card Number 68–62230

Foreword

THIS VOLUME, and the symposium whose proceedings it records, are outgrowths of a recommendation of the NASA Research Advisory Committee on Materials (now renamed the Research and Technology Advisory Subcommittee on Materials). In October 1965 the committee reviewed the research programs of NASA having to do with the technology of bearings and lubrication. It concluded that these programs added up to a relevant and well-rounded effort. However, it felt that additional strength would develop if the talents of scientific workers in other disciplines could be directed toward this technology. It pictured a three-cornered matrix in which the corners were inhabited by (1) engineers concerned with the design and operation of bearings and related devices, (2) scientists conducting applied research dealing with friction, wear, and lubrication, and (3) the less applied scientists, mainly the physicists and chemists, who were performing research on the fundamental nature of surfaces and the interactions between them.

The committee made no claims of having originated these ideas; the same marriage of disciplines to deal with bearing technology is implied in the term "tribology" which has gained rather wide currency in England and elsewhere. However, NASA was particularly fortunate in having all of these classes of individuals on its existing research teams. The problem was not one of introducing new members into the teams but rather one of communication, with the objective of making the work of one group of individuals relevant to the problems of another.

The immediate question was: who shall communicate with whom? It seemed apparent that the number of disciplines is too great for simultaneous communication in all directions. It was decided that a trial symposium could be sponsored by NASA with the objective of bringing together the second and third of the groups defined above. It was obvious that this objective could best be achieved if attendance was limited in number, was restricted to persons who were invited upon the recommendation of a Steering Committee, and was made contingent upon active participation. A limited number of individuals in the remaining category—the users and design engineers—were invited to prepare themselves for possible later symposia which would require such active participation from them as well.

The symposium was received with enthusiasm and marked by a lively response. However, the success of the effort it represents cannot be appraised fully at this time. It will become more evident if gradual but meaningful changes in the scope and direction of research are observed, so that one man's new knowledge relates to another man's problems. This is the true meaning of interdisciplinary communication.

JOSEPH MALTZ
Executive Secretary
Subcommittee for Materials

Preface

THIS VOLUME records the proceedings of the first in a series of interdisciplinary symposia in the field of lubrication, conducted under the sponsorship of the National Aeronautics and Space Administration (NASA). The symposium, entitled Interdisciplinary Approach to Friction and Wear, was concerned primarily with sliding friction and wear under unlubricated and boundary-lubrication conditions. The symposium was held in San Antonio, Texas, on November 28, 29, and 30, 1967. The physical arrangements were made by Southwest Research Institute.

Lubrication is a highly complex and interdisciplinary subject. Satisfactory solutions to many advanced lubrication problems require the concerted effort of lubrication research engineers, design and development engineers, and specialists in many basic scientific disciplines. Yet anyone concerned with such problems cannot fail to be disturbed by the lack of effective dialogue, not only among those within the broad spectrum of the lubrication profession itself, but particularly between the lubrication profession and the related scientific and engineering professions.

The express goal of the NASA lubrication symposia is to foster interdisciplinary dialogue among the various professional groups active or vitally interested in the subject of lubrication and its theoretical and practical implications. In order to satisfy this requirement and keep the discussions within manageable bounds, it was decided to deal with the broad subject of lubrication in several meetings. It was further decided to begin the symposium series with an indepth and interdisciplinary treatment of sliding friction and wear—topics fundamental to the subject of lubrication. Later, depending upon the results achieved, meetings in the same or other areas would be scheduled.*

The symposium on friction and wear was opened with an introductory address presented on behalf of NASA by R. R. Nash, Chief, Materials Science Branch, Office of Advanced Research and Tech-

---

*At this writing, two followup meetings have been scheduled: an Interdisciplinary Workshop on Friction and Wear to be held in Cleveland, O., in Nov. 1968, and a symposium on Interdisciplinary Approach to the Lubrication of Concentrated Contacts to be held in Troy, N.Y., in July 1969.

nology. Following this address, six specific topics were considered in detail:

1. Structure of surfaces and their interactions
2. Topography of solid surfaces
3. Surface interactions in sliding
4. Friction and adhesion
5. Wear
6. Boundary lubrication.

In each case, the topic was introduced by an invited lecture, which was followed by invited discussions and then general discussions. The purpose of each invited lecture was to summarize briefly and comment critically on the subject matter, to delineate the knowns and unknowns, and to provide a starting point for provocative discussions. The invited discussions were meant to complement the lecture and present a balanced viewpoint. It was realized, however, that the value of the symposium depended to a large extent on the contributions of all of the participants; therefore, every effort was made to facilitate and encourage general discussions.

Following deliberations on the six topics, the area of friction and wear was studied in broader perspective from two points of view:

(1) Critical appraisal and research opportunities—the lubrication research viewpoint

(2) Critical appraisal and research opportunities—the materials research viewpoint.

In each case, the study was again introduced by an invited lecture, followed by invited discussions and general discussions.

In an effort to ensure that the deliberations be truly productive, participation was by invitation only. Of the 90 persons who participated in this symposium, about 50 percent were lubrication research engineers, 25 percent design and development engineers, and 25 percent basic scientists.

All participants were registered and briefed almost 3 months in advance on the scope and philosophy of the symposium. A Proceedings Preprint, containing the first six invited lectures, was mailed to each participant about 1 month before the meeting. In addition, a Proceedings Preprint Supplement, containing the last two invited lectures and the invited discussions on the first six lectures, was issued immediately preceding the meeting so that each participant would have an opportunity to become familiar with most of the programmed presentations. Lack of time prevented the preprinting of invited discussions on the last two lectures; however, by the time these lectures and their invited discussions were presented, the participants had already become familiar with the problems being considered so that the general discussions did not suffer appreciably. On the contrary, the general discussions throughout the 3-day meeting were lively and productive. In fact, some of the participants felt that even the rather large amount of time allotted was not sufficient for the discussions to run their full course.

In presenting this volume, I wish to acknowledge the guidance and support given by the following NASA staff: G. C. Deutsch, Associate Director of Research (Materials), Office of Advanced Research and Technology; J. Maltz, Office of Advanced Research and Technology; and E. E. Bisson, Assistant Chief, Fluid System Components Division, Lewis Research Center. Special acknowledgment is due all participants, without whose contributions the symposium would not have been successful. The members of the Steering Committee provided timely assistance and advice throughout the preparation of the symposium, including the technical review of the invited lectures. However, as chairman of the Steering Committee and editor of this volume, I am responsible for any imperfections that may appear in it.

P. M. KU
San Antonio, Texas
June 17, 1968

Contents

Introductory Address

R. R. NASH
National Aeronautics and Space Administration
Washington, D.C.

On behalf of NASA and the Steering Committee for this symposium, I wish to extend a cordial welcome to this unique and important event. The symposium is unique because, so far as I know, it is the first of its kind to address in depth the subject of friction and wear; it is important because we expect real accomplishment.

In making these introductory remarks, it is my intention to think with you informally about the job at hand and particularly about the symposium program designed to make the interdisciplinary concept of an approach to friction and wear come alive in the next 3 days.

The central point I wish to make in this brief introduction is to underscore the importance of the key words *interdisciplinary dialogue*. Perhaps I can best illustrate the meaning of the word "dialogue" by figure 1. We laugh, but I would ask in all seriousness that you carry the image of this all-too-frequent confrontation in your mind for the next 3 days.

In preparation for this presentation, I was reminded of a story told me recently by my son which may have significance to us at this time. In the sand trap of a golf course, two red ants were busy scurrying about amid a violent scene of flying sand, dust, and confusion. Finally, one ant said to the other, "What in the world will we do? Things are getting rough." The other ant replied, "From the evidence we have at this time, it seems that the best idea would be to get on the ball."

What is the ball for us to consider? With the help of figure 2, let us examine the concept of an interdisciplinary approach to the study of friction and wear. If appropriate names of disciplines are placed at the corners of the tetrahedron, it is apparent that there is a definite relationship between all of the disciplines involved in this symposium dialogue. In fact, the picture is incomplete with any discipline omitted. In my view, everyone here sits at some position within the volume of the tetrahedron or within its environment. Where do you sit?

What are we trying to do in this symposium? I believe the central problem is to bridge the gap between basic science and engineering on the one hand and engineering design on the other in connection with friction, wear, and lubrication. In order to build this bridge, cooperation is necessary from all concerned on both sides of the gap. The principal means for building this bridge is meaningful dialogue between the disciplines involved. The vehicle for this dialogue is responsiveness of individuals; and responsiveness of individuals, ultimately, is a matter of open-minded consideration of the contributions of all individuals involved.

The NASA Materials Research Program on this subject is attempt-

"I am NOT giving you any lip, Pop! I'm just trying to conduct a meaningful dialogue!"

FIGURE 1.—Conducting meaningful dialogue (courtesy of Hall Syndicate).

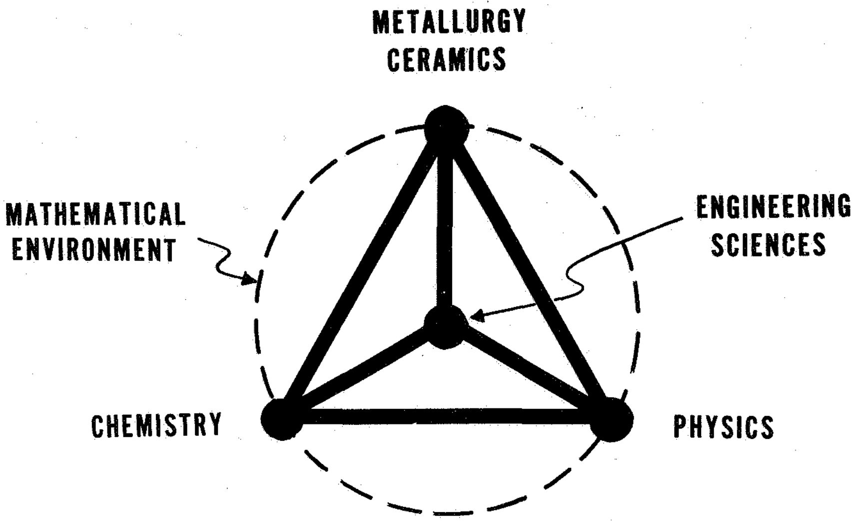


FIGURE 2.—Concept of an interdisciplinary approach to friction and wear.

ing to help build this bridge. Three examples from the NASA Lewis Research Center will illustrate some of our research to study relevant phenomena from the atomic scale of dimensions to the macroscopic. Although the next three figures will present only the essential idea of these studies, specific details of each research topic can be discussed directly with the investigators, who are in the audience.

In figure 3, the defect surface state of magnesium oxide is revealed by an electron spin resonance technique. In this research, the physical nature of an active surface site for adsorption of gases on surfaces was demonstrated for the first time.

In figure 4, the effects of atomic arrangement on friction properties of metals existing in the hexagonal close-packed crystal structure is demonstrated. Although complete understanding of this subject is not at hand, the evidence suggests strongly that the bulk structure as well as surface phenomena are important considerations.

Certain aspects of the effects of environment in bearings and lubrication research are displayed in figure 5.

In all three of the illustrations selected, the main point is that each investigation contributes to greater understanding of the total problem of friction and wear. Interdisciplinary dialogue between the investigators would serve the very useful function of informing, stimulating, and challenging additional research with a high probability of even more successful results.

Let us return briefly to our concept of dialogue, figure 1. Perhaps this illustration is more relevant to the work of this symposium than a casual glance might indicate. If you will allow me to place labels on the two characters, I suggest that the boy represents individuals working in basic physical and engineering sciences. The man represents design and applications engineers concerned with friction, wear, and lubrication. Picture the rage and frustration of each when they do not communicate effectively. Perhaps this situation exists in the subject to be discussed during the next 3 days.

At this point, enter the approach selected for this symposium. The Steering Committee has deliberately mingled basic aspects of surface phenomena with engineering aspects of friction, adhesion, wear, and lubrication. The program places heavy emphasis upon defining the current state of knowledge in these disciplines and also upon identifying the kinds of research needed to extend this knowledge.

In summary, I believe that the productivity of this symposium will begin to bridge the gap between science and technology in friction and wear if we can conduct meaningful dialogue. Every man here is qualified to make a contribution. I challenge you to enter this dialogue in a cooperative spirit and try to hear what other individuals are saying. Try it and see if you can possibly leave empty-handed! Finally, I am confident that I express the intention of NASA and the Steering Committee when I say that such participation is both a challenge and a responsibility.

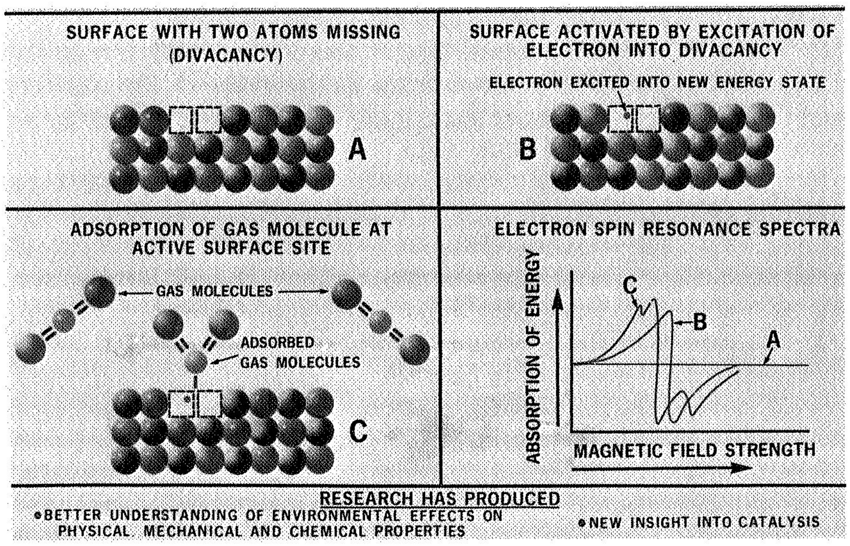


FIGURE 3.—Role of atomic and electronic imperfections in gas adsorption on crystal surfaces. NASA Lewis Research Center.

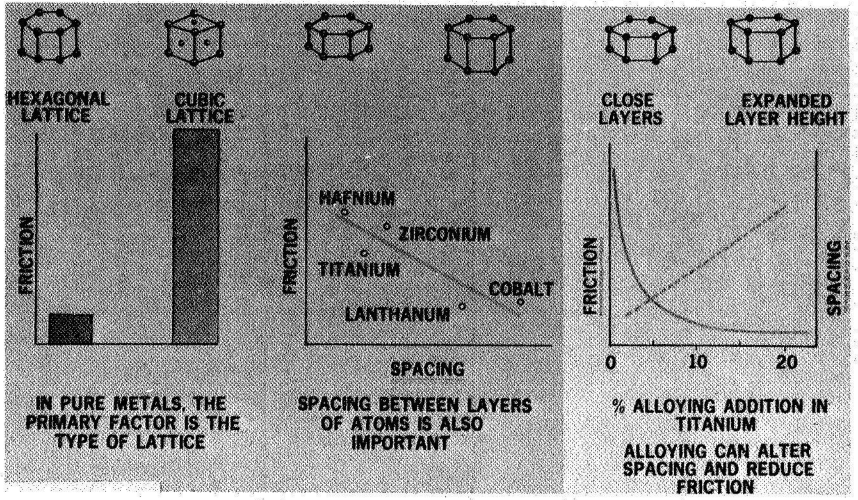


Figure 4.—Effect of atomic structure on frictional properties of metals with hexagonal crystal symmetry. NASA Lewis Research Center.

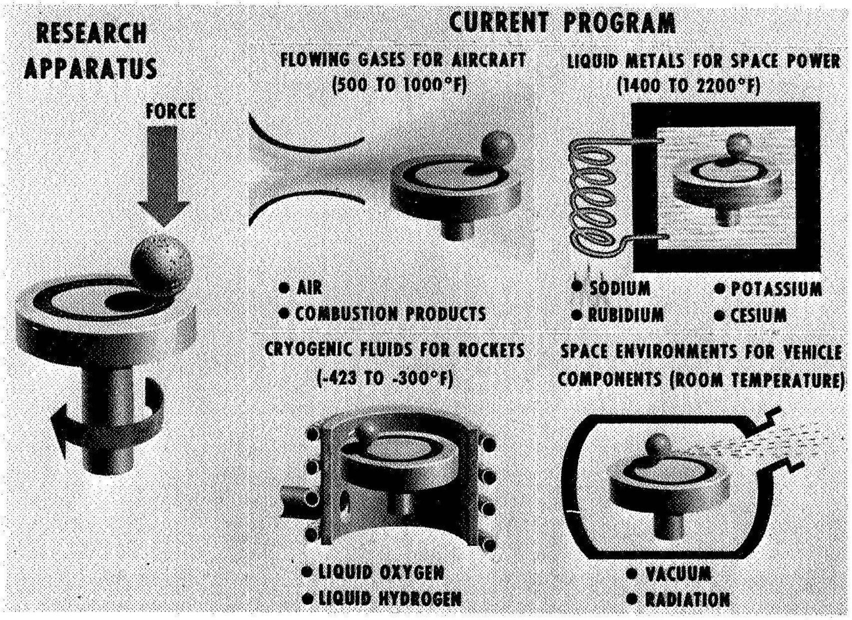


Figure 5.—Investigations of environment effects on friction and wear. NASA Lewis Research Center.

Structure of Surfaces and Their Interactions

H. C. GATOS
Massachusetts Institute of Technology
Cambridge, Massachusetts

The salient characteristics of crystalline solid surfaces are discussed with primary emphasis on their structural aspects. Direct experimental approaches to the study of the structure of surfaces and their limitations are considered together with some theoretical models. The interactions of solid surfaces with gases, liquids, and solids are outlined and assessed in terms of their fundamental and technological significance. A number of surface science areas in need of improved understanding are pointed out in the light of the ever increasing technological importance of surface phenomena.

The structure and behavior of solid surfaces in terms of scientific interest and technological importance continue to occupy a distinct place in modern science and engineering. In fact, surface or interface phenomena bear directly on nearly all facets of basic and applied technology. Indirectly, surface phenomena are strikingly pertinent, even to the science and technology of the solid state proper, because it is through the surface that one must generally communicate with the solid.

In the study of solids, a number of basic quantitative tools have been developed, to a large extent as a result of the extreme periodic order and symmetry associated with crystallinity. These tools are of a theoretical as well as of an experimental nature. Thus, theoretical concepts based on first principles, combined with techniques such as X-ray or neutron diffraction, have brought our understanding of crystalline solids to a very high level indeed. Subsequently, precise magnetoelectrical measurements have brought about further refinements concerning atomic interactions and the energetics of the solid state in general.

The understanding of solids cannot be extended directly to solid surfaces. Surfaces of crystalline solids represent an abrupt termination of the periodicity of the crystalline lattice. They are essentially giant lattice defects and thus cannot be directly submitted to the theoretical

323-472 O—69——2

treatments of crystalline periodicity. Appropriate theoretical treatments must take into consideration the implications of such an abrupt termination of lattice (primarily the relevant energetics). Consequently, they become immensely complex and, by necessity, very approximate.

Experimental approaches to the study of surfaces must overcome two main difficulties of great magnitude inherent in solid surfaces. One of these difficulties is related to the fact that the surface atoms are chemically unsaturated and thus have an intrinsic tendency to react, physically or chemically, with their environment. Consequently, all surfaces are contaminated with foreign atoms. The experimental characterization of chemically "clean" surfaces was virtually impossible until suitable vacuum and related technologies were developed in the 1950's and became moderately available in the 1960's.

The second main difficulty in the experimental study of surfaces resides in the fact that surfaces are one to several atom layers thick. Techniques for quantitative studies on an atomic scale are limited in number; they have rather limited ranges of applicability and present great complexities. The major direct techniques of this type are the electron-, field-, and ion-microscopy and the low energy electron diffraction (LEED).

In considering the structure of solid surfaces, one must be concerned with surface science and technology as a whole, i.e., the structure of surfaces (crystallinity and energetics); their interaction (predominantly but not exclusively chemical) with gases, liquids, or solids; and their physical properties including the changes of such properties resulting from the interactions of the surfaces with their environment.

For the fundamental understanding of surfaces, the experimental information must be obtained on "clean" (frequently called "virgin") surfaces. Achieving clean surfaces is possible only in relatively small numbers of cases, and experimenting with them is a truly complex matter. Nevertheless, data obtained with such surfaces (pertaining to their structures as well as to their reactivity) constitute an essential ingredient in the development of fundamental concepts and theoretical systems for understanding surface behavior.

On the other hand, the greater part of surface phenomena must be dealt with, at least experimentally, in the light of "real" surfaces, i.e., surfaces prepared and maintained in ordinary ambients rather than in ultrahigh vacuum. Unavoidably such surfaces are contaminated and introduce a number of chemical and structural uncertainties. If properly employed, however, they can be well suited for a wide spectrum of significant experimental information. It is important to emphasize that the principal prerequisites for the successful use of real surfaces is that they be prepared reproducibly and that they be sensitive to changes of the parameters or conditions which are being investigated.

Clean surfaces can be meaningfully characterized experimentally or theoretically only in the absence of an ambient, i.e., in ultrahigh vacuum. Accordingly, in considering the crystalline structure of these surfaces and the associated energetics, interactions with external phases are not pertinent. It is apparent that the usefulness of such surfaces is still quite limited.

Real surfaces must be considered in the light of their interactions with external gaseous, liquid, or solid phases. The kinetics and energetics of such interactions are the essential basis in exploring and understanding their structure and the broad spectrum of phenomena in which the surface atoms participate as well as of phenomena which take place on surfaces but without the ultimum participation of the surface atoms. Catalytic processes represent perhaps one example of the latter type of phenomenon.

Regarding the solid-gas interfaces, the various interactions, equilibria or pseudo-equilibria, can be conveniently treated under three general headings: physical adsorption, chemisorption, and oxidation or film formation.

In the case of the solid-liquid interfaces, the solid-electrolyte interfaces are of the broadest fundamental and technological significance. The science and electrochemistry and, of course, electrochemical technology are essentially based on such interfaces. Phenomena such as the electrical double layer, electrode equilibrium, and electrode kinetics, corrosion in general, and many others are based directly on solid-electrolyte interfaces. Although metals and their alloys are electrochemically the most pertinent among solids, perhaps all solids are electrochemically relevant to some extent.

The solid-liquid (nonelectrolyte) interfaces* are certainly not without fundamental or technological significance. Numerous organic and many inorganic liquids are not electrolytes. It is of interest to point out, however, that many of the phenomena pertaining to the interactions of such liquids with solids are special cases of phenomena prevalent in the solid-gas interfaces (notably physical adsorption and chemisorption) or in the solid-electrolyte interfaces (for example, electrical double layer phenomena).

Solid-solid interfaces again are the stage of numerous phenomena of great importance. Grain boundaries, p-n junctions, rectifying contacts, and solid surfaces in physical contact are among such solid-solid interfaces. Of particular technological relevance are the phenomena

---

*The solid-melt interfaces, important as they are regarding solidification and many other phenomena, are outside the scope of the present discussion.

associated with the motion of solid surfaces against one another while in physical contact. Obviously the phenomena of friction and wear fall directly into this category, and atomic interactions are of major consideration. Since the contacting surfaces are exposed to gases or liquids, here also, many aspects of the solid-gas and solid-liquid interfaces come into play. Other important facets of science and technology of solid-solid interfaces include adhesion and surface damage.

Nucleation phenomena represent another important aspect of solid surfaces. These types of phenomena, however, will not be included in this paper.

The above sketch of the classification of the various aspects of solid surfaces represents a convenient working model. There are no sharp delineations among these various aspects, and many arbitrary decisions must often be made as in all classification schemes.

A paper such as the present one cannot, of course, be all inclusive. Neither can it be intensive lest it become extremely specialized; but it should be critical. The criteria chosen here for deciding on topics and coverage hinge on the intent to outline sterile (in terms of progress) aspects of solid surfaces, promising experimental or theoretical techniques and their limitations, areas in need of understanding, and promising areas for future scientific or technological progress. The desire to stimulate discussions has been another criterion.

SURFACE ENERGETICS

Theoretical treatments of solid surfaces, like direct experimental approaches, have encountered great complexities and are still lacking the level of refinement attained for the bulk of crystalline solids.

Gibbs' thermodynamic treatment (ref. 1) was the first attempt to consider theoretically the energetics of surfaces. It still serves as the foundation for the commonly used, macroscopic or phenomenological thermodynamic point of view.

The electronic approach to surface energetics, particularly those of metals, is more recent (one can perhaps attribute its main origin to the early 1930's) and, in spite of numerous concentrated efforts, it remains in a rather primitive stage. In this discussion the theoretical approaches to surface energetics will be briefly sketched. For extensive reviews on the thermodynamics of surfaces, see references 2, 3, and 4.

Thermodynamic Considerations

Surface tension.—In the thermodynamic treatment of systems in equilibrium, it is usually postulated that the only force acting upon them is that due to external pressure P so that the work done is represented by PdV, where V is the volume of the system in question. The other possible forces acting upon thermodynamic systems, such as gravitational, electromagnetic, and others, are usually very small and therefore neglected. However, surface forces (to be defined below) become significant and must be taken into consideration, particularly in systems which have large surface-to-volume ratios. Gibbs' thermodynamic treatment of surfaces (ref. 1) was at first extensively and successfully applied to liquid interfaces where experimental results could be obtained. Reversible changes due to surface forces in solids are significantly more difficult to observe. Nevertheless, high-temperature phenomena and processes involving changes in solid-solid boundaries have acquired increased importance. With the refinements of modern experimentation processes such as sintering, high temperature equilibrium morphologies and others can be submitted to quantitative or at least semiquantitative measurements. It should be pointed out that the thermodynamic approach to solid surfaces remains of a phenomenological nature and does not represent a system based on first principles.

Gibbs introduced a mathematical surface which he called the "dividing surface" and which is supposed to pass through points similarly situated with respect to the two phases on either side. Thus, a given system characterized by a fixed energy content and number of moles of the various constituents is divided by this surface into two volumes, V_1 and V_2. The composition of the system in the vicinity of the dividing surface can vary in some simple or complex way; although there is uncertainty in locating the dividing surface, it need not exceed a few Angstrom units.

In his treatment, Gibbs defines as surface excess of an extensive quantity (for example, free energy or composition) the difference between the actual value of the quantity and the value which the system would have if the composition of the two phases remained unchanged up to the dividing surface. Taking Γ_i, for example, as representing the surface excess per unit area of component i, for an interfacial area A, we have

$$A\Gamma_i = N_i - (C_{1i}V_1 + C_{2i}V_2) \qquad (1)$$

where N_i is the actual total number of moles of constituent i in the whole system and C_{1i} and C_{2i} are the concentrations of species i in the bulk of the corresponding two phases. The surface excess Γ can be positive, negative or zero. It is apparent from equation 1 that, for a single component system, the surface excess is independent of the placement of the dividing surface and essentially vanishes as a thermodynamic variable.

The key thermodynamic quantity for surfaces is the surface tension γ which is the reversible work, W, required to create a unit area of the surface, A, at constant volume, temperature, and chemical potential:

$$\gamma=\frac{dW}{dA}. \tag{2}$$

The surface excess of component i can also be defined as

$$\Gamma_i=\frac{dN_i}{dA}. \tag{3}$$

One can express the change in the Helmholtz free energy, dF, in terms of the work γdA at constant temperature and the chemical potential of species i, μ_i, which is defined as $(\partial F/\partial N_i)_{T,V,N_j,A}$ (N_j represents the number of moles of all components except i):

$$dF=\gamma dA+\sum_{i=1}^{n}\mu_i\Gamma_i dA$$

or

$$\frac{dF}{dA}=\gamma+\sum_{i=1}^{n}\mu_i\Gamma_i. \tag{4}$$

It is apparent from the above equation that dF/dA, the surface free energy per unit area, is equal to the surface tension, γ, only in a single component system where Γ is by definition zero.

The dependence of the change in surface tension on temperature and chemical potential is expressed by Gibbs' well-known adsorption equation, the derivation of which is not presented here (refs. 1 through 4):

$$d\gamma=-S_A dT-\sum_{i=1}^{n}\Gamma_i d\mu_i \tag{5}$$

where S_A is the specific surface entropy dS/dA. It is, of course, assumed here that gravitational and other forces are negligible.

At constant temperature, Gibbs' adsorption equation reduces to his also famous adsorption isotherm

$$d\gamma=\sum_{i=1}^{n}\Gamma_i d\mu_i \text{ or } \Gamma_i=-(\partial\gamma/\partial\mu_i)_{T,\mu_j}. \tag{6}$$

It is apparent that preferential adsorption on the surface (positive Γ) raises μ and thus reduces γ. Conversely, a decrease of the concentration of a component at the surface (negative Γ) leads to a corresponding increase in surface tension. In the case of liquid surfaces, this relationship has been strikingly demonstrated with the well-known fundamental and applied implications. It is also of great significance in solid surfaces at high temperatures. Some experimental results on the pronounced effect of adsorption on the surface tension of solids are shown in table 1.

Surface stresses.—On surfaces, in addition to surface tension, there are forces acting which are referred to as surface stresses. In the case of liquids, the surface stresses reduce to surface tension. In solid surfaces, while the surface tension measures the work required to create new surface, the surface stresses measure the work required to deform the surface. If an infinitesimal deformation of a surface region is brought about, and the deformation changes the local state of the surface, and thus changes γ, then:

$$\gamma\delta A+A\delta\gamma=\gamma A\sum_{i=1,2}\epsilon_{ij}+A\sum_{i,j}\partial\gamma/\partial\epsilon_{ij}\epsilon_{ij} \tag{7}$$

where ϵ_{ij} is the strain tensor of the deformation. This equation can be reduced to:

$$f_{ij}=\delta_{ij}\gamma+(\partial\gamma/\partial\epsilon_{ij})(i=1,2;j=1,2,3) \tag{8}$$

where f_{ij} is the surface stress tensor and $\delta_{ij}=O$ if $i=j$. In the case of liquids, the state of the surface does not change by the deformation; thus:

$$\partial\gamma/\partial\epsilon_{ij}=O \text{ and } f_{ij}=\delta_{ij}\gamma. \tag{9}$$

Thus, the surface stress is tensile force, normal to the surface having a constant magnitude γ per unit length. In the case of solids, however, the infinitesimal deformation will displace the atoms from their regular positions and, in view of their long-range interactions (ref. 2), their new positions will not be energetically the most feasible ones. Consequently, the quantity $\partial\gamma/\partial\epsilon_{ij}$ is not zero. It can be either positive or negative corresponding to compressive or tensile stresses. It should be

TABLE 1.—*Experimental Results of the Effect of Adsorption on the Surface Tension of Solids*

Solid	Adsorbed gas	Pressure mmHg	Temperature, °K	Surface tension, ergs/cm²	
				Without adsorption	After adsorption
Mica [a]	Air	760	300	4500	375
Ag [b]	O_2	150	1205	1140	350
Cu [c]	Pb	0. 1	1070 to 1170	1800	780

[a] See reference 5.
[b] See reference 6.
[c] See reference 7.

pointed out, however, that even in solids, $\partial\gamma/\partial\epsilon_{ij}$ may vanish with time if a mechanism of the restoration of the strain becomes available, as is the case at high temperatures (ref. 3). Unfortunately, it is only at or near the melting point that the resistance to surface deformation reflects directly the surface tension.

In crystalline materials, it is expected that the surface tension varies as a function of orientation. It is convenient to represent the variation of γ with orientation by a polar γ plot, in which the vector emanating from a fixed origin is proportional to the surface tension. For liquids where γ is independent of orientation, the γ plot is a sphere. Depending upon the crystal structure, there are a number of approximations which lead to the construction of a γ plot (ref. 2). A general geometrical solution, which leads to polygonal solids exhibiting a minimum total surface free energy, has been proposed by Wulff (ref. 8). The reader is referred to extensive discussions of the γ plot (refs. 2 through 4).

When γ varies with orientation, the equilibrium shape of the solid will be the one that tends to enlarge the areas of the orientations with lower γ and minimize the areas of the orientations with higher γ. In the equilibrium shape, the surface free energy must be minimized. Unfortunately, the experimental study of the equilibrium shape presents great difficulties. The particles must be very small to attain true equilibrium, and some factors interfering with the equilibrium shape, such as supporting of the particles, are not readily eliminated. Manifestations of the variation of γ with orientation are encountered in thermal etching where planes of low γ or combinations of planes minimizing the surface free energy develop, provided, of course, that there is no interference from chemical reactions with the ambient.

Estimates of surface tension.—There have been numerous attempts to calculate the surface energies of solids. These attempts vary in complexity depending on the type of solid. Thus, in the case of diamond where the atoms are covalently bonded and the valence electrons are localized, the surface energy was taken as one-half of the cohesive energy which in turn was taken equal to the energy of bonds which are cut in creating the surface (ref. 9). In the case of the (111) planes of diamond, for example, there are 1.83×10^{15} bonds per cm^2; taking 90 kcal/mole as the bond energy, the surface energy (at 0° K) is 5400 $ergs/cm^2$.

In the case of rare gas crystals held together primarily by van der Waals forces, the calculations are again based on the energies necessary to separate the atoms in the crystal and the energies needed for the rearrangements of the surface atoms to their equilibrium positions (refs. 10 and 11). The van der Waals forces are estimated from relatively simple potential energy functions where interactions other than those due to nearest neighbors can be neglected. The rearrangement energy can be estimated by allowing the surface atoms to move relative to the adjacent plane below, so that they reach a position of minimum energy. This energy is usually a small fraction (up to several percent) of the surface energy.

For ionic crystals, similar procedures are employed except that charged atoms are involved, and consequently more complex potential energy functions must be used (ref. 12). The situation becomes even more complex in the case of metals, where the valence electrons are not localized and complex wave mechanical treatments are necessary. Some calculated values of surface energy are shown in table 2. The surface energy E_s is related to γ according to

$$E_s=\gamma-T\frac{d\gamma}{dT}. \qquad (10)$$

The available experimental data on surface tension of solids is quite limited. There are, of course, data available on the surface tension of the corresponding liquids at or near their melting points. It is expected that surface energy values of solids should be about 25 percent higher than those of the corresponding liquids. (For an extensive discussion, see reference 13.) On this basis the calculated values (such as those given in table 2) are low, since they are significantly lower (up to a factor of 2) than the surface tension values of their liquids. It is of interest to note that a quantum mechanical calculation (ref. 14) of the surface tension of LiF has yielded a value of 557 erg/cm², which is higher than that of the corresponding liquid.

TABLE 2.—*Calculated Values of Surface Energies for Solid-Vacuum Interfaces at 0° K*

Solid	Surface energy, erg/cm²		Surface energy of corresponding liquid, erg/cm²
	(111)	(100)	
Diamond [a]	5400	9140	
Ne [b]	17. 2	17. 9	15. 1
A [b]	41. 1	42. 7	36. 3
NaF [b]		171	335
NaCl [b]		155	190
KF [b]		160	242
W [c]	6690	6430	

[a] See reference 9.
[b] See references 10 and 11.
[c] See reference 12.

There are no direct methods for the experimental determination of surface energies in solids. The proposed methods are to a small or large extent peculiar to certain solids and special circumstances. (For a survey, see reference 15.) In the case of mica, for example, the surface tension was determined by the work required for cleaving (ref. 5). The surface tension of silver (ref. 6) was determined by a high-temperature wire-pulling technique (ref. 16). For copper the surface tension was estimated from the angles of grain-boundary grooves (ref. 7).

Atomic Theory Considerations

The application of theories based on wave mechanics to the study of solids in general is quite limited and, in fact, has been carried out with any appreciable rigor only in a few simple systems (refs. 17 through 21). In the case of solid surfaces, further complications are introduced by the abrupt termination of the lattice and the associated periodic potential. Assumptions must therefore be made regarding their two-dimensional homogeneity (parallel to the plane of surface) and also regarding variations in lattice spacing and other properties in the direction perpendicular to the surface. However, classical methods such as those employed to estimate the surface energies of covalent, molecular or ionic crystals are not quite applicable in metals where the main contribution to the surface energy is from mobile electrons.

As for the bulk of metallic crystals and solid surfaces, energy calculations are based on the solution of the appropriate wave equation after making suitable assumptions and approximations. The assumptions are based on the electron density at the surface, the associated variation with position, the nature of the existing potential barrier, etc. The basic Schrödinger wave equation for the motion of a single electron of potential energy V is

$$\nabla^2\psi+\frac{8\pi m}{h^2}(E-V)\psi=0 \tag{11}$$

where ψ is the wave function and E is the total energy. This equation is applied in some modified form depending on the particular case under consideration and the choice of the boundary conditions.

It is considered beyond the scope of this paper to discuss the various energy calculations carried out by various researchers which have been based on a variety of assumptions and approximations. The reader may conveniently find detailed discussions of such treatments (refs. 17 through 21). It is, of course, unfortunate that the ultimate merit of the various calculations cannot be assessed on the basis of experimental results, since such results are presently very scarce.

It is of interest to point out that the solution of the proper wave equation may yield wave functions with appreciable amplitude only at the surface. These functions represent electron states localized at the surface which are known as surface states. Such surface states have been experimentally observed in semiconductors (ref. 22), and they were responsible for the discovery of the transistor. Although numerous experimental and theoretical studies have been carried out on the surface states of semiconductors, their exact nature is not yet understood.

The following comment should be made in closing the section: there are three basic problems associated with theoretical treatments of solid surfaces. One of them is associated with the formulation of the surface configuration and the necessary assumptions that must be made for setting up the mathematical apparatus or formalism. The second problem is essentially the great difficulties inherent in the manipulation and exact solution of the basically complex wave mechanical equations. The third problem has to do with the fact that, unlike other areas of science, it is still nearly impossible to bring experimental results to the assistance of theory.

One should feel hopeful, however, that new computation techniques are making significant headway in facilitating the solution of complex mathematical systems. The feedback of such solutions should help refine or improve the assumptions necessary to set up the pertinent mathematical systems. Finally, the advances in various experimental techniques including high vacuum, microscopy, diffraction, electrical measurements, and others give promise of overcoming the major difficulties in obtaining experimental values of surface parameters. Perhaps the most significant among the parameters of solid surfaces to be determined experimentally should be the surface tension. Reliable experimental values of surface tension should prove to be of incalculable value to all theoretical efforts on solid surfaces.

CRYSTALLINE STRUCTURE

There is no direct theoretical way of arriving at the crystalline structure of solid surfaces. Essentially, the problem continues to be an experimental one and as such ranks among the most difficult and challenging experimental problems, requiring advanced theoretical devices.

One would like to know the crystalline structure (on an atomic scale) of the surface layer containing no other type of atom besides those of the particular solid itself. Information on such surfaces, commonly referred to as clean surfaces, can only be meaningfully achieved in very high vacuum (10^{-9} mmHg and beyond), so that interactions of the surface atoms with the atoms of the ambient do not interfere with the measurements. Knowing and understanding the crystalline structure of clean surfaces can readily be viewed as the prerequisite for developing a sound understanding of their energetics.

The interaction of clean surfaces with foreign atoms—driving forces and kinetics—is another fundamental step essential in the understanding of solid surfaces. It is apparent that the initial stages of such interaction should be of special interest. Finally, from a fundamental point of view, it is essential to know what structural rearrangements, if any, accompany the physical or chemical interaction of the clean surfaces.

It must be recognized that even if all the above types of knowledge and understanding were available, there would still remain the problem of mastering the behavior of ordinary or real surfaces. This mastery can best be attained through working with solid surfaces from both the fundamental and applied points of view.

The experimental techniques for the study of the crystalline structure or some crystalline aspects of solid surfaces are based on microscopy or diffraction principles. Among the microscopy techniques, light microscopy and electron microscopy have been, and still are, extensively used in the study of the "microstructure" of surfaces including structural defects. The moiré patterns (ref. 23) revealing directly the presence of dislocations at the surface is one among several outstanding achievements (fig. 1) of direct microscopy. These techniques will not be discussed here as they are widely known. It should be pointed out, however, that despite their broad utilization, their potential has not been exhausted.

The field-emission and the field-ion microscopy are more recent techniques with significantly greater limits of resolution and capabilities but at the same time with greater limitations. These techniques will be discussed here briefly.

Among the diffraction techniques (X-ray and electron diffraction) the application of LEED has added a new dimension to the study of the structure of surfaces in spite of the great experimental and theoretical difficulties it presents. The LEED techniques will also be highlighted here.

Field-Emission Microscopy

The principles on which field-emission microscopy is based and the main features of the field-emission microscope developed by E. W. Müller (ref. 24) can be outlined with the help of figure 2. The specimen under study is in the form of a wire with a sharp point. A high electric field is applied between the specimen (negative potential) and the phosphor-coated positive electrode which is usually a transparent film of tin oxide. Under high vacuum (10^{-9} mmHg and beyond), electrons are emitted from the atoms at the tip of the specimen. They are

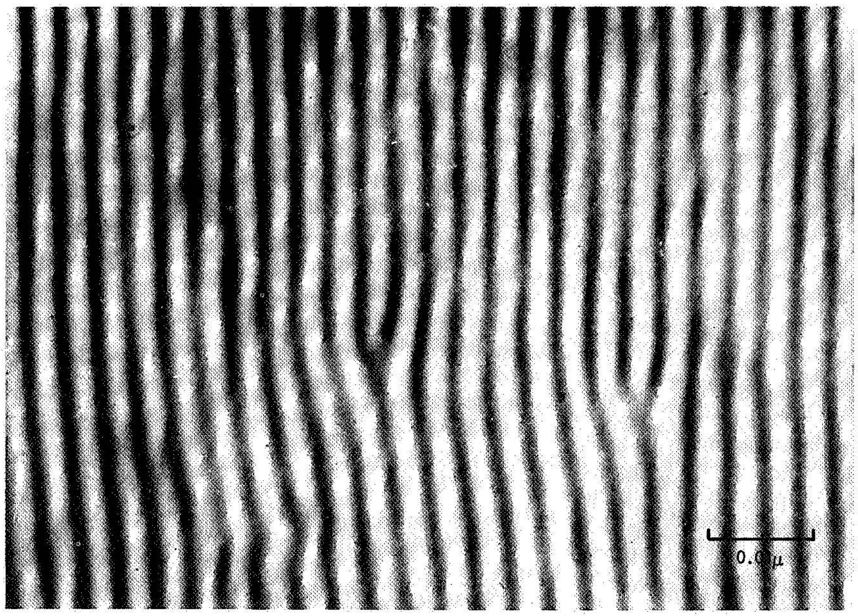


FIGURE 1.—Edge dislocations visible in parallel moiré patterns from overlapping palladium and gold layers in (111) orientation. 1 300 000X.

accelerated under the influence of the existing field and form a magnified image of the tip on the phosphor. The magnification achieved ranges from 10^5 to 10^7 times. No magnetic or electrostatic lenses are required for the production of the image.

A field-emission photograph corresponding to tungsten tip is shown in figure 3. The bright areas in the figure correspond to areas with high electron emission, i.e., relatively low work function. In this case these areas correspond to high index planes. The low index planes with corresponding higher work function appear as dark areas in the photograph.

The emission of the electrons from the surface atoms is not direct in the classical sense, but takes place by quantum-mechanical tunneling. (For a brief discussion, see reference 25.) For the electrons to leave the surface, they must gain energy equal to the work function ϕ over and above their energy at the Fermi level. The applied electric field, $\mathcal{E}$, and the associated surface image potential (which is taken to equal $e^2/4x$, where e is the electronic charge and x is the distance from the metal surface) decrease the height of the surface barrier. At the same time, the thickness of the barrier is decreased so that the probability for the electron to tunnel through the surface poten-

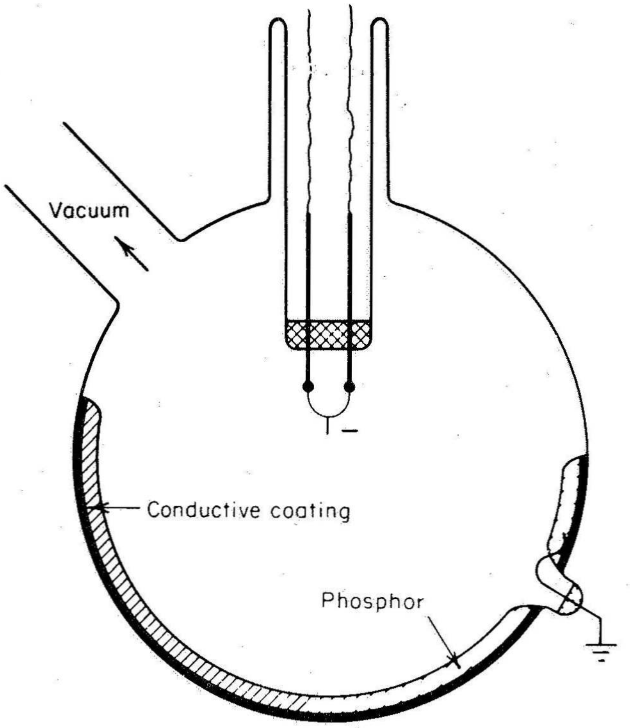


Figure 2.—Schematic representation of field-emission microscope.

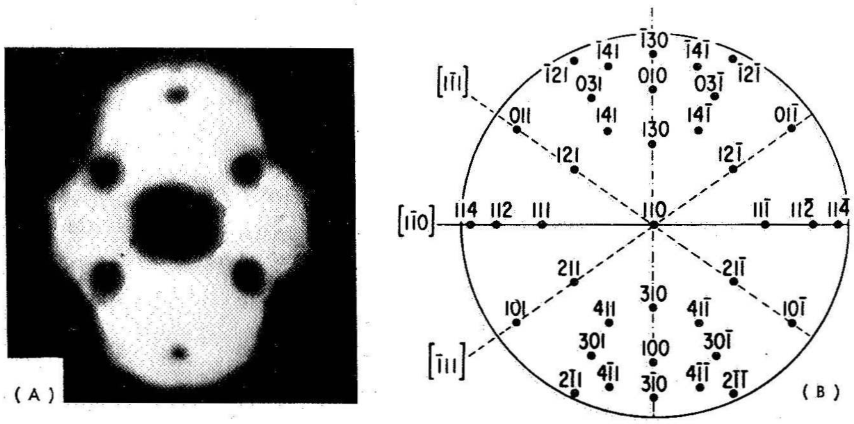


Figure 3.—Field-emission photograph of an (110)-oriented tungsten tip and the corresponding orthographic projection of the cubic lattice.

tial barrier increases. The Fowler-Nordheim expression relates the current density (in amp/cm^2) to the work function (in eV) and the field $\mathcal{E}$ (in V/cm):

$$i=A\mathcal{E}^2 \exp\left[-B\frac{\phi^{3/2}}{\mathcal{E}}\right] \tag{12}$$

where B is a constant (6.8×10^7) and

$$A=6.2\times10^6\frac{(E_F/\phi)^{1/2}}{E_F+\phi} \tag{13}$$

where E_F is the Fermi energy. For values of $\mathcal{E}$ in the range of 5×10^7 V/cm, the value of i is usually of the order of 10^{-4} amp. The resolution of the field-emission microscope is limited to about 20Å because the tunneling electrons have momentum components in directions other than the emission direction.

It is apparent that the atomic structure of individual crystallographic planes cannot be resolved by field emission; however, the usefulness of this technique has been most significant in studying the direct interaction of surfaces with gas atoms. The adsorption of gases causes pronounced changes in the work function of the metal and, thus, to the emission pattern. Since the tunneling electrons are near the Fermi level of the metal, their tunneling is not temperature-sensitive. This characteristic is of fundamental value because it allows the study of surface interactions—including migration of gas atoms on the surface—as a function of temperature.

Emission microscopy has been used successfully to study in some detail the interaction of inert gases (such as Xe) with metals such as tungsten and molybdenum. As a result, some critical aspects of physical adsorption have been resolved (ref. 26). It was found, for example, that the binding energy of xenon on the (111) regions of molybdenum is lower than that on the (100) regions in accord with previous quantitative estimates. For the case of tungsten and molybdenum, assuming no surface structural rearrangements, it was shown that binding energy of an adsorbed atom increases with the number of first and second nearest neighbors surrounding it. It was also demonstrated that interaction of surface atoms with gases (specifically in the case of adsorption of oxygen on nickel) can lead to pronounced surface structural rearrangements.

The two main limitations of field-emission microscopy are: (a) its relatively low resolution of 20Å and (b) the requirements of very low pressures, of the order of 10^{-9} mmHg. In view of the advantages of the two techniques discussed below, it appears that field-emission will continue to be used only within a rather limited scope.

Field-Ion Microscopy

Field-ion microscopy, also developed by E. W. Müller (ref. 27), is similar to the field-emission microscopy. Here again, a field is applied between the metal tip under study and the phosphor-covered electrode, but the polarity is reversed. The chamber of the microscope is filled with the "imaging gas," the atoms of which are ionized at the vicinity of the metal tip and accelerated by the field toward the phosphor-coated negative electrode where they produce an image of the tip. Helium is the most commonly employed gas. Fields of the order of 5×10^8 V/cm are necessary in view of the high ionization potential of helium (24.5 eV). The desirable radius of curvature of the metal tips is in the vicinity of 1000Å, and their preparation entails tedious and complex empirical procedures.

Field-ion formation at the tip takes place by quantum-mechanical tunneling as in the case of field-emission (ref. 25). Here, of course, the electrons tunnel from the gas atom into the metal. There is an optimum distance (of the order of several Å) for tunneling. Below this distance the energy of electrons of the imaging gas is below the Fermi level of the metal and tunneling is prohibited; beyond this distance the surface barrier becomes broader and tunneling improbable. It can be shown that tunneling cannot occur when

$$\mathcal{E}x-\frac{e^2}{x^2}<|I|-|\phi| \tag{14}$$

where I is the ionization potential of the gas.

It is apparent that for optimum imaging of the surface, the ionization of the gas must take place as close to the surface as possible. It can be calculated that with helium as the imaging gas, a field of 4.5×10^8 V/cm at 20° K, and a tip radius of 1000Å, the resolution is 1.5Å. In this case the optimum ionization distance from the metal surface is 2.5Å.

A typical field-ion image photograph is shown in figure 4. The bright spots correspond to individual atoms which can readily be resolved at the edges of the individual planes. A considerably larger number of planes appears on the ion image than on the field-emission image. The planes are outlined by the concentric rings.

The analysis of field-ion patterns is not straightforward. The field configuration, including field compression and the pertinent geometry of the tip, must be taken into consideration. Furthermore, the detailed mechanism of field-ion formation is of basic importance and appears to involve field fluctuations due to protruding atoms. Regarding the low index planes, only the atoms at the plane edges can be observed. In higher index planes, individual atoms can be resolved within the

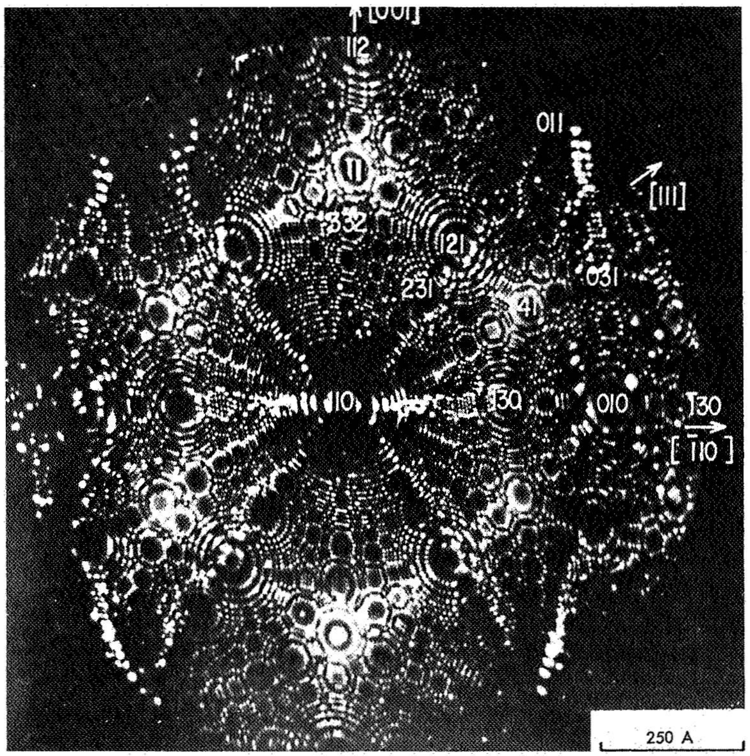


FIGURE 4.—Ion micrograph of tungsten tip at 27° K. Numbers refer to indices of planes.

planes as well as at the edges. The smaller the radius of the metal tip, the greater the number of atoms contributing to the formation of the image. For a radius of curvature of 350Å, about 30 percent of the surface atoms can be observed.

Although the theory of the ion image formation is still being developed, field-ion microscopy has already contributed importantly to the understanding of the structure of surfaces. Field-ion microscopy made it possible to observe directly, for the first time, the atomic structure of surfaces. In the instances studied (such as platinum), the overall structure of the surfaces were consistent with the extrapolation of the bulk crystalline structure to the surface.

By combining field-ion imaging with field-ion evaporation (from the metal tip), it has been possible to see directly individual vacancies and even study their behavior. In the case of platinum, employing the (012) areas, (ref. 24), it was found that there was no particularly high strain acting around the vacancy. The energy of formation of vacancies was also determined and found to be consistent with the value determined by other experimental means (approximately 1.2 eV).

Individual interstitial atoms have also been observed in several instances including the platinum and aluminum surfaces. Interstitial atoms do lead into structural distortion in their immediate vicinity. No vacancy-interstitial pairs have been observed. Other direct observations of lattice defects include dislocations (screw and edge type), grain boundaries, and slip. Figure 5 shows a screw dislocation intersecting the surface of a platinum crystal.

The field-ion microscope is now being employed quite extensively in the study of surface reactions such as adsorption, desorption, surface diffusion, and others. It is in this type of study that field-ion microscopy is most likely to make its most lasting contribution. The technique does present significant limitations over and above the obvious experimental difficulties; thus, only a few metals can be submitted to this technique in view of the severe requirements demanded by the metal tip. The large stresses associated with field-ion imaging or field-ion evaporation cause, in most instances, pronounced surface

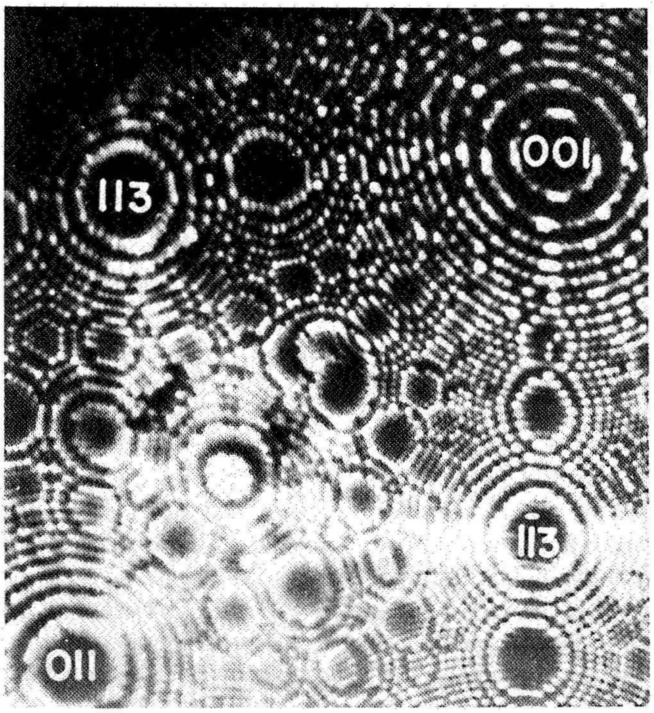


FIGURE 5.—Ion micrograph of a platinum tip. Note screw dislocation intersecting the surface.

distortion. In fact, even with refractory metals it becomes necessary to operate at temperatures below room temperature to insure surface stability and, thus, the necessary resolution.

Low Energy Electron Diffraction

With regard to the study of the crystalline structure of surfaces in general, LEED represents the most powerful tool and holds the greatest promise (ref. 28). The technique involves the recording and analysis of low energy electron diffraction patterns of solid surfaces. The energy of the electron ranges from about 20 to 200 eV. Such low-energy electrons are required because they can only penetrate one or two atom layers, and consequently they are essentially diffracted by the surface layer. The experimental difficulties associated with low-energy electron experimentation are extraordinary: they can be manipulated only in extremely high vacuum lest they be absorbed by the gas atmosphere; the diffracting surface must be clean or rather it must be cleaned in the diffraction tube itself at a vacuum of the order of 10^{-9} mmHg or beyond; two-dimensional diffraction mechanisms and the analysis of the resulting patterns still present great theoretical complexities. Each of these aspects constitutes in itself a field of research and technology and will not be discussed here in any detail. It should be pointed out, however, that their successful intimate interplay is fundamental to the success of LEED as a technique.

Although the diffraction of electrons was discovered in 1927 and although its potential was recognized soon after, only in recent years has LEED been employed in a significant number of laboratories. In fact, for many years H. E. Farnsworth and his students were the major group developing techniques for LEED applications and carrying out research on solid surfaces. This group employed tungsten filaments as sources for electrons, ion bombardment (sputtering) for preparing clean surfaces, and Faraday cage arrangements for the detection and recording of the diffraction patterns (ref. 28).

LEED has become more accessible with the development of high-vacuum technology and a relatively rapid technique whereby the diffraction patterns are displayed directly on a fluorescent screen. The group at the Bell Telephone Laboratories under the stimulation of L. H. Germer has contributed significant improvements.

A schematic diagram of a LEED tube is shown in figure 6. The electrons impinge on the crystal and, after they have been diffracted, pass through two grids. The inner grid rejects all but the elastically scattered electrons as they are the only ones contributing to diffraction. The diffraction patterns form on the fluorescent screen and can be viewed directly or photographed.

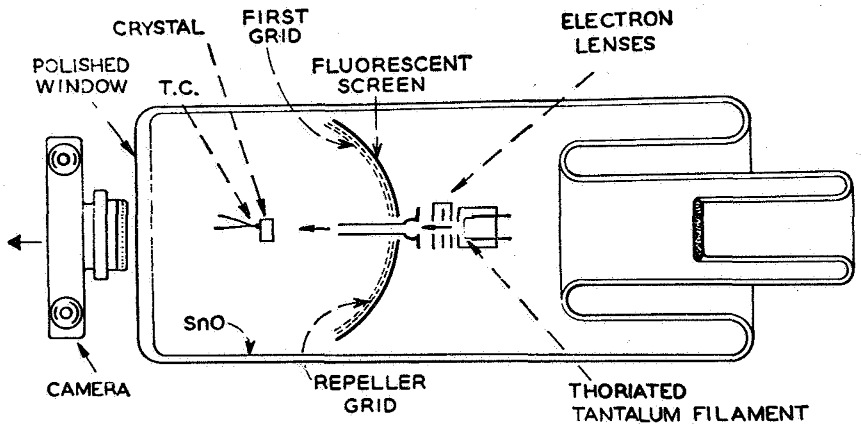


FIGURE 6.—Schematic diagram of a LEED arrangement.

Although the detailed electron scattering mechanisms associated with LEED are still being developed, the main contribution to pattern formation is made by the outermost layer of surface atoms. Such patterns have the characteristics of a two-dimensional grating diffraction. The Bragg diffraction relationship is obeyed except that the interplanar distance d becomes the distance between atom rows on the surface:

$$n\lambda = d_{hk} \sin \phi \tag{15}$$

where λ is the wavelength of the electrons which is actually equal to $\sqrt{150/V}$ in Å (V being the energy of the electrons); ϕ is the angle between the incident and diffracted beam. A schematic diagram of two-dimensional diffraction is shown in figure 7.

The wavelength of 20V electrons is approximately equal to common interatomic distances; however, diffraction effects can be observed with electrons of much smaller energies, even down to 2V. In such cases, larger-scale structural periodicities contribute to the patterns.

Developing an understanding of the LEED patterns and contrasting appropriate theoretical models have been going on in parallel with experimental observations. The experimental studies have been pursued along three main objectives: the determination of the structure of clean surfaces (usually as a function of orientation), the study of structural changes as a result of physical or chemical interactions with gaseous ambients, and the concurrent characterization of clean surfaces with regard to structure and electrical or other properties. This latter objective is, perhaps, the most promising, as it will hopefully help relate directly structure and properties.

Numerous reports have been published on the structure of clean surfaces of metals and semiconductors (ref. 28). Nickel (face-centered cubic) has been extensively studied by the Brown University as well as the Bell Telephone Laboratories groups. Actually, nickel was the material used in the original experiment of Davisson and Germer that demonstrated the diffraction of electrons in 1927. In the case of the (111), (110), and (100) clean surfaces of nickel, the LEED patterns reveal no structural rearrangement or, "superstructures." Thus, there is no difference in translational symmetry between the surface atoms and those in the bulk of the crystal, on planes parallel to the surface. Lack of surface rearrangements does not preclude vertical displacements of the surface atoms or shifts of the entire surface plane. In fact, in the case of the (111) and the (110) surface atoms, there is evidence for a 5 percent outward displacement (ref. 29). The absence of surface rearrangements has also been reported for clean (111), (110), and (100) surfaces of platinum (ref. 30). It appears that in the case of metals there are no atom rearrangements at the surface. This observation is significant because it has been used as an assumption in considering the energetics as well as various surface phenomena, an example of which is epitaxy.

Nickel surfaces have also been employed in studying the structural aspects of adsorption and chemisorption (ref. 31). It is of particular interest to note that adsorption of oxygen (or hydrogen) on the (111) and (100) surfaces brings about no rearrangements of the sur-

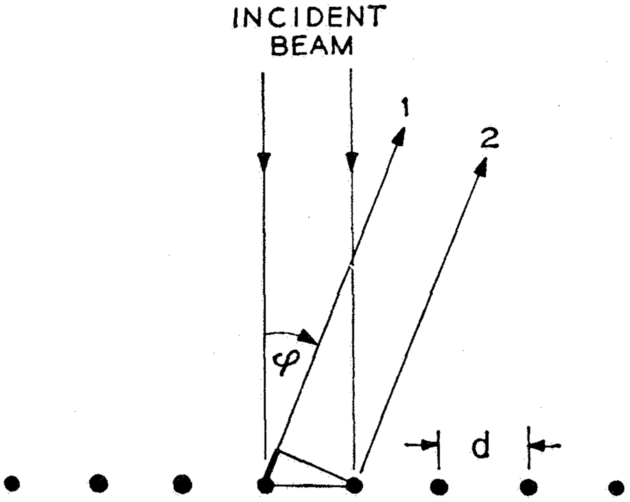


FIGURE 7.—Schematic diagram of two-dimensional diffraction.

face atoms. There is a more or less regular arrangement of the adsorbed atoms on these surfaces. In the case of the (110) surfaces, which are less dense and have a more "open" structure than the other two low-index surfaces, there are reconstructive rearrangements of the metal surface atoms following the adsorption of oxygen or even hydrogen. These rearrangements do lead to lower free-energy configurations. It is apparent that the detailed understanding of these rearrangements, hopefully in the foreseeable future, will settle many of the uncertainties associated with the energetics of physical adsorption and chemisorption.

Semiconductors of the Group IV type (Ge, Si), or the group III–V type (InSb, GaAs), or the group II–VI type (ZnO, CdS), have the diamond or diamond-like structure characterized by tetrahedral coordination (each atom has four nearest neighbors) and highly directional bonds. These characteristics are to be contrasted to those of metals where the number of nearest neighbors is significantly higher (typically six, eight, or twelve) and the bonds are not directional. It is to be expected that the atoms of semiconductor surfaces would undergo constructive rearrangements to accommodate their relatively pronounced unsaturated nature. Extensive LEED work has been performed on clean germanium and silicon surfaces (ref. 28). More recently the work has been extended to some semiconductor compounds. In all instances, the diffraction patterns indicate the presence of constructive rearrangements of varying complexity depending on the orientation. These rearrangements have not been resolved in detail, although much progress has been made in that direction. Also, significant experimental knowledge has been acquired regarding the correspondence of structure and electronic properties.

Although only a sketchy account of LEED as an experimental tool can be given here, it must be stated emphatically that it represents the most powerful tool ever to become available for the fundamental study of surfaces. As the associated experimental techniques become more refined and more versatile and as the theoretical foundation of LEED patterns improves, there is no doubt that our understanding of the surfaces and their interactions will undergo major advances. It is only natural that such advances will reflect directly on the understanding and control of the behavior of real or ordinary surfaces and consequently to the understanding of numerous processes.

Ion Microprobe Mass Spectrometer

This spectrometer is a relatively recent analytical tool (ref. 32); but because of its great sensitivity and applicability to solid surfaces, it should be worthwhile to discuss it very briefly here. The surface being analyzed is bombarded with a beam of inert gas ions having an energy of about 10 kV. The sputtered surface atoms which are ionized, if the bombarding ions are of sufficient energy, are focused into a mass spectrometer analyzer. The sputtering process is applicable to metals as well as to nonmetallic materials and can be controlled so that successive monolayers can be removed from the surface. The sensitivity of the method at present is of the order of parts per million, but it appears that it will be increased to parts per billion in the foreseeable future.

The method is relatively new and was first tested in 1949. Its potential is rather apparent, particularly with regard to the compositional determinations of surface layers. No other means are available for analyzing solid surfaces one atom layer thick for impurities in the range of parts per million to parts per billion. This spectrometer may therefore prove to be the most powerful tool available for the analytical study of real surfaces.

SOLID-GAS INTERFACES

Solid-gas interfaces represent the most important interfaces from the scientific, engineering, and technological points of view of solid surfaces. Surface-gas interactions provide the most direct information regarding the energetics of surfaces and their structural characteristics. In fact, our present understanding of surfaces is based to a very large extent, if not primarily, on the results of surface-gas interactions.

With respect to engineering and technology, it is obvious that surface-gas interactions are critically important in all uses of solid materials. These interactions are bearing either directly or indirectly on the various functions of solids (friction being only one typical instance). Even the mechanical strength characteristics of metal beams for bridge construction cannot be considered without the interaction of their surfaces with the environment. Although a significant segment of the industry is involved in the protection of solid surfaces from gaseous environments, by necessity it is carried out largely on an empirical or semiempirical basis.

General Considerations

Although the detailed understanding or formulation of surface energetics is still unresolved, it is readily recognized that surface atoms are in a state which can be referred to as "unsaturated," "reactive," or "excited." Large amounts of energy, corresponding to a significant fraction of the total cohesive energy, are required for surface formation. It is estimated (ref. 26), for example, that the surface free energy (γ) of the (110) surfaces of copper is 1670 erg/cm^2 and its surface energy ($\gamma - T d\gamma/dT$) 32 kcal/mole. The cohesive energy of copper (heat of sublimation) is 82 kcal/mole.

Solid surface-gas interactions are customarily distinguished in physical adsorption and chemisorption. In physical adsorption, the heat of reaction is usually less than 10 kcal per mole. Forces of the van der Waals type are involved here and are comparable to forces acting in the liquefaction of inert gases. In chemisorption, the nature and magnitude of the forces involved are comparable to those associated with chemical bond formation. It should be underlined, however, that chemisorption need not imply the formation of chemical compounds in the classical sense. Thus, species resulting from chemisorption (of the order of a monolayer) are "two-dimensional" with characteristics (including the energetics) different from those of the corresponding "three-dimensional" chemical compound phases. For example, the heats of chemisorption of O_2 and N_2 on W are 194 and 85 kcal/mole respectively; the corresponding heats of formation of the compounds WO_2 and W_2N are 134 and 34.4 kcal/mole (ref. 26).

In view of the large energy involved, it is apparent that chemisorption changes significantly the characteristics of solid surfaces. The nature of such changes—energetics and kinetics—continues to be the subject of intensive investigations.

With the formation of an adsorbed monolayer, the surfaces are no longer strongly unsaturated and thus exhibit no pronounced tendency for further interactions. However, if a mechanism is provided for the exposure of new surface atoms to the ambient, the solid-gas interaction may proceed beyond a monolayer and, in fact, can led to new three-dimensional phases. Such extended interactions are usually referred to as oxidation. The study of oxidation, particularly oxidation of metals, is a scientific field in itself.

Physical Adsorption

Since the van der Waals forces acting in physical adsorption are of the same nature as those responsible for the liquefaction of gases, the adsorbability of gases should increase with their increasing ease of liquefaction. The above correlation between these two properties has been known for many years. Helium being the most difficult gas to liquefy has the least tendency to adsorb and its heat of adsorption on various surfaces is very small.

Actually, the forces acting among inert gas atoms can serve as a starting point in considering physical adsorption. The potential energy $E(r)$ of two atoms at distance r can be represented according to the Lennard-Jones approximation as:

$$E(r)=4E_o\left[\left(\frac{\sigma}{r}\right)^{12}-\left(\frac{\sigma}{r}\right)^{6}\right] \tag{16}$$

where E_o is the depth of the potential well and σ is the internuclear separation at which the potential energy vanishes. The attractive forces vary as $1/r^6$ and are due to valence electron interactions between the two atoms; the repulsive forces at small internuclear separation result from approaching interpenetration of the closed electron shells which is prohibited by the Pauli principle. The Lennard-Jones approach can be extended to the interaction between solid surfaces and gas atoms. Such treatment is not particularly complicated in the case of covalent and ionic solids in which the electronic configuration of the atoms can be satisfactorily defined. In fact, in those cases the potential energy curve is similar to that for the atom-atom interaction (fig. 8).

In the case of metals where the valence electron configuration is not well defined, the description of the adsorption forces is no longer straightforward. Lennard-Jones considered the metal as a perfect conductor where the fluctuations in the electron cloud around the adatom induce a mirror image, and he arrived at an expression for the attractive interactions. Refinements have been proposed by subsequent investigators. The following expression for the interaction E between an adatom and a metal surface atom is discussed in some detail in reference 26:

$$E=-\frac{6a^3}{\pi n_c}\frac{1}{r^6}\left(\frac{mc^2x}{2}\right)\left[\frac{A^{(2)}(v_1)}{A^{(2)}(0)}+\frac{Ce^2\alpha^{(1)}}{8mc^2xr_s}\right] \tag{17}$$

where a is the lattice spacing, n_c the number of atoms per unit cell, m the electron mass, c the velocity of light, x the diamagnetic susceptibility, $A^{(2)}(v_1)$ the polarizability of the metal per unit volume at frequency v_1, $A^{(2)}(0)$ the static polarizability, C is a dimensionless number (about 2.5), $\alpha^{(1)}$ is polarizability of the adatom, and r_s is the radius of a sphere encompassing a single electron in the metal. The total potential is obtained by summation over all atom cells in the metal.

Although there are questionable assumptions and approximations in the above expression, including the fact that it contains no repulsive terms, it leads into a very significant conclusion: since the high frequency oscillation of the electron cloud of the adatom cannot be followed by the metal valence electron, adsorption on metals is of similar nature as on nonmetals.

Adsorption and Heterogeneity of Surfaces

It is apparent from the discussion of the forces acting in adsorption that the interaction energies must vary with crystallographic orientation. The greater the density of lattice atoms surrounding the adatoms, the greater the interaction energy. Consequently, the higher the atom density on a given crystallographic surface, the more pronounced is

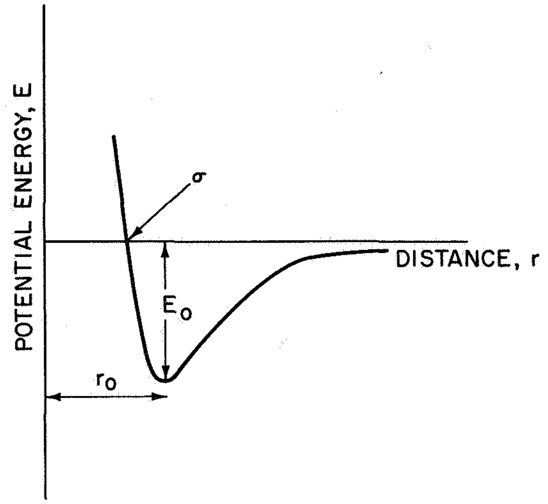


FIGURE 8.—Schematic representation of atom-atom interaction as a function of interatomic distance according to Lennard-Jones approximation.

the adsorption. Similarly, adsorption is most pronounced in lattice holes or edges.

Variations in the adsorption characteristics as a function of orientation has been demonstrated experimentally in a number of instances. In the potassium chloride-argon system, it was shown (ref. 33) that the heat of adsorption of argon is significantly greater in the (111) surfaces than in the (100) surfaces, since in this rock-salt structure the diatomic structure of the (111) surfaces allows the penetration of the argon atoms between surface atoms, leading to stronger adsorption.

With the development of the field-emission microscope, the variation of adsorption with orientation was demonstrated most strikingly and studied in some detail. It was shown, for example, that in the tungsten-xenon system, the binding energy on the (111) regions is lower than on (100) regions, consistent with the fact that the number of first and second nearest neighbors metal atoms surrounding the adatom is greater in (100) regions than in the (111) regions in the b.c.c. metal (ref. 26). As pointed out earlier, field-emission microscopy is uniquely suited for the detailed study of the behavior of the adsorbed layer including surface migration.

Another manifestation of surface heterogeneities is the decrease of adsorption heat as a function surface coverage. This decrease has been demonstrated in a number of solid-gas systems, and it can amount to about an order of magnitude from its initial value (zero coverage) to that corresponding to coverage approaching a monolayer (fig. 9).

The decrease of adsorption heat with surface coverage can, to a certain extent, be attributed to repulsions among the adsorbed layer atoms. Such repulsive forces may be due to electrostatic repulsion of dipoles and to short-range repulsion arising from the overlap of the electron clouds of the interacting particles. In the case of hydrogen monolayers on metals (such as nickel or tungsten), short-range interactions are not present since the hydrogen atom diameter (about 2.0Å) is less than the shortest distance between sites (2.48Å for nickel and 2.73Å for tungsten). Calculation of electrostatic repulsion gives a value of approximately 2.5 kcal/mole for the heat decrease between zero coverage and coverage of one monolayer for tungsten-hydrogen system. The experimentally observed heat decrease is about 42 kcal/mole, suggesting that a major role is played by surface heterogeneities (ref. 34).

A further very important manifestation of surface heterogeneities is the variation of the work function with orientation. It is well known that the low index planes exhibit higher work functions than the high index planes. Actually the field-emission microscopy patterns, as pointed out earlier, are a direct reflection of this characteristic property. Here again, the field-emission microscope provides the means for

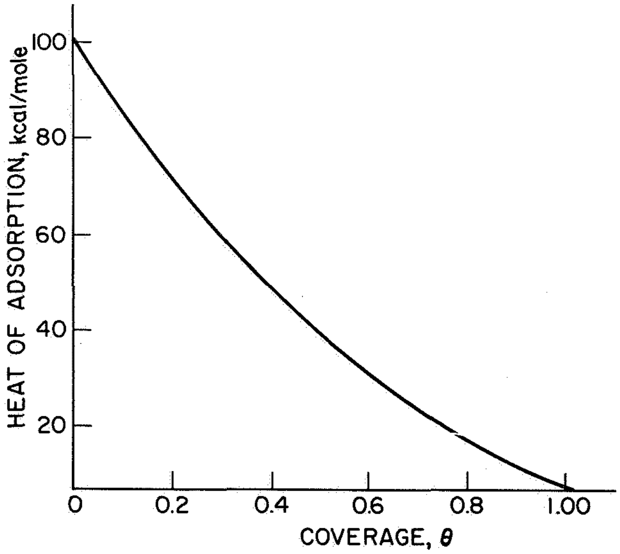


Figure 9.—Typical dependence of heat of adsorption on surface coverage θ in monolayers.

the direct study of the variation of work function with orientation and the influence of adsorption or other important parameters.

Surface heterogeneities have been recognized for some time as being responsible for observed variations in catalytic activity (ref. 35). A small fraction of the surface can in fact be responsible in many instances for the observed catalytic action. Accordingly, it is possible to inhibit catalytic reactions by addition of catalytic inhibitors in amounts far less than required to form a monolayer on the surface.

Although the intrinsic heterogeneous behavior of surfaces is not yet clearly understood, it must be recognized as a very important aspect of surface behavior in the applications of metals. It should be further pointed out that surface heterogeneities can be introduced by localized cold work composition variations, dispersed phases, and other extrinsic factors.

Chemisorption

The nature of the interaction forces in chemisorption, plus their quantitative treatment, have not been developed in detail; however, a great deal of work has been carried out in that direction (ref. 34).

It is possible to diagnose the type of bond at the surface from dipole moment data. Such data must, of course, be obtained at low surface coverage since mutual depolarization may decrease the dipole moment at high coverages. In the case of chemisorbed alkali metals (Na, K, and Cs) on tungsten, the experimentally determined dipole moments are in reasonable agreement with those calculated on the basis of the presence of monovalent ions (i.e., Na^+) on the surface (ref. 34, p. 970). It is thus reasonable to conclude that in this case chemisorption is associated with pure ionic bonding.

In the case of chemisorption of gases on metals, the existing experimental evidence (very small surface dipole) favors the view that the chemical bond formed is covalent. Gases chemisorb to a greater extent on transition metals with partially empty *d*-shells. It has been proposed that in such metals covalent bonds are formed between the chemisorbed gas and the *d*-orbitals. Extensive measurements on transition metals show that their magnetic susceptibility decreases upon chemisorption, indicating that their unpaired *d*-electrons participate in covalent bond formation upon chemisorption. This type of *d*-bond formation is apparently very significant in the mechanisms of catalysis since transition metals are in general the best catalysts.

Covalent bonding with partially filled *d*-shells is not the only type of covalency in chemisorption. Chemisorption of oxygen and other gases on nontransition metals does involve covalent or partially ionic bonding. Actually oxygen chemisorbs on all metals except gold.

The detailed discussion of the chemical bonds involved in surface chemisorption is beyond the scope of the present coverage. It should be pointed out, however, that a significant spectrum of bonds (coordinate links, surface radicals, etc.) must exist between the purely ionic and the purely covalent ones. Chemical bonding in chemisorption will be greatly clarified as the structure of the solid surfaces and their rearrangements during chemisorption are resolved.

The mobility of chemisorbed layers is another very important aspect in the study of surfaces. The motion of such layers is an activated process so that only above a certain temperature do chemisorbed layers tend to be mobile. This temperature is, of course, below the temperature of desorption. In general the activation energy for migration increases with increasing heat of adsorption. Since the adsorption heat decreases with surface coverage, it is apparent that the mobility of surface layers is higher at higher coverages.

Numerous experimental studies have been performed to determine the activation energy of migration in many systems. Table 3 summarizes some results on the mobility of chemisorbed metal layers (ref. 24, p. 205). T_1 is the temperature at which surface mobility can first be detected, and T_2 the temperature at which evaporation just begins. It can be seen that T_1 is roughly ½ of T_2. If it is assumed that the velocity differences in the two processes are due to differences in the corresponding activation energies, then the activation energy of migration can be taken to be approximately one-half the activation energy of desorption. Since the chemisorption process is not activated, it is apparent that the activation energy of migration is about one-half the heat of adsorption.

TABLE 3.—*Mobility of Some Adsorbed Layers**

System	T_1, °K	T_2, °K
Th on W	870	2100
Th on Mo	870	2100
Ba on W	400	900
Ba on Mo	400?	900
Na on W	300	600
Na on Mo	300	600

*Mobility can just be observed at T_1, and at T_2 evaporation just begins (ref. 24, p. 205).

The most direct method for studying surface migration is by means of the field-emission microscope where migration can be continuously monitored as a function of temperature and time.

Oxidation and Surfaces

Interaction of metal surfaces with gases does not necessarily cease with the formation of an adsorbed monolayer. (For an extensive treatment, see reference 36.) If a mechanism is available for the continuous exposure of new surface, the interaction with the ambient proceeds, leading to the formation of a three-dimensional phase. The formation of thick surface films has been most extensively studied in the case of oxide films, although certainly not to the exclusion of other films such as sulfides.

Film formation is obviously very important from the technological point of view and most complex scientifically. The phenomenology of film formation on a macroscopic basis has been well formulated. The detailed mechanisms involved, however, are still on a speculative basis although they have occupied numerous brilliant minds.

From the experimental point of view it has been found that film formation in general (in terms of thickness y as a function of time t) can be described by one of three general equations:

$$y=k_1t+k_2 \text{ (linear)} \tag{18}$$

$$y=k_3\log k_4t+k_5 \text{ (logarithmic)} \tag{19}$$

$$y^2=k_6t+k_7 \text{ (parabolic).} \tag{20}$$

It is apparent that in the case of linear growth, the film itself presents no barrier to its further growth. The surface films are highly porous and populated with cracks so that the same amount of substrate surface is exposed to the gas at all times.

In the other two instances (eqs. 19 and 20), the initial film slows the solid-gas interaction. Communication of the substrate with the gas must take place by migration of gas through the film to the metal-film interface, or by migration of metal species through the film to the film-gas interface. Both mechanisms can, of course, operate simultaneously. These mechanisms have been treated in great detail either in general terms or as applied to specific systems. In some instances, theory and experiment are in agreement; however, many questions remain unsettled. It is of particular interest, for example, to know the role of the initial stages of film formation in the subsequent film growth. The study of the initial stage of oxidation including the surface rearrangements of the metal has been recently pursued by new experimental techniques including the low energy electron diffraction. The results obtained are promising indeed (ref. 37).

Summarizing Remarks

Although solid-gas interfaces are of the most fundamental importance regarding the science and technology of solid surfaces, their detailed understanding remains a matter of the future. Some of the reasons for this situation have probably become apparent from the preceding discussion. Perhaps the most important reason is associated with the limited experimental information on the microscopic configuration of clean surfaces and on the ensuing changes (on an atomic scale) following the exposure to gases. With the rapid development of experimental techniques for the study of interfaces on an atomic scale, coupled with controlled high-vacuum facilities, rapid progress is expected in the next decade. This writer believes that the most important aspects of the solid-gas interfaces lie in the initial detailed interaction of "clean" surfaces with gases well before a monolayer is formed. The energetics, structural changes, and kinetics of these initial interactions should constitute the basis for the subsequent phenomena at the solid-gas interfaces.

SOLID-ELECTROLYTE INTERFACES

As in the solid-gas interfaces, here one is also concerned with the energetics of the associated equilibria, with steady-state configurations, and with the kinetics of the processes involved (refs. 38 and 39). All aspects of the solid-electrolyte interfaces are influenced, if not controlled, by the ions present in the liquid phase and the electrical carriers, electrons or electron deficiencies (holes), in the solid surface. Electrical charge transfer across the interface characterizes electrode processes in general, including such broad technical areas as electrolysis and corrosion.

In spite of the basic and technological significance of electrochemistry and the many electrochemical problems in need of understanding, basic work on many phases of electrochemistry is conspicuously limited, particularly in the United States.

The Electrical Double Layer

The electrical double layer across an interface (ref. 40) was first discussed quantitatively by Helmholtz (1879) for the metal-electrolyte interfaces. He proposed the presence of a fixed layer of charges on the metal surface and a layer of ions of opposite charge held rigidly at some finite distance in the solution (fig. 10). Such a configuration was recognized as a parallel-plate capacitor and was readily treated quantitatively. If x is the distance between the oppositely charged plates and ϵ the dielectric constant, the corresponding capacitance per square centimeter is

$$C=\frac{\epsilon}{4\pi x}. \tag{21}$$

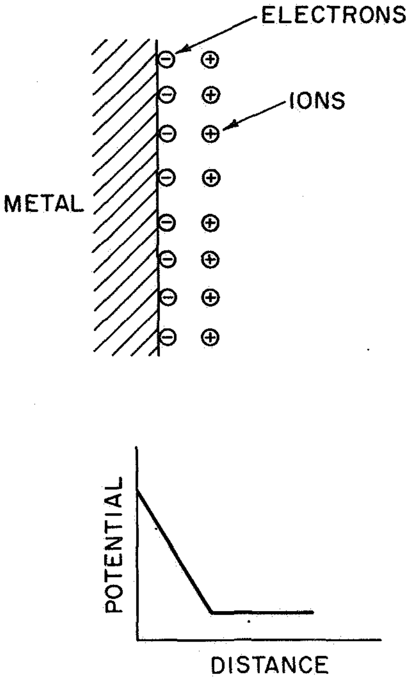


FIGURE 10.—Schematic representation of the Helmholtz double layer.

Thus, if q is the surface charge density, the potential drop ΔV across the interface is

$$\Delta V=\frac{q}{C}=\frac{4\pi xq}{\epsilon}. \tag{22}$$

The limitation of this simple model is the unlikely existence of a "rigid" layer of ions in the liquid phase. A more refined model was proposed by Stern (1924) who assumed a fixed layer of charges on the solid and a nearly rigid layer of oppositely charged ions directly adsorbed. Further in the solution he proposed the existence of a diffuse charge layer, either with opposite or same sign as the adsorbed layer (fig. 11). Thus, there is a sharp potential drop in the fixed part of the layer and a slow drop in the diffuse part. The latter potential, referred to as the zeta potential (ζ), is of basic importance in electrokinetic phenomena. For simplicity, here also it can be assumed that the diffuse double layer corresponds to parallel-plate (x distance apart) condenser; accordingly,

$$\zeta=\frac{4\pi xq}{\epsilon} \tag{23}$$

where q is again the charge density per square centimeter. Accordingly, the capacity in Stern's double layer is considered due to two condensers in series so that the total capacity C is:

$$\frac{1}{C}=\frac{1}{C_f}+\frac{1}{C_d} \tag{24}$$

where C_f is the capacity of the fixed and C_d of the diffuse components.

Under equilibrium conditions (no charge transfer across the interface) the interface tension, γ, is a function of the potential difference (ΔE) across the interface. Actually it has been shown (ref. 41) that

$$d\gamma=-qdE-\Sigma\Gamma_i d\mu_i \tag{25}$$

where q is the interfacial electronic charge density (not including charge of ions), Γ_i is the excess of species i (in moles per interfacial unit area) over that present in the bulk of the phase, and μ_i is the chemical potential of that species.

At constant composition (constant μ) one obtains from equation 25:

$$\left(\frac{d\gamma}{dE}\right)_\mu=-q \tag{26}$$

and

$$\left(\frac{d^2\gamma}{dE^2}\right)_\mu=-\frac{dq}{dE}=C \tag{27}$$

where C is the differential capacitance.

The dependence of γ on E leads to the well-known electrocapillary curve studied extensively in the case of mercury (fig. 12). It is seen that the maximum interfacial tension (electrocapillary maximum) corresponds to the point of zero charge ($q=0$). This point of zero charge is of fundamental significance in the behavior of surfaces, particularly with regard to adsorption from liquids.

The direct determination of electrocapillary curves for solid surfaces would have far reaching basic and applied implications in understanding and controlling the behavior of solid surfaces. As of now such a determination is not possible.

The differential capacity C (eq. 27) can be experimentally determined. It provides one of the most direct means for the study of the electrical structure of the double layer. It has been successfully employed in the study of semiconductor-electrolyte interfaces (ref. 42). Since it is proportional to the real interfacial area, it can serve for the determination of relative surface areas; however, the actual measurements present many difficulties associated with achieving well-characterized interfaces (including steady-state or equilibrium conditions). The reproducibility is generally poor, particularly in interfaces of

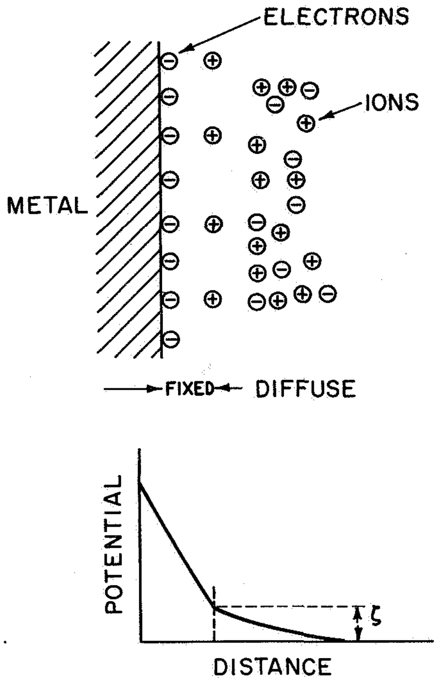


Figure 11.—Schematic representation of the Stern double layer.

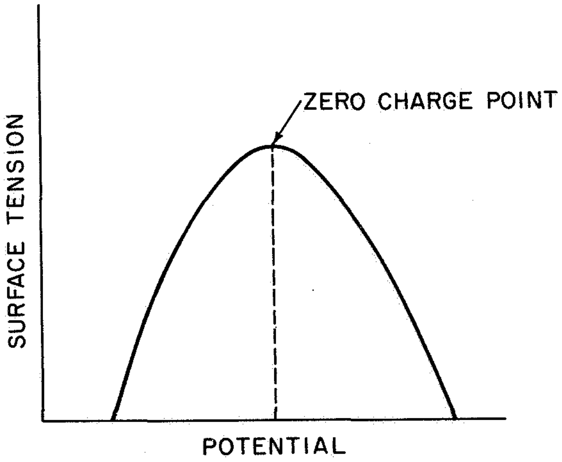


Figure 12.—Schematic representation of the electrocapillary curve.

technological importance where well-defined and reproducible conditions are not common.

Study of the electrical double layer is not receiving the research attention it deserves, perhaps because of its complexity. Yet the key to the macroscopic chemical and mechanical properties of metal surfaces in most of their applications is likely to lie in the detailed structure of the electrical double layer.

Electrochemical Nature of Corrosion

The corrosion of metals in its various forms (including some metal-gas cases) is electrochemical in nature (refs. 43 and 44). Metal atoms in the crystal are oxidized to metal ions as shown schematically in the reaction:

$$M^o \rightleftarrows M^{z+} + ze^-. \tag{28}$$

Here M^o represents metal atoms (zero oxidation state) in the crystal, z is a small integer (valence of the ion), and e^- is an electron. This oxidation reaction occurs on areas of the solid surface called anodes or local anodes, and it is referred to as the anodic reaction. The necessary reduction reaction involving the reduction of a nonmetal N (oxygen, hydrogen, etc.) can be represented as follows:

$$N + ze^- \rightleftarrows N^{z-}. \tag{29}$$

This reaction, referred to as the cathodic reaction, takes place at the local cathode areas. Thus, corrosion proceeds under the action of local electrolyte cells. The two "partial" reactions (eqs. 28 and 29) can be combined into a single oxidation-reduction reaction as follows:

$$M^o + N \rightleftarrows N^{z+} + N^{z-}. \tag{30}$$

The free energy in equations 28 and 29 is

$$G_{(1)} = -zFE_1 \text{ and } G_{(2)} = -zFE_2 \tag{31}$$

where E_1 and E_2 are the respective electrode (oxidation) potentials. The free energy charge (driving force) in equation 30 under thermodynamically reversible conditions is

$$\Delta G = -zF(E_1 + E_2)$$

or

$$\Delta G = -RTlnK + RTln \frac{(\alpha_{Mz+})\ (\alpha_{Nz-})}{(\alpha_{Mo})\ (\alpha_N)} \tag{32}$$

where R is the gas constant, K the absolute temperature, and α is the activity of the species indicated as subscripts.

Under corrosion conditions, the rate of reaction may be controlled by the rate of equations 28 or 29, or both (electrochemical control). If reactions in equations 28 and 29 are very fast, then it is likely that equation 30 is controlled by the rate of diffusion of species N to the metal-electrolyte interface (diffusion control). Both types of control are frequently encountered. The diffusion-type control is obviously very common in the absence of agitation, stirring, or convection at the interface. Regardless of the type of reaction control, the partial reactions (eqs. 28 and 29) play a very important role in the nature of the surface microstructure resulting from corrosion. Furthermore, lattice defects (such as dislocations and grain boundaries), impurities, and other surface heterogeneities may favor or enhance either or both of the partial reactions. Locations with high surface free energy will, of course, enhance the anodic reaction.

During corrosion the rate j_c of the cathodic reaction (current flow from the interface to the metal) is equal to the rate j_a of the anodic reaction (current flow from metal to the interface), and at steady state they are equal to the overall corrosion rate J_{cor}:

$$j_c = j_a = J_{cor}. \tag{33}$$

As a result of the current flow (nonequilibrium conditions) the potentials of the anodic and cathodic elements of the local electrolytic cells are polarized, i.e., removed from their reversible values. The polarization of the local cells during corrosion can be illustrated schematically (fig. 13). The "open circuit" of the cathodic reaction (E_c) and of the anodic reaction (E_a) are the corresponding equilibrium potentials in equations 28 and 29. The corrosion potential lies somewhere between the above equilibrium values and depends on the particular shape of the cathodic and anodic curves, which for simplicity, are shown as straight lines in figure 13. The intersection of the two partial polarization curves determines the corrosion potential.

It is apparent from figure 13 that conditions decreasing the slope of the polarization curves (depolarizing conditions) will shift their intersection point to a higher current value resulting in a higher corrosion rate. Conversely, an increase in their slope (polarization) will lead to a decreased corrosion rate. The various polarization diagrams deriving from the general case of figure 13 are schematically represented in figure 14.

Both the cathodic and the anodic reactions are activated processes, and their rates are usually expressed in the Tafel relationships:

$$j_c = i_o \exp\left(-F\alpha\eta_c/RT\right) \tag{34}$$

$$j_a = i_o \exp\left(-F\beta\eta_a/RT\right) \tag{35}$$

where F is the Faraday constant; α and β are constants whose values depend on the shape of the polarization curves and their interaction point; η_c and η_a are the corresponding activation overvoltages, i.e., the change in the potential from the equilibrium value; it is here assumed that no polarization due to concentration gradients of the ions takes place; T is the absolute temperature, R is the gas constant, and i_o is the exchange current, i.e., the current flowing across a unit area of the electrode in each direction at the reversible potential ($\eta = 0$).

It is apparent that the corrosion theory of local cell action requires the presence of discrete anodes and cathodes. The distribution of the anode and cathode areas need not be fixed; rather, it can vary with time resulting in uniform corrosion attack although corrosion (dissolution) per se takes place at the anode areas. Actually, the concepts of corrosion potential (referred to also as mixed potential) and polyelectrode (referred to also as mixed electrode) can be developed by assuming two or more statistically independent electrode reactions

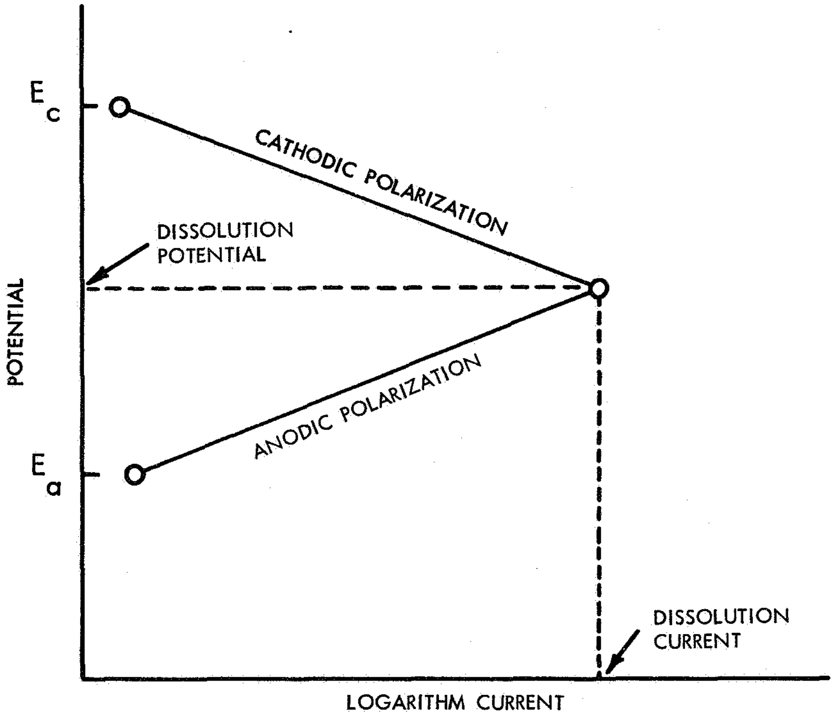


FIGURE 13.—Schematic representation of polarization curves of corrosion local cells.

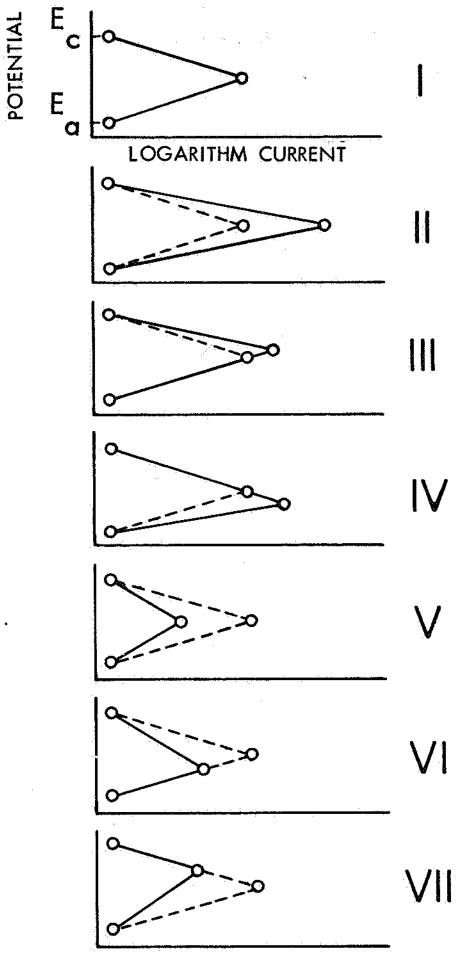


FIGURE 14.—Schematic representation of the polarization curves of corrosion local cells. Cases II–VII are derived from case I (which corresponds to figure 13) by depolarization, or further polarization of the anodic reaction, the cathodic reaction, or both. The diagrams are self-explanatory.

which are related by the common electrical potential difference (mixed potential) between the metal and the electrolyte. Nevertheless, nonuniform corrosion can take place when parts of the corroding surface behave predominantly as anodes or cathodes. "Pitting" is one form of such type of corrosion where the areas leading to pitting serve predominantly as anodic areas. Preferential attack is usually associated with the presence of surface heterogeneities as pointed out earlier.

Although the basic electrochemical nature of metallic corrosion is well established, the actual mechanisms of the various types of corrosion remain obscure. It is of particular interest to point out in the present context that the specific roles of strain, grain boundaries, dislocations, and other structural defects are not understood. Similarly, there are major uncertainties regarding the effects of segregations, adsorbed layers or films, and many other factors commonly present in metallic corrosion. One might reasonably conclude that some major parameters entering lubrication, friction, and wear phenomena are directly related to the microscopic electrochemical mechanisms of metallic corrosion. The initiation of stress corrosion cracking and corrosion fatigue represent an exciting interplay between structural, mechanical, and chemical characteristics of solid surfaces. Here also our knowledge is sketchy.

Solid-Solid Interfaces

Many important aspects of solid-solid interfaces (ref. 13), particularly friction and adhesion, are extensively covered elsewhere in this conference. Only some aspects of general interest will be highlighted here.

When two solid surfaces are brought into intimate contact, their atoms will interact. The nature of the interacting forces is not different from that in the solid-gas or solid-liquid interfaces. In an ideal case where two atomically flat surfaces are brought into intimate contact without exposure to a gaseous atmosphere (clean surfaces), one can make reasonable approximations of the interaction forces. Of course the structure (and bonding) of the surfaces has to be known (or assumed) as well as the relative position of the atoms of the two surfaces. Here again the surface free energy is a good index of the relevant interactions, as it represents the energy of the surface in its activated state. It is of interest to recall here that in many instances the surface free energy, γ, is not very different from the total surface energy (enthalpy, $\gamma—Td\gamma/dt$) because $d\gamma/dT$ is not large.

The actual situation of the solid-solid interfaces becomes further complicated by the presence of gaseous ambients and the associated gas adsorption. Such adsorption significantly decreases the surface free energy by an amount equal to the heat of adsorption. As has been pointed out earlier, the heat of adsorption often represents a significant fraction of the surface free energy. Actually, the decrease in surface free energy due to adsorption can be estimated from the Gibbs adsorption isotherm:

$$\gamma-\gamma_a=\frac{RT}{AM}\int_o^p xdlnp \qquad (36)$$

where R is the gas constant, T is the absolute temperature, A is the specific surface area of the absorbent (surface area per gram), M is the molecular weight of the gas, x is the amount adsorbed in grams per gram of solid, and p is the equilibrium gas pressure.

Friction and wear are very important manifestations of the characteristics of solid-solid interfaces. However, as in other surface phenomena, here also the complexities are far too great (as is pointed out elsewhere in this conference), and consequently no unambiguous quantitative treatment is available.

When friction was first considered systematically it was attributed to the interlocking of surface irregularities (Coulomb). This mechanical aspect of friction probably plays some role in some instances; however, it has been repeatedly demonstrated that adsorbed films (of the order of a monolayer) can decrease friction significantly. At the same time, friction between two solid surfaces increases with increasing load; in fact it was noted by Leonardo da Vinci and rediscovered by Amontons in 1699 that the ratio of the friction over the load (coefficient of friction) remains constant.

Other parameters that must be considered in the friction process are the local temperature, the mechanical characteristics of the materials involved (including hardness) and, of course, the exact surface microtopographies. In considering friction and adhesion, the classical views of plasticity are not applicable if the stressed regions are too small either to contain dislocation or provide the energy to multiply dislocation moving through these regions (ref. 2, p. 236). Thus, stressed regions (of tin, for example) of the order of a micron or less can withstand stresses one to two orders of magnitude greater than the macroscopic yield stress. In the absence of deformation during the motion of two solid surfaces, it is likely that the major role is played by interatomic forces characteristic of interfaces.

Sintering takes place when solid-solid interfaces are exposed to high temperatures (about ¾ of their absolute melting point). Sintering is an activated process in which the surface tension is the driving force. It has been assumed that sintering solids behave as viscous liquids where the strain is linearly related to stress. The viscous flow must result from diffusion processes rather than from plastic flow where slipping of atomic planes takes place. Actually, it has been proposed (ref. 45) that the activation energy for sintering is the activation energy for self-diffusion. This model for sintering has been a useful one but it does not come close to resolving the complexity of the solid-solid interfaces at high temperature.

Among the "internal" solid-solid interfaces, the grain boundaries are of the greatest technological significance. The grain boundaries, most studied in the case of metals, are the regions between crystalline grains of differing orientation. They are manifestations of abrupt (or not so abrupt) changes in crystalline orientation. Like ordinary surfaces, they are lattice defects and are characterized by surface tension. Their exact nature is still not understood, yet they are of paramount importance in the overall behavior of solids and most particularly in the mechanical properties.

For many years the grain boundaries were considered as amorphous cement-like in nature, accommodating grains of different orientation. This point of view is no longer held. Depending upon the difference in orientation between the grains in question, the grain boundaries represent assemblies of dislocations ranging from very regular and well-defined ones (low angle boundaries) to very complex. One can approximate their surface tension which, in fact, is the driving force in the processes involving their motion (grain growth, for example). Their surface tension also accounts for their increased reactivity, as compared with the reactivity of the bulk material.

In some respects and under readily attainable high purity conditions, internal surfaces like grain boundaries lend themselves better to certain types of studies than ordinary solid surfaces for which the conditions of high purity present immensely greater difficulties. The energetics of surfaces and other physical properties may find more direct ultimate experimental success in some type of grain boundaries than in free solid surfaces.

ATOMISTIC APPROACH TO COVALENT MATERIALS

This section will include an illustration from the writer's researches of certain types of atomistic approximations that can be made when rigorous treatments are unavailable. Primarily covalent semiconductor surfaces will be used because such materials with their directional bonds and their relatively small coordination number lend themselves well to atomistic models. Here the atoms are considered as individual entities, and the interatomic forces are viewed as electronic interactions localized between the individual atoms.

The surfaces of the group IV elemental semiconductors, such as germanium and silicon, will first be discussed. They have the diamond structure. In considering their low index planes, it is assumed that the surface atoms undergo no structural rearrangements and that they are characterized by unpaired-unshared electrons (dangling bonds). These surfaces are schematically represented in figure 15. Several conclusions can be drawn from such a simple configuration. The tendency of these surfaces to react can be reasonably taken to be proportional to the density of the surface dangling bonds. Indeed, in certain aqueous solutions the dissolution rate of the various germanium surfaces was found to decrease with decreasing density of the dangling bonds as shown in table 4. The variation of the work function of these surfaces with the density of dangling bonds shown in the table is also of interest (ref. 46).

The tendency of dangling bonds to achieve a more stable configuration is further reflected in the electronic properties of such surfaces including the existence of surface states pointed out earlier. Thus, the *p*-type (electron deficient) character of clean germanium surfaces can be attributed to this tendency:

$$\dot{Ge} \rightarrow \ddot{Ge}^{-} + e^{+}(\text{hole}). \tag{37}$$

Crystallographically Polar Surfaces

The dangling bond configuration of the III–V and the II–VI primarily covalent compounds is not as unambiguous as in the group IV elements with the diamond structure. These compounds have the zinc blende or the wurzite structure where, as in the diamond struc-

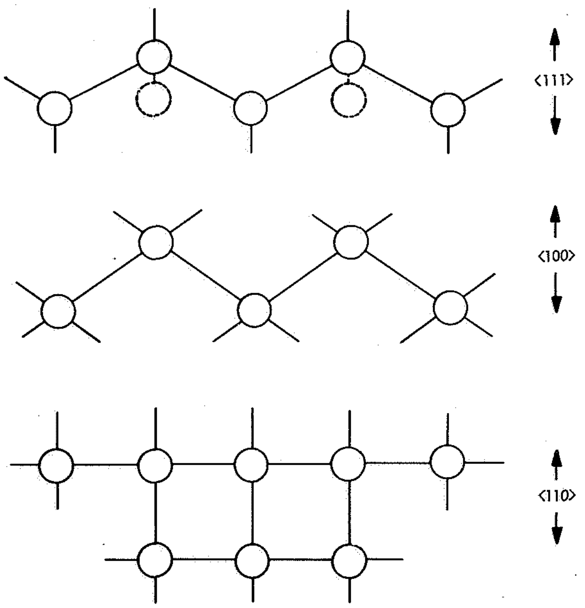


FIGURE 15.—Two-dimensional projection of the outermost and the second layer of atoms of the lower index planes showing the bonding and relative positions of the atoms.

TABLE 4.—*Correlation Between the Density of Free Surface Bonds, the Work Function, and the Dissolution Rates of Germanium in O_2-Saturated H_2O*

Orientation	Free bonds, per cm²	Relative free-bond density	Relative work function	Relative dissolution rate
(100)	1.25×10^{15}	1.00	1.00	1.00
(110)	8.83×10^{14}	0.71	0.95	0.89
(111)	7.22×10^{14}	0.58	0.93	0.62

ture, the atoms are tetrahedrally bonded, each *A* atom having four nearest *B* neighbors and vice versa. An interesting aspect of these structures is their crystallographic polarity along the [111] directions. As a result of this polarity there exist two types of (111) surfaces, those terminating with *A* atoms and those terminating with *B* atoms, as seen in figure 16. If the outermost atom layer is removed from either type of the (111) surfaces, the remaining surface layer consists of atoms with one bond toward the lattice and three unsaturated bonds. This configuration is obviously unstable; therefore, the removal as well as the addition of atoms on the (111) surfaces must proceed through diatomic steps. Accordingly, any steps on the surfaces must be at least two atoms deep. The discussion which follows will be limited to the (111) surfaces of the III–V compounds with the zinc blende structure. The (100) and the (110) surfaces are not considered here because their corresponding crystallographic directions are not polar.

The electronic configurations of the (111) surfaces are perhaps best visualized by considering a cut made perpendicular to the [111] direction between planes *AA* and *BB* as shown in figure 16. Obviously a cut between planes *AA* and *B'B'* is not feasible. It is important to know how the two electrons comprising each bond being broken will be apportioned between the newly created surface atoms. In the case of the diamond type crystals with only one type of atoms (i.e., Ge), each atom will carry with it one electron of the cut bond. In the III–V compounds, however, the isolated *A* atom has three outer shell electrons, whereas an isolated *B* atom has five. Reasoning that the electron affinity of the *A* atom is less than the work function of the *B* atom, one can conclude that the *B* surface atoms must end up with two unshared electrons, whereas the *A* surface atoms end up with no unshared or dangling electrons. Neglecting the partial ionic character of the

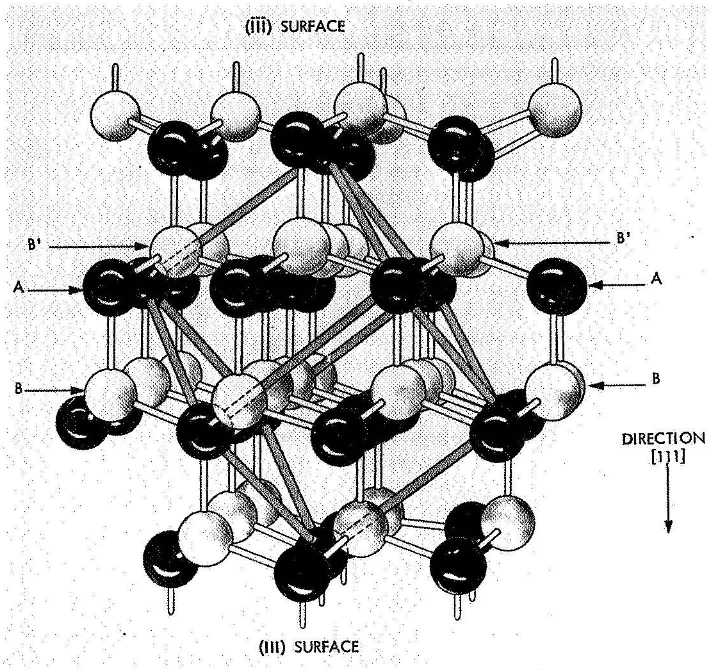


FIGURE 16.—Crystal structure (zinc-blende) of the III–V intermetallic compounds; the unit cell is indicated. •, group III atoms; o, group *V* atoms.

bonds the *A* and *B* surfaces can schematically be represented as shown in figures 17 and 18.

The chemical, structural, and electronic implications of the *A* and *B* surfaces having the above electron configuration will now be considered assuming no structural rearrangements of the surface atom of any major consequence (ref. 47).

Chemical implications.—The *A* surfaces with no dangling electrons are expected to be less reactive toward oxidizing agents than the *B* surfaces. Similarly, adsorption of electron donor molecules or negative ions should be more pronounced on the *A* surfaces than on the *B* surfaces. Conversely, adsorption of electron acceptors or positive ions should be more pronounced on the *B* than on the *A* surfaces.

Differences in chemical reactivity between the *A* and *B* surfaces were found for InSb by employing tetrahedral samples having exclusively *A* or exclusively *B* surfaces (fig. 19). The *B* surfaces were found to react faster in oxidizing media by about an order of magnitude than the corresponding *A* surfaces. This difference in dissolution rates could be altered significantly by employing surface active agents adsorbing preferentially on either the *A* or the *B* surfaces. For instance, the presence of amines in acid solutions decreased the dissolution rate of the *B* surfaces by a factor of about 10, whereas it did not affect to any significant extent the dissolution rate of the *A* surfaces.

The differences in reactivity between the *A* and *B* atoms explained successfully the observation that dislocation etch pits form on the *A* and not on the *B* surfaces of the various III–V compounds. In fact, by altering the reactivity differences through surface active agents it became possible to develop dislocation pits on both *A* and *B* surfaces. Additional chemical observations were found to be consistent with the model indicated above.

Structural implications.—The *B* surface atoms with the two unshared electrons should retain their sp^3 electronic configuration just

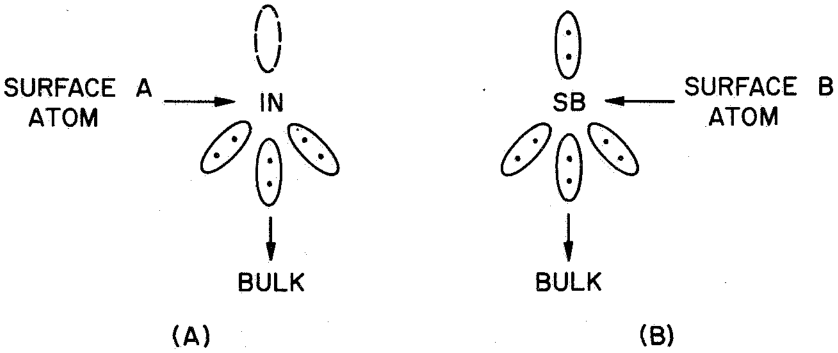


FIGURE 17.—Atomic model of the *A* and *B* (111) surfaces.

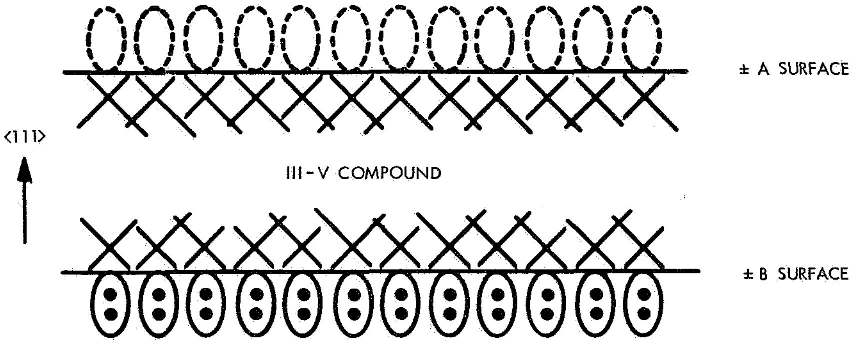


FIGURE 18.—Schematic representation of the dangling bonds on the *A* and *B* surfaces of III–V compounds and the associated surface dipole moments.

as nitrogen does in the ammonia molecule with minor deviations from tetrahedral symmetry. The *A* surfaces, however, do not have sufficient electrons for tetrahedral hybridization. Actually the trivalent *A* atoms tend to acquire an sp^2 hybridization in their AX_3 compounds, such as $InCl_3$. A planar sp^2 configuration, however, is not possible on the *A* surfaces in view of the tetrahedral angles of the bonds underneath the surface layer. One should thus presume that a strained configuration results somewhere between sp^3 and sp^2. Consequently, an elastic strain is expected to be associated with the *A* surfaces but not with the *B* surfaces.

Consistent with this deduction are a number of experimental observations (ref. 46). X-ray diffraction techniques have shown that for the same crystal of InSb the *A* surfaces are less perfect than the *B* surfaces. The *A* are more resistant to abrasion than the *B* surfaces. Crystal growth from the melt is easier in the *B* direction than in the *A* direction. Finally, it was shown that very thin InSb wafers are spontaneously bent as shown in figure 20 because of the elastic strain associated with the *A* surfaces. Actually, these experiments with the very thin (111) wafers (refs. 48 and 49) provided a quantitative approach to the elastic strain of the *A* surfaces. This strain can be calculated from the classical theory of elasticity knowing the elastic constants of the material, the geometric dimension, and the radius of curvature. The elastic energy is only a small fraction (about 10^{-6}) of the corresponding bonding energy. Calculations based on the X-ray diffraction results,

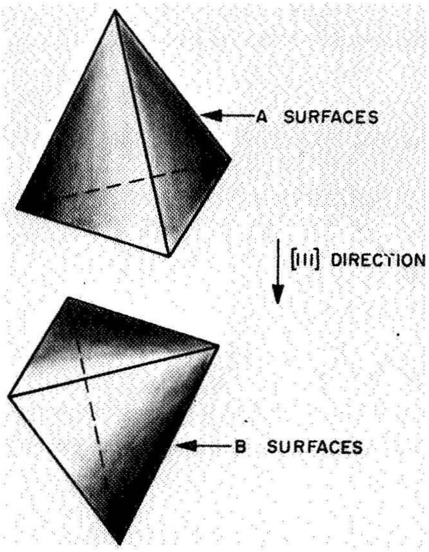


FIGURE 19.—Geometric relationship between tetrahedron *A* having exclusively group III atom surfaces, i.e., (111), $(\bar{1}\bar{1}\bar{1})$, $(\bar{1}\bar{1}1)$, $(\bar{1}1\bar{1})$, and tetrahedron *B* having exclusively group V atom surfaces, i.e., $(\bar{1}\bar{1}\bar{1})$, $(\bar{1}11)$, $(1\bar{1}1)$, $(11\bar{1})$.

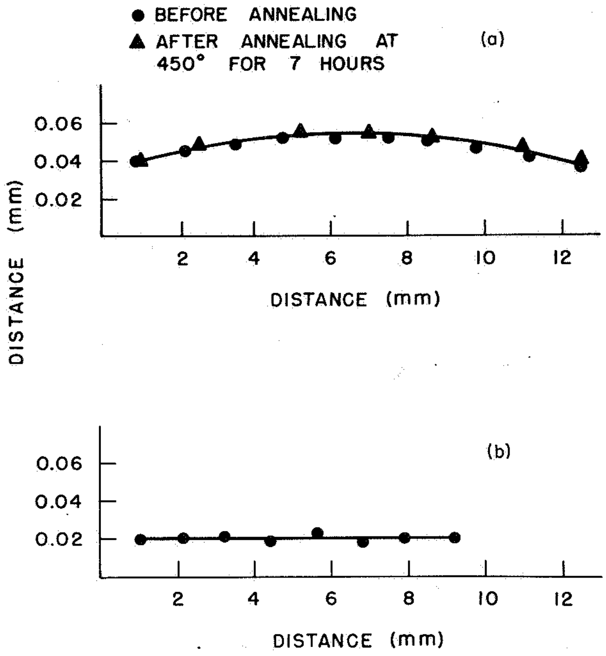


Figure 20.—(a) Curvature exhibited by an InSb (111) wafer 7.8±0.3μ thick before and after annealing. (b) Germanium wafer of the same orientation and thickness.

including some reasonable assumptions, are consistent with the findings of the bending experiments. It should be further noted that the radius of curvature of the spontaneously bent wafers can be increased or decreased by exposure to gaseous ambients that adsorb preferentially on either the A or the B surface.

Experiments with very thin and clean (111) wafers of III–V compounds in a high vacuum arrangement might yield direct information on the heats of adsorption or chemisorption of various gases using the radius of curvature as a measure.

Electronic implications.—B surface atoms with two unshared electrons are expected to act as electron donors, whereas A atoms having no dangling electrons are expected to act primarily as acceptors. Direct observation of differences in electrical behavior continues to present experimental difficulties, although some positive results are now available; however, clean triply bonded A or B atoms are present along the line of edge dislocations. The electronic configuration of the atoms along the edge dislocations should be the same as those of the corresponding surface atoms. Consequently, the electrical behavior of edge dislocations should be similar to that of triply bonded surface atoms. By employing *n*-type InSb samples in which an excess of dislocations with *A* or *B* atoms had been introduced by plastic deformation, it was shown that triply bonded *B* atoms act as electron donors whereas triply bonded *A* atoms act as electron acceptors (ref. 50).

Recent field effect experiments with InSb real surfaces (ref. 51) indicate that both types of its (111) surfaces are *p*-type, the *B* surfaces being significantly more *p*-type than the *A* surfaces. A detailed analysis of the field effect results shows that they are consistent with the present atomistic model.

CONCLUDING REMARKS

In many aspects of science and technology, the importance of solid surfaces is expected to keep increasing. There is a clear trend showing that greater and more refined demands will be made on solid surfaces in electronic, mechanical, and chemical phenomena. Thus, the understanding—basic and applied—of surface behavior will continue to confront fundamental (theoretical and experimental) and applied scientists. No single approach or discipline is likely to claim major breakthroughs in surface science in the foreseeable future. Theory, laboratory experiments, and engineering are the three main fronts along which new work must be fostered.

In scientific fields like nuclear physics and solid state physics, theory can proceed from first principles and advance separately from experiment. In the case of solid surfaces, however, theory will continue to be based on phenomenological and semiquantitative foundations, postulates, and approximations; and it is therefore in acute need of reliable basic experimental information.

In considering experimental work on solid surfaces, one must recognize the advances and shortcomings of clean and real surfaces. Neither type of surface can possibly resolve all the major surface problems. The cleaner the surface, the less one can learn about actual surface behavior; and conversely, the more real the surface, the less one can learn about its fundamental nature. It is this writer's belief that interrelating or reconciling the characteristics of the clean and real surfaces is one of the greatest challenges in surface science and potentially one of the most rewarding.

While work with clean surfaces represents one end of the experimental spectrum, at the other end of the spectrum lies surface engineering. In metallic protection, friction and wear, in lubrication, in catalysis, in conversion of chemical to electrical energy, and in many other technical areas, engineering is directly facing parameters of surface behavior. Sound engineering can, of course, be viewed as the triumph of basic and applied science; but it should not be overlooked that engineering has often played the role of the godfather to science. Here, again, an effective interplay between basic and engineering results holds promise for both.

In conclusion, I believe that success in the understanding and utilization of solid surfaces hinges on the successful interaction and correlation of theory, fundamental experimental data, and engineering results.

ACKNOWLEDGMENT

I wish to express my indebtedness to the Office of Naval Research for supporting my work on surfaces under Contract Nonr–3963(05).

DISCUSSIONS

J. J. Hren (University of Florida, Gainesville, Florida)

Perfect clean surfaces are not only of limited direct application to everyday problems involving real surfaces; they are also difficult to achieve and maintain. Yet it is essential that research efforts be stimulated to study these impractical conditions, even though it is often difficult to see how these data are likely to affect lubrication of a piston or the type of catalyst used to promote a chemical reaction. Of course this is just a special case of the classic argument about the value of basic research to an applied technology; but then that is what this symposium is largely about. There is little doubt that it often requires a giant step to connect the meanderings of basic research with pressing real problems. Although it would please me to assume the mental stature to make this step, I can only offer a tentative step in what looks like one of the right directions. In addition, I will add a few speculations that may be the beginnings of further steps.

Let me hasten to define my subject more specifically. I will concern myself with the kinds of information that can be obtained from field-ion and electron emission microscopy, namely, at the atomic or near atomic level and on clean or cleaned and intentionally contaminated surfaces with and without surface defects. These last are due to the intersection of bulk defects (dislocations, stacking faults, grain boundaries, and the like) with the surface.

Experimental conditions.—When data are obtained about surfaces via field-ion and/or electron-emission microscopes, the specimens have been subjected to a rather unique set of boundary conditions. Of course the information obtained is unique as well, but it must be interpreted with full knowledge of its special applicability. Let us first examine these boundary conditions in some detail. In both microscopes, specimens are generally prepared from small-diameter wire or, to stretch a point, rods. To begin with, these are generally in the 0.002- to 0.010-inch diameter range. The wires are then "sharpened" to a point with a tip radius of perhaps 0.1 to 1.0 microns. This sharpening is normally accomplished by electrochemical polishing but may involve etching or even, in part, ion sputtering. In many cases (particularly in electron emission studies), the specimen is next "flash annealed" to produce a nearly atomically smooth, hemispherical surface. It is already clear that the specimen to be studied has undergone a very special set of treatments, although nearly any state may be attained at this stage if suitable precautions are undertaken. Unfortunately, these preliminary steps are neither uniform nor often reported in detail in publications. The specimen may undergo yet another preparation step; in fact, field evaporation is mandatory in the case of FIM. Field evaporation is the process of removing atoms from the surface under the influence of an electric field with only a limited amount of thermal energy available ($kT > 10^{-2}$ eV). The process is possible because of the enormous electric field at the surface (10^7 to 10^8 V/cm) which reduces the energy barrier for ionization of the surface atoms and permits them to be evaporated. These fields come about, of course, because of the very small diameters which are studied. Thus the field strength is given by:

$$F = F_o\left(\frac{R}{r}\right)^2 \tag{38}$$

and the stresses present normal to the surface are:

$$\sigma_N = \frac{F^2}{8\pi}. \tag{39}$$

In these expressions, F is the field strength at some distance r, F_o is the field at the surface, R is the local tip radius, and σ_N is the stress normal to the surface. During field evaporation, the stresses are of the order of 1000 kg/mm ^{2}. At first glance it is amazing that the material is able to withstand such stresses without being deformed plastically into a pretzel; however, the saving factor is the nearly hemispherical specimen form. Thus, the stresses within the material are substantially hydrostatic. The shear stresses are often sufficient to cause gross deformation, however, and may lead to failure of the specimen. If the field is increased sufficiently, field evaporation may occur. This process is highly controllable (can be viewed literally on an atom-by-atom basis) and is a means of dissecting the specimen atom by atom. Thus although only the surface is "imaged" in any given pattern, a substantial volume may be investigated using a series of observations. Another interesting, and useful, feature is that the process of field evaporation produces a truly clean surface. Thus if a typical field-ion study were conducted, the specimen would be at 20° K, and the high field required for imaging (not as high as for field evaporation, but only about a factor of 2 lower) would "protect" the surface from impurities—they would ionize and be repelled from the specimen before they reached the surface. If an ultrahigh vacuum system were employed (10^{-10} torr or better), the field could be switched off and the surface would remain clean for extended periods. The similarity in the requirements for field-ion microscopy and field-electron emission microscopy thus suggests a number of potential experimental combinations yielding several valuable bits of information about clean or nearly clean surfaces. These "combination" experiments offer great promise of gathering unique experimental information about surface structures. I will describe several studies of this kind which have already yielded significant information about surfaces and will suggest several more that are well within the present state of the art and only remain to be done. Let us first specify what the vital equipment requirements are for each technique.

For field-electron emission, the specimen must have a small radius of curvature (less than 1 micron) and must be placed at a high negative potential in an ultrahigh vacuum system (10^{-10} torr). For field-ion microscopy, the specimen must be somewhat smaller (about 1/10 micron) and must be held at a high positive potential in a high vacuum system (10^{-10} torr). By meeting the most stringent requirements of each, 1/10 micron radius and 10^{-10} torr vacuum, the specimen may be examined both with the FIM and the FEEM by respectively using cryogenic cooling, an imaging gas and a positive potential, and ultrahigh vacuum and a negative potential. Coupling these basic ingredients with a host of possible specimen treatments, e.g., heating, vapor deposition, impurity adsorption, and bombardment with various particles, results in a wide experimental capability for fundamental research on many properties of surfaces.

Contrast mechanisms in electron emission and field ionization.—In general, the kinds of information attainable from FIM and FEEM and pertinent to the properties of surfaces may be deduced from a knowledge of their respective contrast-forming mechanisms. In the case of field-electron emission, electrons must tunnel through a surface barrier determined essentially by the work function ϕ, modified by the influence of a strong applied field F, taking into account the modification introduced by the image potential of an electron at a distance x_c from a conducting surface. These factors are illustrated schematically in figure 21. The probability that an electron incident on the surface may tunnel through this surface barrier is then converted into an expression for the field-emitted current density, J, by integrating over the energy distribution of the electrons incident on the surface barrier from within. The result was first derived by Fowler and Nordheim (ref. 53).

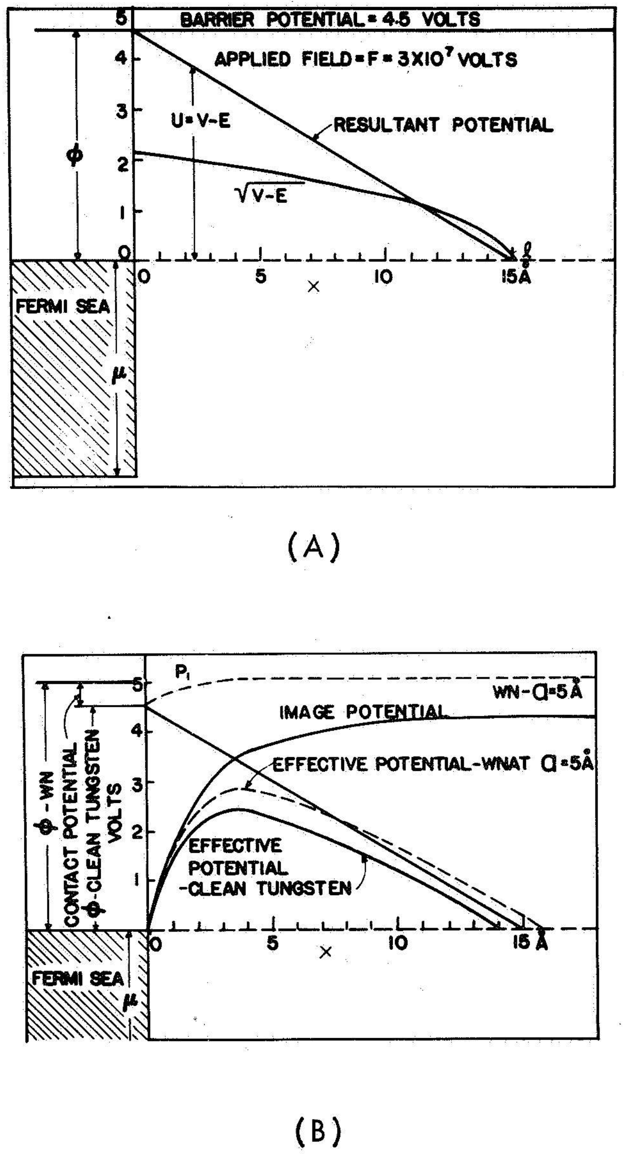


FIGURE 21.—Potential energy diagram for electrons at tungsten surface with an applied field: (a) triangular barrier without inclusion of image potential $-e^2/4x$, (b) with image potential included giving Schottky hum; broken lines show effect of adsorbed gas, here indicated as nitrogen on tungsten (*WN*). After Gomer (ref. 52).

$$J=6.2\times 10^{6}\frac{(\mu/\phi)^{1/2}}{\mu+\phi}F^{2}\exp\left(-6.8\times 10^{7}\frac{\phi^{3/2}}{F}\right)\text{ amp/cm}^{2} \tag{40}$$

where μ is the Fermi energy (in eV) and F the field (in V/cm). The dominant factor in determining the resulting current density is the exponential term $\phi^{3/2}/F$, and the field-emission image is thus very sensitive to local variations of the work function or field strength. The work function is of fundamental interest and varies with crystallographic orientation over the curved surface of the emitter. It is also sensitive to changes in composition at the surface due either to adsorbed impurities from the gaseous phase or impurities diffused to the surface from the interior of the specimen. This, of course, makes the field-electron emission microscope a very sensitive tool in studies of adsorption, oxidation, and corrosion in its initial stages. The sensitivity of the field-emitted electron current to the local field strength may be particularly useful in studies of local topography, especially under conditions where the field-ion microscope may not be ideally suited (e.g., nonrefractory metals). The resolution limit of FEEM is determined by a combination of two effects: the spread of transverse moments of the electrons emitted and the diffraction of the electrons passing through the surface. A theoretical treatment of resolution yields a limit of about 20Å, although small surface protuberances, e.g., several molecules adsorbed on the surface of the emitter, could result in smaller values.

The contrast-forming mechanism in the field-ion microscope is somewhat different. The ions comprising the image are produced above perturbations in the surface potential because of its discrete atomic structure. In fact, FIM arose as a natural consequence of FEEM as the latter became reasonably well understood. For example, a standard technique in FEEM is the desorption of adsorbed atoms either by thermal activation or by applying a positive potential to the emitter. The latter process is known as field desorption, and early researchers tried to image these desorbed atoms but were unsuccessful because of the difficulty in replenishing the supply of desorbed atoms in any reliable way. In 1951 Müller (ref. 54) experimented with a conventional field-emission microscope with the specimen at a high positive potential and an intentional supply of hydrogen gas. A visible hydrogen-ion image was thus produced with resolution in the 5 to 10Å range. Shortly thereafter Inghram and Gomer (ref. 55) used a mass spectrometer to analyze these ions and deduced that adsorption of the hydrogen on the surface prior to ionization was not necessary. In fact the gas molecules could be ionized in the extremely high fields near the surface, a process termed field ionization. The use of helium as the imaging species and the recognition that cryogenic temperatures were required for optimum resolution resulted in the attainment of images of individual atoms by Müller in 1956 (ref. 56). The fields necessary for ionization were found to lie in the 2-to-6 V/Å range for hydrogen and helium, about an order of magnitude higher than required for field-electron emission. Field ionization involves not tunneling of ions, but rather tunneling of an electron from the gas into the solid. Figure 22 depicts schematically the potential energy of an electron in the imaging gas near the specimen surface with a high applied field. Because of the sensitivity of the tunneling probability to the potential barrier at the surface, ionization occurs only over a relatively small region in space above the surface atoms, only a few tenths of an Angstrom unit according to both theory and experiment. It is worth noting that the atom cannot be ionized below some critical distance, labeled x_c in figure 22(C), since the potential energy of the electron in the gas atom will then be below the Fermi level in the metal. This critical distance is from 4 to 8Å for those gases likely to be used in field-ion microscopy, so that the image points on the screen actually correspond to a picture of the peaks in the interatomic potential some distance above the surface. In addition, the imaging gas atoms are not normally ionized before colliding with the surface—although they are not adsorbed. If the gas atom is not ionized in its first pass through the field-ionization zone, it will rebound, thereby passing through again. Kinetic energy is of course lost in the collision with the surface, and the gas atom becomes effectively trapped by the emitter. It is subsequently reattracted to the surface, passing through the field-ionization zone again and spending a greater time in this zone during each succeeding pass. This process has been termed hopping and may involve 100 or more passes such as just described. The ionization and hopping processes are illustrated in figure 23.

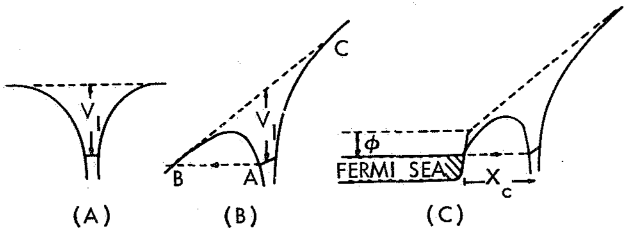


FIGURE 22.—Potential energy diagram of electron in imaging gas: (a) free atom with valence electron at depth V_1, (b) atom in the presence of external field, (c) atom close to surface illustrating critical distance x_c, Fermi level, and work function ϕ in metal.

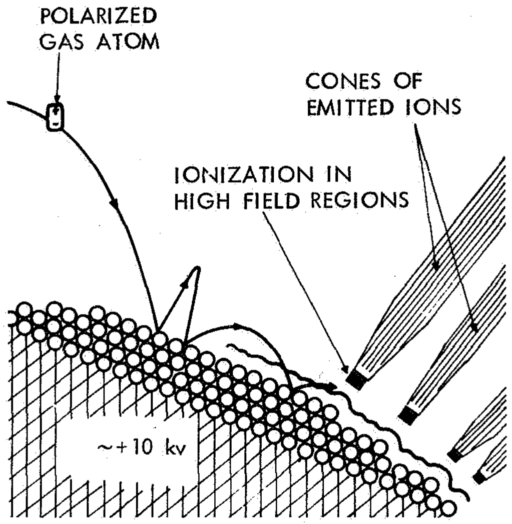


FIGURE 23.—Schematic illustration of field ionization and hopping process of gas atom over a specimen under imaging conditions. Zone of high ionization probability is indicated by heavy curve over surface atoms. Note that appreciable ionization does not occur over all surface atoms.

Finally, the process of field evaporation may occur when the applied field becomes sufficient to permit the surface atoms of the emitter to desorb. If this field for a particular material is below that required for substantial ionization of the imaging gas, it is not possible to obtain a stable image; but this need not always preclude experimental investigation via FIM.

Having described in some detail the experimental conditions pertaining to image formation and resolution, let us now examine some particular kinds of information that can be obtained using both FEEM and FIM.

Studies of surface self-diffusion.—A recent study by Melmed (ref. 57) is part of a series (refs. 52 and 57 through 67) utilizing either field-ion microscopy, field-electron microscopy, or both, to investigate the phenomenon of surface self-diffusion. Thermodynamically, the driving force for mass transport of the surface atoms is based on the higher chemical potential μ_s of a surface with smaller radius of curvature relative to a larger radius of curvature. This may be expressed as:

$$\mu_s=\mu_s^o+\left(\frac{\gamma}{r_1}+\frac{\gamma}{r_2}-\sigma\right)\Omega \tag{41}$$

where μ_s^o is the chemical potential of a flat surface, γ is the surface tension, σ is the normal surface stress (if any), r_1 and r_2 are the principal radii of curvature, and Ω is the atomic volume (ref. 52). If the surface diffusion flux J is taken to be:

$$J=\left(\frac{D_s}{kTA_o}\right)\nabla\,\mu_s\ \text{atoms/cm-sec} \tag{42}$$

where A_o is the area per atom and D_s the surface diffusion coefficient, one can arrive at an expression for the rate of change of the radius of curvature, or the blunting rate at various specimen temperatures, which may be measured quite precisely (ref. 57). Such blunting rates are found to obey Arrhenius-type equations with temperature of the form:

$$t=A\ \exp\left(\frac{Q}{kT}\right) \tag{43}$$

where A is either constant or depends upon temperature as $1/T^2$ and Q is the activation energy of the process. These studies may be conducted either with or without an applied electric field, with apparently good correlation in the deduced activation energy of the process. (Of course the effect of the field must be accounted for, but this can be done quite accurately.)

Surface defects.—There has been considerable recent concern about the structure of surface defects, particularly those produced by the intersection of a bulk defect such as a grain boundary or dislocation with the surface in a field-ion image. The intersection of bulk defects with the surface has long been recognized as significant in influencing such phenomena as crystal growth, whisker growth, catalysis, and a number of nucleation processes. A model has been developed for the growth of whiskers, for example, based on a single screw dislocation up the axis (ref. 68). Experimentally, researchers have had some difficulty testing this and other hypotheses. FIM and FEEM, however, offer prospects of very direct ways of testing the theories and perhaps suggesting other explanations where required. In the case of whiskers, a considerable effort has been expended (refs. 69 through 71). A number of observations purporting to be screw dislocations were reported in the literature very early (ref. 72), using either or both microscopic methods. One rather ingenious experiment was reported by Strayer, et al. (ref. 72), where a pulsed field-emission microscope was used to minimize the influence of the field stresses on the dislocations within the sample. In addition, the specimens were heated so that surface migration could be observed to and from the unending spiral steps formed by the intersection of the dislocations with the surface. By applying a steady electric field independent of the pulsed imaging field and large enough to equal or exceed the stress due to surface tension, no surface migration to or away from the defect occurred because of the difficulty in nucleating new atom layers at the step. On the other hand, with no steady applied field, addition or removal of material could be observed at the dislocation intersection with the surface. This manifested itself in the spiral revolving about the point of intersection in one sense or the other. Although the work just described was done on tungsten, the authors point out that such a study could be accomplished via FEEM on any suitable emitter material.

Other less elaborate investigations have been reported from time to time, but all have been characterized by a lack of quantitative analysis. Most often spiral structures were simply described, without much attempt at analysis, as resulting from the intersection of screw dislocations with the emitter surface. In fact, however, if one looks through the basic literature on the geometry of dislocations, one finds that any dislocation can give rise to an "unending step" or spiral ramp when it intersects a surface. For example, Cottrell (ref. 73) states:

> An extremely important property of a dislocation line is the following one. Suppose that we have a family of parallel planes and project through them a dislocation line with a suitable Burgers vector which also projects through them. Then the structure is changed by the dislocation from a family of parallel planes into a single spiral surface, or a helicoid. This property is true of edge, screw, and mixed dislocations, but can be seen most easily in the case of screw dislocations which project vertically through the family of planes.

The classic picture of a screw dislocation projecting vertically through a set of parallel planes is shown in figure 24; however, one can also visualize a helicoid formed by a mixed dislocation as shown schematically in figure 25. In fact, there are really only two possibilities: either the dislocation intersecting a surface (or parallel set of planes) creates an unending ramp (which implies that the Burgers vector **b** is inclined at some angle to the plane on the surface) or it disturbs the surface only in the immediate vicinity of its core (in which case **b** lies in the surface). In either case, the type of dislocation, i.e., screw, edge, or mixed, need not be specified except for differences in core configuration. These two possibilities may be described concisely by the two vectors **g** and **b**, where **g** is the reciprocal lattice vector corresponding to the set of planes intersected by the dislocation line. The dot product, $\mathbf{g}\cdot\mathbf{b}$, always yields an integer for whole dislocations, the magnitude of which determines the size of the unending step in terms of the interatomic spacings, **d**, of the planes intersected. This may be visualized more easily in terms of real lattice vectors alone, for the case

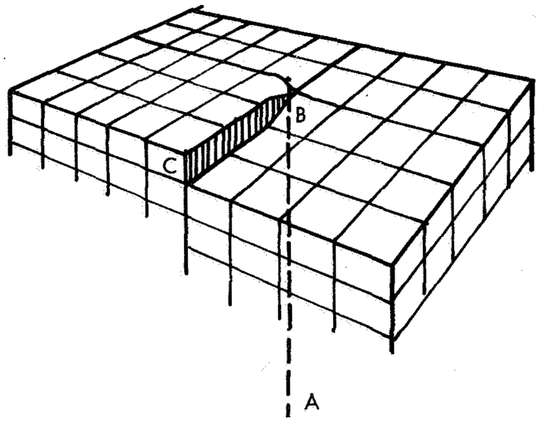


Figure 24.—Schematic of pure screw dislocation, *AB*, intersecting set of planes perpendicular to its line. Note that an unending step is produced, *C*, of height 1 Burgers vector. Planes parallel to the surface are deformed into a helicoid.

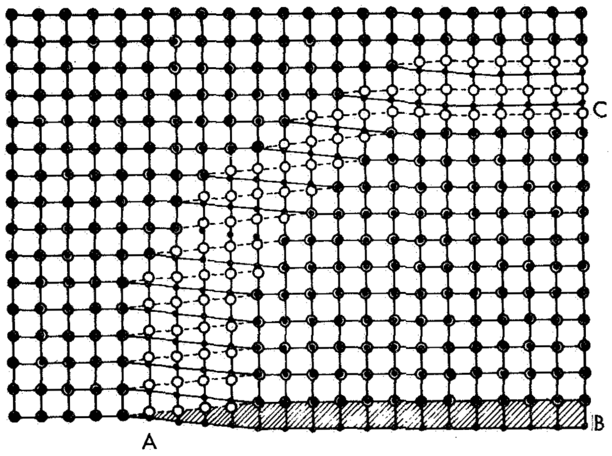


Figure 25.—Curved dislocation *AC* viewed from above in simple cubic structure. Shaded area corresponds to the same kind of step as depicted in figure 24, since dislocation intersects the surface at *A* normal to it. Note that at any position between *A* and *C*, a helicoid is also produced even though the dislocation line is of mixed character, i.e., neither pure screw nor pure edge.

of a cubic crystal structure, as follows. The reciprocal lattice vector $\mathbf{g}_{hkl}$ may be represented by:

$$\mathbf{g}_{hkl}=\frac{\hat{n}_{hkl}}{d_{hkl}} \tag{44}$$

where $\hat{n}_{hkl}$ is a unit vector normal to (hkl) and d_{hkl} is the interplanar spacing. Using the well-known relationships:

$$\hat{n}_{hkl}=\frac{[hkl]}{(h^2+k^2+l^2)^{1/2}} \tag{45}$$

and

$$d_{hkl}=\frac{a_o}{(h^2+k^2+l^2)^{1/2}} \tag{46}$$

where a_o is the unit cell dimension, equation 44 becomes:

$$\mathbf{g}_{hkl}=\frac{[hkl]}{a_o}. \tag{47}$$

Since $\mathbf{b}$ is given in terms of a_o, e.g., $\mathbf{b}=a_o[uvw]$, the dot product $\mathbf{g}\cdot\mathbf{b}$ yields $(hu+kv+lw)$, which is always an integer or zero if $\mathbf{b}$ is a lattice vector. The dot product may be thus thought of as giving the projection of $\mathbf{b}$ on $\hat{n}_{hkl}$ in units of d_{hkl} (illustrated in figure 26) where it may be observed that two different Burgers vectors of the type $\frac{a_o}{2}[110]$ (the usual case for FCC crystals) can yield different sized steps on the surface. Notice that no mention is made of the line direction of the dislocation here, i.e., the character of the dislocation has no influence on the size of the step produced on the surface. If the dot product is zero, the Burgers vector lies in the plane of the surface and no step is produced although small local displacements in the immediate vicinity of the intersection of the dislocation with the surface will exist.

The conclusion to be emphasized is that, except for the relatively rare case where the Burgers vector of a dislocation intersecting the surface lies in the plane of the surface, all dislocations will produce an unending surface ramp which extends over large regions of the surface. Thus if the surface were truly atomically flat, the step would proceed from the dislocation line to the edge of the piece, much like the picture in figure 24. If the surface has a long-range curvature, such as in a field emitter, the step takes on the character of a spiral when viewed from above. The sense of the spiral will be determined by the direction of the Burgers vector, i.e., whether it points out from or down into the surface. These effects are illustrated schematically in figure 27.

The predictions of this analysis may be tested directly with a computer model, i.e., since the **g·b** criterion will describe the nature of the step to be expected upon intersection of any set of crystallographic planes with an arbitrary dislocation and since a good mathematical description of the atomic displacements around a dislocation are available in the literature, the geometry of the unending ramp may be accurately simulated in a digital computer. A typical example of simulation of a field-ion specimen is given in figure 28. The resulting spiral is exactly that predicted by the **g·b** criterion, and this has been true in every case tested (refs. 74 through 76).

The spirals are not always so obvious in a field-ion image, particularly if the dislocation intersects the surface near a high index region (i.e., between prominent, low-index crystallographic planes). Figure 29 is a field-ion image of two closely spaced dislocations intersecting the surface near a (113) plane, and figure 30 is the computer simulation of the image using the principles outlined above. Note that the two dislocations together give a triple spiral (i.e., an unending ramp that is three *d*-spacings of the (113) planes), whereas no single dislocation of the type $\frac{a_o}{2}[110]$ could account for this observation.

In a similar way, the intersection of dissociated dislocations (including the stacking fault) with arbitrary crystallographic planes may

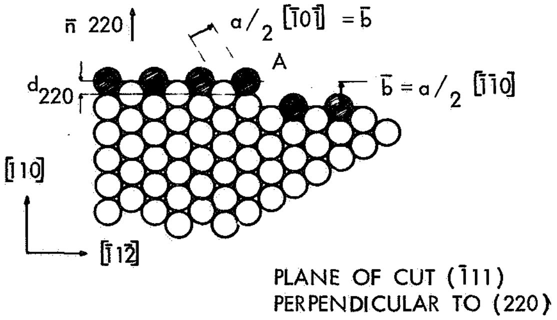


FIGURE 26.—Diagram of ($\bar{1}$11) plane showing the production of a step of either one or two interger *d*-spacings of (220) planes depending on whether the Burgers vector of intersecting dislocation is $\frac{a_0}{2}[\bar{1}\bar{1}0]$ or $\frac{a_0}{2}[\bar{1}0\bar{1}]$.

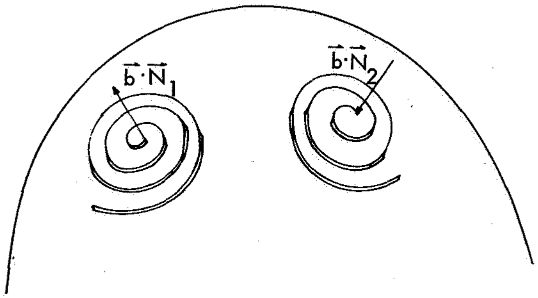


FIGURE 27.—Schematic illustration of the production of clockwise or counterclockwise spirals on hemispherical surface depending on the sign of the Burgers vector projected on the surface normal.

be analyzed using the **g**·**b** criterion and computer simulation combined. Although each partial dislocation alone will not yield a step (or spiral for field emitters) with an integral number of *d*-spacings, the two partial dislocations together yield a whole dislocation with a Burgers vector from one lattice point to another. Thus the long-range effect is an integer *d*-spacing step as just discussed. Only the region between the partials, containing the stacking fault, produces a noninteger step, but even here there is a small step nonetheless. An example of a computer simulation for two partial dislocations (with associated stacking fault) intersecting the surface of a field emitter is given in figure 31. Note that the net long-range effect is a single spiral, as would be predicted for a total dislocation $\frac{a_0}{2}$ [101], and that there is a displacement parallel to the stacking fault between the two partial dislocations. On a perfectly flat surface no spiral would be observed, but the step would proceed to the end of the specimen. In this case the intersection of the stacking fault with the surface would produce only a step of height $d/3$.

Quantitative interpretational work of the kind just described on surface defects in field emitters has been performed only very recently. Ranganathan (ref. 77) first proposed the **g**·**b** description of the spiral surface steps in order to explain the dislocation structure of grain boundaries. Sanwald et al. (ref. 74) verified the validity of the relationship using computer simulation. Sanwald and the author (refs. 75 and 78) were able to reproduce actual experimental images made with the field-ion microscope by computer simulation, thus conclusively establishing the validity of the analysis. Sanwald and Hren (refs. 79 and 80) have also extended the analysis to BCC and HCP as well as FCC, including studies of stacking faults and loops. Newman (ref. 81) has since used combined field-ion microscopy and transmission electron microscopy to study and explain quenched-in defects in platinum using the same approach.

Future studies on surface defects.—Because surface defects of the kind just described (including others not specifically discussed such as anti-phase boundaries in ordered alloys and interface dislocations) can be quantitatively interpreted in field-ion and electron emission patterns, a considerable body of possible experiments of fundamental interest can be meaningfully tackled. I will mention a few here, very briefly because of their speculative nature.

It is easy to speculate that the intersection of a dislocation with a set of crystal planes will alter the work function. It would appear that this might be difficult to measure, however, because the local topography has also been changed, thereby affecting the local field strength F as well. This might be amenable to study if dislocations yielding different step heights can be obtained on the same kinds of planes in different specimens. On the other hand, a single dislocation, if it is inclined to the axis of the specimen, will intersect a number of different planes if the specimen is field-evaporated between electron

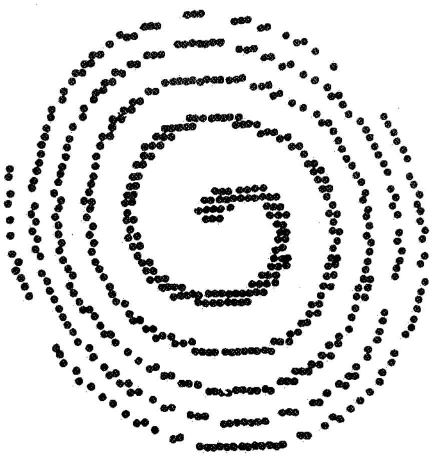


FIGURE 28.—Computer simulation of spiral configuration of dislocation with $\mathbf{b}=\frac{a_0}{2}$ [110] intersecting (204) planes of hemispherical cap approximating field emitter. Long-range effect is the same regardless of direction of dislocation line.

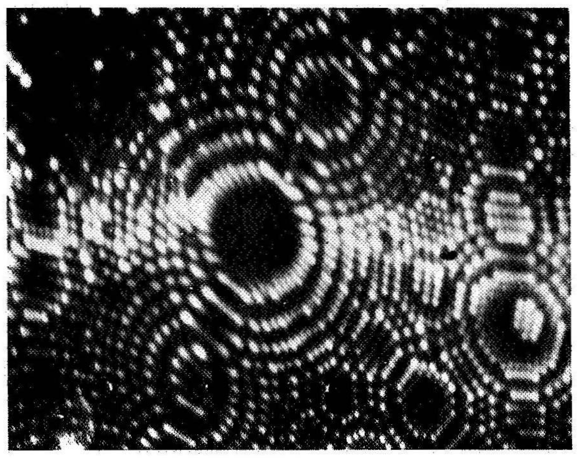


FIGURE 29.—Field-ion micrograph near (113) plane of an iridium specimen showing triple-leaved spiral due to intersection of surface by two dislocations.

emission measurements. Thus the effect of the dislocation on the emission current of a number of crystallographic planes may be studied in one specimen.

The effect of surface defects on adsorption and desorption are also of obvious fundamental importance. One might easily imagine a series of experiments something like the following. A dislocation intersecting a specific set of planes is analyzed via FIM as described above. The work function is measured on the planes of intersection and other perfect planes of the same type on the emitter. A small amount of gas is allowed to enter the microscope in small infinite increments, and the rate of adsorption is measured on the plane with the defect and on the others. Desorption fields are measured making the same comparisons. The specimen is field-evaporated so that the defect intersects another crystallographic set of planes, and the studies are repeated.

Still another prospective study would involve the measurement of surface self-diffusion, again comparing the planes with defects and those without. Similar studies may be made on the nucleation and growth of thin films—particularly nucleation at the dislocation. The heights of the surface steps would no doubt be a crucial parameter in such experiments.

Perhaps the most gratifying thing about studies of the kind just suggested is that they are at least a small step in the direction of understanding something about real surfaces. Clearly, real surfaces are much more complex than the conditions I have considered in field emitters;

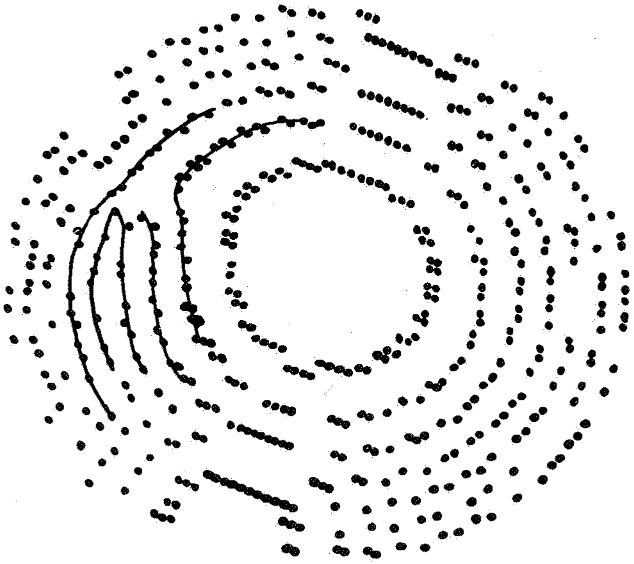


FIGURE 30.—Computer simulated image corresponding to figure 29, with dislocations having $\mathbf{b}_1=\frac{a_0}{2}$ [011] and $\mathbf{b}_2=\frac{a_0}{2}$ [0$\bar{1}$1].

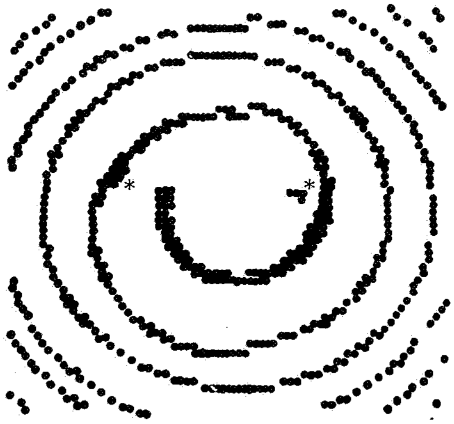


FIGURE 31.—Computer simulation of a pair of Shockley partial dislocations and associated stacking fault intersecting (220) planes in field emitter. Total Burgers vector $\mathbf{b}=\frac{a_0}{2}$ [101].

*Point of emergence of each partial dislocation.

however, they most certainly do contain defects of the kind discussed. The steps caused by the intersections of many dislocations and other bulk defects with the surface undoubtably provide preferential sites for reactions with real environments. These reactions (e.g., oxidation, adsorption, surface migration) can be investigated individually and can serve as a base to predict large-scale effects.

ACKNOWLEDGMENT

The author gratefully acknowledges the use of material from the work of R. W. Newman and R. C. Sanwald and the assistance of Mrs. D. B. Finder and Mrs. S. C. Soong in the preparation of the manuscript and figures. Much of the work in the last section of this paper was supported by the National Aeronautics and Space Administration under grant number 10-005-039.

D. Ramadanoff (Advanced Materials and Technology, Incorporated, Royal Oak, Michigan)

The author is to be congratulated for his brief, clear, and rather complete survey of the literature dealing with the structure and behavior of solid surfaces. More important, the treatment of surface or interface phenomena has been imaginatively presented to cover basic and applied technology and should for this reason be useful to workers in both fields. Dr. Gatos has discussed at length the energetics of clean surfaces such as can be obtained in ultrahigh vacuum and real surfaces where the interactions with the ambient can be gases, liquids, or solids. Finally, the paper adequately covers the utility and limitations of field-emission and ion-emission microscopy, low energy electron diffraction, and ion microprobe mass spectrometer as applied to clean and real surfaces.

In considering fundamental and applied experimental work on solid surfaces, Dr. Gatos recognizes the advantages and shortcomings of the clean and the real surfaces and that neither type of surface can possibly resolve all of the major surface problems. The clean surface teaches us little about actual surface behavior such as occurs in successful practical applications, while the real surface with its complexity does not lend itself to investigating its fundamental nature. Therefore, reconciling the experimental behavior of clean and real surfaces and making them mutually compatible and useful to a better understanding of what happens at the surface is the real challenge facing us today.

Surface engineering must deal in an effective manner with many phases of surface phenomena. Thermionic and photoelectric emission, friction and wear for sliding surfaces in ordinary environment and in high vacuum such as required by ultraspace, lubrication, catalysis, metallic protection, conversion of chemical to electrical energy are a few of the areas for which creative engineering must provide satisfactory solutions.

The purpose of this discussion is not primarily to bring out differences in the authors' points of view, but rather to sight a common parameter which exists between efficient photoelectric and thermionic surfaces and those used for sliding with or without the flow of electric current. The surfaces to be discussed may consist of a thin film from a single metal, such as barium, suitable for high-vacuum applications only, or complex composite surfaces suitable for high-vacuum as well as gaseous environment. The common parameter in all these instances is the electronic work function, which by our present standards is low.

The formation of composite surfaces can be intentional or purely coincidental. The atoms of freshly formed surfaces have unsaturated bonds and have a strong tendency to react physically or chemically with their environment and form composite surfaces. Therefore composite surfaces are heavily populated with foreign atoms. From the practical or utilitarian point of view, it matters much what type of foreign atoms form these composite surfaces. In the case of thermionic emission, for example, the efficiency and general performance of electron emittors is considerably enhanced if pure tungsten filament is replaced with thoriated tungsten, and pure platinum with platinum wire coated with the oxides and sub-oxides of Ba, Sr, and Ca.

In photoelectric phenomena, the skillful preparation of composite surfaces with pure, thick films of the alkali-metals group such as Li, K, Na, Rb, and Cs, is a must if one wishes to get good and efficient photocells. Yet when sliding surface phenomena with carbon as one component is involved, for which low friction and wear are of the utmost importance, the presence of foreign atoms from the alkali group, either metallic or salts, is most unwanted. Lithium in this respect has shown limited exception.

Finally, the presence of atoms from the metals and compounds of the alkaline-earth group such as Ba, Sr, and Ca, and rare-earth group such as Ce and Th can be used effectively in all three instances: thermionic emittors, photoelectric emittors, and sliding surfaces.

In discussing field emission microscopy and its application to the study of surfaces of solids, Dr. Gatos has described the bright areas in the electron emission photograph of figure 3 as those corresponding to high electron emission and relatively low work function. These are the areas that correspond to high index planes. The low index planes, which appear as dark areas, should have higher work function. This photograph was obtained with a pure tungsten point. Indeed it will be most interesting to see a photograph that has been obtained with thoriated tungsten point which has a lower work function than a pure tungsten point.

The work function of a tungsten point can be materially lowered also by condensing on it a thin film of barium metal. But a condensed thin film of barium on the rubbing surfaces of a ball bearing in an X-ray tube has been found to reduce its friction sizably. Here then are two similar surfaces, in so far as the boundary layer of barium metal is concerned, that perform efficiently, each in its own way, in two entirely different phenomena, electron emission and mechanical rubbing. The property common to both of these surfaces is, of course, low work function.

Practical experience has shown that good, efficient thermionic and photoelectric emittors can be made in high vacuum with simple or complex-composite surfaces. Such surfaces can be prepared also in high vacuum and in identical manner for the purpose of low friction, low wear mechanical sliding.

In their present studies of sliding friction and wear of different material combinations, many investigators use a stationary hemispherical ball and a revolving disc in a well evacuated enclosure. Such an apparatus should be very useful for simultaneous measurements of photoelectric emission of the track, and friction and wear of the slider. The track can be cleaned of adsorbed gases with electron bombardment and then coated, while revolving, through evaporation with a thin coat of metallic barium. In fact it is possible to bombard the track during the coating process with barium ions. Such procedure will tend to produce a uniform track surface suitable for more meaningful measurements. Similar procedure can be used for coating the track with more complex material such as the oxides, suboxides and salts of barium, strontium, and calcium; or the metals, oxides, and salts of the rare-earth elements. The photosensitive sulfides, selenides, and tellurides should also be examined and sorted out.

The question now is how to measure the work function of a complex, composite surface. As a first approximation, one should perhaps be satisfied with the measurement of the total photoelectric current with a constant source of illumination. Thus the surface of the sliding track can be examined in high vacuum simultaneously as a sliding surface with or without the flow of electric current and as a photoelectric surface, and all of this without breaking the vacuum of the enclosure. Such an experiment may shed new light on the commonality of these two seemingly different and unrelated phenomena.

The writer hopes that the experimental approach discussed in this study of surface behavior is new and useful. Indeed, in the study of surface behavior there is a need for combining basic and practical measurements, particularly when they can be done on good, identical surfaces and at the same time. It will be especially desirable if these measurements can be made on efficient and practically useful surfaces.

The basic quantitative methods which have been developed for crystalline solids, much as a result of atomic periodicity and symmetry, are no longer applicable to surface or interface phenomena. The surfaces represent an abrupt termination of such periodicity and symmetry; yet, basically, it is through the surface that one must communicate with the solid.

To complicate matters further, freshly uncovered surface atoms have unsaturated bonds and are extremely reactive chemically and physically with the environment to which they are exposed. From the basic scientific point of view, the complexity of the crystalline surface, once it has reacted with its environment, be it gaseous, liquid, or solid, is enormous and interposes so many difficulties and uncertainties as to make a theoretical or even a basic experimental approach look hopeless. But it is in the complexity of this surface phenomena that one senses nature's way of providing us with an infinite number of possibilities by means of which good, practically useful, complex surfaces can be produced. That complex surfaces work better than freshly created ones, in whatever practical endeavor, is almost axiomatic.

Dr. Gatos is not only mindful of the things that have been said here but has discussed in some detail film formation due to gases, liquids and solids. Difficulties in the experimental study of thick surface films several atom layers thick, with electron and ion microscopy and low energy electron diffraction, are also mentioned. Yet this writer has had much success in experiments with sliding surfaces, using thick films with or without the flow of electric current. Desirable thick films can be viewed generally as films of surface pacification and of separation, reducing friction and wear in sliding surface phenomena to very low values. But it is with thick films that an investigator finds himself hopelessly outmaneuvered when he tries to make basic measurements.

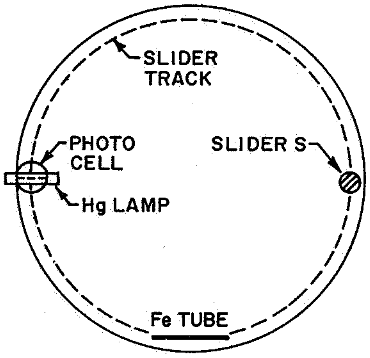
Dr. Gatos has pointed out that the principal prerequisite for the successful use of real surfaces is that they be prepared reproducibly. Perhaps it may be helpful to mention here that films formed on solid rubbing surfaces, particularly with the flow of electric current, are not static. On the contrary, they should be looked upon as dynamic where the film is constantly broken down and repaired. Only when the rate of film repair equals the rate of film destruction can there be stable equilibrium and satisfactory performance.

This discussion will not be complete without mentioning the profound effect which certain gaseous atmospheres have on friction and wear when the sliding components are composite graphite and copper. A pure piece of graphite, when pressed with a force of 8 lb/in.^{2} against a revolving ring moving about 5000 ft/min, with or without the flow of electric current, will seize and wear very rapidly or dust with high coefficient of friction if the surrounding atmosphere is dry air or any of the following gases in dry form: nitrogen, carbon dioxide, argon, and hydrogen. If this piece of pure graphite is made again, if to its composition, before molding and baking, has been added a small amount of carbonate of the alkaline-earth group and a trace of a refractory nitride, and if the experiment is repeated with the above dry atmospheres, the rapid wear will no longer occur and the coefficient of friction will be acceptably small. There is a big difference, however, in performance of the sliding components containing the composite graphite if dry air atmosphere is replaced with dry carbon dioxide. In dry carbon dioxide atmosphere, the wear of the graphite composite decreases by a factor of 50 to 100, and the coefficient of friction decreases by a factor of 3. Since these are lengthy experiments, the beneficial effect of this specific rubbing system, when dry air ambient is replaced by dry carbon dioxide, is apparent.

Details of the suggested method for measuring simultaneously the photoelectric emission and friction and wear of a mechanical rubbing system are shown in figure 32. The plate *P* and the hemispherical slider *S* can be made from mutually compatible metals or alloys, preferably of the type which will be useful scientifically and practically. After proper cleaning of the sliding surfaces of *P* and *S* before assembly and after all component parts are assembled and evacuated as shown in figure 32, an attempt should be made to clean the track surface of plate *P* by electronic bombardment in high vacuum while plate *P* is revolving at a reasonably high surface speed such as 5000 ft/min. This will tend to produce a more uniformly "sensitized" or active sliding surface and minimize the possible formation of different surface areas or patches. This surface then becomes the substrate for the metallic barium coating to be deposited through evaporation.

Without breaking the vacuum and with the pump operating, the track surface of plate *P* can now be coated with pure barium metal by heating a tubular iron wire, 0.025 inch in diameter, with 5 to 7 amperes. This wire tube is filled with barium-aluminum getter alloy designed to evaporate pure barium metal when the tube is heated with the above-mentioned electric current. (One possible source of supply for this wire is the Electronics Division of Union Carbide Corporation in Cleveland, Ohio.) The barium coating should be deposited while the plate *P* is revolving at reasonably high speed. This will tend to produce the most uniform barium coating.

This barium surface can now be examined photoelectrically for uniformity with the slider *S* off the sliding track and the plate *P* revolving at high speed. Any variations in the photoelectric current resulting from surface nonuniformity can now be separated from the steady direct-current component as shown in figure 32 and described briefly in references 82 through 84. Once the alternating-current component of the total photoelectric current is separated, it can be amplified and observed visually on a cathode-ray oscilloscope. If the beam of the ultraviolet light is strong enough to be focused in the form of a thin line parallel to the radius of plate *P*, the sliding track of the plate can be examined for surface uniformity in considerable detail through its photoelectric response along the whole circular path of the sliding surface. Any changes in the photoelectric sensitivity of the sliding track as the rider *S* begins to bear on it will also become apparent.



TOP VIEW OF PLATE P

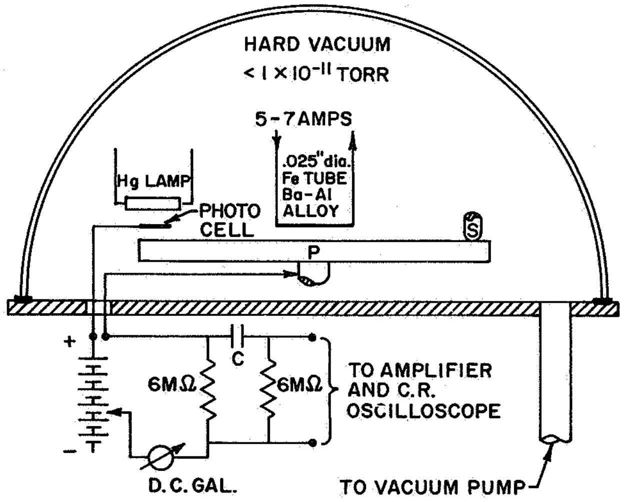


FIGURE 32.—Apparatus for simultaneous measurement of sliding friction and photoelectric emission of a barium surface.

Now, the active wear track can be monitored simultaneously for friction, wear, and photoelectric emission. Any changes in the photoelectric sensitivity of the activated slider *S* track during rotation can be seen and/or recorded.

Should the above method of examining a sliding surface be found useful, then it is possible to coat the sliding track with composite materials derived from the compounds of the alkaline-earth and rare-earth chemical groups and with suitable sulphides, selenides, and tellurides.

D. H. Buckley (NASA Lewis Research Center, Cleveland, Ohio)

Dr. Gatos is to be complimented for bringing to this conference a paper that has set the stage for the papers and discussions to follow. It has helped to orient our thinking and focus our attention on the nature of real surfaces and the means at our disposal for studying these surfaces.

I have two questions which arose in reading Dr. Gatos' paper. The first deals with surface energy and the second with the nature of clean surfaces as determined by LEED. In regard to surface energy, does Dr. Gatos feel that enough reliable measurements have been made in the area of surface energy to warrant its use at the present time in fields such as friction and wear? Even theoretically calculated values do not seem to agree. By way of example, Dr. Gatos presents in table 2 of his paper surface energy values of 5400 and 9140 ergs/cm^2 for the (111) and the (100) planes of diamond, respectively. Gilman (ref. 85) indicates values of 3500 for the (111) planes and 7050 for the (100) planes. Bowden and Tabor (ref. 86) report approximate values of 5000 and 8000 ergs/cm^2 for the same two planes. These differences do not disturb me nearly so much as those I obtained recently while searching for values for various planes of copper. Bondi (ref. 87) showed values for the (111) plane (2499 ergs/cm^2) which was less than for the (100) plane (2892 ergs/cm^2) as I would expect them to be. Yet reference 85 indicates values of 2980 ergs/cm^2 for the (111) plane and 590 ergs/cm^2 for the (100) plane. Here the (100) plane has a markedly lower value. My question to Dr. Gatos is what does this tell us?

In regard to my second question, I wonder if the LEED can in fact tell us when we have a really clean surface. Some months ago I was concerned with the adhesion of clean (100) planes of tungsten to themselves (matched planes and directions). Because of the popularity of ion bombardment for cleaning of surfaces in LEED work, I used this technique to clean the tungsten surfaces. I found that I could not get adhesion to occur on touch contact. Even after prolonged heating to remove the argon used in bombardment, adhesion would not occur. I should mention that the tungsten contained 10 ppm of the

TABLE 5.—*Adhesion and Friction of (100) Matched Planes of Tungsten*

Surface preparation	Adhesion on touch contact	Coefficient of friction
Argon ion bombardment	No	0. 5
Argon ion bombardment with prolonged anneal	No	0. 9
Hydrogen reduction and electron-beam cleaning	Yes	3. 0

interstitials O_2, N_2 and H_2, and carbon. If, however, I heated the specimens in hydrogen for a prolonged time and then electron-beam cleaned the surface, adhesion of the planes occurred on touch contact. These data are summarized in table 5.

These data are not meant to imply that my Edisonian approach gave a completely clean surface; but to me, one who rightly or wrongly has considerable faith in adhesion and friction measurements, it meant that the ion bombardment commonly used in LEED may not give really clean surfaces. Does Dr. Gatos have any comments in this area?

J. J. Bikerman (Horizons Incorporated, Cleveland, Ohio)

I agree with Professor Gatos when he states that "there are no direct methods for the experimental determination of surface energy in solids." I should like to add that none of the proposed indirect methods is convincing. In fact, no effect is known, no observation exists, which would prove to us the existence of a special energy of solid surfaces. Consequently, the values such as given in table 1 should not be taken seriously.

Of the values quoted in the literature, some are simply wrong. The method which employs the creep of a wire is based on the assumption that, when a wire is stretched, only its surface is involved in the process while the bulk of the wire offers no resistance to stretching. To me, this assumption is not admissible. The other values really refer to an energy related to the surface; but this energy is not analogous to the true surface energy as it is known in liquids. It may be called the cuticular energy. It exists because the superficial layer of every solid body is likely to be different from the inside matter. The former may have a different density of dislocations or simply scratches, or may differ from the bulk chemically (as the oxide on a metal surface), and so on. The method of the heat of dissolution belongs to the second group. It is clear that the heat of dissolution of the surface stratum (which consists of an oxide) must be different from that of an equal mass of the bulk which really is metal.

The whole field was reviewed by me in 1965 in the international magazine *Physica Status Solidi* (vol. 10, p. 3), and those who wish to know what the published values of surface tension mean, and what they do not mean, may be advised to read this review.

The values recorded in table 2, i.e., the calculated surface energies, are anything but certain. Depending on the model used, very different numerical values are obtained, and no one can judge which of the models is nearer the truth.

E. E. Bisson (NASA Lewis Research Center, Cleveland, Ohio)

Dr. Gatos refers to the "mobility of chemisorbed layers," pointing out that the chemisorbed layers "tend to be mobile," above a certain minimum temperature and below the temperature of desorption. The mobility or migration to which he refers is, I believe, extremely low and hence would be of limited usefulness in the lubrication process. In the lubrication phenomenon, once the liquid or solid film separating the surface is ruptured, it must be "self-repairing" in order to maintain effective boundary lubrication. The film's self-repairing qualities depend strictly on its mobility. The lubrication process of ordinary mechanisms requires almost instant film repair; this implies that the film's mobility be great. I do not believe that the mobility of chemisorbed layers is sufficient to effect a self-repairing film.

H. M. Davis (Army Research Office, Durham, North Carolina)

The word "surface" is being required to serve multiple functions. When we use it as a noun, we may refer to the geometric surface (the interface with an adjacent phase), but we are more commonly thinking also of the composition and the atomic arrangement, normal or defective, in the first few atomic layers of the material.

We use this word adjectivally with similar imprecision. When we need an adjective denoting the surface and a little of the underlying material, we could use a legitimate word which has been in the literature of geology and geochemistry for decades: we could write or speak of the surficial zone, or the surficial layer.

R. Courtel (Centre National de la Recherche Scientifique, Laboratoires de Bellevue' Bellevue, France)

I would like to ask Dr. Gatos what quantitative, simplified image of the periodical potential of a crystalline surface we presently have. Using such a representation would allow us to progress in solving a problem set by M. Brillouin (ref. 88) in 1899, which has since had no precise answer. Brillouin considers friction force to be the result of energy dissipation accompanying the passage of a series of unstable equilibrium positions of two atomic or molecular layers constituting the interface of rubbing solids. In sketching the map of the distribution of surface potentials, one should therefore be able to find the conclusion that Buckley (ref. 89) has recently given to his own experiments: that all other things being equal, friction is minimal on the planes of higher occupation (slip or cleavage planes) and, in these planes, along the directions of highest atomic density. It would be sufficient, therefore, to establish a relationship between the periodic variations of potential amplitudes (peak to valley) on the one hand, and the degree of instability of the equilibrium positions on the other, in order to justify Brillouin's hypotheses and to explain Buckley's result.

LECTURER'S CLOSURE

Dr. Hren's discussion is pertinent to our topics and shows clearly the strength of some of the recent tools which is brought to bear on the complexities of solid surfaces.

I am pleased that Dr. Ramadanoff chose to accent the importance of the electronic work function and to illustrate its significance. He is making some very interesting points.

Mr. Buckley's first question illustrates vividly the fact that, even in the simplest cases, we are not yet in the position to unambiguously determine surface energies either theoretically or experimentally. In my opinion, however, the prognosis on relevant experimental and theoretical refinements is encouraging.

Regarding the second question, the LEED can indeed tell us when we have clean surfaces. Good LEED patterns of a metal surface indicate that the surface is not contaminated, within a small fraction of a monolayer. Ion bombardment is a powerful surface-cleaning technique, but its effectiveness in each case can only be assessed by actual LEED patterns.

In answer to Dr. Bikerman, I believe that the existing values of surface energies should indeed be taken seriously; but they should not be considered definitive. These values represent our present experimental and theoretical limitations, and they should be viewed in that light.

I agree with Mr. Bisson that chemisorbed layers with relatively high energy of chemisorption are probably not useful as self-repairing films in lubrication phenomena.

With reference to Dr. Courtel's remarks, unfortunately I am not aware of any quantitative, simplified image of the surface periodic potential which might help in solving the problem posed by Brillouin. Hopefully, this question will attract the attention of some thoreticians.

REFERENCES

1. Gibbs, J. W.: Collected Works. Vol. I. Yale Univ. Press, 1948.
2. Herring, C.: The Use of Classical Macroscopic Concepts in Surface-Energy Problems. Structure and Properties of Solid Surfaces, R. Gomer and C. S. Smith, eds., Univ. Chicago Press, 1953, p. 5.
3. Mullin, W. W.: Solid Surface Morphologies Governed by Capillarity. Metal Surfaces, ASM, vol. 1, 1963, p. 17.
4. Cabrera, N.: The Equilibrium of Crystal Surfaces. Solid Surfaces, H. C. Gatos, ed., North-Holland Publ. Co. (Amsterdam), 1964, p. 320.
5. Orowan, E.: Die Zugfestigkeit von Glimmer und das Problem der technischen Festigkeit. Zeit. für Physik, vol. 82, 1933, p. 325.
6. Buttner, F. H.; Funk, E. R.; and Udin, H.: Adsorption of Oxygen in Silver. J. Phys. Chem., vol. 56, 1952, p. 657.
7. Bailey, G. L. J.; and Watkins, H. C.: Surface Tensions in the System Solid Copper-Molten Lead, Proc. Phys. Soc. (London), vol. B63, 1950, p. 350.
8. Wulff, G.: Zur Frage der Geschwindigkeit des Wachstums und der Auflösung der Kristallflächen. Zeit. für Kristallographie, vol. 34, 1901, p. 449.
9. Harkins, W. D.: Energy Relations of the Surface of Solids. J. Chem. Phys., vol. 10, 1942, p. 268.
10. Shuttleworth, R.: The Surface Energies of Inert-gas and Ionic Crystals. Proc. Phys. Soc. (London), vol. A62, 1949, p. 167.
11. Shuttleworth, R.: The Surface Tensions of Solids. Proc. Phys. Soc. (London), vol. A63, 1950, p. 444.
12. Stranski, I. N.; and Suhrman, R.: Uber die Elektronenemission kristalliner Metalloberflächen und ihre Beziehungen zu den Gesetzmässigkeiten des Kristallbaus. Ann. der Physik, vol. 1, 1947, p. 153.
13. Adamson, A. W.: Physical Chemistry of Surfaces. Interscience Publ., Inc., 1960, p. 229.
14. van der Hoff, B. M. E.; and Benson, G. C.: A Quantum-Mechanical Calculation of the Surface Energy of Crystalline Lithium Fluoride. J. Chem. Phys., vol. 22, 1954, p. 475.
15. Geguzin, Y. E.; and Ovcharenko, N. N.: Surface Energy and Surface Processes in Solids. Soviet Physics Usperkhi, vol. 5, 1962, p. 129.
16. Udin, H.; Shaler, A. J.; and Wulff, J.: The Surface Tension of Solid Copper. Trans. AIMME, vol. 185, 1949, p. 186.
17. Ewald, P. P.; and Juretschke, H.: Atomic Theory of Surface Energy. Structure and Properties of Solid Surfaces, R. Gomer and C. S. Smith, eds., Univ. Chicago Press, 1953, p. 82.
18. Heine, V.: Some Theory About Surface States. Solid Surfaces, H. C. Gatos, ed., North-Holland Publ. Co. (Amsterdam), 1964, p. 1.
19. Greene, R. F.: Surface Transport. Ibid., p. 101.
20. Wallis, R. F.: Surface Effects on Lattice Vibrations. Ibid, p. 146.
21. Bardeen, J.: Electron Correlation and Screening Effects in Relation to Surface Physics. Ibid, p. 381.
22. Many, A.; Goldstein, Y.; and Grover, N. B.: Semiconductor Surfaces. North-Holland Publ. Co. (Amsterdam), 1965.
23. Pashley, D. W.: Selected Applications of the Electron Microscope. Modern Development in Electron Microscopy, B. M. Siegel, ed., Acad. Press, 1964, p. 162.
24. Müller, E. W.: Field Ionization and Field Ion Microscopy. Advances in Electronics and Electron Physics, vol. 13, 1960, p. 83.

25. BRENNER, S. S.: Field Ion Microscopy Studies of Surfaces. Metal Surfaces, ASM, 1963, p. 305.
26. EHRLICH, G.: Adsorption and Surface Structure. Metal Surfaces, ASM, 1963, p. 221.
27. MÜLLER, E. W.: Des Feldionenmikroskop. Zeits. für Physik, vol. 131, 1951, p. 136.
28. A number of experimental and theoretical papers with reference to previous work appeared in Surface Science, ASM, vol. 2, 1964, p. 465, and vol. 8, 1967, p. 1.
29. MACRAE, A. U.; AND GERMER, L. H.: The Interatomic Spacings at the Surface of a Clean Nickel Crystal. Ann. New York Acad. of Scs., vol. 101, 1963, p. 627.
30. TUCKER, C. W.: Low Energy Diffraction Study of Adsorption on the (100) Face of Platinum. Surface Science, ASM, vol. 2, 1964, p. 516.
31. MACRAE, A. U.: Low Energy Electron Diffraction. Science, vol. 139, 1963, p. 379.
32. BARRINGTON, A. E.; HERZOG, R. F. K.; AND POSCHENRIDDER, W. P.: The Ion Microprobe Mass Spectrometer. Progress in Nuclear Energy, Series IX, Anal. Chem., vol. 7, 1966, p. 243.
33. YOUNG, D. M.: The Adsorption of Argon on Octahedral Potassium Chloride. Trans. Faraday Soc., vol. 48, 1952, p. 548.
34. TRAPNELL, B. M. W.: Chemisorption. Butterworths Scientific Publ. (London), 1955.
35. BEECK, O.; AND RITCHIE, A. W.: The Effect of Crystal Parameter on Hydrogenation and Dehydrogenation. Dis. Faraday Soc., vol. 8, 1950, p. 159.
36. KUBASCHEWSKI, O.; AND HOPKINS: Oxidation of Metals and Alloys. Butterworths Scientific Publ. (London), 1953.
37. GERMER, L. H.; STERN, R. M.; AND MACRAE, A. U.: Beginning of the Oxidation of Metals. Metal Surfaces, ASM, 1963, p. 287.
38. POTTER, E. C.: Electrochemistry: Principles and Applications. Cleaver Hume Press, Ltd. (London), 1956.
39. KORTUM, G.; AND BOCKRIS, J. O'M.: Textbook of Electrochemistry. Elsevier Publ. Co., 1951.
40. GRAHAME, D. C.: Theory of Electrocapillarity. Chem. Rev., vol. 41, 1947, p. 441.
41. GRAHAME, D. C.; AND WHITNEY, R. B.: The Thermodynamic Theory of Electrocapillarity. J. ACS, vol. 64, 1942, p. 1548.
42. HOLMES, P. J.: Electrochemistry of Semiconductors. Acad. Press, 1962.
43. EVANS, U. R.: Metallic Corrosion Passivity and Protection. Edward Arnold & Co. (London), 1948.
44. UHLIG, H. H.: Corrosion and Corrosion Control. John Wiley & Sons, Inc., 1963.
45. FRENKEL, J.: Viscous Flow of Crystalline Bodies Under the Action of Surface Tension. J. Phys., vol. 9, 1945, p. 385.
46. GATOS, H. C.: Crystalline Structure and Surface Reactivity. Science, vol. 137, 1962, p. 311.
47. GATOS, H. C.; AND LAVINE, M. C.: Chemical Behavior of Semiconductors: Etching Characteristics. Progress in Semiconductors, A. Gibson and R. Burgess, eds., Temple Press Books, Ltd. (London), 1965, p. 1.
48. HANNEMAN, R. E.; FINN, M. C.; AND GATOS, H. C.: Elastic Strain Energy Associated with the "A" Surfaces of the III–V Compounds. J. Phys. & Chem. of Solids, vol. 23, 1962, p. 1553.

49. Finn, M. C.; and Gatos, H. C.: Spontaneous Bending of Thin (111) Crystals of III–V Compounds. Surface Science, vol. 1, 1964, p. 361.
50. Gatos, H. C.; Finn, M. C.; and Lavine, M. C.: Antimony Edge Dislocations in InSb. J. Appl. Phys., vol. 32, 1961, p. 1174.
51. Huff, H.; Kawaji, S.; and Gatos, H. C.: Field Effect Measurements on the A and B (111) Surfaces of Indium Antimonide. Surface Science, vol. 5, 1966, p. 399.
52. Gomer, R.: Field Emission and Field Ionization. Harvard Univ. Press, 1961.
53. Fowler, R. H.; and Nordheim, L. W.: Electron Emission in Intense Electric Fields. Proc. Roy. Soc. (London), vol. A119, 1928, p. 173.
54. Müller, E. W.: The Field Ion Microscope. Zeit. für Physik, vol. 131, 1951, p. 136.
55. Inghram, M. G.; and Gomer, R.: Mass Spectrometric Analysis of Ions from the Field Microscope. J. Chem. Phys., vol. 22, 1954, p. 1279.
56. Müller, E. W.: Resolution of the Atomic Structure of a Metal Surface by the Field Ion Microscope. J. Appl. Phys., vol. 27, 1956, p. 474.
57. Melmed, A. J.: Surface Self-Diffusion of Nickel and Platinum. J. Appl. Phys., vol. 38, 1967, p. 1885.
58. Brenner, S. S.: Metal Surfaces: Structures, Energetics and Kinetics. ASM, 1962, p. 305.
59. Bassett, D. W.: The Thermal Stability and Rearrangement of Field-Evaporated Tungsten Surfaces. Proc. Roy. Soc. (London), vol. A286, 1965, p. 191.
60. Drechsler, M.; and Vanselow, R.: Investigation of Tempering and Growth Structure of Metal Single Crystals With a Field Electron Microscope. Zeit. für Kristallographie, vol. 107, 1956, p. 161.
61. Good, R. H., Jr.; and Müller, E. W.: Field Emission. Handbuch der Physik. Vol. XXI. Springer-Verlag (Berlin), 1956, p. 176.
62. Müller, E. W.: Surface Wandering of Wolfram (tungsten) on Its Own Crystal Lattice. Zeit. für Physik, vol. 126, 1949, p. 642.
63. Sokolovskaia, J. L.: Surface Migration of Tungsten Atoms in an Electric Field. Soviet Phys.-Tech. Phys., vol. 1, 1956, p. 1147.
64. Bettler, P. C.; and Charbonnier, F. M.: Activation Energy for the Surface Migration of Tungsten in the Presence of a High-Electric Field. Phys. Rev., vol. 119, 1960, p. 85.
65. Barbour, J. P.; Charbonnier, F. M.; Dolan, W. W.; Dyke, W. P.; Martin, E. E.; and Trolan, J. K.: Determination of the Surface Tension and Surface Migration Constants for Tungsten. Phys. Rev., vol. 117, 1960, p. 1452.
66. Brenner, S. S.: The Thermal Rearrangement of Field Evaporated Iridium Surfaces. U.S. Steel Rept. PR–241, May 1964.
67. Ehrlich, G.: Field Emission Standards of Metal Surfaces. 12th Field Emission Symp., Penn. State Univ., Sept. 1965.
68. Sears, G. W.: A Mechanism of Whisker Growth. Acta Met., vol. 3, 1955, p. 367.
69. Müller, E. W.: Study of Atomic Structure of Metal Surfaces in the Field Ion Microscope. J. Appl. Phys., vol. 28, 1957, p. 1.
70. Melmed, A. J.: Field-Electron and Field-Ion Emission From Single Vapor-Grown Whiskers. J. Chem. Phys., vol. 38, 1963, p. 607.
71. Melmed, A. J.: Field-Emission Microscopy of Twins in Vapor-Grown FCC Whiskers. J. Appl. Phys., vol. 34, 1963, p. 3325.

72. Strayer, R. W.; Crouser, L. C.; and Trolan, J. K.: Use of the Field Emission Microscope in the Study of Screw Dislocations. Direct Observation of Imperfections in Crystals, J. B. Newkirk and J. H. Wernick, eds., Interscience Publ., New York, 1962, p. 163.

73. Cottrell, A. H.: Theoretical Structural Metallurgy. Second ed., Edward Arnold Publ. (London), 1955, p. 232.

74. Sanwald, R. C.; Ranganathan, S.; and Hren, J. J.: Computer-Simulated Ion-Emission Images of Dislocations: Screw Dislocation at the Center of (420). Appl. Phys. Ltrs., vol. 9, 1966, p. 393.

75. Sanwald, R. C.; and Hren, J. J.: Computer Simulation of F.I.M. Energies of Dislocations. Abstract Bull., Inst. Metals, vol. 2, Feb. 19–23, 1967.

76. Brandon, D. G.; and Perry, A. J.: Computer Analysis of Dislocated Spherical Crystal Surfaces. Phil. Mag., vol. 16, 1967, p. 131.

77. Ranganathan, S.: Field Ion Microscopic Observations of Dislocation Structures at Grain Boundaries. J. Appl. Phys., vol. 37, 1966, p. 4346.

78. Sanwald, R. C.; and Hren, J. J.: Computer Simulation of Field Ion Images. Surface Science, vol. 7, 1967, p. 197.

79. Sanwald, R. C.; and Hren, J. J.: Interpretations of Defect Configuration in Field Ion Images. 14th Field Emission Symp., NBS, Washington, D.C., June 26–30, 1967.

80. Hren, J. J.; and Sanwald, R. C.: Interpretation of Field Ion Images of Undissociated Dislocations. 14th Field Emission Symp., NBS, Washington, D.C., June 26–30, 1967.

81. Newman, R. W.: An Electron Microscope and Field Ion Microscopic Study of Defects in Quenched Platinum. PhD. Dissertation, Univ. of Florida, Dec. 1967.

82. Ramadanoff, D.: A Method of Studying the Effect of Temperature on Photoelectric Currents. Rev. Sci. Inst., vol. 1, p. 768, Dec. 1930.

83. Ramadanoff, D.: Photoelectric Properties of Composite Surfaces at Various Temperatures and Potentials. Phys. Rev., vol. 37, 1931, p. 884.

84. Hughes, A. L.; and BuBridge, L. A.: Photoelectric Phenomena. McGraw-Hill Book Co., 1932, p. 101.

85. Gilman, J. J.: Cleavage, Ductility, and Tenacity in Crystals. Fractures. B. L. Averback, D. K. Felbeck, G. T. Hahn, and D. A. Thomas, eds., Tech. Press, MIT, and John Wiley & Sons, Inc., 1959, p. 193.

86. Bowden, F. P.; and Tabor, D.: The Friction and Lubrication of Solids. Part II, Oxford Univ. Press, 1964.

87. Bondi, A.: The Spreading of Liquid Metals on Solid Surfaces. Chem. Rev., vol. 52, 1953, p. 417.

88. Brillouin, M.: Theorie Moleculaire de Frottement des Solides Polis. Ann. Chim. & Phys., 7eme serie, T. XVI, 1899, p. 433.

89. Buckley, D. H.: The Influence of the Atomic Nature of Crystalline Materials on Friction. NASA TM X-52279, 1967.

Topography of Solid Surfaces

J. B. P. WILLIAMSON
Burndy Corporation
Norwalk, Connecticut

The paper starts with a review of methods of examining surfaces and comments on the advantages and limitations of each. It discusses how the development of our understanding of surface structure has tended to be shaped by the observational techniques available at the time.

The recently developed techniques of numerical surface analysis and micromapping are described, and the new picture they give us of the topography of surfaces and of the nature of static contact is discussed. The approach to contact theory suggested by this new picture emphasizes the equal importance of the shape of individual asperities and the statistics of the height distribution of the entire population of asperities on the surface.

A recent instrumental development—the relocating stage—has permitted the observation of individual asperities before and after contact. The paper describes experiments in which the progressive crushing of asperities has been measured as the load on them was increased to many times their nominal yield pressure. These have revealed some previously unobserved aspects of the behavior of surfaces under extreme loads.

Finally, a possible topographic mechanism of wear is suggested, and a simple experimental investigation of the proposal is described.

INTRODUCTION

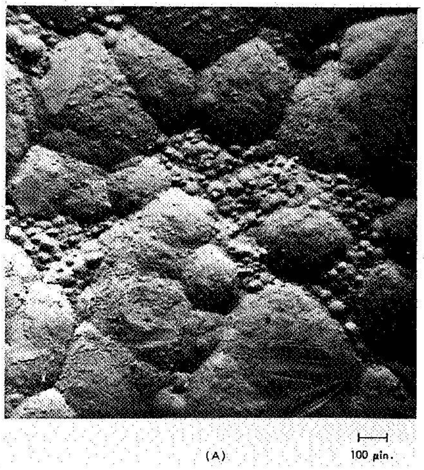
The Principal Methods of Observing Surfaces

PROGRESS IN A SCIENCE is often closely correlated with advances in techniques of observation and measurement. This is especially true in the study of surface contact. The detailed shape of surfaces is very difficult to determine; and although a wide variety of techniques has been developed, none has yet proved completely satisfactory. Each, to some extent, gives a distorted picture—emphasizing certain aspects and obscuring others; therefore, scientific thinking about surfaces has tended to be shaped by the techniques available at the time.

The early metallographic studies of sections through surfaces gave valuable information on large-scale contact phenomena (such as wear tracks and wear particles). They showed that solid surfaces have asperities which are typically 10 to 300 μin. high; that their slopes are shallow, so that their bases can be 50 to 3000 μin. across; and that individual contact areas (as deduced from observations of transferred wear particles) were about 100 to 1000 μin. wide. They also (unique among surface observation techniques) revealed the microstructural state underneath the contact. However, the technique is destructive; and even when using taper-sectioning methods (ref. 1), it is not possible to resolve the detailed topography. The suboptical scale of surface features has led to the development of several interferometric techniques (refs. 2 through 6). These are excellent for stepped surfaces or those with local irregularities within large flat planes; however, on normal rough surfaces they are difficult to operate and interpret. It was only when electron microscopy, both replica and reflexion, was applied to the study of surfaces that the fine structure of asperities was revealed in detail. Figure 1 shows three electron micrographs of a gold surface (taken by Mr. Diehl in my laboratory). The large asperities in figure 1(A) are very shallow domes, roughly 400 μin. across and 50 μin. high; the smallest hills are less than 20 μin. across and 5 μin. high. Figure 1(B) shows a contact area approximately 200 μin. across; and figure 1(C) shows wear marks which extend across several asperities.

These photographic techniques can reveal both macroscopic and microscopic surface features, but they have two major limitations: first, that it is difficult to derive quantitative data—this restricted early work on surface contact to qualitative arguments; and secondly, because of their inherent limited field of view they have tended to encourage scientists to become preoccupied with the behavior of individual hills, whereas in fact the salient point about surface contact is that it involves whole populations of touching asperities.

The second main group of surface observation techniques is based on tactile exploration. Profilometry was introduced by Abbott and Firestone in 1933 (ref. 7); and since then it has been developed into one of the most powerful of surface analysis tools. It has the advantage of presenting a picture of a large population of asperities and showing their relative positions. However, the profiles drawn by these techniques are deliberately distorted: the vertical magnification is often one hundred or more times the horizontal magnification. Consequently the slopes of the asperities are greatly exaggerated. Since the profile is by far the most common representation of surfaces used in this science, a false picture of the magnitude of asperities, arising from this distortion, has crept into scientific thinking and is the root of much confusion—especially in such topics as interlocking during



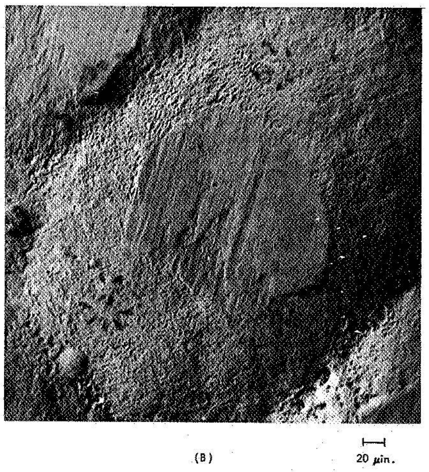
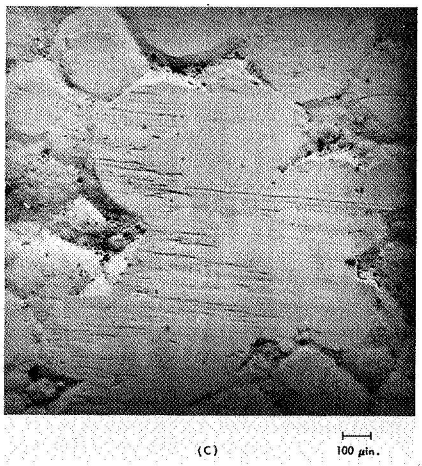


Figure 1.—Electron micrograph of plated gold surface. (a) General view at X 4000. (b) A single contact area X 15000. (c) Contact at heavy load X 4000.



sliding and surface deformation under normal loads. In particular, the question "Does the stylus record the valley bottoms?" often assumes false importance because one tends to think of an undistorted stylus riding over a distorted profile. At the scale of a typical profile, the hemispherical tip of a high-resolution stylus is a line 0.01 in. wide and 1 in. high.

Mr. Hunt and I have recently performed some experiments to determine the size of the track left by the stylus, which is of course related to the question of resolution. Figure 2 shows a network of tracks spaced 0.001 in. apart made on aluminum using a high-resolution profile-measuring instrument (stylus radius 45 μin.; stylus load 60 mg). Aluminum was chosen as it is one of the softest metals on which profiles are made. These tracks are approximately 20 μin. wide; tracks on harder surfaces would be narrower. This indicates that nothing smaller than 20 μin. (horizontal) can be resolved. It can also be shown that for normal surfaces, where the average slope is a few degrees, the "lost region" at the bottom of a vee valley is about 15 μin. These two limitations—the obliteration of minute features due to track deformation, and the loss of bottom-detail due to stylus bluntness—happen to be almost equal. It is thus reasonable to take 20 μin. as the horizontal resolution of the best profile instruments. Another pertinent question is "How deep are the tracks compared to the surface features of interest?" Figure 3 shows four profiles generated by the "east-

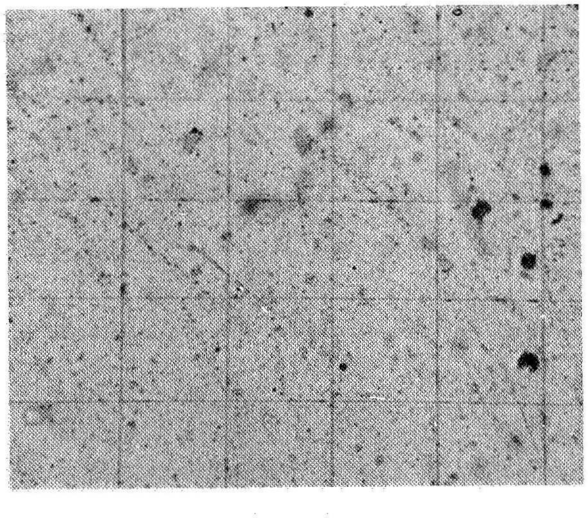


FIGURE 2.—A network of tracks, spaced 0.001 in. apart, made with a high-resolution stylus on aluminum (X 500).

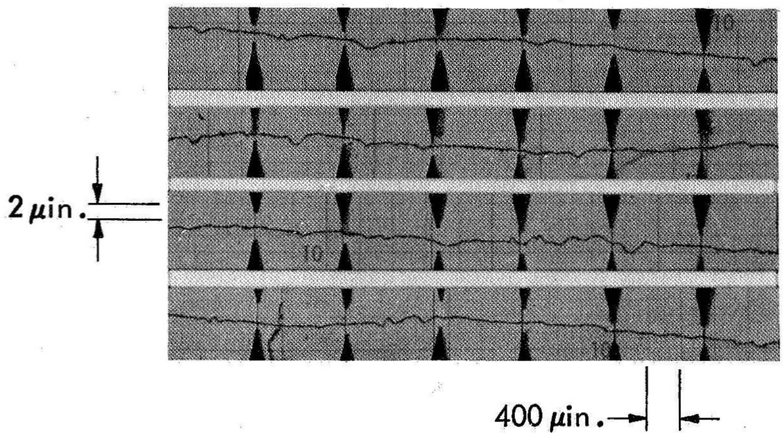


Figure 3.—Profiles of the "east-west" tracks in figure 2, showing the absence of serious surface damage caused by the "north-south" tracks.

west" tracks in the network of figure 2. The positions where these cross the previously drawn "north-south" tracks are marked by the arrowheads. If the stylus tracks were deep compared with the height of the surface features, we would expect this technique to reveal them. Since the stylus is clearly sharp enough to sink into tracks that it itself has made, the east-west tracks should be evident as notches in the north-south profiles. The experiment detected no significant deformation at any of the intersections. This result shows that tracks can be clearly visible (fig. 2) and yet be shallow in comparison with the surface roughness (fig. 3). The depth is small even compared with the asperities on the 1 μin. CLA surfaces studied. A stylus leaving such a shallow track, which in any case is probably of constant depth, will lead to very little error in the profile. The track depth is not more than 1 μin., and the profile will probably reproduce the surface to better than 0.1 μin. In another experiment we have measured the absolute height of a profile as indicated by 30 successive traversals made on exactly the same track. The recorded height of the surface was reduced progressively during the first ten traverses by roughly 1 μin. per traverse; by then the track had grown so that the load could be supported elastically and no further increase in track depth occurred. The same phenomenon of progressive track enlargement occurs when balls roll on flats (ref. 8). It must be emphasized that the above comments on the resolution and on the nondestructive nature of profilometry apply only to high-resolution styli operated under very light loads.

An attempt at finer horizontal discrimination in profilometry (1 μin., for example) would probably not be profitable. The present level of 20 μin. is adequate for the study of surface topography; indeed we have found a resolution of 65 μin. to be satisfactory for most research purposes. Under normal conditions, contact areas are very rarely smaller than 50 μin. across; and, just as the discriminating ability of the observation techniques limits the understanding we may achieve of surface contact, so the size of the contact event limits the scale at which it is profitable to examine surfaces. In this respect it seems that the recent group of experiments in which fine points have been touched lightly together inside electron microscopes and field-emission microscopes (D. Tabor, A. J. W. Moore, private communications) will in fact be of greater value to the solid state physicists than to the surface scientists.

The Development of Our Concept of Surface Contact

So far there have been four distinct stages in the development of our understanding of contact. The first, dating perhaps from Leonardo da Vinci's observations on the contact and friction of rough and smooth bodies (ref. 9), was characterized by the belief that the area of real contact was the same as, or at least very close to, the apparent area. Friction was recognized and studied as an engineering phenomenon in the eighteenth century (refs. 10 and 11), and the two basic laws "F is proportional to W" and "F is not dependent on A" were accurately verified. Three theories of surface interaction emerged as possible explanations of the phenomenon. (Palmer (ref. 12) gives a useful review of the early work on the "interlocking," "adhesion," and "electrification" hypotheses.) Early work was hampered by only a rudimentary knowledge of surface topography and by total ignorance of the concept of surface cleanliness.

The second stage was initiated by the recognition that solid surfaces have asperities which are large compared with the amount of deformation sustained in normal contact. This was perhaps the greatest single advance in the history of the subject. It is not clear exactly when this concept was first introduced (see references 12 through 14, all of which refer to early work); but the first comprehensive study of the implications of this picture is presented in the book Friction and Lubrication of Solids (ref. 15). The distinction between the real and nominal areas of contact immediately suggested a physical explanation of the two basic laws of friction: the independence from the nominal area was obvious; and the assumption that the asperities deformed plastically led directly to the proportionality of contact area and load, and thus to the law "F is proportional to W." Conversely, it was believed that elastic deformation of asperities was inconsistent with this well-established observation. In the following decade, the implications of this "plastic contact" theory were explored in the neighboring fields of adhesion, wear, lubrication, and electric contact.

The third stage probably began in 1957, when Archard (ref. 16) pointed out that although it is reasonable to assume plastic flow for the first few traversals of one body over another, it is absurb to assume this for machine parts which may make millions of traversals during their life; the asperities may flow plastically at first, but they must reach a steady state in which the load is supported elastically. He went on to show that although the simple Hertzian theory does not predict contact areas proportional to load, a generalized model in which each asperity is covered with microasperities, and each microasperity with micro-microasperities, gives successively closer approximations to this relation as more stages are considered. Archard explained that the essential distinction in the argument was whether an increase in load creates new contact areas or increases the size of existing ones. For physically plausible surfaces, any elastic model in which the number of contacts remains constant will give A proportional to $W^{2/3}$; but if the average size remains constant and the number increases, the area will be proportional to the load. This argument did not rule out plastic deformation of the asperities; it merely showed that Amontons' first law could be derived from either plastic or elastic contact.

The latest stage has been the development of a theory of contact which combines the basic contact event (the touching and deformation of two asperities) with its environment (a population of asperities of random—but statistically describable—shapes). Since the early paper of Zhuravlev (ref. 17), who studied the elastic content of an array of aligned spherical asperities, at least a dozen analyses have been published discussing the contact of various assumed asperity shapes, height distributions, and deformation behavior—for example, Ling (ref. 18). Concurrently, several attempts were made to establish criteria based on topographic considerations that would determine the mode of deformation. The most general treatment so far available is given in a series of papers from the Burndy Research Laboratory (refs. 19 through 22). They discuss the contact between two populations of asperities and also offer a concept of a "plasticity index," a parameter which combines both the surface topography and the material properties of the solid in a criterion which determines whether contact will be elastic or plastic. This approach proposes that the nature of surface contact is primarily controlled by two topographic features—the height distribution of the tops of the asperities and their curvature; and two material properties—the elasticity and the hardness. It implies that the center line average is useful only in that it is a loose indication of the summit height distribution; and in retrospect one can see that the long delay in the development of a general theory of contact resulted from lack of an adequate method of measuring surface structure. Our present picture is the direct result of the development of numerical analysis and micromapping techniques for surfaces.

This paper will first discuss the new pictures that these approaches give of surface topography and of the static contact between them; it will then describe some very recent work on the role of asperities in surface contact and finally consider the implications of this new viewpoint to our understanding of friction and lubrication.

NUMERICAL ANALYSIS OF SURFACE PROFILES

It is a straightforward matter to feed the output of a Taylor-Hobson Talysurf through a suitable analog-to-digital conversion and sampling unit into a digital computer. The stylus and the electromechanical transducer produce the voltage analogue of the surface in the usual way. The pen recorder draws the profile, and at the same time a data conversion unit samples this voltage at a rate of 36.7 readings per second and punches it on paper tape. With the usual horizontal magnification of 100, this system records the surface height every 65 μin.; with the 500X magnification, the horizontal resolution becomes 13 μin. In the first case the height recorded is the average value over the previous 20 μin. of the profile; in the second the averaging is over 4 μin. The heights are recorded to three decimal digits. At the maximum vertical magnification available, the low-order digit is in units of 10Å. For precise work, stringent precautions must be taken to isolate the stylus from mechanical vibrations, and the instrument is mounted on an elaborate aseismic table. In practice the system is limited by residual noise, mechanical and electrical, to a resolution of about 50Å. In a typical observation, 1800 height readings are recorded from a 0.12-in. profile. The tape is converted to punched cards for storage; 90 cards are required to carry the information of one profile.

After the sample of heights is entered into a computer it is easy to obtain the center line height, the center line average, the rms of the height deviation, and many other parameters used in the various surface texture standards. Also, long and short wave components of the profile can be separated (in fact with considerably more flexibility than with the conventional analog analyzers). The computer can readily be programmed to evaluate new experimental parameters which can then be obtained retrospectively using the stored tapes of previous Talysurf runs. This flexibility is particularly attractive. But most important, the ability to represent a surface by a series of regularly spaced height readings along a profile offers the theoretician a completely new approach: the computer can locate specific features and produce data referring to them only. This capability provides the crucial advantage of digital analysis over the analog approaches proposed by Myers (ref. 23), Kubo (ref. 24), and Silin and Frederick (ref. 25). For example, the computer can identify the peaks in the profiles and calculate their spacing, curvature, and height distribution. The importance of these parameters is emphasized in recent theories of contact.

An important benefit of this numerical technique has been the ability to describe surfaces in terms of their height distributions. These are obtained by determining the fraction of the surface lying in each stratum (which is also the probability that a randomly chosen point in the surface will have this height). When Mr. Hunt and I first obtained such height distributions in 1960, we were fascinated to find that for many surfaces they were closely Gaussian. This was indeed a windfall for the theoreticians. The early measurements were crude and the sample of heights analyzed was rarely greater than a few hundred. Thus although it was possible to demonstrate that the middle 90 percent of the surface had a Gaussian height distribution, the data were insufficient to reveal whether the extremes of the distribution were also Gaussian. This was a serious drawback because the highest points in the surface obviously play a major role in the contact. Figure 4 shows some recent data we have obtained on bead-blasted aluminum. The results are plotted as a cumulative distribution on normal probability graph paper. This has a distorted scale so that a Gaussian distribution becomes a straight line. The height distribution of the surface (crosses) is indeed Gaussian. This is similar to results reported by Bickel (ref. 26), Tallian et al (ref. 27), and Pesante (ref. 28). However, we were able to record 22 500 height readings while still maintaining an accurate reference plane over the surface; thus the data show that the height distribution remained Gaussian at the high end for at least 3.5 standard deviations from the mean, i.e., at most only 0.01 percent of the surface was non-Gaussian. At the lower end most surfaces tend to be non-Gaussian; this is usually noticeable in the bottom 5 percent of the distribution.

Although most common surface preparations produce Gaussian distributions, some do not; in figure 5 the crosses show the height distribution of a mild steel specimen which has been slid against a copper flat under oleic acid. This figure illustrates a characteristic which occurs in many surfaces: even when the surface as a whole is strongly non-Gaussian the upper 50 percent usually is accurately Gaussian. The circles in figures 4 and 5 show the height distribution of the peaks; these, too, are Gaussian. This is very convenient because the peak height distribution plays an important role in the theory of surface contact.

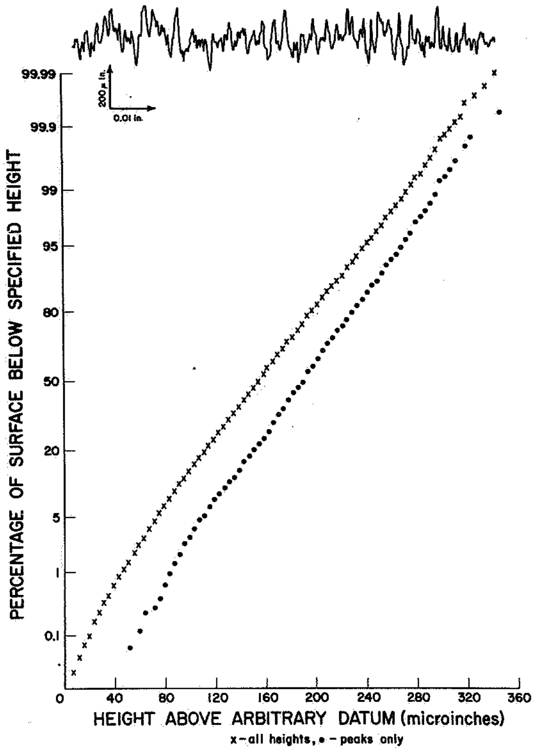


FIGURE 4.—Cumulative height distribution of bead-blasted aluminum. Both the distributions of all heights (*X*) and of peak heights (●) are Gaussian. The profile of the same surface is shown in the upper diagram; the vertical magnification is 50 times the horizontal magnification.

Mr. Hunt and I have analyzed the surface topography of a number of common surface preparations, e.g., lapped, polished, blasted, ground, milled, and turned. We find that in many cases the distribution is approximately Gaussian over nearly the whole height range, and that the height distribution of the upper half, and the peak height distribution, are accurately Gaussian. In other studies, described later in this paper, we have shown that contact between two rough surfaces only rarely occurs below the center line of either, and tends to be on or near the peaks. Thus it is completely justifiable to treat the height distribution as Gaussian, provided that the standard deviation is determined from the upper half of the sample only. This is most convenient and has greatly simplified the mathematical treatment of surfaces. Although the approach used by Greenwood and Williamson (ref. 20) does not assume any particular topographic distribution, it is difficult to see how a generalized contact theory could have arisen had it not been discovered that surfaces were effectively Gaussian.

Surfaces should be Gaussian because the shape of every point on them is the result of many independent deforming processes. It is generally true that any quantity (in this case the height of a surface

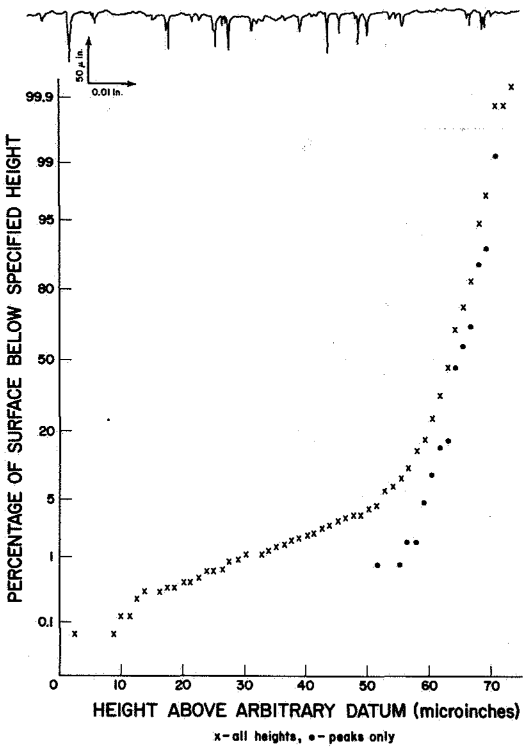


FIGURE 5.—Cumulative height distribution of mild steel specimen. Distribution of all heights, *X*. Distribution of peaks, ●. This specimen was abraded on 400 grade carborundum paper, then slid against a copper block flooded with oleic acid, at approximately 10 kg, 130 cm/s for 30s. Although the distribution at first sight appears highly non-Gaussian, in fact nearly 90 percent of the surface is Gaussian: the surface, with an actual standard deviation of 50μ in., would behave in contact as if Gaussian with a standard deviation of half this. The profile of the same surface is shown in the upper diagram; the vertical magnification is 200 times the horizontal magnification.

at a given point) that is the result of a large number of random increments and decrements will tend to follow a Gaussian distribution. This follows from the well-known Central Limit Theorem of statistical theory. Usually, surface preparation is a repetitive random-event process. Under typical bead-blasting conditions, for example, it can be shown that each point on the surface is hit by a bead more than 1000 times. Clearly, in polishing, lapping, and grinding, numerous impacts occur against each part of the surface. This theorem does not apply to single-process surface preparations, such as cleavage or machining, but it can be extended to cover "double-process" surface preparation. Consider a specimen that is first prepared by a process which leaves a large-amplitude Gaussian height distribution and is then subjected to a second process which operates only on the higher parts of the surface (for example, an abraded specimen which was subsequently worn against a smooth surface). The valley bottoms will still have the Gaussian distribution of the abraded structure, but the upper half of the surface should acquire the Gaussian distribution of the wearing process (as in figure 5). Of course this happens only if the second process is finer than the first.

Certain limitations must always be remembered whenever data obtained from a linear profile are used to describe the topography of a surface. The asperity shape revealed by the profile may be misleading, particularly with strongly oriented surfaces. Further, the profile peaks are not the summits of the asperities; the track will usually cut across the side of a hill. Recently Mr. Hunt and I have been able to determine the height distribution of true surface summits. Our results show that, like the profile peaks, the surface summits have a Gaussian height distribution.

Cumulative height distribution curves such as those shown in figures 4 and 5 are particularly helpful in describing a surface. Abbott and Firestone (ref. 7) pointed out that they are "bearing area curves," i.e., the contact area that would exist if the surface were worn down to a certain height. Several authors (refs. 29 through 31) have recently suggested that these are only "bearing line curves" and that two such distributions from perpendicular profiles must be "multiplied together" to produce a genuine height distribution. This is not true. A height distribution can in principle be obtained from an infinite number of closely-spaced parallel sections—the usual process of integration over a surface. However, when the profiles are long compared with the surface irregularities, they will all contain the same information; so that if 10 percent of one profile lies above a certain height, then 10 percent of every profile does, and the correct interpretation is that 10 percent of the surface, not 1 percent, lies above this height. This is clearly true for randomly structured surfaces; with strongly oriented surfaces, it is necessary to add that the profiles must be constructed to include a representative sample of the topography. In practice this is not an onerous requirement. It is difficult to obtain a rogue height distribution even when the profile location is deliberately chosen to produce one.

It is a paradox that profilometry, which appears to show the shapes of asperities and give only a single cross section of the surface, in fact yields little information about individual asperities but provides exact data on the general topographic distribution.

THE MICROCARTOGRAPHY OF SURFACES

Recent theoretical treatments have shown that the shape of individual asperities and the height distribution of the population of asperities are equally important in determining the nature of surface contact. Records of the surface should thus show the details of individual features but must also cover enough area to include a few hundred asperities. In particular, the asperity height distribution can be obtained only if a reference plane can be held to the necessary accuracy over a large area. The needs of the surface scientist are thus similar to those of a surveyor studying the topography of the earth: local detail and long-range coherence are required simultaneously. As in the large-scale analog, the information can most conveniently be presented in a map.

In principle, a map of a surface may be obtained by taking a number of closely spaced parallel profiles and displaying the data in a two-dimensional array. The principal difficulty is in maintaining a height reference between successive profiles; clearly this must be done to the accuracy with which the height is recorded on each profile (in these experiments 100Å). This can be achieved by mounting the surface in a special holder which has two reference flats—one each side of the specimen. These flats are hardened, coplanar, flat to 1.0 μin. and smooth to 1.0 μin. CLA. The specimen is held so that its surface is approximately coplanar with the flats. (This can usually be achieved to within 1.0 μin. by levelling screws and shims.) The stylus is then drawn across the specimen from flat to flat so that it gives a profile starting and finishing in the reference plane. The holder is mounted on a stage which can be moved in the horizontal plane in small steps perpendicularly to the stylus track. In a typical analysis, 25 parallel profiles are recorded on punched cards. The decks are first roughly synchronized manually in the direction of the profile by using a fine scratch in one flat running perpendicularly to the tracks. The computer then synchronizes the decks precisely by optimizing the correlation between adjacent profiles.

Since all the profiles use the same reference plane, any spurious variations in height due to the raising and lowering of the stylus or to the sideways movement of the specimen can easily be corrected. The central portion of each profile, synchronized lengthways and given as a height variation from the plane defined by the reference flats, is then repunched to create the working deck. The profiles are then printed simultaneously to produce a map of the surface expressed as an array of spot heights. The format prints the data spaced 0.25 in. by 0.5 in. On the surface the readings are made every 65 μin. along the profiles and every 130 μin. between profiles. Thus the magnification of the map is the same in both directions and approximately equal to 4000. Contours can be superimposed on the map and displayed by varying print styles in the computer or by coloring the maps manually.

Figure 6 shows the topography of a bead-blasted aluminum surface. The true shapes of the asperities and their relative positions are clearly illustrated. The major summits are approximately 0.002 in. apart and rise about 200 μin. above the valley floors. The contours are spaced vertically every 40 μin.

Apart from offering the ability to visualize a surface simultaneously in detail and in extent, this mapping technique enables us to measure several topographic parameters not previously available. For example, it is relatively easy to program the computer to search for true summits. (For this purpose a summit is defined as a spot height higher than its eight nearest neighbors.) In 2×10^{-4} square inches of surface, of which figure 6 shows part, there were 400 summits. Figure 7 shows that, like the total surface and like the profile peaks, the summits have a Gaussian height distribution.

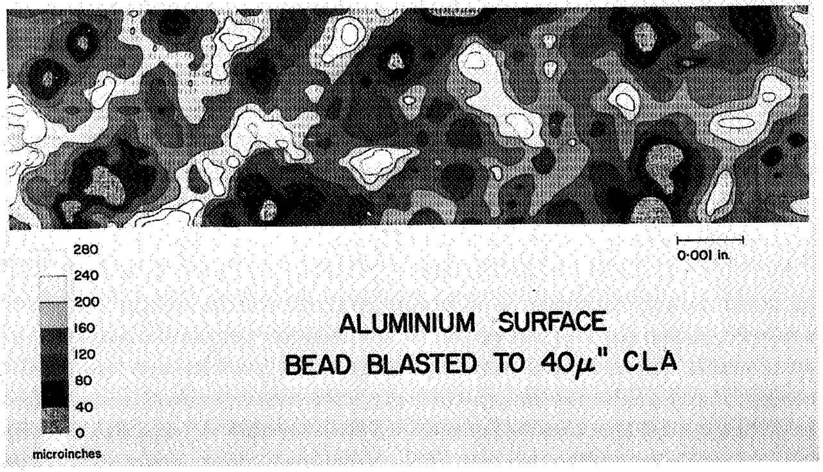


FIGURE 6.—Contour map of a bead-blasted aluminum surface.

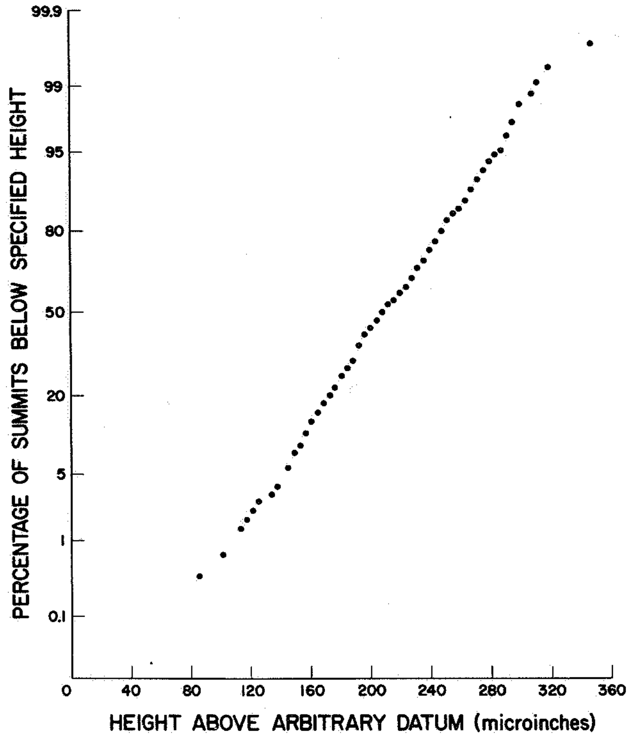


FIGURE 7.—Cumulative height distribution of summits on aluminum surface shown in figure 6.

Microtopography also opens a range of possibilities in the study of contact areas. For a rough surface pushed against a flat, the contact areas at successively increasing loads are simply indicated by contour lines. In practice, the volume of the rough solid above the level of the flat must be accommodated by shape changes in the lower parts of the surface. For the present we shall assume that these changes are negligible; in fact, preliminary results from recent studies suggest that this assumption may be quite reasonable. Contact between two rough surfaces can be discussed in terms of the gap between them. This is merely some arbitrary constant minus the sum of the two surface heights at each pair of corresponding points. As the constant is reduced (which corresponds to the surfaces being brought together), contact occurs wherever the gap becomes zero. If height data from both surfaces are entered into the computer, a map of the gap may easily be produced. Its contours outline the areas of contact occurring at different separations. When these contours are superimposed on a map of one of the contacting surfaces (where, of course, they are no longer loci of equal height) they show how the total area is divided into subareas and how these are located with respect to the surface features. Finally, the separation can be related to the load, using the relations derived by Greenwood and Williamson.

Figure 8 shows how the contact area between the surface illustrated in figure 6 and a similar one grows as they are brought together. Four gap-map contours are shown, superimposed on one of the surfaces. The solid boundaries enclose the areas in contact under the lowest load considered (200g on one cm^2 nominal area). At this load, 0.4 percent of the nominal contact area is touching; and the mean planes are 2.5 standard deviations apart. The dash boundaries, the dot-dash boundaries, and the dot boundaries outline the areas of contact as the separation is progressively decreased. Although contact is usually on or near the summits, it does sometimes occur lower down—presumably because of a hill on the other surface. The height distributions of the contact areas at these four separations, as well as the loads that would give these separations, are shown in table 1. In figure 9 the unshaded Gaussian curve is the height distribution of one of the contacting surfaces. The shaded curves give the height distribution of those parts of this surface making contact at the four different separations. At light loads, contact occurs only on the uppermost 10 percent of the surfaces. Even when the separation has been reduced to one standard deviation, the contact areas are centered around the 85th percentile of the height distribution of either contacting surface; and 90 percent of the contact is above the mean plane. In interpreting these data, it should be remembered that this discussion neglects the deformation

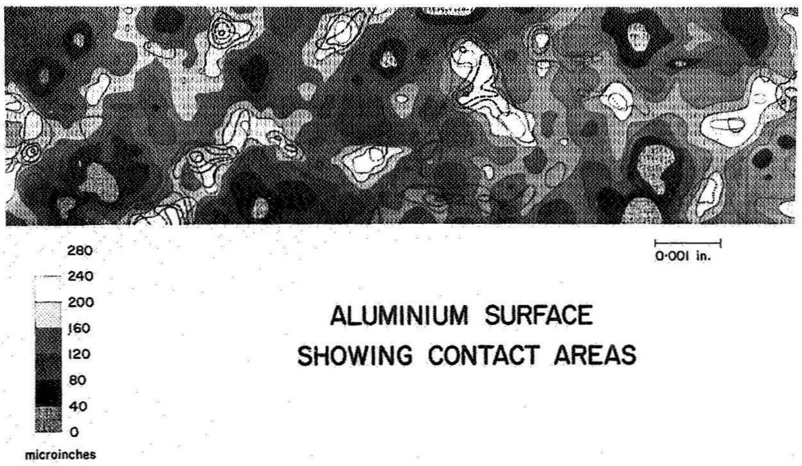


FIGURE 8.—Section of map shown in figure 6 with contact areas superimposed.

occurring as the surfaces squeeze together, restricting the contact to even higher strata of the contacting surfaces.

TABLE 1.—*Contact Between Rough Surfaces (Both Surfaces Have a Standard Deviation of 50 μin.)*

Separation between mean planes in standard deviations, d/σ	Approximate load on 1 cm^2 nominal area,[a] Kg	Percentage of nominal area in contact, %	Location of contact area with respect to the height distribution of either contacting surface		
			10% of contact lies below this percentile	Mean of contact lies on this percentile	10% of contact lies above this percentile
2. 5	0. 2	0. 4	87	97	99. 5
2. 0	0. 8	2. 2	77	95	99. 2
1. 5	3. 0	7. 4	60	90	99. 0
1. 0	8. 0	17. 0	50	85	97. 0

[a] Assuming typical surface topography (ref. 20).

THE DEFORMATION OF ASPERITIES DURING CONTACT

As early as 1948, Moore (ref. 32) pointed out the remarkable persistence of asperities during gross surface deformation. One of the best known pictures in the theory of friction is his microsection of a

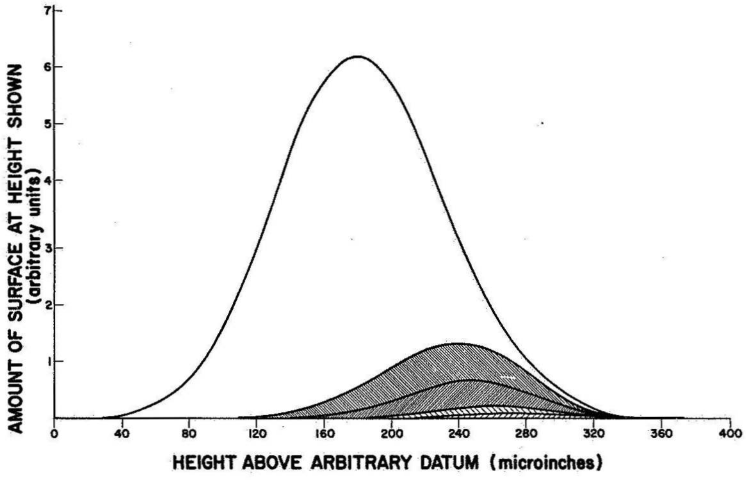


FIGURE 9.—Height distribution of surface shown in figure 8, showing location of contact areas. See also the data in table 1.

deep indentation made into an annealed copper surface. It shows how a series of fine parallel grooves, cut into the surface before indenting, persist almost unchanged in the indentation. Moore found it impossible to make the asperities disappear, however hard he pressed. This is most remarkable, as one would certainly expect the asperities to flow plastically before there is bulk flow. More recently Milner and Rowe (ref. 33) have reported experiments in which roughened surfaces were pressed against smooth ones and, depending on the relative hardness, either the aperities were flattened or they indented the smooth surface; in both cases there was complete mating. Both these experiments have given rise to general pictures of the contact between surfaces: the one based on discrete and small contact areas; the other postulating the possibility of contact areas that approach the nominal area. The first picture offers an explanation of many aspects of friction and its associated phenomena; the second is required to explain the industrial technique of pressure welding.

Why should asperities sometimes collapse completely and sometimes persist even when transmitting forces large enough to cause bulk flow in the solid? Greenwood and Rowe (ref. 34) describe an elegant experiment in which they compressed two cylinders, one tall and one penny-shaped, between hard anvils. The tall one "barreled," and the plastically deforming region did not extend to the interface. When the short one was compressed by the same percentage, bulk plastic flow did occur at the interface, except within a small central dead zone (ref. 35). Greenwood and Rowe found that the asperities, those on the tall cylinder and those in the central dead zone of the penny, collapsed only when they were within a region of bulk plastic deformation. Now the interesting point is not that aperities flatten when within a region of bulk plastic deformation—plasticity theory tells us that they should do so; but rather, how they manage to transmit forces large enough to cause plastic flow in a body which has many times their cross section while remaining virtually undeformed themselves.

Mr. Hunt, Mr. Osias, and I have recently been studying this phenomenon. We have developed an instrument which enables us to obtain profiles along exactly the same track before and after loading the surface. This has considerable advantages over techniques in which independent random profiles are made before and after the loading. Often only a few asperities are deformed; then, even though these may be heavily distorted, the random sample, being predominantly made up of asperities which were not in contact, will show little change. The repeated profile technique permits us to select asperities which have been deformed, and if necessary study their behavior individually. To ensure that the stylus traces exactly the same track, it is necessary to relocate the specimen to a positional accuracy of better than 5 μin. (the track of a high resolution stylus being 20 μin. wide), and to an angular accuracy of 0.001 degree. (Otherwise the tracks will diverge during the 0.2-in. traverse.) Further, it is necessary to remove and replace the specimen by lowering and raising it below the stylus because the mechanism for lifting and lowering the Talysurf arm does not relocate accurately enough. This is achieved with the instrument shown in figure 10 (ref. 36). The main box is built of half-inch thick steel and is firmly attached to the Talysurf base plate. The actual relocating table is also a half-inch thick steel plate, onto which the specimen is rigidly mounted. During the construction of this device, this plate was bolted onto the lid of the box, and three locating holes were drilled and reamed through both at the same time. The relocating table normally rests on two roller tracks halfway up the sides of the box, and from this position it can be removed and reinserted. When the air bag, which lies under it, is inflated, the plate is lifted against the underside of the lid. First the three guide pins (see fig. 10) locate, and the compression rings on them level the plate; then the three ball bearings which rest in hardened seats in the reamed locating holes in the plate are driven into the lower end of the mating holes in the lid to provide the final positioning. When the air is released, the plate falls and can be removed. The instrument relocates to the required accuracy provided two precautious are taken: first, the ball bearings must not be allowed to revolve (their asphericity destroys the alignment); and second, the pressure in the air bag in the raised position must be controlled to 1 torr (so that the distortion in the steel box is constant). The table also relocates to better than 2 μin. in the vertical

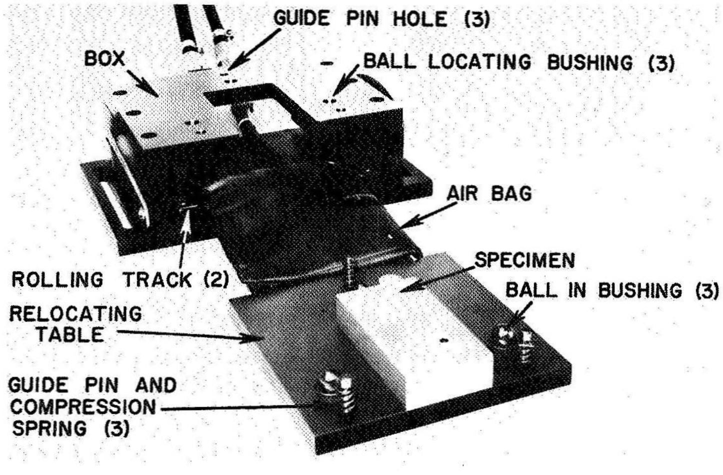


FIGURE 10.—Relocating table.

direction. Figure 11 illustrates the accuracy which is achieved. The first track is 35 μin. wide; the second traverse, made after the specimen had been removed and reinserted, increases the track to 40 μin. and is symmetrical about it at both ends. The tracks shown in this figure were made by a regular stylus; a high-resolution stylus leaves tracks about half as wide.

In our experiments, a cylinder of aluminum is first forced under a pressure of 80 000 psi into a tight-fitting hole in a massive hardened steel jacket. This of course greatly exceeds its yield pressure; and after the aluminum is imbedded in the steel, there are no voids and it can no longer flow plastically. The exposed surface of the specimen lies just below that of the steel and can be pressed with the end of a hardened steel cylinder which fits tightly into the hole in the jacket. Under these conditions the deformation of asperities can be studied without the complication of bulk plastic deformation. The steel-encased aluminum specimen was mounted on the relocating table, and a profile was made of its surface. Initially the aluminum had the smoothness of the anvil with which it had been embedded (top profile in fig. 12). The table was then removed and the aluminum bead-blasted to a roughness of 200 μin. CLA, and a second profile was made. The rough surface was then pressed with the hardened anvil under successively increasing loads up to 9000 lb (giving 80 000 psi over the specimen), and a profile was recorded between each pressing. Figure 12 is a selection from these 30 profiles and shows how the asperities have progressively flattened. In an auxiliary experiment we found that a force of 900 lb

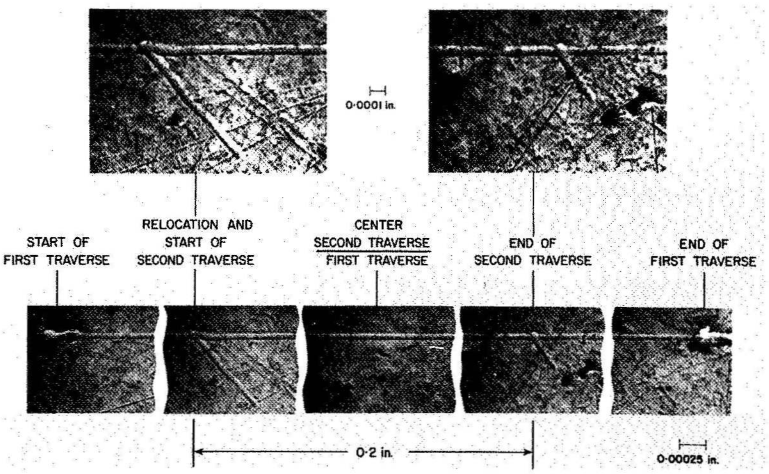


Figure 11.—Profile tracks made using relocating table.

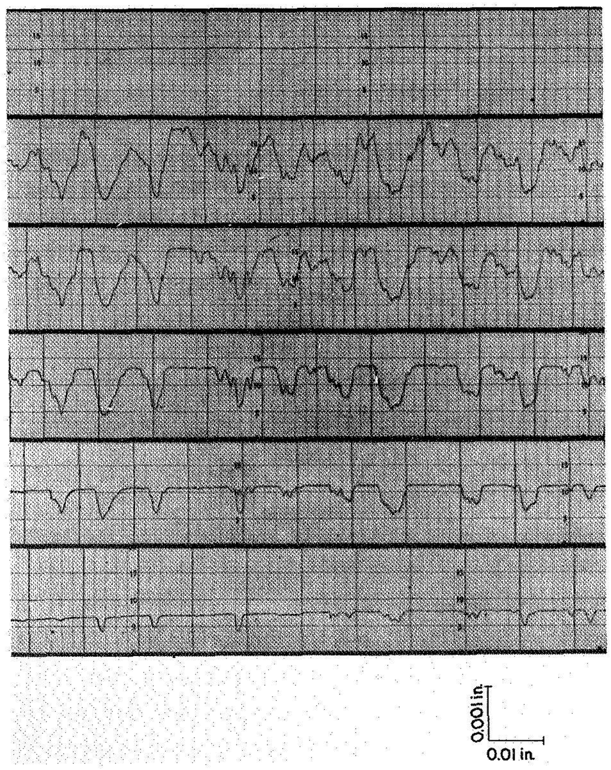


FIGURE 12.—Profiles showing progressive deformation of asperities in high pressure experiment. (a) Virgin surface. (b) Bead-blasted to 200μ in. CLA. (c) After loading to 200 lb (1800 psi). (d) After loading to 1000 lb (9000 psi). (e) After loading to 3800 lb (34 200 psi). (f) After loading to 8000 lb (72 000 psi). The pressures stated are the apparent pressures; the real pressure on the contact spots is much greater.

caused bulk deformation in these aluminum cylinders; yet when this force was applied, the asperities showed little deformation, even up to at least four times the force (fig. 12). Since the real area of contact is then approximately half the nominal area, the asperities must have carried a local pressure of eight times their normal yield pressure. This system has the advantage of permitting asperities to be studied under high forces that would destroy on ordinary specimen.

A very convenient way of describing a surface in deformation experiments is in terms of the deciles of its height distribution, i.e., to state the height below which 10 percent, 20 percent, 30 percent, etc., of the surface lies. In figure 13 these deciles are used to describe how

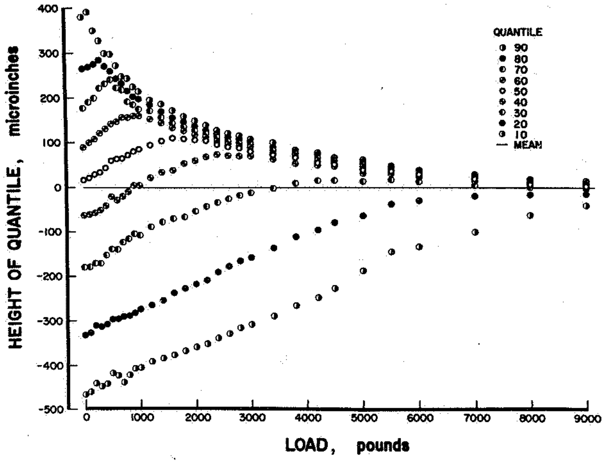


FIGURE 13.—Behavior of aluminum surface in high-pressure experiment, as shown by the changes in the deciles of its height distribution.

the surface changes as the load is increased. At first sight the behavior is surprising. The highest points of the surface flatten with increasing load, as they should. But it was not expected that the rest of the surface would rise as the load increased. Yet the upward slope of all the lower percentiles clearly shows that this is so. Indeed at very low loads, when the area of contact is less than 10 percent of the nominal area, even the 90th percentile rises. As the load increased from 0 to 500 lb, the level of the contact plane was pushed down by 140 μin.; the lowest 10 percent of the surface rose 30 μin., but the height below which 60 percent of the surface lies rose 50 μin. Initially the 50th percentile did not lie on the mean (or center line) because the surface was not symmetrical. At the end of the experiment the surface was flat, and thus symmetrical; the mean and the mode then coincided. The fact that the mean is observed at a constant height throughout the experiment demonstrates the vertical reproducibility of the relocating table.

Because the volume is constant, it is obvious that at very high loads all the surface must lie on the mean plane; but it is not obvious that the only parts of the surface moving downward are those actually in contact with the anvil and that at all loads all the surface not in contact rises. This has an important implication. If the deformation of an asperity were a localized phenomenon confined to the asperity and its immediate roots (as one might assume by thinking about surfaces of the exaggerated shape shown in profiles), then deformation would proceed progressively downward through the deciles and considerable flattening could occur at the summits before the overspill from the asperities affected the valley floors. Figure 13 shows that this picture must be wrong. The region of plastic deformation associated with each asperity must be considerably larger than its base so that as the summit is flattened the redistribution of volume occurs throughout an area covering the asperity, its neighboring valleys, and probably the adjacent asperities as well; otherwise the lower deciles would not be affected by very light loads.

In figure 13 the height of the 90th percentile above the mean is a direct measure of the separation. Further, the percent of the nominal area actually in contact can be determined by analyzing the associated profiles (fig. 12). Since the profile is a random sample of the surface, this is given by the percentage of the profile length contained within the plateaux (assumed to have been caused by contact with the flat anvil). Finally, it can be shown that the number of contact areas per unit area of the surface is proportional to the number of plateaux per unit length of the profile. Thus the experiment on asperity persistence has also provided a means of determining the behavior of the separation, the real contact area, and the number of subareas as functions of the load—and of doing this from light loads to extremely heavy loads.

The top graph in figure 14 gives the clue to the situation. At first the number of plateaux in the profiles increases proportionally to the load; however, above a certain load the number suddenly becomes constant. This break in the curve seems to divide the phenomena of surface contact into two completely different regimes. At light loads, the real area of contact and the number of contact areas are proportional to the load; and thus the average size of each subarea is constant (a condition that leads to Amontons' first law). At loads above this break, the number of contact spots is constant and their size increases with load. The real area of contact (middle, fig. 14) also ceases to be proportional to the load at this point and begins to increase less than linearly with it. This means the real pressure at the contact areas ceases to be constant (it was 20 000 psi) and becomes very great (reaching 90 000 psi at the highest loads used). Finally, the graph of separation versus load (bottom, fig. 14) also shows a sharp kink in this region.

Clearly some major change in the nature of the contact has been revealed by this experiment. But what is it? It cannot be a change from elastic to plastic contact because the entire experiment is performed in the plastic regime. Perhaps the behavior shown in figure 14 marks the point at which the asperities cease to act independently and begin to deform en masse. It has already been suggested, as a deduction from the data of figure 13, that the plastically deforming region associated with each asperity extends over a wider area than the asperity's base. Let us temporarily assume that the deforming area has twice the diameter of the contact area. This would mean that when the real area of contact reaches 25 percent of the nominal area, the deforming regions would cover the entire surface. At this point the resistance to crushing might well increase, causing the bends in the three graphs. Against this hypothesis it can be argued that the bends in the "area" and "number of subareas" graphs occur at 600 lb, whereas that in the "separation" curve is sharply defined at 900 lb. This last force is suggestively equal to that necessary to reach bulk plasticity in the specimen, although why this should coincide with an increased resistance to surface deformation is not at all clear.

FRICTION AND WEAR

The nature of the deformation of the contacting asperities (i.e., whether plastic or elastic) is basic to a discussion of the wear of sliding surfaces. Greenwood and Williamson (ref. 20) have shown that this is controlled by a dimensionless parameter, the plasticity index, which depends on both material and topographic properties. They define the plasticity index as $\frac{E'}{H}\sqrt{\frac{\sigma}{\beta}}$, where E is the elasticity, H the hardness, σ the standard deviation of the height distribution of the summits, and β their mean radius of curvature; and show that contact will be plastic only when it exceeds unity. This immediately implies that the widely accepted view that contact generally is elastic at light loads and becomes plastic with increasing load is wrong; the deformation mode is determined by the material and topographic properties and has only a very minor dependence on load. Most contacts are either always plastic or always elastic over the range of practical loads.

More recently we have measured the plasticity indices of many common engineering surfaces and shown that values ranging from 30 to 0.25 can readily be created by normal techniques; and further that abrading and polishing reduce the plasticity index. This suggests a possible topographic mechanism of wear. The wearing-in process may be the gradual reduction of the plasticity index from an initial plastic value to one in the elastic range. The detailed mechanism of the creation of a wear particle is not relevant here; the hypothesis assumes merely that wear is more probable when asperities touch plastically than in purely elastic contacts. The repeated plastic impact between asperities will reduce their height and increase their radius. This will cause the plasticity index to decrease and, in the absence of any other process, this trend would continue until the contacts become elastic, after which, in principle, no further change should occur and the "worn-in" surface would continue to operate in the elastic regime.

Our first attempts to demonstrate this were unsuccessful. Although

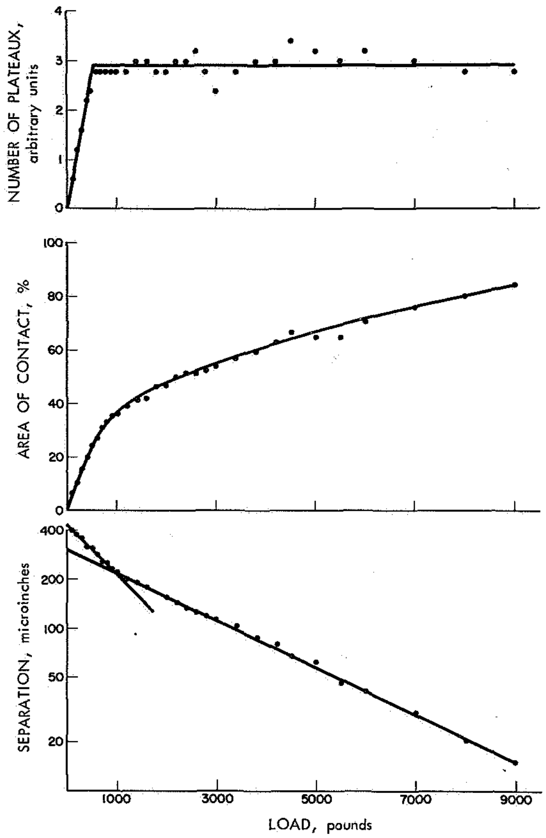


Figure 14.—Behavior of contact under very high pressure. (a) Number of plateaux, showing initial proportionality to load and then limiting value. (b) Real area of contact expressed as percentage of nominal area. (c) Separation, showing break in relation at 900 lb—the force required to cause bulk plasticity in the specimen.

elastic surfaces could readily be made by several fine-polishing processes, we could not generate them by rubbing two plastic surfaces together. Initially the plasticity index fell; but always the process would reverse: the surfaces would become rougher and the plasticity index would become large again. Microscopic examination of the worn specimens revealed small areas that were in fact very smooth. Values of the plasticity index as low as 0.5 (which is definitely elastic) were measured in selected areas; however, the specimens were covered with shallow scratches that increase the spread of asperity heights and caused the surface as a whole to behave plastically. Wear may be the result of two concurrent opposing processes: one which smooths the surface, and one which roughens it. Plastic deformation does not always merely change the shape of the asperity; it can cause a wear particle. Archard and Hirst (ref. 37) suggest that perhaps one in ten thousand contacts results in a detached particle. Wear debris can completely reverse the smoothing process. Even under conditions where the individual wear particles are small, large composite particles can form in the wear process, either by prow-formation (refs. 38 through 40), or by accretion as the debris rolls along the interface. These large particles score the surface as they travel to the edge of the nominal contact area. At first this leads to a surface which is "elastic" between scratches but which, when taken as a whole, is "plastic"; as the process continues, the entire surface becomes scored, and in extreme cases this can lead to seizure.

This suggests that if a means can be devised of removing wear particles rapidly from the interface before they can accrete, the smoothing process will dominate and the plasticity index should fall steadily into the elastic range. It occurred to us that sliding a flat specimen against a woven metal grid would still permit the repeated plastic contact between asperities which leads to smoothing, but would minimize damage by allowing the wear particles to fall through the grid after only very short sliding distances. To test this suggestion three surface-ground mild steel specimens were slid at 100 cm/s under a load of 1.5 kg against a stainless steel woven grid which was flooded with ordinary machining coolant oil. The grid was changed regularly to avoid the development of wear flats. Figure 15 shows that under these conditions the plasticity index does indeed fall until it reaches the elastic range. The rate of change decreases as sliding proceeds, presumably as the fraction of the contacts which are elastic increases. The manner in which the plasticity index falls from an initial high level to a constant is much more significant than its absolute value. The essence of this argument is that plastic contact (indicated by a high plasticity index) tends under lubricated sliding to become elastic contact (giving a lower, constant, plasticity index). This wearing-in is prevented by the roughening action of wear debris. Under special conditions where the debris is removed rapidly from the sliding interface, this natural tendency towards elastic contact can be observed. When sliding was continued, the roughening process eventually became dominant and the plasticity index rose again. At the same time, wear flats appeared on the grid. It is not clear whether these were responsible for the roughening, or whether a transition from mild to severe wear from some other cause was responsible for both.

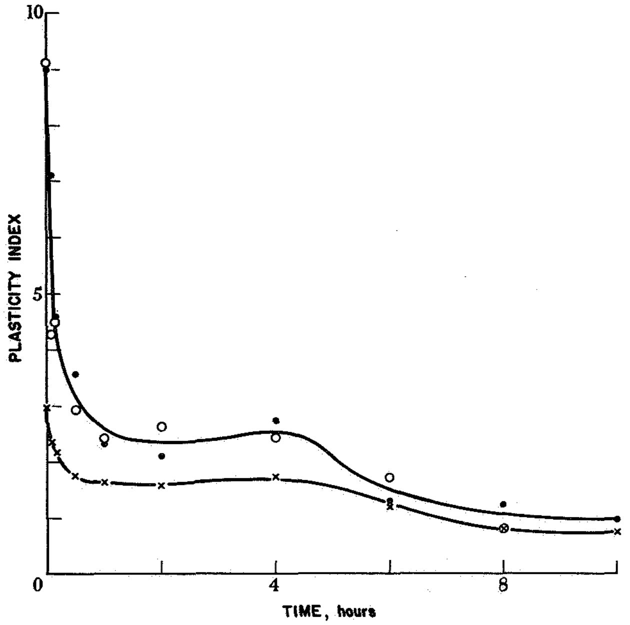


FIGURE 15.—Effect of continued sliding on the plasticity index. The behavior of three mild steel specimens slid against a stainless-steel woven grid at 100 cm/s under loads of 1.5 kg: ● and ○ show results for two specimens with initial plasticity indices of 9; the crosses show the behavior of a specimen which had a lower (though still plastic) value.

CONCLUSION

The new techniques of observation and measurement of surface structure described in this paper have given a picture of the topography of solids which in several ways is different from earlier views. The salient points can be summarized as follows. The first three are based on general theory and observations of a wide range of surfaces; the last three refer to experiments on bead-blasted surfaces, but there seems to be no reason to doubt that these conclusions are equally general.

1. Surfaces typically are covered with asperities which are from 10 to 300 μin. high, and spaced from 50 to 3000 μin. apart. Their slopes occasionally are as steep as 25° but usually lie between 5° and 10°. The range of summit radii is very large. Many surfaces have mean summit radii of 400 to 800 μin.; but values of over 0.02 in. are not unusual. A summit 4 μin. high may have a radius of 0.04 in.; it is certainly not true that the radius is always of the same order as the height, as is sometimes assumed.

2. Many surfaces have height distributions which are Gaussian. In nearly all cases, the height distributions of the upper half of the surface and of the summits of the asperities are accurately Gaussian. (Strictly, the heights of the upper half of the surface do not form a Gaussian distribution; they are the truncated upper half of such a population). Apart from its mathematical convenience, the significance of the Gaussian distribution in contact theory seems to be that over the physically important region it approximates to an exponential distribution. It can be shown that for the exponential distribution there is exact proportionality between the number of contacts, the area of contact, and the load; and thus that the average size (and the distribution of sizes) of a contact area is independent of the load. It follows that the two basic laws of friction are a direct consequence of the topography of real surfaces. Their existence is not the result of any particular material properties.

3. The contact between solids is controlled by two material properties: the plane-stress elastic modulus and the hardness; and three topographic properties: the surface density of the asperities, the standard deviation of their height distribution, and the mean radius of their summits. These combine to give a generalized surface roughness parameter involving both material and topographic properties, the "plasticity index." The commonly quoted parameter "center line average" is useful only in that it is an approximate measure of the spread of asperity heights. For most surfaces, the deformation mode cannot be affected by changes in the load. It will be elastic if the plasticity index is low, and plastic if it is high.

4. When two surfaces touch under light load, contact is most likely to occur above the 90th percentile. Even at high loads, contact is extremely improbable below the center line. The separation is not very sensitive to the pressure; in fact the mean planes of two similar surfaces in contact are usually separated by one to two times the standard deviation, or roughly by the center line average. This means that the average gap between 20 μin. CLA surfaces is, for a wide range of loads, approximately 20 μin. This explains the difficulty of making hermetic metal-to-metal seals between nominally flat surfaces (as, for example, in space vehicle applications).

5. Asperities are remarkably persistent under heavy loads. They will transmit forces large enough to cause the bulk solid to flow without themselves becoming flattened; indeed they can withstand local pressures of at least ten times their nominal yield pressure.

6. At very high loads, i.e., those creating pressures equal to or exceeding the bulk yield pressure of the solid, there seems to be a change in the nature of contact which shows as a bend in the separation-load, real area-load, and number of subcontacts-load graphs. This

may correspond to the stage at which asperities cease to deform independently and begin to react as a coherent block.

These recent observational and experimental techniques have provided a stimulating, new viewpoint on surfaces and surface contact. They have led to several new ideas about contact under static conditions. It is not yet clear, however, how directly this new knowledge of surface topography may be applied to sliding. This is now a vital area of interest in surface physics because we have almost no detailed observations of the topography of sliding contact. These techniques are capable of being extended, and further experiments are in progress in the Burndy Research Laboratory to study the sliding case. Meanwhile, it would be prudent to regard views which have been established quantitatively under static contact conditions as merely qualitative when applied to sliding contact.

DISCUSSIONS

E. P. Kingsbury (MIT Instrumentation Laboratory, Cambridge, Massachusetts)

This paper presents a careful and detailed summary of the study of contacting surfaces now in progress at the author's laboratory, against a background of the historical development of the subject. We are indebted to Dr. Williamson and his colleagues for their contribution toward a useful understanding of this topic.

What quantities are necessary to characterize the fine structure of a flat metal surface? How may these quantities be measured? What conclusions may be drawn concerning real areas of contact, asperity deformation, normal approach, friction and wear processes, when two of these surfaces interact?

One emphasis in Dr. Williamson's paper is on the use of profilometry as an experimental technique. He deals effectively with the questions of surface damage, resolution, and applicability, comparing electron microscopy and interferometry as alternatives. He considers profilometry to be superior partly because it can be coupled with a computer for manipulating large data inputs in various ways, even though there is some sacrifice in resolution. Recently there has been considerable interest in accounting for Amontons' Law, "contact area proportional to load," by means of an asperity deformation theory. Dr. Greenwood has given an elegant proof that for a certain surface topography, the law follows independent of asperity shape or deformation mode. The essential part of the argument is that each increase in load will form new small contacts which compensate for increases in existing contacts so that the average area per contact remains constant. There is then a theoretical interest in those parts of the surface giving small contact areas. Referring to figure 1(A), we see that the smallest asperities are easily resolved by the electron microscope, but that even if the best resolution quoted by Dr. Williamson is invoked, they would be overlooked by his computer.

The author has found many types of surface to have a Gaussian height distribution and claims that their cumulative height distribution curves may be used to describe them. If I understand him correctly, these curves are, or could be, obtained from a single traverse of the surface, yet they show percentage of surface, i.e., area, as ordinate. The transition from linear measurement to areal result is made by arguing that the surfaces are random so that a profile in any direction will give equivalent information. With nonrandom surfaces, the profiles "must be made so as to include a representative sample of the topography." This raises the question of what is meant by a random surface. By itself, a Gaussian height distribution need not imply randomness. One can imagine a surface consisting of a single symmetric continuous asperity having a Gaussian distribution of ordinates. A truly random surface, with each height completely independent of its neighbors but possessing the same height distribution, is harder to visualize. Perhaps it resembles a vapor rather than a solid. The real surface (with the same distribution) lies between these extremes; some further information is necessary for a complete description. As an example, refer again to figure 1(A). Two classes of asperities, large and small, are present. An observer located on a large asperity is most likely surrounded by other large asperities, and conversely. This is not a random topography; how do we choose a profile direction beforehand which will include a representative sample of this topography? A quantity mentioned by the author, but not expanded upon, which could be useful in this connection is the autocorrelation function for the series of height readings making up the profile. The distribution of the slopes of the surface and its deviation could also be useful and should be available from the computer.

A further shortcoming of the height distribution curve as a specification is revealed when it is needed to make a micromap of a surface. In this case the author must use a multiple profile technique which requires a careful height and horizontal reference to be maintained between successive profiles. The author obtains the profile spacing by a consideration of the autocorrelation function along a profile. A three-dimensional description of the surface, which cannot be obtained from its height distribution curve, emerges from the computer.

The experimental difficulties of the multiple profile technique perhaps approach those required for an interferometric analysis (which yields the same information directly). It would be interesting to be able to compare an interferograph of a suitable surface with its micromap as supplied by the profilometer.

The section of the paper on asperity deformation describes some experiments which show that, at least for aluminum, one cannot cause an asperity to flow by the application of the bulk yield pressure to it. One gets the impression that the asperities remain elastic for very large stresses while the bulk material flows plastically under them. Dr. Williamson rejects this view on the grounds that his setup requires the whole specimen to have yielded. Assuming that the asperities are the same chemically as the bulk aluminum, is it fair to attribute the same laws of deformation and the same physical properties to them as are found for the bulk material? In general, materials become stronger as they are tested in smaller sizes; the asperities are very small indeed. It is not inconceivable to realize an increase in yield strength for aluminum by a factor of 8 merely by severe coldworking such as would be suffered by an asperity during bead-blasting. (This point could be checked by annealing after bead-blasting.) It does not then seem unreasonable to consider the behavior of figure 14 as a delayed elastic-plastic transition. If this is a general result, the question of what fraction of asperities would have yielded plastically had they had the bulk hardness is irrelevant.

With regard to friction and wear, Dr. Williamson is concerned with the influence of topography in general, and his plasticity index in particular, on friction and wear phenomena. His index depends on two topographical properties: mean asperity radius and standard deviation of asperity height; and two bulk material properties: elastic modulus and hardness. The theoretical treatment that results in this index is concerned strictly with the normal approach of a rough surface onto a plane. The problem of the approach of two aligned rough surfaces is considered to be equivalent to that of a (different) rough surface onto a plane. But when two real rough surfaces are contacted, most of the asperities will undergo a combined loading, producing stress patterns not included in a Hertzian type analysis. This is true without any overall tangential movement whatever. With gross slip, the divergence is greater. The author does not mention any effect of surface contamination which will certainly be present, although his wear experiments are carried out in the presence of a boundary lubricant (under conditions which make hydrodynamic lubrication a possibility). One can change wear rates by several orders of magnitude through the introduction of a boundary lubricant, yet the lubricant does not change the height distribution, the mean asperity radius, the elasticity, or the hardness of either sliding member. What it does change is the strength of their tangential interaction.

In summary, I feel that the author's comments are well taken insofar as they concern the description of a surface, methods for measurement, behavior during normal approach, estimation of contact areas and their dependence on load, and so on. This work offers an excellent basis for a more detailed study of friction and wear, but as it stands should not be extended too far without some consideration of tangential effects.

A. J. Melmed (National Bureau of Standards, Washington, D.C.)

I wish to compliment Dr. Williamson for writing a very enjoyable and informative exposition. In general, I consider it a good review; however, it contains some logical inconsistencies, raises questions about technical details, and advances conclusions which are not, in my opinion, justified. I will discuss these in the given order, although there may be some overlap.

Logical inconsistencies.—At the beginning, the author states that progress in the study of surface contact is closely correlated with advances in observation and measurement techniques. Later it becomes clear that all of the surface contact problems have not been solved, and there is still room for considerable progress. Yet the author would have us believe that advances in observation and measurement techniques, such as improved resolution afforded by finer horizontal discrimination in profilometry and experiments using electron microscopes and field-emission microscopes, are not really worth pursuing. Now these advances are aiming in the direction of providing us ultimately with the ability to characterize surface contacts and underlying regions in atomic detail. It seems that the achievement of this ideal would indeed shed considerable light on the metallurgical aspects of surface contact.

The author states that "Early work was . . . hampered . . . by total ignorance of the concept of surface cleanliness," and yet surface cleanliness or lack of it is not considered by the author in discussing his experimental procedures and results. Surface cleanliness is a very large factor affecting experimental results in friction studies (refs. 41 through 43).

The author suggests in his conclusion that "the two basic laws of friction are a direct consequence of the topography of real surfaces. Their existence is not the result of any particular material properties." He also states that the contact between solids "is controlled by two material properties . . . and three topographic properties. . . ." These statements are certainly contradictory.

Questions about technical details.—In the discussion of recently-performed experiments on track size in tactile profilometry, aluminum was chosen "as it is one of the softest metals on which profiles are made." However, the surface studied almost certainly was oxidized aluminum. Care was taken to use very light stylus loading. This raises the question of whether the tracks were, in fact, made in aluminum or in an aluminum oxide layer, which might really be a rather hard surface. (Similar experiments on a gold surface would be more convincing.)

Some questions arise concerning the experiment on surface contact using an aluminum specimen under compression. What were the time durations in the applications of the successively larger loads? If the profilometer traces of figure 12 were taken after short-time application of loads, it may be that the asperities were actually continuously deforming and that longer load times would produce further deformation, as in a typical creep experiment. Under such circumstances the results would not be particularly remarkable.

How much work-hardening can one expect for the asperities, and how completely does this explain the quantitative results? Certainly increases in tensile flow strength by factors of 8 have been measured for aluminum at plastic strains of 5 percent (ref. 44).

The author states that the material which disappeared from the asperity tops is accounted for by a redistribution of volume in the asperities, neighboring valleys, and neighboring asperities. How can such extensive material flow occur inside the sample which is under the great compression condition of the experiment? It seems more reasonable that the asperity tops have either fallen away or been distorted in a manner which would not be seen by the profilometer. In this regard, would the profilometer stylus faithfully reproduce the shape of an anvil head or other shape involving reentrant side walls? One further comment here. Caution should be exercised in generalizing the results of this experiment, done under very specialized conditions, to describe the deformation of asperities on real-world surfaces.

Concerning conclusions.—The word "surfaces" as used in the conclusion is far too general. Strong explicit qualifications should be made. Otherwise, some of the statements made by the author are not believable.

The conclusion purporting to explain the "difficulty of making hermetic metal-to-metal seals" ignores factors such as overall flatness, surface cleanliness, and proper choice of materials. When these factors are properly considered, there is no difficulty in mating gas tight metal-to-metal seals, using high loads. In fact, part of the great success of modern metal vacuum systems results from the ease of routinely achieving such seals. The author should qualify the conclusion instead of making a general statement.

I have already commented on the contradiction in the author's conclusions regarding the controlling factors in contact and friction. Furthermore, if one accepts the author's conclusion that "Surfaces typically are covered with asperities" having a specific range of geometries, then it would seem that different materials would have the same friction coefficients if topography alone were the governing factor. Most tables of experimentally measured friction coefficients for different materials clearly show that this is not so. The last conclusion in the paper apparently is derived from the data of one set of experiments (done under the very special conditions of high sample compression). A statement regarding asperities in general, therefore, does not seem appropriate.

Finally, I would like to suggest that two instruments, one still in the development stage and the other commercially available, could contribute significantly to advances in understanding contact and friction. The field-emission ultramicrometer (ref. 45) promises to do profilometry with a vertical accuracy of better than 10^{-6} cm and a horizontal discrimination of about 10^{-5} cm, without touching the specimen. The scanning electron microscope (ref. 46) can look at relatively large surface areas, revealing the shape of surface asperities with sufficient depth of focus to see into reentrant regions.

T. E. Tallian (SKF Industries, Incorporated, King of Prussia, Pennsylvania)

The author gives a progress report of his studies of surface topography, focused on the use of stylus-type surface profile tracers. Known tracing systems are used by the author in conjuntion with two innovations: a coordinate table and a digital method of data recording and analysis. Of these, the more unique is the coordinate table. This permits, for the first time, instrumental analysis of a two-dimensional surface. Digital processing of profile information is also relatively new but has been performed in both the author's and the discussor's laboratory for a few years (ref. 47).

The illustrations given by the author of the use of his surface analysis method are both fascinating and frustrating: fascinating, of course, is the wealth of detail revealed about the surfaces, but frustrating is the fact that the surfaces illustrated are bead-blasted, with a topography so far removed from most typical sliding and rolling surfaces that its selection must have been related to uses for electrical contacts (the main area of interest of Burndy Corporation) rather than for lubricated mechanical contacts.

The author correctly states that it is insufficient to know the rms height of asperities in predicting plastic or elastic contact behavior of surfaces. Williamson, based on his and Greenwood's analysis, favors the plasticity index

$$\frac{E'}{H}\sqrt{\frac{\sigma}{\beta}}$$

This Greenwood index is based on the assumption of hemispherical asperity shapes. Blok and then Halliday (ref. 48), using a sinusoidal asperity model, earlier proposed an index based on the maximum asperity slope θ, and plastic flow occurs when the index exceeds a limiting value, depending on the material.

The latter index has the advantage that the asperity slope θ is easy to measure. As an example, figure 16 shows distributions of slopes for three typical, abrasive-finished surfaces. They were obtained by the tracing and processing technique described in reference 47 for asperity amplitudes, except that an analog differentiator was inserted in the circuitry. From these distributions, it is seen that the typical slope of abrasively-finished hard steel surfaces varies widely. On a ground surface, the 95 percentile slope is $\theta_{0.95}=29°$. Even for much smoother honed surfaces $\theta_{0.95}=3.8°$, and only lapped surfaces show $\theta_{0.95}=0.7°$. Halliday cites $\theta_k=1.6°$ as the limiting slope beyond which plastic flow of asperities occurs on hard steel, and thus there must be significant plastic flow of asperities on new ground surfaces and some on honed surfaces. Only lapped surfaces would be expected to be purely elastic when new. It would be very interesting to compare these results with predictions using the Greenwood index, but unfortunately, the asperity tip radius for these surfaces is not known at this time.

The author quotes Archard to the effect that linearity of total real

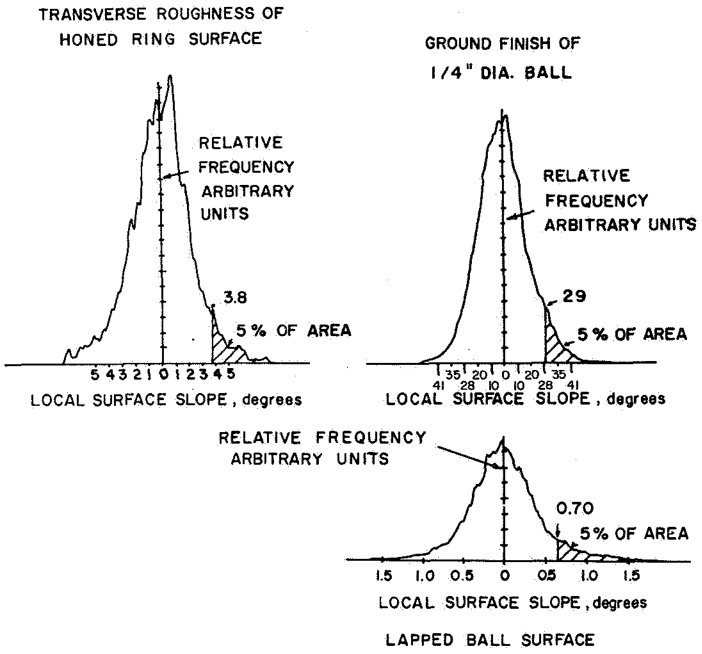


FIGURE 16.—Asperity slope distribution on ground, honed, and lapped surfaces.

contact area with load is to be expected if an increase in load results in more numerous, not larger, individual contact areas. In deriving their plasticity index, Greenwood and Williamson state that the wide applicability of a single plasticity index, independent of load, results from the fact that for a variety of plausible asperity tip height distributions, such as Gaussian, the contact areas (of hemispheric asperities) do become more numerous with load at just the right rate to keep stress within an asperity essentially load-independent.

Another view of the plastic flow criterion is possible and is important in the analysis of partial elastohydrodynamic conditions in rolling contact. It will be shown that a load-independent plastic flow criterion applies to any single asperity, irrespective of the statistical height distribution of a population of asperities and thus independent of the rate of total contact area growth with load, if only the profile of the single asperity satisfies a simple geometric condition. This condition is that the Hertz contact area of the asperity must increase proportionately with its elastic deflection under load (squashing). Figure 17

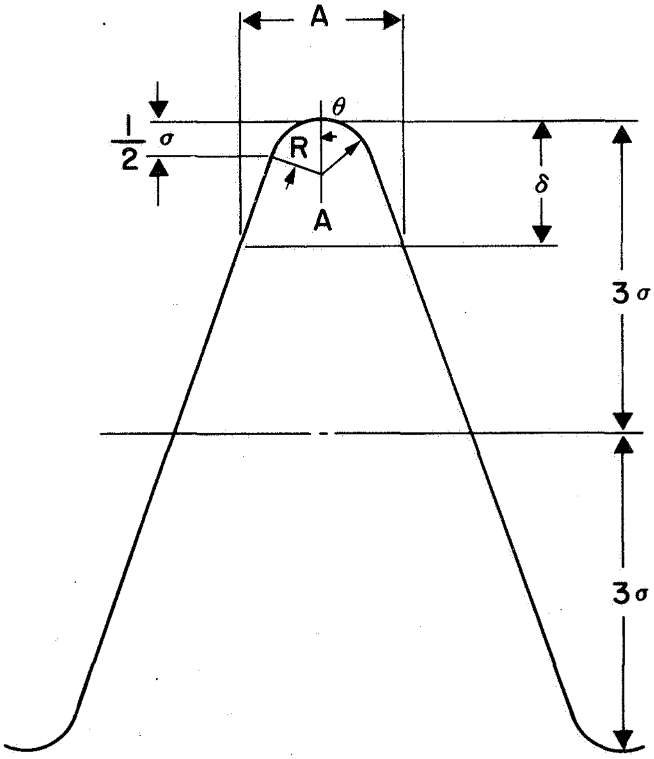


FIGURE 17.—Asperity model cross section.

shows an idealized two-dimensional asperity shape, similar to what one would expect to find on ground, honed, or lapped surfaces, formed by a multitude of cuts by sharp and straight-edged abrasive grains. The asperity is a plane-sided ridge, with a tip radius so selected that the rounding of the tip extends over a small constant fraction of the asperity height ((½) σ of the rms height or $1/12$ of the average total height). Figure 18 shows the elastically calculated maximum subsurface shear stress $\tau_{\max}$ and its depth coordinate $Z_{\max}$ beneath the asperity contact surface, as functions of contact area width (and, approximately, deflection) for three slopes and rms amplitudes. (These slopes and amplitudes are typical of real surfaces described in figure 16.) The stress calculation follows a theory developed by Chiu of the discussor's laboratory and assumes the plane sides to be infinitely long in both directions. Thus, it is valid for deflections which are small compared to the asperity height 6σ. It is seen that the maximum shear stress (which is critical for the onset of plastic flow) increases with load for low loads, as expected for a cylindrical contact. Then, however, the stress levels off as more of the plane sides of the asperity come into contact. Each of the stress curves has a horizontal asymptote which is approached within 20 percent for deflections in excess of 1.5σ. The magnitude of the asymptotic value of $\tau_{\max}$ depends only on the asperity slope, provided that the height occupied by the tip radius is small in comparison to the deflection. Compared with a plastic flow stress in Hertz contact of about $\tau_k = 120\,000$ psi for steel of 60 *RC* hardness (ref. 49), the asymptotic value of the calculated shear stress predicts elastic deflection for all loads for the smoothest surface and plastic flow for the two others, in agreement with Halliday's criterion.

This calculation contains no assumption regarding the height distribution of a population of asperities. A similar result can be obtained for a three-dimensional asperity contact if the profile of the asperity is a suitable nonparabolic curve, giving proportionality between deflection and contact area.

These deformation phenomena suggest a series of questions of central importance to rolling contact performance. As described elsewhere (ref. 50), surface failure in rolling contacts can take three forms: surface fatigue characterized by near-surface plastic occurrences and subsequent microcracking and pitting (often a precursor of spalling failure which, of itself, is not a failure of the surface), wear with loose particle removal, and smearing (metal transfer). Of these, smearing requires substantial macroscopic sliding in the contact and will not be considered here. Surface fatigue and wear have in common the facts that they can occur with little, if any, sliding, and that neither occurs under full elastohydrodynamic conditions, i.e., without asperity contact. But otherwise the two differ.

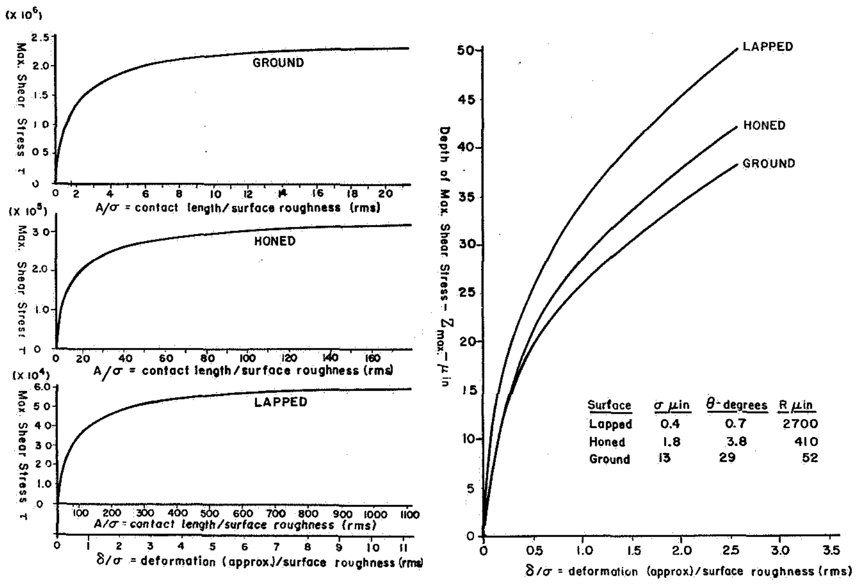


FIGURE 18.—Maximum shear stress and its depth coordinate under the asperity of figure 17.

Surface fatigue does not entail material removal (except in extreme cases) and seems strongly controlled by surface geometry and material. It is secondarily influenced by some lubricant properties. What is called surface fatigue actually initiates in the near subsurface strata (ref. 49) in that plastic deformation zones form at depths of a few hundred μin. in the metal structure, just as they form in the deep subsurface region in the course of common spalling fatigue. Microcracks develop later, apparently originating in the plastically worked near-surface metal but open to the surface (ref. 51).

Wear, of course, is a material removal phenomenon, acting on one asperity at a time, and is strongly influenced by the lubricating properties of the fluid or solid material between the surfaces (ref. 47). By all theories, wear is a traction phenomenon: asperities must weld or strongly adhere so that tensile forces are transmitted before they can wear away, unless abrasive particles intervene, acting by plowing, i.e., again in traction. The presence of even small sliding velocities rapidly increases wear. Surface fatigue appears to be a phenomenon not dependent on macroscopic sliding; in fact, if there is much sliding between the surfaces, it may not develop but be overtaken by smearing failure.

It is intriguing to speculate on the fact that neither wear nor surface fatigue occurs in the presence of an uninterrupted elastohydrodynamic film. In the case of wear, the reason seems obvious: the elastohydrodynamic film cannot transmit the high tractive forces required for a wear particle to form. Considerable shear stresses can be generated in an elastohydrodynamic film, but as pointed out by Smith (ref. 52), they are extremely unlikely to reach the order of flow stresses of hard metals.

This discussor has previously proposed that a full elastohydrodynamic film prevents surface fatigue because of the absence of high tractive forces (ref. 50). This hypothesis is, upon close scrutiny, awkward. It may be true that in any moving contact, when metal touches metal with at most a boundary layer of lubricating material intervening, there always is substantial shearing of the interface because even small sliding velocities impose large magnitudes of shearing on the very thin boundary layers. It is not necessarily true, however, that high tractive forces are generated. If the interposed layer is of low shear strength (and boundary lubrication provides just such layers), then high forces will not act. Thus, high tractive forces will arise only where the boundary layer has been pierced (the typical wear situation). Since surface fatigue often arises over large surface areas in the presence of good boundary lubricants without the accompaniment of massive wear, it is difficult to believe that it results from the few metallic asperity interactions that take place.

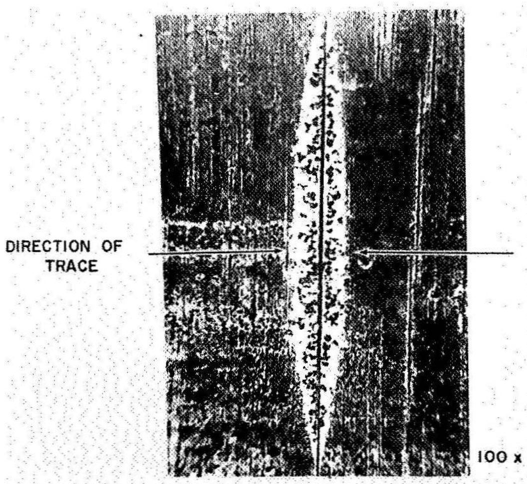
The author's description of compressive asperity deformation under very high pressures offers an explanation for surface fatigue that does not require tractive forces. To begin with, inside an elastohydrodynamic film, in the high pressure region of a rolling contact, the surface pressure does not vary appreciably over distances comparable to asperity spacings, at least over surfaces with moderate asperity slopes, such as the honed and lapped surfaces illustrated in figure 16. Therefore, subsurface shear stresses, which are related to the presence of surface pressure gradients, will be low at depths comparable to asperity spacings. Asperities separated by elastohydrodynamic film thus have no reason to undergo plastic deformation requiring high shear stresses. That they, in fact, do not undergo such deformation is suggested by the as-new appearance of surfaces that have withstood many (10^9) cycles of Hertz stressing with full elastohydrodynamic film. Now, if asperities touch because the film thickness/composite roughness ratio, h/σ, has dropped and negative separation points have arisen in the author's "gap map," then the most obvious change in the surface pressure field at these asperities will be that a sizeable pressure gradient arises; the pressure is higher over the asperity contact areas because these must be depressed to pass over one another. Assuming that continuum mechanics calculations are permissible, high stresses similar to those shown in figure 18 will be generated, even under small total Hertzian loads. Their depth below the surface will be of the order of asperity spacings, as shown in the same figure. At that depth, near-surface plastic flow will arise if the stress exceeds the flow stress of the material. Microscopic observation (ref. 51) has shown that plastic flow does occur at these depths if there is visible evidence of surface fatigue.

Experience shows that such evidence is likely to be found in rolling bearings if $h/\sigma \leqq 1$ and not likely if $h/\sigma \geqq 2$ (ref. 50). Apparently, real asperity shapes are such that a significant amount of plastic flow does not occur at typical asperities unless their deflection is large by comparison to their amplitude.

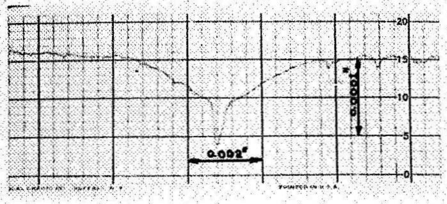
It seems confirmed that the first stage of surface fatigue is plastic flow under contacting asperities. It has been shown that this type of plastic deformation is progressive for hundreds of millions of cycles by involving increasing numbers of favorable martensite platelets (ref. 53). Eventually (and mysteriously) a crack forms in the plastically worked zone. That is the genesis of the second stage of surface fatigue: microcracking.

In the absence of localized geometrical defects and for good (largely elastohydrodynamic) lubrication, it appears that spalling failures in rolling contact (the typical, deep, crater-like spalls which can destroy rolling bearings) originate at the depth of high microscopic Hertzian shear stress (several thousandths of an inch below the surface). However, it is also known (ref. 49) that for low values of h/σ or in the presence of surface defects, spalling is often surface originated. This has led to considerable speculation about hydrostatic crack propagation (ref. 54) and the tractive forces acting in these cases to originate cracks at the surface; the macroscopic Hertzian stress field, in the absence of traction, does not provide the shear stresses at the surface which are believed necessary to start cracks. Recently, workers at A. D. Little Incorporated have shown that if a crack (or void) is already present, a Hertzian compressive stress field can propagate it because there will be shear stress concentrations at the tip of the crack or near the wall of the void (ref. 55). Using this theory, one can see that surface-initiated spalling failure can propagate either from a virgin (unfatigued) surface if there is a preexistent localized sharp discontinuity (crack), or from any surface that has undergone surface fatigue which opens up huge numbers of shallow but sharp cracks. Microscopic evidence exists to show that surface-initiated spalling is generally preceded by surface fatigue (near-surface plastic flow), i.e., that it follows the latter route to failure (ref. 53).

The above description shows how typical asperities of rolling surfaces can initiate surface fatigue when they make contact through a partial elastohydrodynamic film. Another important mode in which surface fatigue can occur is by local thinning of the elastohydrodynamic film. Figure 19 shows intensive surface fatigue surrounding a relatively deep grinding "furrow" on a rolled-over surface. It has been shown by Schoeler in the discussor's laboratory that during cyclic stressing the oval area of surface fatigue acquires a slope towards the furrow (fig. 19). It is not yet known whether the slope results from wear or from subsurface plastic flow, but it was demonstrated (ref. 53) that plastic flow takes place. This series of events at the furrow is attributed to localized thinning of the elastohydrodynamic film in the area surrounding the furrow because of side leakage of oil into the furrow, to the resulting pressure loss, and to the consequent elastic rebound of the material. It appears that a "dynamic asperity" is created near the furrow (much the same as one is created near the exit of every elastohydrodynamic contact) and at the sides of three-dimensional contacts (ref. 56). If this "asperity" makes contact, it may cause surface fatigue. Once surface fatigue has arisen at a furrow, spalling failure can result and, indeed, numerous spalls originate at furrows.



GRINDING FURROW WITH SURFACE FATIGUE



TRACE THROUGH GRINDING FURROW WITH SURFACE FATIGUE

FIGURE 19.—Grinding furrow with surface fatigue.

It is suggested that by using the following parameters one can predict whether surface plastic flow—and ultimately surface fatigue—will occur on rolling contact surfaces.

(1) A quantity giving the variation of typical asperity cross sections as a function of the "elevation" of the section determines the stress-deflection relationship of the asperities and yields a plasticity criterion.

(2) The magnitude of the elastohydrodynamic film thickness/composite roughness ratio, h/σ, determines the magnitude of the deflections imposed on those asperities that come into contact through the elastohydrodynamic film.

Of these, the quantity in (1) can be approximated by the use of a suitable plasticity index, e.g., one related to the asperity slope θ which is readily measured. There are several current (electrical and optical) methods to measure h/σ as stated in (2) in areas where the surface is undisturbed by defects (ref. 47).

A new optical method to elucidate the topography of an elastohydrodynamic contact area in the presence of localized surface defects is illustrated in figure 20. This is an interferogram showing an elastohydrodynamic point contact between a steel ball and a glass flat (ref. 56). It is different from the pioneering interferograms of Archard (ref. 57) and Cameron (ref. 58) in that it was taken with an ultra-

FIGURE 20.—Elastohydrodynamic contact interferogram with imperfection marked with arrow.

viscous (10^6 cs) oil at very slow (10^{-2} in./sec) rolling speed. As a result, one observes clearly a surface defect (scratch) passing through the contact and distorting the elastohydrodynamic film. The horizontal magnification is × 600, so that the resolution is very good. This type of interferogram permits one to follow the action of rare surface features in elastohydrodynamic contacts.

In summary, the author, with his two-dimensional tracing method and with the study of asperity plasticity, has contributed significant and sorely needed solutions to the complex problem of surface geometry and asperity behavior in contact. His imaginative and new approach to the problem has yielded insights that serve well in the interpretation of surface fatigue phenomena. It is hoped that he will continue to have the means and the inclination to pursue these studies further.

J. F. Archard (University of Leicester, Leicester, England)

Dr. Williamson's stimulating paper contains so much of interest that I must necessarily confine my remarks to a single topic. The following discussion is based upon a program of research carried out in collaboration with Mr. D. J. Whitehouse of Rank Taylor Hobson, Limited.

Dr. Williamson shows that for a profile having a Gaussian distribution of heights, the distribution of peak heights is also Gaussian. We ask the question: is this to be expected on theoretical grounds? Recently we have been developing a theoretical analysis of random surface profiles, beginning with the simple assumption that such a surface is defined by two parameters: the height distribution and the correlation length. The correlation length ($\beta_{\max}$) is defined as the length over which the autocorrelation function has declined to a defined, negligibly small value; its physical meaning is that two points on the profile separated by a distance $\beta_{\max}$ can be regarded as statistically independent events, but that points separated by distances less than $\beta_{\max}$ are not statistically independent.

We start by assuming that the height distribution is Gaussian and confine ourselves, here, to outlining the way in which the peak height distribution can be deduced. Consider three events upon a profile separated by distances equal to $\beta_{\max}$, as shown in figure 21. The probability of finding a peak at, or near, a height y can now be defined by three conditions:

(1) Event 1 is lower than y.
(2) Event 2 is at, or near, height y.
(3) Event 3 is lower than y.

The probabilities of each of these events is shown by the shaded portions of the height distribution functions (fig. 21).

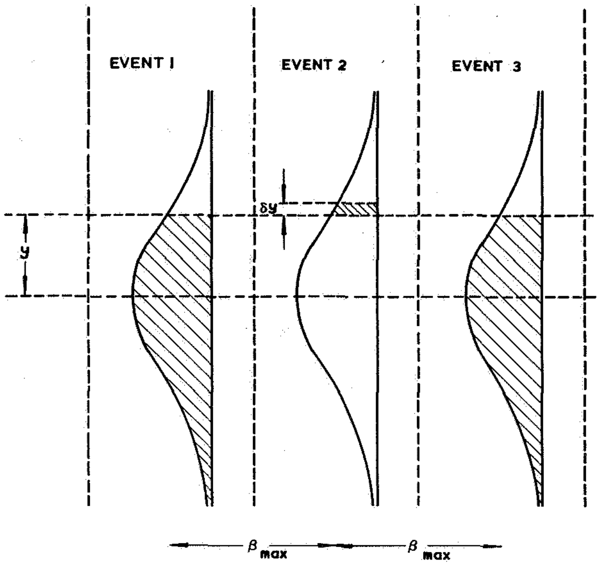


Figure 21.—Illustration of simple theory, the probability of finding a peak between y and $(y+\delta y)$.

Using this approach, it has been shown that if one assumes a profile having a Gaussian distribution of heights, of standard deviation σ, the distribution of peak heights is also approximately Gaussian. This peak height distribution has a mean which lies above the mean height by a distance of 0.81σ, and the standard deviation of the peak height distribution is 0.71σ. These relationships between the height distribution and the peak height distribution are not identical with those measured by Greenwood and Williamson using a bead-blasted surface; however, considering the simplicity of the present analysis, the agreement is sufficiently close to suggest that further development along these lines might be appropriate.

A more complete account of the work just outlined will be published elsewhere. In the present context it is, perhaps, relevant to consider ways of extending the excellent work on the analysis of surface profiles which Dr. Williamson has described. It seems reasonable to suggest that the statistical characteristics of any random surface should be described completely by the height distribution and the autocorrelation function, and that both of these parameters are required for a complete statistical description. (It is of interest to note here that, in his invited discussion, Dr. Kingsbury also suggests the inclusion of the autocorrelation function.) From these two parameters it should be possible to forecast any other statistical characteristic of the surface that might be required, e.g., peak height distribution, peak height radius distribution. A somewhat weaker suggestion is that intuitively we might expect the shape of the autocorrelation function to be basically the same for a considerable range of surface profiles. If this be true, then for this class of surfaces it should be acceptable to simplify the analysis using the correlation length (rather than the full autocorrelation function) somewhat along the lines indicated in the simple analysis outlined above.

Finally, the use of surfaces having a Gaussian distribution of heights has attracted some questioning comment. I would commend Dr. Williamson's emphasis, at this stage of the subject's development, upon the analysis of this type of surface. Indeed, I would suggest that the further development of this same study towards a more complete analysis, perhaps along the lines suggested above, is desirable as the next stage. The completion of this analysis seems to be a necessary prerequisite for the understanding of the characteristics of other surface types less clearly defined in terms of statistical theory.

J. W. Lancaster (Royal Aircraft Establishment, Farnborough, England)

The plasticity index, as defined by Dr. Williamson, and involving both material properties E/H and characteristics of the surface topography σ/β, is a particularly valuable concept. It enables one, at least in principle, to rationalise much previous experimental data which has so far merely been either qualitative or only semiquantitative.

One of the long-term objectives in the fundamental study of wear is to relate the wear rates of materials to their mechanical properties. In most laboratory experiments, rates of wear are measured after the attainment of steady state conditions between surfaces sliding repetatively over the same apparent areas of contact, e.g., pin and disc tests and plain bearings. In such conditions it is the exception rather than the rule to find significant correlations between the wear rates and mechanical properties of different materials. The main reason is that each material generates its own characteristic topography, i.e., the ratio σ/β changes with the material as well as E/H. This point can be illustrated by describing some results obtained with different grades of carbon brushes sliding on copper slip rings.

The elastic moduli of manufactured carbons are, in general, about two orders of magnitude lower than those of metals. The slopes of asperities which can be deformed elastically are thus in the range of about 5 to 15°, in contrast to 0.1 to 0.5° for metals. On most surfaces produced by conventional routine engineering methods, it may therefore be assumed that the deformation of carbons will be predominantly elastic and, if so, an attempt to relate their wear rates to E is reasonable.

The results shown in the upper part of figure 22, obtained with different types of carbons at two loads, indicate that there is no significant inverse relationship between the wear rates and the elastic moduli of the various materials. Microscopic observations showed that some carbons wore the counterface very badly, whereas others hardly affected it at all. The lower part of figure 22 shows results obtained in conditions where a constant surface topography was maintained on the copper slip ring; each brush was worn at a light load (80g) in the wear track generated at a heavier load (1 kg) by one particular grade of brush. Three sets of results are shown, corresponding to three grades of "track-generating" brushes. There are now significant linear inverse relationships between the wear rate and elastic modulus. What remains to be done is to characterize the topography generated during sliding in detail, i.e., to determine σ/β with the aid of the sophisticated techniques described in Williamson's paper. In this way it should be possible to determine which material properties are of most importance in determining σ/β and which determine the wear rate; these two sets of properties are not necessarily the same.

An alternative way to keep the topographical features constant during wear is to spiral a small sample of material over a larger rotating

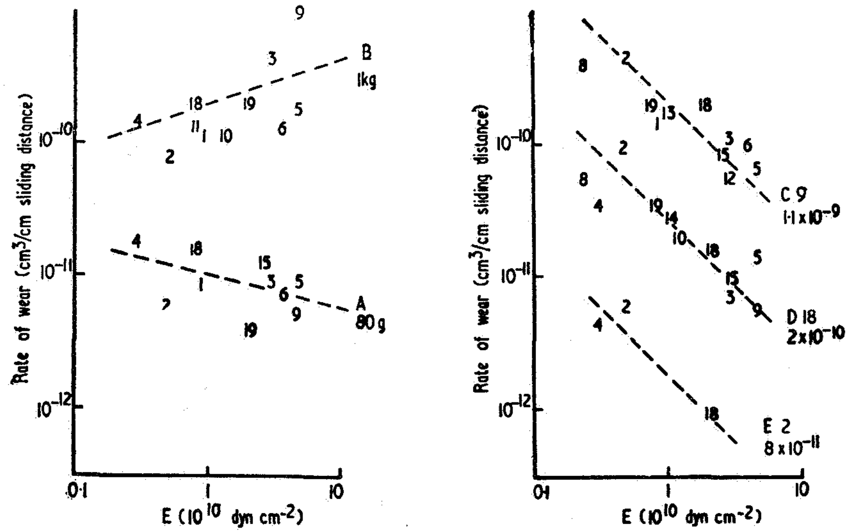
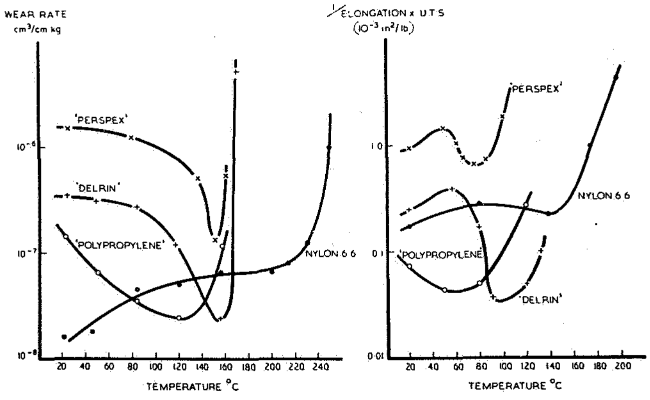


FIGURE 22.—Variation of the rate of wear with modulus of elasticity: A, in own track, 1 kg; B, in own track, 80g; C, in a baked carbon, 9, track, 80g; D, in an electrographite, 18, track, 80g; E, in a natural graphite, 2, track, 80g. The digits are the code numbers of each material, and the figure below it the rate of wear.

drum. This kind of approach has recently been used at R.A.E. with the object of relating the wear rates of polymers to their mechanical properties. In one series of experiments, the drum temperature was varied, and the left-hand side of figure 23 shows the variation of wear rate with temperature for four materials. Values of E/H were obtained from indentation measurements but generally showed little variation with temperature until very near the softening points of the polymers. Other mechanical properties are therefore likely to be involved. Ratner and coworkers (ref. 59) have suggested that one of the most important parameters in polymer wear is the elongation to break (ϵ). Qualitatively it is clear that the magnitude of the elongation together with the ultimate strength must determine whether material is ruptured beneath a moving penetrating asperity, leading to a wear particle, or merely displaced plastically. The right-hand side of figure 23 shows values of $1/\text{U.T.S.} \times \epsilon$ plotted against temperature for the polymers already given. It may be noted that the shapes of the curves are very similar to those for the wear rates. An exact correspondence is hardly to be expected in view of the fact that mechanical properties measured at low rates of strain are being compared with a wear process involving much higher strain rates. The results therefore suggest that some caution must be exercised in extending or extrapolating the use of criteria, such as the plasticity index, from static contact conditions to situations involving sliding.

In a second series of experiments, the surface topography of the drum was changed by abrasion with different grades of silicon carbide paper while the temperature was kept constant (low speeds of sliding).



ABRASIVE WEAR

ELONGATION x U.T.S.

FIGURE 23.—Wear rate and elongation × U.T.S. of four materials vs temperature.

Figure 24 shows that the wear rates of polymers are extremely dependent on the roughness of the counterface, and much more so than those of metals. The relationships obtained with metals are those to be expected assuming plastic deformation and "filing" by relatively hard asperities on the counterface. For the polymers, however, the mode of deformation changes with increasing counterface roughness from predominantly elastic to a mixed elastic/plastic regime. Order-of-magnitude calculations suggest that plastic deformation will become the predominant mode only if the surface roughness is increased to values of the order of at least 1000 μin. CLA. Since it is reasonable, as already mentioned in the paper, to assume that wear is more probable when the deformation is plastic, the observed increase in wear rate with roughness in figure 24 may be interpreted qualitatively as being a consequence of the increasing importance of plastic deformation. Once again, a knowledge of the ratios σ/β for each surface is required to make the analysis quantitative. Perhaps this information is already available for surfaces randomly abraded with different grades of abrasive paper. The methods described by Dr. Williamson for charac-

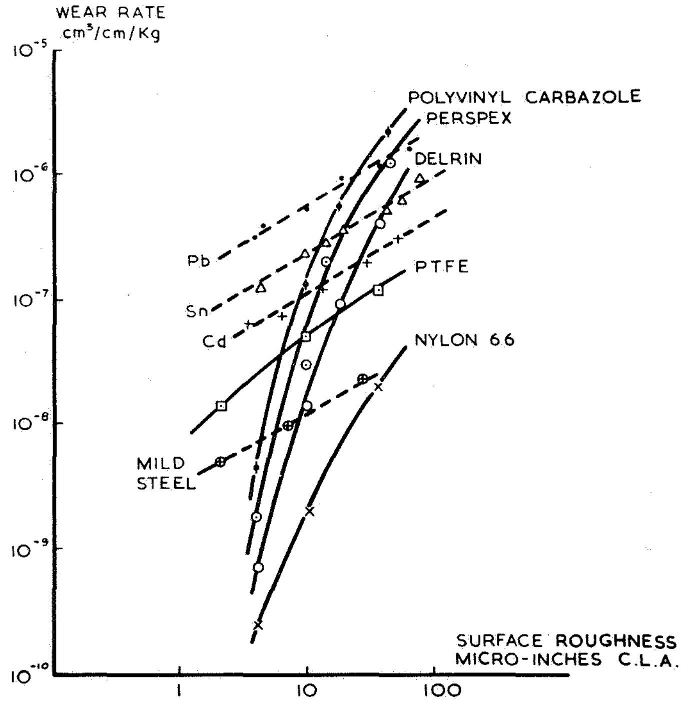


FIGURE 24.—Variation of wear rate with surface roughness during single traversals over steel.

terizing surface topography are clearly going to play an increasingly important role in the future analyses of friction and wear phenomena.

J. J. Bikerman (Horizons Incorporated, Cleveland, Ohio)

Today we heard two papers dealing with solid surfaces, and they had nothing in common. I shall try to establish a link between the two. Professor Gatos spoke of surface energy, and Dr. Williamson, of surface roughness. In my opinion, the existence of surface roughness shows, not the absence of a surface energy of solids, but at least the impossibility to apply rules derived for liquid surfaces to those of solids. Remember the profiles shown by Dr. Williamson. Should we look at a string of this shape, we would immediately conclude that the string is not taut, is not under tension; if tension were present, the string would have been straight.

It is known that the vapor pressure above convex liquid surfaces is higher than above concave ones. This rule was uncritically extended to solids, and it was supposed that the vapor pressure of the hills exceeded that of the valleys. But solids which lay undisturbed for a million years still are rough; a million years was not sufficient for the hills to distil into the neighboring valleys. It is difficult to believe that a process which did not occur in geological time intervals can be responsible for such rapid effects as grain growth or the coarsening of crystals in saturated solutions.

D. G. Flom (General Electric Company, Valley Forge Technology Center, Philadelphia, Pennsylvania)

When two solids are placed in contact, increasing the force between them tends not only to lower the summits of the largest asperities, as might be expected, but more significantly it also tends to raise other portions of the surfaces not yet in contact. (This is shown in figure 13 of Dr. Williamson's paper.) This means that localized asperity deformation involves much subsurface deformation as well.

Subsurface deformation is clearly observed in the deformation of single crystals. Many at this symposium may be familiar with the classic experiments of Gwathmey, Bailey, Dyer, and others showing the production of "square" craters in (001) faces of cubic crystals by deformation with round-tipped indentors. Furthermore, concentric slip patterns form around these deformations owing to interception of (111) glide planes with the surface. Similar considerations apply for other crystal systems (e.g., hexagonal), and the slip patterns can be predicted from a knowledge of the preferred glide planes. I have observed during the indentation of many different crystals that, depending upon the load, these slip patterns can form at distances up to five or more times the radius of the central indentation.

I have found also that with translucent polycrystalline materials, such as pore-free alumina or spinel, the slip systems are not as easily observed but another interesting phenomenon occurs, namely, the production of contrasting "light" (optical) regions within the solids. On sliding an indentor on the crystal, these regions appear at as much as six times the track width away from the track. Presumably they result from intergranular slip within the material causing unique optical effects.

These observations lead me to propose that we should be concerned not only with "areas of contact" but even more significantly with "areas of influence." Admittedly, the latter may be difficult to measure or even to define; but this does not minimize their role in the mechanism of sliding and material deformation.

D. Tabor (University of Cambridge, Cambridge, England)

I should like to raise one point concerning the fascinating results displayed by figures 13 and 14 in Dr. Williamson's paper. It seems to me that one of the factors not considered in the discussion of these results is the constraining pressure exerted by the jacket on the aluminium cylinder. The cylinder is held in a tight fit in the jacket and then subjected to a uniaxial compressive stress of 80 000 psi. Very large hoop stresses must exist in the jacket and when the deforming anvil is removed; these must exert large compressive stresses on the aluminium cylinder. When the surface of the aluminum specimen is later roughened and then flattened by pressing with a hard smooth anvil, these stresses could show themselves in two ways: (1) Plastic flow around individual asperities might release some of the compressive stress imposed by the jacket and so allow the free surface of the specimen to move upwards. This might account for the observed rise of the surface. (2) Because of interaction with the normal stress, gross plastic flow will occur only when the applied anvil pressure approaches the pressure initially applied in forming the specimen, i.e., at a normal stress approaching 80 000 psi.

If this argument is valid, one should obtain a very different behavior if the jacket were split and removed after the aluminum cylinder had been formed so that the constraining pressure no longer exists. If this is not practical, an alternative approach would be to carry out the experiment at different initial compressive stresses. The amount of "rise" of the surface and the bulk plasticity condition should be directly dependent on the magnitude of the initial compressive stress.

H. C. Rogers (General Electric Company, Research and Development Center, Princeton New Jersey)

The early rise in the "floor" of the surface upon "high pressure" contact as measured by Dr. Williamson (fig. 13) is probably the result of the highly stressed asperities acting, in essence, as indenters. It would appear that on initial loading, up to about 500-lb load, a substantial metal flow occurs in the asperities themselves, flattening and broadening them. On further increasing the load, the high surface shear forces restrict continued flow in the body of the broad, flat asperities that actually have, as Dr. Williamson pointed out, a factor of 10 smaller height-to-width ratio than is indicated by the graphs in his figure 12. The asperities with flow restricted are therefore relatively harder than the substrate metal and can act as indenters on both of the contacting surfaces. The deformation zone will be similar to that shown in figure 25 (ref. 60). Plastic deformation takes place within the metal substrate, the metal flowing out to each side of the asperity as the load is increased, filling the gap between contacting asperities. There is also a small amount of flow in the periphery of the asperities because the surface tractions are much reduced there. This further aids in flattening and broadening the asperities.

LECTURER'S CLOSURE

Dr. Kingsbury raises an interesting point when he discusses the problem of developing a description of a surface that takes into account both the short-range order and the long-range order. The "height description" approach does not do this; in the treatment described in the paper, it is supplemented by a "lateral" measurement (e.g., the curvature of the asperity tops). The autocorrelation function is another obvious candidate for the task of providing the lateral

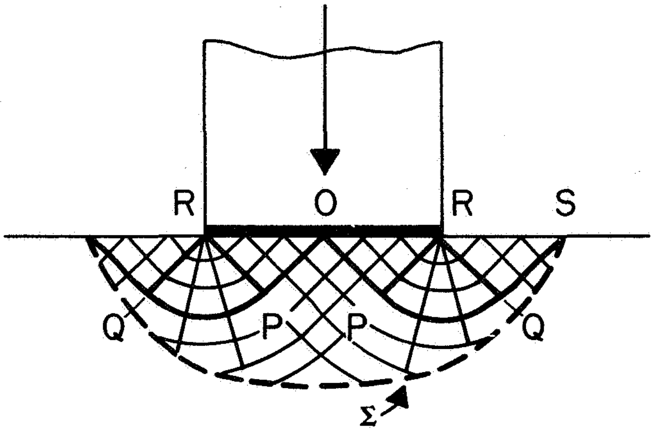


FIGURE 25.—Slipline field and plastic zone in indenting a semi-infinite medium with a flat die (ref. 60).

description. There are, however, two difficulties to be considered. First, although this function reflects the lateral scale of surface features, it does not tell us in which stratum of the surface the features occur. This can be illustrated by a very simple example: the two imaginary profiles have the same autocorrelation function but would obviously possess different contact properties (fig. 26). Second, no one has yet shown how to use the autocorrelation function as the basis of an adequate theory of surface contact. I agree with Dr. Kingsbury that this should be a fruitful area for future work.

It would be most instructive to compare an interferograph of a surface with its micromap. I should be happy to map a surface if anybody will provide one that has been interferographed.

It seems unlikely that the observed asperity deformation can be explained by a delayed elastic-plastic transition. The data in figure 14(b) were obtained at zero load from the profiles of figure 12, and thus all points indicate plastic deformation. The bend in the curve should not be confused with the onset of plasticity in a conventional stress-strain diagram.

Since the average slopes of engineering surfaces are usually only about 10°, it seems unlikely that significant tangential stresses are generated when two rough surfaces touch as Dr. Kingsbury suggests. However, I agree that one must be extremely cautious when applying the static contact theory to friction and wear.

Dr. Melmed begins by discussing my comment on recent surface contact studies using electron microscopy and field-emission microscopy. It is obvious that the paragraph was worded to produce an interpretation on his part which I had in no way intended, and I welcome this opportunity of rectifying that impression. I did not say, and I did not intend to imply, that "advances in observation techniques are not really worth pursuing." I merely wished to sound a note of warning that not all directions of advance may be equally fruitful. To look at things in ever-increasing detail does not necessarily imply that one has become more pertinent to events in the macroscopic world of engineering processes. It seems to me useful to remind ourselves of this

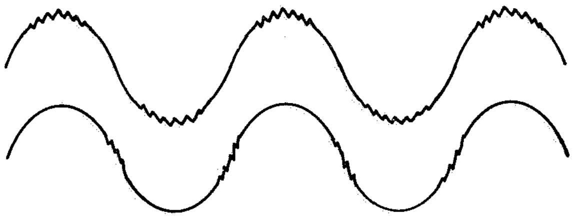


FIGURE 26.—Comparison of two surface profiles.

every now and again; and I feel certain that the scientist whose work I was discussing in the paragraph in question would not object to my view, for he has himself given a very succinct expression of the same opinion:

> The physicist must match the technique he uses to the scale of the problem he is investigating. He must also use all the physical concepts that can be applied to the explanation of his observations. He must know how solids deform elastically, how plastic flow occurs, how contact stresses can produce slip in ductile solids and cracking in brittle solids. He needs to know how surfaces react with environments and how surface films are formed. But he must also be realistic in his approach. If he studies the sliding of single crystals he may be able to explain the gross effects in terms of plasticity theory; if he is lucky he may be able to explain some of the detailed processes in terms of dislocation theory. But if he describes the frictional process in terms of wave mechanics it is probable that though he may possibly understand wave mechanics he does not understand friction.*

As Dr. Melmed says, surface cleanliness is a very large factor affecting experimental results in friction studies; it is also critical in studies of adhesion. However, it has very little influence on the topography of surfaces or on their behavior under static normal contact—to which themes the main part of this paper is limited. I do not believe that the alleged inconsistency exists.

The two laws of friction "f is proportional to W" and "f is not dependent on A" hold independently of any particular mode of deformation of the surfaces. They are the result of surface topography, not of material properties. The property that mild steel and mild cheese have in common, which makes them obey these laws, is not their strength but their topography. The constant of proportionality in the relation between f and W certainly depends on material properties. But the existence of the law does not. I cannot accept Dr. Melmed's contention that my statements are contradictory.

Stylus tracks made on gold are the same width as those on aluminum of the same hardness.

Dr. Melmed asks how the rise in the valley floors can be accounted for by flow occurring inside the sample that is under great compression. I do not see this difficulty at all: the valley floors are the only part of the aluminum cylinder not restrained by the jacket or pushed down by the anvil. Surely this is exactly where we should expect the volume removed from the asperity tops to reappear. The profilometer could not, of course, reproduce an anvil-headed asperity with reentrant side walls; but, on the other hand, it is extremely difficult to see how such a shape could occur when a surface whose average slope is only 10° is compressed under a static normal load. Although at first sight

---

*Written by Dr. Tabor in the Bulletin of the Institute of Physics last September.

the loaded-cylinder experiment appears to study surface deformation under very specialized conditions, the imposed constraint—that there is no macroscopic radial flow—must be very similar to that existing in the middle of a large apparent area of contact, and I believe the results to be of greater generality than Dr. Melmed suggests. But, of course, one must not casually extrapolate the results of static loading experiments into discussions of sliding contact.

It is indeed difficult to make gas-tight seals between two nominally flat metal surfaces. In metal vacuum systems one usually fills the valleys between the contacting surfaces with some soft metal; but if this approach is not permissible, as in certain space applications, it becomes a major problem to ensure a gas-tight seal. If the problem could be dismissed as easily as Dr. Melmed suggests, the Government would not now be spending large sums on contract research to seek a solution.

Bead-blasted surfaces were not chosen for this study because of any particular correlation with the topography of electric contacts as Mr. Tallian has suggested. Our purpose was not to work with "typical" surfaces. (Indeed, I doubt if any can properly be called typical; electric contact surfaces, like those of ball bearings, are special cases.) We chose bead-blasted surfaces because they are intrinsically well suited for basic research on surface contact phenomena. They are simple to create: the short-range texture can be superimposed on any macroscopic shape and, when necessary, even on specimens after they are mounted on jigs. The texture has no orientation and does not reflect bulk features such as grain structure. It is easy to control and reproduce and remains constant over each specimen and between specimens of the same batch. At this stage in our understanding of surface contact, it seemed to us more fruitful to study a simple real surface than to develop the theory of a special geometry, which, however accurately it describes a particular model, can be of little general use.

It is rather misleading to state that the theoretical arguments put forward by Greenwood and me in developing the plasticity index referred to in the paper are based on the assumption of "hemispherical asperities." The theory merely requires that the caps of the asperities (i.e., the surface above the 90th percentile) be approximately spherical. This qualification seems to me to elevate the assumption from the improbable to the highly plausible. However, the important difference between the plasticity criterion discussed in this paper and similar criteria proposed by other workers does not lie in the model of surface geometry chosen (i.e., whether randomly distributed peaks with spherical caps represent a particular surface more realistically than sinusoidal or triangular prisms), but in the fact that the index is computed using only data from the regions where contact is likely to occur. There is clear evidence that rough surfaces touch near the summits of their asperities, and it seems reasonable that it is the geometry there which determines the nature of the contact rather than some parameter averaged over the whole surface and consequently influenced strongly by the topography of parts not involved in the contact.

It is encouraging to hear of the formal similarity found by Dr. Archard and Mr. Whitehouse between the theoretical dependence of the peak height distribution on the surface height distribution and that which Mr. Hunt and I observed experimentally. The numerical ratios quoted depend on the particular assumptions used concerning the nature of the short-term order of the profile, and, as Dr. Archard suggests, the disagreement merely underlines the need for further work. Somehow spatial information such as is contained in the autocorrelation function of a profile must be introduced into the general theory. At the moment, however, it is not at all obvious how this may be done, or even how to extract the pertinent data from the autocorrelation function.

I agree completely with Dr. Lancaster when he comments that one cannot simply extend the use of the plasticity index criterion from static contact conditions to situations involving sliding. The most that can be said at present is that recent experimental techniques have led to some new ideas about static contact and may prove a helpful step in the study of sliding contact. With sliding metals it seems probable that phenomena similar to those described in the paper do occur; but it would certainly be unwise to apply the theory to sliding polymers where other parameters, such as ductility, can play significant and perhaps controlling roles.

Some experiments which Dr. Snowball and I have carried out are relevant to Dr. Bikerman's question concerning the persistence of surface roughness. We took coupons of gold with rough bead-blasted surfaces and measured their topography while they were aged at elevated temperatures. It was necessary to limit the aging temperature to about 600° C to avoid topographic changes due to thermal etching. The surface texture parameters derived from values averaged over the entire surface (such as the CLA) showed no detectible change even after 120 hours at 600° C. However, this does not mean that the surfaces were unchanged; it is another example of how overall surface parameters can give misleading results. Measurements of the asperity radii showed small but reproducible changes. The asperities were blunted; and those with the smallest radii of curvature showed the largest changes, as would be predicted by arguments based on surface energy considerations. These experiments will be reported more fully elsewhere.

It seems likely that the initial rise in the valley floor (shown in fig. 13 of the paper) is caused by a flow pattern of the type suggested by Dr. Rogers. The deformation zone which he discusses might well correspond to the "area of influence" proposed by Dr. Flom. The interesting question is how these individual deformation zones can produce the bends in the three graphs of figure 14. Perhaps the bends occur at the load at which the areas of influence of the individual contacts begin to interact so that the asperities behave collectively.

I agree with Dr. Tabor that the compression stress exerted on the aluminum cylinder by the jacket might have some effect on the surface deformation. Unfortunately, one cannot remove the jacket during the loading as its only purpose is to restrain the aluminum cylinder from flowing under the high loads. But it is straightforward to repeat the experiment using several different pre-loads; we will follow this suggestion and hope to report the result at a later meeting.

REFERENCES

1. Moore, A. J. W.: Taper Sectioning. Metallurgia, vol. 38, June 1948, p. 224.
2. Tolansky, S.: Multiple Beam Interferometry. Oxford Univ. Press (London), 1948.
3. Tolansky, S.: An Introduction to Interferometry. John Wiley & Sons, Inc., 1962.
4. Dyson, J.; and Hirst, W.: The True Contact Area Between Solids. Proc. Phys. Soc. (London), vol. 67B, 1954, p. 309.
5. Nomarski, G.; and Weill, A. R.: Application a la metallographie des methodes interferentielles a deux oncles polarisees. Rev. de Metallurgie, vol. 52, no. 2, 1955, p. 121.
6. Bryan, J. B.; and Mohl, O.: Microinterferometric Scanning for Full Surface Metrology. ASME Paper 64–WA–PROD–15, ASME Winter Annual Meeting, 1964.
7. Abbot, E. J.; and Firestone, F. A.: Specifying Surface Quality. Mech. Eng., vol. 55, 1933, p. 569.
8. Tabor, D.: The Mechanism of Rolling Friction. Phil. Mag., vol. 43, series 7, 1954, p. 1081.
9. MacCurdy, E.: Leonardo da Vinci Notebooks. Jonathan Cape (London), 1938.
10. Amontons, G.: Histoire et Memoires de l'Academis Royale des Sciences. Paris, 1699, pp. 104 and 206; 1700, p. 145; 1703, p. 105; 1704, pp. 96 and 173.
11. Coulomb, C. A.: Memoires de Mathimatique et de Physique de l'Academie Royale des Sciences. Vol. 10 (Paris), 1785, p. 161.
12. Palmer, F.: What about Friction. Am. J. Phys., vol. 17, 1949, p. 181.
13. Holm, R.: Electric Contacts Handbook. Third ed., Springer-Verlag (Berlin), 1958.
14. Kragelskii, I. V.: Friction and Wear. Butterworths Scientific Publ. (Washington, D.C.), 1965
15. Bowden, F. P.; and Tabor, D.: Friction and Lubrication of Solids. Oxford Univ. Press (London), Pt. I, 1954 and Pt. II, 1964.
16. Archard, J. F.: Elastic Deformation and the Laws of Friction. Proc. Roy. Soc. (London), vol. 243A, 1953, p. 190.

17. Zhuravlev, V. A.: On the Physical Basis of the Amontons-Coulomb Law of Friction. J. Tech. Phys. (USSR), vol. 10, 1940, p. 1447.
18. Ling, F. F.: On Asperity Distributions of Metallic Surfaces. J. Appl. Phys., vol. 29, 1958, p. 1168.
19. Greenwood, J. A.; and Williamson, J. B. P.: The Contact of Nominally Flat Surfaces. Proc. 2nd Int. Conf. on Electric Contacts (Graz, Austria), 1964.
20. Greenwood, J. A.; and Williamson, J. B. P.: Contact of Nominally Flat Surfaces. Proc. Roy. Soc. (London), vol. 295A, 1966, p. 300.
21. Greenwood, J. A.: The Area of Contact Between Rough Surfaces and Flats. J. Lub. Tech., ASME, vol. 89F, 1967, p. 81.
22. Greenwood, J. A.; and Tripp, J. H.: The Elastic Contact of Rough Spheres. J. Appl. Mech., Mar. 1967, p. 153.
23. Myers, N. O.: Characterization of Surface Roughness. Wear, vol. 5, 1962, p. 182.
24. Kubo, M.: Measurement and Analysis of Surface Roughness Based on the Information Theory. Acta Imeko, No. 21–JA–174, 1964, p. 115.
25. Silin, R. I.; and Frederick, J. R.: Instrumentation for the Analysis of the Profile of Machined Surfaces. Proc. CIRP General Assembly (Paris), 1966.
26. Bickel, E.: Some Fundamental Problems in the Measurement of Surfaces. Proc. Int. Production Engineering Res. Conf., ASME, 1963, p. 667.
27. Tallian, T. E.; Chiu, Y. P.; Huttenlocher, D. F.; Kamenshine, J. A.; Sibley, L. B.; and Sindlinger, N. E.: Lubricant Films in Rolling Contact of Rough Surfaces. ASLE, Trans., vol. 7, 1964
28. Pesante, M.: Determination of Surface Roughness Typology by Means of Amplitude-Density Curves. CIRP Annalen, vol. 12, 1964, p. 61.
29. Kragelskii, I. V.: and Demkin, N. B.: Contact Areas of Rough Surfaces. Wear, vol. 3, 1960, p. 170.
30. French, H.; and Rohsenow, W. M.: Prediction of Thermal Conductance of Metallic Surfaces in Contact. J. Heat Transfer, ASME, vol. 85C, 1963, p. 115.
31. D'yachenko, P. E.; Tolkacheva, N. N.; Andreev, G. A.; and Karpova, T. M.: The Actual Contact Area Between Touching Surfaces. Consultants Bureau (New York), 1964 (Russian text published in 1963 by Inst. Mech. Eng., Moscow).
32. Moore, A. J. W.: Deformation of Metals in Static and Sliding Contacts. Proc. Roy. Soc. (London), vol. 195A, 1948, p. 231.
33. Milner, D. R.; and Rowe, G. W.: Fundamentals of Solid Phase Welding. Met. Rev., vol. 7, 1952, p. 433.
34. Greenwood, J. A.; and Rowe, G. W.: Deformation of Surface Asperities During Bulk Plastic Flow. J. Appl. Phys., vol. 26, no. 2, 1965, p. 667.
35. Hill, R.: Mathematical Theory of Plasticity. Oxford Univ. Press (London), p. 226.
36. Patent pending.
37. Archard, J. F.; and Hirst, W.: The Wear of Metals Under Unlubricated Conditions. Proc. Roy. Soc. (London), vol. 236A, 1956, p. 397.
38. Cocks, M.: Interaction of Sliding Metal Surfaces. J. Appl. Phys., vol. 33, 1962, p. 2152.
39. Cocks, M.: Role of Displaced Metal in the Sliding of Flat Metal Surfaces. J. Appl. Phys., vol. 35, 1964, p. 1807.
40. Antler, M.: Processes of Metal Transfer and Wear. Wear, vol. 7, 1964, p. 181.
41. Bowden, F. P.; and Hughes, T. P.: Friction of Clean Metals and Influence of Adsorbed Gases. Proc. Roy. Soc. (London), vol. 172A, 1939, p. 263.

42. Tingle, E. D.: The Importance of Surface Oxide Films in the Friction and Lubrication of Metals. I—The Dry Friction of Surfaces Freshly Exposed to Air. II—The Formation of Lubricating Films on Metal Surfaces. Trans. Faraday Soc., vol. 46, 1950, p. 93.
43. Machlin, E. S.; and Yankee, W. R.: Friction of Clean Metals and Oxides with Special Reference to Titanium. J. Appl. Phys., vol. 25, 1954, p. 576.
44. Bodner, S. R.; and Rosen, A.: Discontinuous Yielding of Commercially-Pure Aluminum. J. Mech. Phys. Solids, vol. 15, 1967, p. 63.
45. Young, R. D.: Field Emission Ultramicrometer. Rev. Sci. Instr., vol. 37, 1966, p. 275. See also Machinery, Aug. 1967, p. 102.
46. Smith, K. C. A.; and Oatley, C. W.: The Scanning Electron Microscope and Its Field of Application. Brit. J. Appl. Phys., vol. 6, 1955, p. 391.
47. Tallian, T. E.; Chiu, Y. P.; Huttenlocher, D. F.; Kamenshine, J. A.; Sibley, L. B.; and Sindlinger, N. E.: Lubricant Films in Rolling Contact of Rough Surfaces. ASLE Trans., vol. 7, 1964, p. 109.
48. Halliday, J. S.: Surface Examination by Reflection Electron Microscopy. Proc. IME, vol. 169, 1955.
49. Martin, J. A.; et al.: Second Summary Report on Structural Studies of Bearing Steel Undergoing Cyclic Stressing. U. S. Navy, ONR Contract Nonr 4433(00), SKF Ind. Rept. AL66M031, DDC AD 482 893, 1966.
50. Tallian, T. E.: Rolling Contact Failure Control Through Lubrication. Proc. 3rd Conf. Lubrication and Wear, IME, 1967.
51. Martin, J. A.; Borgese, S. F.; and Eberhardt, A. D.: Third Summary Report on Structural Studies of Bearing Steel Undergoing Cyclic Stressing. U. S. Navy, ONR Contract 4433(00), SKF Ind. Rept. AL67M050, DDC AD 816 316, 1967.
52. Smith, F. W.: Rolling Contact Lubrication—The Application of Elastohydrodynamic Theory. Jour. Basic Eng., ASME, vol. 87, 1965.
53. Martin, J. A.; et al.: First Summary Report on Structural Studies of Bearing Steel Undergoing Cyclic Stressing. U. S. Navy, ONR Contract Nonr 4433(00), SKF Ind. Rept. AL65M030, DDC AD 462 896, 1965.
54. Littmann, W. E.; Widner, R. L.; Wolfe, J. O.; and Stover, J .D.: The Role of Lubrication in Propagation of Contact Fatigue Cracks. ASME Paper 67-LUB-20, 1967.
55. Anon.: Final Report on A Study of Contact Fatigue. Sponsored by the ASME Research Committee on Contact Fatigue of Rolling Elements, Arthur D. Little, Inc., Jan. 1967.
56. Hingley, C. G.; et al.: Final Summary Report on First Project Year of A Study of the Geometry of Elastohydrodynamic Films in Point Contact. U. S. Navy, Contract N0w-66-0284-c, SKF Ind. Rept. AL67L058, 1967.
57. Archard, J. F.; and Kirk, M. T.: Influence of Elastic Modulus of the Lubrication of Point Contacts. Proc. Lubrication and Wear Conf., IME, 1964.
58. Cameron, A.; and Gohar, R.: Theoretical and Experimental Studies of the Oil Film in Lubricated Point Contact. Proc. Roy. Soc. (London), vol. 291A, 1966.
59. Ratner, S. B.; Farberova, I. I.; Radyukevitch, O. V.; and Luré, E. G.: Soviet Plastics, vol. 7, 1964, p. 37.
60. Hill, R.: The Mechanical Theory of Plasticity. Oxford Univ. Press, London, 1950, p. 254.

Surface Interactions in Sliding

F. F. LING
Rensselaer Polytechnic Institute
Troy, New York

Various actions at the surfaces and domain of influence, occasioned by two bodies in sliding contact, are viewed in totality through irreversible thermodynamics. Because of the complexity of these actions, quantitative treatments exist for only one or more of these actions in a single analysis. Topics summarized in this paper are surface temperatures, thermomechanical coupling, surface behavior of composites, model surfaces, and the implication of these on friction and wear.

SURFACE INTERACTIONS IN SLIDING (the title proposed to the author by the Steering Committee) is indeed an imposing subject if one were to expect the totality of what the title embraces. In practice this totality can be studied only in a general way at this juncture in the development of science; of necessity detailed studies are more limited in scope.

For example, various actions at the surfaces and domain of influence, occasioned by two bodies in sliding contact, can be treated in totality through a thermodynamic consideration. That is to say, the chemical, mechanical, and thermal effects, among many others, form the spokes that emanate from the hub which represents irreversible thermodynamics of the system. The wheel is of course the totality in question. Because in practice the individual spokes are sufficiently complicated to handle in a quantitative way, only one or a few of them can be treated at a time. Subsequent sections will cover such segmented topics as surface temperature, thermomechanical coupling, and surface behavior of composites. Final sections will be devoted to model surfaces and implications of the above discussions on friction, etc.

IRREVERSIBLE THERMODYNAMICS

The subject matter is treated elsewhere in a quantitative form (refs. 1 and 2). In view of the interdisciplinary nature of the symposium for which this is written, the treatment here will be qualitative. Some of the ingredients that go into the subsequent thermodynamic consideration are Cauchy's law of motion, the First Law of Thermodynamics, the fundamental postulate of thermodynamics, and the constitutive relationships.

Cauchy's Law of Motion

The balance of momentum leading to a vector relationship between displacements, stresses, and body forces is called the Cauchy law of motion. Displacements measure the change of distances between two neighboring material points in a body as a result of deformation. Stresses are the forces per unit area across a surface (internal or external) of infinitesimal size and a generic orientation. Body forces are those which are proportional to the volume.* These may be gravitational or electromagnetic forces.

First Law of Thermodynamics

The balance of energy, subject to the Cauchy law of motion, leads to a relationship between the product of stresses and strain-rates, heat flux, heat source, and internal energy. This is called the First Law of Thermodynamics.

Fundamental Postulate of Thermodynamics

In irreversible thermodynamics, the basic equations of classical thermostatics are applied to elements of volume in a moving material or in a mixture of materials. Thermodynamics is viewed primarily as a study of entropy. According to Gibbs, the thermodynamic substate is regarded as influencing the internal energy density ϵ. The substate is, in general, the set of parameters $v_\alpha (\alpha = 1, \ldots, n)$. The parameters may be volume, deformation gradient, etc. The substate plus a single further dimensionally independent scalar η suffices to determine ϵ, independently of time, place, motion, and stress. That scalar is called the specific entropy for a given particle. The fundamental postulate is this: ϵ depends on η and the v_α's. $\partial\epsilon/\partial\eta$ gives the temperature T and $\partial\epsilon/\partial v_\alpha$ defines the thermodynamic tension τ_α. For example, when v_1 is the specific volume, $-\tau_1$ is the thermodynamic pressure. In the case of a homogeneous mixture, when the substate includes both the total volume and the masses of the constituents, then the tension τ_α corresponding to the masses of the constituents is called the chemical potential. Where v_α is the deformation gradient, τ_α is the stress vector.

Depending on the generality required, a relationship now connects

---

*Here classical continuum mechanics is used, i.e., the field theory that assumes that the material body is indefinitely divisible while retaining its defining properties. In so choosing the phenomenological method for describing materials, its preference over electron, molecular, or atomic theories is not inferred. Much of the usefulness of many of the theoretical expressions has been established through decades of experience accrued through experimental mechanics.

the internal energy density, temperature, entropy, etc. The equating of this relationship to that derivable under the First Law leads to the Second Law. Note that in this approach the Second Law is a consequence and not a fundamental law. In other terms, in the final analysis, a fundamental postulate is made such as the one on entropy.

Constitutive Relationships

Within materials constants relationship between stresses, strains, temperature, chemical effects, etc. can be derived from the above relationship. Again, the degree of generality depends on the extent desired. This is the first of the constitutive relationships of a given material. Then for each of the relevant variables (e.g., temperature) there is a constitutive equation. In the case of temperature, there is the generalized heat equation. In the same fashion there should be one governing the chemical potential and current potential in the event chemical and electric effects, respectively, are involved.

Summarizing, the above may best be depicted diagrammatically in figure 1. The wheel consists of two parts: (1) field equations such as the Cauchy law of motion, electromagnetic field equations; and (2) constitutive equations such as the thermoelastic stress-strain law, generalized heat equation. These are the ultimate equations to be solved.

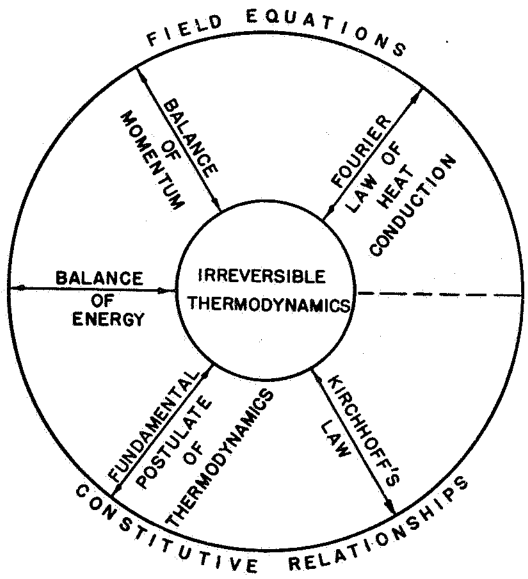


FIGURE 1.—Diagram showing the unifying feature of irreversible thermodynamics.

These equations come forth through the various spokes representing such items as: (1) balance of energy, (2) balance of momentum, (3) fundamental postulate of thermodynamics, (4) Fourier law of heat conduction, and (5) Kirchhoff's law. They merge at the hub through irreversible thermodynamics, which serves, in a very general way, as the unifying and conceptual tie for materials of various properties under the influence of various fields. Theoretically, there is no reason why chemical effects cannot be treated in conjunction with mechanical and thermal effects. In practice, however, two interacting effects are already rather complex to treat quantitatively. Suffice it to say that the above should serve to indicate the path for further challenge in research in a quantitative way; this said, individual facets of surface interaction in sliding will follow.

SURFACE TEMPERATURES

Dynamic Thermocouple

Some theoretical results will be summarized here; the use of dynamic thermocouples will be mentioned briefly later. Figure 2 shows schematically two nonsmooth bodies in contact and the thermoelectricity generation. The Peltier coefficient is π_{AB}, T_S is the interface temperature averaged over the spot of contact, T_0 is the cold junction temperature, and P is the potentiometer reading. It has been shown in an analysis by Gaylord et al. (ref. 3) that if the spots of contact are round and sparsely located and if the leads are placed far away from the contact region, then

$$P=\sum^{n}E_n\sqrt{A_n}/\sum^{n}\sqrt{A_n} \tag{1}$$

where E_n is the emf generated at the nth spot of the contact region,

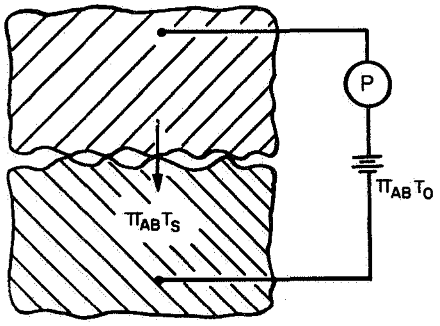


FIGURE 2.—Schematic diagram showing two bodies in contact and thermoelectricity generation.

A_n is the contact area of the nth spot. Equation 1 shows that the potentiometer reading measures the average emf of the spots weighted by the square root of the area of the spots.

Later it was shown by Hughes and Gaylord (ref. 4) that for a continuous contact area, the potentiometer reading gives the average emf weighted by the area divided by the distance from the perimeter of the circle. For a rectangular contact area in sliding contact, the interface temperature based on continuity at the interface has been computed numerically by Shu et al. (ref. 5).

The above shows why experimental results for sparsely distributed contacts do not check with analytical results that are computed on the basis of straight average. Also, for continuous contact, experimental results do not always agree with analytical results computed on the basis of area average.

Juncture Condition at Moving Interfaces

Calculated temperatures are usually obtained by using continuum theory and assuming smooth surfaces. In these calculations, it has been tacitly assumed also that the temperatures due to frictional heating on both surfaces in contact are the same. This condition, which is certainly valid for smooth surface without contaminating films, was first used by Blok (ref. 6). Using known solutions for the given geometries, the surface temperature on both bodies shown in figure 3 within the squares can be found (ref. 7). Uniform and constant heat flux is used: σq for the stationary body 1 and $(1\text{-}\sigma)q$ for the moving body 2, where σ is the fraction of heat flux q which goes into body 1. Examination of the surface temperatures at $y=\eta=0$ (as shown in figure 4), $T_1(x, 0, 0)$ for body 1, and $T_2(\xi, 0, 0)$ for body 2, shows that generally $T_1(x, 0, 0) \neq T_2(\xi, 0, 0)$. However, $(T_1)_{\max}$ and $(T_2)_{\max}$ can be set equal to each other through the parameter σ. Blok argued that the probable temperature is somewhat in between. It is easily shown that

$$\sigma \doteq \frac{1}{1+\frac{K_2}{K_1}\sqrt{2R}} \text{ for } R>5 \tag{2}$$

where $R \equiv Va/4\kappa_2$, V is the speed of movement of body 2, a is the semi-length of the contact area, κ_2 is the thermal diffusivity of body 2, K_1 and K_2 are the thermal conductivities of bodies 1 and 2, respectively. In addition,

$$\sigma = \frac{1}{1+\frac{K_2}{K_1}} \text{ for } R=0 \tag{3}$$

while, in between, the expression would be more complicated. In fact, for the whole range of R (ref. 7),

$$\sigma = \frac{1}{1 + \dfrac{K_2}{K_1} \dfrac{2\pi}{[1 + .414(1 - e^{-1.3R})]I(R)}} \tag{4}$$

where $I(R)$ is shown in figure 5. Strictly speaking, the matching of $T_1(x, y, 0)$ and $T_2(\xi, \eta, 0)$ leads to an integral equation, the inversion of which should give σ as a function of (x, y). This has been done in one instance by Ling (ref. 8), and it was found that the picture shown in figure 4 is not too far off.

So much for the condition of perfect matching of temperature. What of the actual situation? For static contacts, experiments by Konwenhoven and Potter (ref. 9) and Fenech and Rosenow (ref. 10) have shown that there is a temperature break at the interface, i.e., temperature at the surfaces considering the surfaces smooth. In other

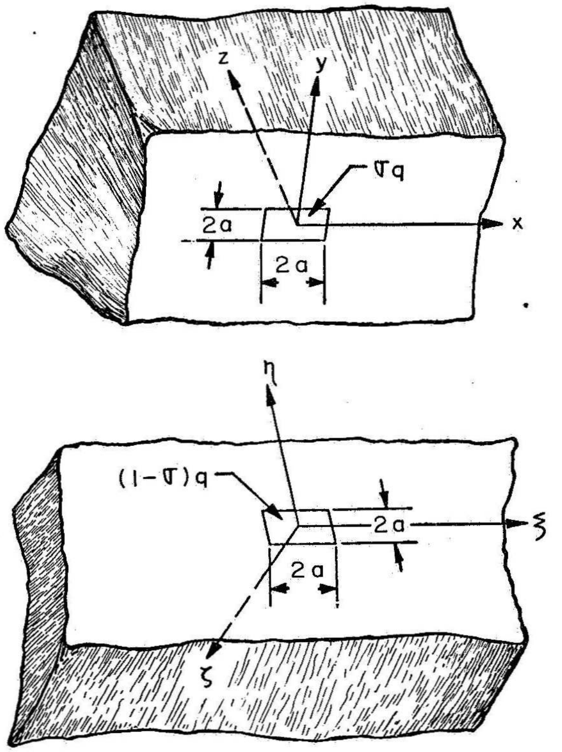


FIGURE 3.—Components of two solids in sliding contact.

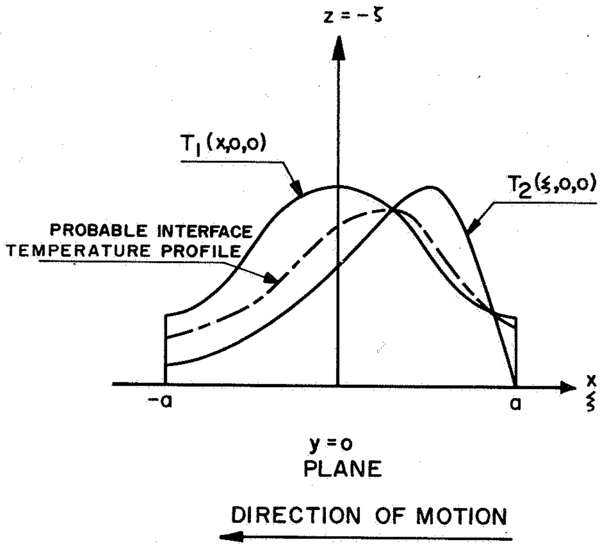


FIGURE 4.—Matching of the surface temperature maxima of two solids.

words, if the mating surfaces are considered smooth, then the temperatures extrapolated by means of continuum theory to the surfaces are not necessarily the same. The basis of extrapolation is measured temperature away from the surface in one instance. The solid curves in figure 6 show the data of h_c, thermal conductance, vs average pressure at the contact region for various values of surface roughness between 3 to 3320 μin. for static contact. The definition of h_c is such that the heat flux is equal to the product of h_c and the temperature jump of the surfaces in contact.

For moving contacts, the measurement of h_c was first carried out by Ling and Simkins (ref. 11) as shown in the schematic diagram, figure 7. The essential components of the apparatus are the friction wheel, *A*, and the friction arm assembly, *B*. Affixed to the drive shaft, *C*, between layers of insulation, is a thin ring plate, *A*, whose exterior surface is in sliding contact with the stationary rider, *D*. The ring plate is the slider and is insulated from the shaft. The rider is insulated on all of its flat surfaces. The concave surface makes at least a line contact with the rider, and the convex surface is cooled by water, thus enabling control of the cooling rate. Thermocouples formed by *D* and pins pressed against it record temperatures at many points. Using the analytical solutions obtained by Ling and Ng (ref. 12) of heat conduction problems associated with the geometries involved, surface temperatures are calculated. It should be emphasized surface temperature or macroscopic surface temperature means the extrapolated temperature to the surface using continuum theory from data obtained away from the surface. Figure 8 shows a typical set of data; i.e., macroscopic surface temperatures vs angular position $-\pi/18 \leq \theta \leq \pi/18$, $\pi/9$ being the extent of the ring sector. Solid curves show the temperature on the rider and dot-and-dash curves show those on the slider. Many of these data were used to obtain average surface temperature and to compute the heat transfer coefficient across the sliding interface. Let Q be the heat flux generated at the interface. Let Q_1 and Q_2 be the heat flux through the periphery of the slider and the back of the rider, respectively. Of course, $Q=Q_1+Q_2$. If ΔT is the area-average macroscopic temperature jump at the interface and h_c is the average heat-transfer coefficient, then the heat transmitted across the interface, Q_h, by conduction, convection, and first-order radiation is $h_c\Delta T$. Since Q_h is not exactly known, bounds on h_c can be calculated from experimental data: $\bar{h}_c$ and $\underline{h}_c$ for the upper and lower bounds, respectively. Where $\bar{Q}_h$ and $\underline{Q}_h$ are for the upper and lower bounds, respectively, $h_c=\bar{Q}_h/\Delta T$ and $\underline{h}_c=\underline{Q}_h/\Delta T$. If Q_1 were entirely generated on the surface or within the rider, then $\bar{Q}_h=Q_1$ is certainly true. For the same materials in sliding contact, the rider wears more than the slider based on the amount of exposure to sliding contact. Using an equal probability argument, let half of the Q be generated on each body. A lower bound is then $\underline{Q}_h=Q_1-Q/2$. In this way, $\bar{h}_c$ and $\underline{h}_c$ for

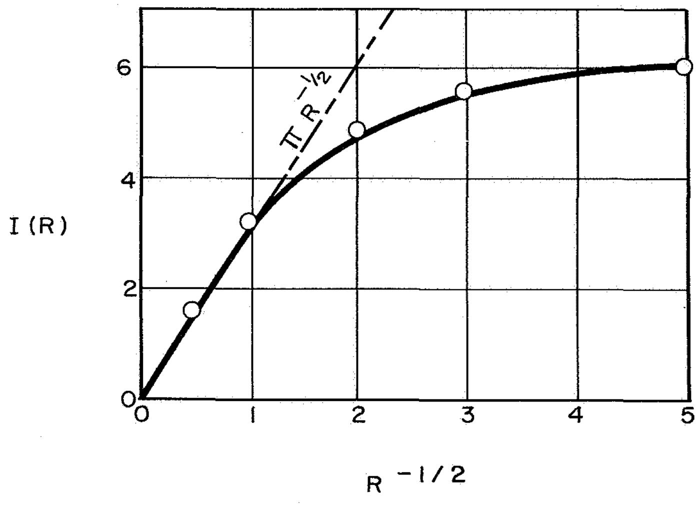


FIGURE 5.—Plot of $(I)R$ vs $R^{-1/2}$.

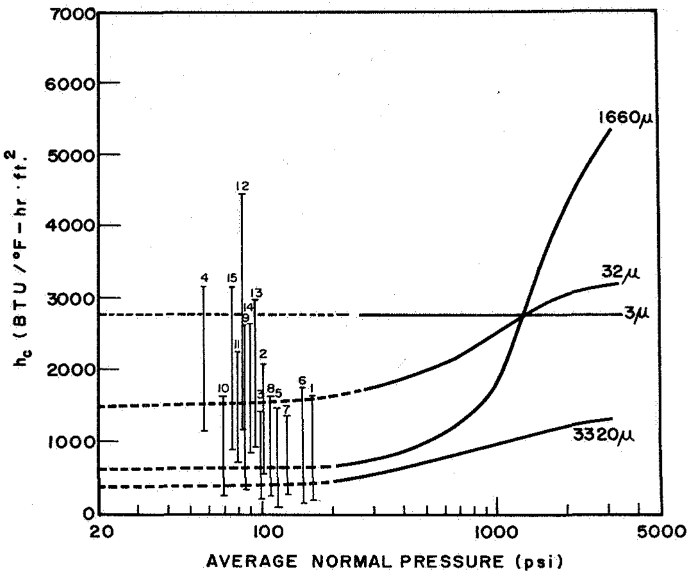


FIGURE 6.—Dynamic $\bar{h}_c$ and $\underline{h}_c$ vs average normal pressure.

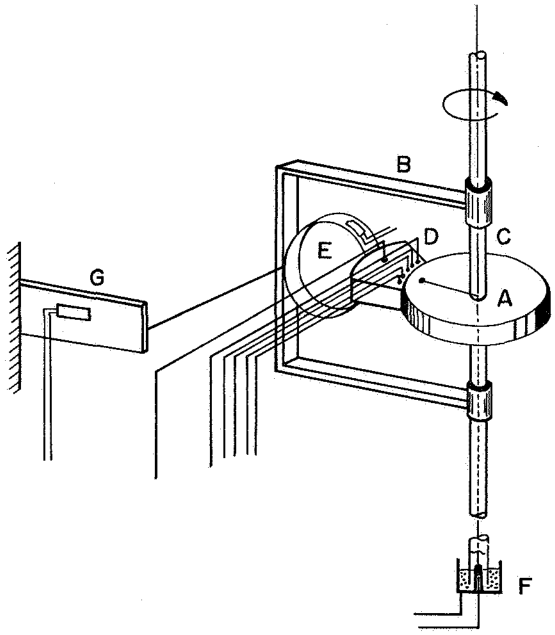


FIGURE 7.—Schematic diagram of apparatus for surface temperature measurement.

15 sets of data are shown in figure 6 together with data for static situation. The horizontal bars of the elongated *I*'s represent the upper and lower bounds on h_c.

Returning to the question of heat generated on the surface or within the body, let λq denote that associated with the stationary body. In other words, λ refers to heat generation (e.g., through deformation and asperity breakage). Then $(1\text{-}\lambda)q$ is that associated with the moving body; q is as defined earlier in connection with the heat removal fraction σ. The following observations have been made by Ling and Pu (ref. 13) and Ling (ref. 14):

(1) In general, frictional heat is not generated in the space between two bodies in sliding contact. Most of the heat is generated on the surface of both bodies because friction derives from either the breaking of adhered junctions, or from the thermodynamically irreversible process of plastic deformation of asperities and the bulk body. That is, $\lambda \neq \sigma$ in general.

(2) Whenever the capacity of one of the bodies to remove heat

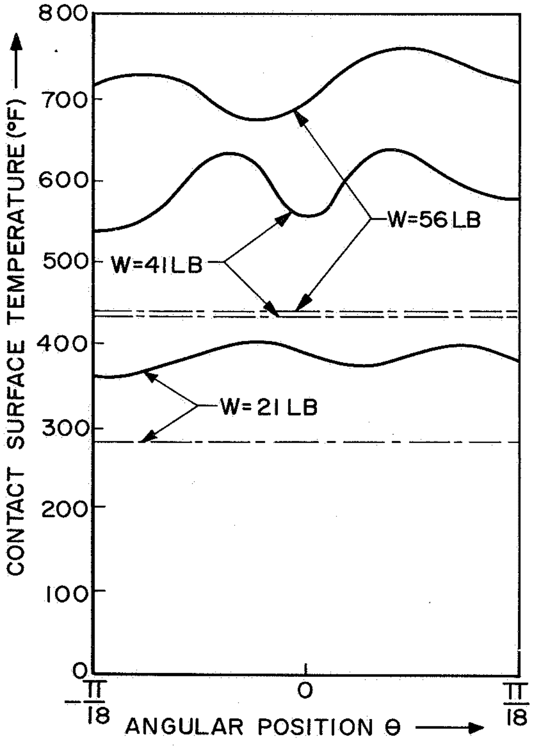


FIGURE 8.—Typical data obtained from the apparatus shown in figure 7.

away from the interfacial zone is less than the amount of heat generated on that surface, there will be a temperature jump across the interface. The surface will then have a higher temperature.

(3) Observations (1) and (2) may be used as quantitative tools for selecting sliding material pairs.

Interface Temperature Measurement

Some methods of measuring interface temperatures include the following:

1. Indirect method for macroscopic temperature with the aid of analytical solution of the heat equation (as above).
2. Dynamic thermocouple method. For very high loads it has found success in metal-cutting investigations as shown by Shore (ref. 15), Gottwein (ref. 16), and Herbert (ref. 17). In these applications, theoretical calculation agrees with the experiment. This method has also been used for gear temperature studies as shown by Nieman and Lechner (ref. 18). For friction studies the checks between theoretical calculations and experiments have not always been good, e.g., the case of ball on cylinder by Furey (ref. 19). Choice of material pair is not altogether free.
3. Change of color method as shown by Schwerd (ref. 20), which requires transparent mating surface.
4. Infrared radiation method, which requires transparent mating surface, as shown by Bowden and Thomas (ref. 21).
5. Imbedding thermocouple beneath the surface. This method is widely used although not very accurate.
6. Kripton gas detection method. Surfaces have to be separated for detection.

Surface Temperature Transients

It is clear from the above discussion that, experimentally, macroscopic surface temperature may be measured indirectly. The usefulness of this measure of surface temperature may be limited or unlimited, depending on the purposes at hand. From fundamental as well as utility points of view, however, it would be desirable to have some measure of the actual surface temperature at points of contact as sliding time goes on. Qualitatively, several of the methods listed may be useful in this regard; but, for metal systems, there does not seem to be any direct method for measuring these transients. All the available methods are indirect. In all cases, the manner of indirect measurement depends on the model of surfaces in contact.

Models creep in as soon as surface details are being discussed (e.g., the dynamic thermocouple). Recently, much attention has been given to the profilometric description of surfaces. The main strength of profilometric representation lies in the reproduction, to the degree of accuracy of the instrumentation, of the actual surfaces before or after sliding engagement. Its drawback is at least twofold: First, it does not give the situation while sliding is taking place. Second, a reproduction of actual events, even with true integrity, is not exactly useful if the events are complicated. If statistical treatment of these events becomes necessary, a typical basic event is all-important. This leads to a more basic modeling of surfaces as far as temperatures are concerned.

Modeling of surfaces for deformation studies, among others, will be treated in detail later. A model for interface surface temperatures, however, will be given here. For this purpose, it is tacitly assumed that insofar as heat transfer is concerned surfaces of the material bodies are smooth, but contacts are specified only where and when physical contacts are supposed to be taking place.

At this point it may be appropriate to digress for a moment to see what sort of error this assumption may cause. Assume that surface asperities are truncated right circular cones of various semiopening angles θ_0. Then according to Pu (ref. 22)

$$\overline{T}=\overline{T}_0(1+\beta/2) \tag{5}$$

where $\overline{T}$ is the average surface temperature on the truncated face for a uniform heat input, $\beta\equiv\pi/2-\theta_0$ is measured in radians, and $\overline{T}_0$ is $\overline{T}$ for $\beta=0$ (i.e., a flat surface). This formula is found empirically from calculated solution; it is good for small β, a condition satisfied by most surfaces. For $\theta_0=80°$, for example, equation 5 indicates that $\overline{T}_0$ is 8 percent higher than $\overline{T}$.

Because in time the population of numerous contact points may move in position and change in temperature magnitude and consequently be very cumbersome to monitor, its statistics may be very useful. To this end, a simple stochastic model has been constructed by Ling (ref. 23) as follows: (1) Given a geometric contact area (figure 9 shows one area of a square ($l\times l$), although it need not be a square); (2) given a load W, based on existing friction theory, the fraction of the geometric area in actual contact does not change with time; (3) subdivide the geometric area into equal unit areas, each of which represents the smallest possible areas; (4) for a given time interval, suitable for the speeds of sliding and computational requirements, the number of units in contact, n, is fixed as discussed above; (5) by random processes, n units are populated; (6) coalescence of units makes larger area of contacts than the basic unit; and (7) the same process is carried out for each succeeding time interval. Given such a model, heat conduction analyses using basic solutions lead to the following observations: (A typical histogram of surface temperature transients is shown in fig. 10.)

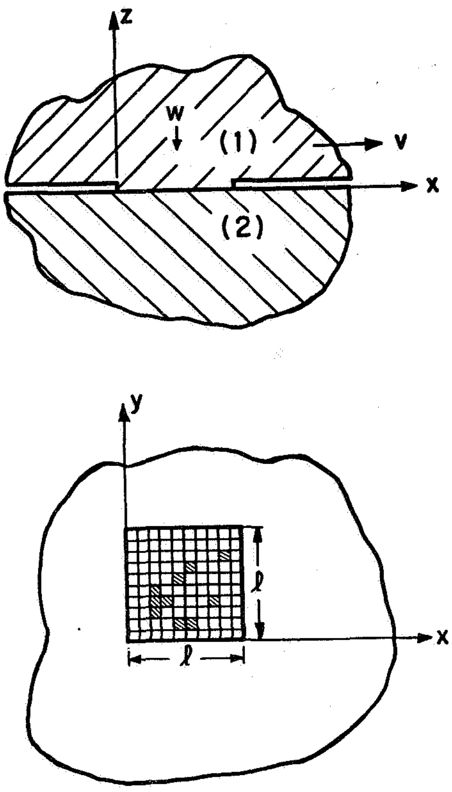


FIGURE 9.—Schematic diagram of sliding contact model.

(1) The average surface temperature of the stationary body differs in general from that of the moving one.

(2) The average surface temperatures of the stationary and the moving bodies, as well as the actual contact temperatures, are low relative to the maximum.

(3) The ratios of standard deviations of surface temperatures of the stationary and the moving bodies to the respective average surface temperatures are between 0.3 and 3.7 with the added observation that (a) the smaller the speeds, the smaller the ratios, and (b) the larger the contact areas, the smaller the range level for the ratios and the lower the levels at which the ratios occur.

(4) The ratios of standard deviations of surface temperatures over contact areas to the average of contact surface temperatures are between 0 and 0.37.

(5) The spectral information on temperatures is insensitive to the particular program of heat-input distribution for a given percentage of contact, a fact which gives credence to the simplified stochastic process.

Interface temperature is important to processes that go on at the interface. One of the main significances of the above is shown in figure 10. While the average temperature may be $T=0.07$, there is a significant peak average at $T=0.37$. It is suggestive that the former temperature may affect the general metallurgical state, while the latter may affect chemical actions.

THERMOMECHANICAL COUPLING

The subject of thermomechanical coupling may be put into two categories. *Interaction* includes the classical thermoelasticity, for example. Referring to the wheel in figure 1, the law of motion is to be solved with stress-strain law incorporated so as to have thermal effects on the one hand and the heat equations that govern heat flow

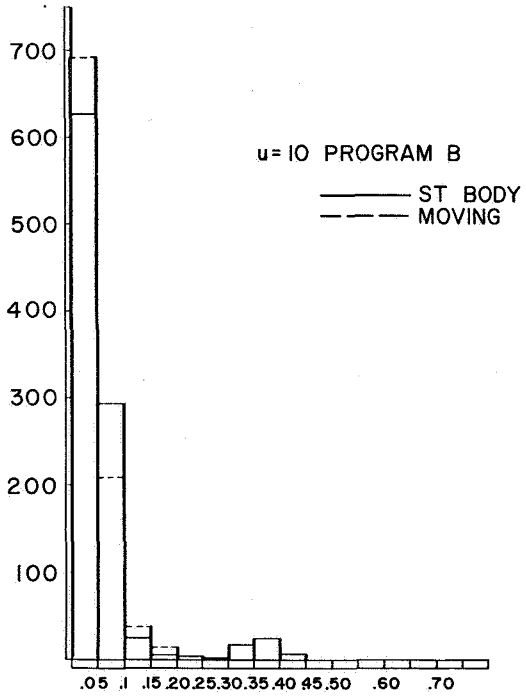


FIGURE 10.—Typical histogram of surface temperature transients.

on the other. *Actual coupling* involves the law of motion with stress-strain law incorporated as before and the heat equation with heat generation term due to mechanical work. In other terms, the relevant equations describing this situation have to be solved simultaneously.

Many investigators have worked in this field and their contributions have been summarized in several recent books (refs. 24 to 27). In spite of these contributions, however, few have dealt with surfaces in motion. Several examples will be cited to indicate the scope of possible coverage of the subject. At the same time, the effects of elasticity, viscoelasticity, and plasticity will be discussed.

Thermoelastic Interaction

Figure 11 shows a large body moving under a distributed heat source which is arbitrary. The surface is allowed to lose heat by convection. This interaction problem (i.e., elasticity problem with allowance for thermal expansion and the heat equation for a moving source) has been solved by Ling and Mow (ref. 28).

The solution has been applied to an elastohydrodynamics problem for which a typical set of heat input data $q(x)$ is given in table 1.

Typically for steel the result on the displacement perpendicular to the surface, U_2, is shown in figure 12. This thermoelastic effect was applied in a problem of elastohydrodynamics. This type of action may well take place on the micro scale in mixed boundary friction regime.

A problem for constant temperature boundary condition over a finite length and impervious surface outside the heat source has been solved by Johanshahi (ref. 29).

Thermoviscoelastic Interaction

The thermomechanical coupling for viscoelastic materials can also be put into two categories, the actual coupling one being more complicated than the interaction category. The theory has been summarized recently by Sternberg (ref. 30). Isothermal viscoelasticity involving surfaces has been examined by Atack and Tabor (ref. 31), Hunter (ref. 32), Flom (ref. 33), May et al. (ref. 34), and Morland (ref. 35).

A problem of the type shown in figure 11 for viscoelastic materials, though amenable to solution, has yet to be solved. However, a simple problem has been worked out and the solution will serve to show an aspect of thermomechanical interaction for viscoelastic materials. The model to be examined is a large plate that is kept at an initial temperature T_0 and insulated on the bottom. It is constrained to expand in the direction perpendicular to the flat surfaces. Now suppose the surface is brought up to a temperature T_1 with $T_1 > T_0$. For this model the amount of expansion is trivial but the normal stresses σ developed along directions in the plane of the plate are of interest.

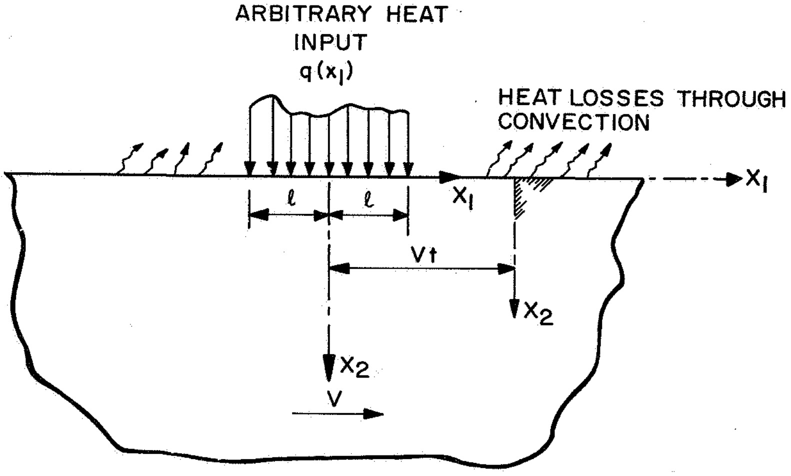


FIGURE 11.—Schematic diagram of the mechanical model.

Specifically σ as a function of time is of interest since the short-time results represent those encountered in a moving plate under a fixed heat load at high speed, while the long-time results are at low speed. In other words, this simple example may give some insight into the problem of the moving plate. Also, while it has been worked out, the solution for the more general viscoelastic stress-strain laws will not be shown. Rather, those corresponding to the simpler but well-known

TABLE 1.—q(x) *Data Encountered in Elastohydrodynamics*

x, in.	q(x), (Btu/in.^{2} hr)$\times 10^{-4}$
−0. 0125	0. 2015
−0. 0875	0. 3129
−0. 005	2. 378
−0. 0025	6. 070
0	12. 53
0. 0025	10. 97
0. 005	16. 20
0. 0075	5. 513
0. 008125	2. 389
0. 00875	1. 243
0. 009375	0. 6064
0. 01	0. 2714
0. 01125	0. 0565
0. 0125	0. 1069

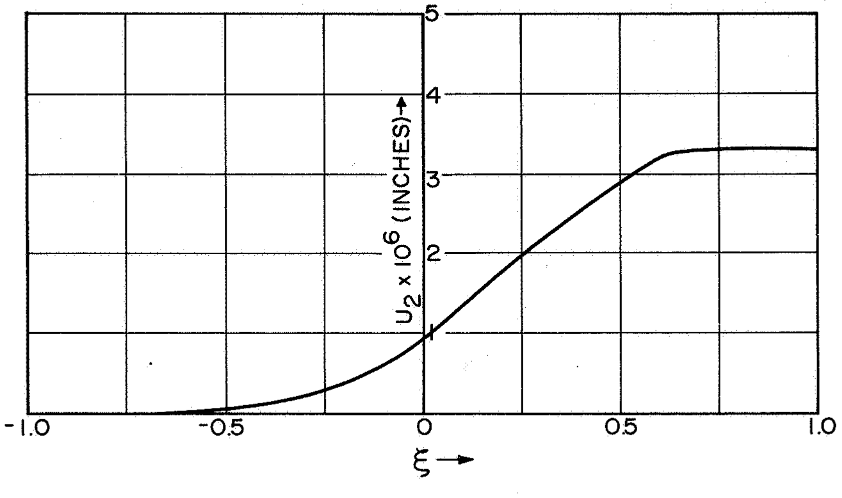


FIGURE 12.—Normal displacement of U_2 (ξ, 0) vs ξ for $-1<\xi>1$ and data given in table 1.

laws (i.e., Maxwell and Kelvin), will be considered. Of course, the standard linear law is a conglomeration of the two, figure 13 shows a plot of a dimensionless normal stress, s, vs a dimensionless time, τ. For the purpose at hand, the exact definitions of those normalizing quantities for arriving at s and τ are not important and therefore not shown. This is done for the idealized Maxwell and Kelvin materials, respectively. For reference, the result for purely elastic case is also shown. The major parameter is the Poisson ratio ν. This example, contained in the paper by Muki and Sternberg (ref. 36), is given to show the effect of temperature on stress for viscoelastic material which often forms friction material.

Thermoplastic Interaction

As with the case of thermoviscoelasticity, the thermoplasticity counterpart of the problem shown in figure 11 has not been solved. Again, a simpler plate problem, which has been solved, might serve to show the character of thermoplastic behavior.

The model is that of a large plate initially at a constant temperature. It is exposed to a uniform heat input that varies with time, while the rest of the faces are insulated. The heat input increases from zero to some maximum over a period of time and then decreases to zero. Therefore, this simple model simulates what a moving plate might see under a heat load due to sliding, for example. As in the last example, the surface expansion for this simple model is trivial, but the extent of the growth of plastic zone is of interest. This is so because the extent of plastic zone is related to the penetration, and therefore deformation friction, when this has to sustain mechanical load as well as heat load. For the analysis, the plastic material is assumed to be elastic and perfectly plastic and satisfies a temperature-independent Mises yield condition. The stress-strain relationship contains thermal stress effect. Figure 14 shows the result as expressed by two curves: $\bar{\xi}$ vs $\bar{q}$ and $\bar{\zeta}$ vs. $\bar{q}$. The dimensionless measure of the maximum heat input rate is $\bar{q}$, and $\bar{\xi}$ is a dimensionless measure of the distance from the center of the plate to the edge of the plastic zone which started at the heated surface (only for $\bar{q}>1.3$ in the case shown). The dimensionless measure of the extent of the plastic zone which started from the center of the plate (only for $\bar{q}>3.3$ in the case shown) is $\bar{\xi}$. This example is contained in the paper by Weiner (ref. 37).

SURFACE BEHAVIOR OF COMPOSITES

Composites embrace a very broad category of material which is formed by more than one material, each component having distinct properties but collectively having still a different set of properties. Because it is not the purpose here to discuss all composites, only three examples will be given.

Layered Elastic Discs

Of concern is an elastic disc with a layer of elastic material on the surface having properties distinct from those of the disc. The surface is loaded with an arbitrarily distributed load over a finite arc and the equilibrating force is concentrated at the center. This center condition is the only model treated that simulates many machine elements. Such an analysis is shown by Beck and Ling (ref. 38). Many analyses on layered elastic systems, mostly having to do with soil systems, have been made in the past decade (refs. 39 to 49).

The main results of this work, without showing tables of numbers, are that in using composites one may achieve a great variety of system behavior both in terms of the surface displacements that are normal to the surface and state of stress within the body and near the surface. The surface displacements influence the deformation friction and the stresses influence the failure of surfaces.

Layered Elastic System Under a Moving Load

This model differs from the previous model in that there is a layer of elastic material which is adhered on a large body. The composite has a flat surface and is moving with a constant velocity V. For low values of V, this model is also akin to those worked out for soil me-

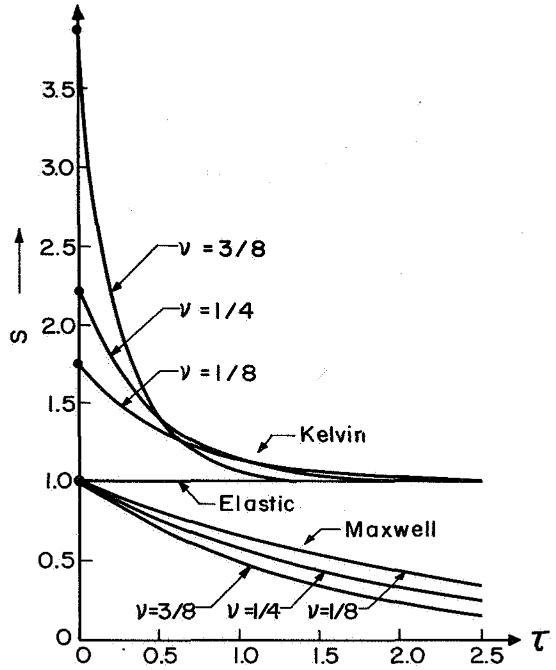


Figure 13.—Time dependence of s at surface for various Poisson's ratios.

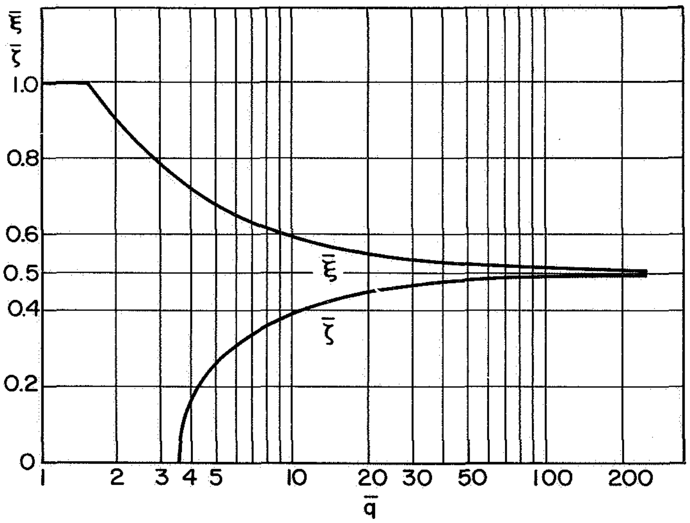


Figure 14.—Growth of plastic zones in the slowly heated free plate.

chanics (e.g., ref. 47). In this problem, the effect of inertia on the shear stress at the interface and normal displacements are sought. Only the quasi-stationary solution (i.e., standing wave problem) is examined, however. This analysis has been carried out by Hsu and Ling (ref. 50) and an experimental counterpart is in progress. Other efforts for fast moving loads have been examined by Sneddon (ref. 51), Cole and Huth (ref. 52), and Sackman (ref. 53).

Two findings of reference 50 were that (1) there is large effect of speed on the surface and surface layer behavior and (2) when the speed exceeds the sonic speed of the layer material, the shear stress at the layer-substrate interface under load undergoes many stress reversals. This indicates that the probability of fatigue failure of the layer is greatly increased. The kinds of failure may be those which Lancaster had discussed (ref. 54).

Layered Elastic System Exhibiting Couple-stresses

For polar materials that exhibit couple-stresses, the stress tensor is asymmetric in general. An analysis similar to that of the elastic disc now shows the stress discontinuity at the interface (fig. 15). This work by Maye and Ling (ref. 55) shows the importance of couple-stress in bonding strength considerations.

Viscoelastic Layer on an Elastic Half-space

The mechanical model that involves a viscoelastic layer on a half-space was studied by Batra and Ling (ref. 56). In particular, surface displacements were computed; this in turn reflects upon deformation friction. This kind of analysis is very complex, but a simplified analysis was made by Abrahamson and Goodier (ref. 57) which preserved qualitatively the behavior as predicted by the above.

MODEL SURFACES

By the very definition of model surface, profilometric studies of surface are excluded from this discussion. Work of this class has been carried out by Greenwood and Williamson (ref. 58), Greenwood (ref. 59), Abbott and Firestone (ref. 60), and Rathbun (ref. 61).

In some work, the profilometric approach is necessary, as indicated by Rathbun (ref. 61) and Schram (ref. 62). In other contexts, model surfaces have been extremely successful.

Undoubtedly, models will enter any discussions on friction in this symposium. Suffice it to say that there have been many contributions in the category—Bowden and Tabor (ref. 63), Dyachenko et al. (ref. 64), Archard (refs. 65 and 66), Ling (refs. 67 to 70), Ling and Lucek (ref. 71), Peterson and Ling (ref. 72), and Green (ref. 73). To these

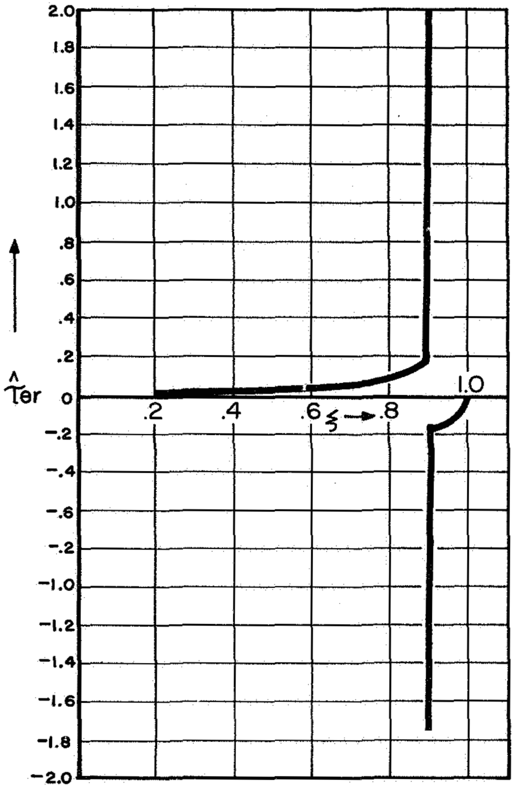


FIGURE 15.—Radial variation of $\tau_{\theta r}$ for θ=0.0500 radians (2.86°).

should be added the models treated in the last section (e.g., layered disc) and the very classic one of Hertz (ref. 74). In the latter case, the model has been used as an ingredient of later models. Where this is inadequate, new ones have evolved like that of Mow et al. (ref. 75).

IMPLICATIONS

This paper has dealt with the use of continuum mechanics. As qualified at the beginning, it is not the intention to preclude molecular, atomic, or electron theories, nor is it the purpose here to show preference. The purpose was, however, to show that continuum mechanics can be useful in a quantitative way because it is a phenomenological approach in the first place.

Having arrived at this stage, the reader may feel that the subject of surface interaction has yet to be touched upon. This is indeed true. The reason is that when interaction is discussed, one would be discussing friction and lubrication, which is not the aim of this paper. Instead, the discussions so far have been viewed as ingredients which have been or can be used.

For example, the discussion on deformation led to quantitative data on surface displacement where the particular model is appropriate; these displacements can lead to significant quantitative information on friction. In the case of the viscoelastic layer material under moving load, the coefficient of deformation friction is proportional to the load and thickness of the film (ref. 56). In the case of the layered elastic disc (ref. 38), calculations have shown friction coefficient comparable to measurements of Tabor (ref. 63). The thermoelastic analysis has been applied to the case of elastohydrodynamics, a case of interacting surfaces.

DISCUSSIONS

W. J. Anderson (NASA Lewis Research Center, Cleveland, Ohio)

Dr. Ling has given us some insight into the current understanding of phenomenological occurrences when surfaces are brought into contact. One's first impression is that the situation is rather discouraging; the war has not been won, nor is victory in sight. As a matter of fact, the victories so far have been in back alley skirmishes, so to speak, rather than in battles of any significance.

The complexity of the subject, by the most conservative estimate, is vast. We are on the brink of an age in which man will travel beyond the boundaries of this planet, but we really do not know with any certainty what constitutes a surface. The author very ably discusses the various facets of the continuum mechanics approach to surface interactions. Since an engineer ordinarily deals with matter whose dimensions are considerably greater than atomic, the use of a continuum mechanics approach certainly appears logical. In a broad sense this represents a statistical averaging of the trillions of individual molecular scale events that occur when two objects slide or roll against each other. The size aspect of the typical surface interaction problem leads to the intuitive conclusion that valid explanations of observed phenomena are more likely to result from a macroscopic approach, such as the use of continuum mechanics, than from a molecular approach. To some extent, history bears this out because the molecular hypotheses of friction set forth by Erving, Tomlinson, Hardy, and Deryagin have not met with wide acceptance.

In defense of the chemists and others who think in terms of molecules, we must ask if a fundamental understanding of the interactions of surfaces (which are really huge masses of molecules) can be achieved without developing an understanding of what happens during the interaction of individual molecules. Undoubtedly there is much merit in both the molecular and macroscopic approaches, but I suspect that the major difficulty may lie in bridging the gap between, say, the viewpoints of the chemist and the thermodynamicist on particular problems.

During early investigations of lubrication phenomena, efforts gravitated into development of hydrodynamic theory or into studies of boundary lubrication. The idea quickly developed that these were the two regimes of surface interaction, with a sharp line of demarcation existing between them. As researchers gained knowledge and sophistication, however, it became increasingly apparent that this demarcation between full fluid film and boundary lubrication is a regime, or several regimes, in itself. This is now commonly referred to as the region of thin film or elastohydrodynamic lubrication. As one would suspect, elastohydrodynamic theory is more complex than hydrodynamic theory because the elastic properties of the materials, as well as the rheological properties of the fluid, become of prime importance.

The study of surface interactions appears to be of least importance in the full film regime. What goes on at the two solid-fluid interfaces may not be well understood, but the assumptions used in developing hydrodynamic theory are apparently valid. In most instances theory can be used to predict quite accurately behavior in the full-film regime.

Surface interactions become considerably more important in the elastohydrodynamic regime. The rheological behavior of the fluid is really not understood. Indications are that viscosity theories developed using a continuum mechanics approach may not be adequate. Brush discusses the problems in developing a satisfactory theory of liquid viscosity because of disparities between the theories of mechanics and the theories of chemistry (ref. 76). Smith points out that the Bernal approach (considering liquids as composed of irregularly packed aggregates of spheres) may have certain advantages in developing a more satisfactory theory to explain liquid behavior (ref. 77). In a sense this work bridges the gap between the continuum mechanics and molecular approaches because the spheres are analogous to molecules. Dr. Ling mentions a viscoelastic layer on an elastic half space, among the various models that have been studied. This analysis is complex, however, and only the mechanical aspects of the problem have been touched.

Finally the study of surface interactions under conditions where the surface layers come into intimate contact—the so-called boundary lubrication regime—is more or less the whole ball of wax. With chemical, mechanical, and thermal effects entering the picture, the problem is truly complex. It is not known yet whether sufficiently realistic models can be devised which are amenable to solution.

The task at hand is not merely a matter of intensified, multidisciplinary analytical effort. The theoretician can judge the value of his work only through experiments, and he is limited by the degree of realism he can maintain in a model while still retaining mathematical tractability. The experimentalist, on the other hand, is limited by his measurement techniques and instrumentation. Transducer size limitations form a barrier to obtaining highly localized measurements. Inherent response rates of measuring, readout, and recording instrumentation determine how accurately transient phenomena can be studied.

Present techniques for measuring surface temperatures are really quite crude. In the area of wear measurement, the use of radioactive tracer techniques has resulted in a significant improvement in detecting and quantitatively measuring minute wear rates. Perhaps a similar method of "tagging" surface molecules and observing their transient thermal behavior could be developed.

H. S. Cheng (Mechanical Technology Incorporated, Latham, New York)

Dr. Ling not only thoroughly reviewed past important contributions on various thermal and mechanical effects related to sliding contacts, but he also gave a totally new perspective on surface interaction in sliding. This perspective involves employing irreversible thermodynamics as the center through which all the thermal, mechanical, and even chemical effects in sliding contacts and the governing laws are connected. It is indeed a great pleasure to comment on this valuable work.

Four methods have been developed by the author to calculate the maximum surface temperature in a sliding contact. The first method is essentially a refinement of the well-known Blok theory in which the heat division is determined by matching the maximum surface temperature at the interface. This approach is conceptually simple and its results are usually compact and usable. For this reason, the Blok theory, and any of its refinements, has been used most extensively in the correlation of surface temperature measurements and in the study of the failure of sliding contacts by scoring. However, as pointed out by Dr. Ling, these methods are incapable of giving the detailed temperature distribution in the contact and are also inadequate for the cases where a large temperature jump exists across the interface.

To remove the first weakness, the author has developed the second method in which the detailed temperature distributions are matched exactly everywhere throughout the interface. This gives rise to the heat partition as a function of (x, y) in the contact instead of a single heat partition factor. While this method represents a definite improvement over the Blok approach, it is somewhat cumbersome and complex for many engineering applications.

The author's third method departs from the classical approach of matching the interfacial temperature for smooth surfaces. Instead, he postulates that the heat generated is concentrated at the asperity contacts and the fraction of heat generated on the interface of each body is not necessarily equal to that flowing into the body. This requires the heat excess to be transferred across the interface throughout the area where asperities are not in contact and results in a temperature jump between the surfaces. The equation developed by the author (ref. 14) provides a quick means to predict this temperature jump and the surface temperatures. However, it must be noted that this method requires the accurate knowledge of λ and h_c, which must be determined independently by two separate experiments. The experiments carried out by Ling and Simkins (ref. 11) establishes a range of h_c for the particular conditions in their experiment. These data seem to indicate that the value of h_c is still highly dependent on λ. Until reliable data on λ are available, the prediction of surface temperatures by using the author's third method may not necessarily be accurate.

The last method represents a most complete treatment of the surface temperature in sliding contact between two rough surfaces. It differs basically from the other three methods in that the heat input is represented by randomly distributed point sources both in space and in time. The location of these sources corresponds to points of real contact. By using a randomly distributed heat input function, the temperature on the surface as well as in the interior of each body at any time interval can be calculated employing the well-known analytical solution of the temperature distribution due to a point source. Two assumptions are still inherent in this method. First, the temperature at the actual contacting point is continuous; in other words, the surface temperatures at the contacting points are matched. Second, the heat transfer across the surfaces where there is no actual contact is negligible. These assumptions certainly appear to be reasonable.

The last method has a special significance because it is the first successful solution of the detailed surface temperature considering two real surfaces with discrete contacting points. It is interesting to note two peaks in the histograms. The first peak at the lower temperature reflects the temperature level throughout the apparent area; the second peak at the higher temperature obviously indicates the temperatures at the contacting points. As suggested by the author, the temperature at the first peak is likely to have a significant influence on the general metallurgical state and the thermomechanical action in the solid.

The significance of the latter peak is difficult to assess. In a lubricated contact, it has been believed that the local temperature during asperity collision, even though it may be many times higher than the so-called transition temperature, will not cause scoring failure until the bulk surface temperature exceeds the transition temperature. In view of recent controversies on this topic (ref. 78), there is a need to reexamine this concept. To this end, the Ling and Pu's general method is certainly an effective tool.

The effect of lubrication is not included in the above four methods. It is interesting to consider the applicability of these methods under various regimes of lubrication.

In the regime of full-film elastohydrodynamic lubrication, the surfaces are completely separated and the heat is generated by the viscous shear of the lubricant. Under these conditions, no matching of the surface temperatures is necessary. The methods based on the matching of surface temperatures of the solids are not applicable here. However, for these cases, the surface temperatures can be calculated directly by coupling the heat transfer equations in the solids with the energy equation in the fluid. Successful solutions have been obtained in references 79 and 80. Figure 16 gives typical curves for a glass disk sliding against a steel disk at various sliding speeds. Note that the temperature of the glass surface is considerably higher than that of the steel because of its lower conductivity and surface speed.

In the regime of partial EHD lubrication or boundary lubrication, there will be actual physical contact between asperities. Under these conditions, temperature continuity must be imposed at the contact, and the author's last method on the general treatment of surface temperature is applicable here. Actually in these regimes there are two other associated problems that are just as important as the surface temperature calculation and for which there is still no good solution.

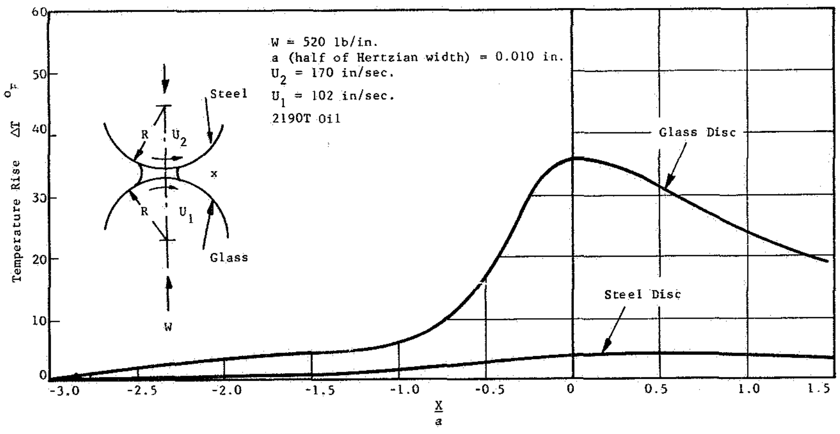


FIGURE 16.—Surface temperature rise of the glass and steel discs.

The first problem is the determination of the bulk temperature of the solids considering the heat transfer of the support. This problem is significant because of the influence of bulk temperature on viscosity and hence friction. Unfortunately, the solution for bulk temperature is coupled with the heat transfer of the entire system of the support. Its calculation is not difficult in principle but is likely to be tedious. Nevertheless it deserves more attention than has been given it in the past.

The second important problem associated with the surface temperature calculation is the local friction. Experimental friction studies yield only an average coefficient of friction but not the local value. To investigate the local friction during asperity collision, one almost has to use the approach described by the author (ref. 68). In the experiments on the collision of two conical tips, Ling and Lucek (ref. 71) found that the effect of lubrication is important for cone interval angles greater than 110°. Since asperities for most machined surfaces have cone interval angles greater than 150°, one can be quite certain that lubrication plays an important role in local friction. Therefore, unless the local sliding friction between two lubricated asperities is understood quantitatively, the calculation of surface temperature is likely to be inaccurate because of uncertain heat input.

Let us take a brief look into the problem of asperity collision in the presence of a lubricant. When two conical or spherical tips slide against each other, a thin lubricant film is possible if the ratio of the asperity height to width is sufficiently small. This mode of lubrication has been suggested by Christensen (refs. 81 and 82) as micro-elastohydrodynamic lubrication. Under this condition, the friction will be governed by the interfacial pressure and the fluid film shear. To determine these two quantities, one must consider the interactions between elastic and plastic deformation and the hydrodynamic action using realistic rheological properties of the lubricant for that condition. The major roadblock to this problem may not be the complexity of the analysis, but the lack of reliable static or dynamic rheological data under high pressure and high shear rate. It appears that this is an important and challenging problem deserving major attention.

Dr. Ling's review of solutions to thermomechanical coupling problems is quite thorough. I merely wish to add that these tools have been used to investigate the effects of temperature on stresses as well as on deformation in sliding contacts with full-film EHD lubrication (refs. 83 and 84). At least for moderately-loaded cases with slide-to-roll ratio up to 25 percent, the effects were found to be not as pronounced as originally speculated.

Dr. Ling has certainly compiled an impressive list of the available tools for studying surface interaction in sliding contacts. The usefulness of some of these tools has been proved in various friction and wear problems as well as in elastohydrodynamic lubrication. One can be certain that they will be used beneficially in future problems. We are indeed fortunate and grateful for his devotion and foresight in this field.

K. L. Johnson (University of Cambridge, Cambridge, England)

Professor Ling has given us a valuable bird's-eye view of surface interactions in sliding, rightly stressing the importance of thermal and mechanical interaction in the behavior of sliding contacts.

I should like, therefore, to turn first to the section on thermoelastic interaction. Ling and Mow (ref. 28) show how the profile at a sliding contact is modified by thermal expansion due to an arbitrary distribution of heat generated by friction at the interface. The analysis is applied to the example of an elastohydrodynamic contact in figures 11 and 12 and table 1. From the point of view of interaction, it is important to ask how their results influence the distribution of contact pressure and, in turn, the estimate of heat generation in the oil film. Although a complete analysis is very complex, it is possible to give a rough answer to the question. If the thermal displacements shown in figure 12 represent a direct change in film thickness, the influence on the pressures in the film would be very great. But this is not the case. The film is very resistant to compression compared with the surface so that we should really compare the displacements in figure 12 with the amount of elastic flattening of the two elastic solids. In the example quoted, I would estimate this to be about 1.0×10^{-4} in., which is almost two orders of magnitude greater than the thermal distortion. In this case the influence of thermal distortion upon the pressure distribution within the contact region will be small. It would appear, therefore, that the interaction between thermal and elastic effects by virtue of expansion in elastohydrodynamic contacts is small.

The predominant thermal interaction in elastohydrodynamic contacts is, of course, through the influence of temperature on the shear properties (effective viscosity) of the lubricant film. However, in dry sliding contacts with conforming surfaces, the thermo-elastic interaction can have an important effect. This is due to concentrations of contact pressure and frictional heating at high spots on the surface. If modification to the surface profile by wear is added to the process of thermal expansion by frictional heating, Barber (ref. 85) has shown that a closed interaction loop is formed that can lead to periodic behavior. The cycle is as follows: (1) The load is carried at a few high spots on the stationary surface; (2) thermal expansion due to frictional heating raises those spots above the surrounding surface and concentrates the pressure at those spots; (3) the high spots wear until new spots come into contact and begin to take a share of the load; (4) the original spots unload, generate less heat and contract, leaving a hollow in the surface; and (5) the new spots take over the load and the cycle is repeated. The period of this cycle is measured in seconds and is long compared with the period of individual asperity contacts.

I should like to turn now to the question of heat generation in sliding contacts. Professor Ling distinguishes between the heats generated within the material of each of the two surfaces in his experimental system, and thereby deduces values for a heat-transfer coefficient h_c across the interface which are compared with static thermal resistance measurements in figure 6. It would appear that for these two coefficients to be strictly comparable, it is necessary that either the heat be generated at a distance from the contact point which is large compared with its size, or that the heat transfer by conduction be small compared with that by convection and radiation. On the other hand, if the depth of heat generation is small compared with the contact dimensions, the heat flow will be almost indistinguishable from that emanating from heat generated at the interface itself. In typical sliding experiments it would seem likely that the depth of heat generation is of a similar magnitude to the asperity contact dimensions. If this is so, the heat flow and bulk temperature distribution will lie between these two extremes.

The existence of a macroscopic temperature jump, as observed by the author, is not conclusive evidence of heat generated below the surface. In fact, a temperature jump will occur whenever the proportion of heat removed from the bodies differs from the "ideal" division given by Blok's solution (ref. 6), even though all the heat was generated at the interface. This conclusion will be appreciated if we consider the ideal case when the extremities of the two bodies are cooled to the same temperature; the heat will then be distributed according to Blok. If now the cooling effect on one body is reduced—a not improbable circumstance is an actual experiment—its temperature will rise and the distribution of heat flow will be disturbed. We can solve this new problem by superposing two solutions: (1) the solution for heat generation and equal temperature boundaries and (2) the solution with no heat generation but with a temperature difference at the boundaries. The second solution introduces a temperature jump. Furthermore, values of h_c calculated from this solution will be comparable with static measurements, since we are here considering heat flow between the distant boundaries of the solids instead of between points near to the contact area. It is possible to calculate the value of h_c due to conduction alone from Blok's solution as follows:

Assume σQ and $(1-\sigma)Q$ are the heat flows into the two solids with zero boundary temperatures and the interface temperature is T (fig. 3).

Since a temperature difference T produces a heat flow σQ into the first body, a temperature difference $-T(1-\sigma)/\sigma$ will produce a heat flow $(1-\sigma)Q$ out of it. Thus a temperature difference of $T+T(1-\sigma)/\sigma$ causes a heat flow of $(1-\sigma)Q$ between the two extreme boundaries, whence

$$h_c=\frac{(1-\sigma)Q}{T+T(1-\sigma)/\sigma}=\frac{Q\sigma(1-\sigma)}{T}$$

Finally, I should like to refer briefly to the problem of an elastic disc having a surface layer of different elastic properties which is loaded on its periphery. Rolling contact between layered discs is frequently encountered in processing machinery. Reference 38 gives only a summary of the work of Beck and Ling, so I should like to ask the author whether he has used the approach outlined in the reference to obtain the numerical solution to particular problems.

We have been interested in the rolling contact of such a system taking into account the influence of friction at the contact interface between the two rollers. If the surface layer is thin compared with the length of the contact area between the rollers, the state of stress in the layer is very dependent upon the magnitude of the friction forces at the interface.*

This communication is the outcome of rewarding discussions of Professor Ling's stimulating paper with my young colleagues J. R. Barber and R. H. Bentall.

D. G. Flom (General Electric Company, Valley Forge Technology Center, Philadelphia, Pennsylvania

Interactions at sliding interfaces are sufficiently complex so that any discussion of them is certain to raise further questions and engender controversy. Professor Ling has stated correctly that irreversible thermodynamics, thermomechanical coupling, and continuum mechanics must be considered in any serious treatment. To the latter must also be added atomic and molecular interactions for a proper physical understanding.

At the expense of clouding the issue even further, I would like to point out that, in calculating surface temperatures, idealized treatments that neglect surface films can give misleading results. This is especially true if the two sliding solids and/or the interposed surface film have widely different thermal properties. Metals sliding against plastics represent a case in point. It is known that in rubbing steel against Plexiglas over a range of speeds and at different ambient tem-

---

*A paper by R. H. Bentall and K. L. Johnson describing the investigation is in preparation.

peratures, rather sharp inflections in the curves are obtained (ref. 86). A possible explanation for the inflections is that localized temperatures within the contact zone can reach the glass transition temperature of the Plexiglas. The interesting point, however, is that the temperatures required for this explanation do not agree with the temperature rises calculated from the equations of Archard (ref. 87), Jaeger (ref. 88), and Blok (ref. 89), if it is assumed that shearing occurs only at the steel-Plexiglas interface. On the other hand, good agreement is obtained if shearing is assumed to take place also within the plastic and if films of Plexiglas of the order of 0.6 to 0.8 micron in thickness are built up on the steel rider. Measurements with a Zeiss interferometer of films formed during sliding show that the observed thickness is closer to 3 microns. The film, however, is not continuous but exists in discrete "islands" covering about ⅓ of the total area.

These results have been obtained for a single sliding system (i.e., steel on Plexiglas), and it remains to be seen how many other systems behave similarly. We do know that Teflon is transferred to metals during sliding even though electrical methods are sometimes required to detect the presence of the polymer film. I suspect that inclusion of film effects in Professor Ling's treatment may not be an easy task, but I heartily encourage the attempt.

LECTURER'S CLOSURE

I am very grateful for the pertinent comments of Mr. Anderson. I agree with him wholeheartedly that there is a gap that needs to be bridged between the viewpoints of the chemist and the engineer. I might add also the gap between those who work at the microscopic scale and those who work on the macroscopic scale. I believe the common ground is modern thermodynamics and through it we can indeed look forward to bridging some of these gaps. With regard to the micro and macro approaches, I believe the modern continuum mechanics which has imbedded feature of the micro structure is in the right direction. Mr. Anderson's comments on possible techniques in measuring surface temperature are certainly welcome and should be attempted.

I am indeed overwhelmed by the complimentary remarks of Dr. Cheng, whose work has been a shining example of the power of continuum mechanics. Aside from my own susceptibility to flattery, I am happy indeed about Dr. Cheng's remarks on behalf of those who have devoted their entire lives to the development of what we know as modern continuum mechanics. To this group we may well point to Maxwell, Hertz, von Mises, and Mindlin. Perhaps it would be appropriate here to say the power of continuum mechanics was again demonstrated in a few minutes by Professor M. C. Shaw, through his comments on the orientation of fractured asperity and the associated friction calculations.* Also, at the risk of being too trite, I believe it is important to point out that continuum mechanics is a very general tool, and the notion that it can only handle homogeneous problems is, of course, a complete misunderstanding.

Dr. Johnson's comments are valuable and pertinent and I am certainly very appreciative of his frank discussions. There are several points I would like to discuss. The first has to do with the sample elastohydrodynamic calculation which I made. Of course, it is only an example that I used to estimate the thermomechanical interaction for that case. It gave me the indication that the thermal interaction is at best weak. However, the main contribution of the paper is that it provides a simple tool which can be used in many other ways. I shall cite an example later. I am pleased that attention is drawn to the interesting paper by Drs. Barber and Johnson in which they use an ingenious experiment to demonstrate the thermomechanical interaction in a sliding process involving wear. I should like to take this opportunity to indicate that the same phenomena occur in the real sliding situation as exhibited by the Barber/Johnson experiment. Of course, for this calculation the main thermoelastic result in the paper which Dr. Johnson discussed would be most useful to obtain quantitative result. Finally, coming to Dr. Johnson's discussion on the coefficient of heat transfer, I should like to point out that the suggested reason for the temperature jump is one of the two reasons mentioned in my paper. Incidentally, this is the favored model.

Finally, I wish to thank Dr. Flom for his provocative discussion. I want to clarify first that continuum mechanics has been effective in describing actions of matter in its usual phenomenological manifestations. Atomic and molecular description, while more basic, has not been as effective for quantitative use. I wish to reiterate that it is not my purpose to ignore the atomic and molecular scale. On the contrary, I wish to emphasize that modern development in micro-continuum is offering a chance to bridge the gap between macro-behavior and microstructure.

Based on our multi-contact model and experience in our laboratory and elsewhere, the simple formulas of Archard, Jaeger, and Blok are not expected to give the proper values. This is the reason for our original model. In fact, considering time and spatial divisions, there are generally two peaks in the temperature spectrum: (1) grand average of values for various space and time divisions and (2) average of values for the divisions where intimate contacts are or have been. The second peak may be one order of magnitude larger than the first. It is

---

*See Dr. Shaw's discussion of Dr. Tabor's lecture on Critical Appraisal and Research Opportunities—The Lubrication Research Viewpoint.

to be noted that the Blok-Archard formula gives the first peak value.

This model, as such, has usually given us good results for cases where films exist. In certain cases the results are good because of the intimate-contact nature of the model. During the last 4 months, we have been developing techniques to incorporate film effects.

NOMENCLATURE

A_n	area of the nth spot of the contact region
a	semi-length of contact square
E_n	emf generated at the nth spot of the contact region
h_c	thermal conductance
$\bar{h}_c$	an upper bound of h_c
$\underline{h}_c$	a lower bound of h_c
$I(R)$	function
$K_{1,\ 2}$	thermal conductivity of bodies 1 and 2, respectively
l	length of contact square
n	number of contact spots
P	potentiometer reading
Q	heat flux generated at the interface
$Q_{1,\ 2}$	heat flux through the periphery of the slider and the back of rider, respectively
Q_h	heat flux through the interface
$\bar{Q}_h$	an upper bound of Q_h
$\underline{Q}_h$	a lower bound of Q_h
$\underline{q}$	heat flux
q	dimensionless measure of the maximum heat input rate
R	dimensionless number $Va/4k_2$
s	dimensionless normal stress
T_0	cold junction temperature; reference temperature
T_s	averaged interface temperature
$T_{1,2}$	temperature on bodies 1 and 2, respectively; specific temperature
ΔT	temperature jump at interface
$\bar{T}$	averaged temperature
$\bar{T}_0$	averaged surface temperature
U_2	displacement perpendicular to the surface
V	speed of moving body
v_α	the αth substrate
W	normal load
x, y, z	coordinates
α	index
β	$\pi/2-\theta_0$
ϵ	internal energy density
$\bar{\zeta}$	dimensionless measure of the extent of plastic zone from the center of plate
η	entropy
θ_0	semi-opening angle of cone
κ_2	thermal diffusivity of body 2
λ	fraction of heat generated on the stationary body 1
ν	Poisson's ratio
$\bar{\xi}$	dimensionless measure of the distance from the center of plate to the edge of plastic zone

ξ, η, ζ	coordinates
π_{AB}	Peltier coefficient
σ	fraction of heat going into the stationary body 1; normal stress
τ	dimensionless time
τ_α	the αth thermodynamic tension

REFERENCES

1. Truesdell, C.; and Toupin, R.: The Classical Field Theories. Encyc. of Physics, Springer-Verlag, Berlin, vol. 3, 1960, p. 226.
2. Ling, F. F.: Surface Mechanics (a forthcoming monograph).
3. Gaylord, E. W.; Hughes, W. F.; Appl, F. C.; and Ling, F. F.: On the Theoretical Analysis of a Dynamic Thermocouple. Trans. ASME, vol. 80, 1958, p. 307.
4. Hughes, W. F.; and Gaylord, E. W.: On the Theoretical Analysis of a Dynamic Thermocouple: 2—The Continuous Area Interface. Jour. Appl. Mech., vol. 27, 1960, p. 259.
5. Shu, H. H. H.; Gaylord, E. W.; and Hughes, W. F.: The Relation between the Rubbing Interface Temperature Distribution and Dynamic Thermocouple Temperature. Jour. Basic Eng., ASME vol. 86, 1964, p. 417.
6. Blok, H.: Theoretical Studies of Temperature Rise at Surfaces of Actual Contact Under Oiliness Lubricating Conditions. Proc. Gen. Disc. Lubrication and Lubricants, IME, 1937, p. 222.
7. Ling, F. F.; and Saibel, E.: Thermal Aspects of Galling of Dry Metallic Surfaces in Sliding Contact. Wear, vol. 1, 1957, p. 80.
8. Ling, F. F.: A Quasi-Iterative Method of Computing Interface Temperature Distributions. Zeit. Ang. Math. und Physik, vol. X, 1959, p. 461.
9. Konwenhoven, W. B.; and Potter, J. H.: Thermal Resistance of Metal Contacts. Welding Research Supplements, vol. 27, 1948, p. 5155.
10. Fenech, H.; and Rosenow, W. M.: Prediction of Thermal Conductance of Metallic Surfaces in Contact. Jour. Heat Transfer, ASME, vol. 55, 1963, p. 15.
11. Ling, F. F.; and Simkins, T. E.: Measurement of Pointwise Juncture Condition of Temperature at the Interface of Two Bodies in Sliding Contact. Jour. Basic Eng., ASME, vol. 85, 1963, p. 481.
12. Ling, F. F.; and Ng, C. W.: On Temperatures at the Interfaces of Bodies in Sliding Contact. Proc. 4th U.S. Nat. Cong. Appl. Mech., 1962, p. 1343.
13. Ling, F. F.; and Pu, S. L.: Probable Interface Temperatures of Solids in Sliding Contact. Wear, vol. 7, 1964, p. 23.
14. Ling, F. F.: The Deformational and Geometrical Aspect of Surfaces in Sliding Contact. Fundamental Phenomena in the Materials Sciences, vol. 2, 1966, Plenum Press, p. 57.
15. Shore, H.: Thermoelectric Measurement of Cutting Tool Temperature. Jour. Washington Acad. Sci., vol. 15, 1925, p. 85.
16. Gottwein, K.: Die Messung der Schneiden temperatur beim Abdrehen von Flusseisen. Maschinenbau, vol. 4, 1925, p. 1129.
17. Herbert, E. G.: The Measurement of Cutting Temperature. Proc. IME, vol. 1, 1926, p. 289.
18. Nieman, G.; and Lechner, G.: The Measurement of Surface Temperatures on Gear Teeth. Jour. Basic Eng., ASME, vol. 87, 1965, p. 641.
19. Furey, M. J.: Surface Temperature in Sliding Contact. ASLE Trans., vol. 7, 1964, p. 133.
20. Schwerd, F.: Uber die Bestimmung des Temperatur felds beim Spanablauf. Zeit. VDI, vol. 71, 1933, p. 211.

21. Bowden, F. P.; and Thomas, P. H.: The Surface Temperature of Sliding Solids. Proc. Roy. Soc. (London), vol. A223, 1954, p. 29.
22. Pu, S. L.: Surface Temperature on a Truncated Right Circular Cone. Proc. 9th Midwestern Conf. Mech, 1966.
23. Ling, F. F.: Heat and Mass Transfer Effects in Sliding Metal Systems Lubricated by Solid Interfacial Films. AF Mat. Lab. Tech. Rept. 66–150, 1966.
24. Melan, E.; and Parkus, H.: Wärmespannungen infolge Stationarer Temperaturfelder. Springer-Verlag, Berlin, 1953.
25. Parkus, H.: Instationäre Wärmespannungen. Springer-Verlag, Berlin, 1959.
26. Boley, B. A.; and Weiner, J. H.: Theory of Thermal Stresses. John Wiley & Sons, 1960.
27. Nomacki, W.: Thermoelasticity. Addison-Wesley Publ. Co., Pergamon Press, 1962.
28. Ling, F. F.; and Mow, V. C.: Surface Displacement of a Convective Elastic Half-Space Under an Arbitrarily Distributed Fast-Moving Heat Source. Jour. Basic Eng., ASME, vol. 87, 1965, p. 729.
29. Jahanshahi, A.: Quasi-Static Stresses Due to Moving Temperature Discontinuity on a Plane Boundary. Jour. Appl. Mech., vol. 33, 1966, p. 814.
30. Sternberg, E.: On Transient Thermal Stresses in Linear Viscoelasticity. Proc. 3rd U.S. Nat. Cong. Appl. Mech., 1958, p. 673.
31. Atack, D.; and Tabor, D.: The Friction of Wood. Proc. Roy. Soc. (London), vol. A246, 1958, p. 534.
32. Hunter, S. C.: The Rolling Contact of Rigid Cylinder with Viscoelastic Half-Space. Jour. Appl. Mech., vol. 28, 1961, p. 611.
33. Flom, D. G.: Rolling Friction of Polymeric Materials—Elastomers. Jour. Appl. Phys., vol. 31, 1960, p. 306.
34. May, W. D.; Morris, E. L.; and Atack, D.: Rolling Friction of a Hard Cylinder Over a Viscoelastic Material. Jour. Appl. Phys., vol. 30, 1959, p. 1713.
35. Morland, L. W.: A Plane Problem of Rolling Contact in Linear Theory. Jour. Appl. Mech., vol. 29, 1962, p. 345.
36. Muki, R.; and Sternberg, E.: On Transient Thermal Stress in Viscoelastic Materials with Temperature-Dependent Properties. Jour. Appl. Mech., vol. 28, 1961, p. 193.
37. Weiner, J. H.: An Elastoplastic Thermal-Stress Analysis of a Free Plate. Jour. Appl. Mech., vol. 23, 1956, p. 397.
38. Beck, S. D.; and Ling, F. F.: Stresses in a Layered Elastic Disc. Proc. 5th U.S. Nat. Cong. Appl. Mech., 1966, p. 244.
39. Lemcoe, M. M.: Stresses in Layered Elastic Solids. ASCE Trans., vol. 126, 1961, p. 194.
40. Marguerre, K.: Spannungverteilung und Wellenausbreitung in der Kontinuierlichgestuetzten Platten. Ingenieur Archiv, vol. 4, 1933.
41. Biot, M. A.: Effects of Certain Discontinuities on Pressure Distribution in a Loaded Soil. Physics, 1935, p. 367.
42. Pickett, G.: Stress Distribution in a Loaded Soil with Some Rigid Boundaries. Proc. 18th Ann. Mtg. Highway Res. Board, vol. 18, 1938, p. 35.
43. Passer, W.: Druckverteilung durch eine elastische Schict. Akad Wissenschaften, Math-Natur Klasse Abt IIa, vols. 5 and 6, Heff, Wein, 1935.
44. Hogg, A. H. A.: Equilibrium of a Thin Slab on an Elastic Foundation of Finite Depth. Phil. Mag. (London), Apr. 1944.
45. Holl, D. L.: Thin Plates of Elastic Foundations. Proc. 5th Int. Cong. Appl. Mech., 1938, p. 71.

46. Volterra, E.: Sul problema generale della piastra poggiata su suolo Elastico. Publicazioni, Dell'Instituto per le applicazioni del calcolo, N. 201, Consiglio Nazionale Delle Ricerche, Roma, 1947.

47. Burmister, D. M.: The General Theory of Stresses and Displacements in Layered Systems . . . . Jour. Appl. Phys., vol. 16, 1945, pp. 89, 126, 296.

48. Sundara, K. T.; Iyengar, R.; and Alwar, R. S.: Stresses in a Layered Half-Plane. Proc. ASCE, vol. 90, 1964.

49. Barovich, D.; Kingsley, S. C.; and Ku, T. C.: Stresses on a Thin Strip or Slab with Different Elastic Properties from That of the Substrate Due to Eliptically Distributed Load. Int. Jour. Eng., Sci., vol. 2, 1964, p. 253.

50. Hsu, Y. C.; and Ling, F. F.: Shear Stresses at the Interface of Layered Elastic Systems Under a Moving Load. Proc. 3rd Tech. Mtg., Soc. Eng. Sci., 1967.

51. Sneddon, I. N.: Stress Produced by a Pulse of Pressure Moving Along the Surface of a Semi-Infinite Solid. Rendiconti Circolo Matematico di Palmero, vol. 2, 1952, p. 57.

52. Cole, J.; and Huth, J.: Stresses Produced in a Half-Plane by Moving Loads. Jour. Appl. Mech., vol. 25, 1958, p. 433.

53. Sackman, J. L.: Uniformly Moving Load on a Layered Half-Plane. Proc. ASCE, vol. 87, 1961.

54. Lancaster, J. K.: Elastic Deformation and the Wear of Electrographite. Nature, vol. 196, 1962, p. 368.

55. Maye, R. F.; and Ling, F. F.: General Solution of the Ring for Cosserat Material. Proc. 5th U.S. Nat. Cong. Appl. Mech., 1966, p. 304.

56. Batra, S. K.; and Ling, F. F.: Deformation Friction and Interface Shear Stress in Viscoelastic-Elastic Layered System under a Moving Load. ASLE Trans., vol. 10, 1967, p. 294.

57. Abrahamson, G. R.; and Goodier, J. N.: The Hump Deformation Preceding a Moving Load on a Layer of Soft Material. Jour. Appl. Mech., vol. 28, 1961, p. 608.

58. Greenwood, J. A.; and Williamson, J. B. P.: The Contact of Nominally Flat Surfaces. Proc. 2nd Int. Conf. Elec. Contacts, Graz, Austria, 1964.

59. Greenwood, J. A.: The Area of Contact Between Rough Surfaces and Flats. Jour. Lub. Tech., ASME, vol. 89, 1967, p. 81.

60. Abbott, E. J.; and Firestone, F. A.: Specifying Surface Quality. Mech. Eng., vol. 55, 1933, p. 569.

61. Rathbun, F. O., ed.: Design Criteria for Zero Leakage Connectors for Launch Vehicles. G. E. Co. Rept. 64GL97, 1964. (Discussion by F. F. Ling and F. O. Rathbun, Jr., On the Mechanism of Contact between Metal Surfaces—The Penetrating Depth and the Average Clearance, T. Tsukizoe and T. Hisakado, Jour. Basic Eng., ASME, vol. 87, 1965, p. 667.)

62. Schram, A.: La Precision dans les mesures en regime moleculaire. Commissariat A L'energie Atomique PA. TE/RT, 1963.

63. Bowden, F. P.; and Tabor, D.: The Friction and Lubrication of Solids, Pt. I, 1954, and Pt. II, 1964, Oxford Univ. Press, London.

64. Dyachenko, P. E.; Tolkacheva, N. N.; Andreev, G. A.; and Karpova, T. M.: The Actual Contact Area Between Touching Surfaces. Consultants Bureau, New York, 1964 (Russian text published in 1963 by Inst. Mech. Eng., Moscow).

65. Archard, J. F.: Contact and Rubbing of Flat Surfaces. Jour. Appl. Phys., vol. 24, 1953, p. 981.

66. Archard, J. F.: Single Contacts and Multiple Encounters. Jour. Appl. Phys., vol. 32, 1961, p. 1420.

67. Ling, F. F.: Some Factors Influencing the Area-Load Characteristics of Semi-Smooth Surfaces under "Static" Loading. Trans. ASME, vol. 80, 1958, p. 1113.
68. Ling, F. F.: On Asperity Distributions of Metallic Surfaces. Jour. Appl. Phys., vol. 29, 1958, p. 1168.
69. Ling, F. F.: The Deformational and Geometrical Aspect of Surfaces in Sliding Contact. Fundamental Phenomena in the Materials Sciences, vol. 2, 1966, Plenum Press, p. 57.
70. Ling, F. F.: Mechanics of Sliding Surfaces. Metals Engineering Quarterly, vol. 7, 1967, p. 1.
71. Ling, F. F.; and Lucek, R. C.: On Model Studies of Metallic Surface Asperities. Jour. Appl. Phys. vol. 30, 1959, p. 1559.
72. Peterson, M. B.; and Ling, F. F.: Friction and Lubrication at Extreme Pressures. Friction and Lubrication in Metal Processing, ASME, 1966, p. 39.
73. Green, A. P.: The Plastic Yielding of Metal Junction Due to Combined Shear and Pressure. Jour. Mech. and Phys. of Solids, vol. 2, 1954, p. 197.
74. Hertz, H.: Über die Beruhrung fester elastischer Korper. Journal für die Reine und Angewandte Mathematik, vol. 92, 1881, p. 156.
75. Mow, V. C.; Chow, P. L.; and Ling, F. F.: Microslips between Contacting Paraboloids. Jour. Appl. Mech., ASME, vol. 34, 1967, p. 321.
76. Brush, S. G.: Theories of Liquid Viscosity. Chem. Rev., vol. 62, 1962.
77. Smith, F. W.: A New Mathematical Basis for Rheology and Its Applications to the Behavior of Lubricants Under Extreme Stress. Proc. IME, vol. 180, Pt. 3B, 1965–66, p. 227.
78. Fein, R. S.; Rowe, C. N.; and Kruez, C. W.: Transition Temperatures in Sliding System. ASLE Trans., vol. 2, 1959, p. 50.
79. Dowson, D.; and Whitaker, A. V.: A Numerical Procedure for the Solution of the Elastohydrodynamic Problem of Rolling and Sliding Contacts Lubricated by a Newtonian Fluid. Elastohydrodynamic Lubrication. IME, 1966, p. 57.
80. Cheng, H. S.: A Refined Solution to the Thermal-Elastohydrodynamic Lubrication of Rolling and Sliding Cylinders. ASLE Trans., vol. 8, Oct. 1965.
81. Christensen, H.: Review of Failure Mechanisms in Highly Loaded Rolling and Sliding Contacts. Chapter IV. Rept. MTI–67TR23, Mechanical Technology Inc., 1967.
82. Christenson, H.: Lubrication and Wear. Lubrication, vol. 51, 1965, p. 60.
83. Mow, V. C.; and Cheng, H. S.: Thermal Stresses in a Convective Elastic Half Space Associated with an Arbitrarily Distributed Fast-Moving Heat Source. Rept. MTI–65TR60, Mechanical Technology Inc., 1965.
84. Kannel, J. W.; Walowit, J. A.; Bell, J. C.; and Allen, C. M.: The Determination of Stresses in Rolling Contact Elements. J. Lub. Tech., vol. 89F, 1967, p. 453.
85. Barber, J. R.: The Influence of Thermal Expansion on the Friction and Wear Process. Wear, vol. 10, 1967.
86. Bueche, A. M.; and Flom, D. G.: Surface Friction and Dynamic Mechanical Properties of Polymers. Wear, vol. 2, 1959, p. 168.
87. Archard, J. F.: The Temperature of Rubbing Surfaces. Wear, vol. 2, 1959, p. 438.
88. Jaeger, J. C.: Moving Sources of Heat and Temperature at Sliding Contacts. Proc. Roy. Soc. (N.S.W.), vol. 76, Pt. 3, 1942, p. 203.
89. Blok, H.: Theoretical Study of Temperature Rise at Surfaces of Actual Contact Under Oiliness Lubricating Conditions. Proc. Gen. Disc. Lubrication and Lubricants, IME, 1937, p. 222.

Friction and Adhesion

M. E. MERCHANT
The Cincinnati Milling Machine Co.
Cincinnati, Ohio

This paper reviews the understanding of the fundamentals of friction and adhesion at present. Adhesion—a major factor in the friction process—is presented as an inherent action of most solid bodies when brought into contact. Its effects, however, are normally masked by surface contamination, very limited area of actual contact, and the effects of elastic recovery. Friction, then, results primarily from the force required to shear the junctions formed (or the contaminant films present) at the areas of actual contact, the area to be sheared being virtually proportional to the applied load in all normal cases.

UNDERSTANDING AND CONTROL of surface friction between solids, and its relation to adhesion, is, and always has been, one of the most practical and necessary pursuits of man in conquering the physical world. Friction, with the burden of useless force and wasteful dissipation of useful energy it places on all of man's mechanical doings and creations, can also be most beneficial.

Imagine, for example, the consequences of the ridiculous situation whereby a huge cosmic switch labeled "friction" could be pulled, "turning off" all surface friction in the universe. As the last traces of friction faded away, man would find himself on the ground or floor almost immediately, unable to rise unless he were fortunate enough to obtain a hook-like grasp on some properly shaped object, not itself supported by friction. He would also be unable to propel himself except by grasp on properly shaped and anchored objects. He would, in fact, find himself and his neighbors, together with numerous objects, gliding swiftly via recurring collisions to the lowest point in the neighborhood (perhaps already occupied by water). There, he would be joined by automobiles, trucks, and buses, whose drivers were turning their steering wheels without avail, while the vehicles disintegrated as their nuts, bolts, and screws unwound under the slightest load. One could go on imagining and perhaps determine whether mankind could survive even in such a situation—but enough of such speculations. Friction is very real and we can be grateful for it, even as we wrestle with the problems of harnessing and controlling it in virtually every phase of modern technical activity. Because a realistic understanding of the fundamentals of friction and adhesion is required, this paper reviews our present state of such understanding for the benefit of those not presently well acquainted with the field.

In investigating the nature of friction, man has found that his approach must be interdisciplinary. The early (15th through 18th century) theories of friction consisted of simple mechanical concepts. However, experimental and analytical methods in subsequent research made it evident that such simple explanations were inadequate. Today's understanding of friction and adhesion draws on knowledge from all fields of physics and chemistry, and continuing development and application of that understanding challenges the best talents and requires some of the most advanced knowledge of physical scientists and engineers in all disciplines.

Similarly, the more technically sophisticated man's world has become, the greater have been his problems in controlling or harnessing friction. Early work was largely concerned with engineering problems related to such simple mechanisms as ropes and belts on pulleys and sheaves, and wheels on axles. Today, complex mechanical equipment and processes operate at temperatures, speeds, and loads that were undreamed of a few years ago, and these variables will increase in the years ahead to meet rapidly rising requirements for performance and utility in manufactured products. Furthermore, man's venture into space has opened up a new environment, characterized especially by the "hard vacuum" in which equipment and processes must now operate—an environment particularly conducive to producing high friction and strong adhesion between solids.

I have not been in close working touch with the research being done on friction in recent years, even though I worked actively in the field in the past. Coming thus "fresh" upon the results of recent research, I can particularly feel its impact. Understanding the fundamentals of friction and adhesion between solids, which have greatly increased in the past decade, can do much to improve the practical use and control of these phenomena. Those who have worked in the field in recent years are to be congratulated on the excellent progress they have made. At the same time, it is satisfying to note that many concepts developed 20 to 30 years ago, e.g., that of the major role played by adhesion in the friction process (one which I helped develop), have proved sound and serviceable and provided a good foundation to the modern understanding of friction.

ADHESION

General Considerations

As intimated above, 20th century research on friction has served to demonstrate with ever-increasing clarity that adhesion between solids is the major factor contributing to friction between them. Thus, I open this review with a discussion of adhesion as essential background for a general understanding of friction. The discussion, then, is not a discussion of adhesion in general, but of those aspects important to frictional phenomena. Since friction between metals is of particular importance, adhesion between metals is emphasized.

Adhesion between solids can be looked upon as an extension of the cohesion (or internal "adhesion" between their various elemental or particulate constituents) holding them together. Ideally, if chemically clean surfaces of two solid crystalline bodies of the same composition are brought in intimate contact in the absence of a contaminant, there should be nothing remaining to distinguish the original interface from, for example, a grain boundary in the original crystalline material. Thus, they should adhere in a manner indistinguishable from the cohesion of the original material. We all know, however, that we do not experience such odd behavior normally. Why do two smooth dimes, when rubbed together, not stick to one another? The two key phrases in the above idealization are "chemically clean surfaces" and "intimate contact."

Surface Contamination

Ordinary solid surfaces are never chemically clean, being the repository of various oxides, adsorbed gases, condensed vapors, and organic films from exposure to their environment. These surface contaminants, normally low in strength or adhesion, will touch and adhere to each other when the solids "contact," keeping the host material separated. That is one reason the dimes do not stick. Secondly, the surfaces of solid bodies are never ideally flat and smooth; they always have some waviness and roughness, even if on a microscopic (or even submicroscopic) scale. Thus, when two apparently flat solid surfaces are brought together after being chemically cleaned in a vacuum, they will touch and adhere to each other at only a few scattered points. Thus, the solids can be pulled apart with little (often scarcely detectable) force, breaking the miniscule junctions.

Bailey (ref. 1), using cleaved mica surfaces, has demonstrated that, if surfaces are chemically clean and molecularly flat, they will indeed unite to produce strengths virtually equal to the cohesive strengths of the parent bodies. Mica has the unusual characteristic of being cleavable along a single plane of atoms over very large areas. The resulting surfaces pick up contamination slowly, even in air at atmospheric pressure. Bailey measured the energy ($\sim$300 erg/cm^2) required to produce such surfaces, then brought them back into contact, whereupon they reunited. The energy required to cleave along the same surfaces a second time was measured at 15 percent lower than that for original cleavage. This small reduction was attributed largely to the difficulty of avoiding entrapping air when the surfaces are reunited.

The Area of Actual Contact

When two conventional solid surfaces (either clean or contaminated) are brought into contact and touch at minute scattered points, the stresses developed at these points of contact are very high. Thus, under all but the most minute load, their value exceeds the elastic limit of the weaker of the two materials and reaches its yield pressure or indentation hardness (mean pressure hardness). Thus, plastic flow of the contacting asperities take place, enlarging the area of actual contact and permitting additional asperities to come into contact. The size of the area of actual contact thus generated can be calculated (ref. 2), knowing that the yield pressure or indentation hardness of most materials is virtually a constant and thus the area is proportional to the load:

$$A=\frac{W}{p} \tag{1}$$

where A is the area of actual contact, p the yield pressure for the weaker material, and W the load.

It can be argued that the deformation of the contacting asperities may remain purely elastic under proper conditions of load and surface roughness, and that, in any case, the asperities undergo not only plastic but also some elastic deformation, which has not been taken into account in the above simple analysis. More sophisticated approaches, taking account of elastic deformation, have therefore been made by a number of investigators in recent years, including Archard (ref. 3), Lodge and Howell (ref. 4), and Greenwood and Williamson (ref. 5); however, these studies also generally conclude that, for usual surface topographies and most materials, equation 1 is a good approximation to reality.

Direct experimental determination of the relationship between actual area of contact and load is not wholly possible at this time, but the best approximations are offered by electrical resistance measurement in the case of metals (refs. 6 and 7), and direct optical observation in the case of transparent nonmetals (refs. 8 through 11). Results of these methods of measurement further serve to confirm the fact that equation 1 is a good approximation to reality. However, since all of the present methods for experimental determination of true contact areas between solids have some degree of uncertainty, this subject requires further research.

Elastic Recovery

Supposing now that our two smooth dimes had chemically clean surfaces and were pressed together with a force great enough to produce a sizeable area of actual contact. After release of the load, it may still be found that, although they did stick together, the force required to pull them apart would be considerably less than the load originally applied. This result may seem somewhat puzzling in the light of equation 1. If the contact area developed by loading is proportional to the load, we might expect that the force required to separate the two bodies by plastic deformation of the multiple minute junctions between them would approximate that required to establish the junctions. In fact, this apparent anomaly has led some investigators to argue that the areas of actual contact between surfaces must have resulted solely from elastic rather than plastic deformation, or that adhesion between contacting asperities does not occur. However, even if the contacting asperities deform wholly by plastic deformation and adhere with a strength equal to the parent material, the above result must be expected in the case of bodies having a reasonably high elastic modulus. This was first pointed out by Ernst and Merchant (ref. 12). Consider, for example, the case of a hard ball indenting a softer plate, as in hardness testing. When the load is decreased, the two surfaces that have been in intimate (though not necessarily total) contact are peeled apart by elastic forces in a process known as elastic recovery. If this process does not occur whenever the load is decreased on an area of contact established by plastic deformation, Hertz's equations of elastic contact, as applied to the final macroscopic surface curvature, would be violated.

Ernst and Merchant reasoned from this fact that equation 1 may be expected to hold approximately, even for decreasing load, when the contacting bodies have high elasticity, the contacting surfaces are initially of (at least slightly) different curvature, and the actual area of contact between them is less than the apparent contact area. Under such conditions (fairly usual for metallic bodies) the peeling introduced by the elastic recovery can produce plastic flow in the scattered junctions between the two surfaces when the load is decreased. The tensile stress necessary to produce plastic flow in a work-hardened metal rod with extremely small cross-sectional area is approximately equal to the yield pressure, p, of the given metal. Thus, the variation of contact area with decreasing load should obey, at least approximately, equation 1. Bowden and Tabor (ref. 13), in a more sophisticated analysis of this situation, show that this result is not true if the material of at least one contacting body is significantly ductile. Their model, figure 1, illustrates that the junctions or "asperity bridges" of ductile materials can elongate appreciably as elastic recovery peels the surfaces apart. Thus, the elongated junctions can bridge at least a portion of the gap, as shown, leaving a significant area of true contact between the bodies at zero load; a detectable force should be required to pull the bodies apart after their chemically clean surfaces have been brought together under load W. This adhesion force should be influenced by ductility, elastic modulus, and load. The greater the ductility (allowing greater elongation of the junctions), the less the elastic modulus (resulting in less elastic recovery), and the greater the initial load (resulting in a larger initial area of actual contact), the greater should be this force.

The Adhesion Coefficient

In the light of the above, it becomes interesting to consider the ratio of force Z (required to separate two adhering bodies) to load W (with which they were originally pressed together). This ratio, Z/W, is known as the adhesion coefficient γ. Equation 1 and the discussion above indicate that, for a given material and a given set of surface conditions (e.g., geometry and cleanliness), the coefficient of adhesion should be relatively constant and independent of load. The area of actual contact may be expected to increase linearly with applied load, and the residual area of actual contact, after load is removed, may be expected to be a constant fraction of the original contact area for a given material and surface geometry. Thus, the force to separate the surfaces should bear a constant relation to the original load. If no elastic recovery occurs and the areas of actual contact adhere fully, force Z required to sever those areas should be approxi-

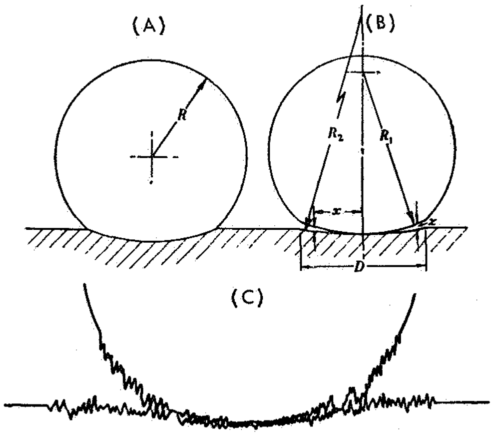


Figure 1.—Diagram of theoretical model of elastic recovery effects. (a) Sphere on flat under load. (b) Effect of elastic recovery on removal of load. (c) Asperities and asperity bridges between surfaces (ref. 13).

mately equal to the load W that joined them. This is true because, as mentioned earlier, the tensile stress necessary to produce plastic flow in a work-hardened metal rod of extremely small cross section is approximately equal to the yield pressure p of the given metal. Thus, for the ideal case of full adhesion with no elastic recovery, the adhesion coefficient may be expected to have a value of the order of magnitude of unity.

McFarlane and Tabor (ref. 14) studied the adhesion of a steel ball to an initially flat indium surface by applying various loads, removing them, and then observing the resulting adhesion. Although the experiments were done in air, the ball was carefully cleaned and the surface of the indium freshly cut. In addition, the thin and brittle surface oxides of indium as well as its low elastic modulus and high ductility facilitate adhesion to other metals. Figure 2 shows the relation between the applied load and the resulting adhesion force. It can be seen that, for a given loading time (indium undergoes severe creep) and ball diameter, the relation between the two is linear. The adhesive force is roughly equal to the original load and thus the adhesion coefficient γ is approximately 1.

When the surfaces were separated, a thin film of indium was found adhering to the ball, showing that the junctions formed were at least as strong as the indium itself, even though the experiment was done in air. Furthermore, the adhesion force was found to be directly proportional to the area of the indentation, figure 3. Similar experiments were conducted with other soft metals of increasing modulus of elasticity and decreasing ductility, namely, lead, tin, cadmium, and aluminum. The results are shown in table 1. It may be seen that the adhesion coefficient decreases as the metals become less elastic and ductile, i.e., harder. It should be remembered that since these experiments were all run in air, the surfaces were protected to some extent by oxide films. Had chemically clean surfaces been exposed in a hard vacuum (10^{-9}–10^{-11} torr), the adhesion coefficients might all (with the exception of indium, which adhered completely) be expected to be higher. Nevertheless, the results validate much of the theory outlined in the previous sections, namely, (1) that surfaces of solids freed reasonably well from contaminating films will adhere, (2) that the area of actual contact between them, and thus the adhesion force, increases linearly with load, and (3) that increased tendency to elastic recovery (higher modulus of elasticity, lower ductility) decreases the residual adhesion remaining after the applied load is released.

In connection with the last point, Buckley (ref. 15) has recently reported some very interesting adhesion experiments on chemically clean surfaces of single pure-copper crystals in a hard vacuum (10^{-11} torr). The experiments were conducted on matched planes with

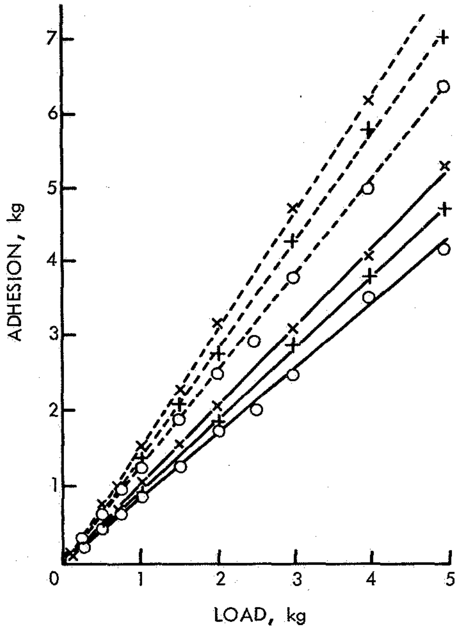


Figure 2.—Adhesion of clean steel spheres on indium. Diameters of spheres: X, ⅛ in.; +, ¼ in.; ○, ¾ in. Loading times: ——— 10 sec; ------ 1000 sec (ref. 14).

matched directions, where the same crystallographic plane was exposed and brought into contact on both specimens; the crystallographic directions in both planes coincided. The specimens were in the form of a hemispherical-ended rider contacting a flat surface. The results are shown in table 2. The adhesion coefficients obtained before sliding, shown in the third column of the table, may be compared with the modulus of elasticity values shown in the second column. Increasing elastic modulus, signalling increasing elastic recovery, again decreases the adhesion coefficient. The data further show that significant adhesion coefficients occur even with high-modulus metals when chemically clean surfaces are brought into contact under load.

The Effect of Tangential Stress or Sliding

The data found in the last column of table 2 illustrate another important feature of adhesion phenomena, namely, that sliding between contacting surfaces of ductile metals under load increases adhesion tremendously. Since the contacting surfaces were in this case chemically clean, the increase must result from an increased area of actual contact brought about by plastic deformation. This is so because under the action of combined shear and normal stresses the phenomenon known as "junction growth" can take place and equation 1 no longer holds. The theory of this phenomenon is described in the section on friction.

For the more usual case of contaminated surfaces in air, sliding has also been found to promote adhesion between metals, because the ability of the sliding action to displace or bury the interfering contaminants under the action of the plastic deformation taking place at the interface. Anderson (ref. 16) and Sikorski (ref. 17) studied this action and obtained considerable information on the mechanisms involved. Anderson, for example, found that when he placed two copper rods, with ends cut in planes normal to their axes, end-to-end under a 5-pound load and then twisted them through an angle of 180° in air, adhesion occurred. The value of the adhesion coefficient varied from test to test but, by making a sufficiently large number of tests, the median value for γ was found to be approximately 1. Anderson also carried out a microscopic examination of the interface between aluminum and copper rods adhered in the same manner and found that aluminum, copper, and the surface oxides of the two were all quite thoroughly mixed at the interface. Aluminum had penetrated into

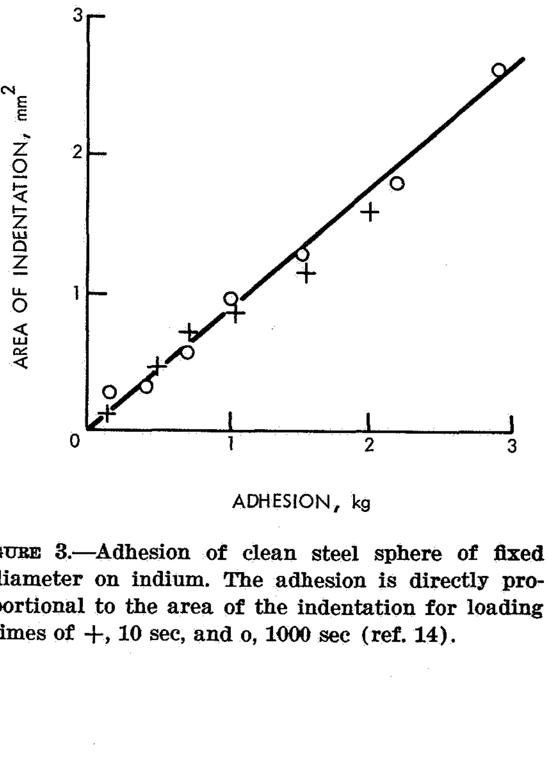


Figure 3.—Adhesion of clean steel sphere of fixed diameter on indium. The adhesion is directly proportional to the area of the indentation for loading times of +, 10 sec, and o, 1000 sec (ref. 14).

the copper up to about ½ mm, and some penetration of the copper by aluminum also occurred. Sikorski used this method to compile data on adhesion coefficients for a large number of pure metals in contact with themselves and correlated these results with measurements on the friction coefficient and with various bulk properties of the metals. He concluded that high hardness, low capacity for strain-hardening, high elastic modulus, and high recrystallization temperature inhibit adhesion and that a high adhesion coefficient indicates a high friction coefficient. (The reasons for this latter correlation become evident in the section on friction.) Sikorski also made some determinations of the adhesion coefficient for dissimilar pairs of metals, the results of which are presented in table 3. The surprisingly low adhesion of the pair Ag–Fe will also be dealt with in the next section.

Atomic and Molecular Considerations

Mechanical considerations play a part in all the adhesion phenomena discussed in the previous sections. For example, the actual contact area over which adhesion can occur is controlled by mechanical (plastic or elastic) deformation, and the breaking of adhesive bonds on release of load is controlled by the configuration, elasticity, and ductility of the contacting bodies. Furthermore, sliding, as discussed in the previous section, produces junction growth, disruption of contaminating films, and mechanical mixing and interlocking. Nevertheless, none of these is the basic cause of adhesion between the solid materials that have been brought into contact. As stated previously, adhesion of solids can be regarded as an extension of their internal cohesion; thus, adhesion phenomena are atomic and molecular in nature.

The different types of forces that can act within (or at the surface of) solid matter to produce cohesion (and thus adhesion) are

TABLE 1.—*Adhesion of Steel on Indium, Lead, Tin, Cadmium, and Aluminum in Air (Clean Surfaces, Load 2 kg, Time of Loading 1000 sec, ⅛-in. Ball)*

[From ref. 14]

Metal surface	Brinell hardness, kg/mm^2	Coefficient of adhesion
Indium	1	1. 5
Lead	4	0. 7
Tin	5	0. 4
Cadmium	22	0. 1
Aluminum	17	0. 0

TABLE 2.—*Adhesion and Friction Characteristics of Single and Polycrystalline Copper (99.99%) in Vacuum*

[From ref. 15]

Copper form and orientation	Young's modulus, 10^{11} dynes/cm^2	Adhesion coefficient before sliding [a]	Friction coefficient during sliding [b]	Adhesion coefficient after sliding [c]
Single crystal (100) matched planes and directions	6. 67	1. 02	>40. 0	>130
Single crystal (110) matched planes and directions	13. 1	0. 61	>40. 0	50. 0
Single crystal (111) matched planes and directions	19. 4	0. 30	21. 0	10. 5
Polycrystal	12. 0	1. 00	>40. 0	100

[a] Load 50 gm, 10^{-11} torr.
[b] Load 50 gm, sliding velocity 0.001 cm/sec, 10^{-11} torr.
[c] Load 50 gm, distance slid in preferred slip directions 0.735 cm, 10^{-11} torr.

well known and have been described concisely by Houwink and Salomon (ref. 18). These may be briefly characterized as follows:

(1) Homopolar or covalent bonding results from atoms sharing pairs of electrons. This is the normal chemical bonding occurring between atoms within molecules.

(2) Electrostatic or ionic bonding results from the purely electrostatic forces produced by one or more electrons transferred completely from one atom to another, thus converting the neutral atoms into electrically charged ions and forming ionic crystals.

(3) Metallic bonding is a "mixture" of covalent and electrostatic bonding. Although the metallic structure is crystalline, the electrons have great mobility and thus are constantly shared by different sets of "ions." (The metal can thus accommodate plastic deformation readily compared to the behavior of brittle ionic crystals.)

(4) Van der Waals forces act between molecules or in interaction of molecules with atoms in the absence of chemical reaction between the two. These forces, although also electrical in nature, arise from the uneven and fluctuating distribution of charge in the molecule, creating momentary and continuously shifting dipoles in the molecule and correspondingly induced dipoles in neighboring molecules or atoms. These forces are therefore much weaker than the previous three types.

Direct and meaningful measurements of adhesion between contacting solid bodies exhibiting these types of bonding are almost nonexistent except for the case of metals (metallic bonding). However, the interesting and convincing observations by Bailey on the readhesion of cleaved surfaces of mica (electrostatic bonding) have already been cited. Bowden and Tabor (ref. 13) have studied directly the adhesion of crystals of rock salt (ionic crystals). They placed the (100) faces of two thin square slabs of freshly cleaved rock salt (about 5×5×1 mm) in contact and compressed them between flat anvils until their original area was doubled. When removed from the anvils, the slabs adhered firmly. When cemented to brass grips and subjected to a tensile test, the interface withstood a stress of 0.4 kg/mm^2—equal to the strength of single rock-salt crystals. This evidence establishes that clean surfaces of ionic bodies brought into intimate contact by plastic deformation will adhere. Further, it demonstrates the identity of that adhesive bonding with the cohesive bonding within the ionic solid. When the same experiment was repeated with a (100) face of one slab in contact with a (110) face of the other, only slight adhesion was observed. This observation seems to agree qualitatively with the known low cohesion at grain boundaries between mismatched rock-salt crystals.

Concerning metals, an analogy with cohesion and plastic flow phenomena would lead us to expect a strong correlation between adhesion phenomena and the lattice structure. The correlation has already been confirmed by Buckley's data presented in table 2. It can be noted that the adhesion coefficient (either before or after sliding) is a marked function of the crystallographic orientation of single copper crystals, even when identical planes are brought together with their crystallographic directions matched. The apparent correlation of these ad-

TABLE 3.—*Coefficients of Adhesion for Dissimilar Metal Combinations in Air*

[From ref. 17]

Metal couple	Crystal structure	Number of adhesion tests	Average of adhesion coefficient
Al–Cu	fcc to fcc	30	0. 60
Ag–Fe	fcc to bcc	23	0. 002
Pt–Fe	fcc to bcc	40	0. 16
Rh–Fe	fcc to bcc	45	0. 03
V–Fe	bcc to bcc	40	0. 07

hesion coefficients with modulus of elasticity has been noted previously in the section on the adhesion coefficient; it was pointed out there that the higher modulus planes should exhibit greater elastic recovery. Thus, the elastic stresses should "peel" such planes apart more completely when the original compressive load is removed and reversed, leaving a smaller area of real contact to be severed by the tensile load. As is known, the spacing between adjacent planes in the face centered cubic system is $(111) > (110) > (100)$. Thus, the force required to rupture junctions along surfaces parallel to the original flat specimen surfaces should be least for the (111) and greatest for the (100) plane. For this reason, the adhesion coefficients for copper can be expected to be $(111) < (110) < (100)$, in agreement with table 2.

Buckley also observed adhesion coefficients for contact between dissimilar planes of single copper crystals (corresponding, in a sense, to grain boundary interfaces in polycrystalline copper); the results are given in table 4. Even though one of the dissimilar planes in contact, is the low modulus (100) plane, the adhesion coefficient before sliding is less than for either of the corresponding similar planes in contact indicating that the junction between unmatched planes is more readily ruptured than that between matched planes.

The possible effects of the type of lattice structure may on adhesion phenomena have been dealt with by a number of investigators, among them Keller (ref. 19) and Sikorski (ref. 17). The latter plotted adhesion coefficients for like pairs of metals as a function of modulus of elasticity, given in figure 4. We know that the adhesion coefficient should decrease with increasing elastic modulus, and the general trend of the data in figure 4 agrees with this conclusion. It is immediately evident, however, that an additional factor, namely, the type of metal lattice structure, is influential. In figure 4, Sikorski drew dotted lines to enclose the points that correspond to metals having cubic or tetragonal crystal structures. The metals having a close-packed hexagonal structure at room temperature (i.e., Cd, Zn, Ti, Mg, and Zr) are in a separate group, exhibiting considerably lower adhesion coefficients than the cubic or tetragonal metals of like elastic moduli.

That close-packed hexagonal metals may exhibit lower adhesion characteristics than cubic had been inferred earlier by Ernst and Merchant (ref. 12) from friction data. Most recently, Buckley (ref. 15) compared the adhesion coefficients for matched (111) planes of single copper crystals, a cubic metal, with those for matched (0001) planes of single cobalt crystals, an hexagonal metal. The results are given in table 5. It is evident that adhesion, being undetectable with Buckley's apparatus, is strikingly lower for the cobalt, even after sliding. Part of this marked difference in the adhesion characteristics of cubic and hexagonal metals may be caused by the latter's lack of ductility (a matter dealt with by Buckley in a later paper (ref. 20)). This low ductility may cause virtually all junctions to be broken by elastic recovery, except in the case of metals having very low elastic modulus. However, the quantitative reasons for these marked effects of lattice structure on adhesion are not at all clear and warrant considerable further study.

Let us now consider adhesion phenomena for contacting solids of two different materials. Here, almost no understanding or information exists except for metals. Pursuing our analogy between adhesion and cohesion, the adhesion of unlike pairs of metals may be related to their well known alloying characteristics. However, alloys have no physical counterpart of the interface between crystal lattices of two different metals (at least when they are soluble in each other) as found at points of actual contact between chemically clean surfaces of unlike pairs of metals (i.e., grain boundaries between crystals of mutually soluble metals do not normally exist internally in solid metals). The reference to solubility suggests, however, that if two different metals will not dissolve in each other to any significant extent (at the temperature used during the adhesion studies) they may exhibit poor adhesion. Ernst and Merchant (ref. 12) first explored this possibility via static friction measurements and found reasonable confirmation. Roach, Goodzeit, and Hunnicut (ref. 21) also explored the concept successfully during sliding friction and wear studies. Table 3 indicates that Sikorski found very low adhesion between silver and iron—an order of magnitude lower than that for any other pair listed there. Iron and silver are highly insoluble, whereas all other pairs are either soluble or form intermetallic compounds.

Buckley (ref. 15), in his study of lattice-structure effect on the adhesion and friction of chemically clean surfaces, also investigated how

TABLE 4.—*Coefficients of Adhesion and Friction for Various Single Crystal Orientations of Copper in Vacuum (10^{-11} torr, 20° C, 50 gm)*

[From ref. 15]

Contacting planes	Adhesion coefficient before sliding	Friction coefficient during sliding [a]	Adhesion coefficient after sliding
(100)/(100)	1.02	>40.0	>130
(110)/(100)	0.25	>40.0	32.5
(111)/(100)	0.20	>40.0	40.0

[a] Sliding velocity 0.001 cm/sec; [110] direction; sliding distance 0.735 cm.

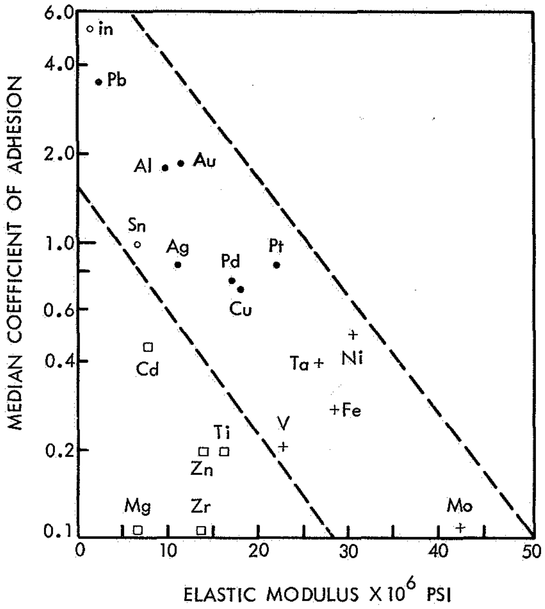


FIGURE 4.—Mean coefficient of adhesion vs modulus of elasticity for like metal pairs; note unusual behavior of metals having hexagonal close-packed crystal structures. Crystal structures: +, B.B.C.; ●, F.C.C.; ○, Tet.; □, H.C.P. (ref. 17).

solid solubility of unlike metals affected their adhesion. The results are given in table 6. The insoluble pair Cu–W shows very low adhesion before sliding and only slight adhesion afterward; the slightly soluble pair Cu–Co (~2 atomic percent Cu soluble in Co near room temperature) also exhibits somewhat low adhesion (this could be due to other factors). Keller (ref. 19) and Spalvins and Keller (ref. 22) also conducted adhesion experiments on chemically clean surfaces of unlike metal pairs in a hard vacuum (5×10^{-11} torr) and correlated the results with known solubility data. Their results are given in table 7. All metals in the "adhesion observed" column form an intermediate phase whereas all those in the second column are immiscible.

In view of the above, knowing how metallic solid solutions and intermetallic compounds are formed may be helpful in developing the theory of adhesion of unlike metals. Hume-Rothery (ref. 23), in developing such a theory, has shown that, when the atomic diameters of solvent and solute (size factor) under equilibrium conditions differ by more than 14 to 15 percent, formation of solid solutions is unlikely, and when the solvent is not significantly electronegative with respect to the solute, or vice versa (electronegativity factor), formation of intermetallic compounds is unlikely. Keller (ref. 19) attempted to apply such a theory in analyzing the adhesion of solid metals, but found that it has severe limitations at present. Not only do equilibrium conditions normally not exist at contacting metal interfaces, but also quantitative understanding of the physics and mechanics of such contact is still lacking. Further, a number of investigators (for example, Machlin and Yankee (ref. 24) find inference of strong adhesion between mutually insoluble metal pairs in studies on sliding friction. Evidently, more research on the quantitative fundamentals of adhesion is needed.

Required Research

Although we now recognize the major factors governing the adhesion of solid surfaces, our knowledge is virtually all qualitative. If we are to understand and control adhesion of solid surfaces, our knowledge must become more quantitative.

The need exists for a quantitative physical-chemical theory that can predict the constitution, arrangement, and properties of surface contamination to be expected from different environmental conditions. Further, the chemistry and mechanics of the reverse process ("decontamination") must be understood and controlled in a practical quantitative manner if adhesion is to be harnessed usefully. The mechanics of disruption, displacement, or removal of surface contaminants require particular attention.

The need for methods for exact quantitative determination of the

TABLE 5.—*Coefficients of Adhesion and Friction for Copper and Cobalt Single Crystals in Vacuum (10^{-11} torr, 50 gm)*

[From ref. 15]

Metal couples, matched poles	Adhesion coefficient before sliding	Friction coefficient during sliding [a]	Adhesion coefficient after sliding
$\frac{\mathrm{Cu}(111)\ [110]}{\mathrm{Cu}(111)\ [110]}$	0. 30	21. 0	10. 5
$\frac{\mathrm{Co}(0001)\ [11\bar{2}0]}{\mathrm{Co}(0001)\ [11\bar{2}0]}$	$<$0. 05	0. 35	$<$0. 05

[a] Sliding velocity 0.001 cm/sec; sliding distance 0.735 cm.

TABLE 6.—*Coefficients of Adhesion and Friction for Various Single-Crystal Metal Couples in Vacuum (10^{-11} torr, 50 gm)*

[From ref. 15]

Metal couples and orientations	Adhesion before sliding	Friction coefficient during sliding [a]	Adhesion coefficient after sliding	Soluble	Crystal structures
Cu(111) Cu(111)	0. 30	21. 0	10. 5	Yes	F.C.C. F.C.C.
Cu(111) [110] Ni(111) [110]	0. 25	4. 0	2. 0	Yes	F.C.C. F.C.C.
Cu(111) [110] Co(0001) [11$\bar{2}$0]	0. 10	2. 00	0. 5	Slightly	F.C.C. Hex
Cu(111) [110] W(110) [111]	<0. 05	1. 40	0. 5	No	F.C.C. F.C.C.

[a] Sliding velocity 0.001 cm/sec; sliding distance 0.735 cm.

magnitude and condition of the actual area of contact between surfaces of solids under load has been stated already. The microscopic and submicroscopic details of such contact requires further study and experimentation. The goal of the study would be a more detailed and quantitative theory of the mechanics of such contact, including the exact role played by plastic and elastic deformations.

Such study would lead naturally into research on the quantitative physics and mechanics of the various elastic recovery events between adhering areas of contact of solid bodies. The semiquantitative theory of Bowden and Tabor (ref. 13) provides at least qualitative understanding of the behavior of the adhesion coefficient for various geometries, materials, lattice structures, and crystallographic planes, and indicates that work to develop a more quantitative theory could be quite fruitful.

More quantitative understanding of how tangential stress or sliding affects adhesion must be gained in respect to junction growth, dealt with in the second part of this paper, and mechanical disruption displacement, or removal of contaminating films, dealt with above.

Research on the quantitative atomic, molecular, and thermodynamic fundamentals of adhesion requires that the present theories of chemical bonding and cohesion be extended to interfacial phenomena such as mismatched lattices and "one-dimensional" solid solutions, and that the nonequilibrium solid-state theory be understood more deeply.

That such improved quantitative understanding and control of adhesion phenomena as outlined above can have far-reaching economic and technical benefits in man's continuing conquest of the physical world will be illustrated by an example from the industrial manufacture of mechanical goods. Most goods today are manufactured by the process of precision material removal—machining. This process removes material from a crude piece of stock to leave behind a precision-shaped useful part—much as a sculptor creates a lovely figure from a crude piece of stone. Manufacturing by this method is, however, most wasteful of material; sometimes as much as 90 percent of the original material is removed and discarded in the process. In time, as relative and absolute material costs rise, such waste may become intolerable. The ultimate method of manufacture in the distant future will be "consolidative processing"—the manufacture of goods (mechanical and otherwise) by assembling material in a controlled, precise manner. Here, controlled adhesion, rather than cutting or dissolution, will be the basic manufacturing process. Crude forerunners of the future commercial processes exist today: electroforming, powder metallurgy, friction welding, pressure welding, diffusion bonding, and adhesive bonding. However, development of versatile, economic consolidative processing for wide application depends on increased understanding and control of adhesion.

FRICTION

Coulomb Friction

Early theories and laws of friction were largely geometrical and mechanical in nature, as might be expected in the absence of detailed understanding of physical, chemical, and mechanical fundamentals. Leonardo da Vinci (ref. 25) stated in his notebook that frictional resistance is doubled when load is doubled and illustrated by sketches and statements the fact that the same force is required to pull a brick-like block along a horizontal plane whether it lies on its back or side

TABLE 7.—*Adhesion Characteristics of Various Metal Couples in Vacuum*

[From ref. 22]

Adhesion observed	No adhesion observed
Iron-aluminum	Copper-molybdenum
Copper-silver	Silver-molybdenum
Nickel-copper	Silver-iron
Nickel-molybdenum	Silver-nickel

or stands on end. He also had a clear idea of the repose angle and a general concept of the friction coefficient, which he considered to have a value of ¼ for smooth surfaces. Such insight was, of course, lost to the world for many years following da Vinci's lifetime. Amontons (ref. 26), in 1699, is thus credited with conducting the first professionally recorded experiments on friction and formulating laws from them. He concluded that the force of friction is (1) proportional to the load with which the upper surfaces presses against the lower and (2) independent of the area of contact. The laws are recognized as generally valid to this day. Coulomb (ref. 27), in 1785, presented further facts about the behavior of friction and also proposed a theory of friction. While he considered that part of the frictional force might be ascribed to cohesion of the molecules at the sliding surfaces, he concluded that "the friction can come only from the engagement of the asperities of the surfaces, and the cohesion should influence it only a little." This supposed mechanism of friction ("interlocking" of asperities) is still known as Coulomb friction in applied mechanics.

The Coulomb theory of friction based on the engagement of surface asperities was developed later to the quantitative idea that, if the surface asperities have a maximum angle of slope α, then the coefficient of friction $f=\tan \alpha$. With time, however, the Coulomb theory was found invalid because it could not explain the loss of energy in sliding; the mass being drawn along is not continually ascending the slope α—it must rise and fall. Thus, the net energy expended will be zero. Further, the usual maximum slope of surface asperities is far too small to explain the usually observed values of the friction coefficient. Also, the theory cannot explain the effect of boundary lubricants.

Basic Mechanism of Friction

Is there a mechanism that explains Amontons' generally valid laws of friction consistent with the foregoing facts as well as other well known features of the friction process? Armed with the knowledge of the first part of this paper, the answer literally "falls out." We have already seen that, when the surfaces of two solids are brought into contact under load, the area of actual contact is proportional to the load and follows equation 1. We have also seen that some adhesion occurs between all contacting solid surfaces, irrespective of their composition; this applies as well to surfaces of contaminating oxide films on the parent surfaces. If, then, sliding is to take place, it must shear the weakest (tangential) planes at the areas of actual contact. This action requires a finite force—the force of friction F. If the mean shear strength of the weakest planes at the contact areas is s and their total area is A, we may write

$$F=As. \tag{2}$$

Substituting the value of A, which we know from equation 1, into equation 2, we obtain

$$F=\frac{s}{p}W. \tag{3}$$

Thus, the friction force is proportional to the load and independent of the area of contact, in agreement with Amontons' laws. Solving equation 3 for the ratio F/W, which, by definition, is the friction coefficient f, we obtain

$$f=\frac{s}{p}. \tag{4}$$

Thus, the coefficient of friction turns out to be the ratio of the mean shear strength of the weakest tangential planes at the areas of actual contact to the mean yield pressure or indentation hardness (mean pressure hardness) of the softer of the two contacting bodies.

Merchant (ref. 28) and Bowden and Tabor (refs. 29 and 30) both arrived at these concepts independently and almost simultaneously. Bowden and Tabor also introduced the concept of additional frictional resistance P caused by the asperities of the harder surface "plowing" through the softer. Thus, equation (2) becomes

$$F=As+P \tag{5}$$

in the case where the plowing force relative to the shearing force is not negligible. To estimate the magnitude to be expected for this force for usual surfaces, we may idealize the asperities of the harder surface to a conical shape of semiapex angle β, as in figure 5, and assume no adhesion. Bowden and Tabor (ref. 31) point out that, if such an asperity plows its way through the softer surface, the plowing force and resulting "friction coefficient" may be readily estimated from the forces required for plastic flow of the softer metal. The vertical load W is, of course, balanced by the yield pressure of the metal acting via contact area A_1. If the width of groove is d, then $W=A_1p=\frac{1}{8}\pi d^2p$. The resistance F to horizontal motion is balanced by the yield pressure of the metal acting over the cross-sectional area of groove A_2. Then, $F=A_2p=\frac{1}{4}d^2p \cot \beta$. Thus,

$$f=\frac{F}{W}=\frac{2}{\pi}\cot \beta. \tag{6}$$

Normal surface asperities seldom have an effective slope of more than 5° or 6°. If we then take a value of 85° for the semiapex angle, we find from equation 6 that the "friction coefficient" caused by plowing will be less than 0.1 in the usual case. With rounded asperities of large

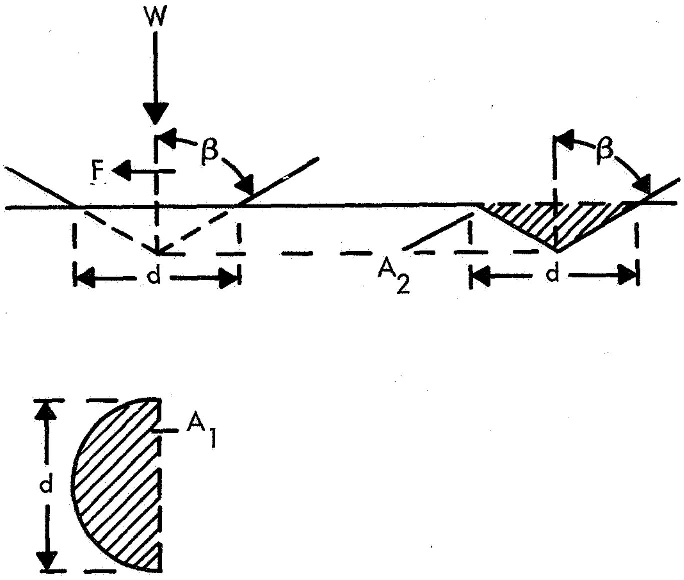


FIGURE 5.—Diagram showing the grooving of a soft metal when a hard conical indenter slides over it. If interfacial adhesion is negligible and no metal pile-up occurs, component A_1 of the contract area supports the normal load W, and component A_2 resists tangential displacement with a force F. In this simple model, $W=pA_1$ and $F=pA^2$, where p is the yield pressure of the softer metal. The effective coefficient of friction resulting from the grooving mechanism is $f=F/W=A_2/A_1$ (ref. 31).

radius combined with surfaces of nearly equal hardness, considerably smaller values can be expected.

How do these values compare with usual values of s/p (equation 4) for the case where adhesion really exists? If chemically clean surfaces of identical materials are brought into contact, the shear strength s of the junctions, in the light of our understanding of adhesion developed in the first part of this paper, should be virtually that of the bulk material. The yield pressure p corresponds to the indentation hardness pressure, as indicated earlier. For metals not work-hardened, the shear strength s of the interface is approximately equal to the critical shear stress τ of the metal, which Tabor (ref. 32) has shown to be approximately 1/5 the harness p. (If the surface is work-hardened superficially, higher values may be postulated.) Thus, for clean metal surfaces, friction from plowing is small in relation to friction from adhesion. However, if the surfaces are covered with low-display shear strength contaminants (such as adsorbed organic films), are dissimilar in hardness, and are somewhat jagged, it may be quite possible that plowing is the major contributor to the observed friction.

Junction Growth

The foregoing suggestion that the friction coefficient for clean metal surfaces may be only of the order of $\frac{1}{5}$ may well have utterly shattered the reader's confidence in the preceding analysis; values of f for clean metals more than 100 times higher can be found in tables 2, 4, 5, and 6. Let us, therefore, hasten to add that $\frac{1}{5}$ is only an estimate of the lower friction-coefficient limit for pure metals. May we then call the reader's attention to the fact that we warned him that equation 1 may be expected to fail whenever tangential stress or sliding were introduced at the contacting asperities because of an action identified as junction growth, which we planned to explain in this part of the paper.

Junction growth under tangential stress has been carefully studied by Courtney-Pratt and Eisner (ref. 33), among others, and analyzed by McFarlane and Tabor (ref. 34). The former, using optical observations and electrical resistance measurements, examined the area of actual contact between a hemispherical slider and a flat surface as tangential force was applied. The results they obtained for platinum (where oxide contamination is no problem) are shown in figure 6. As can be seen, the contact area between clean surfaces continued to grow indefinitely as tangential stress was increased, whereas junction growth between lubricated surfaces terminated after a certain amount of increase of tangential stress, and slip took place.

McFarlane and Tabor's analysis showed that the junction growth took place because, in accordance with plasticity theory, plastic yielding in the junction is determined by the combined effect of normal and tangential stresses. By analogy with the von Mises (or Tresca-Mohr) yield criterion for an ideal two-dimensional model of an intermetallic junction, the yield criterion may be expected to have the form

$$p^2 + \epsilon s^2 = p_0{}^2 \qquad (7)$$

where ϵ is a constant with a value of about 10, and p_0 is the static contact pressure. The metal in the junction is initially subjected to a normal stress equal to its yield pressure $p = p_0$. Thus, as soon as the slightest tangential stress is applied, equation 7 can be satisfied only if p diminishes. This means that the area of contact must increase with the drop in p, in the manner shown in figure 7, while point Q_0 moves to Q. It appears from figure 6, this process continues almost indefinitely for clean surfaces, so that extremely high friction coefficients are attained at slip. However, as seen, if the surfaces are lubricated

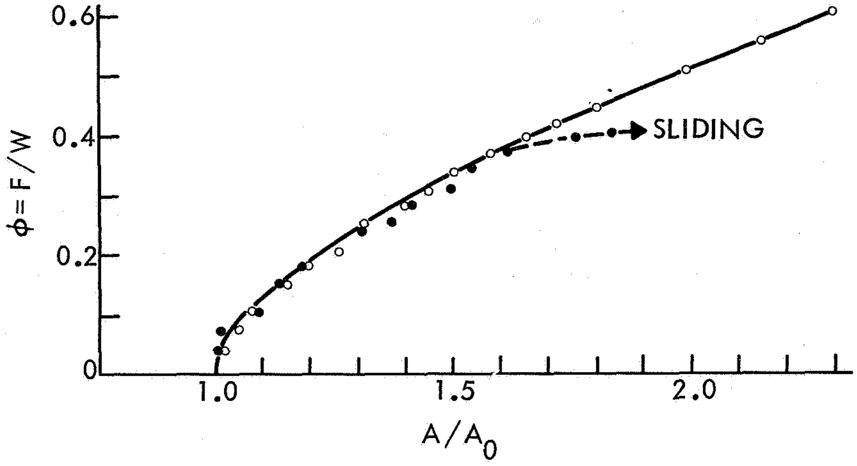
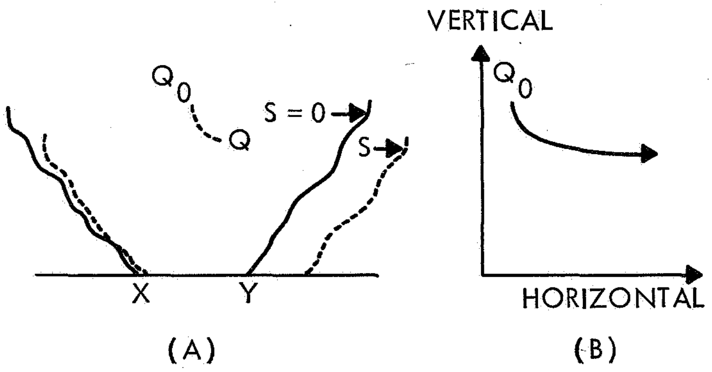


FIGURE 6.—The behavior of contacting platinum surfaces when subjected to a gradually increasing tangential force. Ordinate ϕ is the ratio of the tangential force F to the normal load W. (It corresponds to the coefficient of friction only when sliding actually occurs.) Abscissa is the area of contact A expressed as a ratio of the initial static area A_0. The ratio A/A_0 is calculated directly from the electrical resistance measurements, assuming metallic contact. ○, clean surfaces; ●, lubricated surfaces. The full curve is the theoretical curve plotted according to equation 7 using a value of 12 for ϵ (ref. 33).

(i.e., $s<\tau$), slip occurs at a lower F/W ratio. The qualitative criterion for such slip is that the rate with which the junction size increases becomes less than the rate with which the tangential force increases. Tabor (ref. 35) has analyzed this situation and obtained the result shown in figure 8, which shows the way in which the friction coefficient depends on the strength of the interface. It may be seen that a very slight contamination of the interface, only enough to lower the shear strength to 95 percent of the metal strength, is sufficient to curtail junction growth drastically and limit friction coefficients to values of 1 or less. Thus, for surfaces contaminated normally in air, junction growth may be expected to have only a minor effect on friction. Further, as pointed out to the author by Bisson (ref. 36), junction growth is a rate-sensitive process; at the usual sliding speeds found in mechanical equipment, the short contact time of asperities will severely limit their ability to undergo junction growth. Nevertheless, junction growth is a type of action that, if conditions permit its occurrence, can quickly generate catastrophic scoring of sliding surfaces; in this role, it may have particular practical significance.

Surface Contamination and Thin Films

Since an entire lecture of this symposium is devoted to boundary lubrication, we will here only briefly relate the action of contaminants and nonliquid thin films to the foregoing theory. We have already intimated that, if the interface contains a contaminant that is weaker in shear than either of the contacting bodies, the value of s in equation 4 should be the shear strength of that film. On the other hand, since the film is (normally) very thin, it will not influence the yield pressure p of the contacting solid surfaces, so that the friction coefficient becomes the ratio of the film's shear strength to the yield pressure of the softer of the two contacting solids. This has been successfully illustrated by Bowden and Tabor (ref. 37), who covered the surface of hard metals with a thin film of very soft metal, such as lead, tin, or indium. For example, they observed a friction coefficient of 0.8 for a steel ball sliding on a flat steel plate in the absence of any film. When the flat surface was covered with a thin indium film (10^{-3} cm thick), the friction fell to a value of 0.1. Such thin films are, of course, not highly durable under repeated sliding, but for single applications they can provide a useful method for reducing friction. This effect can be made more permanent by incorporating "particles" of a soft material in a harder material as exemplified by the dispersion of lead in copper in copper lead-bearing alloys. Here, a thin film of lead spreads over the surface of the copper, and replenishment is



(A) (B)

FIGURE 7.—Effect of combined normal and tangential stresses on the behavior of a soft asperity in contact with a hard flat surface. (a) Under a normal load alone, contact occurs across XY; when a tangential stress s is applied, the junction grows as shown by the dotted line and point Q_0 moves to Q. (b) Schematic diagram showing locus of Q as a tangential stress is increased (ref. 13).

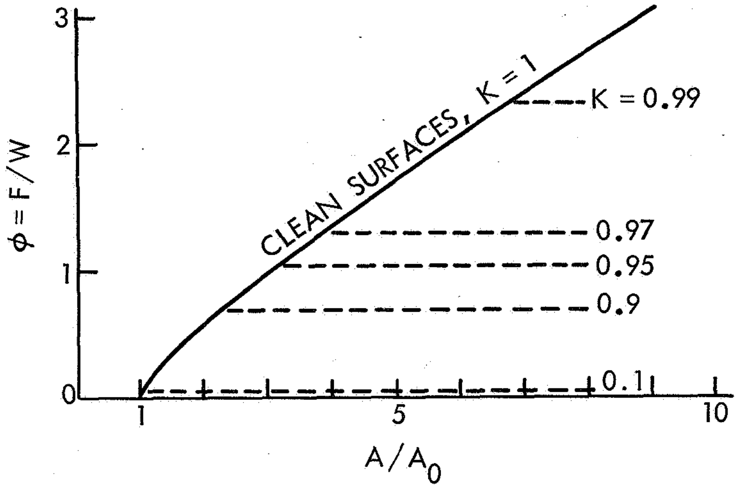


FIGURE 8.—Growth in area of junctions for different values of the interfacial shear strength s, where s is expressed in terms of the critical shear strength τ of the metal ($s=K\tau$).As the tangential force coefficient ϕ ($=F/W$ is increased, junction growth proceeds until at some critical stage, determined by the value of k, gross slip occurs. For $K=1$, junction growth proceeds indefinitely; for $k=0.95$, slip occurs when the area of contact has grown about threefold and the force coefficient (now equal to the coefficient of friction) has a value of about 1 (ref. 35).

continually available at the surface from the fine lead particles dispersed below it.

We have also already intimated, in connection with the discussion of interface shear strength s, that the shear strength may result from a mixture of different material strengths. This, of course, is to be expected, and various studies, such as contact resistance measurements on metals, suggest that this is so. To make this effect quantitative, let us take the simple case of an intermittent film of a low-shear-strength material on otherwise clean solid surfaces of two bodies of identical material. Under these conditions, the two host surfaces will touch and unite over that fraction ρ of the area of actual contact where no film is present on either. This fraction of the contact area will have a shear strength s_1 equal to that of the host bodies. The remaining fraction (1-ρ) of the contact area will have a shear strength s_2 equal to that of the low strength film. The friction force will therefore be

$$F=A[\rho s_1+(1-\rho)s_2]. \qquad (8)$$

Generalization of this approach can be helpful in visualizing, qualitatively, the net effect on friction of complex surface films and layers of contaminants at frictional interfaces.

Rolling Friction

We will not attempt a detailed discussion of rolling friction, which could be a subject for a full review in itself. However, it is important that we briefly consider the basic mechanisms involved in rolling friction and their relation to sliding friction. Early considerations of rolling friction attributed its causes mainly to the sliding friction that occurs over the area of apparent contact between the bodies as a result of the differential stretching of the two surfaces as they pass through that area. However, more recent studies by such investigators as Tabor (ref. 38), Flom and Bueche (ref. 39), and Merwin and Johnson (ref. 40) have shown that energy dissipated by subsurface deformation may constitute the major part of the rolling resistance even for hard metals, except where, as in the case of the latter, very close initial conformity existed between the roller and the mating member (e.g., a rolling ball and a grooved track).

The friction resulting from interfacial slip can be analyzed directly from the principles of friction discussed earlier in this paper, together with appropriate analysis of the differential stretching of the surfaces using Hertzian elastic contact theory. The friction loss caused by subsurface deformation involves energy dissipation from elastic (hysteresis or "internal friction") and plastic deformation. In the case of rubber and rigid polymers, which have relatively high hysteresis-loss characteristics, the energy dissipated by subsurface deformation provides almost all of the observed rolling friction (ref. 38).

Frictional Behavior of Various Materials

The frictional behavior of ductile metals has been referred to frequently in illustrating the foregoing theories to give a reasonable insight into these materials and the manner in which this behavior is kept within reasonable bounds by surface films or lubricants. How poor that inherent behavior is in the absence of surface contaminants becomes evident, for example, from Buckley's friction data in table 2. The problem of operating metallic mechanical devices involving sliding in an outer-space environment, for example, is truly challenging.

The reasons for the poor behavior of ductile metals is now clear from the principles developed above. Figure 8, in particular, demonstrates that avoidance of intolerably high friction and gross seizure requires the interfacial shear strength s to be significantly lower than the shear strength of the weaker metal of the pair or, perhaps, than the yield pressure p of the weaker metal. Evidence for the latter consideration is found in Buckley's data of tables 2, 5, and 6. It must be kept in mind that the yield pressure p associated with the "squashing" together of the points of actual contact derives from three-dimensional plastic flow is not highly sensitive to crystal orientation. On the other hand, shearing the junctions is virtually a one-dimensional plastic flow process and thus very sensitive to the properties of crystallographic planes and directions oriented parallel to the direction of shear.

From the tables mentioned, it can be seen that, as the shear strength of the interface is lowered relative to p by introducing crystallographic planes successively weaker in shear (resulting from increasing interplanar spacing, tables 2 and 5; lattice mismatch, table 6; or lattice incompatibility caused by insolubility, table 6), junction growth is effectively limited and the friction coefficient attains a finite and often surprisingly small value. The case of the behavior of the hexagonal cobalt lattice (table 5) is particularly striking. Ernst and Merchant (ref. 12) obtained evidence early that the interfacial shear strength s of the hexagonal metals, even if polycrystalline, is inherently much lower than their yield pressure p.

The good correlation between friction coefficients and adhesion coefficients noticeable in tables 2, 5, and 6 also supports the above reasoning. As noted in the section on elastic recovery and the adhesion coefficient, the tensile stress required to sever contact junctions or "asperity bridges" is, for ductile materials such as metals, approximately equal to the yield pressure of p of the weaker material. This will only be true, however, when the tensile strength of the junctions produced by adhesion of the original interfaces is at least as great as the yield pressure p of the weaker material. When it is lower, the junctions can fail primarily by "brittle" tensile fracture rather than by prolonged plastic flow in tension. Thus, as the tensile strength of the interface is lowered by introducing into the interface crystallographic planes having successively larger interplanar spacing (tables 2 and 5), lattice mismatch (table 6), or lattice incompatibility caused by insolubility (table 6), the adhesion coefficient will decrease accordingly. Since these same factors also reduce the shear strength of the interface, as noted above, the drop in the friction coefficient parallels that of the adhesion coefficient. Further, the ratio of the interfacial strength (either in shear or in tension) to the yield pressure p of the weaker metal is a very important criterion of the inherent friction and adhesion characteristics of metal pairs; the lower this ratio, the lower the inherent friction and adhesion of the pair.

In respect to brittle materials, such as ionic crystals, one might question whether their frictional behavior would conform to the foregoing theory. It might be expected that the area of actual contact and resistance to shear is conditioned largely by the fracture criteria for these materials, so that equation 4, considerations of junction growth, for example, are not applicable. However, we have seen in the section on adhesion that full adhesion between faces of two rock-salt slabs could be obtained by compressing them between anvils, thus demonstrating capability for both adhesion and plastic flow. Actually even Bridgman (ref. 41) and others have shown that most brittle materials exhibit appreciable ductility when subjected to great compressive stresses or hydrostatic pressure.

Bowden and Tabor (ref. 13), therefore, investigated the frictional behavior of clean rock salt as a typical ionic brittle material. They found that, although there was some fragmentation and cracking of the rock-salt surface along the friction track, the track on a gross scale had the typical grooved appearance characteristic of metals. The area of contact between a hemispherical steel slider and the crystal was found generated by plastic flow of the rock salt; p was of the order of 30 kg/mm^2. Further, the shear strength s of the rock salt in sliding, instead of corresponding to the usual fracture shear strength of about 2 kg/mm^2, was of the order of 5 to 7 kg/mm^2, corresponding to the critical shear stress in plastic compression. The possibility of junction growth was also investigated; evidence was found for growth by a factor of approximately 2 under the action of tangential stress. Further growth is prevented by reducing the compressive stress to the point where brittle fracture occurs, and thereby automatically reducing the junction size to a point where the factor is less than 2. Thus, the friction coefficient does not exceed a value of approximately 1 even for rigorously cleaned surfaces. A variety of other brittle materials studied by Bowden and Tabor exhibited frictional behavior similar to rock salt, thus indicating that the friction of brittle materials, just as ductile metals, is governed by the laws set forth in the earlier sections of this paper.

Let us now consider the frictional behavior of polymers, materials in which the cohesive bonding is primarily covalent. Values of the friction coefficient for these materials generally range between 0.2 and 0.7 (ref. 42), not greatly different from those for clean ionic crystals or clean metal in the absence of junction growth. No significant junction growth is found in frictional processes on polymers. Further, the adhesion at the interface is generally strong, and the relationship $F=As$ holds well (ref. 43), where s is approximately equal to the bulk shear strength of the polymer in most cases. However, with polymers, Amontons' first law (friction force proportional to load) is often found to be a poor approximation. According to studies made by Adams (ref. 9), a small amount of this nonlinear variation of friction with load appears to result from increased shear strength s of the interface with pressure. However, by far the largest portion results from the fact that equation 1 is a poor approximation and the area of actual contact

A is not directly proportional to load W. The relationship is instead very nearly of the form

$$A = kW^m \tag{9}$$

where m is generally between 0.7 and 0.8 (ref. 13). This suggests that elastic deformation (corresponding to $m=0.67$) plays a much more important role in controlling area of contact for polymers than for ionic or metallic materials. Archard's analysis (ref. 3) assumes particular importance here.

As might be assumed from our discussion of rolling friction, internal hysteresis can play a large role in sliding friction of polymers when the adhesion component of friction is small. Further, because polymers are viscoelastic, area of actual contact will increase with loading time and, therefore, friction will vary with speed. Friction of polymers will also vary markedly (as contrasted with metals and ionic solids) with speed from rapid interfacial temperature rise caused by the poor heat conductivity of polymers. The rise in temperature influences shear strength, yield pressure, and viscoelastic properties of the polymer in a complex way, making it difficult to predict variation of friction with speed.

Needed Research

It is evident from the foregoing that present knowledge concerning friction is considerably more quantitative and advanced than that concerning adhesion. It is equally evident that there is much still to be learned for predicting and controlling frictional phenomena.

Concerning friction theory and the basic equations 2 and 4, the real nature of the shear strength s of interfaces and its quantitative relation to the physical properties and yield pressure p of materials still requires investigation. Since shear strength and adhesion strength are not at all the same, yet resistance to shear arises from adhesion, more fundamental work is needed to relate the two quantitatively.

Because of the very important contribution of junction growth in bringing about major departures from equation 4, as well as its role with ductile materials, it needs to be studied more thoroughly. Experimental and analytical study could profitably concentrate on providing a sound basis for quantitative prediction of the constant ϵ in equation 7 as well as on refining this yield criterion if necessary.

To understand surface film effects, a quantitative basis for predicting the shear strength of surface films is needed. Perhaps the somewhat naive thermodynamic approach originally proposed by Ernst and Merchant (ref. 12) may be refined to yield a workable approach.

Concerning the frictional characteristics of various materials, metals still play a leading role in today's manmade world and will for a long time to come. Thus, continuing work on refining the analysis of the friction between metals is justified. In particular, experimental and analytical investigation of the shear strength characteristics of contacting interfaces of different crystallographic planes of identical and different materials (soluble, compound-forming, and insoluble) could prove most helpful in advancing our fundamental understanding of interface shear phenomena.

Because of the increasing importance of polymers in today's and tomorrow's world and our lack of good quantitative understanding of their frictional behavior, analytical research on their friction characteristics should be emphasized.

CONCLUSION

Much progress in understanding the fundamentals of adhesion and friction between solids has been made in this century, and the rate at which that understanding has increased has been exponential. However, in view of the rapidly growing importance to human activity of understanding and controlling adhesion and friction in new environments, the challenge of and need for research in these fields now seems to be growing even faster than our rate of generating new knowledge. Further, as this challenge and need grows, it involves problems from an ever-widening range of scientific and technical disciplines as well as need for the application of knowledge from more and more different disciplines to their solution. Let us hope, therefore, that the interdisciplinary approach fostered by this symposium may be the catalyst that will generate a systems approach equal to the challenge and need.

R. D. Gretz (Battelle Memorial Institute, Columbus, Ohio)

Dr. Merchant's paper evidently constitutes an important contribution to the boundary separating research and development in the field of friction and wear. It is obvious from Dr. Merchant's paper that a proper understanding of friction can only be achieved by assimilating the concepts of basic science into the vast expanse of empirical knowledge presently available; however, a consistent operator does not exist. If this operator existed, the matrix of empirical behavioral patterns might conceivably be transformable into a coherent self-consistent law; this law would allow us to choose a system, select the operational variables that must be monitored, and predict the outcome of the test: this cannot be done.

Considering the omnipresence of friction (in the sense of the wide range of materials, the range of surface topology scale, and the complex chemistry of bounding media involved), an arbitrary and rather limited approach in any particular discussion must be chosen. At the expense of, most likely, your curt dismissal of some things I will discuss, still I will insist on discussing well-defined surfaces (cleavage planes or atomically flat perfect crystals) that may be determined to be clean on an atomic scale. However, as it suits my needs, I will also talk about surfaces that have a macroscopic contour; this will be assumed to be periodic.

The coefficient of friction.—The force required to overcome friction is the product of the coefficient of friction and the normal force acting on the interface between the two bodies. This expression may be generalized to the extent of noting that the shear stress required to overcome the frictional force at the interface is the product of the coefficient of friction and the normal stress acting on the interface (ref. 44). If we consider a perfect crystal, the stress required to separate it normal to a cleavage plane is the theoretical breaking stress (ref. 45)

$$\sigma_{\max}=\left(\frac{E\gamma}{a_0}\right)^{1/2} \tag{10}$$

assuming that the only work done is that against surface energy γ with a material whose Young's modulus is E. The lattice parameter normal to the interface is a_0. If the ruptured material is reassembled so that all lattice planes register on an atomic scale, the normal stress across the interface separating the two halves of the crystal is again given by equation 10. If we slide these two crystal halves parallel to each other, the shear stress required is, in fact, the theoretical shear stress of a perfect crystal, which may be given as (ref. 45)

$$\tau_{\max}=\frac{G}{2\pi} \tag{11}$$

where G is the shear modulus of the material. The coefficient of friction, then, for this extremely ideal case is

$$\mu_i=\frac{\tau_{\max}}{\sigma_{\max}}=\frac{G}{2\pi}\left(\frac{a_0}{E\gamma}\right)^{1/2} \tag{12}$$

Using reasonable values in equations 10 and 11, the coefficient of friction for copper according to equation 12 is 0.05. While the coefficient of friction is attractively small, the shear stress required to produce parallel movement of the crystal halves (equation 11) is 1.2×10^{10} dynes/cm^2, a stress so high that we would say seizure had occurred. Evidently, a coefficient of friction of 0.05, bought at the expense of a shear stress of 180 000 psi, is not an attractive option, even if it were exercisable.

Obviously, equation 1 is wrong because in our thought experiment we did not actually create any new surface at the sliding interface.

Therefore, the stress must be just the load W divided by the contact area

$$\sigma=\frac{W}{A}. \tag{13}$$

That equation 11 may still be applicable can be demonstrated by the following thought experiment. Let us place together two perfect blocks of the same material with identical truncated sinusoidal surface contours. The area of contact is known (the area of the perfectly flat-topped surfaces of the surface profile). Under these test conditions, the materials are bonded at an atomic level so that when we attempt to slide the blocks relative to each other, the force required is the shear stress of the perfect crystal times the contact area

$$f=A\tau_{\max}. \tag{14}$$

Since we know the applied load W and the contact area A, we may use equation 13 to calculate the stress normal to the interface. Alternatively, we may rearrange equation 13 to give

$$A=\frac{W}{\sigma} \tag{15}$$

to arrive at the result

$$f=A\tau_{max}=W\frac{\tau_{max}}{\sigma}=\mu_i W \tag{16}$$

and recover a coefficient of friction

$$\mu_i=\tau_{max}/\sigma. \tag{17}$$

Note, however, that according to equation 13, the σ in equation 17 has a very special meaning; it is the load on the interface divided by the area of contact at the interface. In our thought experiment, we have chosen the contact area to be an invariant. Since we are dealing with a perfect material, we can vary the load without changing the area of contact, that is, we can independently change the stress applied at the interface. If we would choose an applied stress of 1800 psi (e.g., for copper), the coefficient of friction would be 100 and decrease with increasing load.

Let us take another look at equations 16 and 17

$$\mu_i=\frac{\tau_{max}}{\sigma}=\frac{f}{W}. \tag{18}$$

We know that f/W must be a constant, all other things being equal. This implies that $\tau_{\max}/\sigma$ is also constant, yet we have just made the claim that we could independently choose σ for a fixed value of $\tau_{\max}$. Both these statements cannot be true. The dichotomy becomes evanescent when we introduce equations 13 and 14 into the first term of the right-hand side of equation 18 and recover f/W. We conclude, therefore, that we must be careful as to how equations 14 and 15 are used. The stress appearing in equation 15 is a dependent variable (see equation 13) and is not determinable unless we know A and W. Note also that equation 14 is not really the frictional force, it is just the force required to shear the junctions between the crystal halves; this being the case, we should have used equation 11 and the coefficient of friction would indeed have been given by equation 12. In this case we are shearing junctions and creating new surface area, just as we prescribed in our first thought experiment; however, by some means we have managed to convince ourselves that we could have a coefficient of friction (μ_i) that varies by a factor of 2000, depending. But we really have not been talking about friction.

The force given in equation 14 is the force required to shear the material joining the two crystal blocks. The frictional force F on the other hand, by definition (ref. 44), is the force required to slide two materials by each other against an applied load W

$$F=\mu W. \tag{19}$$

The load W is the independent variable, and the frictional force seen at the interface F is a fraction of the normal load on the interface, the coefficient of friction μ. According to this definition, then, the coefficient of friction is a measure of the work that must be done against the normal force and, in this sense, is restricted to processes such as plowing (pushing material along in the interface) or lifting (raising a block to the peak of an asperity). In both plowing and lifting it is easy to visualize that work must be done against a normal force.

It must be recalled that, since a shear force or stress does not and cannot do any work against a normal force or stress (ref. 46), a normal force or stress cannot produce a shear force or stress. We must conclude, therefore, that in the absence of plowing or lifting (or similar boundary conditions that require work against a normal force), a coefficient of friction does not exist. Interposing a two-atom layer of oxide between two perfect crystal blocks will, we assume, result in rupture of the oxide, requiring a force sufficient to meet the required shear stress of the oxide layer, when the blocks are moved. The value of the shear stress is independent of the applied normal stress, so that the frictional force is independent of the applied load. A coefficient of friction does not exist, therefore, unless work is done against normal forces.

Let us consider two blocks, one resting on the other, and visualize the contacting surfaces as being typical engineering surfaces with some asperities and also, presumably, some oxide or other phases on the surfaces. The total force $\mathcal{F}$ required to slide one block on the other will be the sum of any force needed for shear processes (f). We will generalize equation 14 to the extent of considering an arbitrary $\tau (f=A\tau)$ and a frictional force $\mathcal{F}$ needed to do work against normal forces ($F=\mu W$). We will visualize these processes as either plowing or lifting), thus

$$\mathcal{F}=f+F. \tag{20}$$

Inserting the definitions for f and F

$$\mathcal{F}=A\tau+\mu W. \tag{21}$$

We may rearrange this to give

$$\mathcal{F}'=\tau+\mu\sigma \tag{22}$$

by dividing by A. Thus, we have the necessary force per unit area of contact to effect sliding ($\mathcal{F}'$, let us call this a specific force) and a specific stress (σ). We may argue about the details of choosing A, but to a first-order approximation, this procedure is probably correct. We predict, therefore, that a plot of $\mathcal{F}'$ versus σ should be linear, intercept

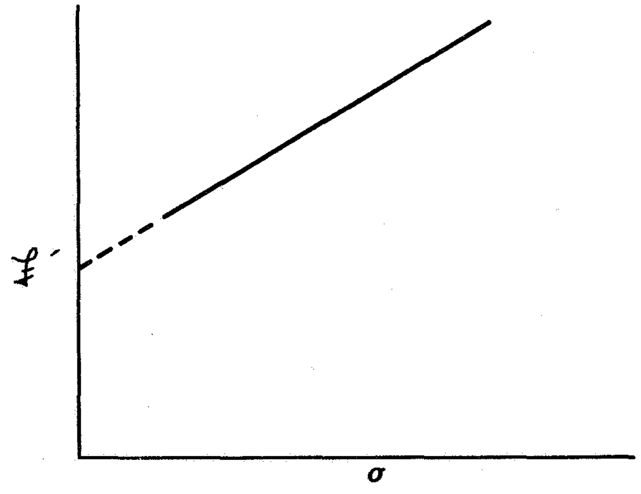


FIGURE 9.—Plot of specific force vs specific stress (ref. 46).

τ, and have a slope μ. The data given by Kragelskii (ref. 46) follow such a law (fig. 9). It is interesting to note that instead of equation 20, it is assumed (ref. 46) that the frictional force F equals the shear force f

$$f=F. \tag{23}$$

Using the same substitutions ($f=A\tau$ and $F=\mu W$) in equation 23, the following result is obtained with the aid of equation 13:

$$\mu=\tau/\sigma. \tag{24}$$

It is concluded (ref. 46), therefore, that the coefficient of friction decreases with increasing load and tends to a limiting value. Analysis of the same data with the help of equations 20 through 22 shows that μ is a constant instead; the slope of the plot is the coefficient of friction. We know that equation 23 is incorrect because it equates a shear force (by definition independent of the normal load) to a frictional force (by definition dependent only on the normal load). In fact, the process of sliding a block over the surface of another block involves both a shearing process and a frictional process.

We conclude, therefore, that friction and adhesion are independent processes. They may occur simultaneously in the same system, but they are separate and identifiable: see equation 22. Returning to our original question, it is now obvious that a coefficient of friction does not exist in sliding atomically perfect surfaces parallel to each other. A force will be needed, but that force will be simply the force needed to shear atomic bonds at the surface. Thus, for an atomically flat surface

$$\mathcal{F}=f. \tag{25}$$

For a surface that has a surface profile, but whose surface bonds are saturated by adsorbed gases (taking the limit where no measurable shear process occurs)

$$\mathcal{F}=F \tag{26}$$

equation 26, evidently, summarizes the case often observed for typically "clean hard" surfaces.

Returning now to equation 22, let us perform a thought experiment where only lifting contributes to the frictional force F. We exclude any plowing of asperities. Suppose we do not impose a compressive stress at the interface (σ) but rather a tensile stress ($\sigma'=-\sigma$) and, further, require that we have chosen a system where $\mathcal{F}'$ is zero

$$\mathcal{F}'=\tau+\mu\sigma'=0 \tag{27}$$

then

$$\mu=\frac{\tau}{\sigma}. \tag{28}$$

In this situation we have the tensile stress working against the shear stress; we have a perfect bearing. We have chosen the conditions of the system so that the frictional force operates to our advantage—it supplies the work for the shear processes needed to move the opposing blocks.

The interfacial free energy.—In the preceding section we have discussed bonding between perfect surfaces and surfaces with a macroscopically measurable surface topography. It should be noted that the full effect of surface bonds will only exist between atomically clean surfaces. If the surface is contaminated by adsorbed gases, the surface forces are effectively masked, and binding between atomic planes is greatly reduced. We know that the intrinsic forces between atomically clean surfaces are quite large. A measure of the level of surface forces is afforded by the activational free energy required for desorption of an adsorbed atom from the surface; this energy is known to decrease markedly with adsorption even to the monolayer level (ref. 47).

To illustrate the masking effect, we consider the ideal case of the (100) surface of a simple cubic structure whose total energy is the sum of the energies contained in the nearest neighbor bonds. If φ is the energy required to break a nearest neighbor bond, and if q is the number of atoms per square centimeter of the (100) face, then the surface energy is (ref. 48)

$$\gamma=\frac{\varphi q}{2} \tag{29}$$

If each surface bond is of strength φ, the surface energy of the lattice plane is given by equation 29, in the nearest neighbor approximation. If each of these bonds are then satisfied, the surface energy of that lattice plane will be reduced effectively to zero; that is, the change in the surface energy of a lattice plane resulting from equilibrium adsorption of gases is reduced to that value given by the Gibbs adsorption equation.

The Gibbs adsorption equation (ref. 49) is

$$d\gamma=-S^{(s)}dT-\sum_i \Gamma_i d\mu_i \tag{30}$$

where $S^{(s)}$ is the entropy of the surface, μ_i is the chemical potential of specie i, Γ_i the surface density of the ith substance, and γ the surface tension. Consider a solid surface (component 1) in equilibrium with a perfect gas (component 2) and position the dividing surface so that

$\Gamma_1=0$. Then the Gibbs adsorption equation becomes, at constant temperature,

$$-d\gamma=\Gamma_2 kT\, d\ln a_2 \tag{31}$$

where a_2 is the activity of the species 2 on the substrate surface. For a monolayer coverage (e.g., a system with a very high desorption energy), the surface density of the second component Γ_2 remains constant, and we may take it to be 10^{15} atoms/cm^2. Choosing $\Gamma_2°$ equal to unity at γ, the surface tension for zero population of adatoms, equation 31 becomes

$$\gamma-\gamma_a=\Gamma_2 kT \ln \Gamma_2 \tag{32}$$

where k is 1.3×10^{-16} ergs/° K. Substituting the pertinent values in equation 32, we have the appreciable change in surface energy of about 1350 ergs/cm^2. If γ for the atomically clean surface was 1350 ergs/cm^2, then γ_a (the surface energy in the presence of monolayer adsorption) would be zero at equilibrium. Thus, small amounts of adsorbed gases will appreciably alter the surface energy of the condensed phase at equilibrium. The latter equilibrium constraint can be taken to mean complete saturation of surface bonds. If bond saturation does not occur within the first monolayer (this is an electronic property of the system and cannot be determined a priori), three or four monolayers may be required to effect saturation. The latter property is not determinable from the simple nearest neighbor model or from a thermodynamic calculation. It is important to realize the consequences of the assumption of thermodynamic equilibrium (equation 32) in the light of equation 29, the nearest neighbor model of a surface. In some systems, however, the thermodynamic assumption of equilibrium seems to be good, as demonstrated by the following example of adsorbed lead or copper.

In terms of the more familiar adsorption notation, $\Gamma_2=n_e$, the adsorbed monomer concentration on the substrate in equilibrium with vapor of the same material. The adsorbed population at equilibrium (ref. 50) is given by the vapor pressure of the solid P_e, and the desorption energy of an adsorbed atom from the surface ΔG_{des},

$$n_e=P_e \exp(\Delta G_{\mathrm{des}}/kT)/(2\pi mkT)^{1/2}v_o \tag{33}$$

where v_o is the vibrational frequency of an adsorbed atom on the surface.

Substituting equation 33 into equation 32 and rearranging terms, the following result is obtained:

$$\ln(\gamma-\gamma_a)\propto\frac{\Delta G_{\mathrm{des}}-\Delta H_s}{kT}. \tag{34}$$

Equation 34 is interesting because it shows that the temperature dependence of the change in solid-vapor interfacial free energy resulting from equilibrium adsorption of atoms of another material is given simply by the difference between the adatom desorption energy, ΔG_{des}, and the bulk heat of evaporation of atoms from its own surface, ΔH_s. Thus, if $\Delta G_{\mathrm{des}} > \Delta H_s$, plots of the form of ln $\Delta\gamma$ versus $1/T$ will have positive slopes; if $\Delta G_{\mathrm{des}} < \Delta H_s$, however, plots of this form will have negative slopes. The data of Bailey and Watkins (ref. 51) for lead vapor adsorbed on solid copper follow equation 34; the resulting positive slope indicates that the desorption energy of lead from copper is 43.8 kcal/mole (fig. 10). This type of calculation affords a new way of obtaining experimental values of ΔG_{des}.

The verification of equation 34 is important because it lends confidence to the equilibrium approach to surface theory. More importantly, equation 34 provides an incisive means of describing the effect of adsorption on surface energy and the temperature dependence of the surface energy. Thus, since heat-of-sublimation energies are readily available, all we need to know is the desorption energy of an atom from a substrate in order to predict whether the solid-vapor interface energy increases or decreases with temperature in an adsorption environment.

Usually, the balance of interfacial free energies between a condensed phase and the substrate on which it is disposed is demonstrated in terms of the contact angle between the condensed phases, the so-called Young equation (ref. 49).

$$\gamma_s = \gamma_{a-s} + \gamma \cos \Theta \tag{35}$$

where γ_s is the crystal-vapor interfacial free energy of the substrate, γ_{a-s} is the interfacial free energy between the adsorbed condensed phase and the substrate on which it resides, γ is the interfacial free energy between the condensed adsorbed phase and its own vapor, and Θ is the contact angle. If the contact angle is zero, perfect wetting is said to have occurred; the condensed liquid phase spreads without limit on the surface of the condensed phase on which it resides. Perfect wetting, known to occur for a number of clean refractory metal surfaces on which gold droplets are melted (ref. 53), is of considerable interest in the analysis of adhesion effects at surfaces.

In the consideration of surface topography, interposition of a perfectly wetting material may have other demonstrable consequences if the particle size is sufficiently small. A line-tension (ref. 54) effect alters the interfacial free energy balance, particularly in a system where the known macroscopic contact angle is zero. This effect arises in the following way: The line-tension interfacial free energy results from the unsatisfaction of surface bonds along a line of atoms, as opposed to a plane of atoms considered in the derivation of equation 29. Invoking an analogous derivation and assuming a surface lattice spacing of a_o, the line tension is then

$$\gamma_t = \frac{\varphi}{2}\frac{1}{a_o} \tag{36}$$

which, upon substituting equation 29 for the right-hand side and, noting that the number of adsorbed atoms per unit area of surface, $q = a_o^{-2}$, becomes

$$\gamma_t = \gamma a_o \tag{37}$$

and the interfacial free energy triangle of Gibbs becomes (ref. 54)

$$\gamma_s = \gamma_{a-s} + \gamma \cos \theta + \frac{\gamma_t}{r \sin \theta} \tag{38}$$

where r is the radius of curvature of a condensed phase whose contact angle is θ. The essence of equation 38 is that if a line tension exists, the actual contact angle between condensed phases for small particle sizes (consider the geometric picture of a liquid droplet on a perfectly flat plane) is increased above the measured macroscopic value. That is, in a system where the known macroscopic contact angle is zero, the line-tension effect may cause the actual contact angle to have a nonzero value (as large as 90°). Thus, in considering wetting effects on a microscale at an interface, the system may differ from that expected on the basis of the known macroscopic behavior. The line-tension effect is new and the consequences of it as yet not generally appreciated. These properties become significant for particles whose radius of curvature is less than about 200Å.

The Effect of Stress on Capillarity.—The preceding analysis has discussed some effects of surface energy on the behavior of surface as well as the creation and destruction of new surface. Since these capillarity processes may be altered by stress, it is of interest to consider the effect of stress on capillarity.

Gurney (ref. 55), following Gibbs (ref. 49), has used thermodynamics to analyze the chemical potential of an atom on the surface of a solid for the general case of a solid in an arbitrary state of elastic stress. Gurney notes that there are two general cases, corresponding to solution and swelling, for describing the effect of elastic strain energy on the solid's chemical potential. It has been shown (refs. 49, 55, and 61) that the analysis is applicable to reversible transport of a substitutional solute (accretion) or an interstitial solute (absorption). The major distinction between accretion and absorption contraints is that in accretion the chemical potential of the atoms on the surface of the solid increases irrespective of the sign (tension or compression) of the applied elastic stress, whereas for absorption the chemical potential of the interstitial solute decreases in tension but increases in compression. Since there is an elastic stress gradient at the surface of an elastically stressed solid in the vicinity of a surface discontinuity, there is also a gradient in the chemical potential. The latter permits highly localized absorption or accretion (in a test environment, this would actually be regarded as dissolution).

Consider a cube of pure solid subjected to a uniform elastic stress on each face by a fluid that may contain the substance of the solid in solution but does not itself dissolve in the solid. The chemical potential of an atom at the surface of the solid for an arbitrary state of stress depends on the face considered.* Assuming that reversible isothermal transport of substance between the $+X$ solid face and an unstressed solid is achieved only if P_x, P_y, P_z, and the dimensions of the solid in the Y and Z directions remain constant,

$$\mu_x = E - TS + P_x\,V \tag{39}$$

where μ** is the chemical potential, E, S, and V are the molar internal energy, entropy, and volume, respectively, of the stressed solid, and T is the temperature. Similar equations can be written for the Y and Z directions. This simply means that the chemical potentials at the three solid faces are different and proportional to the principal stresses P_x, P_y, P_z. Gibbs (ref. 49) next considered the change in chemical potential at the $+X$ face of the hydrostatically stressed solid at pressure $P = P_x$ to be expected upon isothermal application of additional normal stress in the Y and Z directions; thus,

$$\mu_x - \mu^{\circ}_x = \Delta E - T\Delta S + P_x \Delta V \tag{40}$$

$$\Delta\mu = \mu_x - \mu^{\circ}_x = \Delta F + P_x \Delta V \tag{41}$$

where ΔF is the Helmholtz free energy, the total external work done, or the elastic strain energy, whereas the last term is the reversible

---

*The general Gibbs method (ref. 49) is followed, where a μ is defined separately for both the state considered and the standard state, leading to an expression for $\Delta\mu$; in addition, the conventional procedure of defining $\Delta\mu$ and then noting the choice of standard state is dispensed with.

**Note that for a generalized elastic stress system, the chemical potential of a solid is a tensor. The complexities in dealing with tensors (refs. 49 and 55) are reduced by basing the analysis on principal stresses. Note also that in conventional treatments of an elastic-stress effect on chemical potential, a hydrostatic stress is referred to. The present analysis, dealing solely with nonhydrostatic stresses, requires that the chemical potential be a tensor.

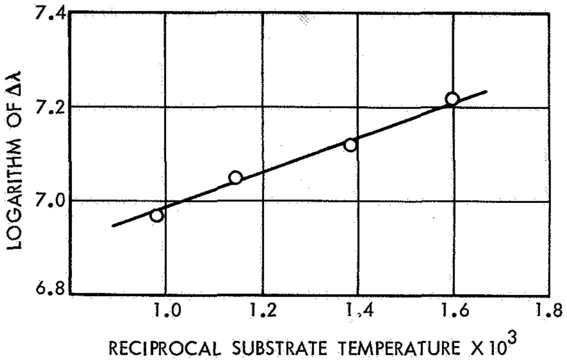


FIGURE 10.—Logarithm of $\Delta\gamma$ vs $1/T$ for lead vapor absorbed on solid copper (ref. 51).

work of displacing the fluid. This quantity is necessarily positive and therefore μ_x is always greater than $\mu^o{}_x$. This means that the chemical potential at the free surface of a pure solid immersed in a fluid media of hydrostatic pressure P^o should always be increased whether the elastic stresses are compressive or tensile. Note that $\Delta F = \sigma^2/2E$, where σ is the applied stress (for example, uniaxial tension or compression in the Y or Z direction), and E is the modulus of the material. Since $P_x\Delta V$ is very small (P_x for a uniaxially stressed sample would just be 1 atmosphere), we have

$$\Delta\mu_x = \sigma^2/2E. \tag{42}$$

The excess solubility of the solid in the liquid resulting from the stress is

$$\Delta C(\sigma) = C_o\Delta\mu_x/RT \tag{43}$$

where C_0 is the equilibrium solubility for the unstressed solid. Equation 43 summarizes the total effect of the accretion constraint on an elastically stressed solid. Actually, since a pure solid is considered, equation 43 gives the driving force for dissolution of the solid. Considering a stress concentration effect at a surface discontinuity, this dissolution effect will be highly localized.

In the context of the force required to move a body relative to another (e.g., a shaft supported by an annular housing), the asperities present will be subjected to a compressive stress, whereupon the propensity for dissolution will increase. This concept may be of importance in the well known effect (ref. 56) of iodine on the lubricating qualities of oils containing a small quantity of iodine. Since the metal is not soluble in the oil, iodine, which is soluble, will serve as a transporting agent for the metal, presumably as an iodine complex. The asperities will be selectively dissolved and the material deposited at a region of zero stress. The net effect results in a smoother surface, so that instead of equation 21 being applicable to the system, after some time, equation 25 is applicable. In this limit, the system behaves in such a manner that it is no longer sensitive to the magnitude of the applied load. Thus, heavy load lubricants (iodine additions to oil) may be viewed as stress-induced dissolution of asperities on contiguous areas, negating the "lifting" contribution of friction and shifting the total force required to move the bodies to that required simply to shear the interposing media (the lubricant). In consideration of crystalline media (say oxide films), the latter process is independent of the applied load but is expected to be somewhat dependent on the applied normal load as it affects the viscosity of the media.

For the majority of metals (ref. 57) iodides are known to form at low temperatures and decompose at higher temperatures (ref. 58). Assuming that there are very hot spots in the envisioned lubricating situation, we would expect, and it seems to be observed, that the iodide is unstable at the asperities. Therefore, the transport mode may be as complexes, resulting in the formation of iodides at the regions of zero stress (the low spots), rather than, as postulated in reference 59, in the formation of lubricating iodide films at the asperities. It is important to note that the distribution pattern of iodides developed in iodine lubricants on steel (ref. 56) is in keeping with the general concepts discussed above. The iodide deposits forming a deposition pattern around the asperity led Furey (ref. 56) to conclude that the lamellar structure of the iodide does not contribute to the mechanism by which friction is reduced in this system. This new concept of iodine action (stress-induced dissolution) is consistent with the recognition that lamellar solids are intrinsically abrasive and, of themselves, cannot improve the lubricating conditions between other solids (ref. 60). Note that the stress-induced dissolution (by iodine), the removal of the "lifting" contribution of friction, and the achievement of a force based solely on the pertinent shear stress (which to a large extent is independent of load) occur whether the stresses are compressive or tensile; i.e., the effect is independent of the sign of the stress. The material removed from a stressed region will be deposited in regions where the stress is zero (neither compressive nor tensile). This feature of itself may be useful in analyzing stress patterns between contacting solids.

At this point it is well worth noting that even interstitial solutes, in addition to the substitutional solute we have been considering, will have a demonstrable behavioral pattern. In the overall analysis of the effect of stress on the chemical potential of solids, it is essential to separate the effects controlled by interstitials from the effects we have already discussed.

The swelling or absorption constraint pertains to reversible transport of interstitials to the solid from the environment; since the interstitial takes a position within the solid lattice, it is effectively absorbed into the solid. This is in contrast to the accretion constraint where the atom is added to or taken away from the surface layer of the solid. Employing an analysis similar to that of Gibbs (ref. 49), Gurney (ref. 55), and Yang, et al. (ref. 61), have shown that, for interstitials,

$$\Delta\mu_i = \Delta\overline{F}_i + [(P_x + P_y + P_x)/3]\overline{V}_i - P_x\overline{V}_i^{\circ} \tag{44}$$

where the bars refer to partial molar quantities. In the usual case of positive partial molar volumes, $\Delta\overline{F}_i$ is always positive, but the PV term may be either positive or negative, depending on the sign of P, and is generally much larger than $\Delta\overline{F}_i$. Therefore, the chemical potential of an interstitial solute tends to increase on compression and decrease on tension.

The clean surface.— There are a number of well known techniques for observing surfaces at a high degree of resolution, each with its own capabilities and its own method of preparing an atomically clean surface. These techniques (e.g., field-ionization microscopy, field-electron emission microscopy, or low-energy electron diffraction) have, as yet, not come into vogue in friction and wear studies; however, the general techniques employed in cleaning surfaces used with these techniques have been to a certain extent assimilated into some of the current literature on friction and related topics. Of these auxiliary techniques, that of observing the properties of a sample in an ultrahigh vacuum has achieved some attention. Dr. Merchant's reference to Buckley's work in an ultrahigh vacuum is an example. I would like to make a few remarks about sample preparation in ultrahigh vacua (10^{-10} torr or better) when oil diffusion pumps are used.

Oil diffusion pumps have had a long history of alleged contamination properties, and I would like to add another observation to that list. The essential point of the new observation is that maintaining some rather low residual gas pressure for long periods of time is not an adequately sensitive criterion for assessing isolation of oil diffusion pumps from ultrahigh vacuum chambers (ref. 62). The newer means of isolation involve primarily molecular sieve trapping, and I would like to show that complete isolation requires that a titanium getter be interposed between the oil pumps and vacuum chamber. The function of this titanium getter is to complete the trapping process of all oil products (light hydrocarbons) not affected by molecular sieves, liquid nitrogen coolants, or, to an unknown extent, titanium getters placed in a side arm of the apparatus (see, for example, Buckley's vacuum system). Samples exposed to an oil-pumped system without proper titanium gettering will have thin but detectable carbide (or carbon) layers on their surface. This layer is generated by decomposition of the light hydrocarbons on a surface annealed above the temperature required for decomposing the hydrocarbon. This effect can be produced in a vacuum system with a 2×10^{-10} torr base pressure maintainable for weeks.

The transfer of oil into the high vacuum part of the system can be monitored by observing the field electron emission patterns developed under various conditions of treating the vacuum system. Good and Müller (ref. 63) demonstrated some time ago the appearance of a typical carbon pattern (carbon on tungsten). The predominant effect of carbon seems to be the formation of collars of high emission intensity about the (110) lattice planes and, at lower concentrations, the development of high work function areas in the (111) region and, to a lesser extent, around the (100) region of the emitter. In the case of oil contamination, carbon is formed on the tip by heating the emitter at low temperature, so that decomposition of the hydrocarbons occurs, leaving adsorbed "carbon" on the tip.

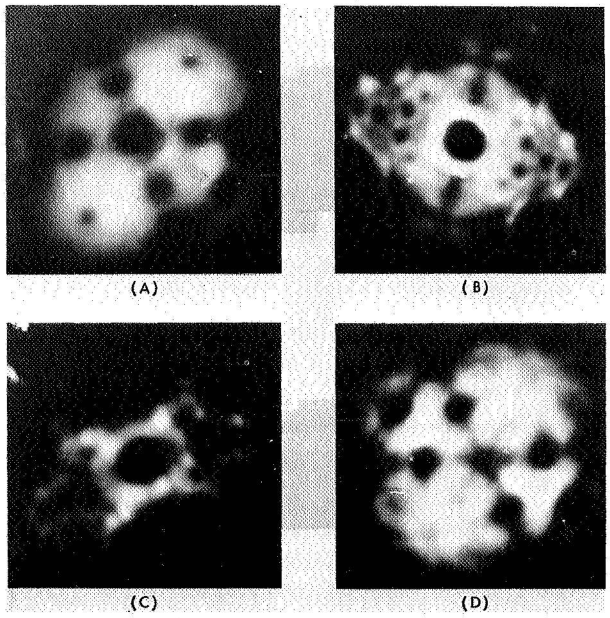


FIGURE 11.—Field-emission patterns showing effects of vacuum techniques on surface contamination.

Figure 11(A) shows a clean tungsten pattern* and figure 11(B) a tungsten FEM pattern produced by heating the tip to 1100° C ("flashed") after bakeout. The latter pattern is identical with the carbon pattern first identified by Müller. The low-temperature heat treatment causes decomposition of the "oil" adsorbed on the tungsten, leaving "carbon" behind. Flashing the emitter to about 1100° C in the condition shown in figure 11(B) results in a clean pattern. However, if the emitter is slowly heated at successively higher temperatures (to determine the exact temperature at which "carbon" desorbs from the tip), the tip becomes carburized and remains unchanged at subsequent flashing near the melting point of tungsten. Based on these tests, it is concluded that the achievement of long isolation times at UHV pressures is a necessary but not sufficient condition for precluding "oil" transfer into the high vacuum part of the system. In the situation where the substrate cannot be processed and maintained under the conditions necessary to remove the contamination, the substrate would be covered either with "oil" or carbon, or be converted to a carbide.

In order to preclude the occurrence of "oil" on substrate surfaces, the existing trap design was modified to include Fiberfrax** on all internal surfaces that could see each other and were not covered with Zeolite pellets. After installing Fiberfrax, using the same bakeout procedure, the carbon pattern, figure 11(B), could not be found. This point was tested with both the trap heated and unheated during bakeout, resulting in gas-covered (because of the low flashing temperature) but carbon-free patterns, figure 11(D). However, under quiescent operating conditions (vacuum gauge operating, but high-vacuum part of system otherwise at ambient temperature), the initially clean emitter, after adsorbing residual gas of the vacuum system for long times (in the range of 20 to 120 hours), displayed a carbon pattern. This pattern was developed as before by allowing the emitter, at ambient temperature and no applied electric field, to collect residual gas and then heating the emitter to 1100° C. This temperature seems to be that required for decomposing the adsorbed hydrocarbon, leaving a carbon residue on the tip. Evidently, hydrocarbons could still be passed through the trap, even under the rather stringent conditions of preparing the internal surfaces of the trap, i.e., surfaces covered with either Zeolite or Fiberfrax.

Modifying the system to include an in-line Ti sublimator does, how-

---

*These FEM patterns were produced using tin oxide as a phosphor (ref. 64).

**Fiberfrax is the registered trademark of Carborundum Corporation for their ceramic paper composed of 50.9 percent Al_2O_3, 46.8 percent SiO_2, supplied in sheets as stock No. 970–H, containing no organic binder. This material was fired at 1000° C for 15 minutes before using.

ever, completely isolate the pumps. The ultimate pressure of the system is reduced from 2×10^{-10} torr to 7×10^{-11} torr; the latter is believed to be the true X-ray limit of the gauge. The Ti sublimator need only be "flashed" for 90 seconds once in 24 hours to maintain isolation of "oil" from the pumps. Exhaustive repetition of the previously mentioned tests failed to show any sign of carbon in the system, even under the conditions of the quiescent tests (collecting residual gas on the emitter for over 100 hours). Evidently, light hydrocarbons may bypass the molecular sieve trap but they are gettered by the Ti sublimator. The tests prove conclusively that the source of carbon is the diffusion pump oil. The use of a titanium sublimator in-line (between pumps and UHV chamber) along with the 90° turn is believed to be a critical design feature (ref. 62). Residual gases coming from the pump zone must see Ti at least once before passing into the UHV chamber.

This contamination work demonstrates that samples whose surfaces are cleaned by long-time annealing at intermediate (relative to their melting point) temperatures in oil-pumped vacuum systems may have contaminant layers on their surface.

It is important to consider vacuum techniques not only because of the alleged contamination of vacuum systems by oil diffusion pumps but to illustrate the power of the field-electron emission technique in identifying surface effects. It is well known and generally appreciated that vacuum gauge readings differ from contamination rates determined by more sensitive techniques; also it is generally observed that vacuum gauges, because of their inherent pumping capability, can read a pressure lower in the gauge than in the system proper where a sample is being subjected to a controlled test. For example, it recently came to my attention that vacuum-pumping methods and pressure measurement were compared to field-electron emission pattern analysis (ref. 65); some fundamental and erroneous conclusions were drawn from the observations. In these tests (ref. 65), the vacuum gauge indicated a system pressure in the 10^{-11} torr range, yet the field-electron emission microscopy patterns shown were characteristic of much higher (by at least 1 order of magnitude) residual gas pressures (ref. 66). In fact, in the work in question (ref. 65), a clean tungsten FEM pattern (ref. 66) was not shown and perhaps never achieved. Furthermore, the overall changes in the FEM pattern, appearance and disappearance of bright spots, was attributed to the presence of individual gas molecules impingent on the tip. The most sensitive criterion of the presence of contaminants on a FEM pattern is the overall appearance of the pattern and its change with time at ambient temperature in the vacuum system. The change is such that, even at a determined pressure of 7×10^{-11} torr, a clean surface cannot be maintained for periods exceeding 1 hour (ref. 62). The overall contaminant rates and FEM pattern appearance (changes in effective work function) agree well with calculations based on the kinetic theory of gases (ref. 62). The bright spots in such patterns cannot be related to individual molecules but rather to residual gas being collected into asperities by the electric field. Therefore, the appearance of a single bright spot cannot be identified with adsorption of single molecules on the emitter, but rather with many molecules, the exact number of which depends on the details of the field enhancement factor applicable to such a process.

An oil decomposition effect is equally demonstrable and occurs with railroad rails in the normal context of their use. Andrews (ref. 67) has found that, after passage of locomotives, railroad tracks near stations have a lower friction coefficient than at other portions on the track. He traced this effect to oil dripping on the track when the locomotives are at rest at the station, subsequent to which the locomotive wheels, spinning when started again, produce an irreversible change (a significant decrease) in the friction coefficient. The decreased friction coefficient cannot be altered unless the rails are surface-ground; all other attempts to clean them do not significantly change the properties. It is evident that Andrews is observing a surface effect most likely produced by oil decomposition which, along with the high local temperature caused by the spinning locomotive wheels, produces a very thin carbide layer on the rail surface. Evidently, the carbide layer has a lower friction coefficient than the normal steel surface. This explanation is consistent with the need to surface-grind the rails and with the propensity of metal surfaces to develop carbide layers when annealed in ultrahigh vacuum systems under the general conditions allowing hydrocarbon decomposition.

The one-dimensional field-emission microscopy.—We have already seen the advantages of properly using field-emission microscopy (FEM) for observing surfaces. I would like to suggest using field emission for friction studies, not in the usual microscope form (where Müller uses a sharply pointed needle cathode) but in the form in which field emission was originally attempted (ref. 68), i.e., a wire rather than a needle cathode geometry. In this geometry, the field-enhancement effect may be used to observe the details of surface topography and the changes with various surface treatments. In addition, an auxiliary wire surface can be brought into contact with the observed surface, the contact resistance measured and the changes in surface topography caused by the intimate contact of the bodies analyzed.

This unique field-emission technique, which I call one-dimensional FEM, seems to have been first reported by Johnson and Shockley (ref. 68). It is highly suited to observing surface characteristics because of its high magnification and resolution and its adaptability to stressing wires in situ while observing the surface-topography characteristics of the wire. Creighton and Hoenig (ref. 69), after using this basic technique to observe the early stages of brittle fracture in 10-mil metal wires, have published dramatic evidence of its ability to detect crack nucleation from flaws and follow their propagation to fracture. An experimental technique capable of detecting microcracks would certainly be of use to friction studies in characterizing surface topography.

The 1–D FEM, if used in conjunction with the vacuum techniques already described, could increase understanding of surface topography effects in crossed-wire adhesion experiments. Johnson and Keller's (ref. 70) use of the crossed-wire technique for adhesion studies would develop an additional level of interpretation and understanding if the 1–D FEM capability were incorporated into the device. In this connection, since the 1–D FEM is basically a simple device composed of a straight wire cathode concentric with an electron impact fluorescent phosphor (the electric field for field emission is applied to the phosphor while the substrate is at ground potential), it would be advantageous to see directly through the phosphor to observe the couple under test. An optically transparent (visual) electron impact phosphor (tin-oxide-coated glass) recently developed (ref. 71) for field emission in general would serve admirably in the context desired.

Conclusions.—An overall look at friction and adhesion as characterized by Dr. Merchant has yielded the following initial thoughts and tentative conclusions and judgments:

(1) The total force required to slide one body on another ($\mathcal{F}$) is the sum of the forces needed for shear processes (f) and for friction processes (F, which is probably the sum of the plowing and lifting contributions that work against the applied normal load). Thus,

$$\mathcal{F}=f+F.$$

For pure frictional processes (against a normal load W and a coefficient of friction μ):

$$\mathcal{F}=F=\mu W.$$

For pure shear processes (on an area A where a shear stress τ is needed):

$$\mathcal{F}=f=A\tau.$$

This may include the pure adhesion case where only atomic bonds must be broken. If the same area is involved for shear and friction processes and the overall stress is σ, the specific force needed for "sliding" ($\mathcal{F}'$) is

$$\mathcal{F}'=\tau+\mu\sigma.$$

This result is shown to agree with experimental data. Thus, friction and adhesion are separate processes.

(2) A nearest neighbor bond model of a crystal surface has been discussed in terms of surface tension characteristics. The thermodynamic equilibrium assumption is equivalent to assuming satisfaction of all surface bonds with one adsorbed layer of atoms. This bond model is extended to include the line-tension effect and its consequences on the characteristics of wetting, expressed in terms of the contact angle. The conclusion is drawn that macroscopic wetting tests may not be of value in determining the behavioral pattern of small aggregates of condensed phases on a surface. In an analysis using adsorbed lead on copper, the temperature dependence of the net surface energy is related to the surface forces seen by the lead; that is, an adatom desorption energy is derived.

(3) The effect of stress on the chemical potential of an atom on the surface of a solid is discussed from the viewpoint of material transport in a "bearing" environment. Theory and experiments show that, in a lubricating environment, the action of iodine in oils may be summarized in terms of iodine complex transporting of atoms at asperities leading to a smoother surface and a total force law dependent principally on the shear stress of the interposing media; this action is largely independent of normal stress. Formation of iodides (as a lubricating media at asperities) is less likely since the material that must be formed can only be formed in opposition to the temperature gradients in the system.

(4) A new method for ensuring development of a clean surface in an ultrahigh vacuum is discussed using FEM as an observational tool. It is shown that time of maintaining a low base pressure in a system (e.g., 2×10^{-10} torr) pumped with oil diffusion pumps is no criterion of isolation. Without an improved getter design, light hydrocarbons may adsorb on the surface. By annealing the sample for long times at modest temperatures, about 1100° C, these hydrocarbons decompose readily and lead to formation of a "carbide" layer on the metal surface. This observation is related to the change in the frictional behavior of steel railroad rails on which hydrocarbons (oil dripping from the train) are decomposed by the heat and pressure of spinning locomotive wheels to form a surface with an irreversibly (except by surface grinding) low coefficient of friction.

(5) Based on the emission properties of a straight wire cathode, the forerunner of the modern FEM, a modification of the FEM, called the 1–D FEM, is suggested for use in friction and adhesion studies (of the crossed-wire form). A recently developed phosphor, along with

improved gettering techniques, could make the 1–D FEM a significant extension of the crossed-wire adhesion experiment.

ACKNOWLEDGMENT

I would like to acknowledge helpful discussions and clarification of essential points provided by C. M. Allen, J. N. Anno, A. R. Rosenfield, and B. A. Wilcox, all of the Battelle-Columbus Laboratories.

D. V. Keller (Syracuse University, Syracuse, New York)

Dr. Merchant should be commended for his extensive review of friction and adhesion and for emphasizing areas that still require extensive investigation, so that friction and adhesion can be placed on a quantitative basis.

In order to expand on this point with particular reference to metallic adhesion, let us examine the steps necessary to put such a process into a quantitative framework. A quantitative analytical approach to metallic adhesion, leading to the development of critical experiments, requires that an ideal model of the process be tested experimentally and compared to similar processes. On the basis of the test results, the model is accepted, rejected, or modified. Most of us are familiar with the models of adhesion as proposed by Bowden and Tabor (ref. 72), involving the interaction of two rough surfaces, (i.e., asperity interaction) and with that attempted by Keller (ref. 73) based on an atomic approach. Although both models are probably as acceptable as the numerous proposals that fall between the limits cited, it would appear that the inherent complexities of the proposed models inhibit a complete quantitative analysis. Consequently, a far more simple model exposing some of the more salient points of the adhesion process without losing contact with the models previously suggested is proposed.

The proposed model of the adhesion process, figure 12, shows two single crystals so oriented that exact atomic coincidence will occur when the ideal metal-vacuum interfaces, i.e., atomically clean interfaces, are brought together (A). Step 2 (B) of the process requires that these surfaces be brought into contact under zero load so that exact atomic coincidence is realized. Step 3 (C) of the process requires that the system be tested in pure tension.

The experimental accomplishment of step 1 is feasible with most metals to within a few minutes of crystallographic orientation if the surface is carefully prepared as established by low-energy electron diffraction studies.

Step 2, on the other hand, requires a thought process, since absolutely perfect orientation and zero contact load are two mechanical steps not readily achieved in the laboratory. If they could be achieved, however,

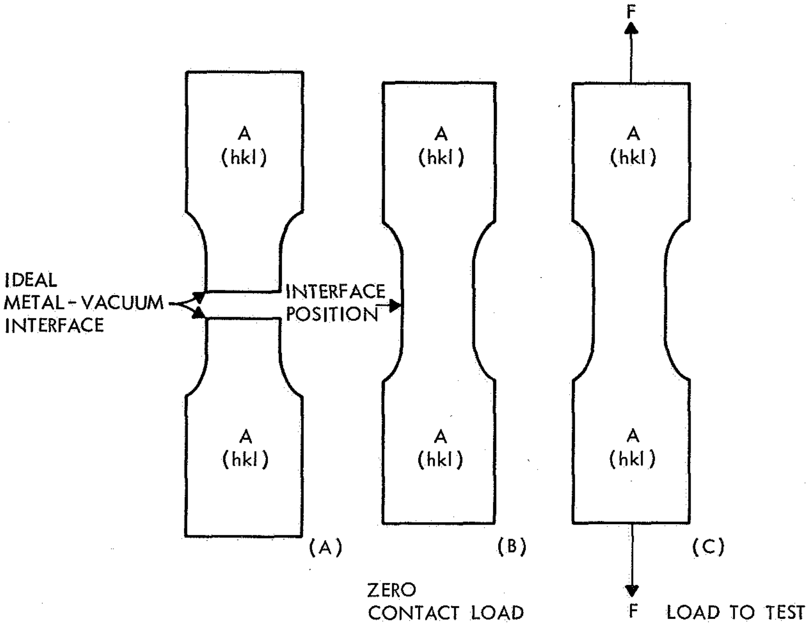


FIGURE 12.—Proposed model for adhesion of metal *A*. (a) Metal-vacuum interfaces for single crystals oriented to contact in atomic coincidences. (b) Surfaces contacted under zero load. (c) Tensile test of system.

one would expect the resulting unit body, shown in figure 12(B), to be indistinguishable from a single crystal. The validity of this very simple model rests on the assumption that the available surface energy release on interface formation is sufficient to establish metallic bonds across the interface so that the interfacial energy is zero. A similar model is quite acceptable in the physical chemistry of surfaces and is frequently used as one step in a reversible process for defining surface energy (ref. 74). Since chemisorption seems to lack an activation energy barrier (ref. 75) and the pair, atom bond strength of the bonds across this interface are of the same magnitude as those in chemisorption, it seem reasonable to assume that no large activation energy of interface formation is involved. As a consequence, the physical structure of 12(B) is that of a single crystal in which the original interface is undiscernible from that of similar planes in the crystal bulk. Recent cleavage studies of germanium (ref. 76) and numerous studies of epitaxial growth from the vapor or liquid state also serve to illustrate that the model, thus far, is reasonable.

By accepting the conditions set forth in step 2, step 3 becomes simply a tensile test of a single crystal described fully in most mechanical metallurgical texts; e.g., a stress-strain diagram for a particular crystallographic orientation can be reproduced and to some degree analyzed on the basis of dislocation theory.

The ideal adhesion test cycle is complete and except for step 2 can be conducted in almost any well equipped laboratory. In order to reduce step 2 to a workable experiment, let us now examine the ideal process with a slight modification, i.e., compressive prestrain in step 2. If the prestrain is ideally elastic, there should be no change in the observations of tensile testing as compared to step 3 in the ideal model. In fact, one would also predict that prestrain slightly above the compressive yield point would only contribute slightly to modifying the resulting stress-strain diagram, shifting the ideal stress-strain curve toward higher crystal strengths. More excessive prestraining will cause a substantial shift in the resulting stress-strain curve, depending on the material and extent of prestraining. Therefore, if we interject a reasonable degree of compressive loading into the ideal adhesion cycle, the resulting tensile strength of crystal should not differ greatly from that of the ideal system.

In further expanding the simple model, let us now rotate the upper bar of figure 12(A) 1 degree-of-arc around the z-axis so that a low angle grain boundary is produced when step 2 is conducted. Since a low angle grain boundary consists of dislocations, the strength variation from the original experiment should not be significant. Larger and more complex rotations simulating high-angle grain boundaries may provide unique slip directions and thus establish a strength-versus-orientation effect. Since the interface formed in the contact of two polycrystalline rods, as shown in figure 12, can be reduced to many unit crystal interactions, the whole strength would be expected to behave as the sum of the unit strengths making up the whole with a slight contribution from the grain boundaries that intersect the metal-vacuum interface before contact. For adhesion cycles between misoriented crystals or polycrystalline metals, one would expect the interfacial energy at step 2, figure 12(B), to be equal to, or slightly greater than, the equilibrium grain boundary energy of the polycrystalline metal, depending on the ability of the interfacial atoms to reach a boundary quasi-equilibrium position which, in turn, is a function of test temperature. In all cases, it is still reasonable to assume that the free energy release in interfacial formation does not exceed that remaining as a residual in the crystal; i.e., in the case of polycrystal metals, the stress-strain diagram, figure 12(C), can be represented as approximately that of a bulk polycrystalline material. If this were not the case, ultraclean grain boundaries would be expected to yield under a much lower force than is generally observed. This extrapolation of the ideal model leads to two conclusions: (1) adhesion results of single crystal systems, e.g., compressive or tensile deformation of a crystal, ought to be structure-dependent, as has been shown by recent studies of Buckley (ref. 77), and (2) adhesion results using polycrystalline metals ought to be closely related to the strength of the respective polycrystalline metals, as has also been observed.

In order to include the generally accepted model of Bowden, et al· (ref. 72) and to place the proposed model on a testing basis, the effects of rough surfaces in figure 12(A) must also be considered as a compressive load is applied to 12(B). This problem has been carefully considered in a recent review by Holm (ref. 78) in which the variation of the contact radius a of the true contact area was shown to vary with load P as follows:

$$\begin{aligned} a &\propto P^{1/3}(\text{pure elastic deformation}) \\ a &\propto P^{1/2}(\text{pure plastic deformation}). \end{aligned} \tag{45}$$

Since an adhesion cycle is likely to proceed from the former to the latter with slight deviations resulting from surface discontinuities and multiple contact effects, we should expect that the contact radius should lie roughly between these two limits. Furthermore, Holm has shown that the constriction resistance R_c in the first approximation for a system with an ultraclean contact point should be

$$R_c = \frac{\rho}{2a} \tag{46}$$

where ρ is the conductivity of the pure bulk metal. Combining these equations we obtain

$$\begin{aligned} R_c &\propto P^{-1/3}(\text{elastic}) \\ R_c &\propto P^{-1/2}(\text{plastic}) \end{aligned} \tag{47}$$

The proposed adhesion model for polycrystalline materials with atomically clean, mechanically rough surfaces, therefore, suggests that during loading the contact radius will expand according to equation 45, and the fracture of this area in tension should proceed as a polycrystalline metal subject to the geometric constraints of the test. The bulk strength of the junction will approach the polycrystalline bulk strength as the notch effects are reduced.

In order to test the model developed above, experiments were devised (refs. 79 and 80) in which crossed-wire samples (1.2 mm o.d.) were cleaned by argon ion bombardment and flashing in an ultrahigh vacuum system so that ultraclean annealed metal surfaces could be brought into contact. The system was also designed so that the samples could be statically loaded to peak loads between 0.01 and 5.0 grams with a load sensitivity of about 0.01 gram.

Simultaneously with the sample loading and unloading process, contact resistance was measured by the cross-wire technique using a Kelvin Bridge; a sensitivity of about 10 microhms was realized. The results of the investigations on silver-silver, titanium-titanium, molybdenum-molybdenum, nickel-copper, nickel-silver, and tungsten-silver samples with ultraclean surfaces and cycled to peak loads varying from 0.020 to greater than 2.0 grams is summarized in figure 13.

The data from these studies presented as constriction resistance versus load during loading, or deformation, usually fell into a band between $R_c \propto P^{-1/2}$ and $R_c \propto P^{-1/3}$ under heavier loads, were small, and probably could be accounted for by asperity effects. Since the sensitivity of the Kelvin technique for resistance measurements was in the range of 10 microhms, the radius of the real contact area could be estimated on a relative basis to about one part in a hundred, which is more than adequate to demonstrate the principle involved. The effects of the more than 700 adhesion cycles on ultraclean metallic surfaces were demonstrated by a technique involving stepwise loading and unloading of the samples; this technique is believed to have caused a degree of scatter and prevented the interpretation of the fine structure during the cycles. Apparatus modification has proved this point conclusively; however, publishable results from the modified system are not available as yet.

Upon unloading the system after establishing a peak load, there was no detectable variation in constriction resistance until reaching a tensile load that initiated fracture. Since automatic recording equipment was not used in these initial experiments, the fine structures of the fracture process were not examined; however, the stability of the constriction resistance to the initiation of fracture indicated that the contact radius was constant to approximately one part in a hundred from peak load to fracture. This observation is comparable to that expected for a tensile specimen with a crack of exceedingly small crack tip radius.

The strength of the junction formed by the various adhesion cycles was estimated from the force required to fracture the samples and the real area of contact as ascertained by the simple constriction resistance equation, equation 46, and/or elastic theory. The fracture strength of the similar metal couples, for the most part, fell in the region between the literature value of the work-hardened strength and the annealed strength. For the dissimilar couples, the strength value could best be represented by the same range for the weaker of the two metals.

The simple model and the expanded polycrystalline model for an

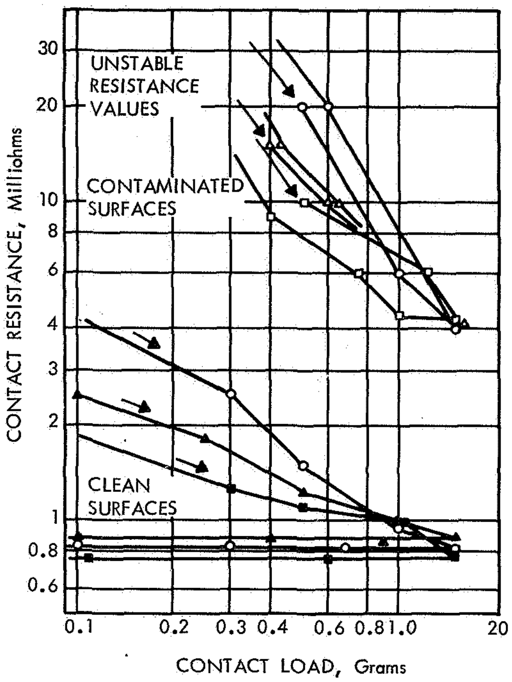


FIGURE 13.—The change in contact resistance between silver couples compared to the contacting load for three adhesion cycles with contaminated surfaces and three ultraclean surfaces.

ideal adhesion cycle proposed above appear to be consistent with the experiments performed on geometrically simple systems (crossed rods) and with other associated physical phenomena. Since our purpose in establishing a model was to make available a route for analysis and/or establishment of critical experiments, let us examine whether the point was achieved. Quantitative analytical data could not have been extracted from the experiments above for two reasons: (1) the relationship between constriction resistance and contact area is not precisely known, and (2) the stepwise technique using polycrystalline metals inherently complex for complete fracture analysis was used to obtain data. These points immediately suggest several obvious critical experiments from which quantitative results could be obtained. The relationship between constriction resistance and contact area could probably be defined precisely by investigating the exposed fracture region after an adhesion cycle was conducted by electron microscopy techniques. This relationship, in turn, would provide a direct route for analyzing strength of the system under various conditions. The major modification in the apparatus necessary to improve reproducibility involves automatic loading and unloading of the crossed rods; it would prevent erratic time delays from point to point and, therefore, remove any time constant effects. The investigation of single crystal systems ought to also be expanded to facilitate detailed fracture analysis. It can be concluded that the routes for detailed quantitative analysis are available.

Thus far no mention has been made of the material effects or contamination effects on the proposed adhesion model, even though these details have been examined and discussed elsewhere (refs. 79 and 80). Briefly, the metal used in the adhesion couple has not affected the shape of the constriction-resistance-versus-load curve. Fracture occurs in the weaker metal within the limits cited above. Contamination of one or more monolayers causes a gross change in the minimum constriction resistance value and the shape of the resulting loading curve (R_c vs P). The change appears to depend on the metal system under investigation and the contaminating agent.

M. B. Peterson (Mechanical Technology Incorporated, Latham, New York)

The search for new knowledge on friction and wear proceeds toward three goals: understanding of the frictional processes, understanding the effects of variables on the friction processes, and understanding the behavior of materials in applications.

Dr. Merchant has directed his attention to understanding the adhesion and the friction processes. He has shown that efforts have been directed primarily to predicting friction—not because it is important in itself but because the ability to predict friction implies a complete understanding of the process. It seems clear that adhesion and friction are essentially measurements of the rheological behavior of surfaces. This being so, a direct correlation between friction and some strength property should be expected. As Figure 14 confirms the expectation by plotting the coefficient of friction against tensile strength for different materials sliding against hardened tool steel at different temperatures. These data have been obtained at a load of 56,000 psi with the area of contact constant. Under these conditions, the adhesion is of little importance. Figure 15 shows the effect of surface roughness on coefficient of friction for tin sliding against a variety of materials. In spite of estimated differences in adhesion, almost all differences in the friction can be explained by differences in roughness.

If friction is primarily caused by surface deformation, one should expect correlation with the rheological properties of the material, for example, the creep rate. One would also expect a strong velocity effect in friction, since in the friction experiment a constant strain rate is

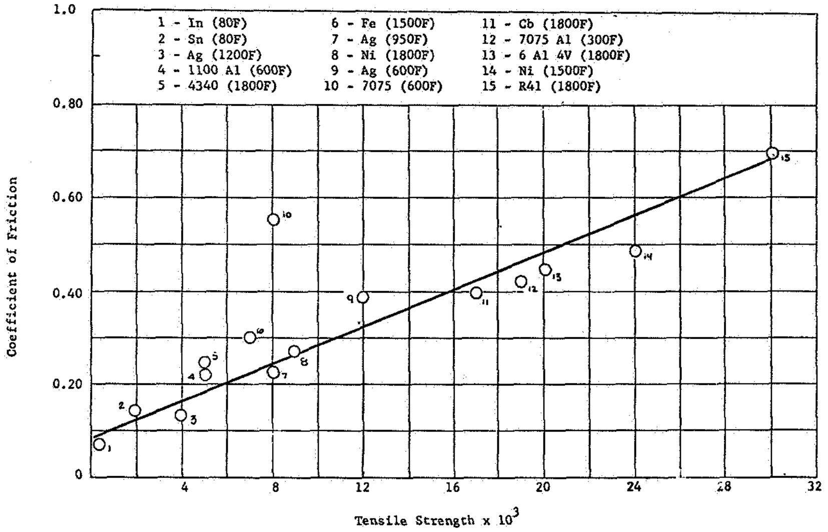


Figure 14.—Effect of material strength on the measured coefficient of friction at various temperatures, p=56 000 psi.

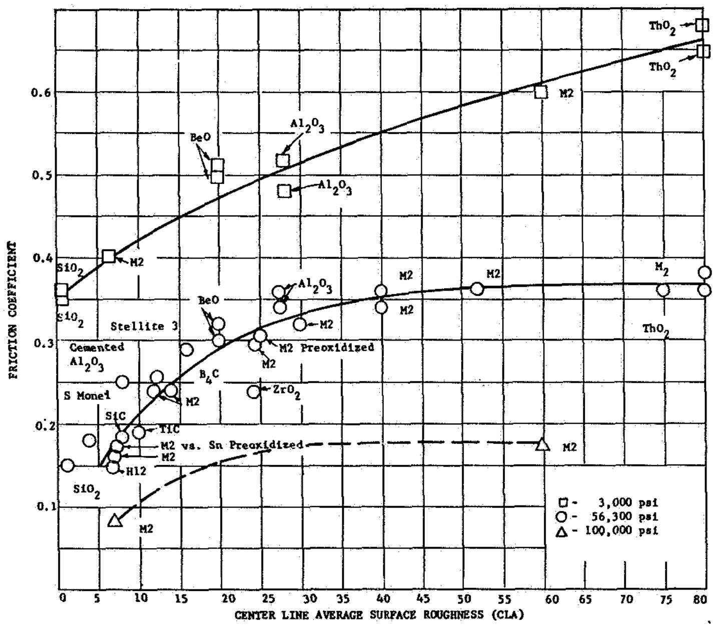


Figure 15.—Result of tin sliding against various materials at room temperature.

applied and the stress is allowed to adjust itself. Thus, if the material is being strained at a greater ratio than the material can flow, friction will increase as shown by Rabinowicz (ref. 81). If junction growth is more pronounced with ductile materials, a correlation should exist between friction and elongation since it is the true measure of ductility. Frictional behavior should be put in terms of material properties, so that the information and understanding of that property can be applied advantageously.

For very hard and some soft materials, the frictional properties are determined by the surface films formed in sliding. Although some work has been done, almost no general conclusions have been drawn. Essentially we should like to know what films are formed under sliding, how do they differ from the natural films, and which films are effective. Much of the corrosion research would be useful here in determining the types of films that should form. There is sufficient evidence to indicate that the films most effective in promoting corrosion resistance are also effective in preventing damage from sliding. Essentially, this characteristic depends on the ductility of the films. Ductile films are most able to repair themselves to prevent rapid corrosion and most effective in preventing damage from sliding. Corrosion scientists could contribute greatly to this field.

The real problem in friction studies is that we are sure of only one variable, i.e., the load; the pressure, area, and surface temperature are always questionable. Furthermore, since we are not sure of what material we are dealing with, its properties are always open to question. It seems that at least the areas or the materials properties must be accurately defined before real progress can be made.

In the studies of variables, considerable work has been done in the following areas: material composition of ductile metals, hard metals and alloys, carbons, and polymers; atmosphere; surface finish; load; velocity; temperature; and time. Areas essential to understanding friction but requiring further research are: material composition of cermets, ceramics, and composites; material properties; area; shape; and pressure.

Studies to understand the frictional behavior of specific material types have provided more information about the soft pure metals than harder metals and alloys. Only a few studies have been conducted on the frictional behavior of cermets and ceramics; much more work is necessary in this area. Carbons and plastics have been studied and their frictional behavior understood in a qualitative way. The frictional behavior of composite materials has not been studied to any degree. Here is a new type of material with promise of controlled properties and of controlled frictional behavior.

Although some work has been done on the effects of material properties on frictional behavior along the lines of Sikorski and others, much more work is needed. This work is difficult since it is not easy to change just one variable at a time. The area of contact and material properties are two significant variables that should be firmly established.

Almost no work has been done on the effect of shape on frictional behavior. The fact that it does not affect the coefficient of friction is relatively unimportant; the effect of shape on surface damage, wear, and metal transfer should be known.

The other variables have been studied or are currently being investigated with the exception of the effect of pressure. At low pressure, frictional behavior is independent of pressure; however, this is not necessarily true at high pressures nor is it expected. At high pressure, nearly the complete area is in contact and the mechanical properties of the material uniquely determine its friction properties.

It is interesting to note that the three areas least understood are precisely those that would be most useful in improving or selecting materials for applications. Furthermore, the information usually acquired is about friction. It would be much more important to characterize the resulting surface damage since this is the means by which materials must be selected.

Great strides have been made in the last thirty years in understanding the basic friction and wear processes and the effect of variables; however, very little of this information has reached the engineers who could most use the information. No simple techniques have been proposed reliable enough for use, even by the experts, to select materials or to determine the adequacy of a material for an application. As a result, the designers still rely on special materials with which they have had some experience, or on surface-hardening techniques. If these do not work, they search for others on a trial-and-error basis. Undoubtedly, when reliable techniques have been found, we will also find a much more welcome audience for our research proposals. We are not saying that some strides have not been made or that finding simple techniques will be easy, but that not enough effort has been expended.

Developing design criteria is difficult for two reasons. First, in any application there are many requirements other than effective sliding characteristics. The material must have adequate strength and corrosion resistance; its thermal expansion must be compatible with other materials used in the application; it must be hard enough to resist wear and be easy to lubricate. In many cases, these requirements reduce the choice to a few materials. For example, where abrasive wear is a factor, hard materials would be specified. What criteria would we then apply to select a hard material that will resist any form of surface damage? Criteria previously proposed often suggest materials such as silver or chromium that do not resist wear or are the most difficult to lubricate.

The second reason is that there are many forms of failure or surface distress rather than one. High friction may be tolerable in one application and not in another; sometimes low friction is undesirable. Some surface damage can be tolerated in certain applications, but not in others. Many factors have to be taken into consideration; however, this in no way diminishes the need for new techniques.

I do not believe that it is necessary to conclude the research before beginning this work. If simple techniques for evaluating the effectiveness of sliding combinations of materials were to be derived, it would seem that, first, the existing failure processes should be studied. These have been fairly well classified as followed: (1) fracture, (2) abrasion, (3) lubricant film wear, (4) critical temperature, (5) surface deformation, (6) fatigue, (7) melting of surface film, (8) fracture of surface film, (9) wear of surface film, and (10) adhesion and metal transfer.

The amount of information on each of these processes varies. Dr. Merchant showed that much work has been done in the area of adhesion fatigue, and critical temperatures. However, very little attention has been directed toward lubricant film wear and failure from deformation and fracture. In order to develop design criteria, the materials commonly available and the geometries of the most common applications should be used to determine the operating conditions under which wear and failure occur, in terms of design variables such as pressure, velocity, and sliding time. Gradually, enough information would be collected to allow generalizations. This was the approach to handling the problem of fatigue; no one suggested that we wait until all our research was complete. This same approach should be taken with each failure process, listed above.

J. J. Bikerman (Horizons Incorporated, Cleveland, Ohio)

We just heard a particularly clear exposition of the majority view on friction. Its clarity makes it easy for me to outline the differences between this concept and that advocated by the minority.

The author raises three objections against the Coulomb theory of friction, but we consider that theory to be correct. The first objection is that the theory cannot account for the loss of energy in sliding, i.e., for the frictional heat. The argument runs as follows: the experimenter spends work for lifting the slider over the hills of the support but then the slider falls down, and an equal amount of work is recovered; consequently, no heat can be evolved.

This argument is astonishingly wrong. Climb up a mountain and then return; see whether the energy you spent in climbing was injected back in you during the descent. A more scientific rebuttal may use the terminology of thermodynamics. According to thermodynamics, the total sum of work done during an isothermal reversible cycle is zero. The cycle must be reversible. Here I lift a weight. In order to make the process reversible, the force applied by my hand must be equal to the weight of the object, implying that the motion must be uniform and extremely slow. Having reached the highest point, I permit the object to descend and to entrain my hand again in a uniform and slow motion. In such a cycle, ideally, no total work is done and no heat is evolved. But let us now perform a different experiment. I lift the object "reversibly" as before but then release it from the highest point. It falls with acceleration, hits the support, and there liberates an amount of heat equivalent to the work performed during the ascent. The potential energy possessed by the object in its summit position now is fully transformed into heat.

The microscopic events in sliding are analogous to the second experiment. When the slider is lifted over an obstacle, the lifting force is practically equal to the relevant component of the slider weight; but when a protuberance of the slider falls into a valley on the support surface, the fall is free, and the potential energy acquired during the climb is dispersed as heat.

The second objection is to the statement that "the usual maximum slope of surface asperities is far too small to explain the usually observed values of coefficient of friction." This argument is also invalid. As mentioned, for instance, in Dr. Williamson's paper, the average slopes of surface hills range between 5° and 25°. The maximum slopes would be 10° to 30°, with tangents of 0.18 and 0.58, respectively. The coefficients of friction should be equal to these tangents, and they are. As Dr. Merchant mentioned, it was believed centuries ago that the coefficient of friction was always near 0.25; this means that the maximum slopes of the solids used then were about 14° (as $\tan 14° \approx 0.25$), i.e., not different from the 20th century surfaces. Thus, the modern quantitative data on surface roughness are a marvelous proof of Coulomb's venerable theory.

The third objection is that the Coulomb theory cannot account for the boundary lubrication. It can. In simple instances, very small amounts of a lubricant have two effects: they prevent the hills of the slider from descending too deeply into the valleys of the support, and they substitute a viscous oil for the "rare" air as the medium that must be sheared between the hills and the slopes of the valleys.

A criticism often voiced by the minority against the majority view pertains to elastic recovery. It has been repeatedly asked—first by William Hardy himself—why, if the slider really adheres to the support, cannot this adhesion be noticed when the slider is lifted up? The answer usually given is that elastic recovery or peeling occurs and destroys the bond. I publicly disagreed with this explanation almost 30 years ago, and now I have a more solid ground to stand on because our knowledge of adhesion is much more profound now. In the book I recently published (of which a second edition will appear early next year), I maintain that our information reached the level of a science (ref. 82). According to this science, the postulated effect cannot occur. If there is an adhesion, it can be measured just as surely by peeling as by shear. If no adhesion can be detected by peeling tests, there is no adhesion to begin with. In 1959, W. P. Mason (ref. 83) published a paper in which the hypothesis of elastic recovery was put in a quantitative form. Unfortunately, the equations that the author took from one of Timoshenko's books are valid only in the absence of adhesion. No wonder that they lead to the conclusion that no adhesion is visible when the slider is raised from the support.

The experiments by Mr. Buckley are a good confirmation of our view. When two solids are pressed together in a vacuum of 10^{-11} torr, the mutual adhesion that takes place can be measured when the pressure is removed. If elastic recovery were as all-powerful as claimed, no normal adhesion would have been detected in Mr. Buckley's tests, either.

P. M. Ku (Southwest Research Institute, San Antonio, Texas)

As Dr. Bikerman has pointed out, the adhesion theory represents what might well be termed the majority view on the theory of friction. I will not attempt to give here a historical account of the development of the friction theories. Suffice it to say, other theories had preceded the introduction of the adhesion theory in the early 1940's and, since that time, the adhesion theory has received wide and enthusiastic support, though not without some strong objections.

We have an unusually good representation of both sides at this symposium. It is not for me, as chairman, to take sides. However, in the spirit of this symposium, I feel I have a clear license to encourage the dialogue. We have now heard Dr. Merchant's exposition of the majority view. We have also heard the dissenting arguments expressed by Dr. Gretz and Dr. Bikerman.

It is our privilege to have with us Dr. Merchant and Dr. Tabor—two originators of the adhesion theory. Dr. Merchant has already presented his view in his lecture and will have another opportunity to defend it in his closing remarks. Therefore, with his permission and Dr. Tabor's indulgence, I take the liberty to call on Dr. Tabor to offer his comments at this time.

D. Tabor (University of Cambridge, Cambridge, England)

The purpose of this symposium is to encourage an interdisciplinary approach to friction and allied subjects. For this reason, I turned with considerable interest to Dr. Gretz's contribution. I have read this several times but cannot say that it has illuminated the subject for me. Dr. Gretz invents definitions that he uses in his own way; furthermore, he carries out "thought experiments" that seem to me to be very far from physical reality.

For example, his definition of coefficient of friction given in equation 12 is the ratio of the ideal plastic shear strength to the ideal brittle tensile strength of a perfect crystal. In friction, we are usually concerned with an applied normal stress and a tangential stress to produce sliding. The normal compressive stress σ so applied has practically no connection with the theoretical brittle tensile stress $\sigma_{\max}$ used by Dr. Gretz. It is a great disservice to the reader to confuse these two quantities as Dr. Gretz does in moving from equation 14 to equation 16. He then rectifies this confusion, after equation 17, but why introduce it at all?

Another part of the thought experiment assumes an area of real contact independent of the normal load. This is a major defect of the Coulomb mechanism of friction: there is no reason why it should be perpetuated 180 years later. Again Dr. Gretz stresses that a tangential force cannot do work against a normal force. Of course this is true and, in the complete absence of adhesion, friction can only arise from some sort of plowing, asperity lifting, or deformation mechanism. In the presence of adhesion, an additional shearing term is involved; however, Dr. Gretz provides no critical thought as to the area of material that must be sheared and how far the area is affected by the applied load (he again assumes that it is independent of the load). It is a great comfort after several pages of text and 15 equations to read that the process of sliding "involves both a shearing process and a frictional (asperity lifting, or deformation) process."

Finally, Dr. Gretz performs a thought experiment in which plowing of asperities is excluded and only lifting contributes to the frictional force. He now applies not a compressive stress at the interface but rather a tensile stress. I presume this means that instead of pressing the surfaces together he pulls the surfaces apart. This thought experiment leads to a "perfect bearing" in which the tensile forces do work against the shear stress. All this means is that the normal tension is pulling the adhesionless bodies up the surface asperities. It is not surprising that in such a system, where the resultant normal force is zero (or negative), the net force necessary to produce sliding is zero (equation 27). This is not a perfect bearing, it is a meaningless abstraction.

I am much more impressed by Dr. Gretz's comments on clean surfaces and the dangers of contamination from oil in the vacuum system. It is a sensible study and a salutory warning to all workers in the field.

I had not intended to record in writing my criticisms of the ideas expressed by Dr. Bikerman on the mechanism of friction. However, since Mr. Ku has asked me to do so, I, somewhat reluctantly, summarize my views in the following terms.

The view that friction is caused by dragging one set of asperities over another was first explicitly exploded by Leslie in 1804. He pointed out quite clearly that the surfaces are not continually ascending; they must alternately rise and fall over the asperities. If there is no adhesion, this process cannot be dissipative. The behavior resembles a roller coaster in which frictionless carriages will continue to move up and down without losing energy. If, however, the rail is broken at the top of a rise, the carriage will fall and all its energy will be lost by impact with the ground. Here, of course, we have a dissipative process—the deformation of the ground and the carriage.

John Leslie obviously did not understand the Coulomb mechanism in this sense—nor indeed do I. It is evident from the original literature that Coulomb made little distinction between climbing up and down asperities, bending over of asperities, or breaking of asperities, nor did he consider the problem of energy dissipation. Dr. Bikerman has interpreted the Coulomb mechanism precisely in terms of the broken rail on the roller coaster. Clearly this provides a mechanism for energy loss by deformation, but even here one has to ask what sort of surface asperities allow one surface to ride up an asperity on one side and fall with a bang as soon as they pass beyond the peak of the asperity. Such asperities would have to be shaped like saw teeth with a slow rise in one direction and a steep or even negative slope beyond; consequently, on reversing the direction of sliding, the friction would be very different. The free fall postulated by Dr. Bikerman is very difficult to envision for any realistic picture of the surface.

J. A. Schey (IIT Research Institute, Chicago, Illinois)

I am somewhat disturbed by the arguments used to repudiate Coulomb's theory of friction. Dr. Merchant mentioned that the theory is unable to explain the loss of energy in sliding, since the energy spent in pulling a mass up the ascending slope of an asperity is recovered on the downhill part of motion. Dr. Tabor drew on the analogy with the frictionless roller coaster.

While these arguments are plausible enough, they oversimplify the problem by reducing a three-dimensional situation to two dimensions. In reality, two surfaces must touch on at least three points; if we assume that, in sliding one body over the other, these three points lie on the ascending slope of three identical asperities, the body would indeed amass sufficient energy by the time it reached the summit to keep on sliding, provided, of course, that the asperity pattern is identical in the path of all three contact points. This is patently impossible on any technical surface; it is more likely that the sliding body will come to a grinding halt simply by geometrical entrapment of the three asperities in nonmatching surface irregularities of the opposing surfaces.

It is possible that Coulomb's mechanism contributes in only minor degrees to the total of friction; however, its absence must be proven on sounder grounds than the analogy of the frictionless roller coaster. This exercise also raises the question whether theories built on analogies with single asperities can be validly applied to the essentially statistical problem of contact between two surfaces of random or almost random topography.

D. Tabor (University of Cambridge, Cambridge, England)

Dr. Schey makes the interesting point that asperities on one surface may be involved in geometric entrapment by nonmatching irregularities on the other. I do not think this raises any difficulties. Euler pointed out long ago that in a purely Coulombic model the top body will ultimately climb the steepest asperity in its path, with the load gradually being relieved from the other contacts. Of course, if one of the blocking contacts is so steep that climbing involves a very large force, it may be easier to shear the asperity completely away, but this is no longer the Coulomb model. It is also true that the friction might fluctuate; the body may ascend a long steep asperity and then descend a short shallow one so that for one instant more energy is lost than gained. At some later stage, the reverse would have to be true if sliding does not involve any overall change of level.

But more important than these arguments is Dr. Schey's general point that models based on individual asperities may be very misleading. Consider the contact between two extended surfaces possessing the types of contours shown in Dr. Williamson's beautiful scale models. How often will surfaces present a series of asperities of identical slope and geometry? Very rarely indeed. As pointed out above, a few dominant engaging asperities will have to take most of the load. In general, however, the surfaces will not behave like an assembly of rigid asperities; the asperities in real life are deformable. The load will be shared over a large number of asperities, some carrying a larger and some a smaller part of the load; the deformations will be elastic or plastic. If they are plastic, an asperity on one surface will either form a groove on the other during sliding so long as contact persists or flatten the tip of the asperity. In both cases, work is done by plastic deformation. If the deformation is elastic, a large part of the deformation will recover as the asperities pass out of contact. The energy loss will be determined by the elastic hysteresis properties of the material. (An elastically deformed asperity could lose all its elastic energy on passing out of contact if the horizontal movement was rapid enough, i.e., if the surfaces moved with a velocity comparable to the velocity of an elastic wave in the material.) In neither the plastic nor the elastic case is it very helpful to describe the process in terms of Dr. Bikerman's version of the Coulomb model. If adhesion between the surfaces is small or negligible, the major part of the friction arises from plastic deformation or elastic hysteresis. Plastic deformation losses are more common with metals, a point much emphasised by the Russian scientist Kragelsky. Elastic hysteresis losses are more common in polymers and elastomers. A loss mechanism by deformation of surface asperities is essentially the one proposed by Leslie rather than the general Coulomb model accepted by his contemporaries.

Of course, the evidence against the Coulomb theory is not simply a matter of polemics. We have taken molecularly smooth mica surfaces, curved them into the form of cylindrical sheets, and slid them across one another. By optical interferometry through the sheets, we may watch the region of contact during the sliding. There are no surface roughnesses even on a molecular scale, yet the friction is very high. If a monolayer of fatty acid is adsorbed on the surfaces, the friction is low. Furthermore, we can now measure the attractive forces between the surface and relate this to the friction. I think Dr. Bikerman will accept these results, he will merely say that, if there is adhesion, this is not "real" friction. But even in those situations that Dr. Bikerman recognizes as "real" friction, radioactive tracers generally show the presence of strongly adhering material transferred from one surface to the other.

Dr. Bikerman is opposed to the concept of adhesion in friction because it does not readily show itself as an adhesive force normal to the surfaces. We have shown that this arises in the following way. If metal surfaces are pressed together, some plastic and elastic deformation will occur at the region of contact. Friction is always measured with the normal load still applied; the junctions are still there and must be sheared if sliding is to occur. On the other hand, the normal load must be removed before the adhesion can be measured. In the course of removing the normal load, the elastic stresses are released and thus peel or snap the junctions apart. Dr. Bikerman will find that Dr. K. L. Johnson (ref. 84) has described the effect of released elastic stresses on two asperities in adhesional contact. He has shown that infinite stresses are produced at the rim of the contact as the load is reduced, resembling a tensile stress applied to a sharp crack in a glue line. Unless a small amount of plastic flow can occur to blunt the tip of the crack, the joint will fall apart. Similarly, a friction junction will not show appreciable normal adhesion unless the material around the junction is reasonably ductile. Further, trapped slivers of oxide occurring in the friction junction may themselves act as sources of weakness when tensile forces act across them. Consequently, adhesion is likely only if the interface is reasonably clean and the junction material is sufficiently ductile to cope with released elastic stresses. These two criteria are satisfied in the cold welding of aluminium, in the adhesion of indium in air, and in the adhesion experiments of Mr. Buckley and Dr. Keller in high vacuum.

J. J. Bikerman (Horizons Incorporated, Cleveland, Ohio)

Two objections against Coulomb's theory of friction were raised during the discussion.

First, there was a repetition of Leslie's (1814) argument that a slider going up and down along a sinusoidal curve, without friction, would not lose any energy and that, consequently, the theory cannot account for the existence of frictional heat. This criticism is invalid. Every process imagined as a reversible process would continue indefinitely without liberating heat (at a constant temperature). Looking at figure 16, where body 1 is the slider and body 2 the support, assume that adhesional forces operate between the hills of the former and the hills of the latter. If the slider moves to the right, as many bonds will be broken between hill A and hill α, as will be simultaneously formed between hill B and hill β. Thus, no overall evolution of heat is possible, and it may be concluded that adhesion cannot be the cause of friction.

This conclusion would be as wrong as that of Leslie. His remark and the present sketch presuppose a degree of regularity that renders a reversible process possible, at least under ideal circumstances. In reality, surface roughness is irregular, and an ideal sinusoidal path is just as unreal as the exact compensation by the increasing $B\beta$ contact for the decreasing $A\alpha$ contact. Because of the random nature of rugosity, both adhesion and surface roughness are equally able to account for the heat liberation. The real difference is not between adhesion and roughness, but between reversible and irreversible events.

The second objection refers to the absolute values of the coefficient of friction $f=F/N$ which, according to our views, is equal to tan φ; F is the force needed to push the slider along the support; N is the normal force on the slider, and φ is the maximum angle of slope on the tall hills. Since the literature mentions values of f equal to 2, for example, it would be necessary to admit that the slope of the above hills may be 63°, which appears unlikely.

To comprehend the frailness of this argument, it should be remembered that the equation $f = \tan \varphi$ is valid only as long as "the" law of friction is valid, i.e., as long as f is independent of N and of the area of contact and, it may be added, is reproducible in repeated experiments. If even one of these three conditions is not satisfied, the ratio F/N is not equal to $\tan \varphi$ and, in fact, may assume any value between $-\infty$ and $+\infty$. A simple example is supplied by two solids in a vacuum of 10^{-11} torr. Let one of them be a ball and the other a plate; when the ball is placed (from above) on the plate, the two solids strongly adhere to each other. Let F_1 be the tangential force needed to overcome the adhesion, that is, to shift the ball along the plate. If a vertical upward force equal to the weight of the ball is applied to the latter, the resultant normal force becomes zero. The value of F_1 will not be significantly altered by this force; hence, the ratio F_1/N will be equal to infinity. Increase now the upward force by an infinitesimal amount so that the direction of the resultant normal force is upward, i.e., N acquires a negative sign. In this instance, the ratio F_1/N becomes very near to $-\infty$. Obviously, the existence of such values of the "coefficient of friction" (if this term is applicable here) is not a contradiction to Coulomb. On the other hand, the above example demonstrates that, if adhesion between slider and support is really present, the coefficient of friction in the usual sense does not exist.

D. Tabor (University of Cambridge, Cambridge, England)

Dr. Bikerman maintains that adhesion cannot be the source of friction for really smooth surfaces, since as many bonds will be simultaneously formed between hill B and hill β as are broken between hill A and hill α (fig. 16). In his view this means that energy cannot be dissipated. This argument is quite erroneous in relation to solids. The error becomes evident if bodies 1 and 2 are made of the same material. The interface between A and α for truly smooth surfaces is then an atomic plane. Even if sliding is restricted to this plane, work will be done in shearing A over α, and additional shearing work will be expended as B comes into contact with β and is itself sheared.

Consider, for example, an even simpler case—the contact between two bodies A and B, which are of the same material, perfectly smooth, and make atomic contact at their common interface (fig. 17). If A is slid over B to position A', the new thermodynamic situation according to Dr. Bikerman is identical with the old. This would imply that no work is involved in shearing a solid, which is quite wrong.

The truth is that thermodynamics cannot be applied to the strength properties of solids without further detailed consideration of the processes involved. It can be applied to elastic deformations, but it cannot be applied to sliding or shear which involves large displace-

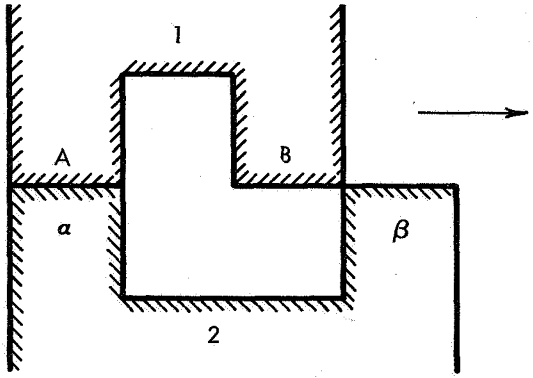


FIGURE 16.—Bikerman's sliding model.

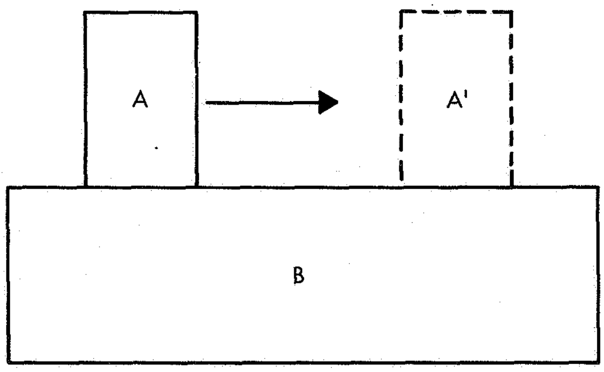


FIGURE 17.—Tabor's sliding model.

ments on the atomic scale. As one plane is moved over the other, the interatomic bonds are stretched and the lattice is distorted elastically. At some critical stress the bonds are finally broken; the strain energy is then dissipated as vibrations in the lattice, and this energy can never be recovered by the subsequent reformation of the next series of bonds. The process is basically nonconservative. The making and breaking of bonds in a solid by shear is fundamentally a dissipative process.

This leads to a second point. Adhesion between solids is as natural as the cohesion that holds atoms together in the solid; it arises from the same sort of forces. The problem of adhesion is not, in essence, why it occurs, but why it often appears to be absent. As mentioned previously, in the sliding of unlubricated surfaces, sensitive studies almost invariably show that material transfer has occurred from one surface to the other even if the adhesion cannot be detected as a normal adhesional force.

The reality of strong adhesion and the part it plays in the contact between surfaces has been very strongly emphasised by Dr. Bikerman. In *The Science of Adhesive Joints*, he points out that failure very rarely occurs in adhesive joints exactly between the glue and the adherend: it nearly always occurs in the adhesive. Dr. Bikerman discusses this in terms of the propagation of cracks, the molecular structure of the adhesive, the roughness of the adherend (resulting in a larger area of the glue-adherend interface), and other mechanisms. He also discusses it in terms of molecular forces. For example, he considers the molecular attraction between a strong solid and a weak solid and shows that the attraction between a "strong" molecule and a "weak" molecule will be smaller than between two identical strong molecules but greater than between two identical weak molecules. Consequently for a metal-polymer joint Dr. Bikerman concludes that

> the attraction between metal and polymer is greater than between polymer and polymer; hence the bond between two polymer molecules is more likely to be severed than any other bond: thus molecular forces favour rupture in cohesion in the adhesive layer

rather than adhesive failure at the polymer-metal interface. Broadly speaking, I favor this explanation, and readers will find it discussed in a little survey article I wrote 16 years ago on the mechanism of adhesion (ref. 85). This article also includes a brief account of the fallacy of attempting to explain the adhesion between solids in terms of conventional thermodynamics.

Dr. Bikerman has provided an excellent explanation of adhesion between different kinds of solids and how transfer can occur when the adhesion is broken. As he points out, even "substances such as polyethylene which contain no dipoles perfectly adhere to solids such as glasses, metals, polymers." Clearly he recognizes the reality of adhesion, even for such unfavorable combinations, when he is discussing adhesion. It seems strange that he should reject with such conviction—and with so much ingenuity—the reality of adhesion between sliding surfaces. How can adhesive forces not be present when polymer slides on metal, or when metal slides on metal, or when oxide slides on oxide? Adhesion may be large or small, or its effect may be swamped by other processes such as released elastic stresses or deformation losses; but to deny the role of adhesion in sliding friction as a matter of principle seems to me scientifically indefensible.

J. F. Archard (University of Leicester, Leicester, England)

Dr. Bikerman has suggested once again that the major component of friction arises from the force required to raise the weight of one sliding body up the slopes of the asperities upon the opposing surface (a concept usually attributed to Coulomb). On this argument the coefficient of friction is given by the mean slope of the asperities involved. Two comments can be made upon this line of argument. First, this process itself does not provide a dissipative mechanism which is an essential part of any acceptable theory of friction. The Coulomb process should provide alternating values of friction with positive and negative amplitudes corresponding to the positive and negative slopes upon the surface. Second, it should be possible to detect the upward and downward movement of the sliding body and to correlate this with the magnitude of the frictional force.

It is therefore appropriate to outline some experiments, along the lines discussed above, that have been performed in my laboratory by D. J. Whitehouse. (Full details of these experiments will be described in a separate publication.) In Mr. Whitehouse's work a smooth steel cylinder was slid against a similar rough cylinder using the crossed-cylinders friction machine (ref. 86). The experiments were performed at slow sliding speeds under conditions considered to be of boundary lubrication.

The following features were recorded:

(1) The coefficient of friction.

(2) The "ride," that is, the upward and downward movement of the upper cylinder (otherwise stationary) which was loaded against the lower moving cylinder.

(3) The surface finish of the rough cylinder, along a given line of the friction track, before and after the experiment.

The results of these experiments which are relevant to the present discussion can be summarized as follows. There exists a small variable component of the friction whose magnitude agrees closely with the slopes of both the recorded ride of the upper specimen and the long-wavelength components of the surface finish of the rough cylinder. To use an analogy drawn from electrical measurements, we conclude that the frictional force consists of a large dc component caused by adhesive mechanisms and a small fluctuating ac component contributed by the Coulomb mechanism. It is also worth noting that the experiments were made with crossed cylinders; with these conditions of contact, nominally at a single point, the magnitude of the Coulomb component of friction should be at its greatest value. With nominally flat rough surface, under the conditions described in Dr. Williamson's paper, the ride, and therefore the Coulomb component of friction, should be considerably reduced. On the disengagement side of any asperity contact, the downward movement of the upper surface should be arrested soon since the load would be taken by other asperity contacts.

E. I. Shobert (Stackpole Carbon Company, St. Marys, Pennsylvania)

Up to this point there has been very little discussion of the effects of the mechanical situation on the results of the various friction and wear measurements. I should like to underline the importance of the mechanical situation in some of the resulting effects by making four points.

The theories of wear and their relation to friction must be considered in the context of specific application. In applications where the wear rates are relatively small and the parts are expected to operate for thousands of hours, the effect of the wear debris on wear and friction will be entirely different than when two metals plow into each other for relatively few operations over relatively short times. I do not believe that we can generalize our concepts of friction, wear, and lubrication to cover these widely different situations. I have previously indicated the various ranges involved in friction and wear problems (ref. 87); but even in the narrow field of carbon brush applications, it can be noted that there may be anywhere from 10 particles of wear per centimeter of travel to one particle of wear per 10 meters of travel within the range of practical applications. Certainly, even within this range, the friction and wear phenomena must be different.

In the field of carbon brushes, we have run an experiment in which a split brush was placed in a massive holder against a massive copper ring. Friction was measured by the power loss in the system under two circumstances. In one case a small spring was placed between the two halves of the brush to press them lightly against the side walls of the holder. In this instance the friction was consistently low and uniform.

In the second case, the spring was removed and the brushes were permitted to move within the normal tolerances within the holder (about 0.005 in.). The friction was four to five times higher than that in the previous case.

If the situation were repeated, the friction rose and fell within minutes after the spring was removed or replaced. The surrounding atmosphere was room air held at constant humidity and with impurities removed by charcoal. A question is then raised about the friction of these combinations. It is my belief that in the case of the free brush the mechanical contact surface can increase to values well above those determined by the yield pressure; in the other case, the brush that is held against the massive holder will not wear into such large values of the mechanical surface because of the inherent eccentricity of the ring system. It is therefore apparent that these factors must be taken into account in friction measurements and in placing values on the coefficient of friction.

With a piece of chalk, I have demonstrated several mechanical aspects of friction phenomena. By holding the chalk near the end away from the blackboard and by considering its vibration as a cantilever spring whose elasticity is determined by my fingers, I show that the chalk rides smoothly on the blackboard for a large leading angle and for any trailing angle. For a small leading angle, the chalk will skip or jump, and the chatter frequency will be determined by the force with which I hold the chalk in this simple mechanical system. It is apparent that there is only a certain region of the leading angle within which chatter will take place. Outside of this angle it is impossible to make the chalk chatter, and within this angle it is possible to eliminate chatter only by holding the chalk solidly along its entire length. Thus this chatter effect has nothing to do with the question of whether the coefficient of friction decreases with velocity. The variation of the coefficient of friction with velocity can either be positive or negative in this situation, and the chalk will chatter within the critical angle. As I pointed out in previous publications (refs. 87 and 88), this chatter will take place only if the friction force can add to the energy of the system. This can take place when there is a normal component of the vibration or motion at the surface which will increase the friction force and the normal force while the amplitude is increasing during a particular cycle of the vibration. The angle within which chatter can take place is related to the coefficient of friction as follows:

$$f = \tan \theta$$

A different mode of chalk vibration can be achieved by holding the chalk within the critical angle mentioned above but with the fingers next to the blackboard. This sets up the high pitched squeal in which the chalk is vibrating longitudinally in a half-wave mode with the maximum amplitude at the center of the chalk. If I increase the energy given to this system by increasing the pressure and the speed, this amplitude can be increased until the chalk will break approximately in the middle and above my fingers where there is nothing touching it. While this represents a different mode of vibration, the principle is the same; that is, that chatter can take place only when the friction force operates to increase the amplitude of the vibration and the normal force at the same time.

The chalk is a simple mechanical system, but the principles are the same for more complicated systems such as clutches, brakes, violin strings, and musical saws. Internal elastic vibrations at the mechanical contact surface of large elastic bodies can influence friction and wear phenomena by jumping or galloping on a microscopic scale.

The fourth item I wish to mention concerns some of the effects that occur in the drastic wear situation that is involved in chalk on the blackboard or a situation such as two clean metals rubbing together.

This can be demonstrated by taking the side of a piece of chalk and passing it with heavy pressure over the blackboard in one direction. A wide, heavy band of chalk is left on the blackboard, but very little chalk dust falls away. On the other hand, if I repeat the motion in the same track, most of the chalk dust now falls down away from the contact surface in a visible cloud. In the first case, we have the adhesion of the chalk dust to the different material of the blackboard; and in the second case, we have the lack of adhesion of the chalk dust to itself on the second passage. It is interesting to note that calcium sulphate, which is the main constitutent of the chalk, can be an effective adjuvant to prevent drastic wear in carbon materials under clean conditions, while by itself and against itself, it is a very poor lubricant. This only illustrates the complexity of the situation and the importance of considering all of the material and environment factors in working with specific operational and practical problems.

M. E. Sikorski (Georgia Institute of Technology, Atlanta, Georgia)

Recent advances in the understanding of basic phenomena such as the effect of crystal structure on friction have led to the formulation of metal alloys with improved friction and wear characteristics.

Figure 18 shows another example of the importance of crystal structure on adhesion. The coefficients of adhesion are given for the platinum-cobalt alloy system as a function of cobalt content. As the amount of cobalt is increased, the crystal structure changes and so does the adhesion coefficient (ref. 89). It is possible that in some cases the transition from a mild to a severe wear regime with a subsequent failure of bearings may be traceable to composition changes in the bearing materials under prolonged use or severe pressure and temperature conditions.

In addition to the Pt–Co system, other relatively simple binary alloys characterized by complete solid solubility have been studied to uncover important parameters governing adhesion and friction. For example, the copper-nickel alloys, which retain the face-centered cubic structure for all concentrations, have been slid against iron and molybdenum to study the effect of low solid solubility of copper in these elements. Similarly, gold-silver alloys have been slid against iron to evaluate the effects of low solubility of silver in iron. The experiments showed that the adhesion was low when the concentration of the immiscible element in the alloy was high. In the above twist-compression tests, the specimens were placed in contact under the combined action of normal and tangential forces.

More recent results on copper-nickel alloys slid against themselves indicate an interesting correlation between the coefficients of adhesion and the work-hardening coefficients, as well as stacking fault energies

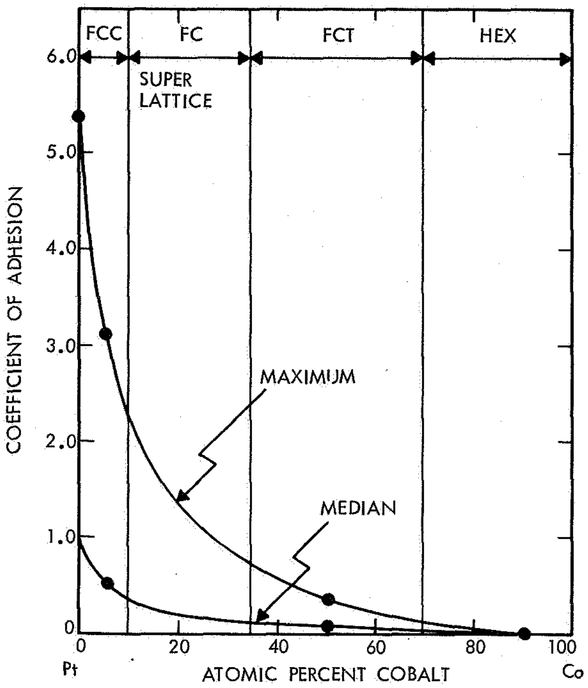


FIGURE 18.—Coefficient of adhesion vs percent cobalt for Pt–Co alloys.

of these alloys. Similar work is being carried out on silver-gold and silver-palladium alloys in an interdisciplinary effort to relate the material properties of these alloys to their adhesion and friction behavior (ref. 90).

J. A. Schey (IIT Research Institute, Chicago, Illinois)

Under Navy sponsorship Dr. Newnham and I have investigated friction and wear in unlubricated contact between typical aerospace structural materials and practical die materials used in deformation processing. The twist-compression test technique was used, and interface pressures were high enough to cause bulk plastic yielding in the specimen. While most of the alloys were too complex to allow direct correlation with earlier work on adhesion, some of the results may be used to test the validity of the solubility theory of adhesion.

It is necessary to make the unproven assumption that the behavior of the alloy follows that of the major constituent. As shown in figure 19, copper gave high friction in contact with a hot working die steel

(H12), probably reflecting the moderate solubility of copper in iron. The high solubility of copper in cobalt could explain the very high friction registered with a cobalt-bonded tungsten carbide (WC) anvil, and the total insolubility of copper in chromium may account for the very low and consistent friction measured against hard, chrome-plated tool steel anvils (CP). The same materials gave a different order of friction values with 70/30 brass, probably as a result of the zinc addition. The very low solubility of zinc in cobalt may account for the drop in friction with the tungsten carbide anvil; data are missing on the Cr end of the Zn–Cr system, and it is uncertain what could have caused friction. Unfortunately, similar correlations could not be found with aluminum as the workpiece material.

This work has also shown that while oxide films generally reduced friction and wear, they are not sufficient to mask the effect of the die-workpiece combination (fig. 20). Friction was higher in an industrial argon atmosphere than in air with all workpiece materials except a nickel-base superalloy (Hastelloy X). The difference was generally more marked when sliding against die steel (H12) than refractory (alumina) anvils. These results are not unexpected since the very high interface pressure combined with relative sliding is bound to break up oxide films and thus facilitate metal-to-metal contact. Nevertheless, it is rather surprising that the effect remains so marked even on the aluminum alloy (7075) which must have had a thin, natural oxide film even in the protected atmosphere.

R. W. Roberts (General Electric Research & Development Center, Schenectady, New York)

With reference to the remarks made by Dr. Gretz, it has been con-

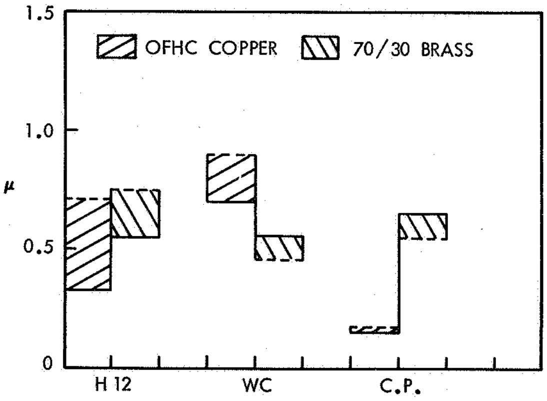


FIGURE 19.—Effect of copper alloying on friction with various anvils at room temperature and yield pressure.

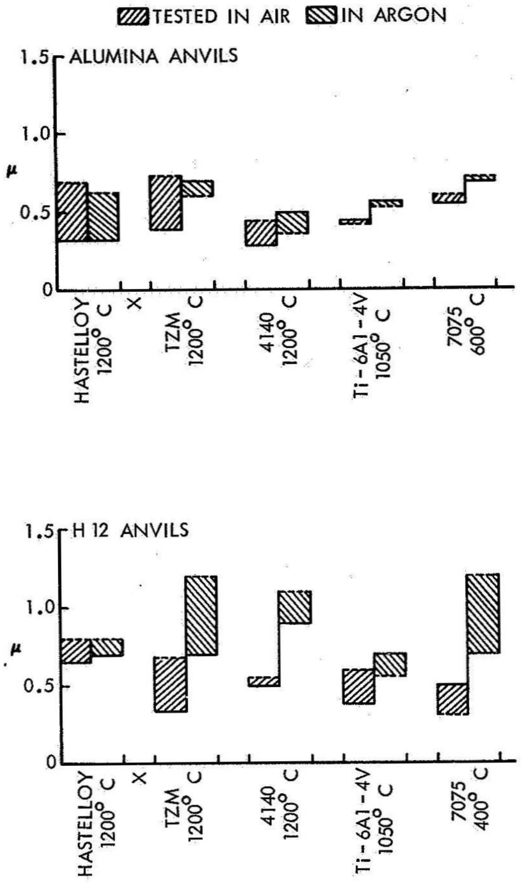


Figure 20.—Variation of friction with atmosphere in twist compression at elevated temperatures and yield pressure.

clusively demonstrated that iodine will function as a lubricant or a lubricant additive for a variety of metals. The iodine can be in the form of a vapor (ref. 91), a charge transfer complex (ref. 92), a solution in oil (ref. 93), and in chemical combination with a silicone or a long-chain hydrocarbon (ref. 94). It is also well known that many lamellar compounds including metal iodides are effective solid lubricants (ref. 95).

We have observed that dramatic reductions in coefficients of friction and wear occur when iodine is used as a lubricant additive for metals for which lamellar iodides exist, i.e., titanium, iron, nickel, cobalt, lead, and cadmium (ref. 96). Similar excellent results are obtained when iodine is used as a lubricant for alloys of these elements—stainless steels and the heat-resistant alloys of nickel and cobalt (refs. 96 and 97). In addition, X-ray diffraction study of surface films (ref. 98) and wear debris for several metals (ref. 94) reveals the presence of metal diiodides. Iodine does not appear to be an effective lubricant for metals that do not form lamellar iodides (ref. 96).

There are a number of ways to prevent oil from a diffusion pump from contaminating a vacuum system and Dr. Gretz has indicated a new one. If the possibility of oil contamination is a major concern in the design of an experimental apparatus, perhaps the best way to avoid it is not to use oil diffusion pumps. Ultrahigh vacuum can be routinely obtained using trapped mercury diffusion pumps, sputter-ion pumps, etc. (ref. 99).

The one-dimensional field-emission microscope using a wire has a major drawback. Magnification of only the circumference of the wire occurs and not of the length; thus most of the detail is lost. Several workers, however, are using the field-ion microscope to study phenomena of fundamental importance to adhesion. Ehrlich has recently measured the diffusion of tungsten atoms along tungsten single crystal planes (ref. 100) and the binding energy of tungsten atoms on tungsten single crystal planes (ref. 101). Müller and Nishikawa have elegantly demonstrated the use of field-ion emission in studying the adhesion of touching clean surfaces (ref. 102).

LECTURER'S CLOSURE

I wish to thank all of the many symposium participants who took the time to prepare these thoughtful discussions. They have contributed materially to meaningful consideration of the present status of the understanding of friction and adhesion and of the needs for greatly broadening that understanding through the interdisciplinary approach.

Most of the discussions were devoted to supplementing the information and concepts given in the paper. Having done this so effectively, they have generated no need for a specific reply on my part other than to express sincere appreciation for their excellent reinforcement of the need for and value of the interdisciplinary approach to this subject. However, Dr. Gretz, Dr. Bikerman and, to a minor extent, Dr. Schey have expressed disagreement with some of the concepts or information which I presented. I deeply appreciate the thoughtful replies already made to them by Dr. Tabor, at Mr. Ku's request, and I thoroughly agree with all of Dr. Tabor's comments. I will then merely try to supplement these.

Concerning Dr. Gretz's theories of friction, it seems that he becomes quite deeply entangled in the semantics of what is or is not friction, with logic somewhat suffering as a consequence. Describing that part of the resistance to sliding that is contributed by the shearing process as non-friction, and defining only that part due to asperity lifting and ploughing as friction, removes one very far from actualities. Friction, as understood by both the layman and the scientist or engineer and as defined by Webster, is "resistance to the relative motion of one body sliding, rolling or flowing over another with which it is in contact." Contrary to Dr. Gretz's impression, there is no requirement that it be associated with a normal force or even be a function of the normal force. However, the practical reality is that in all normal circumstances, the friction force is a linear or nearly linear function of the normal force. Any theory of friction must adequately explain this real situation.

The present state of the understanding of friction shows that this situation is due not only to asperity lifting (a very minor component of friction as demonstrated so well by the experiment described in Dr. Archard's discussion) and ploughing (normally a relatively small component), but also to the fact that the real area of contact between virtually all real-world surfaces increases linearly or nearly linearly with load. Thus experiments that hold this contact area constant and independent of load can readily lead to situations that are irrelevant to the real understanding of friction. It is not difficult to see how these can lead to actual errors in logic such as Gretz's thought that it is necessary to postulate that a normal force or stress must produce a shear force or stress, in order to explain the behavior of real friction arising from resistance to shear at the interface. No such unreal postulate is required; it is only necessary that the normal force influence the shear force, and this it does by influencing the real area to be sheared. The real shear stress remains unaffected. This same quirk of logic leads Gretz to conclude that friction and adhesion are independent processes because a shear force must by definition be independent of normal load. In the real world, where the area to be sheared varies with normal load, nothing could be farther from the truth.

I would also like to draw attention to a possible misunderstanding that could rise from reading that part of Dr. Gretz's discussion dealing with the clean surface. He highlights the well-known fact that oil diffusion pumps are an insidious source of contamination of supposedly chemically clean surfaces in vacuum, and in so doing implies (probably unintentionally) that the data obtained in vacuum by Buckley might have been subject to such error. Actually, Buckley recognized this possible source of error very early in his work on friction and as a result completely eliminated oil diffusion pumps from all his vacuum systems. His data that I cited, as well as those taken several years earlier, were from systems pumped by a combination of an ionization pump and a sorption forepump, together with a liquid-nitrogen or liquid-helium cryopumping coil in the system. He has described this pumping equipment in each of his papers and has discussed the subject in detail in reference 103.

Turning now to Dr. Bikerman's two discussions, it seems that he has been imprisoned in almost the same logic trap as has Dr. Gretz. He recognizes that adhesion does occur between chemically clean metal surfaces in vacuum where sizable values of adhesion coefficient are experienced. He insists, however, that the resulting resistance to sliding, produced by the shear strength of the contact area, is not friction because the friction coefficient is not independent of the normal force. As I stated above, there is nothing in the definition of friction that requires the coefficient of friction to be independent of normal force—it is merely that in most real-world situations it is thus. Is it then logical to conclude that in such real-world situations, when contaminating films are present and the coefficient of friction is independent of normal force, the contact area has no shear strength? Even Dr. Bikerman finds it hard to believe that this is so, since he states that the presence of a boundary lubricant substitutes "a viscous oil for the 'rare' air as the medium which must be sheared between the hills and the slopes of the valleys." Surely this viscous oil must offer more resistance to sliding than the "rare" air. I invite Dr. Bikerman to be fully objective now and substitute the usual thin film of oxide found on a metal surface in air for the thin film of "viscous oil." Will this not offer even more resistance as "the medium which must be sheared between the hills and the slopes of the valleys" if sliding is to take place? Is this resistance not friction? Further, the resistance it offers to sliding is proportional to the normal force; this comes about simply because the contact area over which it must be sheared is proportional to the normal force. I urge Dr. Bikerman to do his utmost to remain wholly objective in following this line of thought to its various conclusions, so that he may be the medium through which the "majority view" and the "minority view" can now be harmonized.

Turning to Dr. Schey's discussion on surface topography, a topic also discussed by Dr. Bikerman, I feel that Dr. Tabor's reply and Dr. Archard's discussion answer their points of concern quite thoroughly. However, I do believe that Dr. Schey's warning against drawing detailed conclusions from two-dimensional surface models without testing them against the statistical factors encountered in actual random three-dimensional surface topography is one to be heeded by all who do research in this field. The invitingly easy simplification of the two-dimensional model can offer traps to the unwary.

REFERENCES

1. Bailey, A. I.: Friction and Adhesion of Clean and Contaminated Mica Surfaces. J. Appl. Phys., vol. 32, 1961, p. 1407.
2. Bowden, F. P.; Moore, A. J. W.; and Tabor, D.: The Ploughing and Adhesion of Sliding Metals. J. Appl. Phys., vol. 14, 1943, p. 80.
3. Archard, J. F.: Contact and Rubbing of Flat Surfaces. J. Appl. Phys., vol. 24, 1953, p. 981.
4. Lodge, A. S.; and Howell, H. G.: Friction of an Elastic Solid. Proc. Phys. Soc. (London), vol. B67, 1954, p. 89.
5. Greenwood, J. A.; and Williamson, J. B. P.: Contact of Nominally Flat Surfaces. Proc. Roy. Soc. (London), vol. A295, 1966, p. 300.
6. Bowden, F. P.; and Tabor, D.: Friction and Lubrication of Solids. Pt. I. Oxford Univ. Press, 1950.
7. Holm, R.: Electrical Contacts Handbook. Springer-Verlag (Berlin), 1958.
8. Bailey, A. I.; and Courtney-Pratt, J. S.: The Area of Real Contact and the Shear Strength of Mono-molecular Layers of a Boundary Lubricant. Proc. Roy. Soc. (London), vol. A227, 1955, p. 500.
9. Adams, N.: Friction and Deformation of Nylon, I. Experimental. J. Appl. Polymer Sci., vol. 7, 1963, p. 2075.
10. Dyson, J.; and Hirst, W.: The True Contact Area Between Solids. Proc. Phys. Soc. (London), vol. B67, 1954, p. 309.
11. Kraghelsky, I. V.; and Sabelnikov, V. P.: Experimental Check of Elementary Law of Boundary Friction (Dry Friction). Proc. Conf. on Lubrication and Wear, IME, 1957, p. 247.
12. Ernst, H.; and Merchant, M. E.: Surface Friction of Clean Metals—A Basic Factor in the Metal Cutting Process. Proc. Special Summer Conf. Friction and Surface Finish, Tech. Press, MIT, 1940, p. 76; Friction—Selected Reprints, Am. Inst. Phys., 1964, p. 34.
13. Bowden, F. P.; and Tabor, D.: Friction and Lubrication of Solids, Pt. II. Oxford Univ. Press, 1964.
14. McFarlane, J. S.; and Tabor, D.: Adhesion of Solids and the Effect of Surface Films. Proc. Roy. Soc. (London), vol. A202, 1950, p. 224.
15. Buckley, D. H.: The Influence of Crystal Structure, Orientation and Solubility on the Adhesion and Sliding Friction of Various Metal Single Crystals in Vacuum (10^{-11} Torr). Adhesion or Cold Welding of Materials in Space Environments, ASTM, 1967, p. 248.
16. Anderson, O. L.: The Role of Surface Shear Strains in the Adhesion of Metals. Pt. I. Wear, vol. 3, 1960, p. 253.
17. Sikorski, M.: Correlation of the Coefficient of Adhesion with Various Physical and Mechanical Properties of Metals. Trans. ASME, vol. D86, 1963, p. 279.
18. Houwink, R.; and Salomon, G.: Adhesion and Adhesives. Elsevier Publ. Co. (Amsterdam), 1965.
19. Keller, D. J.: Adhesion between Solid Metals. Wear, vol. 6, 1963, p. 353.
20. Buckley, D. H.: The Influence of the Atomic Nature of Crystalline Materials on Friction. ASLE Trans., vol. 11, 1968, p. 89.
21. Roach, A. E.; Goodzeit, C. L.; and Hunnicut, R. P.: Scoring Characteristics of Thirty-Eight Different Elemental Metals in High-Speed Sliding Contact with Steel. Trans. ASME, vol. 78, 1956, p. 1659.
22. Spalvins, T.; and Keller, D. V.: Adhesion between Atomically Clean Surfaces. Trans. Vac. Met. Conf., 1962, p. 149.

23. Hume-Rothery, W.: Atomic Theory for Students of Metallurgy. Inst. of Metals (London), 1960.
24. Machlin, E. S.; and Yankee, W. R.: Friction of Clean Metals and Oxides with Special Reference to Titanium. J. Appl. Phys., vol. 25, 1954, p. 576.
25. da Vinci, L.: The Notebooks of Leonardo da Vinci. Compiled by E. MacCurdy. Reynal and Hitchcock, 1938.
26. Amontons, G.: De la Resistance Causée dans les Machines. Roy. Sci. (Paris), 1699, p. 206.
27. Coulomb, C. A.: Theorie dés Machines Simples. Mém. Math. Phys. Acad. Sci., vol. 10, 1785, p. 161.
28. Merchant, M. E.: The Mechanism of Static Friction. J. Appl. Phys., vol. 11, 1940, p. 230.
29. Bowden, F. P.; and Tabor, D.: The Area of Contact between Stationary and between Moving Surfaces. Proc. Roy. Soc. (London), vol. A169, 1939, p. 391.
30. Bowden, F. P.; and Tabor, D.: Mechanism of Metallic Friction. Nature, vol. 149, 1942, p. 197.
31. Bowden, F. P.; and Tabor, D.: Friction, Lubrication and Wear: A Survey of Work During the Last Decade. Brit. J. Appl. Phys., vol. 17, 1966, p. 1521.
32. Tabor, D.: The Hardness of Solids. Oxford Univ. Press, 1951.
33. Courtney-Pratt, J. S.; and Eisner, E.: The Effect of Tangential Force on the Contact of Metallic Bodies. Proc. Roy. Soc. (London), vol. A238, 1957, p. 529.
34. McFarlane, J. S.; and Tabor, D.: Relation between Friction and Adhesion. Proc. Roy. Soc. (London), vol. A202, 1950, p. 244.
35. Tabor, D.: Junction Growth in Metallic Friction: The Role of Combined Stresses and Surface Contamination. Proc. Roy. Soc. (London), vol. A251, 1959, p. 378.
36. Bisson, E. E.: Private communication to the author.
37. Bowden, F. P.; and Tabor, D.: Friction and Lubrication. Methuen & Co., Ltd. (London), 1967.
38. Tabor, D.: The Mechanism of Rolling Friction. II, The Elastic Range. Proc. Roy. Soc. (London), vol. A229, 1955, p. 198.
39. Flom, D. G.; and Bueche, A. M.: Theory of Rolling Friction for Spheres. J. Appl. Phys., vol. 30, 1959, p. 1725.
40. Merwin, J. E.; and Johnson, K. L.: An Analysis of Plastic Deformation in Rolling Contact. Proc. IME, vol. 177, 1963, p. 676.
41. Bridgman, P. W.: Studies in Large Plastic Flow and Fracture. McGraw-Hill Book Co., 1952.
42. Steijn, R. P.: The Friction and Wear of Plastics. ASM Tech. Rept. C6–19.3, 1966.
43. Rabinowicz, E.; and Shooter, K. V.: The Transfer of Metals to Plastics during Sliding. Proc. Phys. Soc. (London), vol. B65, 1952, p. 671.
44. van Rooyen, G. T.; and Backofen, W. A.: Friction in Cold Rolling. J. Iron & Steel Inst., vol. 186, 1957, p. 235.
45. Kelly, A.: Strong Solids. Ch. 1, Oxford Univ. Press, 1966.
46. Kragelskii, I. V.: Friction and Wear. Butterworths Sci. Publ., 1965, p. 236.
47. Hughes, E. L.; and Levenstein, H.: Mean Adsorption Lifetime of Rb on Etched Tungsten Single Crystals: Ions. Phys. Rev., vol. 113, 1959, p. 1029.
48. Mullins, W. W.: Solid Surface Morphologies Governed by Capillarity. Metal Surfaces: Structure Energetics and Kinetics. W. D. Robertson and N. A. Gjostein, eds., chap. 2. ASM, 1963.

49. Gibbs, J. W.: The Collected Works, Thermodynamics. Yale Univ. Press, 1948, p. 260.
50. Hirth, J. P.; and Pound, G. M.: Condensation and Evaporation: Nucleation and Growth Kinetics. Pergamon Press, 1963.
51. Bailey, G. L. J.; and Watkins, H. C.: Surface Tensions in the System Solid Copper—Molten Lead. Proc. Phys. Soc., vol. 53B, Pt. 5, 1950, p. 350.
52. Gretz, R. D.; and Pound, G. M.: Vapor Solid Nucleation. Tech. Rept. NSF Grant G21538, Carn. Inst. Tech., 1963.
53. Gretz, R. D.: Determination of Macroscopic Contact Angles in High Vacuum and the Relationship to Nucleation Theory. Surface Sci., vol. 4, 1966, p. 494.
54. Gretz, R. D.: Line-Tension Effect in a Surface Energy Model of a Cap-Shaped Condensed Phase. J. Chem. Phys., vol. 45, 1966, p. 3160.
55. Gurney, C.: Thermodynamic Relations for Two Phases Containing Two Components in Equilibrium under Generalized Stress. Proc. Phys. Soc., vol. 59, 1947, p. 629.
56. Furey, M. J.: The Action of Iodine in Producing Extremely Low Friction. Wear, vol. 9, 1966, p. 369.
57. Rolsten, R. F.: Iodide Metals and Metal Iodides. John Wiley & Sons, 1961.
58. Gretz, R. D.; Jackson, C. M.; and Hirth, J. P.: Surface Sci., vol. 6, 1967, p. 171.
59. Owens, R. S.; and Roberts, R. W.: Boundary Lubrication of Chrome Steel. Wear, vol. 7, 1964, p. 513.
60. Lancaster, J. K.: Anisotropy in the Mechanical Properties of Lamellar Solids and Its Effect on Wear and Transfer. Wear, vol. 9, 1966, p. 169.
61. Yang, L.; Horne, G. T.; and Pound, G. M.: Effect of Elastic Strain on the Electrode Potential of Metals. Physical Metallurgy of Stress Corrosion Fracture, vol. 4. T. N. Rodin, ed., Interscience Publ. Co., 1959, p. 29.
62. Gretz, R. D.: J. Vac. Sci. Tech., in press.
63. Good, R. H., Jr.; and Müller, E. W.: Handbuch der Physik. vol. 21. Springer-Verlag (Berlin), 1956, p. 176.
64. Gretz, R. D.: Appl. Phys. Ltrs., submitted.
65. Bryant, P. J.; and Gosselin, C. S.: Extreme Vacuum Technology Including Cryopump, Getter Ion Pump, Diffusion Pump and Pressure Standardization Studies. NASA CR-72066, July 1967.
66. Gretz, R. D.: Theory and Methods of Field-Emission Microscopy. High Temperature High Resolution Metallography. Ch. 3, H. I. Aaronson and G. S. Ansell, ed., Gordon and Breach, 1967.
67. Andrews, H. I.: The "Secondary Conditioning" Effect of Surfaces as Observed on Steel Rails. Wear, vol. 6, 1963, p. 262.
68. Johnson, R. P.; and Schockley, W.: Electron Microscopy for Filaments: Emission and Adsorption by Tungsten Single Crystals. Phys. Rev., vol. 49, 1936, p. 436.
69. Creighton, D. L.; and Hoenig, S. A.: Observations of the Early Stages of Brittle Fracture with the Field-Emission Microscopy. Trans. AIME, vol. 233, 1965, p. 1368.
70. Johnson, K. I.; and Keller, D. V., Jr.: J. Vac. Sci. Tech., vol. 4, 1967, p. 115.
71. Gretz, R. D.: Appl. Phys. Ltrs, submitted.
72. Bowden, F. P.; and Tabor, D.: The Friction and Lubrication of Solids, Pt. II, Oxford Univ. Press, 1964.
73. Keller, D. V.: Adhesion Between Solid Metals. Wear, vol. 6, 1963, p. 353.
74. Gregg, S. J.: The Surface Chemistry of Solids. Reinhold Publ. Co., 1961.

75. **Ehrlich, G.**: Metal Surfaces, ASM, 1963, p. 243.

76. **Haneman, D.**; and **Grant, J. T. P.**: Atomic Mating of Germanium Surfaces. Physical Electronics Conference, MIT, Mar., 1967.

77. **Buckley, D. H.**: The Influence of Crystal Structure, Orientation and Solubility on the Adhesion and Sliding of Various Metal Single Crystals in Vacuum (10^{-11} Torr). Adhesion or Cold Welding of Materials in Space Environments, ASTM, 1967.

78. **Holm, R.**: Electric Contacts Handbook. Springer-Verlag (Berlin), 1958.

79. **Johnson, K. I.**; and **Keller, D. V., Jr.**: Effect of Contamination on the Adhesion of Metallic Couples in Ultra-High Vacuum. J. Appl. Phys., vol. 38, 1967, p. 1896.

80. **Johnson, K. I.**; and **Keller, D. V., Jr.**: Factors Affecting the Adhesion of Titanium and Molybdenum Couples. J. Vac. Sci. & Tech., Dec. 1966.

81. **Rabinowicz, E.**: Friction and Wear of Materials. John Wiley & Sons, 1965.

82. **Bikerman, J. J.**: The Science of Adhesive Joints. Acad. Press, 1961.

83. **Mason, W. P.**: Adhesion Between Metals and its Effect on Fixed and Sliding Contacts. ASLE Trans., vol. 2, 1959, p. 39.

84. **Johnson, K. L.**: A Note on the Adhesion of Elastic Solids. Brit. J. Appl. Phys., vol. 9, 1957, p. 199.

85. **Tabor, D.**: Basic Principles of Adhesion. Reports on Progress, Appl. Chem., vol. 36, 1951, p. 621.

86. **Archard, J. F.**: A Crossed Cylinders Friction Machine. Wear, vol. 2, 1958, p. 21.

87. **Shobert, E. I.**: Carbon Brushes. Chem. Publ. Co., 1965.

88. **Shobert, E. I.**: Carbon Brush Friction and Chatter. Power & Apparatus, vol. 30, 1957, p. 268.

89. **Sikorski, M. E.**; **Kleiza, V.**; and **Courtney-Pratt, J. S.**: To be published.

90. **Sikorski, M. E.**; and **Bailey, J. A.**: To be published.

91. **Rowe, G. W.**: Vapor Lubrication of Titanium and Zirconium. Brit. J. Appl. Phys., vol. 7, 1956, p. 152.

92. **Roberts, R. W.**; and **Owens, R. S.**: Titanium Lubrication. Nature, vol. 200, 1963, p. 357.

93. **Furey, M. J.**: The Action of Iodine in Producing Extremely Low Friction. Wear, vol. 9, 1966, p. 369.

94. **Owens, R. S.**; **Roberts, R. W.**; and **Brown, E. D., Jr.**: Iodine in Lubrication. GE Res. & Dev. Ctr. Rept. 67–C–350, Sept. 1967, and Proc. Conf. on Friction and Wear, IME, 1967.

95. **Haltner, A. J.**: Sliding Behavior of Some Layer Lattice Compounds in Ultrahigh Vacuum. ASLE Trans., vol. 9, 1966, p. 136.

96. **Roberts, R. W.**; and **Owens, R. S.**: New Lubricants for the Heat Resistant Alloys. Friction and Lubrication in Metal Processing. F. F. Ling, R. L. Whitely, P. M. Ku, M. B. Peterson, eds., ASME, 1966, p. 244.

97. **Owens, R. S.**; **Roberts, R. W.**; and **Barnes, W. J.**: Iodine as in Extreme Pressure Lubricant Additive. Wear, vol. 9, 1966, p. 79.

98. **Baldwin, D. J.**; and **Rowe, G. W.**: Lubrication at High Temperatures with Vapor-Deposited Surface Coatings. J. Basic Eng., ASME, vol. 183, 1961. p. 133.

99. **Roberts, R. W.**; and **Vanderslice, T. A.**: Ultrahigh Vacuum and Its Applications. Prentice-Hall, Inc., 1963.

100. **Ehrlich, G.**; and **Hudda, F. G.**: Atomic View of Surface Self-Diffusion: Tungsten on Tungsten. J. Chem. Phys., vol. 44, 1966, p. 1039.

101. EHRLICH, G.; AND KIRK, C. F.: Binding and Field Desorption of Individual Tungsten Atoms. J. Chem. Phys., in press.

102. MÜLLER, E. W.; AND NISHIKAWA, O.: Atomic Surface Structures of the Common Transition Metals and the Effect of Adhesion as Seen by Field Ion Microscopy. Adhesion or Cold Welding of Materials in Space Environments, ASTM, 1967.

103. JOHNSON, R. L.; BUCKLEY, D. H.; AND SWIKERT, M. S.: Studies of Lubricating Materials in Vacuum. Proc. USAF–SwRI Aerospace Bearing Conference, P. M. Ku, ed., 1964, p. 60.

Wear

J. F. ARCHARD
University of Leicester
Leicester, England

The development of fundamental knowledge of wear is reviewed with special emphasis upon the unlubricated wear of metals. Against this background selected topics are discussed in further detail. One aspect which has received some attention is the classification of wear, and some classifications are critically reviewed in the light of the experimental evidence of wear studies.

Analytical theories of wear have a value in providing a more exact statement of assumed mechanisms of wear and also provide a basis for comparison between theory and experiment; values of the wear factor, K, of the theories are deduced from such comparisons. There is a need for physical explanations of these K factors which can be confirmed by subsidiary experiments. With a few combinations of materials, the mechanisms of wear are well understood; but the extent to which these same mechanisms apply to a wider range of materials is questioned. Moreover, most of the evidence is drawn from experiments with metallic materials, whereas one might expect the bonding within the material to exercise a major influence upon its wear behavior.

IN SETTING OUT THE OBJECTIVES OF THIS PAPER, the Steering Committee has asked for a broad critical review of the subject. It has also suggested that the discussion be in depth rather than all-inclusive and provocative rather than conservative. Finally, it suggested that the limits of our knowledge of wear be defined where possible.

This is a challenging and interesting task, but it presents many problems. It must be feared that what the author considers to be critical and provocative, the reader may regard as perverse and idiosyncratic. By way of illustration, on another occasion a lecturer of my acquaintance delivered a paper in which some of the limits of knowledge in his subject were quite clearly and interestingly presented. In the ensuing discussion he was rewarded by the comment that "the lecture was rather unusual in that it did not merely admit ignorance about the subject but proclaimed it."

This review has been consciously centered upon a discussion of the unlubricated wear of metals because the most coherent body of fundamental knowledge has been established for these conditions. In the following discussion a few aspects of the subject are selected for further examination and critical comment. Finally some specific conclusions and opinions are offered for further discussion.

GENERAL WEAR STUDIES

The history of wear studies, unlike the study of friction, contains few early classical papers. Because of the inherent difficulties and complications of the subject, most of our knowledge is based upon work carried out during the past 20 years when the fairly sophisticated techniques of modern physical science have been applied to the subject. However, some earlier landmarks in the development of the subject should be noted. Fink (ref. 1) laid stress upon the role of oxidation in the dry wear of steels, and it will be noted that this is a theme to which we shall return later in this paper. In other earlier work by Mailander and Dies (refs. 2 and 3) and by Siebel and coworkers (refs. 4 and 5), the complicated wear behavior of steels was studied and attempts were made to account for changes in the magnitude and character of wear with changes in load and speed.

Undoubtedly the most notable early work in the study of wear was that of Ragnar Holm (ref. 6). Perhaps his most important contribution was a logical basis for the assessment of the magnitude of wear. He showed that when the wear rate was expressed as the volume removed per unit sliding distance (which has the dimensions of an area), the severity of the wear rate might be assessed by comparing it with the true area of contact.

A fresh approach to wear studies was made by Burwell and Strang (ref. 7). The wear of a steel pin rubbing on a flat steel disk under a chemically inert lubricant (hexadecane) was studied; the rate of wear was followed by noting the size of the wear scar on the pin as wear proceeded. Some stress was laid upon the need to arrange the rubbing surfaces in the same orientation when they were reassembled after each observation. It was found that, in general, the rate of wear was proportional to the load and independent of the area of apparent contact. The major deviation from these rules occurred as the pressure over the apparent area of contact approached a value of one-third of the hardness; the wear rate then increased rapidly to a high value. Figure 1 shows a typical set of results from Burwell and Strang's experiments.

Burwell and Strang suggest that this rapid increase in the wear rate might be explained by the fact that the subsurface regions of plastically deformed material, associated with adjacent asperities, could merge when the pressure over the apparent area of contact approaches a value of $H/3$. This could result in an increased area of contact and a lowering of the shear strength of the peaks. Clearly this is a question of some fundamental importance and one worthy of

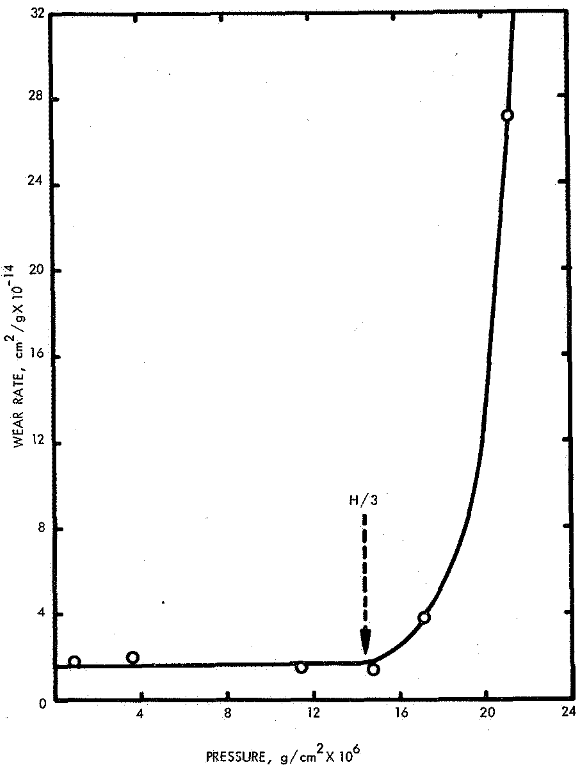


FIGURE 1.—Typical results of Burwell and Strang (ref. 7). The wear rate, expressed as volume removed per unit sliding distance per unit load, is plotted as a function of the pressure over the apparent area of contact.

closer study by other investigators. This is particularly true because in subsequent experimental work Dorinson and Broman (ref. 8) have failed to confirm the form of the results shown in figure 1 but, on the other hand, some theoretical justification for them has been provided (ref. 9).

Archard and Hirst (ref. 10) adopted a somewhat different approach. Their studies were concerned with the wear of a wide range of material combinations (mainly metallic materials) under unlubricated conditions. The results of a few thousand wear experiments with about a hundred different combinations of materials were summarized as follows. At the start of an experiment, the wear rate was not constant but changed as the experiment proceeded; during this period the surface conditions were changing, but eventually they became stabilized.

Under these stabilized conditions the wear rate was constant (i.e., the wear increased linearly with time) and was independent of the apparent area of contact. With most materials the relation between wear rate and load was close to direct proportionality.

As a broad classification, the wear of metals was divided into two types which were called "mild wear" and "severe wear." In severe wear, the contact is metallic, the surfaces are deeply torn, and the worn debris consists of metallic particles up to a fraction of approximately 1 millimeter. In mild wear, the surfaces remain smooth and are usually protected by surface oxide layers generated in rubbing. The worn debris consists of small particles down to a few hundred angstrom units or smaller. Although this classification of wear has been quite widely adopted, it has been criticized as being arbitrary; however, the alternative is to classify wear in terms of the mechanism by which material is removed. The problem here is to decide what the mechanism is.

The classification of wear into "mild" and "severe" is primarily a distinction in the scale of size; it suggests that an event occurring at a rubbing surface can typically have a size of the order of 10^{-2} cm (severe wear), or it can be very much smaller indeed (mild wear). From earlier measurements of the contact resistance between nominally flat surfaces (refs. 6 and 11), it had been deduced that the size of the true areas of contact was of the order of 10^{-2} cm. It seemed possible that these regions might be subdivided into smaller areas of true contact (fig. 2) and in this way one might account for the two types of wear; mild wear when the damage is confined to the individual areas of contact and severe wear when the damage grew to encompass the whole contact region. There is certainly ample evidence from simple experiments in unidirectional sliding to show the growth of severe wear from the coalescence of smaller regions of damage (ref. 12). However, with metals a change in the magnitude of the damage from 10^{-5} cm or 10^{-6} cm to 10^{-2} cm, for example, can easily correspond to a transition from oxide film disruption to severe metallic welding. Therefore, in discussions of the mechanisms of mild and severe wear and the transition from one mechanism to the other, the influence of the protective oxide films has received the major attention; but the possible role of surface topography, as illustrated in figure 2, should not be forgotten.

It is worth noting that these ideas about scales of size had an influence upon the development of the technique of Dyson and Hirst (ref. 13), which has confirmed the existence of the subdivided regions of figure 2. Models of surfaces having asperities of more than one scale of size, such as that of figure 2, have also played a part in theories of friction.

Both Burwell and Strang and also Archard and Hirst (ref. 10) drew upon the earlier work of Holm (ref. 6) and the studies of friction by

Bowden and Tabor (ref. 11) to provide an explanation of their results. It was assumed that the magnitude of wear was related to the true area of contact. From this, analytical theories of wear were developed. About the same time a pioneer study of transfer between rubbing surfaces was made by Rabinowicz and Tabor (ref. 14) using radioactive tracers. They showed that transfer consisted of discrete particles and that the volume of transferred material per unit sliding distance was approximately proportional to the load. This also revealed the important distinction between wear and transfer. In most wear studies, one is concerned with the removal of material from both rubbing surfaces. Transfer implies a stronger bond between the transferred fragment and the surface to which it adheres than that which existed between the fragment and its parent body. The mechanisms by which loose particles are produced were thus shown to require more detailed study. Later developments, discussed below, revealed the role of transfer as one stage of a wear mechanism.

Early studies of wear also revealed one other important feature of the subject. As shown in table 1, wear rates found with common engineering materials under unlubricated conditions can cover a range of approximately 10^5 while the values of the coefficient of friction cover a range of only 3. Moreover, high values of wear are not necessarily associated with high values of friction. Any theory of wear with claims to wide applicability must take this into account.

DETAILED STUDIES OF WEAR

We examine now experiments in which the wear of particular combinations of materials has been studied and show how this work is of crucial importance in our understanding of wear.

In the earlier work the major comment upon the detailed processes of wear was the classification of wear mechanisms into severe and mild. There seems to be an obvious resemblance between the phenomenon of severe wear and the ideas of friction developed in the earlier work of Bowden and Tabor (ref. 11). In this work the role of adhesion at the areas of true contact was emphasised and, insofar as metals were involved, welding was held to play a major role. However, welding does not produce the loose metallic debris which is a feature of severe wear.

The combination of materials which has been studied in the greatest detail is the wear of a free-machining 60/40 brass rubbing on hardened steel or Stellite. Under severe wear, the choice of the harder member is not important because its wear will be negligible. Also, under sufficiently heavy loads, when a brass pin rubs against a rotating ring of hardened material, a layer of brass is transferred to the ring and the wear of the brass is severe. Under these conditions, one has a wear situation that appears to be typical of the severe wear of a soft metal rubbing against a harder one. The particular brass was chosen because its wear behavior was exceptionally reproducible and because the wear rate was proportional to the load (ref. 10). (This result is forecast by simple theories of wear discussed below.) These wear characteristics have been recently confirmed by cooperative experiments carried out in a number of laboratories (ref. 15).

In the experiments of Kerridge and Lancaster (ref. 16), the brass pin was made radioactive and used to follow the history of the material from its original state as part of the pin to its final state as worn debris. In one experiment (fig. 3), an irradiated pin was used and a layer of radioactive brass was established on the ring. When this was replaced by a nonradioactive pin, the radioactive layer was worn away and replaced by nonradioactive material. No radioactive material appeared on the pin, showing that back transfer did not occur. In another

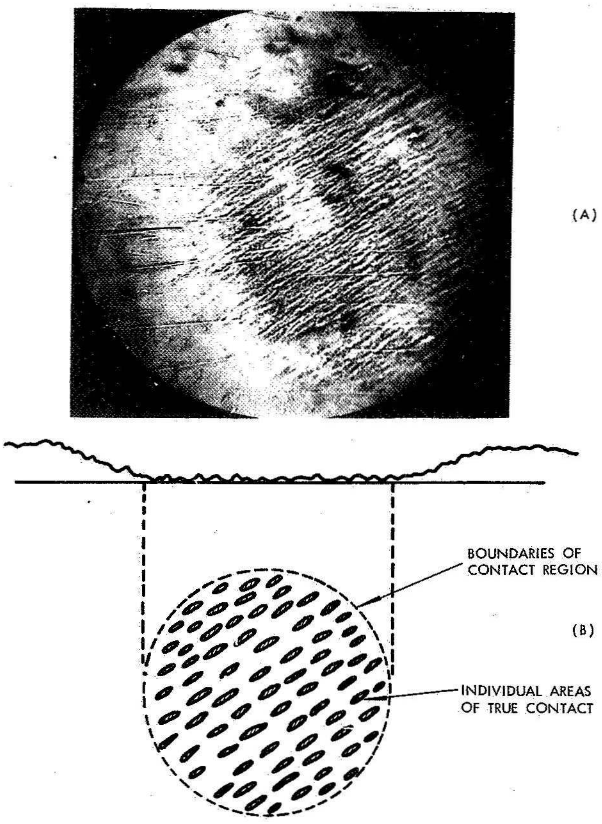


FIGURE 2.—Contact between surfaces. (A) Areas of contact revealed by the method of Dyson and Hirst; (B) Simplified representation of the contact region.

TABLE 1.—*Wear Rates, Coefficients of Friction and Values of K*[a]

Material	Wear rate, cm³/cm	Coefficient of friction	K
Mild steel on mild steel	1.57×10^{-7}	0.62	7.0×10^{-3}
60/40 leaded brass	2.4×10^{-8}	0.24	6.0×10^{-4}
PTFE	2.0×10^{-9}	0.18	2.5×10^{-5}
Stellite	3.2×10^{-10}	0.60	5.5×10^{-5}
Ferritic stainless steel	2.7×10^{-10}	0.53	1.7×10^{-5}
Polythene	3.0×10^{-11}	0.65	1.3×10^{-7}
Tungsten carbide on tungsten carbide	2.0×10^{-12}	0.35	1.0×10^{-6}

[a] Load, 400g; speed, 180 cm/sec. Rings are hardened tool steel except where stated otherwise.

experiment, the procedure was reversed. A nonradioactive pin was used to establish a layer on the ring. When this pin was replaced by an irradiated one, the layer became radioactive. The worn debris produced immediately after this change was nonradioactive, and its radioactivity increased only in accord with the rise in radioactivity of the transferred layer. Throughout these experiments, the wear rate was constant and equal to the equilibrium rate of wear obtained with this combination of materials in a normal wear experiment.

Further experiments showed that the wear was a two-stage process; the transfer from the pin to the ring followed by the removal of transferred material as worn particles. The transfer from the pin to the ring occurred as discrete particles, and the material was removed from the ring as aggregates of about 50 such transferred particles. Figure 3 also shows that the initial rate of increase of ring radioactivity, which corresponds to the rate of transfer, is equal to the rate of wear. Thus, it appears that transfer of material from pin to ring is the first stage of the wear process, and that the rate of transfer determines the rate of wear.

The importance of this work is not that the detailed conclusions have a universal application, but rather it has shown for the first time the complete details of one stage of the multistage wear process in which transfer by welding represents only the first stage. The wear of many combinations of materials involves transfer by welding and, on one system of classification, this has been classed as adhesive wear. The extent to which the severe wear of brass rubbing on hardened steel is typical of adhesive wear will be discussed later.

The effect of speed upon the severe wear of 60/40 brass rubbing on hardened steel has been investigated by Haymana (ref. 17) and by Steijn (ref. 18). The influence of changes in both speed and ambient temperature has been studied by Hirst and Lancaster (refs. 19 and 20).

In general, the wear rate, at any given value of the load, falls with increasing speed up to about 100 cm/sec, where the wear rate starts increasing with speed. The latter has been shown to be a consequence of frictional heating, which has two major effects upon the temperature of rubbing specimens: bulk temperature is raised, and higher transient "flash temperatures" are generated at the true areas of contact. Increased wear rate at high speeds, it was shown, was due to thermal softening of the brass. Moreover, the influential temperature was the mean temperature of the rubbing surfaces and not the much higher transient flash temperatures.

The effect of the speed upon the wear rate of brass was also analysed in terms of the material transfer—the first stage of the wear mechanism. Variations in the transfer (and therefore variations in the wear) arose from changes in the number and thickness of the transferred fragments. It was found that below a speed of 1 cm/sec the fall in wear rate with increasing speed was caused, primarily, by a reduction in the surface area of the transferred fragments, their number and thickness remaining approximately constant. At higher speeds, above 1 cm/sec, the tendency to welding also decreased; the number of transferred fragments thus decreased with increasing speed. The increase in the wear rate due to thermal effects at the highest speeds was due to an increase in the surface area of the fragments and not to an increase in their number of thickness. The finding that the change in wear rate arises from changes in the surface area of fragments of relatively constant thickness, differs from the results of Rabinowicz (ref. 21) who has examined the shape of transferred material using radioactive tracers. He suggests that a typical particle has a characteristic shape (1.7×1.0×0.5) that is substantially independent of the size.

In a further investigation Lancaster (ref. 22) has studied the conditions under which the wear of this same combination of materials changes from the severe to the mild regime. The load, speed, and ambient temperature of the system were varied over a very wide range. It was shown that the transition was determined by the interaction of two opposing processes: first, the mechanism of severe wear exposes clean metal surfaces for a condition which favors its own continuation; and second, the exposure of the surfaces to the atmosphere leads to the formation of oxide films which, if sufficiently protective, lead to mild wear (ref. 23). At low speeds, light loads, and low temperatures, mild wear takes place when there is sufficient time between contacts, at any point of the surface, for a protective film to be established. A similar film is also established when the rate of oxidation is increased, either by raising the ambient temperature of the system or by increased frictional heating. The effect of frictional heating was examined in a manner similar to that used in the investigation of the severe wear of brass. Once again it was found that the influential temperature was the mean temperature of the surface. The transient flash temperatures appeared to have little direct influence.

Thus wear depends upon the combination of load, speed, and ambient temperature. A set of wear rate graphs, for a load of 3 kg, based upon the results of Lancaster, are shown in figure 4. At an ambient temperature of 20° C severe wear occurs over a range of speeds from approximately 10^{-2} cm/sec to 10^{3} cm/sec. As the ambient temperature is increased, the range of speeds over which severe wear occurs is reduced by the increased rate of oxidation.

Lancaster (ref. 22) also examined the surface layers of the brass pin in the mild wear regime. He found that when a transition to mild wear occurred, the brass layers which had been transferred to the steel ring were transferred back to the pin and the surface of the pin then acquired a black oxidized appearance. The structure of the subsurface regions were examined using such things as X-ray microanalysis and electron diffraction. It was found that the surface layers, to a depth of about 2×10^{-3} cm, consisted of mixed oxides including a large proportion of those derived from the ring material. Thus the wear of the harder ring, insignificant under conditions of severe wear, became comparable with, or even larger than, the wear of the pin under mild wear conditions.

Welsh (ref. 24) has carried out a similar study of wear for a range of plain carbon steels. The typical pattern of behavior, when the wear rate is plotted against the load, is shown in figure 5. The heavy full line shows the behavior of 0.52 percent carbon steel tempered to a hardness of 268 DPN. At light loads mild wear occurs, and at a load of about 100 grams the transition (T_1, fig. 5) to severe wear occurs. Severe wear then persists up to a load of about 2.5 kg (T_2, fig. 5) when the wear mechanism reverts to mild wear. It was found, by metallography and microhardness tests, that, above the T_2 transition, the surface layers of the specimens were hardened by a phase change. The value of the T_2 transition load was determined for a number of speeds and for steels of various carbon contents, and in each case the value of the flash temperature was calculated. Bearing in mind a number of uncertainties in these calculations, the overall pattern of results suggest that the surface hardening is caused by a martensitic transformation, and that is caused by the flash temperatures at the true areas of contact.

Figure 5 also shows the effect of tempering on the pattern of wear. As the hardness increases, the range of loads over which severe wear occurs shrinks until, at a hardness of between 360 and 436 DPN, mild wear occurs over the full range of loads.

Welsh also examined in more detail the wear behavior above the upper transition T_2. Specimens which had run, at loads greater than

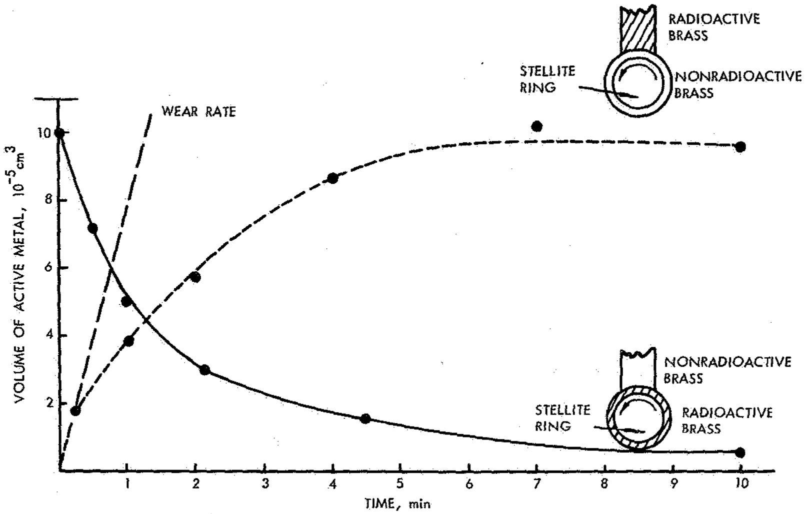


FIGURE 3.—Experiment of Kerridge and Lancaster (ref. 16). The lower curve shows the conversion of the transferred layer from radioactive to nonradioactive when radioactive pin is replaced by nonradioactive pin; the corresponding graph for the reverse experiment is also shown.

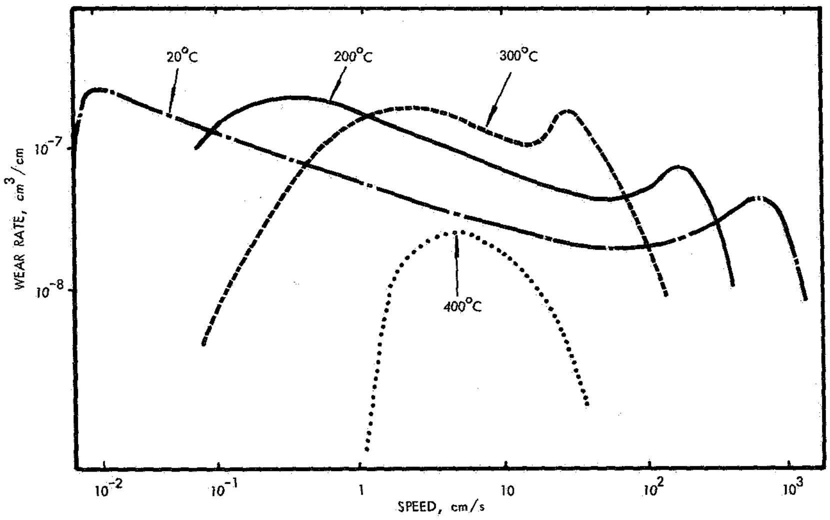


FIGURE 4.—Effect of speed upon the wear rate of brass (ref. 22). The wear rate of 60/40 brass rubbing on tool steel is shown as a function of the speed for various values of the ambient temperature. Load, 3 kg.

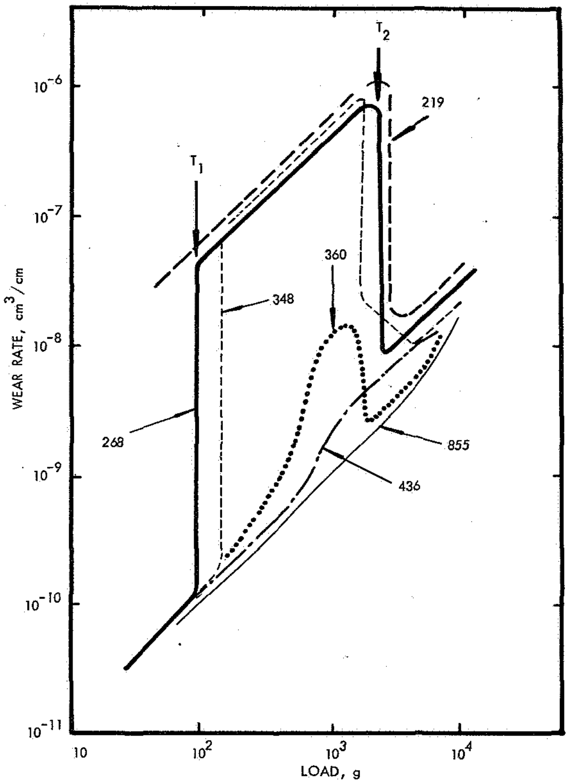


Figure 5.—Effect of heat treatment upon the pattern of wear for a 0.52 percent carbon steel (ref. 24). Sliding speed 100 cm/sec. The wear rate is shown as a function of the load and the hardness (DPN) of the material used in each series of experiments is indicated.

T_2, to attain equilibrium oxide films and hardening, were treated in two ways. For some specimens, the oxide film was removed by etching and then replaced by a new film, established by heating to one of a range of temperatures. For other specimens the oxide film was left intact, but the hardened surface layers were softened to different values by heat treatment in very pure argon. From these and other experiments it was found that at loads immediately above T_2 both hardening and an oxide film were necessary for the continued existence of the mild wear regime. At these loads mild wear continued in the presence of an oxide film, once a critical hardness had been exceeded. For the 0.52 percent carbon steel of figure 5, this critical hardness was approximately 400 DPN.

From Welsh's work the concept of a critical hardness emerges. The general relevance of this concept is illustrated by some experiments carried out in my laboratory.* Crossed cylinders of the same steel were rubbed, for a single traversal, at the slow speed of 0.05 cm/sec. As in Welsh's experiments (fig. 5) the effect of tempering to different values of the hardness was examined. At high values of the hardness there was little damage to the specimens except for small scratches. When the hardness was reduced, large-scale welding appeared at values of hardness in the region of 400 DPN. The appearance of welding was associated with a marked increase in the coefficient of friction, and this is shown by the results plotted in figure 6.

Although in Welsh's published work the theoretical flash temperatures were sufficient to cause phase hardening by well-known mechanisms, some metallurgical features involved in the friction and wear of steels remain obscure. There are indications that phase hardening can occur under conditions where the theoretical flash temperatures are insufficient, in terms of the normal metallurgical behavior, to cause hardening. Moreover, Welsh remarks that the low carbon steels harden more intensively during rubbing than is possible during normal heat treatment. Also, there is ample evidence from engineering practice and laboratory experiments, to suggest that steels containing a high percentage of chromium are particularly susceptible to damage (ref. 25, 26, and 27). When used as shafts in large bearings the damage can be catastrophic, and though palliative measures have been taken against such occurrences, the fundamental reasons remain obscure.

These and other metallurgical aspects of wear, such as the superposition in rubbing experiments of high pressures, high temperatures and transient conditions, are worthy of special study.

Mild wear is considerably more difficult to study than severe wear because of the small scale of size involved. Kerridge (ref. 28) examined the mild wear of annealed steel pins rubbing on hardened steel rings. This work preceded the investigation of the severe wear of brass (ref. 16), but it was in this study of steels that it was first found that metallic transfer from a soft pin to a hard ring could be the first stage of a wear process. With steels, however, it was suggested that the subsequent stages of the wear mechanism were the slow oxidation of the transferred layer and its removal as loose oxidized debris; the material worn away was then replaced by new transferred material. It was concluded that the oxidation and removal of the transferred material was the rate-determining process. The transferred material was found to be in a hardened condition, presumably as a result of the same factors later investigated by Welsh.

---

*I am indebted to P.J.H. Smith who carried out this work as part of his undergraduate project.

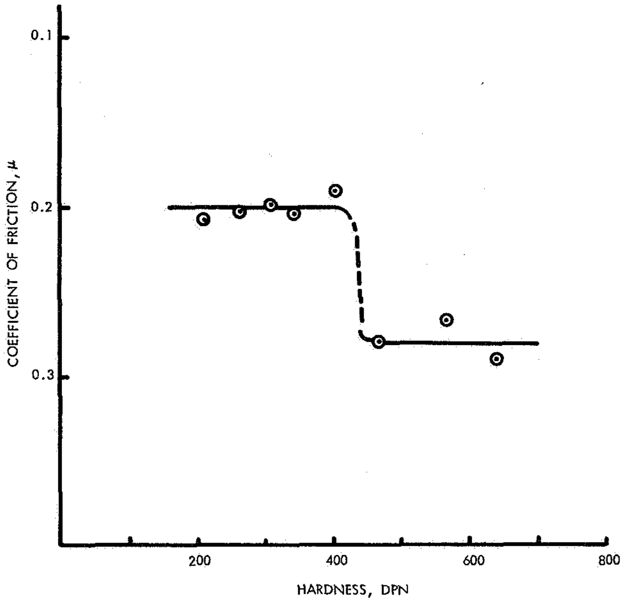


FIGURE 6.—Coefficient of friction μ for crossed cylinders of a 0.52 percent carbon steel as a function of hardness. Load, 24 kg. Speed, 0.05 cm/sec. Specimens lubricated with a plain mineral oil.

The wear of hardened steel rubbing on hardened steel was investigated by Archard and Hirst (ref. 29) using such techniques as radioactive tracers, electron microscopy, and electron diffraction. It was found that metal transfer occurred in the early stages of rubbing, but in the later stages, when the surfaces had run in, the wear process was primarily governed by oxides. This was demonstrated by noting the proportions of the radioactive transferred material which was removed by various methods, both mechanical and chemical, during the run-in. Electron microscopy (fig. 7) showed that the highest features on the run-in surfaces were loose aggregates of oxidized debris, and it was concluded that this would introduce a large element of abrasive wear.

This survey of some detailed studies of wear can now be reviewed against a wider background. The mechanism of the severe wear of brass has been very clearly established, but to the present author's knowledge, no other combination of materials has been studied in comparable detail. Whether the detailed mechanism revealed by Kerridge and Lancaster applies to the severe wear of other combinations of materials is a matter for speculation; but clearly, further experiments of this type are desirable. In particular, it should be noted that the absence of back-transfer in the experiments of Kerridge and Lancaster was a factor which greatly simplified their results. There is evidence to suggest that back-transfer will occur with some other combinations of materials rubbing under conditions of severe metallic wear.

Clearly those factors which favor welding should also favor severe wear. Coffin (ref. 30) showed that metals combinations which form a continuous series of alloys are prone to welding. De Gee (ref. 31) has performed experiments in an atmosphere of argon in which pins of gold-silver alloys were run against steel rings. As the percentage of gold was increased, the onset of severe wear occurred at an alloy composition of 40 atomic percent gold. This is also the alloy composition at which the solubility of iron in gold becomes significant. However, it was found* that the behavior in air was worse than in argon; and even silver, in which iron has negligible solubility, ran under severe wear conditions. Clearly, further experiments are required.

In an attempt to relate severe metallic wear to physical properties Machlin and Yankee (ref. 31) suggested that the important parameter was the ratio of the work of adhesion to the strength of the weaker material. This ratio should determine the tendency exhibited by a given combination of materials towards welding and the formation of transferred particles. Rabinowicz (ref. 33) considers the problem of the production of a loose wear particle and suggests that the influential factor is the ratio of surface energy E_s to hardness H. He suggests that the maximum residual stress which can be stored in an incipient wear particle is related to H and a particle can separate when this stress exceeds the surface energy required to create the particle. On the basis of this argument the linear dimensions of a loose worn particle should be proportional to E_s/H. He considers a wide range of materials and demonstrates an approximate relationship between particle size and E_s/H. However, the correlation within one class of materials (i.e., metals) is much poorer than for the whole range of materials. It has also been noted that a further difficulty is that of obtaining reliable and meaningful surface energy data. Thus, the surface energy concept, although an interesting one, is difficult to apply to detailed mechanisms of wear.

It might be expected that Lancaster's arguments, noted above (ref. 22), about transition from mild to severe wear, would apply to other combinations of materials. Perhaps the need here is that experiments with other materials be combined with a more quantitative

---

* DeGee, A. W. J.; and Begelinger, A.: to be published (see Saloman, G., Proc. Inst. Mech. Engrs., vol. 180, Pt. 3K, 1965–66.)

FIGURE 7.—Electron micrograph of worn surface of high-speed tool steel (ref. 29). Load, 1.25 kg. Speed, 66 cm/sec. The white patches are loose wear debris and the length of the shadows indicates their height compared with other surface features.

approach to the theory. This is also relevant to theories of the mechanism of mild wear by removal of oxide films which is discussed in more detail below.

Finally, some comment upon the broad patterns of behavior, such as those shown in figures 4 and 5, is relevant. Hirst (ref. 34) has suggested that the more limited range of results obtained for other materials will probably fit into the full behavior already recorded. He suggests that the main difference between the wear of different materials may be in the extent to which their behavior is affected by the controlling variables, rather than in differences in the form of the pattern. Clearly, once again, the need is for more information.

ANALYTICAL THEORIES OF WEAR

All theories which forecast the magnitude of wear rate start from the concept of a true area of contact; that is, they are based upon the fact that surfaces will wear at those places where they touch.

We first outline the simplest theories of wear derived from the work of Holm (ref. 6), Burwell and Strang (ref. 7), and Archard (refs. 35 and 36). The essential concept is that the worn volume, V, produced in sliding a distance, L, can be related to the true area of contact, A. It is usually assumed that the unit event which must be considered is the contact of two asperities on the opposing surfaces. Such a unit event, as idealized in figure 8, is concerned with the establishment of an area of contact which, for the sake of convenience, is considered to be a circle of radius, a, and area $\delta A = \pi a^2$.

If such an event leads to the production of a worn particle one assumes that the depth of material removed is proportional to a. This, for example, would be true if the worn particles were roughly equiaxed lumps of material. (Some experimental justification for this assumption is to be found in reference 21). Thus the volume, δV, of the worn particle is proportional to a^3, and the sliding distance, δL, in which this event has occurred will be proportional to the size of the contact a (fig. 9).* Expressing wear rate in terms of volume removed per unit sliding distance, the contribution of this event to the total wear rate is $\delta V/\delta L$ and is proportional to a^2, (i.e., to δA). Thus the contribution of each event to the total wear rate is proportional to its contribution to the total true area of contact. More specifically, if one assumes that the particles are hemispherical, of radius a, and that $\delta L = 2a$, then

$$\frac{\delta V}{\delta L} = \frac{\delta A}{3}.$$

Summing for all the areas of contact, the total wear rate is

$$\frac{V}{L} = \sum \frac{\delta V}{\delta L} = \frac{K}{3} \sum \delta A = \frac{K}{3} A = \frac{KW}{3H}, \tag{1}$$

where H is the hardness, and the summation implies continuous support for the load W. The last part of equation 1 holds when the area of contact, A, is formed by plastic deformation. Moreover, we have introduced the constant of proportionality K, called the wear coefficient. This theory can be applied to wear or to transfer, depending upon whether we regard each event as producing a loose wear particle or a transferred particle.

This theory provides an analysis of wear which parallels the earlier discussion of friction by Holm (ref. 6) and Bowden and Tabor (ref. 11). It is argued that both the frictional force and the wear rate are proportional to the true area of contact. Because the true area of con-

---

*I do not believe it has ever been noted that this assumption is not vital to the argument. One might say that if the event lasts for sliding distance δL the worn volume will be proportional to width a, to depth a, and to length, δL. Thus, $\delta V \propto a^2 \delta L$, and $\delta V/\delta L \propto a^2$.

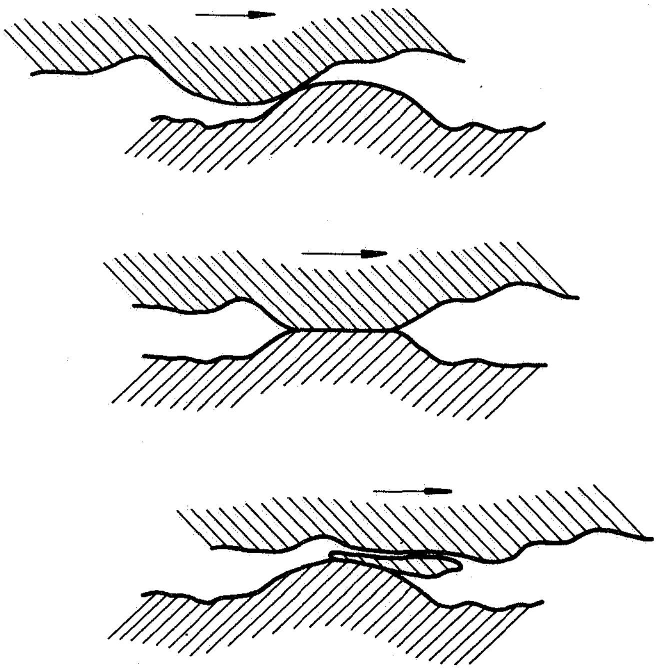


Figure 8.—Schematic representation of a unit event in the process of friction and wear.

tact is proportional to the load, so also are the friction and the wear rate. It need not be argued, at this stage, whether the proportionality between area of contact and load arises from plastic or elastic deformation of the contacting asperities.

The first conclusions to be drawn from the theory are that it forecasts wear to be proportional to sliding distance and load, but independent of the apparent area of contact. These are results obtained from general studies of wear such as those discussed above. The theory also suggests that, if the area of contact is formed by plastic deformation, the wear is inversely proportional to the hardness. This has been shown to hold for a limited range of materials.

The second major comment to be made upon equation 1 concerns the magnitude of wear rate forecasted. When measured values of wear rates are compared with this theory (table 1), it is found that the deduced values of K range from approximately 10^{-2} to 10^{-7}. (A similar result has been obtained by Rabinowicz (ref. 21) for a representative sample of 172 wear measurements.) This means that, if all events are assumed to produce a worn particle, the discrepancy between theory and experiment ranges between 10^2 and 10^7. Therefore, it is usual to assume that K represents the probability that any given event will produce a worn particle. For example, for tungsten carbide rubbing on tungsten carbide (table 1), only one in every million events results in a worn particle.

It is possible, of course, that other explanations can be provided for K; some will be discussed later. However, the most important fact about equation 1 is that it is an order of magnitude calculation and represents the first attempts to provide a quantitative basis for wear. Comparing the theory of wear with the theory of the strength of materials, equation 1 corresponds to Frenkel's calculation of the theoretical shear strength of perfect crystals (refs. 37 and 38) where large discrepancies were also shown to exist between theory and experiment, and these were eventually resolved by the dislocation theory. Similarly we require physical explanations of the magnitudes of K which can be confirmed by appropriate experiments.

We now outline the consequences of an alternative assumption about the shape of the worn particles. It is possible, particularly when considering the existence of protective films on metals, that each worn particle consists of a layer of material of thickness d. We may now proceed with arguments similar to those used in the derivation of equation 1. At any event, $\delta A=\pi a^2$, $\delta V=\pi a^2 d$, $\delta L=2a$. Thus $\delta V/\delta L=\frac{1}{2}\pi ad$, and

$$\frac{V}{L}=K\frac{\pi}{2}\sum da. \tag{2}$$

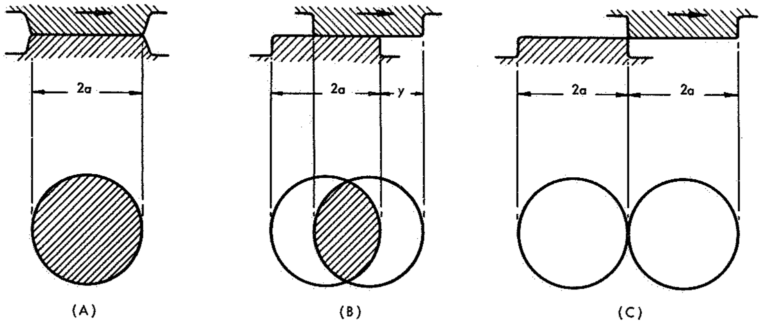


Figure 9.—Idealized representation of a single contact in sliding surfaces: (A) maximum contact area of radius a; (B) after sliding through a distance y; (C) after sliding through a distance $2a$, the contact area is just reduced to zero.

We note that, if d has a constant value (or represents a mean value which does not change during the wear experiments under consideration)

$$\frac{V}{L}=K\frac{\pi d}{2}\sum a.$$

The wear rate is now proportional to the load only when an increase in load causes a proportional increase in the number of contacts, which, if their mean size remains unchanged, than $A=n\pi a^2$, $n \propto W$, and

$$\frac{V}{L}=K\left(\frac{d}{2a}\right)A. \tag{3}$$

Equation 3 is similar to a formula discussed by Holm (ref. 6). As an order of magnitude calculation he imagined the worn volume to be distributed as a thin layer, m atomic layers thick, over the true area of contact ($d=m\alpha$, where α is the interatomic spacing). Comparison of equations 1 and 3 suggests that if the wear mechanism is the removal of the material as a thin layer rather than a lump, the K factor of equation 1 is reduced by a factor of approximately (d/a). A crucial factor here might be whether one takes as the value of a, the size of the overall region of figure 2, or the size of the individual regions of true contact.

Uhlig (ref. 39) and Yoshimoto and Tsukizoe (ref. 40) have developed the theory of wear using the assumption that the layer thickness d is the thickness of an oxide film. It is further postulated that this thickness can be derived from the theory of the growth rate of such films. Assuming a logarithmic growth rate,

$$d=\eta ln\left(\frac{t}{\tau}+1\right), \tag{4}$$

where η and τ are constants, and t is the time available for growth. For small values of t/τ, a useful approximation is

$$d=\eta t/\tau. \tag{5}$$

The time, t, in equations 4 and 5, is the time that elapses between two events at any given point on the surface. After such an event wear has denuded the surface over an area, δA, of its oxide film, and during the time, t, the film grows again to a thickness, d, given by equations 4 or 5. During this time, t, the surfaces slide a distance, s, where $s=vt$ and v is the sliding velocity. Thus equation 5 becomes,

$$d=\frac{\eta s}{\tau v}, \text{ when } \frac{s}{\tau v}<<1. \tag{6}$$

The crucial question in the next stage of the theory is to assign a value to s. The following simplified discussion is based on important physical assumptions (refs. 39 and 40). First, the total true area of contact, A, consists of n areas each of size $\delta A = \pi a^2$. For the sake of simplicity these areas are assumed to be spaced in a regular rectangular array with a spacing l (fig. 10).

It is further assumed that s is equal to l, the mean linear spacing of the contact areas. Now $nl^2 = A_{\mathrm{app}}$, where A_{app} is the apparent area of contact of the rubbing surfaces. Then using $A = n\pi a^2$, and putting $l = s$ in equation 6, equations 2 and 3 become,

$$\begin{aligned} \frac{V}{L} &= K\frac{\eta}{\tau v}(\pi A_{\mathrm{app}}\, A)^{1/2} \\ &= K\frac{\eta}{\tau v}\left(\frac{\pi A_{\mathrm{app}}}{H}\right)^{1/2} W^{1/2}, \end{aligned} \tag{7}$$

where the second half of the equation is based upon the assumption that the area of contact is formed by plastic deformation.

In the opinion of this writer, the assumption, $s = l$, is incorrect. Con-

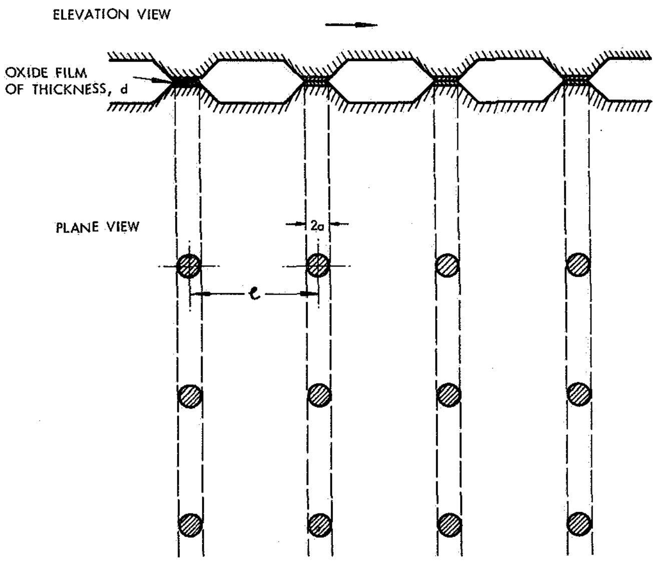


FIGURE 10.—Idealized distribution of the real area of contact.

sider the following argument. In deriving equation 2 it was assumed that an area $\delta A = \pi a^2$ is rubbed during a sliding distance $\delta L = 2a$. There are n such areas making a total area $A = n\delta A$. The question which needs to be asked is: How far will the surface slide, on an average, before these same areas are rubbed again? Now, in the course of a long period, all points in the apparent area of contact should be rubbed the same number of times. Therefore these same areas will be rubbed again after a sliding distance in which all the apparent area has been rubbed. Since an area A is rubbed in a sliding distance $2a$, the whole apparent area, A_{app}, is rubbed in a sliding distance

$$s = 2a(A_{\mathrm{app}}/A). \tag{8}$$

Then inserting this value of s in equation 6 and using equation 2 or 3, one finds that

$$\frac{V}{L} = K\frac{\eta A_{\mathrm{app}}}{\tau v}, \tag{9}$$

and that the wear rate is independent of load, inversely proportional to sliding velocity, and proportional to the apparent area of contact.

A number of workers have studied abrasive wear of materials, usually using abrasive papers or loose abrasives (refs. 41 and 42), The theory of abrasive wear is relatively simple, but its application, thus far, is less certain.

In the theory of abrasive wear (refs. 41 and 42), it is assumed that each abrasive particle is in the form of a cone that ploughs out and removes material from the abraded surface (fig. 11). Only minor changes in the constants of proportionality occur if a pyramid is assumed instead of a cone. Burwell (ref. 43) suggests that abrasive wear occurs most readily when the abrasive material is significantly harder than the abraded material, and when the abrasive particles are sharp, rather than rounded. This provides a justification for the assumption of a conical or pyramidal indentor. The physical conditions under which a spherical indentor might remove material have been investigated by Kragelskii (ref. 44) and this is discussed below.

If the load, δW, on an abrasive particle (fig. 11) is supported only over the leading half of the contact

$$\delta W = \tfrac{1}{2}\pi H z^2 \tan^2 \theta,$$

where z is the depth of penetration, θ is the cone half angle and H is the hardness of the material. The volume of material δV removed in sliding a distance δL is

$$\delta V = (z^2 \tan \theta)\, \delta L.$$

Thus, substituting for z from above

$$\frac{\delta V}{\delta L}=\frac{2\delta W}{\pi H \tan \theta},$$

and once again, assuming that only a proportion, K, of all the contacts produce worn particles, one has

$$\frac{V}{L}=K\left(\frac{2 \cot \theta}{\pi H}\right)W, \tag{10}$$

where θ is taken as a mean value for all the particles.

Equation 10 may be compared with equation 1. The problem of abrasive wear theory is to arrive at a correct value of cot θ, and to deduce the value of K. Mulhearn and Samuels (ref. 45) made measurements of the effective value of cot θ for the grits on their abrasive papers, and deduced a value of K of about 0.1.

It would be appropriate here to discuss briefly the extent to which the above theories and the assumptions have been justified by experimental evidence. Insofar as equation 1 has been used in the analysis of metallic transfer and severe metallic wear, there are a number of experimental investigations which justify its assumptions. In particular, the analysis of worn and transferred particles by Rabinowicz and Tabor (ref. 14), by Rabinowicz (ref. 21), and by Kerridge and Lan-

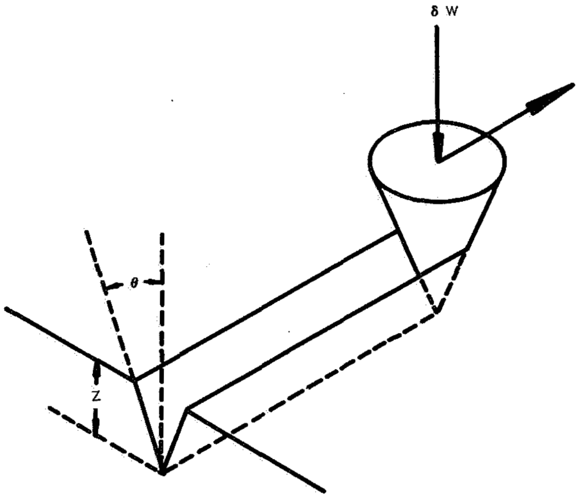


Figure 11.—Geometric assumptions in the theory of abrasive wear. Similar geometric relationships hold if the indentor is assumed to be in the form of a pyramid.

caster (ref. 16), using the techniques of radiotracers has already been cited.

In investigations of mild wear—in particular the idea of wear by removal of oxide films—no direct experimental evidence is known to this writer. Indeed, because of the very small scale of size, it might be a formidable task to provide such evidence. It seems more likely that the next stage in the development of this aspect of the subject is to devise experiments in which the wear rate is measured under controlled conditions. This is discussed later in this paper. Therefore, it should be realised that the theory associated with equations 7 and 9 is largely conjectural. At a number of points in the derivation of these equations alternative assumptions might be made. For example, the logarithmic growth rate law, equation 4, was used in the first paper on this subject (ref. 39), and has therefore been adopted here, but its replacement by a parabolic law is a relatively simple matter.

The assumptions of the theory of abrasive wear have been examined in larger scale tests with single indentors, notably by Kruschov and Babichev (ref. 46) and by Sedricks and Mulhearn (refs. 47 and 48). The extent to which the results of such large-scale experiments are applicable to the smaller scale and the multiple-contact conditions of abrasive wear is open to question. However, some justification for this procedure has been provided by Mulhearn and Samuels (ref. 45) in their analysis of abrasive papers.

CLASSIFICATION OF WEAR

A number of authors have attempted to provide a broad classification of wear. Perhaps the most commonly used classification is that originated by Burwell (ref. 43). It involves wear by

(1) adhesion or galling
(2) corrosion
(3) the presence of loose abrasive particles
(4) cutting or ploughing of a soft material by a harder rough surface
(5) less common effects such as erosion and surface fatigue.

Burwell's classification is essentially a list of the mechanisms of wear. Its major function is to provide a framework so that "at last the engineer and designer will be provided with a rational approach to the wear problem in their machinery."

Adhesive wear is described primarily in terms of metallic welding and, in the light of work discussed above, can probably be taken to represent the first stage of the severe wear of metals. Abrasive wear includes the two mechanisms given in (3) and (4) and the distinction between them should be quite clear; the two situations have sometimes been described as three-body and two-body conditions. In abrasive wear the ability of a material to withstand a given magnitude of strain may be an important consideration, and Oberle (ref. 50) first pointed out that the relevant physical property of the material was H/E, the ratio of the hardness to the elastic modulus. The same parameter has occurred in the analysis of the ability of contacting rough surfaces to withstand plastic flow (ref. 51 and 52).

Corrosive wear requires the presence of both corrosion and rubbing. Specifically it was noted that "oxides or hydroxides of the metals in question are formed by exposure to the environment, and since with few exceptions these compounds are relatively loosely adherent to the metal base, even the mildest rubbing serves to remove them." Burwell (ref. 43) described fatigue wear as occurring primarily when surfaces are in rolling contact. Pitting or flaking of the surfaces was principally in mind. However, as discussed below, subsequent work suggests that fatigue mechanisms could well play even a larger part.

The state of knowledge, which existed when this classification was proposed, was such that these mechanisms could be largely considered as separate and distinct ways of removing material. However, to quote again from Burwell's original paper, "In any particular instance of wear, one may have any of these mechanisms operating either singly or in combination. For example, the hard metal oxide formed on a steel surface by corrosion may then act as a fine abrasive to wear both the surfaces."

Kragelskii (ref. 53) has defined five "types of destruction of the frictional bond." These are shown in figure 12 and are classified in table 2. The nature, and perhaps the function, of Kragelskii's classification is different from that of Burwell. It is a description of the various types of event which can occur between asperities in contact.

In this classification Kragelskii attempts to define the conditions under which various types of events will occur. A spherical indentor, or an asperity whose shape can be represented by a spherical cap, is assumed. The first important criterion which determines the character of sliding is the ratio h/R, where h is the depth of penetration of the indentor or the height of the asperity and R is the relative radius of curvature of the surfaces. It has been shown (ref. 54) that, when a hard spherical indentor is pressed into a flat surface, plastic flow will occur when

$$\frac{h}{R} \geq C\left(\frac{H}{E}\right)^2, \tag{11}$$

where C is a constant. (Alternatively one can derive an equation in which H is replaced by σ_s, the yield stress). Similarly, if h represents the height of an asperity of radius of curvature R, equation 11 expresses the condition under which the asperity, when flattened, will

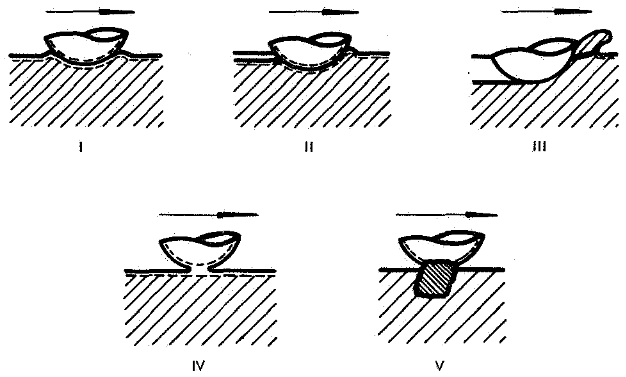


FIGURE 12.—Schematic representation of the main types of destruction of frictional bonds (ref. 44). Also see table 2.

incur plastic flow. This same criterion can be expressed in terms of the height to base dimension ratio called the mean slope (tan θ) of the asperity (ref. 51 and 52). Thus if no plastic flow occurs when the asperity is flattened

$$\tan\theta \leq C'\left(\frac{H}{E}\right), \tag{12}$$

where C' is a constant that depends upon the particular assumed shape for the asperity. In table 2, the type I frictional bond represents pure elastic deformation.

With plastic flow the bond can be of type II or type III, the important distinction being that under type III conditions material is removed from the surface. Kragelskii (ref. 44) gives a simple derivation of the conditions for the transition from type II to type III, and they involve the equilibrium of the material at the front of and in contact with the indentor. If, as the spherical indentor moves forward, this material can slip beneath it, the sphere will plough out a track with displacement but no removal of material. If, on the other hand, material is unable to move with respect to the indentor, "pile up" will occur in front of the indentor and eventually material will be removed. In this analysis the important factors are the limiting friction between the indentor and the work material (determined by μ, the true or molecular coefficient of friction between indentor and work material), and the h/R ratio. As μ increases, the transition to material removal occurs at lower values of h/R. This is shown in table 2 for dry and lubricated conditions. Kragelskii's simple analysis suggests that, as μ rises to unity, the transition to

TABLE 2.—*The Main Type of Destruction of Frictional Bonds*

[From ref. 44]

Characteristics	Type of bond				
	I	II	III	IV	V
Nature of deformation	Elastic displacement	Plastic displacement (repeated deformations)	Cutting of the material	Destruction of surface films	Destruction of bulk material
Number of cycles, n, leading to failure of base	$n = \infty$	$1 < n < \infty$	$n \rightarrow 1$	$n \rightarrow \infty$	$n = 1$
Conditions	h/R [a] < 0.01 (ferrous metals)	$h/R < 0.1$ (dry)	$h/R < 0.1$ (dry)	$d\tau/dh$ [b] > 0	$d\tau/dh < 0$
	0.0001 (nonferrous metals)	$h/R < 0.3$ (lubricated)	$h/R > 0.3$ (lubricated)	—	

[a] The notation h/R refers to depth of penetration, h, of a spherical indenter of radius, R, or the ratio of the height, h, to radius of curvature, R, of an asperity (ref. 51).

[b] The notation $d\tau/dh$ refers to the variation of shear strength, τ, with depth, h, below the surface.

removal of material tends to a value of $h/R=0$. Presumably, this corresponds to type V of his classification: metallic welding, scuffing of the surfaces, and removal of material in depth.

For the case of a pyramidal, rather than a spherical indentor, Sedricks and Mulhearn (refs. 47 and 48) consider the mechanics of the two situations: sliding with displacement of material, and sliding with removal of material. The first is analysed in the very simple terms used by Bowden and Tabor (ref. 11) to deduce the coefficient of friction associated with the ploughing mechanism. The second is analysed by methods similar to those used in the theory of metal cutting. From comparison of these two analyses, the critical condition for the removal of metal is shown to depend upon the "attack angle," and the true coefficient of friction μ between work material and indentor. The "attack angle" is the angle between the leading face of the indentor and the work material and is equal to ($\pi/2+$the rake angle), in metal-cutting terminology. Mulhearn and Samuels (ref. 45) have explained the existence of cutting and noncutting grits in terms of the range of "attack angles" exhibited by grits in abrasive papers. In this way they give an explanation of the value $K=0.1$ (equation 10) deduced from their experiments.

The final comment on Kragelskii's classification is the distinction he draws between type IV and type V. The critical parameter here is taken to be the variation of the shear strength τ with depth. If the shear strength of the surface film (and, one might add, of the film-substrate interface) is less than that of substrate, the frictional bond is of type IV; but if the film, or the surface layer without the film, is the stronger the bond will be of type V. The implication of this classification appears to be that the removal or destruction of the surface film is an adhesive phenomenon.

In the discussion of detailed studies of wear, it was emphasized that experiments have shown many mechanisms of wear to be multistage processes. Here the term "mechanism of wear" is used to describe the event or events which result in material, originally part of one rubbing surface, becoming a part of the loose worn debris. On this basis it may be difficult to define wear mechanisms by Burwell's classification. For example, in the severe wear of brass (ref. 16), the first stage of metallic transfer is adhesive. Kerridge and Lancaster suggest that "the most plausible explanation for the removal of wear particles is that the bond between them and the original surface of the ring becomes gradually weakened as a result of the alternate compressive and tensile stresses to which they are subjected"; that is, the second stage is a fatigue process associated primarily with subsurface rather than purely surface adhesive effects. Therefore, in considering a fundamental approach to wear, this writer is inclined to adopt an approach, similar to that of Kragelskii, in which the classification is one of events rather than of complete wear mechanisms.

The classification put forth here, and shown in table 3, is readily adaptable. One first considers the nature of the contact between the opposing surfaces; the deformation will be elastic or plastic. The results which then follow from the imposition of sliding upon the system are associated with surface effects and/or phenomena occurring in the bulk of the material. It should be noted that in an elastic contact the maximum shear stress occurs below the surface. Likewise in a static contact involving plastic deformation the plastic flow is below the surface. Moreover, experiments have shown that this is also true when tangential motion occurs, provided the tangential forces are moderate (refs. 55, 56 and 57). Theory suggests that this situation will continue until the tangential forces reach values equivalent to a coefficient of friction of approximately 0.25 (refs. 58 and 59). It seems likely from this form of classification, that subsurface effects can play a part in wear only when they become linked with the surface effects. For example, a detailed analysis may well show that the transition from Kragelskii's type II to type III condition corresponds to plastic flow, entirely subsurface in type II, reaching the surface.

It is worth noting that what has been described in this paper as mild wear can possibly cover a number of different mechanisms. The most obvious have been summarized by Waterhouse (ref. 60) in his review of fretting:

TABLE 3.—*Classification of Mechanisms Contributing to Friction and Wear*

	Surface effects	Subsurface or bulk effects
Elastic deformation	Amontons' law of friction obeyed with multiple contacts. Adhesive wear likely to be involved with protective surface films (i.e., mild wear).	Elastic hysteresis component of friction. Fatigue wear includes rolling contact fatigue (pitting) and perhaps some smaller scale mechanisms.
Plastic deformation	$\mu = s/H$ (Bowden and Tabor) Adhesive wear can be mild or severe dependent upon degree of protection offered by surface films. If severe wear, surface and subsurface effects become linked.	Ploughing term of friction. Abrasive wear, and its equivalents, involve direct removal of material. Ploughing can occur without removal of material but fatigue mechanisms are then probable.

(1) Removal of metal in a finely divided form with possibly subsequent oxidation.

(2) Removal of metal which when oxidised gives an abrasive powder to continue the action.

(3) Oxidation of the metal surface and the removal of the oxide which exposes fresh metal for oxidation.

The classification of wear can serve many purposes, but the emphasis of this discussion has been upon the evolution of systems of classification which are useful in furthering basic research. However, this is not to deny other very important functions that such classifications can serve. Perhaps the greatest dangers are in a failure to reassess these systems in the light of new knowledge, and in the use of a classification outside the intentions of its author. For example, the severe wear of 60/40 brass has been described (ref. 15) as "pure adhesive wear," which implies something unique. However, it has already been shown that severe wear (involving adhesive wear as the first stage) can, possibly, cover a range of mechanisms in which the second or subsequent stages could differ widely. Only further research can answer such questions, and, perhaps this will always be done independently of how we classify and reclassify wear on the basis of imperfect knowledge.

THE WEAR OF NONMETALS

The vast majority of investigations of wear have been concerned with metals. One is thus preoccupied with the behavior of a ductile substrate covered with a thin protective film. In the terms of materials science, it might be argued that the research into mechanisms of wear has confined most of the evidence to materials in which the bonding forces are metallic, and therefore, strong and nondirectional. Moreover, in the terms of the classification of table 3, one might argue the effects of changes in the rubbing materials.

In using nonmetallic materials, the changes in the deformation characteristics can be of two main types. For materials with the ionic or covalent bonding forces, one might expect brittle rather than ductile failure. By using molecular solids, such as high polymers, one is concerned with materials of low elastic modulus and large elastic range. This is important in terms of the shape of asperities which can be pressed flat without exceeding the elastic limit (refs. 51 and 52).

It can also be argued that nonmetallic materials might have less adhesion at the interface because of the absence of strong metallic welds. The experimental evidence, however, is mixed. Finally we note that the nonmetallic materials will be notably poorer thermal conductors so that the thermal effects of sliding are expected to be greater.

A number of limited investigations have been made into the wear behavior of nonmetals, and some general conclusions can be drawn.

First, there seem to be some indications that the wear of polymers and rubbers (refs. 61 to 71) occurs on two major scales of size somewhat analogous to the mild and severe wear of metals. Kragelskii (ref. 67) produces evidence that, on a small scale, the wear is a fatigue process. On a large scale, Schallamach (ref. 62) shows how the damage is built up from small tears in the surface. These arise from large tensile stresses existing at the edge of the contact region under the combined effects of normal and tangential loads. It may be noted that this situation requires a large adhesion.

A number of authors have noted the importance of thermal effects in the wear of polymers; a large increase in wear has been shown to correspond to the conditions under which the surface temperature reaches the softening point of a thermoplastic polymer. In other situations it has been suggested (ref. 68) that thermal and oxidative effects result in the formation of reactive low molecular weight components. It has been shown that wear rate of PTFE moldings of varying crystallinity (ref. 66) is approximately proportional to the elastic modulus and inversely proportional to the yield strain. However, it would appear that in order to understand these observations, one needs to establish clearly the mechanism of nonmetallic wear.

The wear rate of electrographite is of particular interest because under many conditions its wear rate is amongst the lowest recorded. The questions arising from the layer lattice structure of graphitic carbons will not be discussed here, beyond noting that most carbons seem to combine a low elastic modulus with a low shear strength and fairly high thermal conductivity. A number of investigations (refs. 72 to 78) suggest that the wear of carbons on metals is critically dependent upon the surface finish of the metal surface. When this surface is rough, the wear rate is comparatively high (although still low by severe wear of metals standards) and the mechanism of wear is by a filing or micro cutting process probably involving brittle fracture of the carbon (ref. 74). With smoother surfaces of copper, the surfaces become smooth by transfer and wear of the copper. The wear rate under these conditions can be very low indeed, and it is suggested that the deformation of the asperities is entirely elastic and that the wear mechanism is a fatigue process caused by repeated elastic deformation of the regions of contact (refs. 75 and 76). Indeed, it is suggested that under these conditions the wear rate of the carbon material is inversely proportional to its elastic modulus.

It has also been found that nongraphitic carbons, rubbed against themselves or against mild steel in an atmosphere of dry carbon dioxide, exhibit a curious pattern of an increased friction at regular intervals. It was shown (refs. 77 and 78) that the periods of high friction corresponded to a period when the transferred material formed a very smooth surface. In this situation the area of contact, and therefore the frictional force, increased. The increased shear forces then caused a breakdown of the surface layers and the surfaces became roughened for the process to start once again. It is also worth noting that the wear and transfer of graphitic materials rubbing on metals involves some questions worthy of further study. There is evidence (ref. 73) that oxides on the surfaces of metals play a beneficial role in the establishment of the transferred layers.

The wear of hard brittle materials has been investigated by a number of workers (refs. 79 to 90). It is suggested that the wear of abrasive grits (ref. 85) on metals is of two types: "attritious wear," probably due, at least in part, to chemical reactions between the metal and the abrasive grit; and, fragmentation, which is presumably a brittle fracture phenomenon. The wear of artificial sapphire rubbing on metals has been investigated in detail (refs. 82, 86, and 87). It has been found that the wear is dependent upon the orientation of the sliding direction to the crystal axes. The relative roles of chemical and mechanical factors have not been fully established.

It might be thought that the mechanism of wear of all brittle materials would show some similarity, but it seems likely that these materials can show a range of wear mechanisms dependent upon the operating conditions. Thus Bowden and Hanwell (ref. 88) have found that in a high vacuum denuded of adsorbed films, diamond rubbing on itself shows high adhesion and severe wear caused by brittle failure. On the other hand, King and Tabor (ref. 89) have shown that brittle solids can exhibit a ductile behavior under contact stresses. This arises from the existence of large superposed hydrostatic stresses in the subsurface regions. Finally, Atkins and Tabor (ref. 90) have recently shown that hard brittle materials become ductile at high temperatures. Thus when used at high ambient temperatures, or perhaps under the influence of temperatures generated by frictional heating, these materials may well display the ductile behavior found in rubbing metals.

WEAR UNDER LUBRICATED CONDITIONS

Perhaps the major problem in the study of wear under lubricated conditions is in deciding how much of the load is borne by a fluid film, how much by boundary lubricated contact, and how much by unlubricated contact. (In this connection the term "boundary lubrication" is used to describe lubrication by films whose properties and thicknesses are determined by the molecular rather than the bulk properties of the lubricant). The possible existence of fluid films has often been suggested by experimenters even though the application of Reynolds' equation to their experimental conditions imply that such films would be very thin indeed. For example, in some early experiments with the 4-ball machine, Beeck, Givens, and Smith (ref. 91) measured electrical resistance between the specimens. In this machine the load between one upper ball and each of three lower balls is concentrated, nominally, at a single point. Beeck, Givens, and Smith concluded that there existed between the balls a quasi-hydrodynamic film. However, assuming the bulk viscosity of the lubricant at ambient temperature and pressure, and using the undeformed shape of the surfaces, it can be shown that the theoretical thickness of the film generated by hydrodynamic mechanisms would be comparable with or smaller than the size of a lubricant molecule.

These persistent suggestions, that fluid films exist where theory says they are impossible, have been justified, in recent years, by the development of the subject of elastohydrodynamic (ehd) lubrication (refs. 92, 93, and 94). It has been shown that in systems where the load is concentrated, such as gears, cams and tappets, and rolling-contact bearings, a new mechanism of lubrication occurs. The local deformation of the surfaces and the large increase of lubricant viscosity with pressure combine to provide the conditions for a fluid film, where previously this was thought to be impossible.

Using lubricants of conventional viscosities (say, a fraction of 1 poise under laboratory conditions) and steel specimens, having a radius of curvature on the order of an inch, it has been shown that this mechanism of lubrication can persist down to surface speeds of a fraction of one centimeter per second. Under these adverse conditions, film thicknesses of the order of a few hundred Angström units have been measured (refs. 95 and 96). Under conditions more normal in engineering, the film thickness may be approximately 10^{-4} cm. In ehd lubrication the coefficients of friction often lie in the range normally attributed to boundary lubrication (0.05 to 0.10). These coefficients, high for normal conditions of fluid-film lubrication, arise from the shearing of a fluid film of high viscosity. An increase in viscosity of lubricants in the load-bearing region, by factors as high as 10^4 or 10^5, have forecast by ehd theory and confirmed by experiments.

The development of the subject of ehd lubrication, and the fact it has been shown to persist over a wide range of conditions, adds considerably to the study of wear under lubricated conditions. For example, it seems likely that when the wear of nominally conforming surfaces is studied, ehd mechanisms can provide lubrication of individual asperities. These mechanisms would also provide most of the mechanical features involved in the unlubricated contact of asperities, including the existence of appreciable tangential forces. However, features associated with intimate contact, such as metallic welding, should be absent. Perhaps, under these conditions, wear by fatigue mechanisms and wear by the removal of protective films could occur.

The theory of ehd lubrication assumes that the surfaces are perfectly smooth. Likewise, the experiments in which the ehd film thicknesses have been measured, used very smooth surfaces. More recently experiments have been performed with rougher surfaces and the extent to which solid contact occurs through the fluid film has been studied (refs. 97 and 98). To specify the character of these partial ehd conditions, a dimensionless parameter, the ratio of the ehd film thickness to some measure of the surface roughness, has been used. Dawson (refs. 99 and 100) uses a different method of specifying the surface roughness from Tallian and co-workers (ref. 97), but the essential physical ideas are the same. Wear and failure phenomena in lubricated rolling contacts apparently depend upon this parameter (ref. 101).

In conditions of partial ehd lubrication, one is concerned with the behavior of asperities, both in the lubrication of rough surfaces whose general shape is nonconforming, and in the lubrication of conforming surfaces whose surface roughness contains superposed features of widely differing wavelengths (fig. 2). The theory of ehd lubrication has, as yet, been concerned with the mechanism of lubrication only when the larger of these features is present. A fundamental question which remains to be answered is the extent to which ehd mechanisms can depress, and therefore protect from contact and damage, the smaller scale features within the main load bearing region of an ehd contact.

THEORIES OF WEAR

One important question in any analytical theory of wear is the significance of the factor K. In the earlier discussion K was interpreted as the probability that any event would produce a worn particle. According to this, a typical severe wear situation, where the deduced values of K (table 1) can be as low as 10^{-3}, implies that each part of the rubbing surface must be rubbed 10^3 times before a worn particle is produced. This, of course, would explain why wear rates can vary over a very wide range (table 1), while coefficients of friction show little variation. In all normal rubbing conditions the majority of events, and those which therefore determine the frictional force, are contacts between asperities which separate without damage.

Perhaps the most important problem is in relating the further development of analytical wear theories to real, rather than hypothetical, mechanisms of wear. Also, a part of this development must yield a more exact meaning for the value of K. There are a number of possible explanations that might be advanced to give the factor K some physical significance. For the following discussion assume that K is 10^{-3}.

(1) On statistical grounds, one in every thousand events is sufficiently severe to produce a worn particle; in other words, one in every thousand deformation events is sufficiently severe to initiate damage.

(2) On an average, under equilibrium conditions, each asperity must be rubbed a thousand times before it is damaged. This implies that wear is a microfatigue process, a view expressed by Kerridge and Lancaster (ref. 16) and by Kragelskii (ref. 102).

(3) In severe metallic wear with back transfer, material is transferred to and fro between the rubbing surfaces until the bond becomes so weak that the material is removed rather than transferred. It seems unlikely that this could be the total explanation of the K factor because the highest values of K, deduced (refs. 35 and 36) from transfer experiments, are in the range 4 to 16×10^{-2}.

(4) In mild wear, if one is concerned with removal of a protective film, then, under the conditions defined by equation 3, the wear rate, and therefore the deduced value of K, is reduced by a factor d/a (d is the film thickness, and a is the radius of the individual contact areas).

One further factor that can influence deduced values of K has been discussed recently by Cocks (ref. 103). Both Cocks (ref. 104) and Antler (ref. 105) have discussed a mechanism by which loose particles are produced when a rider, typically of the noble metals, slides against a flat surface of a similar metal. Clumps of displaced particles build up at the leading edge of the rider and the whole of the load can be borne on this 'prow' of material until it breaks away to form a loose particle. More detailed mechanics have been studied by Cocks (ref. 103) using concepts derived from Greenwood and Tabor (ref. 106). Cocks shows that in this model situation the wear rate becomes

$$\frac{V}{L} = A \sin \phi, \tag{11}$$

where a is the true area of contact, and ϕ is the angle at which shear in the junction is inclined to the surfaces. In this argument the probability factor K of equation 1 is replaced by $\sin \phi$. Since K for severe wear is typically in the range 10^{-2} to 10^{-3} (table 1), Cocks' argument could be only a partial explanation of the observed wear rates. Another factor still seems to be needed.

Rowe has applied wear theories to boundary lubrication conditions and has produced a theory (ref. 107) that is essentially a modification of equation 1. His approach may be outlined briefly as follows:

(1) Under conditions of boundary lubrication, the area of metal-to-metal contact, A_m, is a proportion, α, of the total true area of contact, A, ($A_m = \alpha A$).

(2) The "fractional film defect," α, is determined by the chance which an asperity has of trapping, as it passes over the surface, a site unoccupied by adsorbed boundary lubricant molecules.

(3) α can be calculated from adsorption theory and is a function of known constants, such as the heat of adsorption, the time of fundamental vibration of the molecule in the adsorbed state, the dimensions of the adsorbed site, and the sliding velocity.

(4) Wear occurs only at tha area of metal-to-metal contact $A_m = \alpha A$, while negligible wear occurs at the proportion of the true area of contact, $(1-\alpha)\ A$, protected by a boundary lubricant film.

(5) For the area of metal-to-metal contact, there is a true value of K designated K_m.

Rowe also describes some experiments measuring the wear of copper rubbing on steel. The agreement between these results and the theory is remarkably good. Tabor (ref. 108) comments that some of the results, for example, the variation of wear rate with speed, might arise from the influence of hydrodynamic lubrication. This comment illustrates the problems involved in the interpretation of wear experiments under lubricated conditions.

To this writer there are two other problems in Rowe's theory. First, according to (4), the wear rate is proportional to A_m even though this may be made up of desorbed regions of molecular dimensions; on the other hand wear is generally concerned with regions of much larger dimensions. Second, the theory appears to be concerned with desorbed sites on one surface only, which means that if one calculated the fractional film defect, α, as the area where desorbed sites on both surfaces coincide, one might arrive at a very different result. Perhaps also under these conditions, Rowe's surprisingly good agreement between theory and experiment might no longer apply. Nevertheless, Rowe's approach is a pioneer attempt at producing a modified theory of wear in which the value of K is at least partially attributed to measurable physical factors. Rowe has also applied these concepts to the wear of graphite (ref. 109).

A modified theory of wear, appropriate to the mild wear of steels by oxide film removal, has been suggested by Quinn. In a preliminary paper (ref. 110), he discussed the influence of transient flash temperatures, generated at the true areas of contact, upon the oxide films formed on rubbing surfaces. Some correlation was claimed between the composition of the oxide debris, as observed by various investigators, and the calculated flash temperatures.

In a later paper Quinn (ref. 111) outlined a theory of wear based on the following concepts:

(1) Oxide film growth occurs during the generated flash temperature phase, but growth at other times is negligible.

(2) Oxide film growth continues for N traversals, over any given point on the surface, each traversal having a duration τ.

(3) At this stage a film of thickness ξ has been generated that the substrate can no longer support. The film is then removed.

(4) On this basis the value of K in equation 1 is $1/N$.

This last statement seems to this writer to be in error, because, according to the argument of equation 3, the removal of a layer of thickness ξ, instead of a hemispherical lump of radius a, is equivalent to changing K by a factor of approximately ξ/a. However, the most interesting question raised by Quinn is the implication of assumption (1) in which flash temperature is considered the influential factor in the generation of surface protective films. Other evidence, though, such as that of Lancaster (ref. 22), suggests that the mean surface temperature is the important factor.

Consider this simple order-of-magnitude calculation. Assume that a surface of apparent area of contact A_{app} is rubbing against a similar surface. We take, as did Quinn, a parabolic growth rate law,

$$(\Delta m)^2=kt \tag{12}$$

where Δm is the increment of film growth, t is the time, and k is a constant which depends upon the temperature T and given by

$$k=\beta \exp(-Q/RT), \tag{13}$$

where β and Q are constants for the given material and R is the gas constant.

If t_1 is the time for which the surface is exposed to the flash temperature T_1, and t_2 is the time for which it is exposed to the mean surface temperature T_2, then, according to the argument used in the derivation of equation 8, one may write the approximate relationship

$$\frac{t_1}{t_2}=\frac{2a}{s}=\frac{A}{A_{\text{app}}}. \tag{14}$$

Thus if Δm_1 is the increment of growth during the time t_1 (flash tempature) and Δm_2 is the increment of growth during the time t_2 (mean surface temperature) equations 12, 13, and 14 give

$$\left(\frac{\Delta m_1}{\Delta m_2}\right)^2=\frac{A}{A_{\text{app}}}\frac{\exp\ (-Q/RT_1)}{\exp\ (-Q/RT_2)}. \tag{15}$$

Now, assume $A=10^{-4}\text{cm}^2$, $A_{\text{app}}=1\ \text{cm}^2$, $T_1=1000°$ K, and $T_2=500°$ K. From reference 111, $Q=4.6\times 100^4$ cal/°K/mole, and $R=1.99$ cal/mole. Thus from equation 15, $(\Delta m_1)/(\Delta m_2)\sim 10^3$.

This calculation appears to justify the assumption that flash temperature is the influential factor. However, apart from the experimental evidence of Lancaster (ref. 22), some support for the importance of mean surface temperature has been provided in recent experiments by Clark and Pritchard,* who studied the mild wear of hardened steel as a function of load, speed, ambient temperature, and atmosphere. They found that load or speed variations affected mean surface temperature. In addition the variation of wear rate with load and speed was found to be consistent only when the mean surface temperature rather than the ambient temperature was maintained constant. It is perhaps of interest to note that under these conditions the wear rate was found to be independent of the load and to decrease with speed rather more slowly than inverse proportionality (cf. eq. 9 although the agreement may be coincidental).

The role of these temperatures remains an open question. It is suggested that appropriate theories for the mild wear regime must be based upon experiments in which the mean surface temperature is known and can be controlled over an adequate range of values. Also, flash temperatures must be estimated with reasonable accuracy: a difficult task. Quinn has followed a number of other workers, including this writer (refs. 112, 113, and 114), in estimating the maximum possible flash temperature, assuming that all the load is borne at a single region. The difficulties and dangers of this procedure can be illustrated by referring to the work of Welsh (ref. 24). He found that for loads greater than the T_2 transition (see fig. 5), mild wear occurred. But with soft steels this was always preceded by a period of severe wear in which phase hardening of the surface layers occurred. As explained earlier, the conditions at the T_2 transition were consistent with the idea that phase hardening was caused by a martensitic transformation. This argument was also based on the assumption that flash temperature was produced when all the load was borne at a single region. However, Welsh found that for loads greater than T_2, after the establishment of mild wear, the intense hardening of the surface layers was not maintained. As mild wear proceeded the hardness of the surface layers declined, implying that, under mild wear conditions, the load was more widely distributed.

CONCLUSIONS

This review has stressed fundamental studies of wear, and although it may be regarded as a counsel of perfection, the aim of such studies should be to reveal, in detail, the ways in which wear occurs. Nor-

---

*Clark, W. T.; Pritchard, C.: unpublished. (See Pritchard, C., contribution to discussion, Lubrication and Wear Convention, Inst. Mech. Engrs., paper 21, May 1967).

mally this implies discovering the way in which material, originally a part of one rubbing member, eventually becomes loose detached debris. Judged on these standards, our knowledge of wear is less than our knowledge of the allied subjects of friction and lubrication.

Wear is a complex subject, though. For example, the aspect of wear about which most is known is transfer by metallic welding (adhesive wear). But, in many rubbing situations, this is only one part of the complete wear mechanism. Moreover, it seems possible that the majority of "normal wear" in successful engineering systems does not involve metallic transfer at all. Under these conditions mild wear occurs, and here again, fundamental knowledge is meager. It is also clear that chemical reaction between the environment and the rubbing surfaces often plays a vital role. For example, the importance of oxide films, in the unlubricated wear of metals in air, has been established in the transition from severe wear to mild wear. However, the detailed role of oxide films, once mild wear is established, is not clear.

One possible mechanism of mild wear has been examined in more detail: the growth of oxide films and their removal at the areas of contact. However, experimental evidence does not support the corresponding theories, and further study is needed.

It seems that we have not yet discovered how to establish the details of all wear mechanisms. Even when worn surfaces and loose debris are examined using sophisticated techniques, the sequence of events by which they were produced cannot always be established.

Against this limited knowledge of mechanisms of wear, the classifications of wear have been reviewed. Although such classifications have a useful role to play as a broad statement of present knowledge, their value in a more basic approach has been questioned.

Finally, it should be emphasized that part of our interest in wear arises from the potential it offers for the application of knowledge from a wide range of scientific disciplines. For example, a contact of size 10^{-2} cm, rubbing at a speed of 100 cm/sec, could well involve the following combination of physical conditions: pressures, greater than 10^5 psi; temperatures, 500° to 1000° C above ambient; duration, 100 microseconds. Perhaps the application of a range of scientific disciplines to the study of wear would also broaden those disciplines.

DISCUSSIONS

E. Rabinowicz (Massachusetts Institute of Technology, Cambridge, Massachusetts)

As is invariably the case with his papers, Dr. Archard's contribution to this symposium is interesting, important, and relevant. He has produced a very comprehensive evaluation of the wear field. In my discussion, I want to make a few brief additions to some of his topics, and then make a more detailed assessment of his conclusions.

In regard to the history of wear, I have recently discovered that the first person to discuss a wear problem, that of material removal during polishing, was none other than Newton in 1704 (ref. 115). His conclusion, that polishing is simply abrasion on a reduced dimensional scale, is currently controversial (and in my opinion incorrect), but has the powerful support of Samuels (ref. 116). All in all, the pedigree of wear is seen to be about on a par with that of friction, whose ancestors include Leonardo da Vinci, Euler, and Coulomb.

Dr. Archard makes the point several times that fine oxide deposits lead to abrasive wear. This comment is frequently made, especially in regard to fretting (I have done so myself on many occasions). But two types of information have recently become available which make me doubt that oxide particles cause any appreciable amount of abrasion. The experiments of Rabinowicz and Mutis (ref. 117) show that abrasive particles of small grain size give greatly reduced abrasive action; further, some unpublished calculations of mine show that values of K which arise during the fretting of lubricated surfaces are generally about 10^{-6}. This is typical of adhesive wear situations, but much too low to be abrasive wear as we generally define it.

Dr. Archard quotes Oberle's work and his use of the H/E ratio. Oberle gives a table of H/E ratios for various metals, and finds that metals with the highest H/E ratios are most resistant to wear. His procedure is attractive, but logically it is suspect. The ratio H/E is essentially a constant of value 1% for all metals not subject to creep. The metals which do creep at room temperature are a special class with a number of unusual properties (low hardness, low melting temperature). Thus, the fact that these metals behave differently may not be related to a lower H/E ratio at all. The point is that the strength properties of metals are so highly intercorrelated that it is difficult to prove by considering a batch of metals that a given effect arises from a given cause and no other cause. Much other work in the literature is open to the same objection (e.g., Bowden and Hughes (ref. 118), work on polishing, in which case, because of the high correlation between hardness and melting temperatures, it is difficult to judge whether powders polish because they are hard or because they resist melting).

My main point deals with Dr. Archard's conclusions that studies of detailed wear processes are needed. As he admits, this is a counsel of perfection but I go further, and regard it as the kind of counsel one gives to other people rather than to oneself, bearing in mind the relatively meager practical results which have come out of such studies in the past.

Before examining wear on a microscopic scale, we should direct our attention to macroscopic aspects, of which we have to date only scratched the surface. Equation 1, of Dr. Archard's paper,

$$\frac{V}{L}=\frac{KW}{3H},$$

seems to me to be the most fruitful jumping off point. To extend it, we note that K is nondimensional, and write

$$K=f(r_1,\ r_2,\ r_3,\ \text{etc.}),$$

where the r terms are nondimensional parameters, representing ratios of like-dimensional quantities.

Now we are launched into a hunt for r terms which might influence wear. For hexagonal metals, one such ratio has been identified, that between the c and a dimensions of the crystal lattice (ref. 119). Table 4 shows the relatively good correlation we have measured between K and the c/a ratios of a number of hexagonal metals.

Another potent ratio is that between W_{ab}, the interfacial energy of adhesion at contact, and the surface energies γ_a and γ_b of the two contacting materials. As Machlin and Yankee (ref. 120) have postulated, especially significant is the ratio of W_{ab} to the surface energy of the softer material. This ratio is high for compatible materials (in the metallurgical sense of high solubility) and low for incompatible materials (ref. 121). This correlation between K and compatibility is strikingly borne out in wear studies (table 5). Further work on compatibility effects, along the lines embarked on by DeGee (ref. 122), is very obviously called for.

A geometrical ratio which seems to be significant is the plasticity

TABLE 4.—*Wear Coefficients, K, of Hexagonal Metals Arranged in Order of Increasing c/a Ratio* [a]

Metal	c/a ratio	Wear coefficient, K
Hafnium	1. 583	68×10^{-4}
Titanium	1. 586	23
Zirconium	1. 588	42
Thallium	1. 600	2
Rhenium	1. 615	. 04
Magnesium	1. 620	12
Cobalt	1. 623	. 81

[a] Riders were slid against flats of the same metals, under 200 g load, at a speed of 10 cm/sec.

TABLE 5.—*Comparison of Coefficients and Compatibility Ratings of Metal Pairs*

Metal pair	Compatibility	Wear coefficient, K
Copper-lead	Incompatible	0.10×10^{-4}
Nickel-lead	Incompatible	0.21
Iron-silver	Incompatible	0.68
Nickel-silver	Incompatible	0.68
Iron-lead	Incompatible	0.69
Aluminum-lead	Incompatible	1.4
Silver-lead	Limited compatibility	2.5
Magnesium-lead	Compatible	2.5
Zinc-lead	Incompatible	2.6
Silver-silver	Compatible	3.4
Aluminum-zinc	Compatible	3.9
Aluminum-nickel	Compatible	4.7
Aluminum-copper	Compatible	4.8
Aluminum-silver	Compatible	5.2
Aluminum-iron	Compatible	6.0
Iron-zinc	Compatible	8.4
Silver-zinc	Compatible	8.4
Nickel-zinc	Compatible	11
Zinc-zinc	Compatible	12
Magnesium-aluminum	Compatible	16
Zinc-copper	Compatible	18
Iron-copper	Limited compatibility	19
Silver-copper	Compatible	19
Lead-lead	Compatible	24
Nickel-magnesium	Compatible	28
Zinc-magnesium	Compatible	29
Aluminum-aluminum	Compatible	30
Copper-magnesium	Compatible	30
Silver-magnesium	Compatible	32
Magnesium-magnesium	Compatible	36
Iron-magnesium	Limited compatibility	38
Iron-nickel	Compatible	59
Iron-iron	Compatible	77
Copper-nickel	Compatible	81
Copper-copper	Compatible	130
Nickel-nickel	Compatible	290

index developed by Williamson. This influences the way the load is carried during sliding and thus determines the transition between severe and mild wear behavior.

In conclusion, I would say that our present state of knowledge is the result of the peculiar history of the study of wear. Back in the 1920's, 1930's, and 1940's, when all the tools necessary for the study of wear on a macroscopic scale were available, no research was done, possibly because everyone assumed that the available tools were inadequate. In the 1950's, when radiotracer techniques for studying wear on a microscopic scale became available, much fundamental research was carried out with highly gratifying results. However, we cannot fit these results into perspective because macroscopic data are still lacking. We do not know enough about the effect of geometry or of compatibility, and we know next to nothing about the wear properties of multiphase or composite materials, or nonmetals, elastometers, or otherwise. Rectifying this deficiency seems to me to be the first order of business.

C. N. Rowe (Mobil Research and Development Corporation, Princeton, New Jersey)

Dr. Archard has presented a rather comprehensive and balanced appraisal of the literature on wear; I am in general agreement. The fact that the paper focuses largely on the descriptions and mechanisms of unlubricated wear testifies that knowledge of the critical parameters is limited, and a need exists for a deeper insight into wear processes.

The purpose of these discussions is to stress viewpoints that complement or deviate from those of the principal paper. Using these guidelines, I shall describe a few viewpoints that may hopefully stimulate discussions or suggest routes leading to further advances in the study of wear.

Mild vs Severe Wear

The author discusses several proposed models for differentiating between mild and severe wear. First, the transition from mild to severe wear is associated with the merging, by subsurface deformation, of local asperities. Second, metal oxide is removed in the mild regime while large metal particles are removed in the severe region. The latter concept has received the most attention because oxide debris has been found in the mild region.

An alternate proposal not discussed is that small metallic particles, in contrast to oxide particles, are generated in the mild regime and then undergo rapid oxidation, thereby giving the appearance of metal oxide removal. Larger particles removed under severe conditions would be less rapidly oxidized and appear as metal. Consider this order-of-magnitude calculation using the same values for oxidation that Dr. Archard used. For a wear particle 10^{-6} cm in diameter coming off the surface at 1000° K (flash temperature), the time for complete oxidation is 0.001 second (β in the equation $k=\beta \exp(-Q/RT)$ is 1.2×10^2 gm^{2}cm^{-4}sec^{-1}). For a particle of 10^{-5} cm in diameter, the time is still only 0.1 second. If this calculation has merit, then small particles oxidize very rapidly, and the argument that oxide debris implies that oxide has been removed from the surface may be fallacious.

The formation of large particles in the severe region can result from the model discussed above (merging of local asperities), or may result from the aggregation of smaller wear particles in the contact region. Surprisingly, the aggregation of wear particles under sliding conditions has received little attention as a possible step in the sequence of events leading to severe wear.

Wear Factor K

Dr. Archard favors two interpretations for the apparent low values of K under unlubricated conditions: (1) one in about every thousand events results in a wear particle; and (2) each asperity must be rubbed about 1000 times before it results in a wear particle. As he states, we reported a modification of his (refs. 123 and 124) and Burwell and Strang's (ref. 125) expression for adhesive (metallic) wear in which it was assumed that wear occurs only at regions of metallic contact, and that the function of the lubricant is to control the amount of metallic contact (ref. 126). In this way the first-order contributions of the critical metallurgical and lubricant factors in K are segregated, thereby making it possible, in theory, to investigate each separately. Because this model leads to a different physical meaning of K than that described by Archard, it seems appropriate to review the model in greater detail and to examine it in light of available experimental results.

The basic expression is

$$\frac{V}{L}=K_m A_m=K_m \alpha A, \tag{16}$$

where K_m is a true proportionality constant characteristic of the sliding surfaces, and A_m represents the true metal-metal area of contact and may be related to the real area of contact A by α. It follows that α is a characteristic property, or a combination of properties, of any lubricant species within the area of contact. That is to say, α is determined by what happens to the lubricant film under sliding conditions, and is not necessarily determined by the nature of the lubricant molecules at the lubricant-metal interface outside the contact area. The weakness of this model is that it does not describe the sequence of events that leads to film disruption and the consequent metal-metal contact. It focuses on those local areas in which metal-metal contact has occurred.

The ensuing problem, as to what lubricant properties are critical in α, is a difficult one. Intuitively, the rheology of the surface or boundary film is important, but such knowledge of molecular films at the existing pressures and shear rates in a contact area is scanty. Because adsorbed films are the last line of defense against metal-metal contact a model was constructed in which the heat of adsorption of the lubricant is a critical parameter. Using this model for α, an average value of about 0.2 for K_m was calculated for copper sliding on steel with conditions of varying load, sliding velocity, and temperature, using n-hexadecane as lubricant (ref. 126).

In a nitrogen atmosphere the wear rate of a copper pin against a steel disk was 3.3×10^{-6} cm³/cm, giving a value of 7×10^{-3} for $K=K_m\alpha$. Again, using the heat of adsorption model for α and the value of 0.2 for K_m ($\alpha=3.5\times10^{-2}$), the calculated heat of adsorption of the nitrogen on the sliding metal surfaces is about 7 kcal/mole, which is reasonable for initial coverage of metals (ref. 127). Thus, these results suggest that the concept of unlubricated sliding is not strictly correct because adsorbed gases can reduce the metal-metal area of contact and thus function in some degree as boundary lubricants. If this supposition has merit, then the values of K reported, for example, in table 1 of Archard's paper, may be low because of adsorbed gases or vapors.

Because the proposed model for α has not been firmly established, a more direct determination of K_m that does not involve α is desirable. (Needless to say, wear rate measurements with clean surfaces, where $\alpha\rightarrow1$, at high vacuum would be of no use because the wear process would most likely be catastrophic and thus is unlike the mild wear process about which we are concerned.) One approach is to examine K_m in terms of the dimensions (shape factors) of wear particles.

Letting N be the total number of wear particles in a sliding distance of L, and n_m be the number of metal-metal junctions making up the true metal-metal area of contact at a given instant, equation 16 can be expressed as

$$\frac{Nv}{L}=K_m n_m a_m, \tag{17}$$

where v is the average volume of a wear particle, and a_m is the average cross-sectional area of a metal-metal junction. A reasonable assumption is that the junctions are initially circular, but assume an elliptical shape with the major axis along the direction of sliding because of tangential stresses and adhesion (ref. 128). Letting l equal the length of the junction and defining $L=N_0 l$, where N_0 represents the number of times the metal-metal area of contact is repeated in sliding distance of L, then the total number of metal-metal junctions, N_m, in a sliding distance of L, becomes $N_0 n_m$. Substituting and rearranging gives

$$K_m=\frac{N}{N_m}\cdot\frac{v}{a_m l}. \tag{18}$$

It is evident that the ratio N/N_m is the probability or frequency, P_m, that a given metal-metal junction will result in a wear particle. It is apparent that K_m and probability are not synonomous, as Archard stated, but are related through a wear particle shape factor.

Rabinowicz (ref. 128) observed that transferred wear fragments of copper adhering to steel have dimensions approximating a semiellipsoid. Finkin (ref. 129) studied the sizes and shapes of wear particles and found that they have a length, width, and thickness, all of which are significantly different. Most significant is the finding by both workers that there is a relation among the dimensions that is independent of the size of the fragments or particles (ref. 128). Assuming the particle shape to be an ellipsoid or a semiellipsoid, the volume will be $\pi lwh/6$, where l, w, and h are length, width, and thickness, respectively. The area of the junction a_m can be expressed as $\pi lw/4$. Thus equation 18 becomes

$$K_m = \frac{2h}{3l} P_m. \tag{19}$$

If P_m is indeed unity, K_m can be estimated from the shapes of wear particles. Conversely, if K_m can be determined by some other method, such as proposed in reference 126, then P_m can be estimated from the shape factors of wear particles.

Rabinowicz (ref. 128) analyzed the size and shape of transferred copper fragments from a radioactive copper pin sliding on an unlubricated steel surface. Table 6 gives the relative dimensions, which are independent of fragment size, and the calculated value of K_m using equation 19, and assuming P_m is unity. The agreement of the value of 0.20 with the value of 0.2 for copper sliding on steel in the presence of n-hexadecane lubricant tends to confirm the assumption that P_m is near unity, and may be taken as supporting evidence for the proposed adsorption model for a in reference 126.

Finkin (ref. 129) measured the size and shape of over 100 particles that resulted from a ring rubbing against an unlubricated flat for each metal in table 6. The calculated values of K_m, from the relative dimensions in table 6, show that K_m is dependent on the metal. The excellent agreement of the value for copper, with the calculated value from the copper-on-steel results of Rabinowicz, is noteworthy and reassuring. The overall agreement is suggestive that the probability P_m for a metal-metal contact is near unity, and that adsorbed gases or contaminants are probably responsible for the low values of K calculated from wear rate measurements under dry sliding conditions. In other words, this analysis suggests that two kinds of events occur simultaneously during dry sliding: one when the opposing load-supporting asperities have a metallic interface, and one when a boundary film exists at the asperity-asperity interface.

The weakness of this approach for estimating K_m is that the size of the particles are on the order of 50 to 1000 microns, which is several orders of magnitude larger than those found in the mild wear regime; therefore, it must be questioned whether the shape factors are similar in both regimes. It is thus apparent that a study of particle shape factors with different nonreacting lubricants would be helpful in evaluating K_m. It should be noted that Rabinowicz (ref. 130) has found that particle size does depend on the lubricant.

The factor K_m represents in part the tendency of materials to wear by adhesive bonds (welding). We now need to identify the metallurgical or physical properties that influence this tendency. Rabinowicz (ref. 131) considers the ratio of the surface energy to hardness to be a critical factor. Archard discussed the ratio of hardness to elastic modulus in abrasive wear and in elastic deformation of asperities. Table 7 compares the parameter H/K_m with H/γ; the values are expressed in this manner since H/K_m denotes "adhesive wear resistance," LW/V, the product of sliding distance and load per unit wear volume. Also tabulated are the relative abrasive wear resistance values reported by Khruschov (ref. 132) and the hardness/elastic modulus ratios. Notice that all four parameters exhibit the same trend of increasing resistance to wear. This qualitative agreement appears worthy of further investigation using other materials. That the adhesive and abrasive wear resistances of these four materials exhibit the same trend is of interest in itself, and requires further study. Apparently a close relation between adhesive and abrasive wear properties of metals has never been quantitatively demonstrated.

Another system in which a K_m value has been calculated is in the lubrication of graphite by vapors (ref. 133). Again, the adsorption

TABLE 6.—*Calculation of K_m*

Metal	l/w	l/h	K_m[a]
Copper-on-steel [b]	1.7	3.4	0.20
Copper [c]	1.59 ± 0.088	3.82 ± 0.305	0.18
Aluminum	1.71 ± 0.144	4.16 ± 0.328	0.16
Mild steel	2.11 ± 0.205	6.49 ± 0.519	0.10
Zinc	1.55 ± 0.077	7.38 ± 0.580	0.09
Copper-on-steel [d]	-------------	-------------	0.2

[a] Except for last value, the calculated values assume P_m is unity.
[b] From ref. 128.
[c] 95 percent confidence level (ref. 129).
[d] From ref. 126.

TABLE 7.—*Comparison of Wear Resistance Parameters*

[From refs. 131 and 132]

Metal	Adhesive wear resistance, H/K_m, kg/cm^2	Relative abrasive wear resistance, ϵ	Hardness/ surface energy, H/γ, cm^{-1}	Hardness/ elastic modulus, H/E
Aluminum	1.7	3.6	3.3×10^6	4.3×10^{-3}
Zinc	4.2	6.5	4.8×10^6	4.2×10^{-3}
Copper	4.4	10.0	7.1×10^6	6.7×10^{-3}
Mild steel	27.0	22.0	--------------	14.0×10^{-3}

model was used for α, and the calculated value of K_m is about 3.5×10^{-8}. Because K_m represents the height worn per unit sliding distance, the calculated thickness of the graphite removed, in a sliding distance equivalent to the diameter of the rider, is on the same order of magnitude as the interlamellar spacing. Thus, one may be attracted to the thought that the graphite wears by the removal of individual lamellae, which certainly would be consistent with the concept of the function of graphite in friction phenomena. However, Savage and Brown (ref. 134) found that the graphite debris consisted of platelets of approximately 3500 Angstroms in diameter and 20.4 Angstroms thick. Thus, the probability of forming a platelet would be $P_m=(3500/20.4)K_m=6\times10^{-6}$. This low value provides quantitative evidence for the fatigue process of graphites mentioned by Archard. Further evidence that the value of 3.5×10^{-8} is reasonable can be found in a wear rate study of graphite compacts in ultrahigh vacuum (10^{-9} torr) by Bryant and Gutshall (ref. 135). They report a wear rate of 0.006 mm/min. The calculated wear rate *assuming* that the apparent area of contact is the true area of contact is 0.005 mm/min ($h=K_mL=3.5\times10^{-8}\times2.4\times10^3\times60$ mm/sec$\times$sec/min$=0.005$ mm/min). Thus, the calculated value of K_m from wear rate data under lubricated conditions appears to support what may be considered conflicting observations. Clearly, additional wear rate studies are needed to achieve a fuller understanding of the graphite wear process, especially to determine if the wear rate is independent of the apparent area of contact for this system. It is also worth noting that pyrolytic crystalline graphite showed negligible wear in high vacuum (ref. 135) so that the K_m value must depend on the nature of the graphite substrate as well. In comparison with the metals, the low value of K_m may be indicative of the wear characteristics of lamellar solids.

As a final comment on K_m it must be stated that since the critical physical properties of the sliding surface are unknown, it is not possible to predict the effect of temperature (phase transformations, etc.), work hardening that may occur during sliding, minor changes in composition, or possible diffusion of lubricant or lubricant fragments into the substrate. These phenomena can most likely be investigated provided K_m per se can be experimentally measured.

Chemical Wear

Archard discusses in some detail the chemical wear theory proposed by Quinn (ref. 136). It is agreed that the analysis of wear in an oxygen atmosphere is worth pursuing. As Rowe and Fowles pointed out in a discussion of Quinn's paper (ref. 137), and as Archard stated independently in his discussion, the probability must be changed by the factor ξ/d, the ratio of oxide film thickness and junction diameter. Upon making this change, Quinn finds that ξ, and not ξ^2/d, may become constant at 10^{-6} cm at high flash temperatures. This constant value would be in agreement with recent work of Rabinowicz (ref. 138) thus lending support to the concept of a critical oxide film thickness.

As Archard stated, the question remains as to what temperature is important, flash or mean surface. The importance of this question is indirectly borne out in Quinn's work in that the K factors fall within a range of $10(7\times10^{-6}$ to $7\times10^{-5})$, while the calculated oxide film thicknesses vary over a range of $10^8(\times10^{-14}$ to 10^{-6} cm). It may be expected that if the oxide film is important in preventing wear (chemical wear will take place if the film becomes very thick), then a larger variation in the K factor should result. Moreover, Quinn observes the highest K factors when the calculated oxide film thickness approaches 10^{-6} cm. In contrast, Rabinowicz (ref. 138) believes a minimum thickness of 10^{-6} cm is required to prevent metallic contact. On this basis, films less than 10^{-6} cm should result in higher K factors because of metallic contact. Clearly, more work is needed to resolve the role of oxide films in wear. In all probability, mathematical models of both adhesive and chemical wear must be considered simultaneously in order to achieve better understanding. For example, by considering both wear processes one may be able to predict the minimum in the curve of the wear rate vs. oxygen concentration.

C. M. Wayman (University of Illinois, Urbana, Illinois)

I am pleased with the invitation to participate in this interdisciplinary symposium and to present some comments on Dr. Archard's paper on wear. It is always pleasant to expose one's self to a new subject through a broad, well focused review such as Dr. Archard's. As do most interesting subjects, Dr. Archard's account makes clear the numerous processes in wear. The complexity of wear reminds me of the welding process, where, as a former professor pointed out, steelmaking, solidification, heat treatment, and deformation all occur at essentially the same time. It seems clear that the questions raised by wear far outnumber those which can satisfactorily be answered at this time. But this is a healthy situation, and it is gratifying to see here a mustering of forces to engage the problem in an interdisciplinary way. My participation here finds me previously unexposed to the general subject of wear; thus I am fortunately not involved in the existing controversies, but unfortunately some of my comments will reflect my ignorance.

In this discussion, aspects of lubricated wear are not considered, in accordance with the mainstream of Dr. Archard's paper.

In general, I find myself wondering about the condition of various test specimens; they are not well specified in many cases. For example, terms such as *smooth* and *flat* are not well defined, whereas these can be resolved to a scale of at least a few hundred angstrom units by interference microscopy. Although the general wear situation may change when surfaces have settled down or run in, it is not clear that the initial surface condition of pins, rings, etc., is not important. Certainly the apparent contact area would more closely approach the true contact area if the rubbing members were carefully lapped, with, say, a vibratory polisher or an electropolisher. Such techniques can produce flatness to within a few hundred angstrom units, and in terms of Dr. Archard's size classification relevant to mild vs. severe wear, may not involve the blending of small-scale contacts into large-scale contacts. One thus wonders about the wear behavior of two pin-ring combinations, the only difference otherwise being in initial flatness, perhaps suggesting a difference between atomic and engineering wear. One should also consider that even conventionally well-machined surfaces are severely work-hardened to begin with, whereas a spark machined surface, subsequently electropolished, is relatively soft. The recently developed technique of scanning electron microscopy should prove invaluable in studying surfaces initially as well as after various degrees of wear. Likewise, the wear particles themselves could be studied in greater detail. The great depth of focus of the scanning electron microscope (coupled with its high resolution), employing secondary electron emission, is naturally suited to metallic specimens.

Along the same lines, I could not determine if the pins used in rubbing experiments were taken from a cold drawn or swaged rod (having characteristic grain orientation), or were removed from an annealed block of material having essentially randomly equiaxed grains. In the former case one not only has to contend with the fact that the specimen is *initially* coldworked (rather than during wear), but also that a preferred grain orientation may exhibit a pronounced consequence, especially since friction is related to rubbing plane orientation. Likewise, the grain size per se must be important since atoms in a grain boundary, having on the average fewer neighbors, have a tendency to acquire more neighbors. This tendency would be a source of friction. (Dr. Archard points out that wear rates can vary by 10^5 whereas frictional coefficients may vary by only a factor of three or so. In this connection it is not certain that the measured coefficients of friction are the true ones because of the difference between the true and apparent areas of contact.) Also, the well annealed (soft) vs coldworked nature of rubbing specimens should be important in an overall wear experiment, as evidenced by the work on steels which suggests that only mild wear occurs once the rubbing members reach a certain hardness, (or flow stress). In fact one wonders how wear results would compare if, other things being equal, both annealed and coldworked pins were made to rub on a given ring. Thus it would seem important to characterize thoroughly the initial state of rubbing members at least as to grain size and randomness, hardness, and initial surface flatness.

Much useful information has been and will be obtained from pin-ring tests, but one wonders if this method tends to obscure certain trends. Since the pin is usually small relative to the ring or rubbing substrate, the bulk of the thermal effects due to rubbing may be concentrated in only one member of the couple (the pin) to the extent that dynamic conditions are always present, and that both softening, diffusion effects, and hardening processes are manifest concurrently. One also wonders about the effect of the thermal conductivity of the rubbing members. Apart from other differences, one might expect different effects from, say, brass, a good conductor, and stainless steel. In the latter case, thermal effects due to rubbing would likely be more localized due to relatively poor conduction. Perhaps a systematic investigation based on thermal conductivity would yield interesting information.

The formation of oxides during wear raises questions about which little is apparently known. The view has been presented that mild wear may correspond to the rupturing away of small oxide particles, $\sim 10^{-5}$ cm on a side, whereas severe wear involves welding and the formation of much larger (metallic) wear particles. One could question to what extent oxides behave like jeweler's rouge: smoothing out the rubbing surfaces and preventing any large scale grab or adherence leading to severe wear. It seems clear that the absence of oxides would exclude mild wear, and thus one may think about some useful experiments performed under refrigerated conditions. Almost any rubbing experiment could be carried out at liquid nitrogen temperature, thus deterring oxide film growth. Would only severe wear then result? It would be interesting to compare two identical experiments, one carried out at room temperature and the other at, say, 77° K. Equivalently the same kind of information may result from experiments carried out in a reducing atmosphere, or by working with noble metals such as gold. Noble metals, though, as Dr. Archard points out, encounter special difficulties because of a tendency for clumps of particles to build up at the leading edge of a rider; the load is borne by the clumps rather than by the rubbing pin. This kind of effect, however, may not really be a point of peculiarity, but may only reflect certain metallurgical differences. For example, a room-temperature wear comparison of, say, gold vs cartridge brass is not a fair one, even though both are fcc. In the first place gold is generally produced of 99.99 or 99.999 percent purity, and is soft because of no appreciable solution-hardening. Cartridge brass, on the other hand, containing ~30 percent zinc, is highly solution-hardened. (For example, the critical resolved shear stress of 99.9 percent Ag is four times (200 psi) that of 99.999 percent Ag.) Likewise, it would not seem fair to compare say Au to Pb of comparable purity in room-temperature wear experiments because the temperature at which the experiments are carried out would be $0.2T_m$ for Au in contrast to 0.5 T_m for Pb (where T_m is the melting temperature in °K). Thus when an allowance is made for the purity, test temperature, and load, relative to T_m, the wear situation may not be so different after all. Of course the oxide film problem can also be minimized in the experiments now being conducted in ultrahigh vacuum. Much will be learned about surfaces and adhesion in these UHV experiments.

Numerous effects can be superimposed during a typical wear experiment, as evidenced by the observation that a low carbon steel may be made harder during wear than is possible by heat treatment alone. This probably indicates the combined effects of martensitic hardening (which is reported), work-hardening, and/or the formation of brittle phases such as oxides and intermetallic compounds. The high temperatures in some of the wear experiments may cause significant diffusion leading to both hardening and softening processes. Softening processes would be those such as recrystallization, tempering, and overaging; hardening processes would include solution hardening by alloying, the formation of new phases such as brittle oxides and intermetallics, and martensitic transformation. Although it has been well taken that rubbing steels can form martensite, it should also be considered that the temperatures involved may also cause tempering and hence softening. In view of the many likely diffusion effects which can occur during wear experiments, it would seem that greater insight might be obtained by microprobe analysis of various worn surfaces and subsurfaces.

Another puzzling question about wear is the nature of the deformation processes involved. The severe in-rubbing of surfaces leaving gouges and resulting in the gross removal of particles (or clumps of particles), coupled with the observation that worn debris is in a hardened condition, suggests the importance of plastic deformation. This also follows from the correlation of wear rate to the ratio of yield strength and elastic modulus in certain experiments. However, the brittle rupturing or fracture of oxides and metallic asperities may be an important mechanism. One might also consider that the combined effects of load and heat would enhance the "squashing out" of asperities by plastic deformation, and in this way the true contact area could significantly increase, leading to consequences in the wear behavior. On the other hand, in certain works, it is indicated to be unclear whether the wear process involves elastic or plastic deformation, and there are suggestions (particularly for rolling contact) that fatigue may be an important process in wear.

As far as deformation mechanisms are concerned, there are several experimental avenues which might produce results. It seems important to learn more of the deformed state of surface and subsurface regions after wear has occurred (e.g., depth of the wear affected zone). Transmission electron microscopy might be valuable here. This is particularly so now that techniques are available for removing sections from bulk pieces to reveal conditions at or near the surface (e.g., dislocation density and cell structure). Such experiments have recently been performed in my own laboratory to study surface oxidation effects. Also to be considered in greater depth is the fatigue deformation. One wonders about the outcome of wear experiments when the surfaces have been preconditioned with favorable stress distributions. For example, fatigue life can be extended substantially by shotpeening. Such treatment produces residual compressive stresses that partially nullify cyclical tensile stresses. One might similarly expect a difference in wear behavior in, say, peened vs unpeened rubbing members if indeed microfatigue mechanisms are responsible for the removal of worn particles. Following the argument presented by Dr. Archard, if $K = 10^{-3}$ and if microfatigue is involved, then a typical asperity must be rubbed 10^3 times before it is removed. Fatigue failure at 10^3 cycles, by comparison to typical *S–N* curves, indicates that the stresses thus involved would not be greatly different from static yield stresses, signifying again the importance of plastic deformation in wear processes.

In Dr. Archard's paper, several experimental points that were not clearly resolved may be worth brief mention since they were sources of confusion to me as an uninitiated reader. It is reported that worn particles are significantly harder than the members from which they are removed. One wonders how this is established, since it is difficult to determine the microhardness of such small particles or aggregates, even though their chemical activity should be higher in the cold-worked, or hardened, state.

The transfer of material from one rubbing member to another, such as brass to stellite, is an interesting problem. One wonders about the basic principles involved: Does this vary according to the chemical dissimilarity, or alloying tendency, of rubbing members, or does a material, once mechanically embedded in an opposing surface, attract species of its own kind from, say, a pin which is comparatively hotter due to frictional heat? It is reported that welding (which should favor severe wear) is most pronounced when materials are mutually soluble. Thus should the rubbing of two metals which are mutually insoluble lead to the absence of severe wear? In addition, one wonders what is involved in transfer in cases where the substrate is covered by an oxide film. If, as is suggested, mild wear is associated with oxide film disruption, the K factor should be reduced by d/a (where a is the radius of the contact area, and d is the film thickness). Is this borne out by experiments where oxide films are definitely established?

It is also shown that once a critical hardness is reached, only mild wear occurs, and thus one wonders if a severely work-hardened pin does not exhibit severe wear. Since oxide particles (or intermetallics) are evidently generated in profusion during wear, it appears that dispersion hardening effects due to such phases should not be overlooked; and, it seems possible that while some oxides definitely exert an abrasive influence, others may perform a lubricating function and thus lead to contrastingly different wear behavior.

Finally, it is encouraging to see work like the recent investigations reported by Buckley in which fcc, bcc, and hcp crystals of known orientation are rubbed in ultrahigh vacuum. Systematic correlations between friction, welding, and crystal structures are evolving. Likewise, the importance of the relative orientation of close-packed crystal planes has been established; in each case those of highest atomic density result in lowest friction. There is much to be said for the single crystal type of investigation, and especially those involving nonstandard materials. For example, much information on dislocation mobility and motion has been obtained by studying lithium fluoride, not a well-known compound. Our present knowledge of work-hardening of metals has been increased substantially by single crystal investigations. It also seems that the study of a particular phenomenon in a material such as steel may not be efficient. As a case in point, our knowledge of martensitic transformations has been extended substantially by studying martensitic reactions in materials other than steel. Indeed, because of various complexities involved, one is not certain today why martensite in steels is hard, although hardened steels have been employed technologically for centuries.

R. S. Fein (Texaco Research Center, Beacon, New York)

Dr. Archard pointed out that one of the most completely understood wear mechanisms is the severe wear of 60/40 brass rubbing against hard materials. He also indicated that this severe wear mechanism may be typical for a soft metal against a harder one. The purpose of this discussion is to point out: (1) some aspects of this mechanism which need further study, and (2) the implications of these aspects for both basic research and screening tests for bearing materials and lubricants. These points can best be made by reviewing the results of a four ball machine study.

The study was conducted with 52 100 steel rotating balls and "bronze" stationary balls.* This study under lubricated conditions indicated the same mechanism of severe wear as the dry sliding studies (refs. 139 to 141) discussed by Dr. Archard. Thus, transferred "bronze" was observed on the steel ball and, with a light solvent-neutral oil lubricant, "bronze" ball wear and friction were shown to depend on the history of the steel ball surface (not on the "bronze" ball surface history). Runs under dry conditions, and with cetane as a lubricant, produced wear coefficients and smooth wear-versus-time curves in good agreement with Kerridge and Lancaster. This agreement was achieved despite the differences between machine configurations and, probably, test specimen materials.

Lubrication by the solvent-neutral oil reduced the steady-state wear coefficient by over two orders of magnitude from that under dry or cetane-lubricated conditions. Further, as illustrated in figure 13, the lubricant introduced additional features in the "bronze" ball wear-versus-time relationship.** In the range of sliding distance marked stage II, changes in slope (which appear small on a log-log scale) were observed corresponding to changes in volumetric wear rate of about two orders of magnitude. Friction coefficient changes were generally found correlating with changes in slope of the wear curves; however, there were significant sliding distance displacements. An example of the latter is indicated by the large increase in friction near the end of stage II and the large increase in wear; in this case the friction increase precedes the wear increase by about 20 000 cm of sliding. Similar data

---

*The quotation marks on "bronze" are used because subsequent analysis of these commercial balls indicated a brass-like composition.

**The data in figure 13 represent 23 individual runs. The sequence of runs was randomized and each run was made with new test specimens and oil. Fewer than 23 points are shown due to essentially exact replication of wear and friction results near the sharp maximum and minima in the friction curves and loss of one friction point by equipment failure.

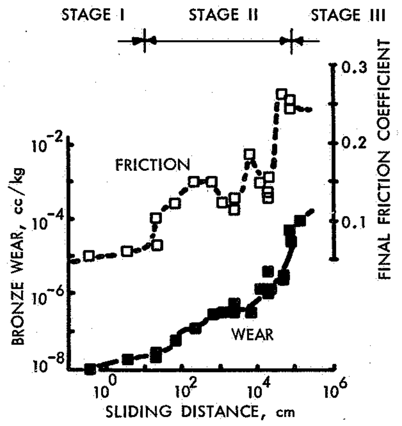


Figure 13.—Friction and wear transients of 52100 steel on "bronze," using solvent-neutral oil; load, 40 kg; 9.23 rpm; 210° F.

were also obtained at speeds of 0.34 rpm and 1800 rpm and loads of 20 and 40 kg. In all cases, qualitatively similar but quantitatively different stage II transients were observed, and the limits of the stages were different. By contrast, stage I and stage III (steady-state) data on friction and wear per unit load as a function of sliding distance, appeared quantitatively independent of operating conditions.

It is apparent that even for this fairly well understood wear mechanism, much remains to be learned about how the sequence of wear steps can lead to the observed stages and transients and why these depend so much on operating conditions. Further, it would be of interest to know the reasons for the indirect relationship between the friction and wear processes sometimes indicated by high wear rate and low friction (and vice versa) in stage II.

Table 8 presents data taken with three lubricants and using the average steady-state (i.e., stage III) wear coefficient, on the constitution of the "bronze" at three locations in the wearing system. Obviously, there is depletion of the "bronze's" zinc which has been transferred to the steel, and then further depletion in the loose wear debris. Further, the wear appears to decrease with increasing zinc depletion. These results indicate that there is a chemical factor in the wear process in need of further study.

Further elaboration of Dr. Archard's severe wear mechanism should conceivably follow the direction outlined above. Much more important than the indicated future study of severe wear, however, are the needs for more careful basic research and materials screening. As Dr. Archard pointed out, all wear processes seem to occur by a series of both concurrent and consecutive physical and chemical steps. As a consequence of the competitive and sequential nature of these steps, transients and sensitivities to operating conditions would be expected.

Unless these phenomena are recognized, misleading experimental results will follow. To illustrate, prior to obtaining the above results (table 8), an exploratory study of the effects of lubricant constitution was carried out using fairly arbitrarily chosen operating conditions of load and run time. The results showed a rational correlation of friction with lubricant composition at the end of the run. However, when the wear data were subsequently examined, they were not found to be rationally related to the lubricant. This observation led to the study discussed above (fig. 13), showing that the original screening test conditions gave results in the stage II or transient region. Thus, the friction (and wear) values obtained from this screening test provided misleading lubricant comparisons. This illustrates the need to exercise considerable caution before attaching significance to results from arbitrary tests. This is already widely recognized by those sophisticated in friction, wear, and lubrication research and development. However, it appears to need emphasis because this recognition often is not evident in publications or in the use of screening or specification tests. Consequently, people tend to be mislead by results which may be essentially meaningless.

D. G. Flom (General Electric Company, Valley Forge Technology Center, Philadelphia, Pennsylvania)

I am interested particularly in Dr. Archard's last paragraph in which he cites the special conditions at rubbing contacts; namely, contact size of 10^{-2} cm, speeds of 100 cm/sec, pressures greater than 10^5 psi, temperatures of 500 to 1000° C, and durations of 100 microseconds. These conditions are strongly suggestive of those obtained during solid-solid impact which in turn can result in spallation.

During the past ten years, many laboratories have developed tech-

TABLE 8.—*Effect of Lubricant on Wear and "Bronze"* [a] *Constitution* [b]

Lubricant	Average wear coefficient × 10^6	Atomic Ratio [c], Zn/Cu		
		Wear debris in oil	Steel wear surface	"Bronze" wear surface
3,9-diethyltridecanol–6	0. 58	0. 35	0. 43	0. 63
Light solvent neutral oil	1. 4	0. 45	0. 47	0. 63
SAE 90 mixed base oil	4. 5	----------	0. 50	0. 63

[a] "Bronze" of brass-like composition.

[b] Steel on "bronze" at 1800 rpm for 15 hr (3.7×10^6 cm); load, 20 kg; 100° F.

[c] New "bronze" Zn/Cu=0.63.

niques for producing high-speed impacts. One method is the "flying plate" technique in which a 6-inch compressed-gas gun is used to propel a flat plate or "slapper" within the barrel of the gun so as to impact a stationary flat plate or "target" situated at the middle or end of the gun. Ordinarily, a thin slapper of lower inertia than the target is used if one wants to determine the threshold of spallation for a material. What happens at impact is that a compressive stress pulse is propagated through the target at shock velocity, and, on reflection from an interface of lower impedance (e.g., air), a tensile stress is developed which can be sufficient to cause fracture, or spallation, if the initial impact velocity is high enough. The fact that spallation usually occurs from the rear surface of a target should not deter us from considering it in connection with wear since most compressive pulses develop a tensile tail; consequently, for this and other reasons, front surface spallation is not uncommon.

For typical plastics and composites, spallation occurs at pressures of 0.5 to 5 kilobars (i.e., $\sim$ 7000 to 70 000 psi) at impact velocities of perhaps 0.05 to 0.2 mm/μ sec (165 to 660 ft/sec) and with impact durations of a few microseconds. The exact numbers depend on the dynamic equations of state for the materials involved.

For metals, spallation occurs at pressures of greater than 10 kilobars but usually less than 100 kilobars. The temperatures during impact are more difficult to specify although it is certain that these can be high. Also, if the gun is not sufficiently evacuated ahead of the flying plate, gas temperatures owing to compression just prior to impact can be high enough to scorch the face of the target. It is interesting to note that the results of hypervelocity particle impact experiments are explained best by hydrodynamic equations, indicating that the solids in impact behave like liquids.

The idea that fracture and spallation can contribute to wear has occurred to others (ref. 142). However, it does bring to the subject a viewpoint somewhat different from that customarily employed. I think one of the important points is that, in wear, we should be thinking in terms of dynamic properties and equations of state of the materials involved rather than in static properties.

J. J. Bikerman (Horizons Incorporated, Cleveland, Ohio)

It is interesting to imagine what Coulomb would have said on the elementary mechanism of wear in air. When two metal asperities are "in contact" in air, there is no atomic metal-to-metal contact between them. They are always separated from each other by adsorbed oxygen, water and so on. However, stress is readily transferred through the adsorbed film.

If the average tilt of the support asperities were, say, 10°, then a tangential force proportional to tan 10° would be sufficient to raise the slider over the obstacle and to continue sliding. On the other hand, if a particular hill had a slope of, say, 30°, the above force would be insufficient. Consequently, in the latter case, the force which was previously spread over many contacts would now be concentrated at the steeper hill. As the asperities are small to begin with and as the "contact" area (through which the stress is transmitted) is small compared with the surface area of an asperity, the local stress (i.e., the ratio of the concentrated force to the microscopic "contact" area) is very high. It may be high enough to initiate a crack in one of the two hills pressing each other.

This mechanism would account for the fact that only one in many "contacts" results in the formation of a wear particle. It also leads to the expression $\delta V \sim a^3$ (where δV=volume of removed material, and a=radius of contact area) as the chunk broken off an asperity, on the average, may be expected to be greater when the hill is more voluminous. It may, moreover, supply the mechanism for the decrease of the wear rate on an increase in the speed of sliding. When this speed increases, the depth to which a hill of the slider descends in a valley of the support gets smaller (as the hill simply has no time to descend far) and, consequently, the probability of hitting a particularly steep slope decreases.

J. A. Schey (IIT Research Institute, Chicago, Illinois)

It is rather regrettable that some of the most sophisticated work on the mechanisms of wear has been conducted with 60/40 brass as the specimen material. While this material has many desirable technological properties in addition to low price, it is rather poorly defined, and it is unlikely that results could ever be reproduced on a different batch of material. For varying reasons, the value of much of the work on friction and wear is greatly reduced by the unknown or unidentified condition of the experimental materials, especially steels and other alloys. Material properties are highly dependent on prior thermomechanical history.

I would suggest that a few simple steps be taken to insure greater reproducibility and relevance of experiments and to allow for the integration of the work of several researchers into a logical, generalized treatment. First, a list of preferred experimental materials (both metals and lubricants) should be drawn up. These materials should be used unless there is some pressing reason to select a less well-defined material. Second, a basic stock of pedigreed materials, accessible to all researchers, should be established encompassing steels and at least the most important copper nickel and aluminum base alloys. Third, whether as writer, editor, or referee, we should assure that only work with properly identified materials and experimental conditions be published.

LECTURER'S CLOSURE

I am greatly indebted to the discussers of my paper for their contributions; they have effectively broadened and supplemented my presentation by their suggestions, comments, and questions. In this reply it would be appropriate to isolate and discuss those items which are regarded as most important in the further development of the subject.

In my first draft, wear was described as the Cinderella of the trio of subjects friction, lubrication, and wear. In his review Dr. Tabor has responded by referring to it as "a wayward, capricious child of mixed and uncertain parentage."* This well expresses the reaction of those involved in the study of wear. Where else in basic research would one expect to find useful information from a 60/40 leaded brass as opposed to high purity copper?

It is probably fair to comment that this particular 60/40 brass was first used for a detailed study because of its exceptionally reproducible results. Indeed this has been confirmed by experiments carried out in different laboratories under the auspices of the OECD Group on Wear of Engineering Materials (ref. 15); a brass of essentially similar composition to that used in the original work has been included in this cooperative program. It is, perhaps, through programs such as this that the need for standard materials, as suggested by Dr. Schey, can best be met. It is also worth recording that Welsh's work on steels, outlined in my paper, was performed with a special range of steels reserved, by the British Iron and Steel Research Association, for experimental purposes. The description of Welsh's work in my paper was very much condensed, but the reader is referred to the original papers for a detailed description of the experimental conditions including the thermomechanical history of the specimens.

I think that Lancaster's meticulous study of the wear of 60/40 brass should be defended as being relevant and timely at that particular stage in the development of our knowledge of wear. It showed that wear can be a "competitive and sequential" process, as Dr. Fein has stressed. As emphasized in my review, one of our present problems is to determine the extent to which the results of this particular study (brass on steel) are applicable to a wider range of materials. Only further experiments can provide the answers. Dr. Fein suggests that the introduction of a lubricant can also make radical changes. In my review, the role of the lubricant in affecting the way in which the load is distributed was emphasized. Dr. Fein also comments on the possible

---

***See Dr. Tabor's lecture on Critical Appraisal and Research Opportunities—The Lubrication Research Viewpoint.**

role of the lubricant in controlling the chemical nature of the environment.

Much emphasis at this symposium has been on the desirability of well defined surface conditions. For example, Professor Wayman has discussed the advantages that might accrue from the use of single crystals, very smooth surfaces, variations in grain size, and strain free surfaces. The point worthy of emphasis here is that such experiments are probably relevant to only one stage of a multistage wear mechanism (for example, the adhesion stage). Moreover, under equilibrium wear conditions, the rubbing process generates a characteristic surface condition which is the result of many traversals of each point on the surface. This surface condition, complicated as it may be, is often a vital part of the wear process. Experiments made under carefully controlled conditions have a part to play in wear study, but their relevance to wear processes resulting from repeated sliding remains to be proved.

An alternative, and complementary, approach to wear studies is the application of the full range of modern physical techniques to the study of equilibrium wear conditions. Both the surfaces and the debris might be studied in this way, as Professor Wayman has outlined. He also makes a very important point when he refers to the assymmetry present in pin and ring (and pin and disk) experiments because of size differences. The interested reader is referred to Welsh's work on steels for a discussion of some possible influences of this type.

In discussing thermal effects in wear the major effects are the overall temperature rise of the rubbing members and the "flash temperatures" occurring at the true areas of contact. The hardening effects discussed in connection with steels may arise from rates of heating and quenching which are rapid compared with those found in normal metallurgical heat treatment. The superposed conditions of high pressures were stressed in the closing stages of my paper and Dr. Flom suggests that the same combination of conditions might well be obtained in other ways. The extent to which the flash temperature directly affects the experiment through thermal softening is raised by Professor Wayman. The evidence of Lancaster's experiments, discussed in my paper, is that the bulk rather than the flash temperature is the influential variable. This conclusion and its explanation clearly deserve further examination. In any discussion of thermal effects, the contribution of Dr. Johnson in connection with Professor Ling's paper should also be noted. He cites the work of Barber who demonstrated that conventional ideas of the duration of flash temperatures may not always be valid.

Some broad questions of mild and severe wear also deserve comment. Dr. Rowe makes a valid point when he suggests that aggregation of small metallic wear particles could be a mechanism in the scale of size transition from mild to severe wear. However, I am not aware of any experimental evidence which would support this suggestion. In discussing the three possible mechanisms of mild wear (or fretting) proposed by Waterhouse, Dr. Rowe stresses the role of flash temperatures in the oxidation of small metallic particles. In this process sliding speed, upon which flash temperatures are markedly dependent, should be an influential factor. Professor Rabinowicz doubts whether loose oxide debris could be an important cause of abrasive wear. If this conclusion could be fully substantiated, it would clearly help in the development of the subject by removing one possible mechanism of mild wear from the list. In the meantime a number of investigators have suggested that Al_2O_3 plays an abrasive role in fretting when aluminium alloys are used.

In my paper the need for a physical interpretation of K factors was stressed. Both in his discussion and in his published papers, Dr. Rowe has emphasized the idea that the K factor, as used in my paper, can be divided into two parts. Firstly, he postulates a value of K (namely K_m) that is truly representative of metal-to-metal contact. Secondly, he introduces a factor, α, which is characteristic of the lubricant species within the area of contact. Thus his theory appears to be applicable to a wear process concerned with the removal of metallic particles. In my review the value of α, derived from the adsorption model, was emphasized because it appeared to provide a basis for the calculation of α (and therefore K) in terms of physical quantities derived from independent measurements. In his discussion Dr. Rowe points out that other interpretations of α are possible, but whether they can similarly be calculated is not yet clear. Dr. Rowe's discussion also highlights, in my view, the difficulties which are met in assigning a true value to K_m.

In considering the major wear factors, both Dr. Rowe and Professor Rabinowicz have suggested further investigation of factors such as H/E and H/γ. It is relevant to point out that H/E forms part of the Greenwood and Williamson plasticity index discussed in Dr. Williamson's paper. H/E is therefore a relevant factor for comparison only when the surface topography remains constant. Professor Rabinowicz extends the discussion of these factors to suggest a strategy of wear analysis based upon the inclusion of a number of such nondimensional parameters. Two difficulties seem to arise, however. Firstly, as Professor Rabinowicz himself points out, many of the factors which have been proposed are interrelated. Secondly, these factors are difficult to assign values which are appropriate to a wearing interface under equilibrium conditions. It is much easier to see the relevance of these factors when considering the more idealized experiments described in Dr. Merchant's paper. Perhaps our greatest need is for a firmer bridge between these fundamental studies of friction and adhesion and the more complex array of facts which must characterize any complete description of our present knowledge of wear.

REFERENCES

1. Fink, M.: Wear Oxidation, A New Component of Wear. Trans. Am. Soc. Steel Treating, vol. 18, 1930, p. 204.
2. Mailander, R.; and Dies, K.: Contributions to the Investigation of the Processes Taking Place During Wear. Arch. Eisenhutt Wes., vol. 16, 1943, p. 385.
3. Dies, K.: Processes Taking Place During Wear in Conditions of Pure Sliding and Dry Friction. Reibung und Vershleiss, VKI Verlag (Berlin), 1939.
4. Kehl, B.; and Siebel, E.: Arch. Eisenhutt Wes., vol. 9, 1939.
5. Siebel, E.; and Kobitzsch, R.: Verschleisserscheinungen bei gleitender trockener Reibung, VKI Verlag (Berlin), 1941.
6. Holm, R.: Electric Contacts Handbook, Springer-Verlag (Berlin), 1958, p. 199.
7. Burwell, J. T.; and Strang, C. D.: On the Empirical Law of Adhesive Wear. J. Appl. Phys., vol. 23, p. 18.
8. Dorinson, A.; and Broman, V. E.: Extreme Pressure Lubrication and Wear. The Influence of Contact Stress Upon Metallic Wear. ASLE Trans., vol. 3, 1960, p. 170.
9. Ling, F. F.; Whitely, R. L.; Ku, P. M.; and Peterson, M. B.; eds.; Friction and Lubrication in Metal Processing, ASME, 1966.
10. Archard, J. F.; and Hirst, W.: The Wear of Metals Under Unlubricated Conditions. Proc. Roy. Soc. (London), vol. A236, 1956, p. 397.
11. Bowden, F. P.; and Tabor, D.: The Friction and Lubrication of Solids, Oxford Univ. Press, 1950.
12. Cocks, M.: Proc. Phys. Soc. (London), vol. B67, 1954, p. 238.
13. Dyson, J.; and Hirst, W.: The True Area of Contact Between Solids. Proc. Phys. Soc. (London), vol. B67, 1954, p. 309.
14. Rabinowicz, E.; and Tabor, D.: Metallic Transfer Between Sliding Metals; An Autoradiographic Study. Proc. Roy. Soc. (London), vol. A208, 1951, p. 455.
15. DeGee, A. W. J.: Review of the Activities of the OECD Group of Experts on the Wear of Engineering Metals. Lubrication and Wear Group Convention, Inst. Mech. Engrs., May 1967.
16. Kerridge, M.; and Lancaster, J. K.: The Stages in a Process of Severe Metallic Wear. Proc. Roy. Soc. (London), vol. A236, 1956, p. 250.
17. Haymana, F.: Rept. Castings Res. Lab., Waseda Univ., 1958.
18. Steijn, R. P.: Investigation of Dry Adhesive Wear. J. Basic Eng., ASME, vol 81D, 1959, p. 56.
19. Lancaster, J. K.: The Influence of Temperature Upon Metallic Wear. Proc. Phys. Soc. (London), vol. B70, 1957, p. 112.
20. Hirst, W.; and Lancaster, J. K.: The Influence of Speed on Metallic Wear. Proc. Roy. Soc. (London), vol. A259, 1960, p. 288.
21. Rabinowicz, E.: Friction and Wear of Materials. John Wiley & Sons, 1965, Chap. 6.
22. Lancaster, J. K.: The Formation of Surface Films and the Transition Between Mild and Severe Wear. Proc. Roy. Soc. (London), vol. A273, 1963, p. 466.

23. Johnson, R. L.; Swikert, M. A., and Bisson, E. E.: NACA TN 1442, 1947.
24. Welsh, N. C.: The Dry Wear of Steel. I. The General Pattern of Behavior; II. Interpretation and Special Features. Phil. Trans. Roy. Soc. (London), vol. A257, 1965, p. 31.
25. Dawson, P. H.; and Fidler, F.: The Behaviour of Chromium Steel in Large High Speed Bearings. AEI Engineering, vol. 2, no. 2, 1962.
26. Dawson, P. H.; Fidler, F.; and Rowley, P.: Wire-Wool Type Failures in 3% Cr–½% Mo Steel Bearings—A Preliminary Evaluation of Some Palliatives. Proc. Inst. Mech. Engrs., vol. 179, Pt. 3J, 1964–65, p. 27.
27. Hother-Lushington, S.: Failure of Turbine Bearings Using Chromium Alloy Steel Journals. Proc. Inst. Mech. Engrs., vol. 180, Pt. 3K, 1965–66, p. 243.
28. Kerridge, M.: Metal Transfer and the Wear Process. Proc. Phys. Soc. (London), vol. B68, 1955, p. 400.
29. Archard, J. F.; and Hirst, W.: An Examination of a Mild Wear Process. Proc. Roy. Soc. (London), vol. A238, 1957, p. 315.
30. Coffin, L. F.: A Study of the Sliding of Metals With Particular Reference to Atmosphere. Lub. Engr., vol. 12, 1956, p. 50.
31. DeGee, A. W. J.: The Friction of Gold-Silver Alloys Against Steel. Wear, vol. 8, 1965, p. 121.
32. Machlin, E. S.; and Yankee, W. R.: Friction of Clean Metals and Oxides With Special Reference to Titanium. J. Appl. Phys., vol. 25, 1954, p. 576.
33. Rabinowicz, E.: Friction and Wear of Materials, John Wiley & Sons, 1965, ch. 4.
34. Hirst, W.: Wear. Met. Rev., vol. 10, 1965, p. 145.
35. Archard, J. F.: Contact and Rubbing of Flat Surfaces. J. Appl. Phys., vol. 24, 1953, p. 24.
36. Archard, J. F.: Single Contacts and Multiple Encounters. J. Appl. Phys., vol. 32, 1961, p. 1420.
37. Frenkel, J.: A. Physik, vol. 37, 1926, p. 572.
38. Kittel, C.: Introduction to Solid State Physics. John Wiley & Sons, 1962, p. 536.
39. Uhlig, H. H.: Mechanism of Fretting Corrosion. J. Appl. Mech., vol. 21, 1954, p. 401.
40. Yoshimoto, G.; and Tsukizoe: On the Mechanism of Wear Between Metal Surfaces. Wear, vol. 1, 1957–58, p. 472.
41. Avient, B. W. E.; Goddard, J.; and Wilman, M.: Proc. Roy. Soc. (London), vol. A258, 1960, p. 159.
42. Kruschov, M. M.; and Babichev, M. A.: Friction and Wear in Machinery, English translation by ASME, vol. 22, 1956, p. 1, and vol. 12, 1958, p. 5. A review of experiment is also to be found in ref. 21, chap. 7.
43. Burwell, J. T.: Survey of Possible Wear Mechanisms. Wear, vol. 1, 1957–58, p. 119.
44. Kragelskii, I. V. (L. Ronson and J. K. Lancaster, trans.): Friction and Wear. Butterworths Scientific Publ. (Wash., D.C.), p. 98.
45. Mulhearn, T. O.; and Samuels, L. E.: The Abrasion of Metals: A Model of the Process. Wear, vol. 5, 1962, p. 478.
46. Kruschov, M. M.; and Babichev, M. A..: Research on the Wear of Metals. Ch. 17, NEL Translation No. 892.
47. Sedricks, A. J.; and Mulhearn, T. O.: Mechanics of Cutting and Rubbing in Simulated Abrasive Processes. Wear, vol. 6, 1963, p. 457.
48. Sedricks, A. J.; and Mulhearn, T. O.: The Effect of Work-Hardening on

the Mechanics of Cutting in Simulated Abrasive Processes. Wear, vol. 7, 1964, p. 451.
49. Kragelskii, I. V. (L. Ronson and J. K. Lancaster, trans.): Friction and Wear. Butterworths Scientific Publ. (Wash., D.C.), ch. 5.
50. Oberle, T. L.: J. Metals, vol. 3, 1951, p. 438.
51. Blok, H.: Proc. Roy. Soc. (London), vol. A212, 1952, p. 480.
52. Halliday, J. S.: Proc. Inst. Mech. Engrs., 1957, p. 647.
53. Kragelskii, I. V. (L. Ronson and J. K. Lancaster, trans.): Friction and Wear. Butterworths Scientific Publ. (Wash., D.C.), p. 24.
54. Tabor, D.: The Hardness of Metals. Oxford Univ. Press, 1956.
55. Crook, A. W.: Simulated Gear-Tooth Contacts; Some Experiments Upon Their Lubrication and Subsurface Deformations. Proc. Inst. Mech. Engrs., vol. 171, 1957, p. 187.
56. Welsh, N. C.: Structural Changes in Rubbed Steel Surfaces. Proc. Inst. Mech. Engrs., 1957, p. 701.
57. Hamilton, G. M.: Plastic Flow in Rollers Loaded Above the Yield Point. Proc. Inst. Mech. Engrs., vol. 177, 1963, p. 667.
58. Poritsky, H.: Stresses and Deflections in Cylindrical Bodies With Application to Contact of Gears and Locomotive Wheels. J. Appl. Mech., ASME, vol. 72, 1950, p. 191.
59. Hamilton, G. M.; and Goodman, L. E.: J. Appl. Mech., ASME, vol. 89E, 1966, p. 371.
60. Waterhouse, R. B.: Fretting Corrosion. Proc. Inst. Mech. Engrs., vol. 169, 1955, p. 1157.
61. Flom, D. G.; and Porile, N. T.: Friction of Teflon Sliding on Teflon. J. Appl. Phys., vol. 26, 1955, p. 1088; J. Appl. Phys., vol. 28, 1957, p. 1361.
62. Schallamach, A.: Friction and Abrasion of Rubber. Wear, vol. 1, 1967, p. 384.
63. Ratner, S. B.; Gool, V. E.; and Klitenick, G. S.: On the Abrasion of Vulcanized Rubber Against Wire Gauze. Wear, vol. 2, 1958–59, p. 127.
64. Vinogradov, G. V.; and Bezborodko, M. D.: Friction, Wear and Lubrication of Plastics. Wear, vol. 5, 1962, p. 467.
65. Rudakov, A. P.; and Kuvshinskii, Ye. V.: Mechanism of Wear of Rubber. Polymer Science (U.S.S.R.), vol. 4, 1963, p. 1081.
66. Lontz, J. F.; and Kumnick, M. C.: Wear Studies on Moldings of PTFE Resin. ASLE Trans., vol. 6, 1963, p. 276.
67. Kragelskii, I. V.; and Neponmyashchi, E. F.: Fatigue Wear Under Elastic Conditions. Wear, vol. 8, 1965, p. 303.
68. Vinogradov, G. V.; Mustafaef, V. A.; and Podolsky, Yu. Ya.: A Study of Heavy Metal to Plastic Friction Duties and of the Wear of Hardened Steel in the Presence of Polymers. Wear, vol. 8, 1965, p. 385.
69. Savkoor, A. R.: Some Aspects of Friction and Wear of Tyres Arising From Deformations, Slip and Stresses at the Ground Contact. Wear, vol. 9, 1966, p. 66.
70. Ludema, K. C.; and Tabor, D.: The Friction and Visco-Elastic Properties of Polymeric Solids. Wear, vol. 9, 1966, p. 329.
71. Wisander, D. W.; Maley, C. E.; and Johnson, R. L.: Wear and Friction of Filled PTFE Compositions in Liquid Nitrogen. ASLE Trans., vol. 2, 1959, p. 58.
72. Hady, W. R.; Allen, G. P.; and Johnson, R. L.: Wear and Friction of Mechanical Carbons in Liquid Oxygen as Influenced by Transfer Films. ASLE Trans., vol. 6, 1963, p. 201.

73. Buckley, D. H.; and Johnson, R. L.: Mechanism of Lubrication for Solid Carbon Materials in Vacuum at 10^{-9} mm Mercury. ASLE Trans., vol. 7, 1964, p. 91.
74. Lancaster, J. K.: The Wear of Carbon and Graphitic Materials Sliding Against Rough Metal Surfaces. Lubrication and Wear Convention, Inst. Mech. Engrs., 1963, paper 16.
75. Lancaster, J. K.: The Relationship Between the Wear of Carbon Brush Materials and Their Elastic Moduli. Brit. J. Appl. Phys., vol. 14, 1963, p. 497.
76. Clark, W. T.; and Lancaster, J. K.: Breakdown and Fatigue of Carbons During Repeated Sliding. Wear, vol. 6, 1963, p. 467.
77. Midgley, J. W.; and Teer, D. G.: An Investigation of the Mechanism of the Friction and Wear of Carbon. ASME paper 62–Lub–15, 1962.
78. Longley, B. A.; Midgley, J. W.; Strang, A.; and Teer, D. G.: Mechanism of the Frictional Behaviour of High, Low and Non-Graphitic Carbon. Lubrication and Wear Convention, 1963, paper no. 17.
79. Glaeser, W. A.: Wear Characteristics in Non-Metallic Materials. Wear, vol. 6, 1963, p. 93.
80. Coffin, L. F.: A Fundamental Study of Synthetic Sapphire as a Bearing Material. ASLE Trans., vol. 1, 1958, p. 108.
81. Zeman, K. P.; and Coffin, L. F.: Friction and Wear of Refractory Compounds. ASLE Trans., vol. 3, 1960, p. 191.
82. Steijn, R. P.: On the Wear of Sapphire. Wear, vol. 5, 1962, p. 308.
83. Sibley, L. B.; and Allen, C. M.: Friction and Wear Behaviour of Refractory Materials at High Sliding Velocities and Temperatures. Wear, vol. 5, 1962, p. 312.
84. Amateau, M. F.; and Glaeser, W. A.: Survey of Materials for High-Temperature Bearing and Sliding Applications. Wear, vol. 7, 1964, p. 385.
85. Duwell, E. J.; and McDonald, W. J.: Some Factors That Affect the Resistance of Abrasive Grits to Wear. Wear, vol. 4, 1961, p. 372.
86. Duwell, E. J.: Friction and Wear of Single Crystal Sapphire Sliding on Steel. J. Appl. Phys., vol. 33, 1962, p. 2691; Wear, vol. 6, 1963, p. 312.
87. Duwell, E. J.; and Butzke, H. C.: The Effect of Interface Composition on the Wear Rate of Sapphire. ASLE Trans., vol. 7, 1964, p. 101.
88. Bowden, F. P.; and Hanwell, A. E.: Friction and Wear of Diamond in High Vacuum. Nature, vol. 201, 1964, p. 1279.
89. King, R. F.; and Tabor, D.: The Strength Properties and Frictional Behavior of Brittle Solids. Proc. Roy. Soc. (London), vol. A223, 1964, p. 225.
90. Atkins, A. G.; and Tabor, D.: Hardness and Deformation Properties of Solids at Very High Temperatures. Proc. Roy. Soc. (London), vol. A292, 1966, p. 441.
91. Beeck, O.; Givens, W. J.; and Smith, A. E.: The Mechanism of Boundary Lubrication. Proc. Roy. Soc. (London), vol. A177, 1940, p. 9.
92. Dowson, D.; and Higginson, G. R.: Elastohydrodynamic Lubrication: The Fundamentals of Roller and Gear Lubrication. Pergamon Press (Oxford), 1965.
93. Dowson, D.: Elastohydrodynamic Lubrication: An Introduction and a Review of Theoretical Studies. Proc. Inst. Mech. Engrs., vol. 180, Pt. 3B, 1965–66, p. 7.
94. Archard, J. F.: Experimental Studies of Elastohydrodynamic Lubrication. Proc. Inst. Mech. Engrs., vol. 180, Pt. 3B, 1965–66, p. 17.

95. Archard, J. F.; and Kirk, M. T.: Lubrication at Point Contacts. Proc. Roy. Soc. (London), vol. A261, 1961, p. 532.
96. Dyson, A.; Naylor, H.; and Wilson, A. R.: The Measurement of Oil Film Thickness in Elastohydrodynamic Contacts. Proc. Inst. Mech. Engrs., vol. 180, Pt. 3B, 1965–66, p. 119.
97. Tallian, T. E.; McCool, J. I.; and Sibley, L. B.: Partial Elastohydrodynamic Lubrication in Rolling Contact. Proc. Inst. Mech. Engrs., vol. 180, Pt. 3B, 1965–66, p. 169.
98. Christensen, H.: Nature of Metallic Contact in Mixed Lubrication. Proc. Inst. Mech. Engrs., vol. 180, Pt. 3B, 1965–66, p. 147.
99. Dawson, P. H.: Effect of Metallic Contact on the Pitting of Lubricated Rolling Surfaces. J. Mech. Eng. Sci., vol. 4, 1962.
100. Dawson, P. H.: Further Experiments on the Effect of Metallic Contact on the Lubrication of Rolling Surfaces. Proc. Inst. Mech. Engrs., vol. 180, Pt. 3B, 1965–66, p. 95.
101. Tallian, T. E.: Rolling Failure, Contact Failure Control Through Lubrication. Conference on Lubrication and Wear, Inst. Mech. Engrs., 1967, paper 14.
102. Kragelskii, I. V. (L. Ronson and J. K. Lancaster, trans.): Friction and Wear. Butterworths Scientific Publ. (Wash., D.C.), p. 126.
103. Cocks, M.: Shearing of Junctions Between Metal Surfaces. Wear, vol. 9, 1966, p. 320.
104. Cocks, M.: Interaction of Sliding Metal Surfaces. J. Appl. Phys., vol. 33, 1962, p. 2152.
105. Antler, M.: The Lubrication of Gold. Wear, vol. 6, 1963, p. 44.
106. Bowden, F. P.; and Tabor, D.: The Friction and Lubrication of Solids, Part II. Oxford Univ. Press, 1965, ch. 4.
107. Rowe, C. N.: Some Aspects of the Heat of Adsorption in the Function of a Boundary Lubricant. ASLE Trans., vol. 9, 1966, p. 100.
108. Waterhouse, R. B.: Fretting Corrosion. Proc. Inst. Mech. Engrs., vol. 169, 1955; Discussion by D. Tabor.
109. Rowe, C. N.: A Relation Between Adhesive Wear and Heat of Adsorption for the Vapor Lubrication of Graphite. ASLE Trans., vol. 10, 1967, p. 10.
110. Quinn, T. F. J.: The Role of Oxidation in the Mild Wear of Steel. Brit. J. Appl. Phys., vol. 13, 1962, p. 33.
111. Quinn, T. F. J.: The Effect of "Hot-Spot" Temperatures on the Unlubricated Wear of Steel. ASLE Trans., vol. 10, 1967, p. 158.
112. Archard, J. F.: The Temperature of Rubbing Surfaces. Wear, vol. 2, 1959, p. 438.
113. Furey, M. J.: Surface Temperatures in Sliding Contact. ASLE Trans., vol. 7, 1964, p. 133.
114. Dayson, C.: Surface Temperatures at Unlubricated Sliding Contacts. ASLE Trans., vol. 10, 1967, p. 169.
115. Newton, I.: Opticks. Roy. Soc. (London), 1704; reprinted by Dover Publ., 1952, p. 265.
116. Samuels, L. E.: The Nature of Polished Metal Surfaces. Australian J. Sci., vol. 21, 1959, p. 163.
117. Rabinowicz, E.; and Mutis, A.: Effect of Abrasive Particle Size on Wear. Wear, vol. 8, 1965, p. 381.
118. Bowden, F. P.; and Hughes, T. P.: Polishing, Surface Flow and the Formation of the Beilby Layer. Proc. Roy. Soc. (London), vol. A160, 1937, p. 575.
119. Buckley, D. H.; and Johnson, R. L.: Friction and Wear of Hexagonal

Metals and Alloys as Related to Crystal Structure and Lattice Parameters in Vacuum. ASLE Trans., vol. 9, 1966, p. 121.

120. MACHLIN, E. S.; AND YANKEE, W. R.: Friction of Clean Metals and Alloys With Special Reference to Titanium. J. Appl. Phys., vol. 25, 1954, p. 576.
121. BONDI, A.: Spreading of Liquid Metals on Solid Surfaces. Chem. Rev., vol. 52, 1953, p. 417.
122. DEGEE, A. W. J.: The Friction of Gold-Silver Alloys Against Steel. Wear, vol. 8, 1965, p. 121.
123. ARCHARD, J. F.: Contact and Rubbing of Flat Surfaces. J. Appl. Phys., vol. 24, 1953, p. 981.
124. ARCHARD, J. F.: Single Contacts and Multiple Encounters. J. Appl. Phys., vol. 32, 1961, p. 1420.
125. BURWELL, J. T.; AND STRANG, C. D.: On the Empirical Law of Adhesive Wear. J. Appl. Phys., vol. 23, 1952, p. 23.
126. ROWE, C. N.: Some Aspects of the Heat Adsorption in the Function of a Boundary Lubricant. ASLE Trans., vol. 9, 1966, p. 100.
127. BRUNNAUER, S.: The Adsorption of Gases and Vapors. Princeton Univ. Press 1942, p. 230.
128. RABINOWICZ, E.: Friction and Wear of Solids. John Wiley & Sons, 1965, p. 132.
129. FINKIN, E. F.: The Wear of Copper, Aluminum, Mild Steel and Zinc, and Their Wear Particle Shape Factors. ASLE Trans., vol. 7, 1964, p. 377.
130. RABINOWICZ, E.: Friction and Wear of Solids. John Wiley & Sons, 1965, p. 157.
131. RABINOWICZ, E.: Friction and Wear of Solids. John Wiley & Sons, 1965, ch. 6.
132. KHRUSCHOV, M. M.: Resistance of Metals to Wear by Abrasion, as Related to Hardness. Proc. Conf. on Lubrication and Wear, Inst. Mech. Engrs., 1957, p. 655.
133. ROWE, C. N.: A Relation Between Adhesive Wear and Heat of Adsorption for the Vapor Lubrication of Graphite. ASLE Trans., vol. 10, 1967, p. 10.
134. SAVAGE, R. H.; AND BROWN, C.: Chemical and Physical Adsorption of Gases on Carbon Dust. J. Am. Chem. Soc., vol. 70, 1948, p. 70.
135. BRYANT, P. J.; AND GUTSHALL, P. L.: Friction and Wear Behavior of Solid Films. AF Mat. Lab. Tech. Rept. 66–8, 1966.
136. QUINN, T. F. J.: The Effect of "Hot Spot" Temperatures on the Unlubricated Wear of Steel. ASLE Trans., vol. 10, 1967, p. 158.
137. QUINN, T. F. J.: The Effect of "Hot Spot" Temperatures on the Unlubricated Wear of Steel. ASLE Trans., vol. 10, 1967; Discussion by C. N. Rowe and P. E. Fowles.
138. RABINOWICZ, E.: Lubrication of Metal Surfaces by Oxide Films. ASLE Trans., vol. 10, 1967, p. 400.
139. KERRIDGE, M.; AND LANCASTER, J. K.: The Stages in a Process of Severe Metallic Wear. Proc. Roy. Soc. (London), vol. 236A, 1956, p. 250.
140. STEIJN, R. P.: An Investigation of Dry Adhesive Wear. J. Basic Eng., ASME, vol. 81D, p. 56.
141. STEIJN, R. P.: Sliding Wear and Metal Transfer Under Unlubricated Conditions. J. Basic Eng, ASME, vol. 81D, 1959, p. 67.
142. RINEHART, J. S.: Fracturing by Spalling. Wear, vol. 7, 1964, p. 315.

Boundary Lubrication

D. GODFREY
Chevron Research Company
Richmond, California

This review article discusses boundary lubrication in terms of the lubricating films on sliding surfaces. Films are formed by physical adsorption, chemisorption, and chemical reaction. The physical properties of the films, particularly melting point, shear strength, hardness, and adhesion to the surface, are important to lubrication. Film failure mechanisms include desorption, melting, solution, discontinuity, detachment, and rupture. This review also reveals that there is no adequate theory for boundary lubrication, and that there is a lack of knowledge of the chemical composition of films on sliding surfaces and on their physical properties under conditions of high pressure, high shear rate, and high temperature.

IN 1922, Sir William Hardy (ref. 1) wrote that boundary lubrication is that condition where the "friction depends not only on the lubricant, but also on the chemical nature of the solid boundaries." Recently, Dowson (ref. 2) defined boundary lubrication as follows: "Surface interaction between monomolecular layers of boundary lubricants and the solids dominate the operation of the contact. Hydrodynamic effects are negligible, and there is considerable asperity contact." The concept is represented in figure 1 which shows a microscopic cross section of films on two surfaces and areas of asperity contact. This writer considers that the films may be multimolecular as well as monomolecular.

In practical machines, boundary lubrication conditions can be recognized by the occurrence of wear and by the effectiveness of small quantities of surfactant additives in reducing wear. The cam-hydraulic lifter contact in an automobile is an example. In laboratory apparatus, boundary conditions are induced by high loads, small areas of contact, and low sliding speeds; they are recognized by wear and coefficients of friction greater than about 0.03. A small pin sliding on a slowly rotating disk is an example.

Three other states of lubrication have been recognized: (1) hydrodynamic, (2) mixed hydrodynamic and boundary, and (3) elastohydrodynamic. Hydrodynamic lubrication, where the surfaces are

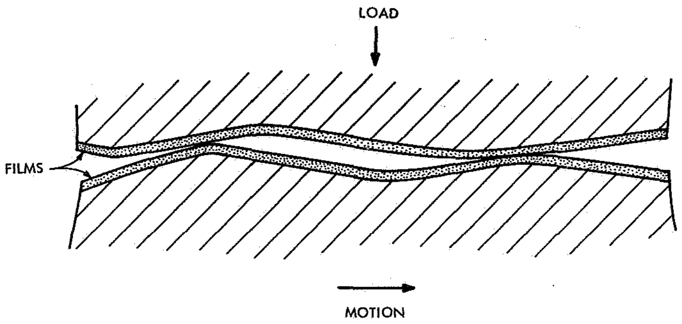


FIGURE 1.—Schematic representation of boundary lubrication.

completely separated by a fluid film, is the most desirable state of lubrication and is dependent on fluid viscosity, speed, and bearing geometry. The journal bearing in an automobile operates hydrodynamically at high speeds.

Elastohydrodynamic lubrication is a recently proposed state of lubrication which possibly exists in cases (such as gears) previously labeled as boundary. In elastohydrodynamic lubrication, local elastic distortion of the solids provides coherent hydrodynamic films, and asperity interaction is prevented (ref. 2). The thin oil films are subjected to high pressures, and, as a result, local viscosity increases markedly.

In lubrication research there is a tendency to analyze results in terms of one state of lubrication or one mechanism. Undoubtedly, other conditions are imposed, such as local high pressures, hot spots, and short exposure times. These conditions cause states and mechanisms to occur other than the one being investigated. Further complications arise, especially in practical machines, by the use of complex impure lubricants where various molecules compete for the surface. Also, variations in the atmosphere occur, such as water content. Larsen and Perry (ref. 3) listed 19 different reactions that could occur with the action of tricresyl phosphate in mineral oil on a steel surface. The happenings in the intimacy of a sliding surface are indeed complex.

In the art of lubricant development, terms such as oiliness, lubricity, extreme pressure, mild extreme pressure, antifriction and antiwear are used. They are related to boundary lubrication, but will not be used in this review because they are not specific.

The lubrication research literature reviewed here indicates that under boundary conditions, the physical properties of surface films govern performance. These films may be extremely thin, such as a monolayer of chemisorbed soap, or thick, such as a 1000Å film of iron sulfide. The important physical properties are melting point, shear strength, and hardness. Other properties are adhesion or tenacity, lack of cohesion, and rates of formation. This review points out the absence of adequate theories and the shortage of experimental facts in boundary lubrication. Melting points of films appear to be one unifying physical property. There is a need to conduct experiments to determine the chemical composition and physical properties of films under conditions of sliding. Where appropriate, contrary views, areas of confusion, and controversial mechanisms are mentioned.

Other recent and pertinent reviews of boundary lubrication have been written by Akhmatov (ref. 4), Rowe (ref. 5), Davies (ref. 6), Bowden and Tabor (ref. 7), Barwell (ref. 8), Zisman (ref. 9), and Bisson and Anderson (ref. 10).

MECHANISMS OF FORMATION OF SURFACE FILMS

The literature and patents on the chemistry of boundary lubricants is voluminous. A great variety of compounds has been added to petroleum and synthetic oils to improve their performance. In most cases, the mechanism of additive action was not known. Only in the last ten years has some understanding emerged. The chemistry of modern lubricating oil additives has been summarized by Stewart and Stuart (ref. 11). They discuss boundary lubricant additives such as esters, acids, and soaps which are adsorbed on the surface and reduce friction; zinc dialkyldithiophosphate, which reduces wear; and mercaptans, which react and inhibit scuffing. Each additive action is sensitive to concentration, temperature, and the presence of other additives and contaminants such as water and oxygen.

This writer recognizes the importance of this chemistry but feels that it is ultimately evaluated in the physical properties of the resultant film. Even corrosive wear, where chemical films are removed by mechanical action, is a result of the weakness of the film. Therefore, physical properties will be emphasized in this paper.

The classifications of film formation to be made are due to Uhlig (ref. 3) and Akhmatov (ref. 4). The latter discussed film formation comprehensively.

Physical Adsorption

The film is composed of molecules concentrated on the solid surface from the liquid or vapor state. Molecules are weakly bonded, and formation is characterized by reversibility and monomolecular or multimolecular layers. Polar molecules, particularly long-chain hydrocarbons, adsorb with preferred vertical orientation.

Perfect nonpolar molecules are rare, so adsorption in random positions may be difficult to achieve. Even if molecules are originally perfectly symmetrical, they can be polarized by the process of adsorption itself (ref. 4) or by heat, catalysis, or sliding (ref. 12). In flowing systems, such as in viscosity measurements or a journal bearing, the first layer is strongly adsorbed; otherwise, the liquid would slip over the solid. Levine and Zisman (ref. 13) imply that adsorbed films are solid only above a chain length of 10 carbon atoms. The physical adsorption of hexadecanol with vertical orientation is represented in figure 2.

Chemisorption

Chemisorbed films are composed of molecules that were first adsorbed physically on the surface and then reacted chemically to form a new entity. The chemisorbed film is limited to a monolayer, and formation

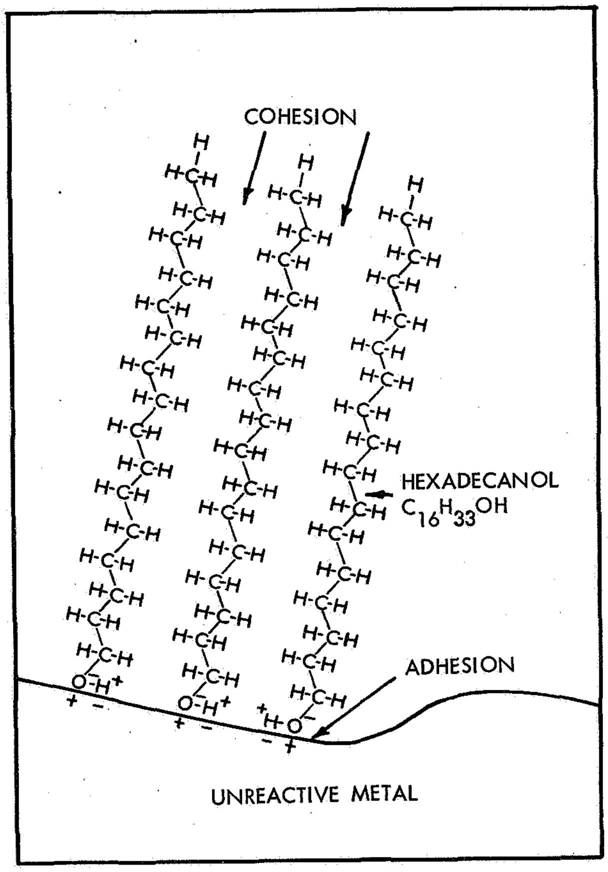


FIGURE 2.—Schematic diagram representing the physical adsorption with preferred orientation of three polar molecules of hexadecanol to a metal surface.

is characterized by irreversibility and strong bonding energies compared to physical adsorption. An example is a monolayer of zinc stearate from the chemisorption of stearic acid on zinc (ref. 14).

Holm (ref. 15) considered these films to be solid and to behave "as a rigid continuation of the solid body." Akhmatov (ref. 4) cites several papers which show by diffraction that the chemisorbed films of long-chain hydrocarbons are crystalline. Andrew (ref. 16) established that strongly oriented films correlated with efficient boundary lubrication. The chemisorption of stearic acid on an oxidized steel surface to form a monolayer of iron stearate is represented in figure 3.

Chemical Reaction

For lubrication, the term "chemical reaction" will be used to describe the formation of films of inorganic salts on metals. The term "corrosion" could also be used. Films are unlimited in thickness and characterized by high activation and bonding energies and irreversibility. An example is the reaction of sulfur with iron to form a film of iron sulfide, shown ideally in figure 4.

The formation of adsorbed and reaction films is most dependent upon temperature, which is an operating variable in a lubricated machine. The weak bond of a physically adsorbed film is further reduced by increased temperature, whereas the chemically bonded films increase in rate of formation with increased temperature. The work of Sakurai and Sato (ref. 17) is an example of studies of chemical reaction rates. They show that the rate of reaction of additives, such as dibenzyldisulfide and monochlorobenzene, with iron follows a parabolic law and is diffusion-controlled. Other chemical factors influencing film formation are pH of the solution, adsorption of solvent, precipitation, and chelation (ref. 18).

The composition of the film in the above mechanisms has been ideally represented as a single pure compound. This is not generally true on sliding surfaces. For example, this writer has found (ref. 19) that sulfurized oil forms a mixture of iron sulfide and iron oxide on steel. On gear teeth run in sulfurized oil, the oxygen was found in high concentrations several microns below the surface. Fein and Kreuz (ref. 20) and others have written on the importance of oxygen in oil-lubricated systems. Minor concentrations of impurities, such as polar compounds, have a marked influence on boundary lubrication (refs. 21 and 22). For example, Tingle (ref. 23) showed that a fatty acid will not react with a metal to form the soap film unless the metal is oxidized and water is present.

MEASUREMENT OF FILM PERFORMANCE

Performance of surface films is judged by measurement of friction,

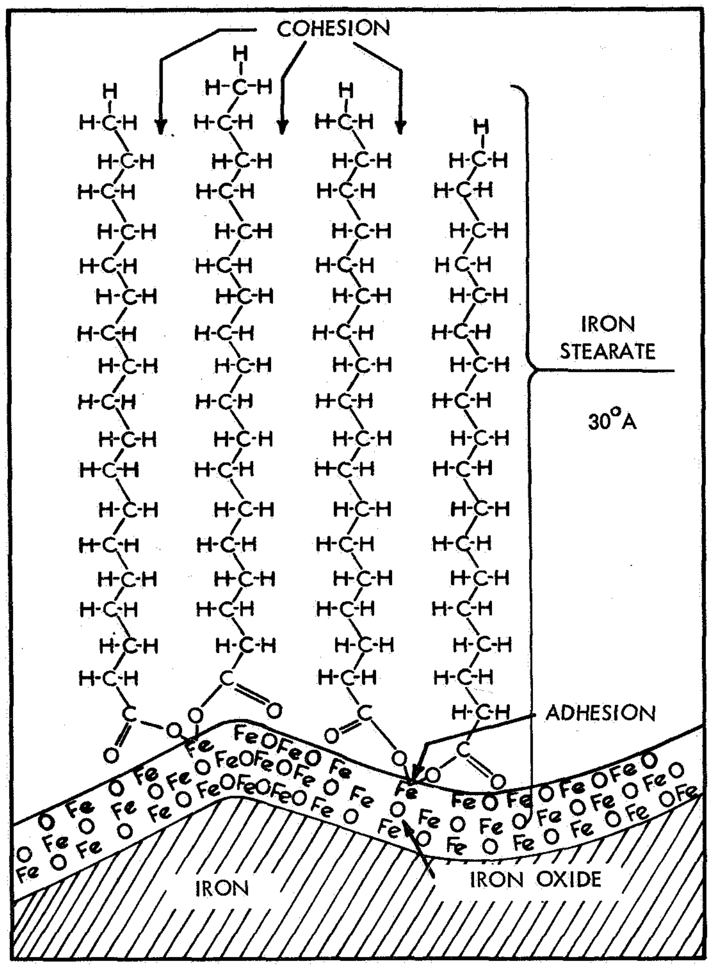


Figure 3.—Schematic diagram representing the chemisorption of stearic acid on an iron surface to form a monolayer of iron stearate, a soap.

wear, load carrying capacity, and failure temperature. In these days of good lubrication and ample power, friction is of lesser practical significance than wear, but it is significant to the understanding of lubrication. Under boundary conditions, the friction force has been considered to be the sum of the force required to shear the lubricant film and that required to shear solid-to-solid junctions. For example, with oleic acid on smooth steel the friction of the lubricant film has been estimated to be 0.194 kg and the friction of the solid junctions

0.002 kg (ref. 24). Wilson (ref. 25) also showed that during effective boundary lubrication there is very little solid-to-solid contact. Friction is a sensitive index of film action in one range of operation, such as the reduction of the coefficient of friction from infinity to unity by the admission of oxygen to a clean surface. But as Rabinowicz (ref. 26) has pointed out, it is an insensitive index in the more practical range of values. For example, the addition of a fatty acid to a base oil may lower the coefficient of friction from 0.15 to only 0.09.

Other definitions are: (1) "wear" is the removal of solid material due to mechanical action; (2) "load carrying" capacity is the maximum load, in a specified machine, which can be applied before a lubricant fails; and (3) "failure temperature" is the temperature at which a lubricant film fails. Wear is of great practical significance and is sensitive to effective boundary lubricants (ref. 26). Some authors recognize failure temperature as the most rewarding way to measure the effectiveness of boundary lubricants (refs. 27 to 29). Increasing loads and speeds resolve into temperature effects (ref. 29) provided the environment is constant.

PHYSICAL PROPERTIES OF FILMS

Melting Point or Thermal Stability

The melting point or thermal stability of surface film appears to be one unifying physical property governing failure temperature for a wide range of materials, including inorganic salts. It is based on the observation (refs. 7, 13, and 15) that only a film which is solid can properly interfere with potentially damaging asperity contacts. Conversely, a liquid film allows high friction and wear. Under practical

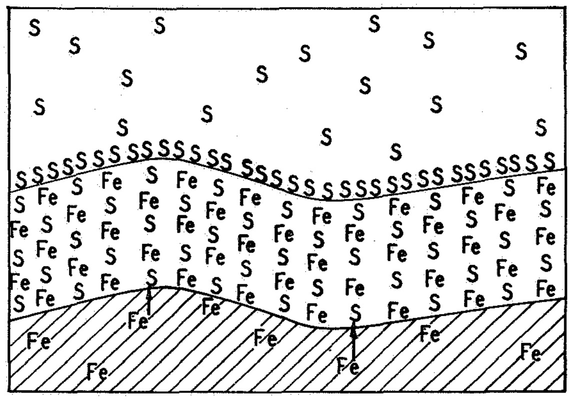


FIGURE 4.—Schematic representation of an inorganic film formed by reaction of sulfur with iron to form iron sulfide.

conditions, physically adsorbed additives are known to be effective only at low temperatures and chemisorbed additives at moderate temperatures (ref. 30). High melting inorganic materials are used for high-temperature lubricants.

A temperature effect to be considered before melting is desorption of physically adsorbed molecules. Brophy and Zisman (ref. 31) have shown that the temperature of desorption was 40° to 100° C for alcohols, 70° to 130° C for acids, 100° to 150° C for amines, and 140° to 170° C for amides.

The correlation of melting point with failure temperature has recently been established for a variety of organic films by Russell, Campbell, Burton, and Ku (ref. 32). Their data for hydrocarbons are reproduced in figure 5. The correspondence of melting points and sudden increases in friction is striking. The failure temperature of hexadecane has been measured by others (refs. 33 and 34).

Films of soap also fail to lubricate above their melting point (refs. 14, 26, 33, and 35). Chemisorbed fatty acids on reactive metals fail to lubricate above the melting point of the soap film that is formed. The correlation of increasing friction and melting point found by Barwell and Milne (ref. 35) for copper stearate is shown in figure 6. The failure temperature of soft metal films, such as lead and zinc, occurs at their melting point (refs. 7 and 15).

For materials melting above approximately 150° C, the correlation is not established. Rowe (ref. 36) showed a failure temperature of about 800° C for MoS_2 (which decomposes at 1200° C), and 300° C for tungsten hexachloride (m.p., 275° C) in vacuum. Braithwaite (ref. 37) stated, but did not support, that the performance of solid lubricants is limited by melting. In a review of high-temperature lubrication, Peterson (ref. 38) stated that thermal softening is one of the predominant factors leading to failure. He showed how simple films such as silver, gold, and polytetrafluorethylene cannot be used above their melting points.

The highest failure temperature in petroleum lubrication has been observed with sulfurized mineral oils. Provided a thick film consisting of iron oxides and iron sulfides is allowed to form, the SAE lubricant tester can be loaded until the steel rings are thermally distorted. The bulk melting point of the oxides and sulfide is above 1000° C.

Lack of correlation of failure temperature and melting point frequently can be explained by decomposition, oxidation, or vaporization. For example, Rowe's (ref. 36) titanium dioxide (m.p. 600° C) decomposed at 450° C into volatile tetraiodide. The effect of pressure on the melting point of a solid could confuse the correlation. Oxidation of oils and catalytic polymerization mask failure temperatures (ref. 39). In wire drawing, soaps are known to lubricate above their melting

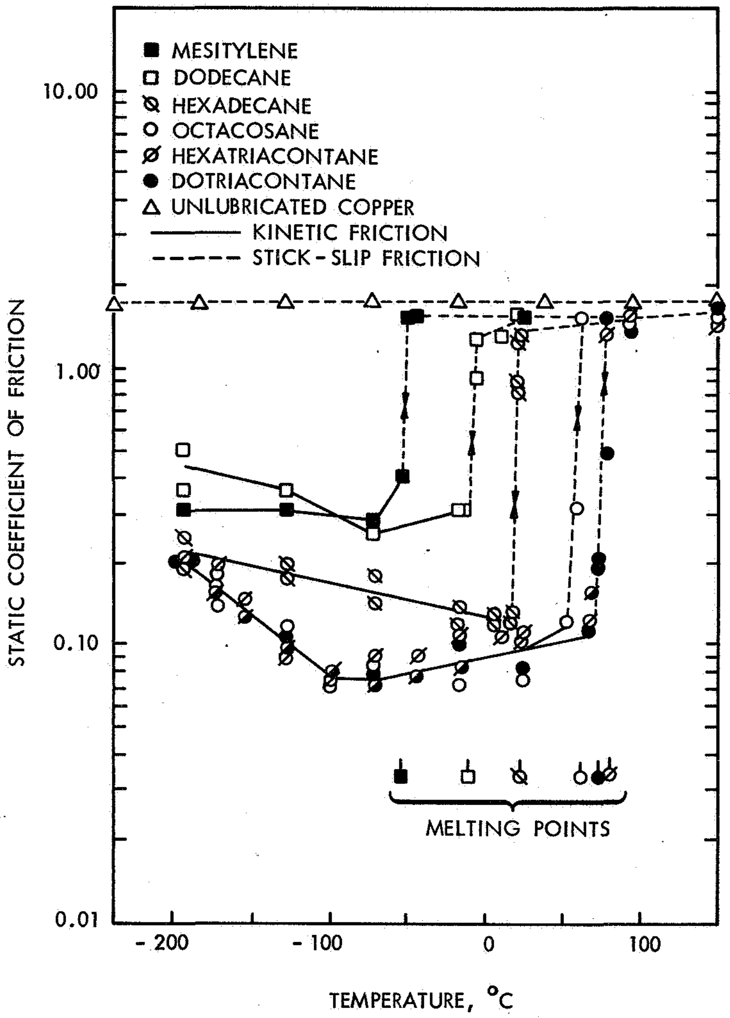


FIGURE 5.—Variation in friction with temperature for copper pairs lubricated with hydrocarbons in dry helium (ref. 32).

point. Wistreich (ref. 40) believes this is possible because of the short time (for example, 75 microseconds at 4000 feet/minute) that the soap film is in the hot zone. High-temperature performance of soaps is improved by the addition of fillers with higher melting points.

Proof of the importance of melting point in boundary lubrication might be found in the results of experiments where films of different composition or structure but the same melting point show the same failure temperature.

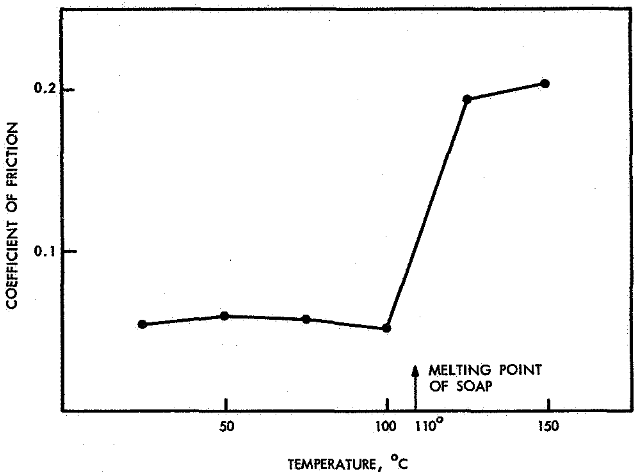


FIGURE 6.—Change in coefficient of friction with temperature for a film of copper stearate on a steel sphere sliding on flat steel. Load, 3.6 kg. Speed, 0.01 cm/sec. (ref. 35).

Shear Strength

Presuming the films are actually sheared during sliding, one would predict that low shear strength would cause low friction. The shear strength of a calcium stearate film has been found to be only 0.25 kg/mm^2 with a corresponding low friction coefficient of 0.05 (ref. 41). White (ref. 42) has measured the shear strength of soaps. Table 1 lists the shear strength of some film materials used in lubrication showing values ranging from 0.25 to 130 kg/mm^2 and different values obtained by different investigators. For example, the shear strength of lead has been measured at 0.75 and 1.6 kg/mm^2. The difference is probably due to the different anvil pressures used.

Merchant (ref. 43) has calculated the shear strength of inorganic salts from the derived equation

$$S=13930\ Q_f\rho \log_e \frac{T_m}{T},$$

where S is the strength of the crystalline solid in dynes/cm^2; Q_f is the latent heat of fusion, cal/g; ρ is the density; T_m is the melting point in degrees absolute; and T is the bulk temperature of the solid in degrees absolute. His results are included in table 1. They do not agree with the observed data of Bridgeman (ref. 44), and Boyd and Robertson (ref. 45). The expected friction reduction for given combinations in a cutting operation was observed. Merchant and his

TABLE 1.—*Observed and Calculated Shear Strength of Solids*

Material [a]	Shear strength, kg/mm^2	Anvil pressure, kg/mm^2	Reference
Stearic acid	0.25	2.5	4
Calcium stearate	0.25	?	41
Stearic acid	0.30	?	25
Silver sulfate	~0.7	~100	45
Lead	0.75	?	7
Stearic acid	~1.4	~100	45
Ferric stearate (100° C)	1.41	109	42
Lead	1.6	100	44
Ferric stearate (27° C)	3.16	109	42
Stearic acid	4.4	169	42
Silver sulfate	4.9	100	44
Lead sulfide, PbS	5.0	100	44
Molybdenum disulfide, MoS_2	5.6	~100	45
Lead oxide, PbO	6.5	100	44
Lead	6.9 (Calc.)	?	43
Gold	7.0	100	44
Zinc	8.0	100	43
Lead sulfate, $PbSO_4$	8.5	100	44
Aluminum chloride, Al_2Cl_6	9.0 (Calc.)	?	43
Silver	10.2	100	44
Cuprous chloride	10.2 (Calc.)	?	43
Zinc sulfide	13.0	100	44
Silver sulfide, Ag_2S	15	100	44
Iron chloride, Fe_2Cl_6	15.2 (Calc.)	?	43
Copper	16	?	7
Iron sulfate, FeS_2	17	100	44
Lead chloride, $PbCl_2$	18.6 (Calc.)	?	44
Zinc sulfate	19	100	44
Iron sulfide, FeS	20	100	44
Zinc oxide, ZnO	21.0	100	44
Iron oxide, Fe_2O_3	23	100	44
Iron sulfide, FeS_2	23	?	44
Graphite	29	100	44
Iron chloride, $FeCl_2$	37.2 (Calc.)	?	43
Aluminum	40 (Calc.)	?	43
Copper sulfide, Cu_2S	41.2 (Calc.)	?	43
Iron sulfide, FeS	60.7 (Calc.)	?	43
Steel	90	?	7
Copper	96 (Calc.)	?	43
Iron	130 (Calc.)	?	43

---

[a] Repeated values are by different investigators.

coworkers identified films of inorganic salts formed during a cutting operation and proved that a low ratio of shear strength of the film to the substrate metal gives a reduction in friction. Conversely, a high shear strength film on a low shear strength metal, such as lead chloride on lead, gives an increase in friction over lead. Another example of the wrong combination is silver sulfide on silver. Friction coefficients greater than 0.8 (ref. 46) have been measured for silver sulfide films on silver.

The high shear strengths of some high melting point inorganic salt films result in high friction values. Rowe (ref. 36) has measured a coefficient of friction of 1.8 for copper oxide film. Friction values for thick films of iron sulfide on steel have been found to be at least 0.6 (refs. 7 and 46) as shown in figure 7. It was speculated that the first rise in the curve of figure 7 was due to increasing metal-to-metal contact. Near 75° C, sulfur started to react with the metal to form a sulfide film of optimum thickness giving a friction of 0.3 at 100° C. Above that temperature, the film was too thick and its lubricating effect on steel was lost.

These data support Tabor's junction growth theory (ref. 7) of friction which states that friction is a function of the ratio of shear

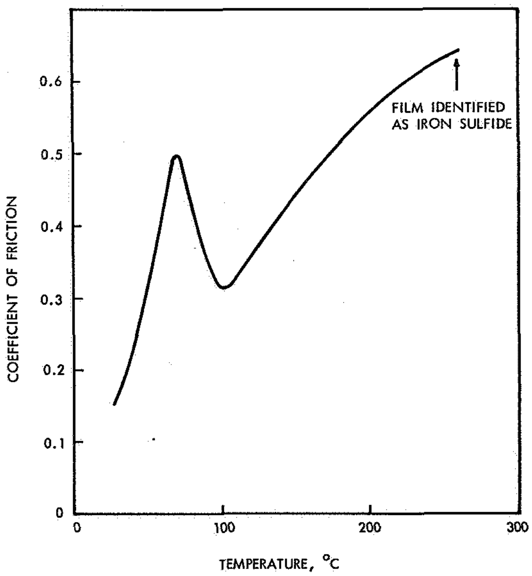


FIGURE 7.—Change in coefficient of friction with temperature for steel sliding on steel lubricated with deaerated sulfurized oil. Load, 0.5 kg. Speed, 2.68 cm/sec.

strength of surface films to the shear strength of the underlying solid. The relationship is shown in figure 8, where the ratio is plotted on the horizontal axis with a value of 1 at the left and zero at the right. The theory is in agreement with experience. For example, a MoS_2 film on steel gives low friction and Fe_2O_3 gives high friction. The theory also indicates how the same friction value can be obtained with various combinations provided the ratio is the same. For example, in order to obtain low friction with copper, a low shear strength film such as stearic acid is necessary. With this combination, the ratio of shear strengths is 0.09. The same friction can be obtained with materials of a higher level of shear strength, such as lead sulfate on steel, which also has a ratio of 0.09.

Knowledge of the chemical composition of a film on a sliding surface is needed in order to apply shear strength analysis. Further, there is a need to know the composition of the film as it is being subjected to high pressures, high temperatures, and shear. Part of these data might be obtained by X-ray diffraction of a thin film of material between small loaded rotating anvils.

The location of the shear plane is an unknown and presents a challenge to the experimentalist. In the thinnest film consisting of two hydrocarbon layers, slip probably occurs between the outermost terminal groups (usually methyl) of the molecules. Wilson's (ref. 25) electrical resistance data indicated that adsorbed fatty acid molecules were not sheared but bent over during sliding. If shear occurs between the lamina of a multilayer film, flow would occur and low friction values would be expected. If sliding causes rupture of molecular bonds, very high friction values would be observed. Since adhesive bonds are the same for a homologous series of fatty acids on a given metal, the coefficient of friction should be the same. The work of Hardy (ref. 1), Bowden and Tabor (ref. 7), and Levine and Zisman (ref. 13) shows, however, that the friction is not constant for a series of acids. If slip occurs at the interface of the film and the metal, adhesive bonds would be broken. Patrick (ref. 47) reports that rupture of the interface of an oriented monolayer on a metal would be very difficult to realize.

The location of shear planes in thick films of inorganic salts is also unknown. Under conditions of mild wear of steel, the wear fragments are Fe_3O_4, a few hundred angstroms in size. The small size and the absence of iron (by electron diffraction) indicated that only the oxide film has sheared.

This writer suggests that the location of the shear planes is essential to further understanding of boundary lubrication. The knowledge would lead to the development of better film materials, where the shear would be clearly fixed in the films rather than at the metal-film

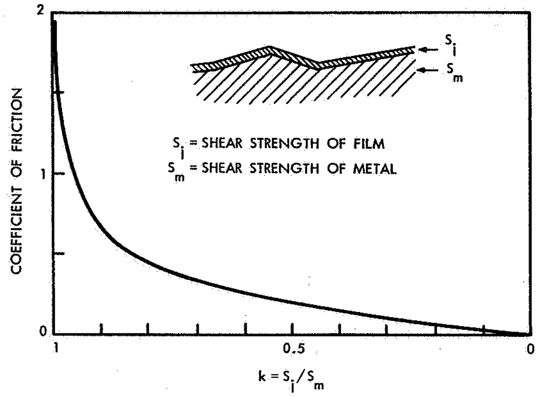


Figure 8.—Coefficient of friction as a function of the ratio of shear strengths of film and metal (ref. 7).

interface. Sensitive and sophisticated analytical tools are needed to locate shear zones.

Hardness Limitation to Melting Point and Shear Strength

A limit to increasing the melting point and shear strength in order to achieve high failure temperatures is imposed by such undesirable physical properties as increasing brittleness and abrasiveness.

A hard particle can plow a furrow in a softer solid, so one must assume that the attrited solid films are softer than the material being lubricated in order to avoid abrasive wear. The hard film may be the oxide of the metal, such as tin oxide, as described by Dies (ref. 48). Piggott and Wilman (ref. 49) conducted wear experiments with films of Fe_3O_4, α-Fe_2O_3, and FeS on mild steel disks under paraffin oil. The wear observed was abrasive caused by the attrited films which were identified in the debris.

Combined Properties

The foregoing indicates that a high melting point and thermal stability allow high failure temperatures, and low shear strength results in low friction. Unfortunately, high melting point usually is combined with high shear strength. Iron sulfide films have this characteristic and show high friction and high failure temperatures. Molybdenum disulfide is an exception possessing both high thermal stability and low shear strength. Since low shear strength and high melting point are desirable, compromises must be made or other combinations sought. Reasonably low melting points are needed to realize the low friction of low shear strength materials. Similiarly, high melting point is compromised with moderate hardness.

FAILURE MECHANISMS OF SOLID FILMS

A film fails to lubricate when metal-to-metal contact occurs. Failure mechanisms should be considered with the view that asperities on normally machined surfaces are not sharp peaks but elevations with gentle slopes (refs. 50 and 51) as shown in figure 1. These low slopes give surface films a better opportunity to prevent metal-to-metal contact than if the asperities were sharp peaks. The mechanism by which such an asperity actually penetrates a film is not clear.

Melting

An earlier section indicated that failure temperature correlates with melting point. During melting there is a reduction of molecular cohesion and orientation. Zisman (ref. 21) has commented that as the temperature of a monolayer is raised, the film melts and the molecules flow away from the asperity contact. The evidence was not presented. If the liquid is not strongly bonded to the surface, high surface tension may draw it up into droplets; then the metal would be unprotected. But, if melted chemical films are still bonded to the metal and resist flow, how then would metal-to-metal contact occur?

Incomplete and Defective Films

Discontinuous or incomplete films allow asperity contact at the site of the discontinuity. Peterson and Johnson (ref. 52) confirmed this by showing that effective solid lubricants such as MoS_2 are those which adhere to surfaces and form a continuous film on the sliding track. Films formed in situ may be incomplete as a result of chemical factors such as too low a concentration of the solute molecules, kinetics such as too low a temperature for chemisorption or chemical reaction, too short a time of exposure for reaction, or physical factors such as over extension because of plastic deformation. Discontinuity and defects in the crystal structure could also exist in a film by the presence of competing adsorbents. For example, the hexadecane solvent may be adsorbed in a local area to the exclusion of a fatty acid (ref. 13) if the acid concentration is too low. Further, the solid film can be removed from the surface by solution in the base oil (ref. 33). This mechanism is recognized (ref. 7) for soap films with increased solubility at higher temperatures.

Rate of formation determines film completeness. Cook (ref. 53) has shown that in one minute about 40 percent of a monolayer of 2-hydroxystearic acid was adsorbed from 50° C solution onto a gold and iron coated window of a Geiger counter. This rate seems slow when applied to a sliding surface, which is probably at a higher temperature.

The completeness of the film during boundary lubrication also appears to be time dependent because of the rheological properties of film material. Tabor's junction growth theory assumes sufficient time for film coverage and for plastic deformation, which may not be available in lubricated machinery. Milne (ref. 54) has considered the possibility of lubricant film migration or creep under load. He found that static friction was high after prolonged static loading, for example, a soft grease film gave a static friction of approximately 0.02 after one hour loading and 0.26 after 16 hours' loading. Apparently, the grease film was removed in the latter case.

Some data (refs. 18 and 55) indicated that less than a monolayer could provide protection during sliding. However, the films were Langmuir films lifted from water surfaces (which Zisman refers to as artifacts) and loads were low. At higher loads, such as 2 kg, less than a monolayer allowed failure (ref. 55).

A question on the need for a continuous film on a sliding surface arises when one realizes that only the tops of the asperities need to be covered to prevent metal-to-metal contact. Could effective lubrication occur if only the tips of the asperities were covered with a film and the depressions were not? Since the potential area of real contact is a small fraction of the area of apparent contact (ref. 7), the film coverage could be a very small percent of the total area and still be effective. This question could be answered by experiments using autoradiographs from radio-tagged film material on a sliding surface.

Lack of Cohesion and Adhesion

The question arises on how an asperity penetrates a double, close-packed, continuous chemisorbed film if it does not melt. The situation is represented schematically in figure 9 which shows two ideally smooth contacting asperities covered with oxide and hydrocarbon films. Levine and Zisman (ref. 13) have stated that the cohesional energy of

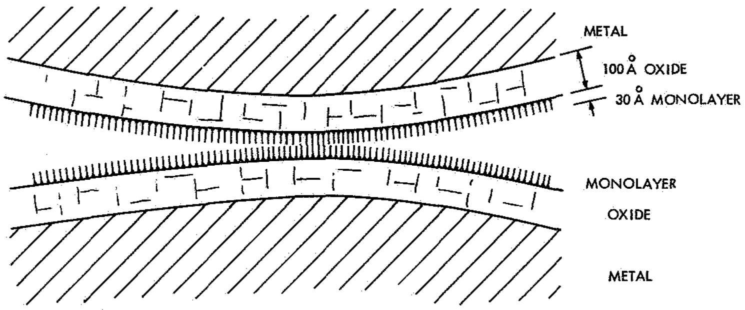


FIGURE 9.—Schematic representation of solid chemisorbed film on two metal asperities.

long-chain hydrocarbon films is considerable, and Akhmatov (ref. 4) reports high shear resistance for very thin films. Therefore, high friction alone, without evidence of surface damage, does not prove film failure. Also the presence of furrows, observed microscopically after sliding over a monolayer (ref. 7) could be caused by plastic deformation of the metal through the unruptured film. Lack of cohesion among molecules in a close-packed film seems to resolve into other mechanisms such as desorption, melting, or incomplete films.

Lack of adhesion of the film to the metal would result in film detachment and exposure of base metal. Research on the lubrication of gold (ref. 56) has shown that adhesion is necessary for effective lubrication. Monolayers of long-chain polar materials do not lubricate gold, whereas halogenated fluids do because of the strong adhesion of the halogen atom to the gold. Levine and Zisman (ref. 13) explain the lack of durability of films by discontinuity and lack of adhesion, but they do not provide direct evidence for the mechanism of failure.

The adhesion of boundary lubricants under extreme loads was demonstrated by Lancaster and Rowe (ref. 57) in drawing mild steel bars ½ inch by ½ inch in cross section. They showed that fatty acids and soap could sustain extreme die loads, such as 1 to 5 tons, provided there was an ample supply of lubricant to cover the newly formed surface.

The failure mechanisms of films of inorganic salts has not been satisfactorily resolved. For example, does an iron oxide or an iron phosphate film melt or break off of the steel surface? Gulbransen, et al (ref. 58) have speculated that cubic oxides are nonadherent and could allow wear, whereas oxides that form flexible whiskers and platelets could resist wear. His electron micrographs and sketches of these forms of iron oxide are shown in figure 10. Hirst and Lancaster (ref. 55) supported this by showing that the rate of metal oxide film formation affected friction during boundary lubrication with stearic acid. They concluded that the protection given by a lubricant film complements that afforded by the oxide. Osias and Tripp (ref. 59) showed that simple cracking of a brittle film does not allow metal-to-metal contact because the film is still present. They studied a model of two contacting asperities consisting of shellac on plasticine and observed that plasticine-to-plasticine contacts occurred only if the cracks matched. A smear of petrolatum prevented the cracks matching.

A close examination of films during various stages of sliding would contribute to identification of the physical properties. The mechanism of film failure could be resolved with sensitive tools such as optical and thermal devices, radioactive tracers, or electron diffraction and electron microscopy, which would show the state of the film during or after failure. For example, autoradiographic evidence, that a film was

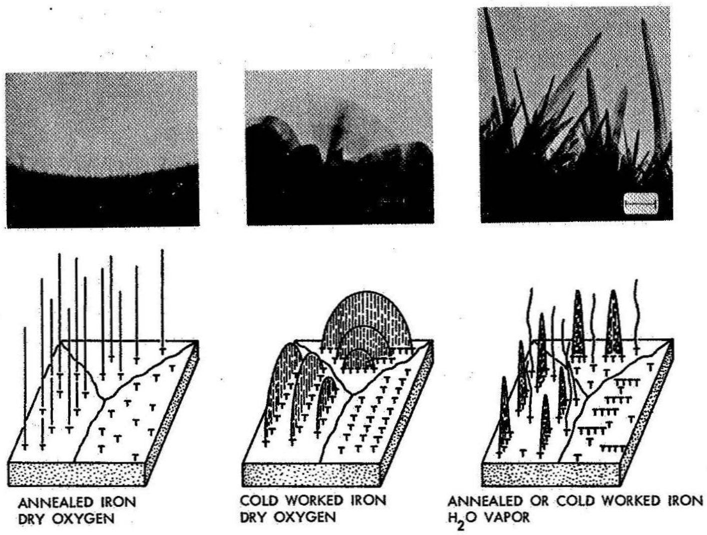


FIGURE 10.—Electron micrographs and schematic representations of oxide film growth at 400° C on pure iron under three conditions (ref. 64).

still present in the wear track after a failure with metal transfer, would suggest that it must have melted and flowed away when traversed by the rider and then flowed back and resolidified.

EFFECT OF CONSTITUTION VARIABLES IN BOUNDARY LUBRICATION

The premise of this paper is that the physical properties of surface films determine the nature of boundary lubrication. Any constitutional variable, such as the chain length of a fatty acid, becomes effective as it influences the film properties. Ample data, for example, those of Pomey (ref. 60), show different friction and wear values with different metals and lubricants used.

Less obvious variables include the concentration of air in the lubricant and in the atmosphere. Removal of air from white mineral oil has a marked effect on the kind of wear obtained in the sliding of steel on steel. Feng (ref. 61) reported severe wear of cast iron balls in a four-ball machine when air was removed from white oil. This writer found that with untreated white oil, mild wear was observed. The mild wear was evident by the formation of microscopic plateaus and the generation of wear fragments about 100Å in size and composed of Fe_3O_4. With deaeration of the oil and operation in an N_2 atmosphere severe wear was observed, evidenced by scuffing and the generation of very large metallic wear fragments. Clearly, one component of air was reducing wear during boundary lubrication.

Another example of the importance of minor components is the lubrication mechanism of tricresyl phosphate additive. Film analyses showed the film formed on steel to be iron phosphate (ref. 62). The phosphate was formed by phosphoric acid contamination in the commercial product or hydrolysis during use to form acidic or polar materials (refs. 22 and 63).

Sulfur and oxygen in mineral oil act upon steel cooperatively during high-temperature lubrication (ref. 46). Iron sulfide alone on a surface causes high friction and wear. A mixture of Fe_3O_4 and FeS reduces friction and wear.

CONCLUSION

This review of the literature on boundary lubrication revealed that no entirely adequate theory exists, and that there is a shortage of experimental facts. The formation of a solid surface film appears to be essential for effective lubrication. The shear strength of the film determines friction and the melting point correlates with failure temperature. Melting point appears to be the one unifying physical property which determines film performance. Other factors can and do mask or complicate the basic mechanisms.

Necessary areas of research include:

(1) Knowledge of the chemical composition of films on sliding surfaces

(2) Data on the shear strength of film materials under conditions of high pressure, high shear rate and temperature

(3) Data on the melting points of film material under conditions of high pressure and high shear rate

(4) Examination of films before and after sliding to indicate the location of shear planes and the importance of adhesion and cohesion

DISCUSSIONS

W. E. Campbell (Rensselaer Polytechnic Institute, Troy, New York)

In studying the literature on boundary lubrication, it is easy to become confused by the terminology. For example, many investigators, starting with Hardy (ref. 65) who coined the term boundary lubrication, refer to lubrication by a film resulting from interaction of a solid and a lubricant, usually an organic liquid. This has been extended by some to cover lubrication by films, such as oxides and sulfides, formed by chemical reaction with gases or even predeposited on the surface by other means such as electrodeposition or by rubbing against a lubricating solid (ref. 66). Many others use the term for lubrication in the four-ball, Timken or Falex machines operated at speeds where the load is obviously borne in significant part by a hydrodynamic film, operation is therefore mixed lubrication.

Mr. Godfrey very sensibly started his interesting review with two definitions. However, the first is very unspecific, being an incomplete quotation, while the second, which limits boundary lubrication to the action of monomolecular films, would seem to be too restrictive, in that the available evidence indicates that effective boundary films, especially those of inorganic composition, are hundreds or even thousands of monolayers in thickness (refs. 67 and 68). Even in the case of adsorbed, polar, long-chain organic films, several molecular layers are often necessary for effective friction reduction and wear prevention (refs. 69 and 70). A more complete quotation of Hardy is:

> In what is often called complete (hydrodynamic) lubrication, the solid surfaces are completely floated apart by the lubricant. There is, however, another kind of lubrication in which the solid faces are near enough to influence directly the physical properties of the lubricant and the friction depends not only on the [properties of] the lubricant but on the chemical nature of the solid boundaries. Boundary lubrication differs so greatly from complete lubrication as to suggest that there is a discontinuity between the two states. In boundary lubrication the surfaces have the property of static friction and the [frictional] resistance is some inverse function of the viscosity of the lubricant. In complete lubrication static friction is absent and the resistance varies directly with the viscosity of the lubricant.

Essentially the same definition is given more briefly by Bowden and Tabor (ref. 71). It is this discusser's belief that the essence of this definition is still basically sound and that confusion in communication of research results would be greatly reduced by retaining it. The following definition is therefore offered as a basis for discussion:

> Boundary lubrication is lubrication by a *liquid* under conditions where the solid surfaces are so close together that appreciable contact takes place between the asperities. The friction and wear are influenced predominantly by interaction between the lubricant and the solid, and the bulk flow properties of the lubricant play little or no part in friction and wear behavior.

This definition separates boundary lubrication from solid lubrication (defined as lubrication by a solid film) on the one hand, and from hydrodynamic (including elastrohydrodynamic) and mixed lubrication on the other. It would include, however, hydrodynamic lubrication resulting from lubricant-solid interaction to produce highly viscous products as proposed by Fein and Kreuz (ref. 72). Since, as Mr. Godfrey has pointed out, a solid film at the metal-liquid interface is the lubricating agent in many boundary lubrication situations, and the same solid applied as a film alone may be equally effective as a solid lubricant, the distinction implied in the definition may appear arbitrary. Aside from the convenience of the distinction for classification purposes, there is often a basic difference in the two situations because of changes in the film properties in the two different environments. The basis for this belief is outlined below.

In view of the uncertainities in definition and the inadequate knowledge of mechanism clearly brought out by Mr. Godfrey, one wonders why the term "boundary lubrication" has any more specificity than the terms "oiliness" or "lubricity." It is generally understood that they refer to lubricating ability under boundary conditions. Since they are less clumsy to use in general descriptions, it is hard to understand the rather common objection to their use.

It is also hard to see how an adequate discussion of boundary lubrication can be given without reference to the chemical mechanism of film formation and to the reaction kinetics involved. If, for example, the rate of film formation were lower than the rate of its removal by frictional action, the film could have ideal physical properties but would not provide effective boundary lubrication. Rather than any weakness of the film, competition between rate of film formation and rate of film solution in the lubricant can be the controlling factor in so-called corrosive wear situations. Thus copper oleate is an effective boundary additive for steel surfaces; but oleic acid in mineral oil is highly corrosive to copper or brass surfaces under boundary lubrication conditions because of the high solubility of the copper oleate formed by reaction with the surface oxide film. If the oleic acid were used in a medium which did not dissolve copper oleate, it is likely that good lubricity with low corrosiveness would be provided.

Although there is a great deal of evidence that solid films play a dominant role in many boundary situations, much of it has been obtained at high loads and very low speeds designed to eliminate viscosity and surface temperature effects. These conditions are not representative of many practical situations and there is mounting evidence that films of highly viscous polymers, developed by complex oxidation reactions, are involved in boundary lubrication at higher speeds (refs. 72 to 75). The effectiveness of these films is obviously critically dependent on chemical reactions involving metal oxide, lubricant, and oil oxidation intermediates such as organic peroxides.

There is also evidence that wear may be related to oxidation of the metal by reaction with organic peroxides. Vinogradov et al (ref. 76), in four-ball machine studies of petroleum fractions in a deoxygenated and an oxygenated condition, conclude that friction is governed by the interaction of metal and hydrocarbon oxidation products that accelerate chemical wear but reduce seizure and facilitate running-in. Similar explanations have recently been advanced to explain reduction in fretting corrosion of splines by addition of antioxidants to the lubricant (ref. 77). It is suggested that the metal is oxidized by organic peroxides and that the continuous formation and removal of the oxide film is the cause of the wear. The peroxides react preferentially with the antioxidant when it is present and thus prevent metal oxide formation.

Another interaction which may be of importance and which accords with Mr. Godfrey's emphasis on the necessity for studying film action under actual operating conditions, is that of the oxide or other film with the environment. For example, oxide films on copper 300 to 400Å thick have been shown to reduce the friction coefficient from 1.8 to 0.4 when measured in air at low loads (ref. 78). When measured in helium, other workers using comparable loads find films of Cu_2O up to 2000Å thick to give values varying from a low of 1.0 to above 2, greater in some cases than that of clean Cu (ref. 79). This suggests that absorbed oxygen and/or water films on the oxide are a factor in its effectiveness as a friction reducer.

There are several other ways in which oxide film interactions can influence boundary friction and wear. One of these, reactions with an additive which is unreactive to the substrate, was cited by Mr. Godfrey. Another interesting case is the lubrication of aluminum as reported by St. Pierre et al (ref. 80). The lubricant hexadecane, an unsaturated hydrocarbon, while reducing friction, is relatively ineffective at reducing wear. On the other hand allyl stearate, a long-chain ester with an unsaturated ester group, reduces both friction and wear. It is concluded that the former additive adsorbs only on the aluminum, leaving the hard oxide particles free to cause abrasive wear, while the latter adsorbs on the oxide surface as well as the metal.

Another factor influencing the behavior of metal oxides is their defect structure. Many are defect semiconductors which conduct by movement of anions or cations into lattice vacancies. Since oxidation rate is also frequently controlled by diffusion of cations through a cation deficient lattice, it is likely that chemisorbed films, by attracting cations, can influence their mobility and therefore the rate of film formation. Entry into the oxide lattice from the substrate metal of impurity ions or ions of minor alloying elements having different valence from the oxide cation can influence both the rate of film formation and the physical properties of the film.

The adhesion theory, as modified to take into account the effect of combined stresses to produce junction growth, gives a satisfactory semiquantitative explanation for a wide variety of phenomena, especially under conditions, such as those cited by Mr. Godfrey, where the proportion of metallic contact in the load-bearing area is relatively small. However, there is a sizable body of evidence indicating that bulk solid deformation, rather than adhesion, is the principal factor in frictional action more often than the adhesion theory predicts. Thus Feng (ref. 82) believes that the work necessary to deform interlocked steps, produced presumably by emergence of dislocations in the mating surfaces, is the primary source of frictional energy loss. His ideas are supported by Russian work (refs. 83 and 84). The prevailing friction theory of a large school of Russian workers, spearheaded by Kraghelskii (ref. 85), is that friction is a process involving both adhesion and metal deformation with the latter generally assuming the more important role. Whatever the merits of these theories, as applied to boundary lubrication in general, it would appear likely that under extreme conditions, such as sliding of a hard surface on a soft surface, or in metalworking operations, metal deformation accounts for most of the friction. Under these conditions, it is believed by many that the lubricant functions to weaken the metal at points of high stress. The exact mechanisms are still in dispute, but they do not involve load support by an interfacial film (refs. 86 and 87).

Thus Shaw has conducted experiments on the lubricating action of carbon tetrachloride which indicate that its action as an effective cutting lubricant is not dependent on its ability to lower tool-chip friction, which is poor, but on its ability to absorb in invisible microcracks in the freshcut chip surface and to prevent them from rewelding under the high compressive stresses near the tool-chip interface (ref. 88). This explanation is in accord with the work of Kohn (ref. 89) who believes that the lubricant's role in low-speed metal cutting is primarily to promote shear at the tool tip by the stabilization of microcracks at the tool tip enabling shear to occur at reduced levels of stress. On his theory lubricant action along the tool-chip interface is basically similar, in this case resulting from embrittlement of asperity welds. This action is believed to result from the blocking, by an adsorbed film, of the egress of dislocations, in much the same way as impurity atoms and foreign particles have been shown to block them in the bulk of the solid.

In stating that for lubrication use "chemical reaction" refers only to inorganic reactions, it is not clear whether the author is defining a usage for his review or is referring to what he considers established practice. The purpose of the distinction is not clear. The formation of metal soap, a common constituent of load-bearing films, is certainly a chemical reaction, as are the organic reactions from which many of the complex components of a boundary film result. In fact it is this discusser's belief that the action of a monolayer is the exception in boundary lubrication, whereas chemical reaction to form films of finite thickness is the rule.

Two other points relating to interfacial interaction appear worthy of comment. The first has to do with orientation of chemisorbed films. The author has depicted such a film in figure 3. This representation gives the impression, commonly held, that effective organic boundary films are oriented with the chain axis normal to the surface. This is by no means always the case, even for Langmuir-Blodgett type films. For example, stearic acid films on chromium are found to be inclined at 35° to the surface normal (ref. 90), and the dibasic sebacic acid is believed to be oriented parallel to the surface. Furthermore, it should be remembered that most of the diffraction work has been done on films floated onto a metal from a water surface, or deposited from a solution. Data on friction and contact resistance measurements (ref. 91)* indicate that the film at the sliding interface is always bent parallel to the surface; its thickness is reduced to that of the width of the molecule, ∼5Å, thus allowing conduction by electron tunnelling.

The other point has to do with the relation between polarity and lubricity. Although it is conceded that many effective boundary lubricants are polar, polarity in a molecule does not ensure the adhesion necessary for effective action. Thus the long-chain liquid alcohols are relatively ineffective. On the other hand, as has already been pointed out, a nonpolar molecule with an active group, such as carbon tetrachloride, can be very effective. The good lubricity of the relatively nonpolar perfluorocarbons is probably also related to the high electronegativity of the fluorine atom.

It is realized that most of the points discussed in these remarks are known to the author, who deliberately chose to stress film properties. They reinforce his main thesis that much more effort needs to be addressed to the extremely difficult task of studying the boundary film in situ.

R. S. Fein and K. L. Kreutz (Texaco Research Center, Beacon, New York)

Mr. Godfrey concludes that "no entirely adequate theory exists" for boundary lubrication and "there is a shortage of experimental facts." Even a quick literature perusal confirms a paucity of theory but reveals an abundance of facts. The critical need is for a theory which will explain the facts. Consistent with the symposium objectives, this discussion is addressed to this need rather than explicitly to Mr. Godfrey's review.

The present lack of adequate theory to explain boundary lubrication facts seems to result from: (1) inadequate boundary lubrication definitions, and (2) lack of appreciation of the dynamic characteristics of boundary lubrication.

Boundary lubrication has been defined in terms of experimental ob-

---

*Also unpublished data taken by the discusser.

servations (e.g., friction response to compositional and operational variables) or in terms of apparent solutions to the problem (e.g., monolayers, corrosion products, etc.). The experimental observation definitions fail to exclude observations resulting from elastohydrodynamic lubrication without asperity interactions (ref. 92). The apparent solution definitions fail because their limited scopes exclude many boundary lubrication phenomena and they presume knowledge of molecular scale conditions. It is suggested that the problem be approached by defining the boundary *situation*, specifically in physical terms which can be adequately tested.

Beyond this, it appears necessary that the boundary situation, and particularly boundary lubrication, must be treated as a dynamic system. The act of sliding brings about a number of interacting processes occurring both concurrently and consecutively at the bearing surface conjunction. The controlling features at the boundary cannot be understood by employing the essentially static model of the pioneer workers in the field, except as a point of departure when sliding begins.

Boundary lubrication effects occur whenever bearing surfaces of a sliding system approach each other so closely that their prominences or asperities interact. Thus, to encompass these effects, the boundary lubrication situation may be defined specifically as one in which bulk lubricant film thickness between bearing surfaces is less than the sum height of opposing asperities. Verification of this situation for a bearing is generally possible by means of experimentally confirmed theory (particularly hydrodynamic theory) and available surface roughness measuring equipment.

Using this definition of the boundary lubrication situation, the dynamic interactions of asperities must be considered. Specifically, these interactions should be amenable to the same type of analyses proved effective for full hydrodynamic lubrication but on a "micro" level. Thus, elastohydrodynamic principles (ref. 93) should apply to phenomena occurring at asperities, as long as the scale is not so small that the materials cannot be treated as continua. On this basis, this discussion first treats boundary lubrication, of a single asperity, by an idealized Newtonian liquid. This treatment leads to quantitative expressions relating micro-elastohydrodynamic (micro-ehd) film thickness, and stresses on the asperity to the physical properties of the liquid and surface; the asperity topography; surface separation as determined by bulk lubricant film thickness and/or surface compliance; and sliding velocity. Qualitative generalization of this idealized micro-ehd model provides what may be termed a micro-rheodynamic (micro-rhd) model of boundary lubrication. Examples are given of how this model permits rationalization of some of the boundary lubrication phenomena. It is suggested that this model may be useful for explaining the large number of confounding boundary lubrication results. Lastly, because of its apparent wide-ranging potential value, research needs indicated by the model are listed. At the outset, it should be emphasized that this model merely consists of an integration of concepts proposed by other research workers.

Boundary lubrication model.—The idealized model is illustrated in figure 11. This shows a single hemispherical asperity with undeformed crown radius r and height Z sliding along an infinitely stiff plane at velocity U_s. The basal planes are separated by a distance H. The space between the basal planes is occupied by a Newtonian liquid with viscosity η_0. The viscosity of this liquid at pressure P is $\eta_0 \exp (\alpha P)$ while temperature is assumed constant. The asperity is deformed by elastohydrodynamic forces and separated from the opposing plane surface by a fluid film of thickness h. Macro-elastohydrodynamic studies (ref. 93) show that over a wide range of conditions the dimensions of flattened area on the bearing surfaces will be approximately the same as would be obtained with static elastic contact (i.e., "Hertzian contact") under the same load. The studies also show that, over this range of conditions, the film thickness between the surfaces in the flattened area may be considered approximately constant and independent of load. Application of these approximations to the model asperity of figure 11 permits calculation of the pertinent parameters; thus the mirco-ehd film thickness h may be calculated from Archard and Cowking's macroscopic "point" conjunction formula (ref. 94)

$$h=1.07\ (\alpha\eta_0 U_s)^{2/3}(r)^{1/3}. \tag{1}$$

If the basal plane separation H is fixed, then, for a given undeformed asperity height Z, changes in the micro-ehd film thickness must

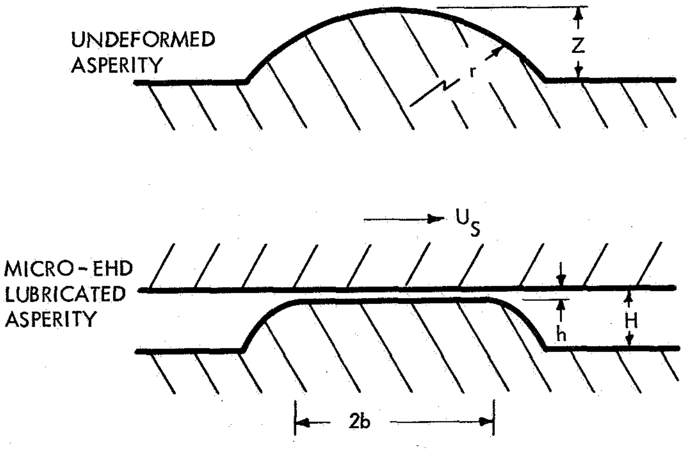


FIGURE 11.—Idealized single asperity model.

change the asperity deformation. The Hertz elastic theory (ref. 93) yields the radius of the flattened zone, b, the asperity load, W, and the average pressure on the flattened zone, $\bar{P}$, as

$$b=(rZ)^{1/2}\left(1+\frac{h}{Z}-\frac{H}{Z}\right)^{1/2}, \tag{2}$$

$$W=\frac{2}{3}E'r^{1/2}Z^{3/2}\left(1+\frac{h}{Z}-\frac{H}{Z}\right)^{3/2} \tag{3}$$

and

$$\bar{P}=\frac{2}{3\pi}E'\left(\frac{Z}{r}\right)^{1/2}\left(1+\frac{h}{Z}-\frac{H}{Z}\right)^{1/2}, \tag{4}$$

in which $E'=\text{effective elastic modulus}=\frac{\text{Young's Modulus*}}{1-(\text{Poisson Ratio})^2}$. Viscous shear of lubricant produces a much greater shear stress within the flattened area than outside because of the high shear rates and the increase of viscosity with pressure. The average shear stress at the surface of the flattened area is

$$\bar{\tau}=\frac{\eta_e U_s}{h}, \tag{5}$$

in which η_e=effective viscosity.

Some feeling for the significance of these equations can be obtained by making some reasonable substitutions. Thus, substituting α as $10^{-9}\text{cm}^2/d$ (typical of a mineral oil) and $r=10^{-2}$ cm (a finely ground steel bearing surface after brief run-in) yields from equation 1

$$h=0.23\times10^{-6}(\eta_0 U_s)^{2/3}\ cm, \tag{1a}$$

in which viscosity is in units of poises and sliding velocity in centimeters per second. This equation is shown graphically in figure 12. The equation indicates that at very low sliding speeds a high viscosity lubricant is necessary to achieve a reasonably thick micro-ehd film. For example, at $U_S=0.01$ cm/sec, η_0 must be 100 poise to provide a film thickness equivalent to about a monolayer of stearic acid. Viscosities greater than this occur in macro-elastohydrodynamic conjunctions lubricated with oils having typical lubricating oil viscosities because of the effect of pressure on viscosity (ref. 93). Thus, even at low sliding velocities, asperities in macro-ehd conjunctions should be separated by micro-ehd films. Obviously, at more practical sliding speeds (e.g., several cm/sec), viscosities in the range of typical lubricating oils would provide reasonably thick micro-ehd

---

*When one surface is not infinitely stiff, $E'=E'_1E'_2/(E'_1+E'_2)$.

films (even in conjunctions without appreciable macro-ehd load support).

Equations 2, 3, and 4 all indicate that b, W, and $\overline{P}$ increase with undeformed asperity height and

$$1+\frac{h}{Z}-\frac{H}{Z},$$

while b and W increase with asperity crown radius and $\overline{P}$ decreases with it. The effective elastic modulus has no effect on b but directly affects both W and $\overline{P}$. Among these parameters affecting b, W, and $\overline{P}$,

$$1+\frac{h}{Z}-\frac{H}{Z}$$*

is the most critical since it becomes nil when H/Z equals $(1+h/Z)$.

Substitution of values of $E'=2.3\times 10^{6}$ kg/cm^{2} (for steel), $Z=100\times 10^{-6}$ cm (for a high asperity on a ground steel bearing surface), and again $r=10^{-2}$ cm gives

$$b=10^{-3}\left(1+\frac{h}{Z}-\frac{H}{Z}\right)^{1/2} cm \tag{2a}$$

$$W=150\left(1+\frac{h}{Z}-\frac{H}{Z}\right)^{3/2} g \tag{3a}$$

and,

$$\overline{P}=490\left(1+\frac{h}{Z}-\frac{H}{Z}\right)^{1/2} kg/mm^{2}. \tag{4a}$$

Figures 13 and 14 graphically illustrate these equations for micro-ehd film thickness, h, ranging from zero (i.e., dry contact to 10×10^{-6} cm, and basal plane separations, H, from 60 to 110×10^{-6} cm. From figures 13 and 14 it is evident that asperity interaction effects become appreciable when the ratio of basal plane separation to asperity height, H/Z, is in the vicinity of unity. Further, as this ratio is reduced below unity the effect of micro-ehd thickness h becomes smaller on a percentage basis. In other words, b, W, and $\overline{P}$ with a micro-ehd film tend to approximate the values for dry contact ($h=0$) when H/Z becomes less than unity. Thus, the micro-ehd model yields approximately the same values of b, W, and $\overline{P}$ as those in the static boundary lubrication model with elastic contact. For an assemblage of asperities,

---

*It can be shown that equation 1 is valid only when H is at least approximately $5h$. Equations 2 through 5, however, should be valid to somewhat lower H/h ratios.

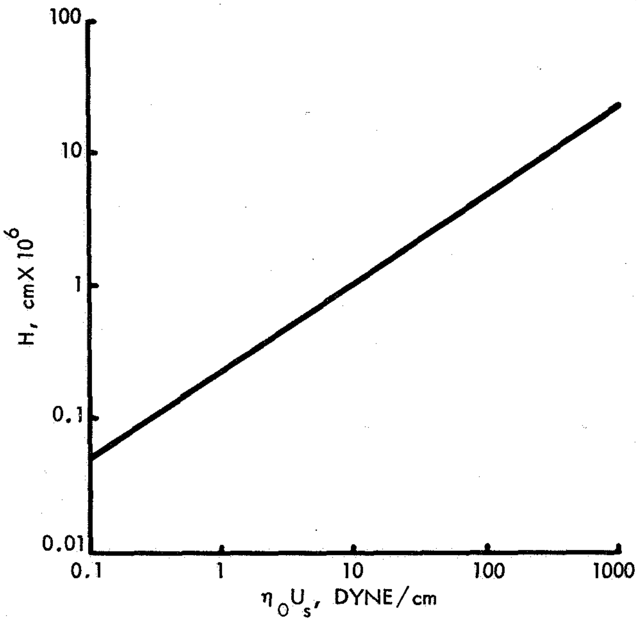


FIGURE 12.—Micro-elastohydrodynamic film thickness.

in this static model, the total area supported elastically by asperity contact is essentially proportional to load (ref. 96) as in the classic static plastic asperity model (ref. 97). Further, from figure 13, it is apparent that micro-ehd can produce pressures sufficient for full plastic deformation of steel. In this case the value of h should depend on the plastic asperity deformation in the micro-conjunction entrance region; in the absence of junction growth, b, W, and $\bar{P}$ should be essentially the classical values (ref. 97).

Consideration of the average asperity shear stress calculated from equation 5, however, shows a considerable deviation from the classical model. Substitution of h from equation 1 and incorporation of the Bell, Kannel, and Allen (ref. 98) effective viscosity relationship yields the average shear stress as

$$\bar{\tau}=\frac{(\eta_0 U_s)^{1/3}}{\left(\frac{3}{2}\alpha\bar{P}\right)^2}\left[\left(\exp\frac{3}{2}\alpha\bar{P}\right)\left(\frac{3}{2}\alpha\bar{P}-1\right)+1\right]. \tag{5a}$$

Hence, the model indicates that the shear stress should increase slightly with $\eta_0 U_s$ and exponentially with $\alpha\bar{P}$. Further, shear stresses calculated with reasonable values of $\eta_0 U_s$ and $\alpha\bar{P}$ are ridiculously high, being greater than the shear modulus of elasticity for organic materials (ref. 99). The explanation, probably, lies in the high rates of change of pressure and rates of shear in micro-ehd lubrication. Comparisons of these rates with estimated relaxation times under high pressure indicate that both compressional (ref. 100) and shear viscoelasticity should produce effective viscosities much lower than those given by the Bell, Kannel, and Allen (ref. 98) relationship used in equation 5a. At present, available rheological knowledge seems to be too limited to permit substitution of a more realistic effective viscosity in equation 5.

Consider now another ideal case illustrated in figure 15. Here the ideal lubricant is present as a film of thickness $d<H$ on the plane surface. Two steady-state possibilities would seem to exist; d is either sufficient or insufficient to permit development of the full micro-ehd film thickness. The limiting value of the film thickness on the plane surface can be obtained by a material transport balance, assuming lubricant incompressibility and negligible conjunction side leakage*. That is, in the limiting case, the lubricant transport to the conjunction, $U_s d_1$, must equal the transport of lubricant through the conjunction, $U_s h/2$. Thus, the minimum lubricant film thickness at which full micro-ehd lubrication can be achieved is $d_1 = h/2$.

Although d_1 is less than h on the surface far ahead of the conjunction under steady-state conditions, the lubricant film immediately in front of the flat on the asperity tip must be greater than h. This is obvious from the hydrodynamic considerations which require some filling of the gap between the surfaces to provide hydrodynamic pressure buildup in the entrance region of the conjunction. Elastohydrodynamic analyses suggest that the vast majority of the pressure buildup occurs within about

$$\frac{3(rh)^{2/3}}{b^{1/3}}$$

ahead of the Hertzian flat. Thus, to achieve steady state, the gap between the surfaces must be filled to about this distance ahead of the Hertzian flat. This gap is approximately

$$h_0 = 3^{3/2} h. \tag{6}$$

Thus, with the limiting lubricant film thickness far ahead of the conjunction, $d_1 = h/2$, the film thickness in the micro-conjunction

---

*The incompressibility is usual in elastohydrodynamics since it is reasonably valid (ref. 93). The work of Archard and Cowking (ref. 94) suggests that negligible side leakage is not too bad an assumption for a circular elastic load support zone.

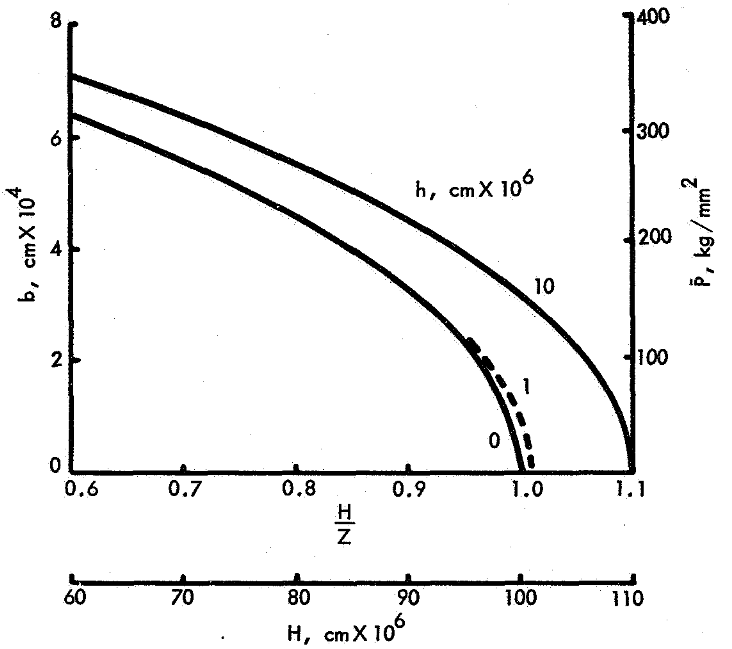


FIGURE 13.—Asperity load support zone radius and average pressure.

entrance region must be about $h_0=10d_1$. In other words, there is a pileup of the lubricant film ahead of the asperity; this pileup must be accumulated during an initial transient condition in which the micro-elastohydrodynamic film thickness is less than the steady-state value.

In the case that the lubricant film on the plane surface is less than d_1, then under steady-state conditions, the micro-ehd film is determined by the material transport. This "starved" micro-ehd film thickness is

$$h_s=2d, \text{ for } d<d_1. \tag{7}$$

Material transport consideration indicates that the steady-state micro-ehd film thickness should be obtained after very short sliding distances; this initial sliding is required to accumulate sufficient ideal liquid at the entrance to the micro-conjunction. In the case of non-ideal liquids, such as ideal plastic solids, similar accumulation should occur in front of the micro-conjunction.

General model for effective boundary lubrication.—The preceding section discussed the component parts of micro-ehd lubrication in terms of a single asperity on one surface. In general both surfaces will have many asperities. The same micro-ehd principles should

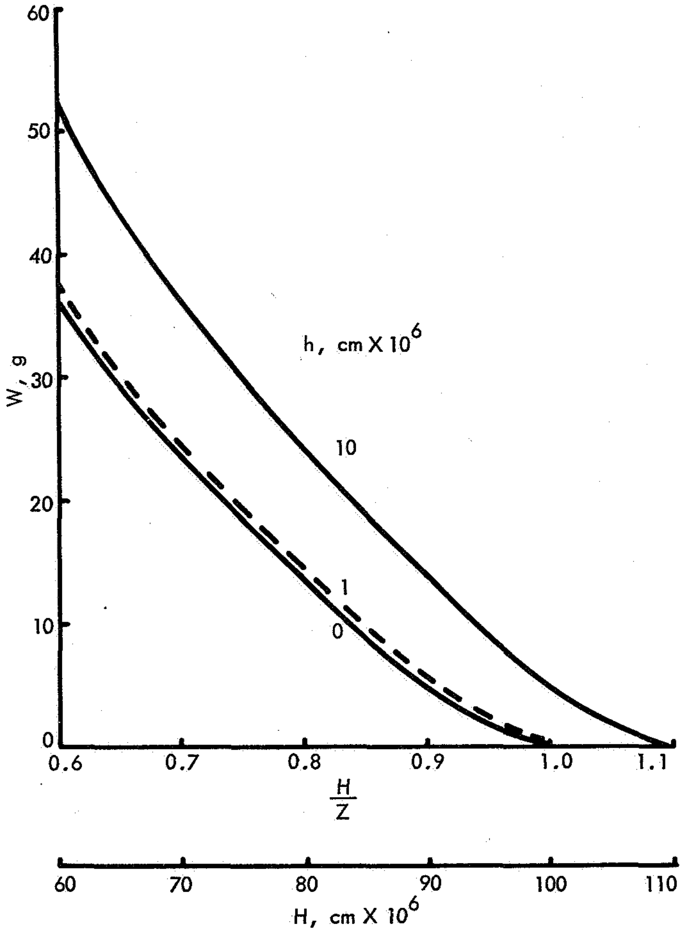


FIGURE 14.—Asperity load.

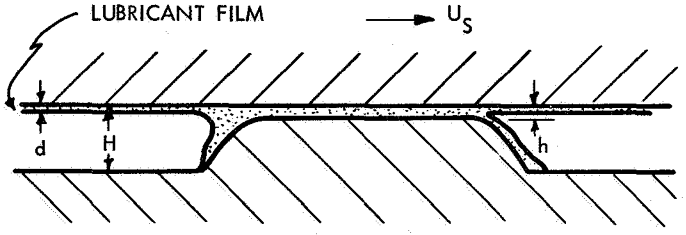


FIGURE 15.—Thin-film micro-elastohydrodynamic model.

apply at any instant. However, the asperity crown radius becomes the relative crown radius, and the asperity height the total for two interacting asperities. These along with the effective basal plane separations will change as sliding progresses. Hence, the micro-ehd film thickness, radius of the micro load-support zone, asperity load, and stresses will change continuously. Further, as previously pointed out, asperities can deform plastically and thus affect the micro-film thickness.

Also, in general, the asperity substrates will be covered with a solid film, a more or less fluid boundary film, and a bulk lubricant fluid film, none of which have ideal rheological properties. This general picture would be expected to appear somewhat like figure 16. Obviously, some of the films separating the bearing surface asperities could be absent or the films could be combined into a single film with a continuous gradient in constitution and rheological properties. Since the rheological properties and thickness of the films, along with the surface topography and separation, control the micro-ehd lubrication, any variable affecting these factors should influence the effectiveness of the lubrication. Melting or other abrupt physical change (e.g., a glass transition) would abruptly change rheological properties of the films while dissolution would change thickness; thus the model would account for the classical effects covered by Mr. Godfrey. Surface topography would change with bearing surface finishing operations, plastic flow of the surfaces, and wear processes. Separation (i.e., effective basal plane) would change with macro-ehd film thickness (refs. 93 and 94) and with compliance of the bearing on a scale much larger than a single asperity (refs. 96 and 101). Since dynamic deformation of asperities and the films separating them is the essence of this general model, it may appropriately be called a micro-rhd

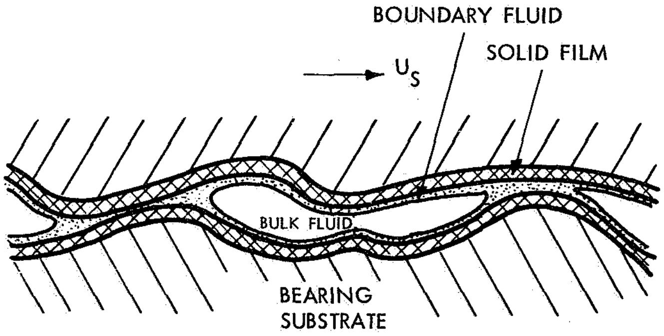


FIGURE 16.—Generalized micro-rheodynamic model.

model. Although quantitative treatment of this general micro-rhd model is prohibited presently by its complexity, it may be qualitatively treated with the aid of the idealized single-asperity ideal-fluid micro-ehd model.

An example of the usefulness of this micro-rhd model is lubrication by solid monolayers and multilayers of fatty type materials. Here the film thickness will be less than the basal plane separation and may be "starved" in the terms of the ideal model. Also, the film material probably more closely approximates an ideal plastic than an ideal liquid. Sliding a short distance should plow up sufficient film in front of an asperity to provide lift. This increases the film thickness to a level greater than the applied layer,* and disorients the film in the load support zone. Assuming the load-supporting micro-rhd film is an ideal plastic, the average shear stress reduces to the classical shear strength S. However, actual films would be expected to show a viscous component of shear stress. Equation 5 indicates that this should lead to an increase in shear stress with sliding velocity at each microconjunction, and consequently an increase in friction coefficient with sliding velocity; this agrees with experiment (ref. 96).** Since the micro-rhd film thickness depends on the thickness of the film applied to the surface under starved conditions (equation 7), equation 5, for the small viscous component of shear stress, also indicates that friction coefficient should decrease as the applied film becomes thicker due to longer molecules and/or multilayers. This also agrees with experiment.

Melting or disorientation of the fatty films decreases their rheodynamic lifting ability and abruptly changes their properties towards those of an ideal liquid. Thus the sudden degradation of lubricating ability with these phase changes is consistent with the micro-rhd picture and the classical picture covered by Mr. Godfrey.

Similar considerations might be expected to apply to thin films of any material which flows coherently under stresses that are low compared to those of the substrates. Thus for thin soft metal films, the shear stress (and friction) should decrease with increased applied film thickness until this applied film bridges the basal plane gap. When the gap is bridged, film also will be sheared outside this microconjunction and friction should increase. Thus, the existence of an optimum

---

*This increase is assumed in the classical interpretation of monolayer lubrication (ref. 102); thus, the interpretation assumes that asperity tips on both surfaces are covered by a monolayer although initially a monolayer was applied to only one surface.

**With mineral oil lubricants, experiment shows that friction coefficient decreases with sliding velocity (ref. 96). Presumably in this case, viscoelastic effects decrease the effective viscosity of the micro-rhd film with increased sliding velocity and the liquid becomes progressively less ideal.

applied film thickness for minimum friction (ref. 97) appears consistent with the model.

Wear and lubrication failure.—The natural question at this point is how wear and lubrication failure occur when asperities are always separated by a micro-rhd film. The answer appears to lie in the facts that: (1) some wear processes can occur in the presence of the film, and (2) plastic flow of the surfaces probably can produce film disruption leading to adhesion and sometimes to failure.

Consider first the four basic wear processes (ref. 92): corrosion, fatigue, plowing, and adhesion.

Corrosion occurs when a reaction product of the bearing material is removed from the surface. The reaction product may be a soluble or dispersible salt such as metal soap or other metal organic compound, an inorganic material such as an oxide, or a mixture of materials. If the reaction-product film on the asperities is soluble, dispersible, or readily flows coherently, then micro-rhd lubrication would be expected to take the film from the surface, pile it in front of the asperities, and make it subject to removal by shear stresses. If the reaction product is a brittle material, shear stresses in a supernatant micro-rhd film could be expected to fracture the film and remove it as wear particles. Neither of these processes require contact of asperities.

Repeated stressing of asperities through micro-rhd films would be expected to produce "fatigue" with consequent spalling from the surfaces. No asperity contact appears necessary for this microspalling.

Plowing occurs when an exceptionally large and hard asperity or wear particle plastically "plows" a groove in the bearing surface. Providing the asperity or wear particle has a sufficiently large curvature, the "particle" probably could be separated from the surface by a micro-rhd film. If the curvature is too small, films would be too thin to flow coherently and this would be considered as contact.

Adhesion of materials requires approach within atomic dimensions over an area large compared to atomic dimensions (ref. 103). Thus, contact between clean surfaces is necessary for adhesion. The problem in this case is a paraphrase of the question raised by Tabor (ref. 104); how is the contaminating film removed from a sufficient area over two opposing surfaces for them to adhere? A reasonable answer might be provided by the "creation" of new surface areas when only a thin contaminating film is between the rubbing surfaces. Such creation could result either from plastic "junction growth" (ref. 105) when a thin coherent micro-rhd film separates the surfaces, plastic deformation of a wear particle as it plows a surface, or even from brittle "breaking" of a wear particle as it plows a surface. In all cases, adhesion should occur only when the freshly formed surface area is large compared to the thickness of the contaminating film. For an ideal liquid or plastic film, this appears to require extension of the area by at least a factor of 10^8h, when h is in centimeters. This factor reduces the micro-rhd film thickness to less than one atomic radius.

The failure of lubrication is readily indicated by bearing surface distortion in the form of macrospalling, plastic flow or metal transfer, and tearing (i.e., scuffing). Macrospalling might be affected by micro-rhd lubrication through the sum of shear stress contributions from individual microconjunctions. This contribution increases markedly as H/Z decreases in the single-asperity model (since the single asperity contribution to the total is $\pi b^2\tau$). Since H should be directly dependent on the macrosurface separation, it is not surprising that decreasing the ratio of macro-elastohydrodynamic film thickness to asperity height beyond a certain limit produces macrospalling (ref. 106). Nor, in the light of the effects of surface films on micro-rhd lubrication, is it surprising that lubricant constitution influences pitting (ref. 107). Similarly, on the basis of Tabor's analysis (ref. 104), gross plastic flow of the bearing surface material would be expected to depend on the shear stress contribution from the microconjunctions. Thus, "wiping" of bearings under macropressures that are low compared to their hardness might be expected when the micro-rhd lubrication leads to high friction.

Scuffing of bearing surfaces leads, via metal transfer and plastic flow during the tearing of microwelds, to an increase in asperity heights (i.e., Z in the single asperity model) and a decrease in asperity curvatures (i.e., r in the model). Since both of these changes in asperity topography would be expected to decrease the effectiveness of micro-rhd lubrication, the self-accelerating nature of scuffing is consistent with this model. Initiation of scuffing appears to require microdistortion of the surfaces by plastic flow but not necessarily by high temperature. Such microdistortion towards higher and/or sharper asperities might be expected as a result of plowing or adhesion. Conceivably, it could result also from the action of the micro-rhd stresses on an asperity made plastic by these stresses. Figure 13 indicates that average pressures may be sufficient to initiate plastic deformation of asperities (ref. 96), and any micro-rhd shear stress should then lead to increased plastic deformation (ref. 105). Thus all of these initiation phenomena would be expected to depend on the stresses on asperity produced under micro-rhd conditions.

The importance of initiating scuffing by micro-rhd produced stresses permits explanation of the many factors affecting load-carrying capacity, for example, so-called "critical" or "transition" temperatures. Thus surface film melting point and "shear strength" should affect load-carrying capacity as Godfrey states. In addition, experimentally observed load-carrying capacity can be affected by macro-ehd film thickness (ref. 108), bulk lubricant viscosity, sliding velocity (ref. 109), and Blok's flash temperature (ref. 110). The effect of Blok's flash "temperature" in producing scuffing is rationalized as resulting from reduced viscosity at the entrance to micro-ehd lubricated conjunctions. The effects of the other observed parameters are explained by the previously discussed greater or lesser effects on the asperity stresses in the presence of micro-rhd films. The factor (or factors) that determines whether scuffing is initiated depends intimately on the overall bearing system configuration, operating variables, microtopography, and rheological properties of the surfaces and surface films. Since the surfaces and surface films change during the course of sliding, changes in load-carrying capacity with test procedure should and do (ref. 111) occur. Further, the asperity stresses under starved conditions should depend on the film thickness (e.g., fatty material chain length for an oriented monolayer and the number of monolayers).

Environmental effects on rheology of solids.—The previous section pointed out that plastic flow of the bearing material is an essential feature of lubrication failure and all the types of wear except corrosion. Evidence reviewed by Kramer and Demer (ref. 112) indicates that environment can affect both the general stress-strain, strain rate, and fatigue properties of metals. Most often inorganic surface films have an embrittling effect while adsorbed polar organic molecules have a plasticizing effect. The latter is often called the "Rehbinder Effect" and has been demonstrated with amphipaths which are effective boundary lubricants (e.g., oleic acid).

Published evidence that extreme pressure (EP) lubricants also plasticize metals seems to be primarily based on metalworking studies (ref. 113). In these studies the observed results suggesting a change in the metal ductility or shear strength could also be explained by a change in friction coefficient. Therefore, to eliminate possible friction effects, a statistically designed torsion study was carried out with the torsion specimens immersed in an oil filled heated capsule.* The mild steel test specimens were cylindrical rods with "necks" midway between the ends to assure that fracture occurred midway between the grips. The SAE 30 oil environments were a solvent neutral oil and a sulfurized oil; the latter did not visibly stain (i.e., corrode) the bright steel surfaces even after several days at room temperature. Stress-strain curves were recorded for each run.

Two parameters, the strain at which fracture occurred, and consequently the energy to fracture, varied significantly with the variables studied. The effects of all the variables except specimen-to-specimen differences are summarized in figure 17. Summarizing briefly; first,

---

*Data by H. T. Marshall and R. S. Fein, unpublished.

the strain-to-fracture, or ductility, with the mineral oil environment is greater than that with the sulfurized oil environment, and secondly, the relative ductility with the two oils changes with torsion (i.e., strain) rate and temperature. In other words, the EP sulfurized oil tends to embrittle the specimens relative to the mineral oil effect, and the amount of embrittlement depends on operating conditions (the maximum and minimum shown in figure 17 are not statistically significant). Thus, at least for these conditions, lubricant environment can affect the rheology of metals and the effect of an EP oil is the opposite of that reported for fatty acid lubricants.

The most important result of varying metal rheology would appear to be control of the consequences of plowing and adhesive contact between boundary lubricated surfaces (ref. 114). Plasticizing the surface material would be expected to lead to increased surface distortion due to plowing and increased junction growth under micro-rhd stresses. On the other hand, embrittlement should decrease damage. As shown by table 2, this explanation appears to correlate with the extent of the wear and friction jumps accompanying initial seizure (or "transition") in the four ball machine. The initial seizure transition has been associated with the failure of micro-rhd lubrication (ref. 108); and presumably reflects the onset of microconjunction growth. The amount of wear occurring after the initial seizure is one of the measures of EP oil action included in the so-called "Mean Hertz Load."

In addition to the influence of EP oils and other nonflowing films on the direct control of microjunction growth, the films should influence the stresses which initiate microconjunction growth. This should occur because a solid film can change the effective elastic modulus of an asperity and, according to equation 4, the average pressure on an asperity is directly proportional to effective elastic modulus. Experimental data (ref. 115) indicate that coatings with different elastic moduli begin to change the effective elastic modulus when coating thicknesses approach the Hertzian radius of the substrate material. As Godfrey (ref. 116) and others (ref. 117) have pointed out, EP films often are of the order of 10^{-4} cm in thickness. From figure 13 and from experiments (refs. 118 and 119), b is also of the order of 10^{-4} to 10^{-3} cm. Hence, EP films should be sufficiently thick to affect the effective elastic moduli of asperities. Since oxides and other inorganic compounds of metals usually have lower elastic moduli than the parent metal, the inorganic portion of the EP films should favorably influence the effective elastic modulus. Further, as illustrated in tables 3 and 4 by carbon and hydrogen content as well as the "organic" portion, EP films generally seem to contain a major organic component. This component would be expected to reduce greatly the elastic modulus of the film and, therefore, the

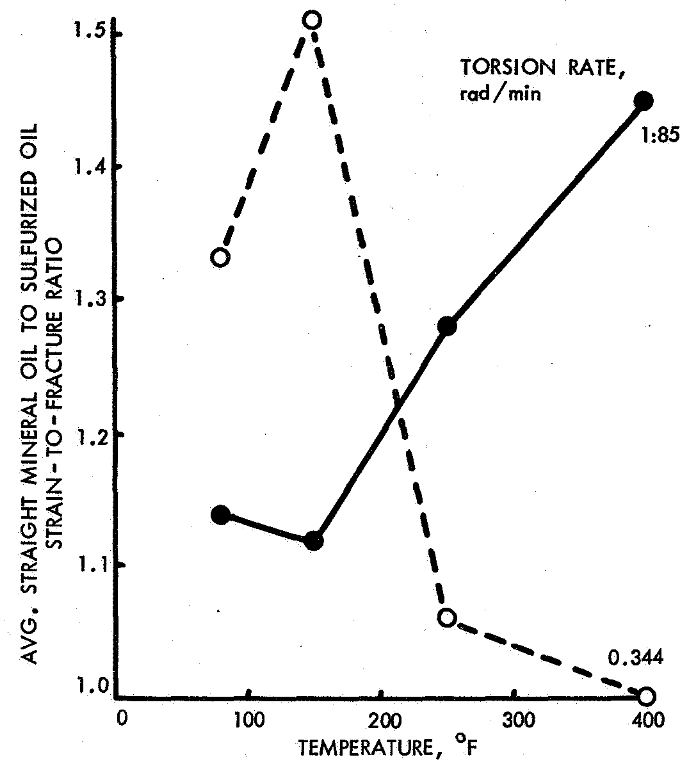


FIGURE 17.—Effects of temperature and torsion rate on relative torsional ductility of straight mineral and sulfurized oils. SAE 1080 steel.

TABLE 2.—*Effect of Fatty Acid and EP Additives on Extent of Wear and Friction "Jumps" at Transition* [a]

Additive	Wear, (cc/kg) × 10^9		Friction coef.	
	Below transition	Above transition	Below transition	Above transition
0.43% stearic acid	4. 5	8300	0. 08	0. 29
None	5. 9	290	0. 08	0. 11
0.375% zinc diisopropyldithiophosphate	11	32	0. 12	0. 14

[a] Four ball machine; load 40 to 50 kg; 1800 rpm; 10 sec; base oil, squalane.

average pressure $\overline{P}$ on an asperity. In turn, the decreased pressure should, because of the junction growth mechanism (ref. 104), reduce the tendency for microconjunctions to grow and adhere.

Mr. Godfrey indicates that shear strength, melting point, and hardness of solid films are important to lubrication. However, elastic modulus, hardness, and shear strength are all fairly directly related (ref. 121). Thus, the micro-rhd model suggests that these apparent effects of solid film properties might be equally well explained in terms of elasticity; until temperatures become sufficiently high to allow film flow.

Summary.—The foregoing micro-rhd model qualitatively describes boundary lubrication in terms of (1) dynamically shearing lubricant films between asperities, and (2) the influence of these films on asperity stresses and the tendency to fatigue and plastically deform the asperities. The quantitative plausibility of the model is demonstrated with an idealized single asperity picture. The *qualitative* plausibility of the model is demonstrated by a few examples illustrating how it permits explanation of boundary lubrication phenomena.

Admittedly, the model is not as simple as that of classical boundary lubrication or smooth-surface macro-elastohydrodynamic lubrication. However, the model produced by integration of these concepts, coupled with less ideal material rheology, appears to have the generality necessary to account for the diversity of experimental boundary lubrication results. If indeed the micro-rhd model can account for these results, then its lack of simplicity is justified. The question of justification requires further amplification and quantification of the model, and these requirements lead to the research needs listed below. Implicit in this suggested research is the development of new knowledge relating

TABLE 3.—*Constitution of EP Films and Approximate Atomic Ratios in EP Films*

Component	EP lubricant type	
	Sulfur[a]	Borate[b]
Iron	1. 0	1. 0
Oxygen	1. 2	4. 7 (by difference)
Carbon	0. 5	1. 0
EP element	0. 02 (sulfur)	0. 82 (boron)
Hydrogen	0. 8	Not determined

[a] From ref. 116.
[b] From ref. 117.

TABLE 4.—*Constitution of EP Films and Ratio of "Organic" to "Inorganic" Phosphate in Effective Phosphate-ester Film*

[From ref. 120]

Surface	Ratio "organic" to "inorganic" phosphate
Chrome-plate	3. 1
Hardenable iron	1. 3
Chilled cast iron	0. 37

physical properties of bearing and lubricant materials to their constitution and to their chemistry related to boundary lubrication.

Indicated research needs.—Amplification and quantification of the micro-rhd model points to the need for considerable research beyond that indicated by Mr. Godfrey. Roughly, this research may be divided into elaboration of the model and development of techniques to measure pertinent physical properties of surfaces and surface films. The research should include:

(1) Analyses and experiments with realistic assemblages of asperities to determine the influence of surface topography

(2) Analyses with idealized constitutive equations for non-Newtonian liquid and elastic-plastic solid micro-rheodynamic lubricant films

(3) Analyses and experiments with gradients of lubricant-film rheological properties normal to the surfaces; a statistical mechanical analysis, of the influence of a solid surface on the shear viscosity of a particulate liquid near the surface (the particulate liquid, of course, is a model assemblage of molecules); and the rheological factors affecting adhesion to the substrate

(4) Analyses and experiments to determine how an asperity plastically changes in shape under the influence of forces generated by micro-rheodynamic lubrication

(5) Consideration of the model in planning and interpreting experiments on boundary lubrication, load-carrying capacity, etc.

Pertinent physical properties to be determined should include:

(a) Surface topography descriptions with at least statistical distributions of asperity heights and crown radii consistent with the results of Item 1 above (It should be noted here that the Greenwood and Williamson's description (ref. 101) certainly is an improvement over common practice and may even be adequate.)

(b) Rheological properties of bearing surfaces consisting of the stress-strain and strain-rate characteristics in appropriate stress fields and in the presence of various environments including solid films; these properties necessarily include more than just elastic modulus, shear strength, and hardness

(c) Rheological properties of solid films applied or formed on the bearing surface

(d) Rheological properties of more or less "liquid" lubricant films formed on bearing surfaces; these properties include more than just viscosity and shear strength

P. M. Ku (Southwest Research Institute, San Antonio, Texas)

As indicated in the remarks made by Mr. Godfrey, Dr. Campbell, and Drs. Fein and Kreutz, there is, as it has been for many years, considerable disagreement with regard to the definition of boundary lubrication. I agree with Mr. Godfrey and Dr. Campbell that Dowson's definition, limiting the lubricant film to monolayer thickness, is too restrictive. I like to think of boundary lubrication simply as that regime of lubrication where substantial asperity contacts take place through the lubricant film, regardless of the film type or its interactions with the surfaces. It appears to me that the latter relates to the mechanism of boundary lubrication, rather than to its definition.

Turning to the case of liquid lubricants, many investigators have considered that there are three regimes of operation: hydrodynamic lubrication, mixed lubrication, and boundary lubrication. Recent advances in elastohydrodynamics have clearly established that a goodly portion of the so-called mixed lubrication regime, that portion immediately adjacent to the hydrodynamic lubrication regime, is in fact a regime of elastohydrodynamic lubrication. One might then speak of four regimes of lubrication: hydrodynamic lubrication, elastohydrodynamic lubrication, a regime of uncertainty, and then boundary lubrication. In the regime of uncertainty, one might include the micro-elastohydrodynamic lubrication proposed by Drs. Fein and Kreutz, and perhaps others. I refer to it as a regime of uncertainty, rather than as mixed lubrication, because not enough is known about it at present.

It appears to me that there is a continuous progression of lubrication regimes from boundary to hydrodynamic. Our difficulty is not only in definitions, but in learning more about the transition from one regime to another as well as the mechanisms involved. Once the nature of the problem is better understood, definitions may hopefully evolve by themselves.

J. F. Archard (University of Leicester, Leicester, England)

I would like to commiserate with Mr. Godfrey in the difficulties he faces in attempting to define boundary lubrication. This is a persistent problem somewhat akin to the difficulties in the classification of wear which I have discussed. In my paper, in discussing the question of wear under lubricated conditions, it was suggested that boundary lubrication could be described as "lubrication by films whose properties and thicknesses are determined by the molecular rather than bulk properties of the lubricant."

However, it must be accepted that all such definitions or systems of classification are an oversimplification of the true situation; more probably there exists a continuous spectrum of conditions ranging from a full fluid film to solid contact. Thus, it has been shown (refs. 122 and 123) that, with sufficiently smooth surfaces, the full fluid-film regime persists under conditions of elastohydrodynamic lubrication, down to film thicknesses of some hundreds of Angstrom units. Moreover, under these conditions, the coefficient of friction was indistinguishable from that observed under boundary lubrication with monomolecular films; thus measurement of the coefficient of friction provided no reliable indication of the lubrication regime.

Between such conditions of thin, full, fluid films and those defined as "surface interaction between monomolecular layers," lies an interesting region which has not been mentioned at this symposium. For example, Cameron and co-workers (refs. 124 and 125) have suggested that, if the fluid film becomes sufficiently thin, the molecular architecture can have a significant influence upon its thickness and other properties. In this connection they might have cited the work of Deryaguin (ref. 126) which invokes the existence of long range molecular forces. The range of such surface forces is a subject of some controversy. However, further examination of the properties of such thin lubricant films is a subject of both considerable fundamental interest and great practical importance.

The discussion by Drs. Fein and Kreutz is a well argued and provocative statement of the case for the existence of ehd effects at asperity contacts. I agree with their conclusions insofar as isolated asperities are considered; however, when groups of asperities upon nonconforming surfaces (such as those of rolling disks) are considered, the situation becomes more complicated, and it is difficult to provide a solution which satisfies both the elastic and hydrodynamic equations. I have argued (ref. 127) that appreciable flattening of an asperity can occur only if (in the terminology of figure 11) the Z/r ratio is less than the Z/r ratio of the Hertzian load bearing region between the disks.

R. W. Roberts (General Electric Research & Development Center, Schenectady, New York)

Over the past several years, we have developed a useful approach to predicting boundary lubricant additives for difficult-to-lubricate metals (refs. 128 and 129). During sliding under boundary conditions, metal surfaces wear and expose clean metal. By studying the surface chemistry of similar clean metal surfaces, using ultrahigh vacuum and associated techniques, it is possible to identify chemical species which can function as boundary lubricants. New practical boundary lubricants for aluminum were a direct outgrowth of such an approach (ref. 130). Currently these techniques are being used to predict and investigate new lubricants for stainless steels and the heat-resistant alloys of nickel and cobalt (ref. 131).

LECTURER'S CLOSURE

Drs. Campbell, Fein, Kreutz, Ku, and Archard have submitted definitions of boundary lubrication. Dawson's definition, which I used, is open to some criticism. First, limiting boundary lubrication to monomolecular films is questionable. Certainly multimolecular films are influential. Secondly, the wording is confusing: "Surface interaction between monomolecular layers . . . and solids . . ." is not clear because the monomolecular layer is a result of the interaction of the lubricant and the solid. The use of the word "boundary" in the definition of boundary lubrication seems unwise. Lastly, asperity contact implies metal-to-metal contact, which is not necessary. As Archard has shown, many contacts of film-to-film or oxide-to-oxide may occur without metal-to-metal contact. My definition, like the others, shows how they are tailored to each author's concept and analytical approach. I prefer the following: Boundary lubrication is that state of lubrication where surface films influence friction and wear.

I appreciate Dr. Campbell's comments on the chemistry of boundary lubrication because they balance my view on the importance of physical properties of resultant films. However, I can rationalize some of the mechanisms he discusses as physical phenomena. For example, if rate of film removal exceeds rate of film formation or if the film is dissolved, an incomplete or discontinuous film would result. Is not the "hole" in the film the important result? Similarly, if the metal oxide film formed by organic peroxides had more resistance to removal, would not fretting wear be reduced?

I agree that use of the term "chemical reaction" to denote the formation of a film of inorganic salt on metal, such as iron sulfide, is not satisfactory. A better term is needed.

Drs. Fein and Kreutz have offered a theory of boundary lubrication involving no asperity contact which they use to explain numerous observations. This theory needs experimental support as the authors state and indicate in their list of areas of needed research. I do not mean to contest the theory, but in future writings I would appreciate explanations of observations of low contact resistance and additional discussion of the marked effect of a small concentration of additives on friction and wear.

REFERENCES

1. Hardy, W. B.; and Doubleday, I.: Boundary Lubrication. The Paraffin Series. Proc. Roy. Soc. (London), vol. A100, 1922, p. 550.
2. Dowson, D.: The Transition to Boundary Lubrication From Elastohydrodynamic Lubrication. Res. Rept., Univ. of Leeds, Dept. of Mech. Eng., Nov. 1965.
3. Larsen, R. G.; and Perry, G. L.: Chemical Aspects of Wear and Friction. Mechanical Wear, J. T. Burwell, Jr., ed., ASM, 1950, p. 73; Discussion by H. H. Uhlig, p. 90.
4. Akhmatov, A. S.: Molecular Physics of Boundary Lubrication. Israel Program for Scientific Translations (Dist. by Daniel Davey & Co., New York, 1966).
5. Rowe, G. W.: Boundary Lubrication. Principles of Lubrication, A. Cameron, John Wiley & Sons, 1966, chap. 20.
6. Davies, C. B.: A Review of Boundary Lubrication With Particular Reference to the Conference Papers. Proc. Conf. on Lubrication and Wear, Inst. Mech. Engrs., 1957, p. 319.
7. Bowden, F. P.; and Tabor, D.: The Friction and Lubrication of Solids, Pt. I, 1954; Pt. II, 1964, Oxford Univ. Press.
8. Barwell, F. T.: Some Further Thoughts on the Nature of Boundary Lubrication. Review of Roumaine des Sciences Techniques, Series de Mecanique Applique, vol. 11, 1966, p. 683.
9. Zisman, W. A.: Historical Review, Lubricants and Lubrication. Synthetic Lubricants, R. C. Gunderson and A. W. Hart, eds., Reinhold Publ. Corp., New York, 1962, chap. 2.
10. Bisson, E. E.; and Anderson, W. J.: Advanced Bearing Technology, NASA, Washington, D.C., 1964.
11. Stewart, W. T.; and Stuart, F. A.: Lubricating Oil Additives. Advances in Petroleum Chemistry and Refining, J. J. McKetta, Jr., ed., vol. 7, Interscience Publ., John Wiley & Sons, New York, 1963.
12. Dunken, H.: Chemisorption of Hydrocarbon on Metal. Ges. Schmiertech Symposium Esson, Sept. 22–23, 1966 (Abst. in Erdoel und Kohle, vol. 19, 1966, p. 910).
13. Levine, O.; and Zisman, W. A.: Physical Properties of Monolayers Adsorbed at the Solid Air Interface. II. Mechanical Durability of Aliphatic Polar Compounds and Effect of Halogenation. J. Phys. Chem., vol. 61, 1957, p. 1188.
14. Mentor, J. W.; and Tabor, D.: Orientation of Fatty Acid and Soap Films on Metal Surfaces. Proc. Roy. Soc. (London), vol. A204, 1951, p. 514.
15. Holm, R.: Electric Contacts, Almqvist and Wiksells Akademiska Handböcker, Hugo Gebers Förlag (Stockholm), 1946, p. 198.
16. Andrew, L. T.: Electron-Diffraction Analysis of the Orientation of the Molecules of Lubricating Oils. Trans. Faraday Soc., vol. 32, 1936, p. 607.
17. Sakurai, T.; and Sato, K.: Study of Corrosivity and Correlation Between Chemical Reactivity and Load Carrying Capacity of Oils Containing Extreme Pressure Agents. ASLE Trans., vol. 9, 1966, p. 77.
18. Schulman, J. H.; Waterhouse, R. B.; and Spink, J. A.: Adhesion of Amphipathic Molecules to Solid Surfaces. Kolloid Zeitschrift, vol. 146, 1956, p. 77.
19. Godfrey, D.: Chemical Changes in Steel Surfaces During Extreme Pressure Lubrication. ASLE Trans., vol. 5, 1962, p. 57.

20. Fein, R. S.; and Kreuz, K. L.: Chemistry of Boundary Lubrication of Steel by Hydrocarbons. ASLE Trans., vol. 8, 1965, p. 29.
21. Zisman, W. A.: Chemistry of Organic Compounds on Surfaces of High and Low Energy. Proc. Sym. on Chemistry of Solid Surfaces, Durham, N.C., Office Ord. Res., U.S. Army, 1958, p. VIII–1.
22. Klaus, E. E.; and Bieber, H. E.: Effects of P^{32} Impurities on the Behavior of Tricresyl Phosphate-32 as an Anti-Wear Additive. ASLE Trans., vol. 8, 1965, p. 12.
23. Tingle, E. D.: Influence of Water on the Lubrication of Metals. Nature, vol. 160, No. 4073, 1947, p. 710.
24. Grunberg, L.; and Campbell R. B.: Metal Transfer in Boundary Lubrication and the Effect of Sliding Velocity and Surface Roughness. Proc. Conf. on Lubrication and Wear, Inst. Mech. Engrs., 1957, p. 290.
25. Wilson, R. W.: The Contact Resistance and Mechanical Properties of Surface Films on Metals. Proc. Phy. Soc., (London), vol. B68, 1955, p. 625.
26. Rabinowicz, E.: An Investigation of Surface Damage Using Radioactive Metals. Brit. J. Appl. Phys., 1951, p. 82.
27. Leach, E. F.; and Kelley, B. W.: Temperature—The Key to Lubricant Capacity. ASLE Trans., vol. 8, 1965, p. 271.
28. Fein, R. S.; Rowe, C. N.; and Kreuz, K. L.: Transition Temperatures in Sliding Systems. ASLE Trans., vol. 2, 1959, p. 50.
29. Cowley, C. W.; Ultee, C. J.; and West, C. W.: Influence of Temperature on Boundary Lubrication. ASLE Trans., vol. 1, 1958, p. 281.
30. Miller, A.; and Anderson, A. A.: Investigation of Oil Additives for Boundary Lubrication of Railroad Journal Bearings. ASLE Trans., vol. 1, 1958, p. 273.
31. Brophy, J. E.; and Zisman, W. A.: Surface Chemical Phenomena in Lubrication. Ann. New York Acad. Sci., vol. 53, 1951, p. 836.
32. Russell, J. A.; Campbell, W. E.; Burton, R. A.; and Ku, P. M.: Boundary Lubrication Behavior of Organic Films at Low Temperatures. ASLE Trans., vol. 8, 1965, p. 48.
33. Frewing, J. J.: The Influence of Temperature on Boundary Lubrication. Proc. Roy. Soc. (London), vol. A181, 1942, p. 23.
34. Kingsbury, E. P.: Some Aspects of the Thermal Desorption of a Boundary Lubricant. J. Appl. Phys., vol. 29, 1958, p. 888.
35. Barwell, F. T.; and Milne, A. A.: The Lubrication of Rough Steel Surfaces by a Series of Metallic Soaps. Brit. J. Appl. Phys., 1951, p. 44.
36. Rowe, G. W.: Vapour Lubrication and the Friction of Clean Surfaces. Proc. Conf. on Lubrication and Wear, Inst. Mech. Engrs., 1957, p. 333.
37. Braithwaite, E. R.: Solid Lubricants and Surfaces, MacMillan Co., 1964, p. 55.
38. Peterson, M. B.: High Temperature Lubrication. Proc. International Sym. on Lubrication and Wear, D. Muster and B. Sternlicht, eds., McCutchan Publ. Corp., Berkeley, 1965, p. 941.
39. Rowe, C. N.; Fein, R. S.; and Kreuz, K. L.: Relationship Between Lubricant Oxidation and Wear, Part I: Friction and Wear Reduction by Fatty Acid Lubricants. Part II: Lubricant Oxidation in Wearing Systems. Sym. on Chemistry of Friction and Wear, ACS, vol. 3, No. 4A, 1958, p. A111.
40. Wistreich, J. G.: Lubrication in Wire Drawing. Proc. Conf. on Lubrication and Wear, Inst. Mech. Engrs., 1957, p. 505.
41. Bailey, A. I.; and Courtney-Pratt, J. S.: The Area of Real Contact and the Shear Strength of Monomolecular Layers of a Boundary Lubricant. Proc. Roy Soc. (London), vol. A227, 1955, p. 500.

42. White, J. R.: Frictional Properties of Soaps at High Pressures. Lub. Eng., vol. 10, 1954, p. 340.
43. Merchant, M. E.: Cutting Fluid Action and the Wear of Cutting Tools. Proc. Conf. on Lubrication and Wear, Inst. Mech. Engrs., 1957, p. 566.
44. Bridgeman, P. W.: Shearing Phenomena at High Pressures, Particularly in Inorganic Compounds. Proc. Amer. Acad. Arts and Scs., vol. 71, 1937, p. 387.
45. Boyd, J.; and Robertson, B. P.: The Friction Properties of Various Lubricants at High Pressures. Trans. ASME, vol. 67, 1945, p. 51.
46. Godfrey, D.: The Effect of High Temperature on Friction and Wear. Proc. 5th World Petroleum Congress, 1959, p. 345.
47. Patrick, R. L.: Orientation and Adhesion of Monolayers. Proc. Sym. on Chemistry of Solid Surfaces, Durham, N.C., Office Ord. Res., U.S. Army, 1958, p. IV–1.
48. Dies, K.: The Friction Oxidation as a Chemico-Mechanical Process. Archives für das Eisenhüttenwesen, vol. 16, 1943, p. 399.
49. Piggott, M. R.; and Wilman, A.: Nature of the Wear and Friction of Mild Steel on Mild Steel and the Effect of Surface Oxide and Sulfide Layers. Proc. Conf. on Lubrication and Wear, Inst. Mech. Engrs., 1957, p. 613.
50. Williamson, J. B. P.: Recent Studies of Physics of Electrical Connector Surfaces. International Convention of IEEE, 1966, Special Convention Record Part 9.
51. Reason, R. E.: Significance and Measurement of Surface Finish. Grinding and Finishing Magazine, no. 4, Engis Equipment Co., Morton Grove, Ill.
52. Peterson, M. B.; and Johnson, R. L.: Factors Influencing Friction and Wear with Solid Lubricants. Lub. Eng., vol. 11, 1955, p. 325.
53. Cook, H. D.: Adsorption of Radiolabeled 2-Hydroxystearic Acid From Solution Onto Solid Surfaces. Sym. on Chemistry of Friction and Wear, Am. Chem. Soc., vol. 3, No. 4A, 1958, p. A15.
54. Milne, A. A.: Friction Experiments With a Soft Solid. Wear, vol. 2, 1958, p. 28.
55. Hirst, W.; and Lancaster, J. K.: The Influence of Oxide and Lubricant Films on the Friction and Surface Damage of Metals. Proc. Roy. Soc. (London), vol. A223, 1954, p. 324.
56. Antler, M.: The Lubrication of Gold. Wear, vol. 6, 1963, p. 44.
57. Lancaster, P. R.; and Rowe, G. W.: A Comparison of Boundary Lubricants Under Light and Heavy Loads. Wear, vol. 2, 1959, p. 428.
58. Gulbransen, E. A.; Copan, T. P.; van Rooyen, D.; and Andrew, K. F.: Crystal Habits of the Corrosion Products of Gas-Metal Reactions. Sym. on Chemistry of Friction and Wear, Am. Chem. Soc., vol. 3, No. 4A, 1958, p. A39.
59. Osias, J. R.; and Tripp, J. H.: Mechanical Disruption of Surface Films on Metals. Wear, vol. 9, 1966, p. 388.
60. Pomey, J.: Friction and Wear (Translation of Le Frottemont el L'Usure, Office National d'Etudes et de Recherches Aeronautics Rapport Technique, No. 36, 1948), NACA, Washington, D.C., 1952.
61. Feng, I. M.: Mechanisms of Wear. Proc. Sym. on Chemistry of Solid Surfaces, Durham, N.C., Office Ord. Res., U.S. Army, 1958, p. VI–1.
62. Godfrey, D.: The Lubrication Mechanism of Tricresyl Phosphate on Steel. ASLE Trans., vol. 8, 1965, p. 1.
63. Barcroft, F. T.; and Daniel, S. G.: The Action of Neutral Organic Phosphates as EP Additives. J. Basic Eng., ASME, vol. D87, 1965, p. 761.
64. Gulbransen, E. A.: Model of Oxidation Processes Occurring in Oxide

Whiskers and Platelets on Iron at 400° C to 500° C. Processus de Nucleation Dans Les Reactions Des Gas Sur Les Metaux et Problems Connexes, Editions Du Centre National de La Recherche Scientifique, Paris, No. 122, 1965.

65. Hardy, W. B.; and Doubleday, I.: Boundary Lubrication. The Paraffin Series. Proc. Roy. Soc. (London), vol. A100, 1922, p. 550.
66. Bisson, E. E.; and Anderson, W. J.: Advanced Bearing Technology, NASA SP–38, Washington, D.C., 1964, Chap. 2.
67. Campbell, W. E.: Studies in Boundary Lubrication. Trans. ASME, vol. 61, 1939, p. 633.
68. Bisson, E. E.; Johnson, R. L.; Swikert, M. A.; and Godfrey, D.: Friction, Wear and Surface Damage of Metals as Affected by Solid Surface Films. NACA TN 3444, 1955.
69. Bowden, F. P.; and Leban, L.: The Friction of Lubricated Metals. Proc. Roy. Soc. (London), vol. A239, 1940, p. 1.
70. Greenhill, E. B.: The Lubrication of Metal Surfaces by Mono- and Multi-Molecular Layers. Trans. Faraday Soc., vol. 45, 1949, p. 631.
71. Bowden, F. P.; and Tabor, D.: The Friction and Lubrication of Solids. Pt. I, Oxford Univ. Press, 1950.
72. Fein, R. S.; and Kreuz, K. L.: Chemistry of Boundary Lubrication of Steel. Trans. ASLE, vol. 8, 1965, p. 29.
73. Fein, R. S.; and Kreuz, K. L.: Lubrication and Wear. Lubrication, vol. 51, 1965, p. 61.
74. Vinogradov, G. V.; Lian, Go-Lin, and Paolovskaia, N. T.: Anti-Wear and Anti-Friction Properties of Lubricants Under Severe Friction Conditions. (in Russian), Trenie I Iznos V Mashinakh, vol. 15, 1964, p. 393.
75. Rounds, F. G.: Some Environmental Factors Affecting Surface Coating Formation With Lubricating Oil Additives. Trans. ASLE, vol. 9, 1966, p. 88.
76. Vinogradov, G. V.; Arkharariva, V. V.; and Petrov, A. A.: Anti-Wear and Anti-Friction Properties of Hydrocarbons Under Heavy Loads. Wear, vol. 4, 1961, p. 274.
77. Weatherford, W. D., Jr.; Valtierra, M. L.; and Ku, P. M.: Mechanisms of Wear in Misaligned Splines. J. Lub. Tech., ASME, vol. 90F, 1968, p. 42.
78. Whitehead, J. R.: Surface Deformation and Friction of Metals at Light Loads. Proc. Roy. Soc. (London), vol. A201, p. 109.
79. Russell, J. A.; Burton, R. A.; and Ku, P. M.: Atmospheric Effects on Friction and Wear of Oxidized Copper, Wear, vol. 8, 1965, p. 173.
80. St. Pierre, L. E.; Owens, R. S.; and Klint, R. V.: Chemical Effects in the Boundary Lubrication of Aluminum. Nature, vol. 202, 1964, p. 1204.
81. Grünberg, L.: The Study of Freshly Deformed Metal Surfaces With the Aid of Exo-Electron Emission. Wear, vol. 1, 1957, p. 142.
82. Feng, I. M.: Plastic Roughening and Wear. Proc. Conf. on Lubrication and Wear, Inst. Mech. Engrs. 1957, p. 635.
83. Elin, L. V.: Mutual Intrusion of the Surface Layers of Metals as One of the Causes of Wear During Imperfect Lubrication. Trenie I Iznos V Mashinakh, vol. 13, 1959 (English translation: Friction and Wear in Machinery. ASME, 1961).
84. Kostetskii, B. I.; and Nazarenko, P. V.: Interaction of Surfaces in the External Friction of Solids. Soviet Physic Doklady, vol. 10, 1965, p. 59.
85. Kraghelskii, I. V. (L. Ronson and J. K. Lancaster, trans.): Friction and Wear. Butterworths Scientific Publ., (Wash., D.C.), 1965.
86. Likhtman, V. I.; Rebinder, P. A.; and Karpenko, G. V.: Effect of Surface-

Active Media on the Deformation of Metals. Chemical Publ. Co. (New York), 1960.

87. Benedicks, C. W.: Proc. International Symp. on Reactions of Solids. (Gothenberg), 1952, p. 477.
88. Shaw, M. C.: On the Action of Metal Cutting Fluids at Low Speeds. Wear, vol. 2, 1959, p. 217.
89. Kohn, E. H.: A Theory on the Role of Lubricants in Metal Cutting at Low Speeds and in Boundary Lubrication, Wear, vol. 8, 1965, p. 43.
90. Germer, L. H.; and Storks, K. H.: Rubbed Films of Barium Stearate and Stearic Acid, Phys. Rev., vol. 55, 1939, p. 648.
91. Bowden, F. P.; and Tabor, D.: The Friction and Lubrication of Solids, Pt. II, Oxford Univ. Press, 1964, p. 399.
92. Lubrication and Wear. Lubrication, vol. 51, 1965, p. 60.
93. Dowson, D.; and Higginson, G. R.: Elasto-Hydrodynamic Lubrication, Pergamon Press (Oxford), 1966.
94. Archard, J. F.; and Cowking, E. W.: Elastohydrodynamic Lubrication at Point Contacts. Elastohydrodynamic Lubrication, Inst. Mech. Engrs., 1965–66, p. 47.
95. Timoshenko, S.; and Goodier, J. N.: Theory of Elasticity, McGraw-Hill Book Co., 1951.
96. Archard, J. F.: Elastic Deformation and the Laws of Friction. Proc. Roy. Soc. (London), vol. A243, 1957, p. 190.
97. Bowden, F. P.; and Tabor, D.: The Friction and Lubrication of Solids. Pt. I, 1954; Pt. II, 1964, Oxford Univ. Press.
98. Bell, J. C.; Kannel, J. W.; and Allen, C. M.: The Rheological Behavior of the Lubricant in the Contact Zone of a Rolling Contact System. J. Basic Eng., ASME, vol. 86D, 1964, p. 423.
99. Barlow, A. J.; and Lamb, J.: The Visco-Elastic Behavior of Lubricating Oils Under Cyclic Shearing Stress. Proc. Roy. Soc. (London), vol. A253, 1959, p. 52.
100. Fein, R. S.: Possible Role of Compressional Viscoelasticity in Concentrated Contact Lubrication. J. Lub. Tech., ASME, vol. 89F, 1967, p. 127.
101. Greenwood, J. A.; and Williamson, J. B. P.: Contact of Nominally Flat Surfaces. Proc. Roy. Soc. (London), vol. A295, 1966, p. 300.
102. Zisman, W. A.: Friction, Durability and Wettability Properties of Monomolecular Films on Solids. Friction and Wear, R. Davies, ed., Elsevier Publ. Co., 1959.
103. Keller, D. V.: The Analysis of Metallic Adhesion Data. Adhesion or Cold Welding of Materials in Space Environments, ASTM, 1967, p. 181.
104. Tabor, D.: Effects of Lubricants on Transition Temperatures. ASLE Trans., vol. 8, 1965; Discussion, p. 66.
105. Tabor, D.: Junction Growth in Metallic Friction: The Role of Combined Stresses and Surface Contamination. Proc. Roy. Soc. (London), vol. A251, 1959, p. 378.
106. Dawson, P. H.: Further Experiments on the Effect of Metallic Contact on Pitting of Lubricated Rolling Surfaces. Elastohydrodynamic Lubrication, Inst. Mech. Engrs., 1965–66, p. 95.
107. Barwell, F. T.; and Scott, D.: Effect of Lubricants on Pitting Failure of Ball Bearings. Engineering, vol. 182, 1956, p. 9.
108. Dudley, D. W.: Elastohydrodynamic Behavior Observed in Gear Tooth Action. Elastohydrodynamic Lubrication, Inst. Mech. Engrs., 1965–66, p. 206.

109. Fein, R. S.: Effects of Lubricants on Transition Temperatures. ASLE Trans., vol. 8, 1965, p. 59.
110. Leach, E. F.; and Kelley, B. W.: Temperatures—The Key to Lubricant Capacity. ASLE Trans., vol. 8, 1965, p. 271.
111. Fein, R. S.: Effect of Operating Procedures on Critical Temperatures. ASLE Trans., vol. 10, 1967, p. 373.
112. Kramer, I. R.; and Demer, L. J.: Effects of Environment on Mechanical Properties of Metals. Prog. in Materials Sci., vol. 9, 1961, p. 133.
113. Epifanov; et al.: The Action Mechanism of Active Media in the Cutting of Metals. Doklady Akad. Nauk SSSR, vol. 97, 1954, p. 277.
114. Greenwood, J. A.; and Tabor, D.: The Properties of Model Friction Junctions. Proc. Conf. on Lubrication and Wear, Inst. Mech. Engrs., 1957.
115. Waters, N. W.: The Indentation of Thin Rubber Sheets by Spherical Indentors. Brit. J. Appl. Phys., vol. 16, 1965, p. 557.
116. Godfrey, D.: Chemical Changes in Steel Surfaces During Extreme Pressure Lubrication. ASLE Trans., vol. 5, 1962, p. 57.
117. Kreuz, K. L.; Fein, R. S.; and Dundy, M.: EP Films From Borate Lubricants. ASLE Trans., vol. 10, 1967, p. 67.
118. Rabinowicz, E.: A Quantitative Study of the Wear Process. Proc. Phys. Soc. (London), vol. 66B, 1953, p. 131.
119. Rabinowicz, E.: Autocorrelation Analysis of the Sliding Process. J. Appl. Phys., vol. 27, 1956, p. 131.
120. Barcroft, F. T.; and Daniel, S. G.: The Action of Neutral Organic Phosphates as EP Additives. J. Basic Eng., ASME, vol. 87D, 1965, p. 761.
121. Rabinowicz, E.: Friction and Wear of Materials, John Wiley & Sons, 1965.
122. Archard, J. F.; and Kirk, M. T.: Lubrication at Point Contacts. Proc. Roy. Soc. (London), vol. A261, 1961, p. 532.
123. Dyson, A.; Naylor, H.; and Wilson, A. R.: The Measurement of Oil Film Thickness in Elastohydrodynamic Contacts. Elastohydrodynamic Lubrication, Inst. Mech. Engrs., 1965–66, p. 169.
124. Askwith, T. C.; Cameron, A.; and Crouch, R. F.: Chain Length of Additives in Relation to Lubricants in Thin Film and Boundary Lubrication. Proc. Roy Soc. (London), vol. A291, 1966, p. 550.
125. Cameron, A.; and Gohar, R.: Theoretical and Experimental Studies of the Oil Firm in Lubricated Point Contact. Proc. Roy. Soc. (London), vol. A291, 1966, p. 520.
126. Deryaguin, B. V.: The Mechanism of Boundary Lubrication and the Properties of the Lubricating Film. Wear, vol. 1, 1957–58, p. 277.
127. Archard, J. F.: Discussion of papers by Poon and Haines and by Garnell, Proc. Inst. Mech. Engrs., vol. 181, Pt. 1, 1966–67, p. 377.
128. Roberts, R. W.; and Owens, R. S.: A Chemical Approach to Lubrication. Science Journal (British), vol. 2, 1966, p. 69.
129. Suits, C. G.; and Bueche, A. M.: Cases of Research and Development in a Diversified Company. Applied Science and Technological Progress, Nat. Acad. Sci. (Wash., D.C.), 1967, p. 297.
130. St. Pierre, L. W.; Owens, R. S.; and Klint, R. V.: Chemical Effects in the Boundary Lubrication of Aluminum. Wear, vol. 9, 1966, p. 160.
131. Roberts, R. W.; and Owens, R. S.: New Lubricants for the Heat Resistant Alloys. Friction and Lubrication in Metal Processing (F. F. Ling; R. L. Whitley; P. M. Ku; and M. B. Peterson, eds.), ASME, 1966, p. 24.

Critical Appraisal and Research Opportunities—the Lubrication Research Viewpoint

D. TABOR
University of Cambridge
Cambridge, England

According to the organizers of this symposium the main purpose of this review is twofold: first to spell out, particularly for workers in other disciplines, the main problems in the fields of friction, lubrication, and wear; and secondly to indicate those areas where, as specialists, they might find worthwhile and challenging problems of fundamental interest, to which they could make useful contributions.

Neither of these tasks is easy. As is apparent from the interesting and stimulating papers already prepared, the field is too large for any individual to claim expert knowledge of the whole of it. Problems that may seem to one worker to be fundamental and of challenging importance may seem trivial or boring to the specialist. For these reasons this review must necessarily be imperfect and reflect to some extent the personal viewpoints of the writer.

STRUCTURE OF CLEAN SOLID SURFACES

Dr. Gatos has produced a very full and comprehensive account of current ideas concerning the solid surface. This review discusses six of the points he has raised.

Surface Energy

Surface energy and surface tension are not the same thing. The situation may be summarized by saying that if γ is the free surface energy, the excess energy for a surface area A is γA. The work done in producing an incremental change in A is given by

$$s=\frac{d}{dA}(\gamma A)=\gamma+A\frac{\partial\gamma}{\partial A},$$

where s is the line tension in the surface. For liquids $\partial\gamma/\partial A$ is zero so that s and γ are identical. For solids $\partial\gamma/\partial A$ is not zero; in some cases theory suggests that it may be negative and may even be so negative that $\gamma + A\ \partial\gamma/\partial A$ will be negative; the surface forces will then be compressive rather than tensile. Thus the forces in the surface may actually tend to expand the surface rather than contract it. It is sad to note that experimental study of this has been almost entirely neglected.

Of course the free surface energy is thermodynamically the more important parameter—it enters directly into solid-solid, solid-liquid, and solid-vapour interactions. As Professor Rabinowicz has often pointed out, it may play an important part in the mechanism of wear particle detachment, even though the curling up of wear fragments depends on the surface tension rather than surface energy. In any case direct measurement of the surface tension, perhaps along the lines indicated at the end of Dr. Gatos's paper, is a challenge to the experimental worker, and could be of great value to the theoretician.

Surface Energy and Adhesion of Solids

The free surface energy of metals can now be measured with reasonable accuracy at temperatures a little below the melting point by studying the noncreep equilibrium of a suspended wire. With nonmetals, particularly materials which are brittle at low temperatures, the surface energy can be determined by cleavage methods. The measured values are in fair agreement with values calculated from simple theoretical models. If two atomically smooth surfaces could be placed in contact, the adhesion should depend only on the surface energies and the interfacial energy. There is, as far as I know, no way of determining the interfacial energy of solids. From the point of view of frictional mechanisms, however, contacting surfaces are never atomically smooth (except in the case of mica); further, the contact between surfaces always involves appreciable stresses normal to the surface so that some deformation occurs at the contacting asperities. This introduces a new factor and it is questionable as to whether surface energies per se are important in determining adhesion under these conditions. This is discussed further below.

Range of Action of Surface Forces

Apart from some work in Russia and Britain during the last decade, there has been little systematic study of surface forces. These forces have great importance in interactions between solids and between lubricants and solid surfaces. Our own studies with mica show that surface forces can be very strong for distances of the order of 5 to 10Å but very weak for distances greater than say 20 or 30Å. More detailed study in this area would be of great theoretical and practical value.

The Wetting of Surfaces

Associated with surface energy measurements is the problem of the wetting of a clean solid by a liquid. A great deal of beautiful experimental work by Dr. Zisman has stimulated a more detailed analysis of wetting than that provided by the classical Gibbs equations. It is now clear from the work of Fowkes and Goode that the wetting of a solid by a liquid involves specific factors such as the ionic and Van der Waals contributions to the surface energies. For example, Fowkes suggests that the interaction of water with clean mercury is almost completely (over 95 percent) due to Van der Waals interactions at the interface, while the remainder arises from dipole: image-dipole: interactions. Even though this conclusion is difficult to believe, it is supported by experiments at the Bell Laboratories (ref. 1) showing that water does not wet clean gold. This raises another question: How clean are clean surfaces?

Structure of Clean Surfaces

Vacuum technology has improved to such an extent during the last decade that it is now possible to prepare surfaces that remain free of adsorbed gases for hours on end. It is generally assumed that if surfaces in a very high vacuum are thoroughly denuded by ion bombardment, or similar means, the surface must be clean and must remain clean for a prolonged period. To some extent this view has been fostered and indeed strengthened by low energy electron diffraction (LEED) studies. However recent research by Sewell (unpublished) using high energy electron diffraction, X-ray spectrometry, and mass spectrometry has shown that surfaces apparently clean according to LEED may, in fact, be covered over a large part of the surface by undetected contaminants. It is, of course, not the fault of the vacuum, but rather that of the specimen. Even the purest specimens contain slight impurities (say, one part in 10^7); thus when a specimen is ion-bombarded to remove contaminants, and then annealed to remove the damage produced by bombardment, impurities diffuse to the surface. In the future much greater care will have to be exercised in preparing clean pure surfaces and in categorically claiming that they are clean on LEED evidence. This is an experimental challenge of very great importance.*

There is also a theoretical problem of even greater importance which is, at last, attracting some attention. The diffraction of low-energy electrons by crystalline solids is not as simple as the early experimental work suggested. The interaction of electrons possessing energies com-

---

*Incidentally, the gold surfaces used in the Bell work were all heated to at least 250° C. If the specimen was then treated with concentrated nitric acid, it could be wetted by water. This was attributed to the formation of a wettable gold oxide film. Sewell has remarked (private communication) that concentrated nitric acid is an effective treatment for removing diffused surface contaminants.

parable with those within the solid is a complicated business. Current work suggests that subsidiary peaks in the diffracted intensities may occur; some of the super-lattice structures observed may arise from this or similar causes. We need far more understanding of the scattering of low-energy electrons by crystalline solids. The theoretician could find this a fruitful and challenging field.

"Real" Surfaces Resulting From the Interaction of Clean Surfaces With Liquids and Gases

Dr. Gatos has made a very pertinent point here. We do not often have to deal with really clean surfaces except perhaps in outer space. Ordinary surfaces are covered with films formed by oxidation or corrosion. Many of the simplest processes, even the initial growth of the surface film, are not fully understood. We know even less about the mechanical properties of these films, or the way in which their properties are affected by conditions of formation and growth, or the way in which these films can be ruptured either by pressure or sliding.

TOPOGRAPHY OF SOLID SURFACES AND SURFACE INTERACTIONS

Dr. Williamson's elegant survey of his recent work on surface topography marks a very valuable advance in our knowledge of the contour of real surfaces and of the way in which contact occurs when surfaces are pressed together. Dr. Ling's ingenious rendering shows how the analysis of thermal effects between sliding surfaces can yield information about the nature of interfacial contacts. There are three points here that I wish to raise.

Local Hardness

The first concerns the local hardness of solids when deformation is restricted to a very small volume. Some years ago Professor Milton Shaw carried out experiments on the fine machining of metals and suggested that when the dimensions of the deformed material are very small, the strength and hardness are very high. Recently Dr. Gane in our laboratory has studied the indentation of metals at very low loads (less than a fraction of a gram) and has indeed observed very large indentation hardness values. In some cases the hardness values approach the "theoretical" hardness for a perfect crystal. It may be that this is a pure size effect corresponding to the conditions that might arise if the deformed volume is too small to generate mobile dislocations. On the other hand it is possible that this is due to surface films of diffused impurity. Whatever the mechanism, these results suggest that over small regions of contact solids may be much more resistant to plastic deformation than their bulk hardness would suggest. Dr. Gane is pushing ahead with this work—it is clearly of very great importance.

Area of Contact

My second point concerns experimental methods used in determining the area of *true* contact between touching solids. This is a difficult problem and a challenge to the ingenuity of the experimental physicist. With metals, electrical resistance measurements have been used but they are of limited validity. With transparent solids optical methods have been applied but again these have their limitations. Dr. Williamson has shown that a detailed topographical study of two surfaces before and after they have been placed in contact can give a direct indication of the amount of true contact that has occurred. This is a very valuable new method, but it requires extremely fine measurements and location techniques: it is also limited to surfaces in stationary contact. Dr. Ling has shown how thermal measurements can be used to study the contact between sliding surfaces, but his method is not always practical, or possible, to apply. Is it not possible to derive a method of determining the area of real contact that will be simple, reliable, and applicable to both sliding and stationary surfaces?

Thermal Expansion and Wear

Dr. Ling has pointed out that local heating will produce normal displacements due to thermal expansion. In a multiasperity contact any individual asperity that becomes hotter than its neighbours will expand more and relieve the load from the nearby asperities. This will lead to a runaway situation where the individual hotspot gets hotter and hotter and carries a greater and greater proportion of the load. Clearly this could be an important factor in the wear mechanism. This should be studied further, both theoretically and experimentally.

FRICTION AND ADHESION

Tangential Stresses

As Dr. Merchant points out in his comprehensive survey paper, the friction between solids arises mainly from the need to shear the junctions formed by adhesion at the regions of real contact. He shows that junction growth occurs as a result of both normal and tangential stresses. One theory for this effect is that although asperities may deform elastically under static contact, a small amount of tangential stress can readily lead to plastic deformation. Consequently elastic deformation is possible when the tangential stresses are small, as in rolling contact or in lubricated sliding. Elastic deformation is very unlikely if there is strong interfacial adhesion. Tangential stresses, even below those necessary to produce gross sliding, may also play a very important role in three other fields: First, in metals they may lead to fatigue and fretting of surface asperities; second, with solids covered with hard surface films they may lead to rupture of the surface films; and third, with brittle solids they can be extremely effective in generating fracture in and below the sliding interface. Dr. Merchant has rightly stressed the important role of tangential stresses in sliding—it merits further emphasis.

Adhesion

The strength of the interface depends crucially on the adhesion between the surfaces. For two perfectly clean surfaces of the same metal the adhesion must necessarily be very strong and the interface will have a strength comparable to that of the metal itself. For dissimilar metals the adhesion is more difficult to specify. As Dr. Merchant points out, metal pairs which are mutually soluble will be of the right atomic size and electronic configuration to give strong adhesion (unless the compounds they form are basically brittle), and insoluble pairs should give poor adhesion. Such insoluble pairs, however, as gold-and-germanium and indium-and-diamond give strong adhesion. The work of Spalvins and Keller showed that there was poor adhesion between clean iron and silver when in static contact, and they suggested that such a pair of metals might be a very satisfactory combination for sliding in a very high vacuum. Some preliminary frictional studies by deGee seemed to support this conclusion, but later he found that the frictional and wear behavior were crucially dependent on small amounts of surface oxide. In his most recent work (private communication) he concludes that metals which are mutually soluble will always adhere well, while metals which are mutually insoluble will not adhere so well but they will usually show enough adhesion in sliding contact to give appreciable friction, metallic transfer, and wear.*

Two other adhesion factors Dr. Merchant points out, are released elastic stresses and crystal structure. A further factor is the crystallographic orientation of one surface relative to the other. If there is marked mismatch, the adhesion may be small. It is probably misleading, however, to think of this simply in terms of a mismatch in atomic bonding; for example, strong adhesion can occur at grain boundaries where mismatch between grains can be very marked. If adhesion between mismatching contacting crystals is small, it is probably the result of incompatible deformations of the surfaces as they are pressed together. In such circumstances, as Semenoff has suggested, the ad-

---

*At the time of writing this article I was not aware of the 1967 paper by Johnson and Keller in the *Journal of Applied Physics* in which Dr. Keller showed that strong adhesion may occur between metals even if they are mutually insoluble. The lack of adhesion which Spalvins and Keller observed in their 1963 paper is attributed to the presence of contaminant films. This implies that some combinations of materials are more susceptible to contamination than others since the older paper showed strong adhesion between mutually soluble metals, but not between mutually insoluble metals.

hesion can be greatly increased by providing extra energy at the surface, either thermally, or by the work of plastic deformation. Thermal activation resembles the sintering process; plastic activation, that of cold welding.

We see that adhesion, particularly in a sliding system, is not a simple issue of surface energies or mutual solubilities. Vestiges of surface oxide, released elastic stresses, and unfavorable crystallographic conditions can markedly reduce adhesion. On the other hand, sliding always involves some plastic deformation and some frictional heating, and both of these will increase adhesion.

Adhesion is of basic importance in friction. With lubricated surfaces it determines whether one pair of materials will scuff more readily than others. More information, both theoretical and empirical, is needed for the better understanding of adhesion.

The Deformation Term

If a hard asperity ploughs its way through the surface of a softer metal, the ploughing force can be easily calculated if adhesion between the surfaces is assumed negligible. A typical analysis for a conical asperity has been quoted by Dr. Merchant, and has been greatly elaborated by Wilman and his colleagues in their studies of abrasive wear. Their experiments show that the rate of metal removal by abrasion is inversely proportional to the hardness of the metal, a conclusion in close agreement with earlier results obtained by Krushchov. It is interesting to note that if we assume that ten percent of the volume of the ploughed grooves appears as removed metal, there is a very good quantative agreement between the calculated and observed abrasive wear rates. Mulhearn and Samuels have suggested that this can be explained in terms of the average orientation of the abrasive particles. Wilman has also studied, using electron diffraction, the orientation of surface layers produced by the abrasive process. This and later papers show that an oriented texture, similar to that produced in compression or rolling, is obtained.

Some very beautiful work has also been carried out by Bailey (ref. 2), and more recently by Courtel (ref. 3), on the deformation and ploughing of single metal crystals. The slip processes can, to some extent, be followed and explained in terms of dislocation movement and glide. But the most striking observation is that material piles up ahead of the slider in some directions, and on either side of the slider in other directions.

This type of material displacement and the simple ploughing model described by Dr. Merchant are relatively easy to understand. Indeed, if interfacial adhesion is negligible, the ploughing term can be calculated with a fair degree of reliability. But what if adhesion is not negligible? The ploughing term will react with the adhesion term and will probably lead to an increase in the deformation component itself. This type of interaction is probably very common in unlubricated sliding, yet there is no satisfactory theoretical solution. One approach might be in terms of the "model junctions" described some years ago by Greenwood and Tabor (see figure 1). Rheologists and plasticians could find this a most challenging area of study.

In most sliding situations the deformation term is small compared with the adhesion term. It can, however, be important in the abrasion of metals, as mentioned above; in the grinding and pulping of wood; in the sliding friction of polytetrafluorethylene where the adhesion component is generally small, and in lubricated friction of those materials for which the deformation losses may be relatively large.

The Deformation Term in Tire Friction

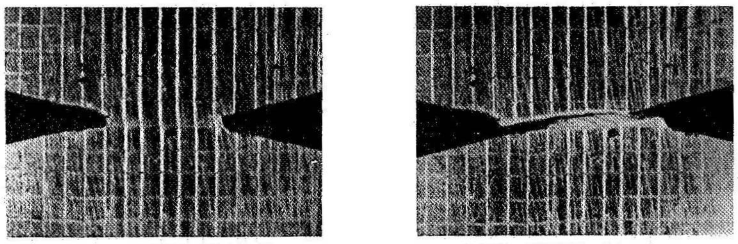
The deformation term can be particularly important in the friction of rubber. For example, if a hard asperity traverses a well-lubricated rubber surface, as much as 80 to 90 percent of the total friction may arise from the energy lost in deforming (or ploughing) the rubber. These losses are proportional to the hysteresis loss property of the rubber. Consequently, under conditions of poor adhesion, the larger the hysteresis losses in the rubber the greater the friction of hard sliders over the rubber. This concept has stimulated the use of high hysteresis rubber on automobile tires. If road surfaces are wet or greasy so that direct contact between the rubber and the road is prevented, the major part of the friction may well arise from the "ploughing" of the tire tread by the road asperities.

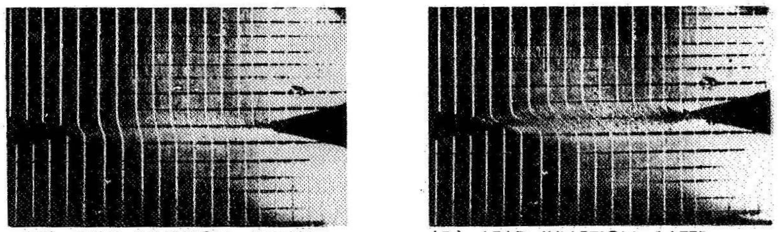
Here again problems are raised, this time by the interaction of the deformation or grooving mechanism with a small amount of adhesion.

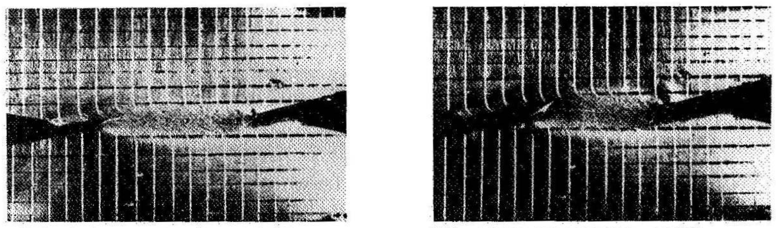
Intermittent Motion

None of the papers in this series has referred to intermittent motion. Yet it is of wide occurrence in many sliding mechanisms, often as a series of "sticks" and "slips." This is always associated with a system in which the static friction is higher than the kinetic friction, or where the kinetic friction decreases markedly with increasing speed. Most engineers have been content to describe intermittent motion in terms of the natural frequency of the system, its damping characteristics, its stiffness in relation to the magnitude of the frictional force, and the friction-velocity characteristics of the system.

Another way of looking at this problem is to recognize that there is a time effect such that the longer the surfaces are in contact, the higher the static friction. Rabinowicz has presented an interesting treatment


(A) HARD COPPER JUNCTION: INITIAL YIELDING.
(B) HARD COPPER JUNCTION: FRACTURE.


(C) LEAD JUNCTION: INITIAL YIELDING
(D) LEAD JUNCTION: LATER


(E) LEAD JUNCTION: LATER
(F) LEAD JUNCTION: LATER

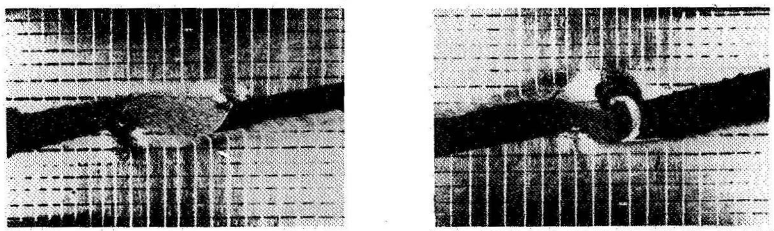

(G) LEAD JUNCTION: LATER.
(H) LEAD JUNCTION: FINAL STAGE

FIGURE 1.—Model junctions.

in these terms and has shown how this explains the main characteristics of stick-slip motion.

Although this approach is very satisfactory in a descriptive way, the fundamental question remains as to why static friction increases with time of contact. It may be attributable to creep as a result of which the area of contact increases with loading time, or there may be a time factor involved in the breakdown of contaminant surface films. It is also possible that diffusion across the interface leads to a strengthening of junctions. Again Courtel (ref. 3) suggests that the interaction of the ploughing and adhesion term and the formation of a "frontal bulge" may lead to an instability producing intermittent motion. Finally, in a recent publication Tolstoi (ref. 4) has shown that stick-slip motion in a horizontal direction is associated with a slight vertical movement of one surface relative to the other. If this vertical motion is suppressed, the intermittent motion is reduced or eliminated. This is a very important observation; unfortunately, Tolstoi does not discuss how this vertical movement is produced in terms of the formation and shearing of the interfacial junctions. It may well be that the model-junction study by Greenwood and Tabor is relevant here. In view of the basic importance of avoiding stick-slip motion in most engineering mechanisms, it is surprising that more fundamental work on this has not been done.

There are many mechanical systems in which intermittent motion is the result of poor design rather than actual friction characteristics. This can occur in poorly designed brakes. Spurr has given an example of this in his account of a "spragging," or jamming mechanism.

Friction in Very High Vacuum

A large amount of interesting work has been done by Bisson and Anderson and by Buckley and Johnson in the NASA laboratories, on the friction and wear of metals in a vacuum of 10^{-10} torr. The system is baked out at 200° C, and the surfaces are repeatedly slid over one another for periods on the order of one hour. It was found that certain alloys and metals give relatively low friction and wear even under these severe conditions. For example, a titanium alloy containing 21 percent aluminum, sliding on a stainless steel surface, gives a friction coefficient of only about 0.4. On the other hand, pure titanium sliding on the same surface results in gross seizure ($f>1.6$). The authors suggest that this is connected with the structure of the metal or alloy and particularly with a favorable ratio of the c to a dimensions in hexagonal structures; a large c to a ratio yields low friction and wear. Dr. Merchant has given a simple and plausible explanation of this in his survey paper. There is room here for further experimental and theoretical study, particularly in view of the connection between these results and the behavior of lamellar solids.

The effect of surface film removal by sliding in high vacuum is clearly of great importance in space applications. One recent result of interest, observed in our laboratory, concerns the behavior of diamond surfaces. These are difficult to clean by heating in a vacuum because prolonged heating causes graphitization; in fact with graphitization the coefficient of friction does not exceed about 0.4. If the diamond surface is repeatedly traversed in a vacuum of 10^{-10} torr at slow sliding speeds, so that frictional heating is negligible and graphitization does not occur, the friction coefficient rises to a value of about 1. In addition the strong adhesion and the increased tangential stress at the interface lead to catastrophic fragmentation of the diamond.

Friction of Nonmetals—Brittle Solids

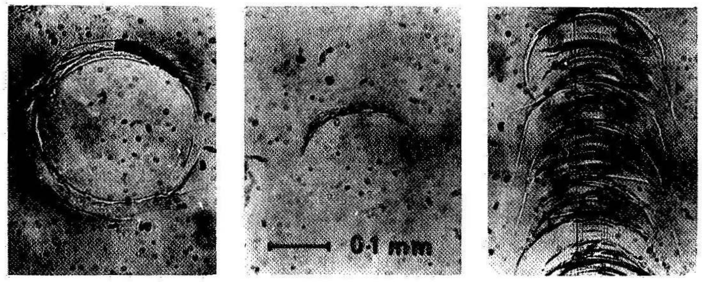
Dr. Merchant's survey has shown why the friction of brittle solids resembles that of metals. One important difference, though, arises from the limited ductility of the material which in turn restricts junction growth. Another important difference is the way single crystals of brittle solids can manifest marked frictional anisotropy. This is shown most simply by studying the friction of hard sliders on the surfaces of single crystals. When the deformation is gentle, the behavior resembles that of ductile materials and metals; the higher friction being associated with the piling up of a hill of work-hardened material ahead of the slider. When the deformation is predominantly brittle, frictional anisotropy again occurs and is due largely to an increase in the ploughing term. However, the high friction observed in certain crystallographic directions is due not so much to a piling up of material as to an increase in penetration by the slider into the brittle crystal (ref. 5). In addition the cracking around the slider is often caused by a pile-up of the dislocations associated with plastic displacement of material. This mechanism may also account for the frictional anisotropy observed earlier on diamond, but the evidence is by no means conclusive. In view of the importance of diamond as a "low friction" material and its use as a cutting and grinding tool, more work in this area would be of value.

Although friction with brittle solids appears to be due to both adhesion and ploughing mechanisms, wear is usually dominated by cracking and fragmentation rather than adhesion and transfer. In this process, tangential stresses are of primary importance. This may be illustrated by considering the behavior of a noncrystalline material such as glass where neither flow nor brittle failure is associated with specific planes or directions. Figure 2 (A) shows the appearance of a glass surface after a hard spherical indenter has been pressed against it with a load of nearly 7 kg. The static load is just sufficient to produce a ring crack around the Hertzian area of contact. If the normal load is reduced to 2 kg, no such cracking occurs. However, if now a tangential force of only 1 kg is applied (insufficient to produce sliding), the tangential stress generated in the surface layers produces cracking over an arc lying just behind the contact area (fig. 2(B)). At slightly higher tangential forces, where gross sliding occurs (fig. 2(C)), typical arc cracks are observed over the whole friction track (ref. 6). These can naturally be the cause of very heavy wear. The interaction of normal and tangential stresses in the failure of brittle solids is a very interesting and challenging theoretical problem; two independent treatments, by Hamilton and Goodman and by Frank and Lawn, have recently appeared.

Friction of Nonmetals—Lamellar Solids

The low friction of lamellar solids is generally attributed to the marked anisotropy of their strength properties: strong in compression but weak in shear. With graphite the friction and wear are high unless adsorbed water vapor or oxygen is present. The earlier work suggested that the cleavage faces of graphite are low energy surfaces, the edges, high energy surfaces. On this view, the adsorbed vapors are effective because they reduce the surface energy and bonding strength of the edges of the graphite crystallites. Recent work by Bryant, however, suggests that the cleavage face of graphite is a high energy surface. If this is so, the amount of adsorbed vapor necessary to reduce the friction and wear, would be much larger than has been observed. A critical study of the friction of graphite is now overdue.

With molybdenum disulphide there is more general agreement. The low friction is essentially a property of the structure of the material, molybdenum disulphide being extremely weak along the shear planes. The presence of adsorbed films and other impurities tends, if anything, to increase friction.



W = 7, F = 0
STATIONARY
(A)

W = 2, F = 0.5
STATIONARY
(B)

W = 2, F = 1
SLIP
(C)

FIGURE 2.—Deformation of glass by a hard spherical indenter (0.63 cm diameter). In (B) and (C) the slider is being pulled in the downward direction relative to the glass.

One interesting research development is the effect of crystal structure on the frictional properties of lamellar solids of the molybdenum disulphide type (chalcogenides). Some evidence suggests that, as with the hexagonal metals and alloys studied by Buckley and Johnson, a large *c* to *a* ratio is advantageous. Other experiments, however, suggest that the effectiveness of lamellar compounds depends mainly on the extent to which they are stoichiometric; the more perfect their structure the lower their friction. Probably the most important single factor is the type of bonding between the lamellae. This has not been fully explored and deserves further study.

A very promising line of research of both academic and practical interest is the possible use of chemical attack to form, in situ, surface films possessing a lamellar structure. This was first described in detail by Rowe in 1957. He showed, for example, that titanium could be effectively lubricated by forming titanium iodide on the surface; similarly, tungsten can be lubricated by forming tungsten chloride. This has recently led to a number of practical applications; for example, Roberts and Owens have found that a lubricant containing an activated iodine complex can provide very effective lubrication for titanium (and chromium). There is room for more work in this area.

Friction of Nonmetals—Rubber and Rigid Polymers

Dr. Merchant has pointed out that the friction of polymeric solids differs from that of metals in that the deformations are elastic or, more correctly, viscoelastic. In many situations this can lead to a dependence of the coefficients of friction depending on load and geometry so that Amontons's laws are not obeyed. There is a further, more important, difference. The deformation and strength properties of polymeric materials are markedly dependent on temperature and frequency of deformation. One would therefore expect to find a corresponding dependence of friction on temperature and speed of sliding. This has been well demonstrated for rubber by Grosch (ref. 7). He studied the friction over a wide range of temperatures (-15 to $+85°$ C) and speeds (10^{-4} to 1 cm/sec.). (By restricting his sliding speeds to less than 1 cm/sec, he was able to avoid complications due to frictional heating.) His results show that the friction, as a function of speed and temperature, can be fitted to a "master" friction curve, by using the same Williams-Landel-Ferry transform that correlates the viscoelastic properties of the rubber. At room temperature the friction reaches its peak at a sliding speed, v, of about 1 cm/sec. At this temperature the peak in the viscoelastic losses occurs at a frequency, ν, of about 10^6 cps. If the friction and viscoelastic losses are to be attributed to the same mechanism these results imply that, in the sliding process, there is a significant distance, λ, given by $\lambda = v/\mu$. This quantity has the value of about 100Å which is approximately the length of a segment of the rubber molecules which other types of experiment show to be involved in the relaxation process.

One of the characteristic features of rubber friction is that, over an appreciable range of conditions, there appears to be no transfer of rubber to the other surface; that is, sliding appears to occur truly at the interface. This suggests that the rubber molecules are temporarily attached to and then detached from the other surface. This has led Schallamach (ref. 8) to develop a molecular theory of rubber friction involving three variables: bond formation between the rubber molecule and the other surface; dwell time; and stress-activated detachment of the molecule. The treatment is essentially an extension of the Eyring molecular theory of viscous flow and gives reasonably good qualitative agreement with experiment. However, there is room for a more critical analytical model.

With rigid polymers such as perspex, polyethylene, PVC, and nylon, it has not been possible to obtain a similar clear-cut correlation between friction and viscoelastic properties. Ludema and Tabor (ref. 9) have suggested that this is because conventional study of viscoelasticity is carried out at very small strain levels and at ambient pressure. By contrast, the material around the friction junctions is subjected to a very high local pressure and the shear strains, before sliding occurs are enormous. Thus although sliding friction often varies markedly with speed and temperature, there is no reason why this should reflect viscoelastic properties measured under much gentler conditions. There is room here for a study of viscoelastic properties at high pressures and for large strains, but professional rheologists do not seem very keen to tackle this area of work.

With polymers adhesion at the interface is generally strong; shearing occurs within the polymer itself, and there is marked material transfer. Under carefully controlled conditions strong interfacial adhesion can occur even with PTFE, though the friction remains small. The low friction must be attributed to the anisotropic yield properties of the crystallites within the PTFE. The details are not certain, but it is interesting to note that PTFE of low crystallinity gives a friction that is generally 10 to 15 percent higher than that observed with highly crystalline material. It is, of course, true that in many practical situations the adhesion between PTFE and other surfaces is small, and this is accompanied by a further small but significant drop in friction. Nevertheless, the friction experiments suggest that more attention should be paid to the influence of polymer structure on frictional behavior.

Effect of Speed, Temperature, and Environment

Sliding friction depends on environment, on the nature of surface films present or formed during sliding, on temperature, and on speed. At moderate speeds, the friction tends to decrease as speed increases, probably because the junctions are weaker if they are formed (and destroyed) in a very short time. However, there is no satisfactory mechanism for this and, as mentioned above, it merits further study.

At higher sliding speeds, the main effect is frictional heating. With metals this produces a thin film of soft or molten metal at the interface and results in lower friction. If large-scale melting occurs the friction again rises because of the large increase in the area of contact. The rate of metal removal is determined by the rate at which the molten front diffuses into the metal ahead of it. With nonmetals there is a similar decrease in friction with increasing speed. With brittle materials thermal shock may lead to fragmentation and excessive wear. With polymers, on the other hand, the friction and wear remain low even at the highest speeds. This is because a considerable part of the frictional energy may be expended in raising the temperature of the molten interface rather than expanding the molten zone. In addition, there is some evidence that decomposition of the polymer may provide some measure of gas lubrication, but more work on this is necessary. The most striking result is that at sliding speeds of 2000 ft/sec (600 m/sec) the wear of nylon may be less than that of copper or aluminium and scarcely greater than that of molybdenum.

The high flash temperatures and rapid cooling that takes place as soon as the hot spot passes out of the contact region, can lead to important phase changes. For example, Welsh has shown that with ferrous materials, severe conditions can produce a martensitic surface layer which, because of its extreme hardness, can lead to a marked reduction in wear.

The effect on friction of extremely high or low ambient temperatures has received considerable impetus from nuclear power developments and space research. Most of the fundamental studies have been carried out in a vacuum or inert gas to prevent oxidation at the high temperatures, or condensation at the low. With pure metals the main effect of high temperature is to increase the ductility of the material at the interface. This greatly facilitates junction growth and readily leads to gross seizure. For this reason the use of pure metals and even steels is impractical above about 1000° C.

In this higher temperature range, the materials mainly studied have been the carbides, borides, and nitrides of the transition metals. Their friction at room temperature, even when clean, is relatively low, probably because of their low ductility. Above 1200° C, however, a marked increase in volume diffusion takes place, sintering at the interface is facilitated, friction rises, and surface damage increases greatly. Environmental factors may aid sliding surfaces by the in situ formation of a suitable lattice-layer (lamellar) compound as mentioned above. Another fruitful approach is that of Buckley and Johnson who have fabricated composite materials that can operate successfully at very high temperatures. Some of these materials can also function at only a few degrees absolute. At a more fundamental level it would be interesting to explore the effect of low-temperature embrittlement on friction; a decrease in ductility should reduce friction, although, if the substance became too brittle, wear might increase catastrophically.

WEAR

A recent survey of industrial research in Britain shows that in the general fields of friction, lubrication, and wear; by far the greatest effort is expended on wear testing. This reflects the fact that it is difficult, if not impossible, to anticipate the wear behavior of a particular pair of solids from more fundamental studies. The reasons for this are evident from the interesting survey that Dr. Archard has presented at this symposium. There are several wear mechanisms which are fairly well understood in themselves, but in any particular sliding mechanism, the way in which they interact is almost unpredictable.

In this connection I should like to quote from a draft report by the Organization for Economic Cooperation and Development (OECD) Group on Wear of Engineering Materials. It describes a study of wear results obtained in different laboratories using the same materials. The relative wear rates of copper and bronze against steel, measured at a load of 10 kg, gave values which varied from 0.5 to 95; that is to say, using identical materials different workers obtained results which differed by a factor of nearly 200.

The authors (A. Begelinger and A. W. J. de Gee) conclude that if only one wear mechanism occurs, reasonable reproducibility may be obtained. However, "seemingly minor factors, amongst which the thermal conditions at the friction interface are the most important, may have a considerable influence on the type of wear mechanism that predominates, or rather on the relative importance of the different mechanisms that may be operative successively or simultaneously."

A few of the wear mechanisms will be considered below.

Adhesion Mechanism

First there is the adhesion mechanism which occurs when asperities break through the protective lubricant or oxide film and make contact. Junctions are formed, and when these are broken, particles are transferred from one surface to the other. The detailed behavior depends on the running conditions, particularly the load, speed, and temperature, since these can lead to changes in the strength properties of the surface layers as well as to phase changes in the underlying material. Again the adhesion may depend critically on such parameters as mutual solubility, relative orientation, and surface energy properties. Some of these factors have already been discussed in the preceding sections. At a later stage the transferred fragments become detached and then become true wear particles. The mechanism by which the transferred fragments are freed is still the subject of discussion. Rabinowicz has made a most useful contribution in suggesting that the fragments are released when the stored energy (resulting from repeated deformation) exceeds the necessary surface energy. The first difficulty here is to make the concept quantitative. We do not really know the interfacial energies of solids. Secondly, most wear fragments are such mashed-up particles of metal, oxide, and other contaminants that we do not know what interfacial energy to choose. Thirdly, there is some uncertainty as to the relation between strength properties and surface energy for ductile solids (even when they are highly worked), since a small amount of plastic deformation can easily swamp the surface energy term. There is a need here for a basic study exploring Rabinowicz's mechanism in greater detail. Another approach is to avoid detailed mechanisms and to describe the detachment of wear particles as a fatigue process.

Oxidation and Surface Films

A second important mechanism is surface oxidation. Oxides generally reduce adhesion but the way in which they are penetrated and broken down is not properly understood. Osias and Tripp have studied the deformation of a plasticine sphere covered with a thin film of lacquer, and have suggested that this resembles the breakdown of a hard oxide on a softer metal; here the cracking of the film plays a very important part. There is little doubt that in the real sliding process tangential tractions on a brittle surface layer will greatly facilitate cracking and fracture (fig. 2). This raises a point of theoretical and practical importance: How will a ductile oxide film behave? In general a crystalline oxide will be more ductile than an amorphous one. Does this mean that a high-temperature oxide formed on aluminium will give better protection than the oxide normally formed at room temperature?

The mechanical properties of hard protective surface films are probably of critical importance. Hard plasma-sprayed coatings which are able to withstand the stresses in the sliding system can greatly reduce wear. If, however, they are too brittle and are fragmented by the sliding process, the detached fragments can enormously increase wear by abrasion.

Abrasive Wear

Abrasive wear has been discussed briefly in a previous section. Although it is probably the main cause of material removal in many sliding mechanisms, it is not generally a serious problem unless the designer has been very unwise in his choice of materials. A far more serious situation is sudden seizure. This type of scuffing or galling is often observed in well-lubricated systems. One suggestion is that it may be caused by hard particles, larger than the oil-film thickness, which somehow get trapped between the sliding surfaces. These particles may be carried by the oil supply from one part of the machine to another and may escape the most stringent filtration procedures. On the other hand, they may be produced by the sliding process itself. For example, repeated asperity contacts (even if they are protected by an unbroken film of lubricant) may gradually cause fatigue resulting in a fragment. For this reason the correct choice of materials is of great importance even in well-lubricated mechanisms if scuffing and seizure are to be avoided.

There have been a number of attempts to derive from short-time experiments wear equations that can be extrapolated to longer periods. Some of these equations (refs. 10 and 11) are based on reasonable physical models, and they work surprisingly well. This is largely because they are restricted to specific types of wear mechanism. If for any reason the balance of wear mechanisms should change, the equations would probably be misleading. In some ways this treatment resembles the prediction of long-term creep from short-term experiments.

Dr. Archard referred to wear as a Cinderella.* I would refer to it as a wayward, capricious child of mixed and rather uncertain parentage. In the basic study of wear, metallurgists and material scientists have not (with some notable exceptions such as the group at NASA) played the part they could in the selection and development of materials for sliding pairs. There is particular scope for thoughtful consideration of composition, structural factors and of types of surface films that can be formed.

BOUNDARY LUBRICATION

The survey by Mr. Godfrey follows so closely my own thinking on the subject of boundary lubrication that there is little point in repeating his views in detail here. There are, however, a few concepts that I would emphasize in a slightly different form.

Boundary and Elastohydrodynamic Lubrication

First is the issue of boundary lubrication and elastohydrodynamic lubrication. We now know that when mineral oils are sub-

---

*This expression was used in the original draft of Dr. Archard's manuscript, not in the final version appearing in this volume.

jected to very high pressures, they undergo a prodigious increase in viscosity. As a result the oil film may remain trapped between the solid surfaces and prevent metal-metal contact. For example, with very carefully aligned steel surfaces operating at contact pressures on the order of 50 to 100 kg/mm^2 (so that the pressures are not far from those capable of producing plastic flow of asperities), classical elastohydrodynamic lubrication can be obtained with a simple mineral oil, even though the coefficient of friction may be on the order of 0.06 to 0.08 (K. L. Johnson, private communication). We have recently found similar effects even at relatively slow sliding speeds. This is not boundary lubrication and some workers have gone as far as to suggest that boundary lubrication does not even exist.

The reality of boundary lubrication, however, is revealed if experiments are carried out with mineral oils containing small quantities of surface active materials. For example, Naylor (private communication) studied the behavior of a typical disc machine in the presence of a pure mineral oil; he has been able to operate the system under ideal ehd conditions with a limiting film thickness of less than 200Å. Adding a small quantity of oleic acid to the oil did not affect the film thickness; furthermore, there was no change in the viscous resistance of the film. On the other hand, in a practical cam and tappet rig operating at similar pressures and speeds, it was found that with the oil containing the small quantity of oleic acid the load at which scuffing occurs is much higher than with the pure mineral oil. We may conclude that a very thin film, perhaps only one molecule thick, of oleic acid is adsorbed on the metal surfaces: It plays no part in the bulk viscous properties of the system but provides protection to the surfaces when the ehd film breaks down. It is evident that we have returned to a type of boundary lubrication very similar to that described in the earlier classical work. We may define it as that type of lubrication which cannot be attributed to the bulk viscous properties of the lubricant (whether the system is operating under hydrodynamic or elastohydrodynamic conditions), but arises from a specific solid-lubricant interaction.

Cooperative Effects in Boundary Lubrication

Mr. Godfrey has indicated that dissolved air in a mineral oil can greatly reduce wear between heavily loaded surfaces. It is probable, as his own experiments and those of Vinogradov suggest, that oxygen reacts with the wearing surfaces to form a protective oxide film. There is also evidence that if long-chain fatty acids are dissolved in a mineral oil, their effectiveness as boundary additives is greatly reduced if the oil is completely de-aerated. It may be that the fatty acid does not react effectively with the naked metal but does so much more readily with the metal oxide. On the other hand, it is possible that a single monolayer of an adsorbed fatty acid monolayer is not sufficiently protective; it may need the additional protection of a thin oxide film. This is worth exploring further especially since it suggests that noble metals will not be easily lubricated by adsorbed polar molecules alone. This may explain why gold can only be lubricated effectively if halogenated materials are also present.

Exo-Electrons

Little mention has been made in this symposium of an area of work which appeared very promising some ten years ago; namely, the study of the electrical charges liberated from freshly formed surfaces when exposed to appropriate atmospheres (e.g., exo-electrons, the Kramer effect). This work has not been continued in its original form, and merits a 'new look.' In a related area, very valuable work has been carried out on the high reactivity of fresh metal surfaces to organic compounds. As Mr. Godfrey points out, this may be of great importance to lubricant performance, especially under severe conditions of load and speed. No account is given here, though, of the influence of surface-active materials on the mechanical strength of solids. This may prove to be of importance in the machining and cutting of metals as well as in Rehbinder's classical field of rock-drilling.

"Weeping" Lubrication

Mr. Godfrey hinted at two lines of recent lubrication research; they merit further attention. The first is the observation of McCutchen (ref. 12) that porous materials may extrude lubricant when subjected to pressure, thus providing a 'weeping' type of lubrication. He has suggested that this may be one of the ways in which the cartilage and synovial fluid lubricate bone joints. The second is somewhat similar but operates under much higher pressures. If a metal surface is lightly shot-peened before lubrication, it can be subjected to very heavy deformation in subsequent rolling or drawing without the occurrence of scuffing; the lubricant is squeezed out of the cavities during the metal-working operation and prevents metal-to-metal contact (ref. 13). In this way even titanium may be successfully worked. It would not be inappropriate to refer to this as plastohydrodynamic lubrication.

Areas of Needed Research

I cannot do better than repeat the final remarks of Mr. Godfrey. In this field we need:

(1) Knowledge of the chemical composition of films on sliding surfaces

(2) Data on the shear strength of film materials under conditions of high pressure, high shear rate and temperature

(3) Data on the melting point of film material under conditions of high pressure and high shear rate

(4) Examination of films before and after sliding to indicate the location of shear planes and the importance of adhesion and cohesion

GENERAL OBSERVATIONS

Any simpleton can measure friction—a great many of us do. Accurate knowledge of frictional coefficients are fundamental to engineering. For example, in the action of brakes where a high steady friction is required; in the operation of an automobile door-lock where a lady's finger must be strong enough to actuate a very ingenious and powerful mechanism; in a gyroscope where extremely low reproducible frictions are needed; and, at the other end of the scale, in enormous power generators where a small diminution in friction can lead to appreciable economic gains.

The measurement of wear, which is so widespread in industrial laboratories, is of immense practical importance in estimating the reliability and viability of a particular sliding mechanism.

The Challenge to the Scientist

Nevertheless it is, I think, true that for the scientist the measurement of friction and wear is a relatively simple, almost trivial, operation. The problem, though, is in trying to understand how friction and wear arise, and, once known, how they can be controlled. For this purpose the chemist, physicist, metallurgist, and materials scientist must bring to bear all their most powerful techniques in order to understand how surfaces interact when in close static and dynamic proximity. Optical microscopy, optical interference, electrical measurements, and other techniques are needed to determine the area of contact. Also needed are electron microscopy and probes to study surface deformation and material transfer during sliding. Electron diffraction can yield the structure and composition of surfaces and surface films. Radioactive tracers can help study wear and transfer, and mass spectroscopy can indicate surface reactions. And in addition to all this, they must match the technique they use to the scale of the problem they are investigating.

Scientists must also use all the scientific concepts that can be applied to the explanation of their observations. For example, the physicist must know how solids deform elastically, how plastic flow occurs, how contact stresses can produce slip in ductile solids and cracking in brittle solids. He needs to know how surfaces react with environments and how surface films are formed. But he must also be realistic. For example, if he studies the sliding of single crystals he may be able to explain the gross effects in terms of plasticity theory; and, if he is lucky, he may be able to explain some of the detailed processes in terms of dislocation theory. But if he describes the frictional process in terms of wave mechanics, it is probable that though he may understand wave mechanics he will not achieve an understanding of friction.

The Need for a Cooperative Effort

In all this work the scientist needs the tools and the techniques of his profession. But above all he needs scientific insight and a sense of reality. Without these attributes he will amass expensive equipment, but meaningless data. In the long run, therefore, the contribution of the scientist to our understanding of the sliding process, both with and without surface films, will depend to a very large extent on the maturity of outlook and quality of mind he brings to bear.

These comments can be summarized by saying that the most effective contribution to tribology will be from the collective effort of the chemist, physicist, materials scientist, metallurgist, and engineer. In this work they will all need refined equipment. But, taking the longer view, men and ideas may well prove to be more important than machines and computers.

DISCUSSIONS

R. L. Adamczak (Air Force Materials Laboratory, Wright-Patterson Air Force Base, Ohio)

The authors and discussers thus far have done an excellent job of outlining further research needs. All I shall do, therefore, is to use some of the material discussed to develop several basic points.

The first point is that the field of friction and wear, lubrication, or tribology (whichever one wishes to call it) is a sorely disjointed science. There has been very little that actually relates to or ties each of these reviews together. Thus the current and most apparent need is for a unified, and unifying theory.

The second point is that of treating this entire subject of lubrication as an interdisciplinary science. In reality it is a scientific field onto itself and should be considered as such both academically and professionally.

These concepts assume that lubrication is a disjointed science which must be put together. I shall expand on this using the texts of the other presentations at this meeting. Reviewing the papers I found that terms, phrases, and concepts used in any one presentation, are ambiguous and/or differently defined in others. Some ideas have almost no consistency of definition.

Professor Gatos stated that, "Reliable experimental values of surface tension should prove of incalculable value to all theoretical efforts on solid surfaces." Let us presume that surface tension can be precisely determined, then what would be the implications on the theory of surfaces? My question, of course, is not directed at the authors, but at the audience and those working in our field. Even Dr. Tabor brings to bear his concepts of surface energy and surface tension.

Dr. Gatos discussed, and Dr. Tabor amplified, the subject of "clean" and "real" surfaces. I feel that the quotes are wrongly placed and should be on the word "surface," since any surface which is not "clean" is "real." This whole area of surface structures and interactions seems to be plagued with the inability to define a surface. Granted a number of surface types have been suggested as a result of numerous investigations, but even these cannot be related.

Dr. Williamson's treatment, as are all those presented at this meeting, is measured in degrees of elegance. He mentions that microinch resolution is quite adequate for the study of surface topography, and that more refined techniques will be of greater value only to physicists rather than to the surface (or lubrication) scientists. This of course is interesting since in the preceding paper, microscopy and diffraction techniques are used for resolutions of 2 to 20Å, whereas other papers speak of resolutions that are visible to the eye. Here, as with all parameters, we are attempting to measure finer or more accurately; but in most cases it is to find, rather than to prove or disprove, a theory. The concept of size or dimension is often inconsistent and sometimes not even applicable to the treatment of the problem.

Professor Ling has contributed a sound mathematical, and experimentally verified, approach for the determination of surface temperatures. If we assume that the two contacting surfaces are atomically flat, is this interfacial temperature that of the surface molecular layer, or of the last 100Å, or of the last microinch? Since surfaces do have asperities, a wide range of contact geometry can be expected. Thus we can assume that there would be little temperature rise where there is only minimal contact, and maximum temperature rise at points of greatest asperity contact. If we assume no deformation, can this maximum temperature be accurately determined? Since it is fairly well accepted that frictional heating is the result of junction breaking and deformation, can the deformation term be exploited as to its maximum temperature contribution? Such reasoning might yield bounds within which temperature effects would be most reasonably expected.

In Dr. Merchant's paper, one clearly sees the problem of semantics, interrelationships, and disciplinary concepts. In addition, there are ill-defined terms such as "chemically clean surfaces," "intimate contact," and especially "true area of contact." In this review there is also a vacillation between micro- and macroscale. In a way this reflects the academic background of the researcher.

To adumbrate Dr. Archard in his introduction, we have gleaned from this meeting an overview of the present state of the art. I am sure that if this conference were to fully relate our problems, at least several weeks more would be required. But let this not hang heavy on our conscience for our problems provide us and our successors with a living. Wear seems to be the culmination of everything we study in the field of lubrication; and yet, it is predicated on all the phenomena, theories, and work covered in this symposium. Simply, if we cannot predict the nature of what precedes wear—its uncertain percentage—then obviously wear must remain elusive.

Mr. Godfrey's presentation of boundary lubrication indicates the lack of contributions from the field of chemistry. Only in the last 15 years have specialized disciplines contributed usefully to the field of lubrication. This accounts for the recent advances in hydrodynamics and elastohydrodynamics. But boundary lubrication still suffers for lack of interdisciplinary exposure.

In spite of this, I feel that lubrication and all of its ramifications should not be considered an interdisciplinary science, but rather as a distinct scientific field both academically and professionally.

I thank our European friends and especially people such as Mr. Jost and Professor Blok for fostering and propagating the concept of tribology. As definitive tribological theories are developed and useful textbooks are written, independent tribology departments will develop, with subdepartments eventually emerging, as in the other sciences. I believe that this measure of respectability is as sorely needed as the need for a unified theory of lubrication.

H. Gisser (Pitman-Dunn Research Laboratories, Frankford Arsenal, Philadelphia, Pennsylvania)

It is at once evident from Dr. Tabor's paper that lubrication research, whatever its other attributes, is and of necessity must be interdisciplinary in nature. At the risk of repeating what is axiomatic, long range success in efficiently solving lubrication problems requires knowledge of the fundamentals of the processes and materials involved. That this has received recognition in the recent past is evident by the extensive scope of current research activities which are part of and contribute to lubrication science. Dr. Tabor has done a masterful job of appraising the current status, and pointing out the general direction for further lubrication research. This is achieved by his usual economy of language and intuition for those items and areas which would most probably bring fruitful and valuable findings. His summary in "the challenge to the scientist" is a useful general guide and approach to anyone probing the nature of the fundamentals of lubrication.

I should like to comment on several points concerning boundary lubrication—that is, lubrication where the chemical nature of the lubricant film is a major determinant of lubrication characteristics.

There has been an extensive empirical study of the interaction of organic materials with metal surfaces, together with, in many cases, indentification or at least indication of the nature of the products. Some of this work has emphasized the role of surface oxide, particularly where fairly reactive organic lubricants are of interest. There is good reason to expect that the mechanism of the reactions with oxide, and the kinetics of such reactions, are of importance. The previous researchers did not in general cover nonpolar compounds except in relatively extreme conditions under which fairly extensive oxidation leads to the formation of reactive or polar compounds.

The development of electron spin resonance spectrometry during the past decade has stimulated extensive studies on oxide surfaces and, by providing a means of identifying the nature of reactive sites on the surface, has made accessible a more detailed study of interactions of oxides with organic compounds. Very recently we have shown, using ESR (ref. 14), that pure aliphatic hydrocarbons react chemically with zinc oxide surfaces, thus changing the nature of the reactive species. Since the existence of oxides, and very probably their structure, is of importance in rubbing phenomena with metals, further exploration of the nature of these oxides and their reactivity, is germane. A body of knowledge already exists on the study of these materials as "semiconductors", i.e., surface species with an excess or deficiency of electric charges. The detailed study of such oxides lends itself to ESR techniques; further studies including mechanism and kinetics of reaction with environmental media, such as reactive gases together with organic compounds, are needed.

Further exploration of the nature of lubricant films in the boundary region is needed. None will doubt the role of solid films in such lubrication; for example, the failure of soap films above their melting point has been reported more than once, and the activity of organic phosphates has been shown to be due to inorganic phosphates on the surface. The empirical conclusion that a solid film is required for boundary lubrication has recently been shown to extend to nonpolar materials, e.g., hydrocarbons, even under relatively mild conditions in the absence of reactive atmospheric gases. The nature of the lubricating film with nonpolar materials in the presence of atmospheric gases is of interest. While oxidation offers a ready explanation, and indeed has been demonstrated to take place at elevated temperatures (which may occur locally because of rubbing), the reactions leading to the film and the nature of the film formed, even under mild conditions, are still to be demonstrated.

Since metal surfaces are oxide layers, and since lubrication failure means metal wear or destruction, then the behavior of the oxide surface, and the chemical and physical mechanisms of its formation and destruction during the rubbing process, need exploration. Here ESR spectrometry and both high and low energy electron diffraction should be of value.

Interesting new techniques have recently become available for study of thin films. For example, there are ellipsometry (ref. 15) which permits study of the thickness and optical constants, and reflectance spectrometry (ref. 16) which permits study of chemical structure. These techniques, when combined with modern high-vacuum technology, may make available in greater detail the physics and chemistry of monolayer-substrate interactions during rubbing. Ellipsometry and reflectance spectrometry provide means of studying, in situ, films adsorbed from solvent. Approaches using thin films are particularly attractive because the experimental systems are in principle simpler than when relatively large bulks of material are under study; the pertinent parameters are under more precise control. As a consequence the pertinent phenomena may be more readily "sorted out" and the data more readily related to fundamentals.

In his summary Dr. Tabor mentioned approaches and techniques which should be explored. But I believe that a key point of his presentation, and one in which I believe should be stressed, is that in the long run the men and their research ideas may prove to be more important than their equipment.

R. L. Johnson (NASA Lewis Research Center, Cleveland, Ohio)

Throughout the papers of this symposium, the critical appraisal of Dr. Tabor, and the earlier discussions, there are repeated references to phenomenological considerations; for example, elastic modulus, various mechanical strengths, thermal properties, and surface energies. Very often these are interrelated phenomena that make it possible to analyze data (of limited scope) using any one of several models and achieve reasonable numerical correlations. For example, surface energy correlations use numbers derived from hardness measurements (ref. 17), so we may wonder if the result is a surface energy or a hardness effect. The facts are that we do not have an adequate definition of hardness so that analysis on the basis of either hardness or surface energy is representative of the empiricism that must be set aside if real scientific progress is to be achieved. The suggestion of Dr. Tabor for a new approach to determining surface energy is also open to question because "non-creep equilibrium" is virtually unobtainable. I agree with Dr. Archard to the effect that, in general, the surface energy concept is interesting but of questionable usefulness.

The continuum mechanics approach of Dr. Ling and others provides models for sophisticated mathematical analysis. At best, however, these represent a first approximation to true phenomena because the materials of lubrication are mostly anisotropic. It is generally appreciated that anisotropic phenomenological behavior can be rationally explained using atomistic mechanics.

The importance of crystal structure on the friction, wear, and adhesion of metals has been documented in studies conducted in vacuum (refs. 18, 19, and 20). The influence of crystal structure can be illustrated with metals that are polymorphic as a function of temperature, for example, cobalt (fig. 3). Friction and wear increase dramatically with temperature as hexagonal cobalt assumes the f.c.c. structure; the effect and its reversibility have been demonstrated for a number of metals. Considering hexagonal metals as a group, a correlation involving coefficient of friction and the ratio c to a of the interbasal planar spacing to the lattice parameter has been documented (fig. 4). Further, it was shown (ref. 18) that alloying of hexagonal metals (e.g., 25 percent molybdenum in cobalt) stabilizes the crystal structure and eliminates the transition illustrated in figure 3. Also, alloying can be used to change the crystal lattice parameters to achieve reduced friction. Figure 5 shows the effect of such alloying additions of aluminum or tin to titanium as mentioned by Dr. Tabor; the result is a good example of how crystal structures, hence atomistic mechanics considerations, achieve useful slider alloys.

The question is often posed as to the applicability of simple crystal concepts to the polycrystalline materials used in sliding contact. During sliding, orientation or texturing occurs as a result of recrystallization and/or plastic deformation of surfaces. Fortunately, such orientation allows shear to develop on the plane of minimum shear strength. Figure 6 shows the characteristic effects of load on friction coefficient, for single crystals oriented for minimum shear force, and for polycrystalline materials. Friction of the polycrystalline material decreases with orientation at greater loads, and the friction of single crystals increases when grain boundaries appear with recrystallization. The magnitude of grain boundary impedence to shear is illustrated by the increase in friction for the single crystal. The two curves converge at a characteristic load. Table 1 relates recrystallization temperature to the load for which a series of metals converge to the same approximate coefficient of friction. The energy inputs for the conditions represented in this table are very small compared to engineering applications. Thus orientation or texturing is likely in most mechanisms.

Largely on the basis of NASA data, Dr. Merchant suggested that on certain crystal planes of copper, elastic recovery to "peel" the planes apart might play a large role in causing low adhesion. Dr. Tabor gives credence to that view. Considering the experience, however, at NASA-Lewis using a substantial number of materials, and especially the results for the platinum metals (ref. 22), a strong caution must be made against applying that view broadly to friction and adhesion. Also, in the study of titanium, friction was found to decrease with increasing temperatures as the *c* to *a* lattice ratio becomes larger; in the same range of conditions, elasticity of the material decreases. Further, for beryllium, the opposite effect of temperature on friction was observed but this also could be explained on the basis of *c* to *a* ratio (ref. 23).

Although discussed in the presentations of this symposium with particular reference to solid solubility or cold welding, it is well to reemphasize that material transfer is likely to occur whenever there is solid contact between load supporting surfaces in relative motion. Adhesion sufficient to cause material transfer occurs not only because of (1) diffusion bonding as mentioned above, but also because of (2) mechanical bonding, and (3) chemical bonding. The strength of these bonds determines if shear occurs at the original interface or in one of the substrate materials (ref. 18). The coefficient of friction (shear resistance) of slider metals has been directly related to critical resolved shear stresses and inversely related to stacking fault energies (refs. 21 and 22).

The summation of these observations on unlubricated sliding suggests that the phenomenological basis for continued study of slider

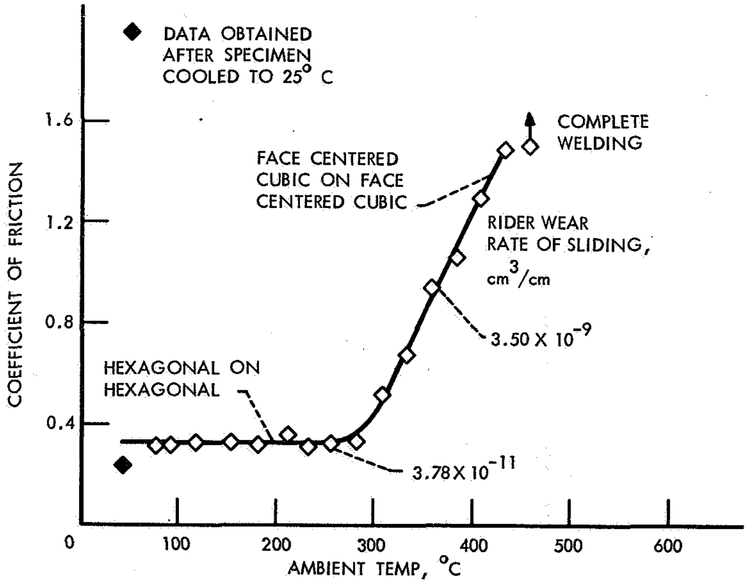


FIGURE 3.—Coefficient of friction for cobalt sliding on cobalt. 1000g load, 198 cm/sec sliding velocity, 10^{-9} torr.

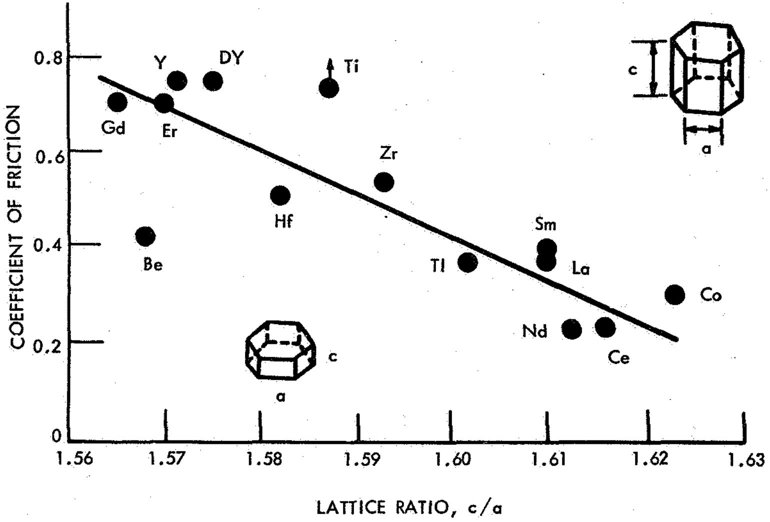


FIGURE 4.—Friction of various hexagonal metals on 440C steel. 1000g load, 390 ft/min sliding velocity, 10^{-9} torr.

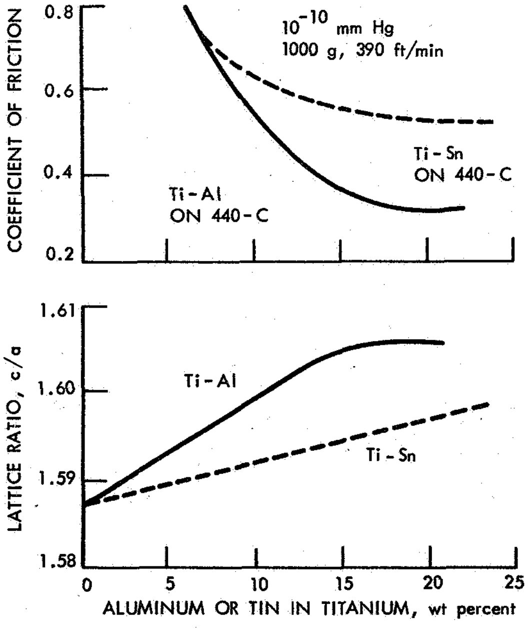


FIGURE 5.—Friction and lattice ratio (c/a) for titanium alloys.

materials should be in fracture mechanics and must involve atomistic considerations.

Better methods for the study and production of slider materials are urgently needed. Low energy electron diffraction as well as field-emission and field-ion microscopy are promising techniques for the definition of surfaces on an atomic basis. Inconsistencies in the use of such techniques are most likely the fault of our inadequate interpretation of the results rather than the techniques alone. A better theoretical basis for these devices is most certainly required.

Consideration for intermittent motion was mentioned by Dr. Tabor as having been overlooked in the primary papers of this symposium. There is no question that stick-slip phenomena are very important to engineering. An illustration from our own studies is the case of dynamic instability of bellows type face-contact seals reported in reference 24. Small accelerometers mounted on the nose-piece retainer of the seal measured oscillations at frequencies above 40 000 cycles per second. Instability contributed to diametral rocking of the seal, accelerated wear by impacts, and increased leakage past the seal. Studies to understand and control stick-slip must concentrate on rheology and especially the shear behavior of materials. Systems dynamics characterize stick-slip but have little to do with its origin.

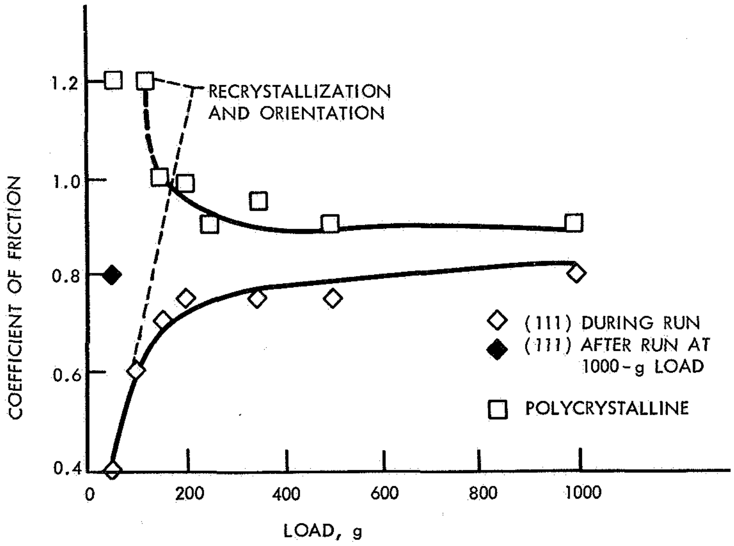


FIGURE 6.—Coefficient of friction for single crystal and polycrystalline copper sliding on polycrystalline Al_2O_3. Cu oriented with (111) plane parallel to sliding interface, 0.001 cm/sec sliding velocity, 10^{-11} torr, no external specimen heating.

The question of frictional behavior of materials, and especially polymers, at very low temperatures has been raised. Studies by D. Wisander at NASA-Lewis with PTFE in sliding contact with austenic stainless steel gave results as shown in table 2. From such data, it may be deduced that reduced ductility or other characteristics associated with lower temperatures of liquid nitrogen can give reduced friction; but the trend is not consistent as the data for liquid hydrogen illustrate. Clearly, we need to understand more fully the behavior of materials at cryogenic temperatures and ductility alone is not an adequate criterion for considering friction changes as a function of temperature.

Mr. Godfrey and Dr. Tabor neglected to discuss one of the considerations in boundary or thin film lubrication known to be one of Mr. Godfrey's major concerns, the photographic observation of cavitation-like behavior in sliding contacts. This writer's interest in the two-phase fluid phenomena of cavitation is, however, for different reasons. Cavitation has been visually documented in thick-film lubrication of journal bearings (ref. 25), in thin-film lubrication of rolling contacts (ref. 26), and in the boundary lubrication of sliding contacts by Mr. Godfrey.

Figure 7 taken from reference 26 shows how the wake of a rolling contact exhibits cavitation streamers with two phases. Concern for this phenomenon is its effect on cooling, because if the liquid does not wet the load-bearing surfaces under the conditions shown in figure 7, cooling may be seriously impaired. We have learned that many of the fluids, such as polyphenyl ethers, considered for advanced lubrication applications will not wet the films formed by their contact with metal surfaces (e.g., friction polymers). The apparent inadequate lubrica-

TABLE 1.—*Recrystallization Temperatures and Loads at Which Equivalent Friction Coefficients Were Obtained for Single Crystal and Polycrystalline Metals. Al_2O_3 Disk, 0.001 cm/sec Sliding Velocity, 10^{-11} Torr, No External Specimen Heating*

Metal	Load at which friction is equivalent for single and polycrystalline metals, *g*	Recrystalline temperature, ° C
Copper	200	100
Nickel	300	350
Iron	400	450
Titanium	1500	700
Beryllium	3500	900
Tungsten	3500	1200

tion by many new fluids (e.g., polyphenyl ethers) may result from heat transfer deficiencies rather than poor lubrication properties per se. It is interesting that polyphenyl ethers perform better in air than in inert environments. Contact angle measurements show that the fluid has a lower contact angle on surface films formed on steel in air than those formed in nitrogen. This is a good area for study.

Visual studies of thin-film lubrication are very useful. The work of Cameron and his associates (ref. 27) and that of Hingley (ref. 26) are good examples. Analytical studies of fluid-film seals at NASA-Lewis indicate that analyses of thin films should include the inertia terms from the Navier-Stokes equations (ref. 28). The limiting conditions for significance of inertia effects are not clear. With cooperation of Dr. Cameron, however, some data were obtained with high-density fluids (fluorocarbons) in two viscosity levels; it is considered significant that the low-viscosity (226 cp.) fluid gave about 50 percent greater film thickness than the corresponding high-viscosity fluid (1175 cp.). These data also suggest the more general consideration of inertia effects in extreme thin-film lubrication.

These comments are to supplement those of Dr. Tabor and the other authors. It is hoped that they will stimulate study of fundamentals applicable to lubrication technology.

M. C. Shaw (Carnegie-Mellon University, Pittsburgh, Pennsylvania)

While the elementary theory of friction proposed independently by Merchant and Bowden over a quarter century ago (refs. 29, 30, and 31) is qualitatively useful, it is far less satisfying quantitatively.

TABLE 2.—*Friction Coefficient for Polytetrafluoroethylene on Austenitic Stainless Steel. 3/16 in. Radius Hemisphere in Sliding Contact With the Flat Surface of a Rotating Disk at the Given Environments. 1 kg Load, 2300 ft/min. Sliding Velocity*

Materials	Temperature		Environment	Friction coefficient
	°F	°K		
PTFE on 304 stainless steel	+160	344	Gaseous nitrogen	0. 18
	−320	77	Liquid nitrogen	0. 07
	−423	20	Liquid nitrogen	0. 13
25% glass fiber filled PTFE on 304 stainless steel	+160	344	Gaseous nitrogen	0. 25
	−320	77	Liquid nitrogen	0. 12
	−423	20	Liquid hydrogen	0. 22

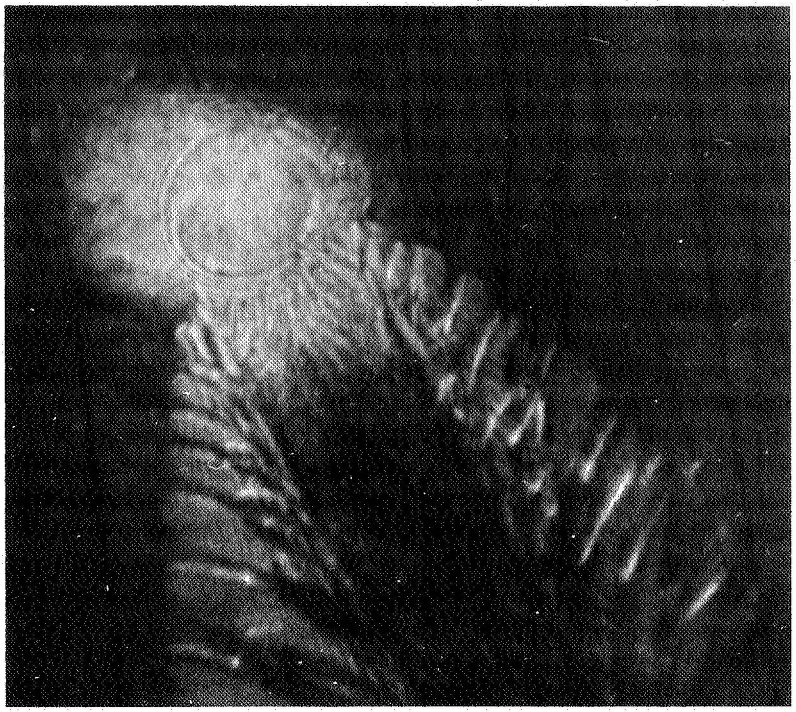


FIGURE 7.—Optical interferogram showing cavitation streamers in the wake of rolling contact (ref. 26).

In equation 4 of Dr. Merchant's paper, coefficient of friction is given by

$$f=\frac{S}{P}, \tag{1}$$

where S is the shear strength of the material and P is the hardness. This is consistent with the excellent results obtained with a bearing surface consisting of a weak layer on a hard substrate. However, when applied to a slider consisting of two homogeneous ductile metals, it gives quantitative results that are extremely poor. It is well established that the hardness, P, of a homogeneous metal is approximately three times the uniaxial flow stress, σ_0, for the same metal, and by the maximum shear theory, shear stress S will be one-half the uniaxial flow stress. It therefore follows from the equation above that the maximum value of f for clean surfaces would be

$$\frac{\sigma_0/2}{3\sigma_0} \text{ or } 0.167.$$

However, ductile metal surfaces that are clean are experimentally found to exhibit coefficients of friction far in excess of one.

The first observation to be made concerning equation 1 is that P is actually the unrestrained flow stress and not the hardness. Figure 8 shows two asperities just before contact as they are brought together in a normal direction. The degree of restraint that is present is far less than that in an indentation hardness test. The flattening of asperities more closely resembles the flattening of a cylinder—the uniaxial flow stress condition—than it resembles a hardness test which is equivalent to three times the uniaxial flow stress. The maximum value f could have for a homogeneous material, according to this interpretation, is then

$$\frac{K}{2K} \text{ or } \frac{1}{2},$$

by the Tresca or maximum shear criterion of plasticity, where K is the flow stress in shear. But experimentally observed values of f still remain unexplained.

An attempt has been made to explain the failure of equation 1 in terms of the "junction growth" that accompanies a tangential or frictional force F. In the derivation of equation 1 the real area of contact, A_R, was considered to be proportional to the normal force, W, on the slider, and no account was taken of the presence of the frictional force, F. The two forces F and W will, of course, combine to satisfy the criterion for plastic flow, and a given W will give rise to a larger area in the presence of a frictional component. We may define two hardnesses

$$P_1 = \frac{W}{A_{R1}} \text{ (in the absence of } F\text{)}, \tag{2}$$

$$P_2 = \frac{W}{A_{R2}} \text{ (when } F \text{ is present)}, \tag{3}$$

where W is the normal applied load. The hardness P_2 will obviously be less than P_1 since F helps induce plastic flow; hence $A_{R2} > A_{R1}$, which is what is meant by the "junction growth" that occurs when a shear force F is introduced. "Junction growth" is discussed in references 32 and 33.

While this idea is qualitatively appealing it fails to explain the discrepancy between the observed and predicted values of coefficient of friction for clean surfaces, as the following analysis will show. This analysis employs elementary principles of strength of materials—an engineering approach in which stresses are assumed uniformly distributed. This treatment considers a pair of similar asperities consist-

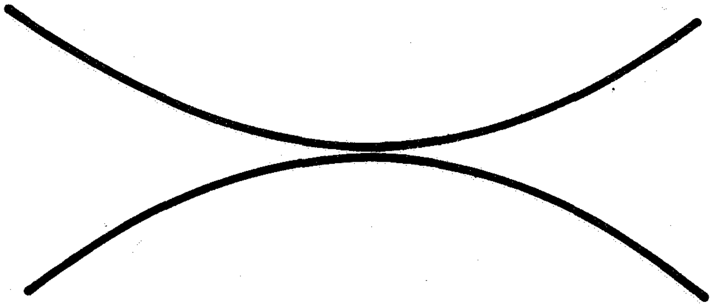


FIGURE 8.—Two asperities before contact.

ing of a homogeneous isotropic material that deforms without strain hardening.

Figure 9(A) shows the two asperities after equilibrium has been established. The applied load is W_1 and the area of contact is A_1. In figure 9(B) is the corresponding Mohr's circle diagram—a stress-space diagram giving the relative magnitudes of shear and normal stresses for cubes oriented in different directions. Plane X in the real space diagram (fig. 9(A)) is clearly a plane of principal stress and will occupy the position indicated in the stress-space diagram (fig. 9(B)). The radius of Mohr's circle is K, the shear flow stress for the material.

Figure 10 shows diagrams equivalent to those of figure 9 but with shearing force, F, present. The principal plane of stress will be perpendicular to the resultant force vector (plane Y in figure 10(A)), and this will plot as point Y on the Mohr's circle diagram. Plane X is subjected to both shear and normal stresses. If A_2 is the final area developed, these stresses will be

$$\tau_x = \frac{F}{A_2}, \tag{4}$$

and

$$\beta\sigma_x = \frac{W_2}{A_2}, \tag{5}$$

while angle QOX(fig. 10(B)) will be the friction angle, β, whose tangent is the coefficient of friction f.

Point Z in figure 10(B) is the plane of maximum shear stress, and is located at an angle 2α from plane X in stress space but at an angle α counterclockwise from X in real space (fig. 10(A)).

Figure 10 will hold for all cases for which f is less than unity ($\beta < 45°$), and

$$W_2 = \sigma_x A_2 = (OQ)A_2, \tag{6}$$

where OQ is the normal stress on plane X (fig. 10(B)).

But

$$OQ=2K \cos^2 \beta. \tag{7}$$

Letting

$$W_1 \text{ (fig. 9)} = W_2 \text{ (fig. 10), then}$$

$$2KA_1=2K \cos^2 \beta A_2, \tag{8}$$

or

$$\frac{A_2}{A_1}=\frac{1}{\cos^2\beta}=1+f^2, \tag{9}$$

where f=coefficient of friction=tan β.

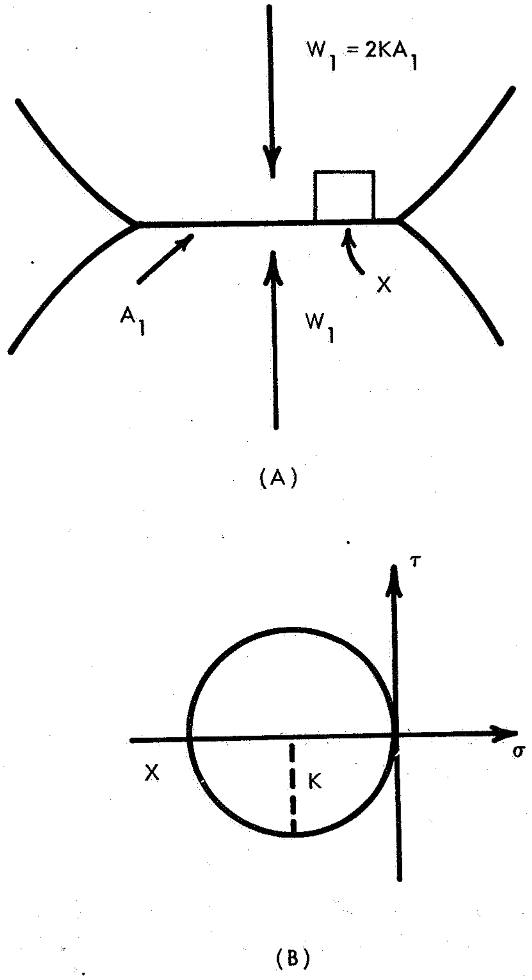
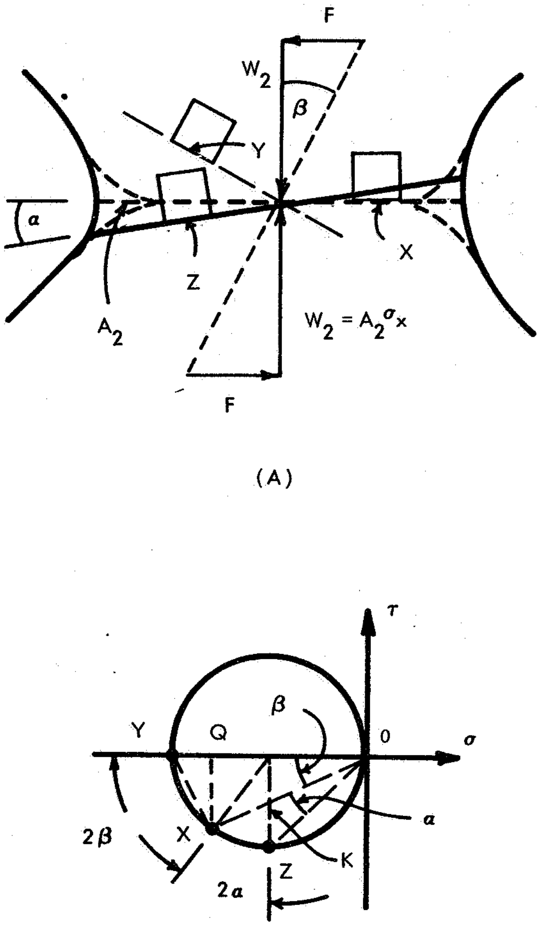


FIGURE 9.—Two asperities under normal load.



(A)

(B)

Figure 10.—Asperities in contact with f less than unity.

Equation 9 provides the relation between values P_1 and P_2 in equations 2 and 3 such that

$$\frac{P_2}{P_1}=\frac{1}{1+f^2}. \tag{10}$$

If the two junctions weld completely so that they form a single homogeneous isotropic body, we should expect shear rupture to occur on the plane of maximum shear stress and on the plane of shortest extent. This would mean that points X (shortest plane) and Z (maximum shear stress) in figure 10(B) would coincide. The coefficient of friction would then be one and this would be the greatest value possible under these conditions. If the two asperities were only partly welded (e.g., due to contamination), then rupture would still be expected to occur on the shortest plane but at a lower value of shear stress. Point X would then be to the left of point Z as in figure 10(B) and the shortest plane X would not be the plane of maximum shear stress Z. The coefficient of friction would then be less than one.

The relation between the coefficient of friction with junction growth, f, and without junction growth, f_0, may be derived starting with equation 1.

$$f=\frac{S}{P_2}, \tag{11}$$

where S is the shear stress at rupture for plane X.

From equation 10

$$f=\frac{S}{P_2}=\frac{S}{P_1}(1+f^2)=f_0(1+f^2), \tag{12}$$

thus

$$f_0=\frac{f}{1+f^2}. \tag{13}$$

Figure 11 shows the relation between f and f_0.

While the concept of junction growth explains coefficients of friction as high as one, it does not explain the values in excess of one when very clean ductile metals slide together. Further alteration in the basic model is obviously needed.

In looking for other explanations two possibilities come to mind:

(1) The normal load, W, on an asperity, could fluctuate and the maximum value of W might occur before the maximum value of F.

(2) The shear strength of the material near the interface of two asperities (V_1 in figure 12) could be greater than the flow stress K for the bulk material of the asperity (V_2 in figure 12).

Either of these possibilities would lead to values of f greater than one.

The first possibility calls for a vibrational motion perpendicular to the sliding direction. If the frequency of vibration were high enough, the normal load could be reduced to a lower value before shear took place. The maximum normal load would thus determine the area to be sheared, while the lower mean value of normal load would be present during shear. The frequency of vibration required is too high for the entire plate to vibrate, while the shape of asperities (very broad and shallow) make it appear unlikely that individual asperities could vibrate with sufficient amplitude to satisfy this mechanism.

There appear to be four possible ways in which the second mechanism might operate:

(1) Contaminating materials such as oxides might lead to dispersion strengthening of the immediate surfaces.

(2) Strain hardening of the surfaces might exceed that of the bulk metal.

(3) The rupture stress for surface material (V_1 in figure 12) might be greater than the flow stress for the material in the central region of an asperity (V_2 in figure 12) due to a size effect.

(4) Asperities might be subjected to a large component of hydrostatic compressive stress in addition to a uniaxial component of stress. It has been clearly shown by Bridgman and others that hydrostatic compression greatly increases rupture stress but has only a small influence on flow stress. This could also lead to the difference in properties for the surface and bulk regions.

The contaminating mechanism may be eliminated on grounds that the highest values of friction are observed under conditions of high vacuum and surface cleanliness.

The differential strain hardening mechanism appears to be too weak an effect to account for the large values of friction observed with clean surfaces. Even the difference in shear strengths of fully worked and annealed materials is not sufficient to account for the difference in strength required: an order of magnitude or more in some cases. This mechanism appears the more unlikely when it is realized that the difference in strengths of two plastically strained regions (V_1 and V_2 in figure 12) is involved. Since strain hardening tends to saturate with strain, any difference in such strengths would be small.

This leaves only the size and Bridgman effects to explain the anomaly.

A size effect, in which shear strength increases with decrease in specimen size due to the decreased probability of finding an imperfection or stress concentration, has been clearly demonstrated to exist in the case of glass fibers, a wide variety of whiskers, fine grinding chips, and microhardness indentations. It therefore does not appear unreasonable to expect a similar effect to exist at the tip of an asperity. The fact that the volume of an entire asperity will be of the order of 10^{-10} in.^{3}, while the volume of a chip or whisker showing a strong size effect will be about 5×10^{-9} in.^{3} and 10^{-8} in.^{3} respectively makes the possibility even more likely.

It can be shown that an asperity will be subjected to a large hydrostatic compressive stress. However, this requires consideration of the stresses on an asperity from the point of view of elasticity instead of

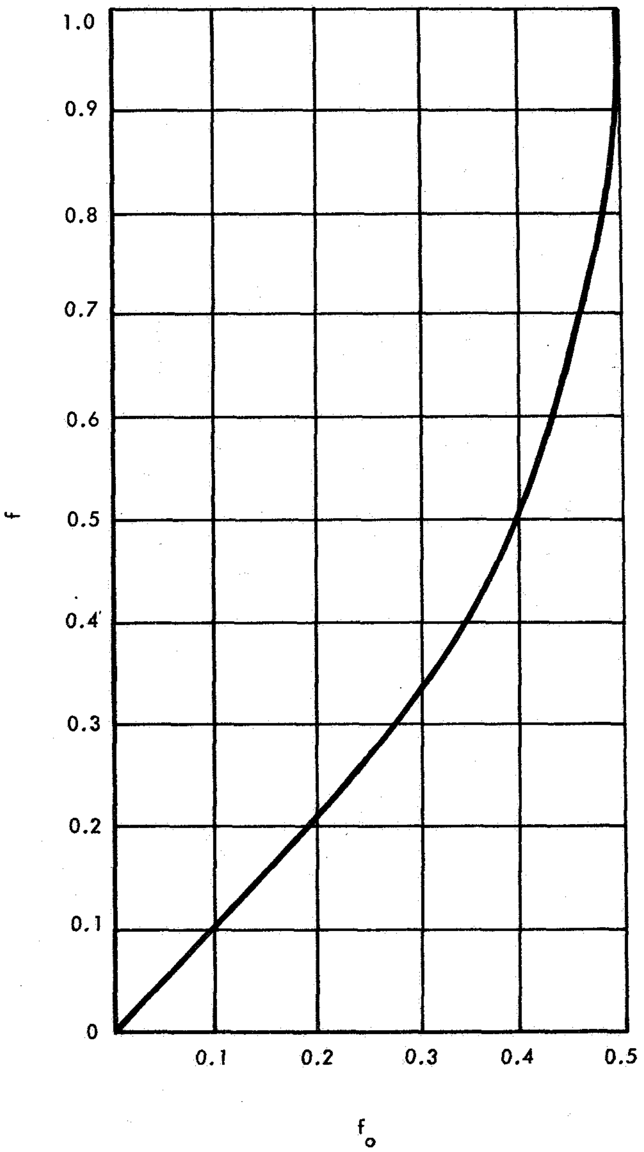


FIGURE 11.—Relation between f (including junction growth) and f_0 (ignoring junction growth).

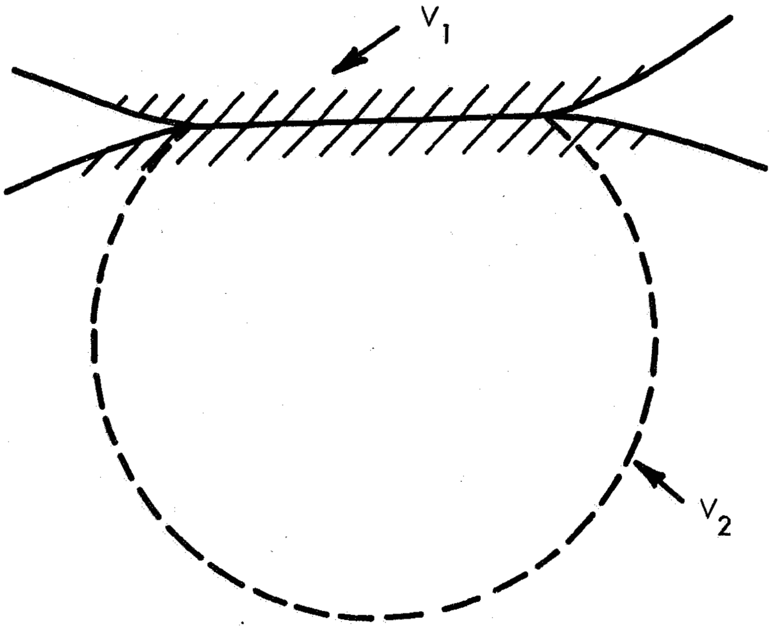


FIGURE 12.—Contacting asperities showing region of rupture stress at surface V_1 and region of flow stress enclosed within V_2.

the strength of materials approach employed here, and therefore will be deferred for a later publication.

Nevertheless it might be concluded that values of coefficient of friction in excess of unity can be explained in terms of a combination of size and Bridgman effects. Both of these effects lead to a rupture stress, K_r, for the surface of an asperity that is greater than the uniaxial flow stress, K, for the main body of the asperity.

The diagram corresponding to figure 10 for values of f greater than unity is given in figure 13.

Figures 10 and 13 are useful in indicating the direction of maximum shear stress corresponding to any value of f. This will be at an angle, α, to the direction of motion, while α will be considered positive for values of f less than one and negative for values of f greater than one. From the Mohr's circle diagrams it is evident that

$$\alpha = 45^\circ - \tan^{-1} f \tag{14}$$

Since we are dealing with ductile materials we should expect rupture to occur in shear and to follow the plane of maximum shear stress. It is thus to be expected that asperities having a value of f less than unity will fracture with an inclination such as that shown by the solid line in figure 10(A) (positive α), while asperities having an f greater than unity will fracture as shown in figure 13(A) (negative α).

By carefully tracing a surface that has been slid in contact with another similar dry clean surface, it should be possible to estimate the coefficient of friction for individual asperities that collectively determine the value which has previously been the only one possible to measure. The appearance of a surface carefully traced after sliding should be expected to appear as shown in figure 14, where a local

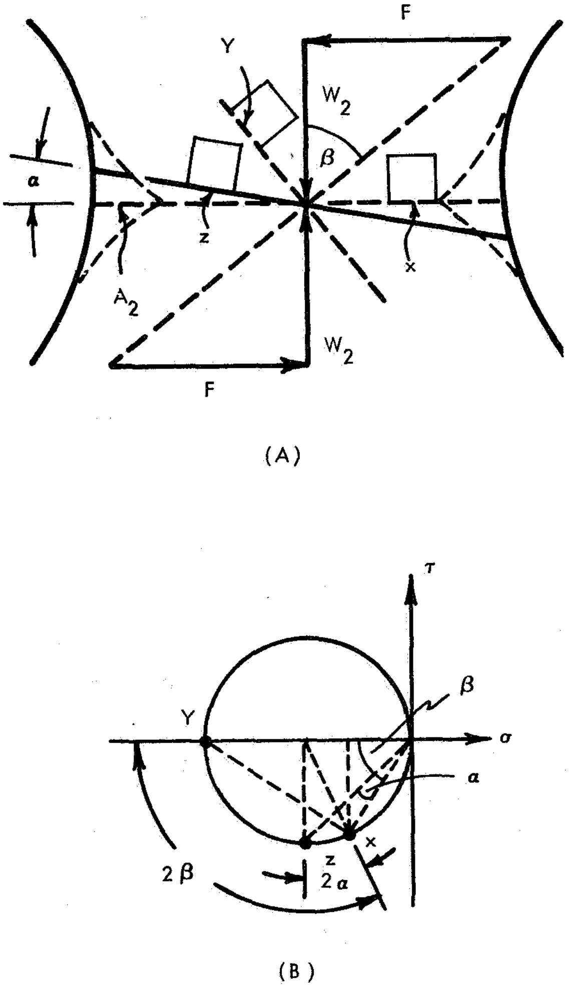


FIGURE 13.—Asperities in contact with f greater than unity.

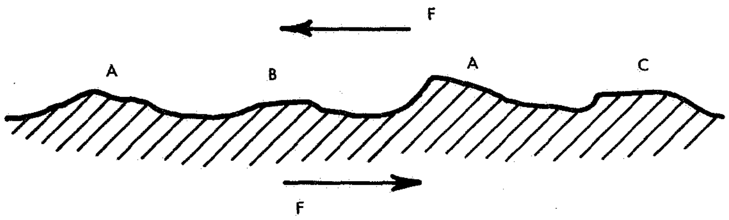


FIGURE 14.—Profile of interface showing asperities which have sheared at A ($f>1$), B ($f<1$), and C ($f=1$).

coefficient of friction is indicated to be greater than unity at *A*, less than unity at *B*, and equal to unity at *C*.

It is of interest to note that the large model asperity shown in figure 1(B) of Dr. Tabor's paper has an angle of α of about plus 5 degrees. This would suggest a condition of loading in this case corresponding to a coefficient of friction of tan $(45^\circ - 5^\circ) = 0.84$.

R. Courtel (Centre National de la Recherche Scientifique, Laboratoires de Bellevue, Bellevue, France)

The papers presented at this symposium are of remarkably high vision and exceptional interest.

I have been invited to discuss the basic principles enunciated and to present some further commentaries. I shall also submit some reflections made in that same spirit.

It is recognized that the study of friction takes its main elements from the various fields of mechanics, physics, solid state, physical chemistry, and metallurgy. Therefore, the overview of friction will vary according to both training of the given researcher as well as the scale of size employed. Phenomena such as molecular adhesion, frictional compatibility, physical adsorption, and chemisorption must ultimately find their explanation on the scale of short-distance interactions between atoms, molecules, ions, and electrons of the rubbing bodies present at the interface.

On the micron level one finds plastic deformation and material transfer mechanisms in crystalline solids (actual "macrowear" processes), denoting both the affinity of solids in contact as well as their local heterogeneity.

On the millimeter level, and above, one tries to describe phenomena as if they happened between two semi-infinite, continuous, homogeneous, isotropic solids, where the laws of elasticity, plasticity, and viscoelasticity can be applied. If one deals with machine parts, one uses formulas based on strength of materials, and one studies wear occurring during lengthy experiments for varying conditions of speed, load, temperature, environment, etc.

In the absence of a coherent doctrine which would give an understanding to the very small scale friction processes, it is the considerable practical importance of wear phenomena that establishes the relative de facto unity between friction theory and experimentation: each developing according to the field and training of the given researcher.

Under these conditions, it would appear that a classification of friction phenomena, because of its multidisciplinary basis, should be interdisciplinary in structure. Such a classification would clearly define goals as well as state of the art, but would not be restrictive as to areas of basic research.

Since friction and wear phenomena vary according to the size level (and time level) of observation, such a consideration should be made in any final classification.

I suggest that an attempt in this direction be made as the topic for a forthcoming symposium. This implies, of course, starting a difficult and, at first, a nonrewarding intellectual exercise, but it would have the advantage of shuffling ideas and setting forth certain poorly examined aspects of the friction phenomena.

LECTURER'S CLOSURE

I should like to thank Dr. Adamczak and Dr. Gisser for their interesting discussion on my own and other papers in this symposium. I have already emphasized the need for a broad-based attack if we wish to solve those problems of friction, lubrication, and wear confronting industry and technology. It seems to me that there are two features that should be borne in mind: First, that good people with good ideas should receive support, and second, that a mixture of disciplines within a given research group can prove very fruitful as well as stimulating to the members themselves.

Does this involve, as Dr. Adamczak suggests, the establishment of independent tribology departments? The answer must depend on the country concerned, on the economic and industrial environment, on the financial situation, on the intellectual and technological climate, and on the availability of suitable men to run such departments. As a visitor to this country I cannot say more than this. I can only repeat that in the long run men and ideas will prove crucial.

I would like to make two comments on Mr. Johnson's discussion. First, the friction of metals depends primarily on two factors: surface films and the deformation characteristics of the metals. With clean surfaces the second factor is the major one. In the beautiful systematic series of investigations carried out at NASA, the emphasis has been placed on crystal structure, *c* to *a* ratio, and "atomistic considerations."

Naturally all deformation characteristics are related to these parameters. However, we feel it more realistic to emphasize the deformation properties rather than the structural factors. When clean surfaces are slid over one another, the friction and surface damage will be very large if the metals can easily deform. This occurs with face centered cubic materials, particularly if they do not work harden too rapidly. If plastic deformation is restricted either because of intrinsic structural factors or because the material has a steep work hardening characteristic, the friction and surface damage will be less. This has been shown in recent work in our laboratory by T. C. Childs on the friction of metals down to 30° K; here the main effect of low temperature is to increase the work hardening index resulting in a decrease in friction and damage.

My second comment concerns the behavior of PTFE sliders at low temperatures. This material has puzzling frictional properties. At one time we considered that the low friction was due to its poor adhesion. We have since found that in some cases strong adhesion can occur; nevertheless the friction still remains low. Apparently a very thin film (about 200Å thick) is drawn out of the PTFE during the sliding process and the low friction must be associated with the special drawing properties of PTFE. I am happy that Dr. Steijn of duPont has confirmed and extended these observations. The speed and temperature dependence of PTFE friction reflects the way in which hardness (which determines the area of contact) and drawing properties of PTFE depend on the rate of deformation and temperature. Some typical results extracted from the paper by Ludema and Tabor which illustrate this are shown in figure 15. The sliding speeds were all very low so as to minimize the effects of frictional heating. The figure also includes Mr. Johnson's data obtained at a much higher sliding speed where behavior is probably very much complicated by frictional heating of the polymer.

Professor Shaw's contribution, as one might expect, is an interesting and stimulating one.

First I would like to say that in experiments carried out between an indium slider and a flat glass surface, very marked junction growth can occur. Further, it is possible to determine the area of real contact at any stage of the frictional process either by direct optical observation or by measurement of adhesion. It turns out that frictional force increases steadily with junction growth. If the surfaces are reasonably clean it is easy to obtain a coefficient of friction exceeding unity. The important points here are that the junctions are large, little work hardening occurs, and contact pressures are relatively small. Thus none of the mechanisms suggested by Professor Shaw needs to be invoked to explain a large coefficient of friction. If the surfaces are

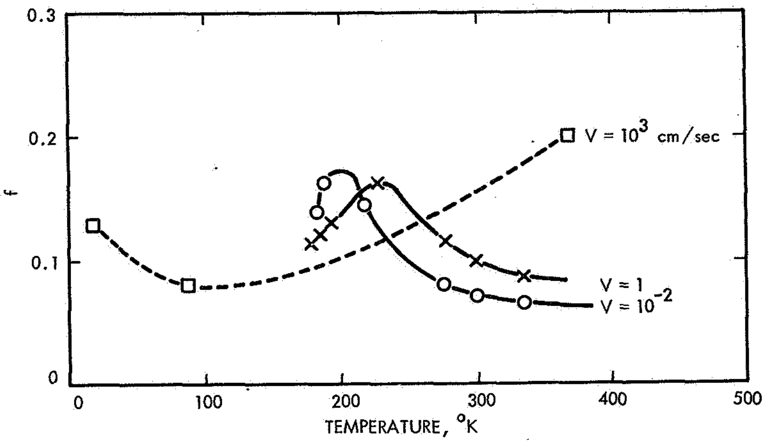


FIGURE 15.—Friction of steel and PTFE as a function of temperature. The low-speed results, $v=1$ and $v=10^{-2}$ cm/sec, are taken from the paper by Ludema and Tabor. The high-speed result, $v=10^{3}$ cm/sec, is from data given by Johnson; here frictional heating is probably important.

clean, if the materials are ductile, and if the geometry is appropriate, junction growth can occur on a large scale as a result of the combined normal and tangential stresses. These results show that junction growth can be sufficient to account for very large coefficients of friction. This does not mean, of course, that the factors described by Professor Shaw may not be important in other situations.

To my mind the main value of his contribution is the stimulation it may provide to professional plasticians for more detailed study of the yield conditions around a friction junction. His analysis, admittedly an approximate one, suggests that the direction of shear will vary with the coefficient of friction. The direction of shear may be of very great significance in at least one practical field. The recent work of Tolstoi suggests that intermittent motion in sliding is generally accompanied by minute up-and-down movement of the surfaces. If this up-and-down movement is damped out the intermittent frictional movement may be suppressed. The causes of up-and-down movement are clearly of very great importance. It may be connected with the direction of shear described by Professor Shaw or with the "bourrelet frontal" described some time ago by Dr. Courtel. Professor Shaw has exposed here an issue of importance and a challenge to the materials scientist.

In his comment Dr. Courtel submits some reflections on research in tribology. I fully agree with his general point—that in our interdisciplinary approach to friction, lubrication, and wear, both our theoretical concepts and our experimental techniques must be matched in scale to the particular problem we are investigating.

REFERENCES

1. White, M. L.: The Wetting of Gold Surfaces by Water. J. Phys. Chem., vol. 68, 1964, p. 3083.
2. Bailey, J. M.: Surface Deformation and Friction on Single Crystals of Copper. Ph. D. thesis, University of Virginia, 1959.
3. Courtel, R.: Modes d'intervention de la troisième dimension dans les processus de frottement intermetallique. Rôle du bourrelet frontal. Bull. d'Information Scientifiques et Techniques du Commissariat a l'Energie Atomique, vol. 90, 1965, p. 1.
4. Tolstoi, D. M.: Significance of the Normal Degree of Freedom and Natural Normal Vibrations in Contact Friction. Wear, vol. 10, 1967, p. 199.
5. Bowden, F. P.; and Brookes, C. A.: Frictional Anisotropy in Nonmetallic Crystals. Proc. Roy. Soc., vol. A295, 1966, p. 244.
6. Billinghurst, P. R.; Brookes, C. A.; and Tabor, D.: The Sliding Process as a Fracture-Inducing Mechanism. Conference on Physical Basis of Yield and Fracture, (Oxford), 1966, p. 253.
7. Grosch, K. A.: The Relation Between the Friction and Viscoelastic Properties of Rubber. Proc. Roy. Soc. (London), vol. A274, 1963, p. 21.
8. Schallamach, A.: A Theory of Dynamic Rubber Friction. Wear, vol. 6, 1963, p. 275.
9. Ludema, K. C.; and Tabor, D.: The Friction and Viscoelastic Properties of Polymeric Solids. Wear, vol. 9, 1966, p. 329.
10. Bayer, R. G.; Clinton, W. C.; Nelson, C. W.; and Schumacher, R. A.: Engineering Model for Wear. Wear, vol. 5, 1962, p. 378.
11. Clinton, W. C.; Ku, T. C.; and Schumacher, R. A.: Extension of the Engineering Model for Wear to Plastics, Sintered Metals, and Platings. Wear, vol. 7, 1964, p. 354.
12. McCutchen, C. W.: Animal Joints and Weeping Lubrication. The New Scientist, vol. 15, 1962, p. 412.
13. Lancaster, P. R.; and Rowe, G. W.: A Comparison of Boundary Lubricants under Light and Heavy Loads. Wear, vol. 2, 1959, p. 428.
14. Codell, M.; Weisberg, J.; Gisser, H.; and Iyengar, R. D.: Electron Spin Resonance Studies of Adsorbed tert-Butyl Hydroperoxide on Zinc Oxide. National ACS Meeting, San Francisco, March 31 to April 5, 1968.
15. McCrackin, F. L.; Passaglia, E.; Stromberg, R. R.; and Steinberg, H. L.: Measurement of Thickness of Very Thin Films and the Optical Properties of Surfaces by Ellipsometry. J. Research, NBS, vol. 67A, 1963, p. 363.
16. Wendlandt, W. W.; and Hecht, H. G.: Reflectance Spectroscopy. Interscience Publishers (New York), 1966.
17. Rabinowicz, E.: Friction and Wear of Materials. John Wiley & Sons, 1965.
18. Johnson, R. L.; and Buckley, D. H.: Lubrication and Wear Fundamentals for High Vacuum Applications. NASA TM X-52271, 1967.
19. Buckley, D. H.: The Influence of Crystal Structure, Orientation and Solubility on the Adhesion and Sliding Friction of Various Metal Single Crystals in Vacuum. Adhesion or Cold Welding of Materials in Space Environments, ASTM, 1967, p. 248.
20. Buckley, D. H.: The Influence of the Atomic Nature of Crystalline Materials on Friction. ASLE Trans., vol. 11, 1968, p. 89.

21. Buckley, D. H.: Possible Relation of Friction of Copper-Aluminum Alloys with Decreasing Stacking-Fault Energy. NASA TN D-3864, March 1967.
22. Buckley, D. H.: Sliding Friction of Some Members of the Platinum Metals Group. NASA TN D-4152, 1967.
23. Buckley, D. H.; and Johnson, R. L.: Friction and Wear of Hexagonal Metals and Alloys as Related to Crystal Structure and Lattice Parameters in Vacuum. ASLE Trans., vol. 9, 1966, p. 121.
24. Hudelson, J. C.: Dynamic Instability of Undamped Bellows Face Seals in Cryogenic Liquid. NASA TN D-3198, Jan. 1966.
25. Cole, J. A.; and Hughes, C. J.: Oil Flow and Film Extent in Complete Journal Bearings. Proc. Inst. Mech. Engrs., vol. 170, 1956, p. 499.
26. Hingley, C. G., et al.: A Study of the Geometry of Elastohydrodynamic Films in Point Contact. SKF Prog. Rept. No. 1, U.S. Navy Contract NOO019-67-C-0206.
27. Foord, C. A.; Cameron, A.; and Hammann, W. C.: Evaluation of Lubricants Using Optical Elastohydrodynamics. ASLE Trans., vol. 11, 1968, p. 31.
28. Zuk, J.; Ludwig, L. P.; and Johnson, R. L.: Flow and Pressure Field Analysis of Parallel Groove Geometry for an Incompressible Fluid with Convective Inertia Effects. NASA TN D-3635, Sept. 1966.
29. Merchant, M. E.: The Mechanism of Static Friction. J. Appl. Phys., vol. 11, 1940, p. 230.
30. Bowden, F. P.; and Tabor, D.: The Area of Contact between Stationary and Between Moving Surfaces. Proc. Roy. Soc. (London), vol. A169, 1939, p. 391.
31. Bowden, F. P.; and Tabor, D.: Mechanism of Metallic Friction. Nature, vol. 149, 1942, p. 197.
32. Courtney-Pratt, J. S.; and Eisner, E.: The Effect of Tangential Force on the Contact of Metallic Bodies. Proc. Roy. Soc. (London), vol. A238, 1957, p. 529.
33. McFarlane, J. S.; and Tabor, D.: Relation Between Friction and Adhesion. Proc. Roy. Soc (London), vol. A202, 1950, p. 244.

Critical Appraisal and Research Opportunities—the Materials Research Viewpoint

K. R. LAWLESS
University of Virginia
Charlottesville, Virginia

FRICTION AND WEAR have been familiar to man since prehistoric times. However, scientific or engineering studies of friction date only from the time of Leonardo da Vinci (1452–1519), and modern studies of wear have been done primarily in the last thirty years (ref. 1). The purpose of this paper is to attempt to make an overall evaluation of the current state of knowledge in these fields, particularly as revealed in the first six lectures of this symposium and as viewed from the standpoint of a materials scientist. On the basis of this evaluation, certain areas in which research is badly needed will be suggested.

The phenomena of friction and wear, considered by many as belonging to the realm of mechanical engineering, are in fact truly interdisciplinary. Friction is often placed under the heading of physics, wear is placed under metallurgy, and adhesion and lubrication are often placed under chemistry. As a matter of fact, physics, chemistry, and metallurgy must all be considered in any complete treatment of any one of the above subjects. This is, of course, very clear from the previous papers.

Friction, adhesion, wear, and lubrication are generally thought of as surface phenomena. This is certainly an oversimplification, unless one considers that the "surface" has a very remarkable depth. It is preferable to think of the initial processes as surface interaction phenomena, since in all cases either a solid-gas, solid-liquid, or solid-solid surface contact is involved. However, the bulk properties and processes of the involved materials must play an important role at any time after the very initial surface contact. The mechanisms of these processes must therefore be extremely complex and involve the interaction of surface and bulk phenomena. This point has already been strongly emphasized during this symposium.

The complexities of the interactions involved suggest very strongly that engineering type measurements of coefficients of friction, adhesion, or wear will reveal very little of a fundamental nature that will help our understanding of the basics of friction and wear. Such coefficients, as usually measured, are by no means fundamental materials properties, although many engineers treat them as such; they are at best a crude description of a complex interacting system of two material surfaces that are almost always covered with contaminants. It is quite clear that an understanding of friction and wear must involve the determination of a number of fundamental bulk and surface properties, as well as their mutual interactions. Volume properties must include elastic and plastic deformation parameters, while surface properties must include structure, surface energy, adsorption, chemical reactivity and epitaxy.

The major emphasis in these lectures has been on metal friction and wear systems, with only a very minor account of inorganic or polymeric systems. This paper will continue in this vein with only a minor discussion of nonmetallic systems.

STRUCTURE OF SURFACES AND THEIR INTERACTIONS

It was clear from the lectures of both Gatos and Merchant that the detailed nature of the solid surface is of great importance in the processes of friction and adhesion. In this section, I shall concentrate on the current status of our knowledge of solid surfaces.

First of all, what is a surface? The usual definition states that a surface is an interface or boundary between two phases, or two grains of the same phase with different orientations. Because of the rapid structural changes which occur near or at phase boundaries, such as the sharp termination of the periodic structure at the surface of a crystal, surface properties are nearly always different from bulk properties. When we talk of solid surfaces we must always keep in mind the distinction between the ideal or clean surface, and the real or contaminated surface. The real solid surface is actually a series of interfaces between different phases; metal oxide layers, adsorbed water on oxide, and adsorbed water-gas interfaces may all be present within less than 50Å of a true metal surface. Further complicating things, all of these boundaries may be diffuse due to interpenetration of atoms from one phase into another. It should be quite clear that real solid surfaces are by no means simple.

Our current knowledge of solid surfaces is based almost entirely on experimental information, much of which is highly unreliable because of the lack of characterization of surface cleanliness. Surface energies and the atomic positions and spacings in a surface have been, and still are, extremely difficult to determine either experimentally or analytically. It is only in recent years that the improvement in vacuum technology has made possible the attainment of clean surfaces remaining that way long enough for experiments to be performed. A major question which arises is: How does one determine if a surface is truly clean? I shall discuss this later.

The atomic structure, i.e., the arrangement of the atoms in solid surfaces has been observed directly by means of the field-ion microscope (ref. 2). The overall structure of actual surfaces are remarkably consistent with models based on extrapolation of the bulk crystalline structure, as determined by x-ray diffraction. A model of the surface of a face-centered cubic crystal is shown in figure 1, and should be compared with the field-ion picture of a platinum crystal shown by Gatos. Despite these observations by field-ion microscopy, which were made on metals and alloys, it should not be concluded that the surface arrangement of atoms is the same as the bulk for all materials. Low energy electron diffraction (LEED) and high energy electron diffraction (HEED) studies have been made of supposedly "clean" surfaces of many metals, semiconductors, and nonmetals (ref. 3). These studies have indicated that for most "clean" metals, the surface arrangement of atoms is the same as the bulk, although there is considerable controversy over the surface structure of gold and platinum (ref. 4). For semiconductor materials, however, the LEED data indicate a more or less complex rearrangement of the surface atoms, apparently taking place in order to satisfy the unsaturated surface bonds (ref. 5). The data on nonmetallic materials is not yet very extensive, but indications are that the surfaces do not represent the volume.

The controversy on the surface structure of gold and platinum is related to the more important problem of experimental determination of when a surface is absolutely clean. The LEED pattern in itself is clearly not sufficient for at least two reasons: First, randomly situated clumps of contaminants, such as oxide, may not show up in the pattern (ref. 6) (fig. 2); and second, amorphous structures, such as a random arrangement of adsorbed molecules, will in most cases contribute mainly to the general background scattering and to modifications in the intensity of spots from the ordered substrate; these effects are not readily obvious in the diffraction pattern. It is, therefore, necessary to use more refined techniques to determine true surface cleanliness. Measurements of spot and background intensities in the diffraction pattern can be helpful, but perhaps the best technique is to use work function measurements in conjunction with the diffraction studies, since work function is extremely sensitive to trace impurities on the surface regardless of whether they show an ordered or disordered structure.

The ion-microprobe mass spectrometer described by Gatos provides a powerful tool for characterizing the cleanliness of a surface, but its sensitivity is not yet as good as would be desired. It is also not the type of instrument which can readily be used in conjunction with structure determining techniques such as LEED. Recent investigations by Sewell (personal communication) indicate that partial monolayers of elements as low in atomic number as carbon can be detected and identified by using a high energy (30 to 50 kev) electron beam striking the surface at a very low glancing angle and analyzing the x-rays generated.

Our current understanding of the energetics of solid surfaces is poor. The thermodynamic treatment of Gibbs (ref. 7) has been applied very successfully to liquid interfaces, and has in recent years been utilized for studies of solid surfaces. Other authors such as Herring (ref. 8), Mullins (ref. 9), and Cabrera (ref. 10) have extended the phenomenological treatment of surfaces and applied it to many surface problems. Detailed discussions of such equilibrium problems as the determination of the equilibrium shape of small crystals and intersecting surfaces have been given.

Atomistic or quantum mechanical treatments of solid surfaces are, however, with the possible exception of semiconductor surfaces, quite

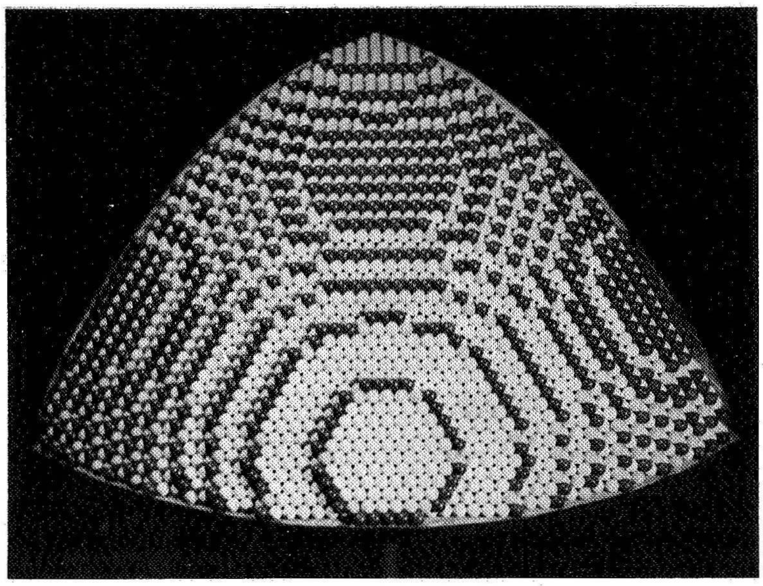
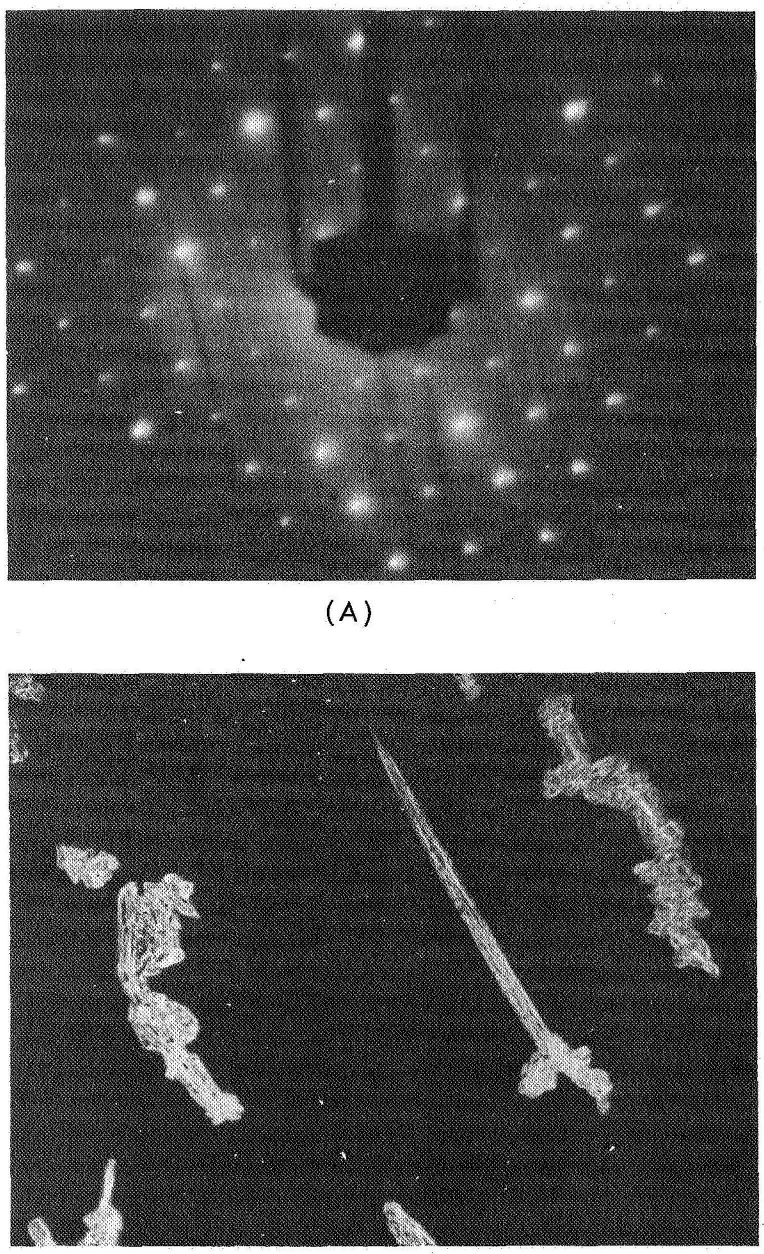


FIGURE 1.—Hard-sphere model of a face-centered cubic crystal.



(A)

(B)

FIGURE 2.—(A) Low-energy electron diffraction pattern showing only a centered (6×2) oxygen adsorption structure on a (110) face of copper. (B) Dark field optical micrograph of (110) copper surface showing oxide growths which were not revealed in the LEED pattern. × 200.

scarce, mainly because of the great mathematical difficulties involved in the solution of the equations. Hopefully this situation will improve with the help of modern computing techniques. Meanwhile experimental determinations of such quantities as surface stress and surface tension are difficult to make and only a very few determinations have been made on really clean surfaces. Considerable emphasis has been put on determinations of the variation of surface tension, γ, with crystal orientation. There is a tremendous opportunity for research in this area, but despite the availability of techniques for obtaining clean surfaces, the experimental difficulties are great.

Our understanding of the interactions of solid surfaces with gases, liquids, or other solids is to a great extent dependent on our knowledge of surface energetics and structure, which we have seen is improving, but is not yet good. Yet in practice, much of the available information on surface energetics has been derived from studies of surface-gas interactions. These experiments have been carried out mostly on poorly characterized surfaces and have led to a proliferation of unreliable data. Experimental methods such as field-emission or field-ion microscopy and electron diffraction used in conjunction with ultrahigh vacuum techniques have improved this situation.

The first stage of a solid-gas interaction is adsorption. Adsorption is usually classified as either physical adsorption, for which the binding energy is relatively weak (10 k cal/mole or less), or chemisorption, where the bond is much stronger, e.g., of the same order of magnitude as a chemical bond. The variation in both physical adsorption and chemisorption with crystal orientation has been proved beyond doubt for many different materials by the above techniques, but the details of the interactions are still not clear. It is thought that heterogeneities within a crystal plane, such as a dislocation or an impurity atom, may play an important role in adsorption and related processes. The positive evidence of such a role is still lacking and it is likely that steps or kink sites on a surface are, after crystal orientation, the most important surface heterogeneities.

Adsorption is one of the most fruitful and rewarding areas of research in surface science today. Many sophisticated techniques including field-emission and field-ion microscopy, LEED and HEED, flash filament desorption, work function measurements, calorimetry, and electron spin resonance are being used to obtain detailed information on surface-gas interactions. The field is wide open for further studies.

I would like to cite one example of a controversial area of study which is of considerable importance in adhesion and friction studies. LEED data from experiments on the adsorption of oxygen on the (110) face of nickel has been interpreted in terms of a surface rearrangement, or reconstruction of the nickel surface atoms (ref. 11). This interpretation is based on the assumption that scattering from the oxygen atoms is weak compared to that from the nickel. This assumption has been questioned by Bauer (ref. 12) who suggests that scattering of low energy electrons from oxygen may be quite large, and that the above interpretation is therefore subject to doubt. This controversy has not been resolved as yet, but it seems clear that additional measurements, such as that of work function, would clear this up. An important point is that in studying surface-gas interactions, several different measurement techniques should be used (e.g., LEED plus work function measurements).

It is unfortunate that the mechanisms involved in the oxidation of most materials are still subject to doubt (ref. 13). This is due primarily to an overemphasis on the determination of the kinetics of oxidation reactions, most of which were carried out on surfaces which were either poorly or not at all characterized, and to a lack of attention regarding the detailed structures of the oxide films formed. In recent years only a few scientists have given much attention to the development of suitable theories of oxidation (ref. 14), and in most cases adequate experimental data are not available to provide a good test of the theories.

The initial stages of the oxidation process have been particularly neglected up to the last few years, and there is essentially no knowledge available concerning the important transition from an adsorption layer on a surface to a three-dimensional oxide phase. Detailed epitaxial studies have been made on a number of metal-oxide systems (ref. 15), but more data are needed on other systems since the oxide orientation may play an important role in friction processes. There is a marked scarcity of information on the mechanical properties of oxide films on metal surfaces. These type of data are absolutely essential for an ultimate understanding of the oxidation process as well as for understanding the mechanisms of friction and wear. It seems strange that so little research has been done in this critical area, and although the experimental difficulties are considerable, particularly for very thin films, the research opportunities are great.

The environment-sensitive mechanical behavior of materials is really a surface problem which has become the subject of great interest in recent years (ref. 16). The mechanical behavior of most crystalline materials is determined by the generation and/or motion of dislocations. The presence of an oxide film, or of an adsorbed film on the surface, can influence the mechanical properties of a solid by affecting either the generation or movement of dislocations at or near the surface. Environmental effects on the generation and operation of surface sources of dislocations have been clearly demonstrated for nonmetals,

323-472 O—69——29

but are still the subject of considerable controversy for metallic materials (ref. 17). Even in cases where these effects have been very clearly demonstrated, there seems to be no simple single explanation. This subject deserves further study, but requires very careful experiments to distinguish between the effects of surface dislocation sources and barriers.

The presence of water on a surface, either as an adsorbed vapor film or as a liquid containing various ions, plays a large role in a tremendous variety of technologically important processes. For example, stress corrosion cracking and corrosion fatigue, which may be important in many wear processes, have been studied extensively, but the mechanisms involved are still the subject of great controversy. This is, of course, a fascinating area of research, and studies aimed at understanding not only the basic mechanisms involved, but also the relationship of these processes to wear, should be very fruitful.

Adsorbed water on a surface may also give rise to photomechanical and electromechanical effects for semiconductor materials; many of the oxide films on metal surfaces are semiconductors. Perhaps of more importance is the observation of anomalous indentation creep (ref. 18). This effect, in which hardness is lowered and is also time dependent, has been confirmed for nonmetals but is still subject to question for metals. It may, however, play an important indirect role on yielding or creep of metal specimens where an oxide film is controlling the mechanical behavior of the materials. Satisfactory mechanisms for these important effects, usually called Rebinder effects, are not known and should be an interesting subject for further research.

I have been talking so far about surface structures from an atomistic point of view. I will now consider surfaces from a more macroscopic, and perhaps more practical, point of view. Field-ion microscopy gives a highly detailed view of the atomistic arrangement of solid surfaces, but the technique cannot be extended to give a view of a large surface area. This is most unfortunate because friction and wear processes may involve quite large areas and it would be highly desirable to know the corresponding detailed surface structure.

A number of observational techniques are available today for examining surface topography, and most have already been mentioned by Williamson. Standard optical microscopy techniques, although widely used, have two major disadvantages: First, the resolution is limited by the numerical aperture of the lenses and by the wavelength of the illumination to no better than about 0.1 to 0.2μ; and second, the depth of field at high magnifications is very small because of the large numerical apertures of the objective lenses; thus, with the exception of extremely smooth surfaces, it is never possible to get the whole surface in focus. Taper section methods improve the situation but the taper cuts are destructive and difficult to make accurately. Interferometric techniques are really suitable only for quite smooth surfaces.

Electron optical techniques are currently the most satisfactory available for the detailed examination of surface topographies. Standard carbon replica techniques are capable of a resolution of 20Å or better and for this resolution the depth of field is of the order of 7000Å. Quantitative data on the heights and shapes of asperities can easily be obtained by placing a few latex spheres of a known size on the surface and by shadowing at two known angles. Quantitative stereo electron microscopy of replica surfaces can provide detailed three-dimensional information of the shapes of surface irregularities, but the technique is substantially more complex than standard microscopy.

The reflection electron microscopy technique has been used by a number of workers (ref. 1) for friction and wear studies. This technique is limited, however, because of its poor resolution and magnification distortion. A much more powerful technique, scanning electron microscopy (ref. 1), the principles of which were first developed nearly thirty years ago, has been used increasingly for the examination of surface topography in the past couple of years. The best resolution for the reflection mode use of this equipment is around 100 to 150Å which is not as good as the replica techniques. Scanning microscopy, however, has the simplicity of use of the optical microscope, and a depth of field at least 300 times as great for comparable magnifications. Magnifications may be varied over a wide range from × 20 to × 50, 000. It might be possible to carry out continuous observations during a friction experiment. The resolution will undoubtedly be improved considerably in the near future and I believe the scanning microscope will be the most useful tool available for the detailed study of surface topography over large areas.

Williamson and his co-workers have made very sophisticated use of profilometric techniques for the examination of surface topography. Using a direct connection from a profilometer to a computer, detailed numerical analyses of surface profiles can be made and surfaces described in terms of height distributions. The computer has also been utilized to prepare detailed contour maps of surfaces which show clearly the shapes of individual asperities and their arrangement on a surface. The profilometer has a horizontal resolution of only about 5000Å, compared to 20Å for the electron microscope replica technique, but an ability to reproduce the surface to better than 25Å accuracy (100Å for the microcartography). Williamson believes that it would not be profitable to strive for finer horizontal discrimination since contact areas are rarely smaller than about 12 700Å (50μin.) For contact area studies the resolution of the profilometric technique is probably adequate. However, for detailed studies of any plastic deformation of asperities and the surrounding bulk material, the surface features, such as slip steps, would not be resolved with this technique. For such studies it is desirable to define the surface on as close to the atomic level as possible, and electron optical techniques or field-ion microscopy are the only methods currently available for this sort of resolution. It should be noted that preliminary friction studies, i.e., the effect of contact, have already been tried in the field-ion microscope (ref. 20).

Investigations of static contact areas and the deformation associated with the contact have only become possible with the development of the techniques discussed above. Williamson finds that at light loads both the number of contact areas and the real areas of contact increase proportionally with the load when a hardened anvil presses against a bead-blasted aluminum surface. Above a certain load, however, the number of contacts becomes constant and the area still increases with load but no longer proportionately. These experiments were carried out in the region of plastic deformation and Williamson has suggested that this change in behavior is the result of en masse rather than individual asperity deformation. This suggestion seems very likely to be correct, but needs detailed information on the deformation of the asperities and the surrounding bulk area for confirmation.

The deformation problem is a complex one, both experimentally and analytically, particularly if polycrystalline samples are used because there will not necessarily be any specific relationship between the slip systems associated with adjacent asperities. Grain boundaries complicate the problem by introducing constraints to dislocation motion and the exact manner in which each grain deforms may be compromised by its orientation with respect to its neighbors. The simplest experiment would utilize single crystal surfaces of known orientation with a controlled spacing of asperities. The deformation of the asperities and the adjacent bulk area could then be studied metallographically and the interaction of the various slip systems about each asperity could be easily studied as a function of load. It is, however, still difficult to associate quantitatively the dislocation behavior of single crystals with the macroscopic behavior of a polycrystalline material.

The deformation of asperities is probably even more complex than indicated above. The size of the initial contact regions on most asperities are usually very small and the stressed regions may, in the manner of metal whiskers, be able to withstand stresses several orders of magnitude greater than the macroscopic yield stress. This means that the stresses could be transmitted through the asperities and cause deformation in the bulk before plastic flow occurs in the asperities. On the other hand, the asperities may be unusually hard as a result of prior work hardening of the metal; this also could result in yielding in the bulk before it occurs in the asperities. Oxide films could complicate the behavior even more. A great deal of work on the details of deformation is needed. I will come back to this subject again.

It should also be noted that the relationship of load and contact area is generally expressed as

$$A=\frac{W}{p},$$

where p is the local indentation hardness. This is based on the assumption that the contact pressure on an asperity is just sufficient to cause plastic flow. This is certainly just an approximation. A more detailed analysis, however, has been given by Greenwood and Williamson (ref. 20) and they have shown that for very smooth surfaces and moderate loads, asperities may still be within the elastic limit, whereas for rougher surfaces or higher loads plastic flow occurs; in both cases the real contact area is very nearly proportional to the load. They thus conclude that this linear relationship is not proof of plastic deformation. It has also been shown (ref. 21), for a multiple asperity model involving purely elastic deformation, that the contact area is proportional to W^m, where m is equal to 26/27.

Greenwood and Williamson (ref. 20) assert that the contact between solids is controlled by what they call a "plasticity index," defined as

$$\frac{E}{p}\sqrt{\frac{\sigma}{\beta}},$$

where the variables are two material properties, the elasticity, E, and the hardness, p, as well as two topographic properties, the radius of curvature, β, of the peaks of the asperities, and the standard deviation, σ, of the height distribution of the summits. There is no doubt that all these material and topographic properties are important in determining the nature of surface contact; however, the value of such an empirical index has yet to be fully determined. The influence of surface films must be considered in any study of solid contact, but no mention is made of this factor. For example, in using the "plasticity index" for a solid with a 60Å film of oxide, should the values of elasticity and hardness used be those of the oxide, or of the metal, or some particular combination? Values of the hardness will depend on the detailed history of the material as well as environment, and will vary with the load because of work-hardening, so that the specific values to be used in the index will be difficult to determine. Considerable experimental evaluation of such an index is still needed.

Solid-solid interactions become even more complex as soon as a tangential force is applied and sliding results. At least three major interacting effects—mechanical, thermal, and chemical—play an important role in sliding interactions, and in practice it is difficult to treat even two interacting effects quantitatively. Ling, among others, has considered the problems of surface temperatures and thermomechanical coupling utilizing the methods of continuum mechanics.

From the point of view of the materials scientist, temperature changes are extremely important in that they may bring about changes in the structure and/or mechanical properties of the materials involved. The sliding interaction of two solids generates heat on the surface of both bodies as a result of the plastic deformation of the materials and the breaking of adhesional bonds. This heat is distributed from the points of contact primarily into the bulk of the solids: the transfer depends on the thermal properties of the materials. Calculations of the temperature rise are very difficult because of the complexity of the boundary conditions imposed on the heat flow equations. Calculations based on simplified models were first made by Blok (ref. 22), and later by Ling, et al. The reliability of these calculations is hard to evaluate because of difficulty in measuring transient surface temperatures, and because each pair of interacting surfaces is different in detail and requires its own solutions of the heat flow equations.

The transient temperatures generated at the points of contact and immediately below the surface may be very high, in fact, under certain conditions may be exceeding the melting point of the material (ref. 1). These high temperatures may bring about phase transformations, recrystallization, oxygen solution into the metal, or enhanced oxide formation on the surface, and decomposition of lubricants, as well as the usual decrease of shear strength and hardness. The bulk temperature rise is, of course, much less than that of the surface but can still be important in terms of the metallurgical characteristics of the solid.

The importance of thermal effects is quite clear and it is indeed fortunate that workers like Ling are making such vigorous efforts toward both the calculation and the measurement of interface temperatures during sliding contact.

In this section on surfaces and surface interactions, I have emphasized the necessity for fundamental studies on clean, well characterized, surfaces. I do not intend to imply that all experiments on other type surfaces are worthless, but it is essential that the multitude of variables involved in studies on "real" surfaces be carefully controlled. The most challenging problem of surface science, and indeed of materials science, is to relate the basic atomic properties of a material, as determined under clean controlled conditions, to its "real" or engineering properties under complex conditions.

FRICTION AND ADHESION

The two well-known laws of friction which state that the friction is proportional to the load and that it is independent of the size of the area of contact, were first enunciated by Leonardo da Vinci (ref. 23) around the beginning of the sixteenth century, and were reformulated and experimentally confirmed by Amontons (ref. 24) at the end of the seventeenth century. These laws are still widely used and must obviously have merit. The question of the current status of friction resolves itself into two other questions: First, how far have we come in the past 500 years in our understanding of the mechanisms of these laws of friction; and second, how generally valid are these laws?

In the period of about 250 years following Amontons' studies, work on friction was primarily concerned with the role of asperities. The French school made the contribution that contact between surfaces occurred only at discrete points. They failed to realize, though, that the role of the asperities was more than just geometric and the unfortunate concept of interlocking asperities due to Coulomb (ref. 25) is still found in some modern texts. The idea that adhesion played a role in friction was suggested as early as 1724 by Desaguliers (ref. 26) in England, but was rejected by most workers in the field until the beginning of this century.

In the past thirty years most workers have accepted the concept that friction between unlubricated surfaces is the result of two major factors: adhesion at the points of contact, and ploughing of one surface by asperities on the other. Assuming essentially no interaction between these processes, the friction force is often written as

$$F = F_{\text{adhesion}} + F_{\text{ploughing}}$$

This may be somewhat misleading because the so-called adhesion term, F_{adhesion}, really has nothing to do directly with adhesion. It is actually a shear term which enters because once the bonds are formed between two surfaces, they must be broken for sliding to occur. This part of the frictional force is usually expressed as $F = As$, where A is the real area of contact, and s is the average shear strength of the weaker material. The applicability of this term has been questioned (ref. 27), and it will be discussed in more detail further below.

I will discuss now the role adhesion plays in friction. The very considerable literature on adhesion has been summarized in several recent books by Houwink and Salomon (ref. 28), Eley (ref. 29), and Weiss (ref. 30). The specific case of metal-metal adhesion has been considered in detail in several recent articles by Keller (refs. 31 to 34) whose results and discussion are particularly apropos.

Consider first an ideal case in which two absolutely clean and atomically flat surfaces of single crystal metal are brought intimately together in perfect crystallographic alignment. Without the addition of external force, bonding or adhesion would clearly take place and the strength of the junction would be the same as the fracture strength of a similarly oriented, well-annealed single crystal. The driving force for the bonding could be considered as the reduction of surface area and hence of total free energy. If the two crystals are disoriented with respect to each other, bonding will still occur, but the fracture strength may be somewhat modified by the presence of a grain boundary. The energy of the system will be modified also by the interfacial or grain boundary energy, but this will be a small effect. A coefficient of adhesion—the ratio of the force required to separate two adhering bodies to the load with which they were pressed together—has no meaning whatsoever in this case since the load is zero. Indeed the coefficient of adhesion is an empirical quantity which has some practical engineering usage, but is of almost no use from a basic science point of view.

The above situation, of course, is ideal and has been approached (and essentially confirmed) experimentally only in A. I. Bailey's studies on mica (ref. 35). Under less idealized conditions, the problem in making adhesion strength measurements is mainly one of determining the actual surface area of contact.

Keller (refs. 31 to 34) has made detailed studies of the adhesion of clean metal couples (Ag, Mo, Ti) and found the junction strength to be in the range of the bulk annealed strength. The actual contact area was determined by mechanical considerations ($A=W/p$) or by very ingenious contact resistance measurements. Similarly, the adhesion strengths of Ag and Ni (immiscible), and Cu and Ni (mutually soluble), couples were in the range of strength of the weaker metal, even at very light loads. Buckley (ref. 36) has also done extensive work with single clean copper crystals showing the high adhesion of clean metal surfaces.

Transmission electron microscopy may be used to study the details of metal interfaces. A typical micrograph by this writer of copper deposited on nickel under ultrahigh vacuum conditions is shown in figure 3. In many cases the arrangement of interfacial dislocations may be readily observed. This is an exciting area for future research.

Keller (ref. 33) found no evidence using clean metals for the breaking of adhesion bonds by the unloading of elastic stresses. This is certainly expected if the metals show sufficient ductility. This evidence does not substantiate the postulates of Bowden and Rowe (ref. 37) that the failure of metallic bonds on unloading is due to elastic forces. It seems probable that the presence of contaminants is very important in helping to bring about a decrease in bonded area on unloading.

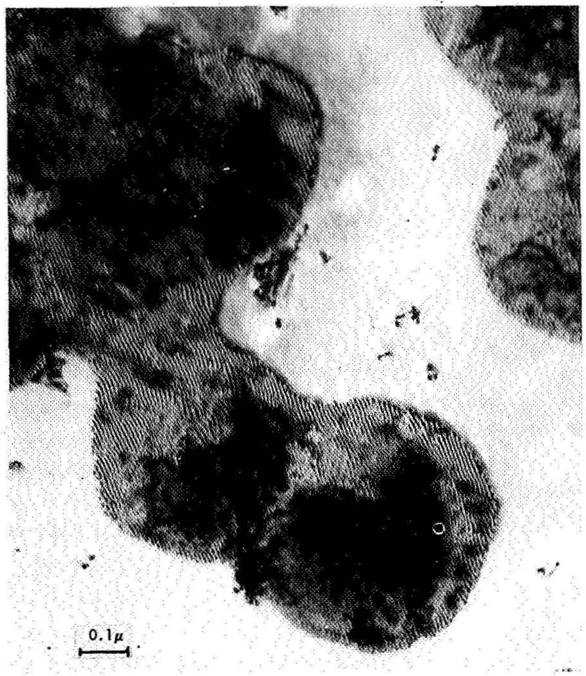


FIGURE 3.—Transmission electron micrograph of adherent nuclei of copper on nickel, prepared by evaporation in an ultrahigh vacuum system, with nickel substrate at room temperature. Moiré fringe pattern shows plainly with many dislocations.

We might ask now what the determining factors are for the strength of adhesion for clean metals. Experimental data are still very scarce. The concept that two different metals which are immiscible in the bulk will show poor adhesion does not seem to be valid. As mentioned above the insoluble Ag and Ni couple shows strong adhesion. Another insoluble pair, Cu and *W*, has been shown to have low adhesion (ref. 36), but results by other workers suggest that the adhesion is quite high (ref. 38). It seems quite clear that we cannot apply bulk criteria of solubility to the surface layer.

What about crystal structure and orientation? Buckley's data (ref. 36) on copper for matched crystal planes show an apparent correlation of adhesion coefficients with the modulus of elasticity. In light of Keller's results on the effects of released elastic stresses, however, this correlation is not meaningful; in any case, in view of the high ductility of copper, we would not expect the elastic stress relief to break the bonds. Metallic bonding shows little or no directionality and the actual bond strength would be very nearly the same for different orientations of the bonded crystals. What is being measured in determining the adhesion coefficient of clean metals, is really some function of the ultimate tensile strength, and it will vary in a complex way depending on the orientation of the crystals. Buckley (ref. 36) finds the adhesion coefficients for three pairs of similar planes of copper to be in the order $(100) > (110) > (111)$, which is, as it has been suggested in this symposium, because the spacing between adjacent planes in the face centered cubic system is in the order $(111) > (110) > (100)$. The argument is that the force to rupture junctions in surfaces parallel to (111) should be less than that for the (100) simply because the spacing between (111) planes is greater than the spacing between (100) planes. This is, of course, not correct. We are concerned here initially with a problem of resolved shear stresses on (111) type slip planes; with extensive interaction of different slip systems the analysis becomes very complex—not a simple pulling apart of planes. The rate of work-hardening is in the order $[111] > [100] > [110]$. We might expect fracture to occur at a lower stress for adhered (111) planes, but the order is not right to explain the (100) and (110) coefficients. Very few detailed studies have been made in which a single crystal with a given orientation was pulled to fracture. The analysis of such an experiment would be very difficult because of the interacting slip systems, the rotation of the lattice, and the resulting complex tangles of dislocations.

When two dissimilar planes are bonded the values of the adhesion coefficient are considerably less than the values for either of the planes concerned (ref. 36). This was interpreted as an indication that the junction between unmatched planes is more readily ruptured than that between matched planes because of the presence of a grain boundary. However, grain boundaries may be stronger than the grains themselves in many cases so the explanation is by no means simple.

It has been suggested by a number of workers that hexagonal close-packed metals should exhibit lower adhesion characteristics than cubic metals. Recent evidence is conflicting: data on cobalt (ref. 39) showing relatively low adhesion, and data on titanium showing junction strengths of the order of ½ that of the bulk metal. It is likely that contaminants affected both of these results, and in fact even with hexagonal metals we should expect the bond to show the bulk strength of the metal as suggested by Keller (ref. 32).

It seems likely to me that the most important factor in the specific value of the adhesion coefficient as measured under clean conditions is the rate of work-hardening. Work-hardening varies widely with orientation, and is influenced by the presence of grain boundaries. The study of the processes leading up to fracture should be a fruitful, but very complex, area for further research, even though it is of minor importance to the problems of adhesion per se.

While the adhesion of clean metal surfaces provides a fine basis for theory, the great mass of data on adhesion has been obtained from surfaces which were grossly contaminated. If true metal-to-metal adhesion occurs, obviously the contaminant layer must be removed in the spots where bonding takes place. The contaminant layer, in most cases, may be broken or pushed aside by the normal load on the asperities. The details of rupture, though, are not clear. The film may be brittle or ductile and its rupture is undoubtedly affected by the nature of any slip processes occurring in the underlying metal. Whatever the method of film removal, it is very difficult to visualize a mechanism by which the metal-to-metal contact area would even begin to approach the contact area given by the equation $A = W/p$ for normal loading.

Relief of elastic stresses on unloading would certainly break any bonds which might be present between the oxide or contaminant layers on the two surfaces—assuming these bonds are weaker than the metallic bonds. Another possibility is that the surface film may be intermixed with the metal and incorporated as impurities in the boundary. The bonding forces would be complex in this case and a theoretical evaluation of the strength of the adhesion would be very difficult. Another situation which may lead to removal of oxide or certain other contaminants is heating of the surface, whether by friction or other means. Certain oxides such as NiO or Cu_2O may dissolve in the metal; other oxides such as PtO_2, WO_3, and Ag_2O are unstable above certain temperatures and decompose; thus either of these mechanisms may remove oxide films from the surface, usually, however, leaving an adsorbed monolayer of oxygen on the surface. It seems very likely that adsorbed monolayers could be incorporated in the boundary of an adhesive bond, but this, of course, needs experimental verification.

Sikorski (ref. 40) has obtained extensive data on the adhesion of metals in air, using a twist-compression type experiment. These experiments differ from those considered previously in that both normal and tangential forces were applied. Coefficients of adhesion determined this way were compared with such physical properties of the metal as crystal structure, hardness, surface energy, elastic modulus, work-hardening, recrystallization temperature, purity, and atomic volume. This type of experiment provides a shearing force which can create fresh surfaces of metal that readily bond. It is greatly complicated by the mixture of oxide, contaminants, and metal formed in the interface region during the bonding process, as observed by Anderson (ref. 41). The nonreproducibility of the coefficients of adhesion also complicate matters. These results show, though, that metals with high hardness, elastic moduli, recrystallization temperature, and surface energy show low coefficients of adhesion; they also provide useful data for the practicing engineer. The apparent correlations are very important, but similar studies under better controlled surface conditions are necessary for confirmation.

The overall picture of adhesion in air is even more complex than I have indicated so far. Consider two polycrystalline metal surfaces covered with low angle asperities of different sizes and orientations, all of the surfaces being covered by an irregular polycrystalline oxide, perhaps 20 to 60Å thick; the oxide itself is covered with a mixed layer of adsorbed gases and particularly water vapor. Suppose we press these together with a load sufficient to cause metal-to-metal bonding, however this happens, at some of the asperities, so that we have adhesion of a sort. Now we can pull these two metals apart and measure the force of fracture, and then try to analyze the results. We need an area of contact to use in our evaluation, so what do we use? Perhaps we use a contact resistance measurement, or calculate the area from $A = W/p$, but is this really the area we want to use? It would seem not, because in this complex case we most likely have bonding other than at metal-to-metal junctions: surface tension forces in the narrow regions between the two surfaces, oxide-to-oxide bonding, metal-to-oxide, or general dispersion force bonding of metal-contaminant-metal. This all seems too complicated even to think about. But let us try to complete the analysis assuming a reasonable value of the area, and no bonding other than metallic. Look at the metal-to-metal contact closely. Because of the nature of contact of asperities we have a built-in crack or notch at each contact area. Now we need to know the fracture mechanics involved, and this depends on a multitude of new variables related to the bulk properties of the solid, dislocation mechanisms, impurity effects, orientation effects, and environmental effects; all must be considered. We have probably by now given up in despair any hope of completely analyzing what is meant by a simple determination of a coefficient of adhesion. Obviously to obtain fundamental data for understanding this complex process, we must simplify our experiments.

It seems clear that data from two types of adhesion experiments are needed. The first type involves the determination of the strength of adhesion using ultra clean conditions, such as those used by Keller (ref. 33) and Buckley (ref. 36) and single crystals. Such variables as load, bulk crystal structure, crystal orientation, surface energy, rate of work-hardening, solubility, purity, elastic modulus, and surface structure can only be properly investigated under these refined experimental conditions.

The second type of experiment is really a modification of the first and involves the same type of experiment on single crystals, but with a controlled layer or layers of contaminant on the surfaces. This experiment, though difficult, is essential to our understanding of "real" surfaces. Most metal surfaces are covered with an oxide film and much of the contact may be of the unknown oxide-to-oxide type. Oxide films of controlled thickness and orientation can be prepared on many metals, and the adhesion of such surfaces could be readily studied over a range of loads that go up to and beyond oxide fracture. This would be a fruitful area of research.

The theoretical basis of adhesion is currently not very good because of the lack of well defined, reliable, experimental data, and because of the analytical complexity. Surface energy considerations have been used by several workers, but with almost no reliable solid surface energy measurements available, this does not seem a profitable line of attack (ref. 32). Dispersion force mechanisms have been suggested (ref. 42), but it is unlikely that these are important for metal-to-metal contact, although they may be important for adhesion between contaminant layers. The use of regular solution theory applied to interfaces has been suggested (ref. 32) and shows some promise worth following up in future research.

Having considered the current status of the metal-to-metal adhesion problem, we are now in a position to examine friction. The friction force is usually expressed as

$$F=As+P,$$

where A is the real area of contact, s is the shear strength of the weaker material, and P a deformation term due to the ploughing action.

The relative influence of the two friction terms depends on several factors. For a given load on relatively smooth surfaces, the adhesive term will be most important, and with increasing roughness of the surface the ploughing term increases in importance. For a given size indentor (or asperity), ploughing action is favored for higher loads. Contaminants will also play a role.

Consider now a typical friction experiment performed in air with surfaces which are sufficiently smooth that we can neglect the ploughing term. The surfaces are complicated by low-angle asperities, oxide, and adsorbed gas. When a normal load is applied, deformation of the surface asperities will take place so that the actual area of contact is approximately proportional to the load. At this point there will probably be little if any pure metal-to-metal contact, but some adhesion will result nevertheless. The shear strength of these joints will generally be weak compared to that of metal joints.

Since $A=W/p$, approximately, we can substitute this into the equation $F=As$, obtaining

$$F=\frac{s}{p}W.$$

This essentially explains Amontons' laws that the friction force is proportional to the load and independent of the area. The ratio F/W, which equals s/p, is defined as the coefficient of friction.

Several problems arise immediately. The first concerns the values of s and p to use in the equation for the coefficient of friction. If we assume no metal-to-metal contact we really have a lubrication problem and we need to know the shear strength for the lubricant—in this case oxide, or adsorbed film. These types of data are almost nonexistent. The indentation hardness, p, is related in an extremely complex way to the yield strength. If two different metals are involved, we are talking about the hardness of the softer metal. The value of p will be difficult to obtain, since once the surface is loaded some work-hardening will occur (p is some unknown function of the load). Calculations of friction coefficients in this case seem out of the question.

A second problem appears if the surface film is broken during the sliding process so metal-to-metal bonding occurs. Some sort of average shear strength for the interface must then be used and this is another unknown.

If we assume a situation with considerable metal-to-metal contact after normal loading, apply a tangential force, and measure the friction, the same equation for the friction force should apply as in the previous case. Here, however, s is the shear stress of the weaker component of the junction and p is the local indentation hardness of the softer metal. Calculation of the coefficient of friction using reasonable values of s and p yields a result which is far too small.

What is the reason for this failure of the theory? It is clear that the model adopted assumes that the normal yield stress and the tangential shear stress are independent strength properties. Experimental evidence (ref. 43) and plasticity theory show that this assumption is not correct and, indeed, plastic yielding of the contacts occurs under the combined effect of the tangential and the normal stress.

In almost all practical cases of unlubricated surfaces, when a tangential force is applied, the combined normal and tangential stresses at the point of contact lead to junction growth before macroscopic sliding begins. Since the new contact area seems to be proportional to the original static contact area (ref. 21), Amontons' laws still seem to hold for contaminated surfaces. However, for clean surfaces, junction growth may continue until gross seizure occurs and friction rises to a very large value.

The phenomena of junction growth and junction failure are very complex and there is a great need for detailed studies in this area. As sliding begins, the deformation about the areas of contact will generally produce severe work-hardening. The critical shear stress will tend to vary in a very complicated manner from point to point on the surface. We cannot hope to understand the mechanisms involved without detailed studies of the dislocation distributions about the junctions. These can, of course, be studied most readily using single crystal specimens. The mechanism by which a contaminant film stops junction growth is by no means clear.

The deformation or ploughing term is generally thought to make only a small contribution to the total coefficient of friction in most sliding experiments; e.g., 0.1 to 0.2 (ref. 21). Asperities on polished surfaces usually have a slope of the order of 5° to 6° and so their contribution is considered small. Even in special cases where a sharp pointed cone is used and the initial contribution of the deformation term may be large, with repeated traversals, the track width grows, the metal work-hardens, and the ploughing term decreases. Perhaps more important than the ploughing term by itself is its interaction with the adhesion or shear term. Only a few studies have been made of this important interaction (ref. 21), but it seems likely that the interaction will serve to bring about an increase in the deformation term. A detailed study of this interaction is badly needed.

A number of very important sliding experiments in which the ploughing term was predominant have been carried out using single crystal surfaces and a hard slider (refs. 44 and 45). These investigations have shown very clearly that the friction depends markedly on crystal plane and the crystallographic orientation. Steijn (ref. 45) found that the coefficient of friction was a maximum in the [110] directions and minimum in the [100] directions on the cube face for a number of materials of the face centered, body centered, and NaCl type cubic structures. This seems strange at first sight because in all three of these crystal structures plastic deformation occurs by different slip systems. The explanation lies in the experimental fact that a hillock is pushed up ahead of the slider in the [110] azimuths, whereas two hillocks to either side of the front of the slider are present for the [100] azimuths. Figure 4 shows interferograms by Dyer (ref. 46) for a ball rolling in the [110] and [100] direction on a (100) surface of copper illustrating the same effect. This hillock formation on crystals other than f.c.c. metals is not completely understood in terms of dislocation mechanisms, despite the fact that a number of theoretical studies have been made of the dislocation patterns about indentations. More work of this type is desirable.

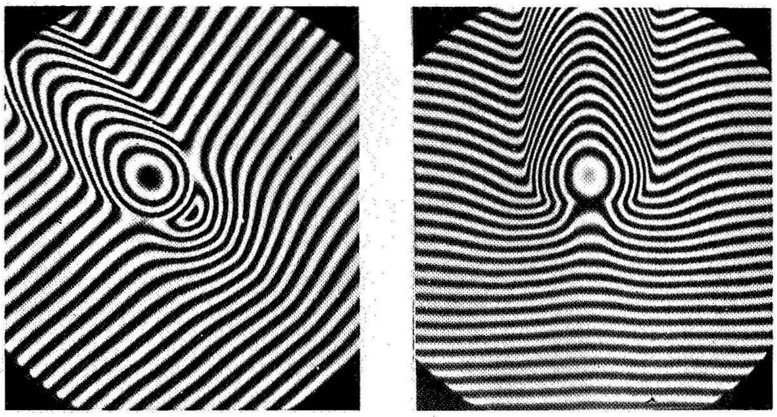
Another feature observed in these single crystal studies is the variation with sliding direction of track width; this is also apparent in figure 4. The track is narrow in the high-friction direction, partially as the result of work-hardening, and partially as a result of the ridge of material pushed up ahead of the slider. Surface and bulk deformation around wear tracks in single crystal copper have been studied in some detail by J. M. Bailey (ref. 44), and dislocation distributions have been determined by metallographic techniques.

Transmission electron microscopy is a powerful technique for studying deformation processes in solids, but it has not yet been applied to detailed studies of friction processes. As an example of the potentialities of this technique, I will show some results from a crude friction experiment carried out eight years ago on a (100) oriented single crystal copper film.

A small hemispherical sapphire cone under a very light load was slid across a 1000Å single crystal (100) copper film which was grown epitaxially on a rock salt substrate. Two tracks were made: one sliding in the [001] direction, and the other in the [011] direction. The copper film was then removed from the rock salt by dissolving the substrate and examined by transmission electron microscopy.

Relatively low magnification micrographs such as those shown in figures 5 and 6 for the [001] and [011] sliding tracks give an overall picture of the area deformed in relationship to the surrounding area. Very little deformation is seen outside the actual track area, and this seems to be greater for the $[0\bar{1}1]$ sliding direction. Deformation faults are visible in both micrographs, and it is apparent that rotation of segments of the crystal lattice have occurred. This is most obvious for the [011] direction of sliding.

An interesting and perhaps important observation is the enhanced



(A) (B)

Figure 4.—Interferograms from rolling friction tracks on a (001) copper surface. (A) [110] track end. (B) [100] track end.

formation of oxide in the track area, visible as the multitude of small black specks in the micrograph of figure 6. Diffraction patterns can be readily obtained to give detailed information on the occurrence of break up and rotation of the lattice.

Higher magnification micrographs can give much more detailed information on the types of deformation produced. Figure 7 shows a heavy density of dislocations, mostly arranged as tangles in walls parallel to the sliding direction, and deformation faults, apparently both twins and stacking faults, formed on (111) planes which originally intersected the (100) plane in the [011] direction. The beginning of cell formation is readily seen in the micrograph. If the deformed area is heated, the oxide dissolves, and the cell structure sharpens up as shown in Figure 8. Finally, Figure 9 shows several deformation faults extending into the area outside the track and the marked rotation occurring within the track.

I want to emphasize that, despite the qualitative nature of the results presented here, detailed and quantitative studies of the deformation associated with the friction process can be made using this technique. Friction induced deformation on bulk samples can be studied equally well using well-known thinning techniques to get suitable samples. The availability of high-voltage electron microscopes capable of examining samples 2 to 3 microns thick makes transmission electron microscopy even more attractive for such studies.

Interaction between the adhesion and ploughing terms is often considered negligible, and as has been shown it is possible to study the ploughing term under conditions where adhesion is negligible; however, it does not seem likely that these terms can be separated in any case where metal-to-metal adhesion occurs. In almost any practical case of this sort, the normal load will produce some plastic deformation and hardening at the surface of the asperities, so that when a tangential force is applied, shear may be more complex in the surface than in the bulk. The dislocation interactions involved here will be terribly complex; and more detailed studies are necessary for an understanding of bond fracture.

The complex interactions of adhesion and ploughing terms must be analyzed on clean surfaces in order to obtain the data necessary to understand the friction process. The more complex situations can be studied with carefully controlled surface films of known structure.

The discussion above has been limited entirely to metals and their associated surface films. Lack of space and time prohibits a discussion of friction on polymeric, ceramic, or composite materials. Basically the same types of information are needed though, if we are to understand friction on them. Adhesion forces here may be even more complex, and here again, bonding should be studied under clean conditions.

For nonmetals further study is needed to understand mechanisms of deformation and shear, interactions of normal and tangential stresses, and the important role that elastic forces undoubtedly play in the friction of polymers.

WEAR

Wear has been defined as "the unwanted removal of solid material from rubbing surfaces." It is an even more complex subject than friction, and our knowledge of the fundamental mechanisms is correspondingly more meager. This is indicated by the fact that the type of wear depends not only on the basic properties of the rubbing materials, but also on the load, the speed of rubbing, the surface roughness and the environment. The complexity is further indicated by the multitude of different classifications of wear mechanisms (ref. 47).

Burwell and Strang (ref. 48) have suggested the following classification, and this seems like a good place to start looking closely at the status of wear:

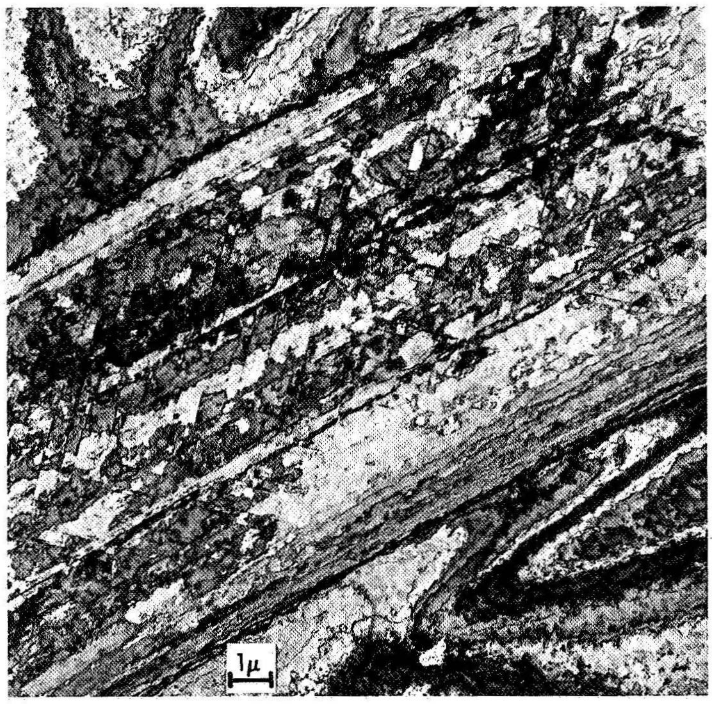


FIGURE 5.—Wear track formed by sapphire stylus sliding in [001] direction on the (100) face of copper single crystal.

FIGURE 6.—Wear track formed by sapphire stylus sliding in [011] direction on the (100) face of copper single crystal.

(1) Adhesion or galling
(2) Corrosion
(3) Loose abrasive particles
(4) Cutting or ploughing by a harder rough surface
(5) Erosion and surface fatigue

A further somewhat arbitrary classification into "severe wear" and "mild wear" is often made on the basis of wear particle size. In general, for any particular example of wear, any of the above mechanisms may operate singly or more likely in combination, further complicating the analysis. In some cases, as indicated by Archard, multistage processes of wear may be operative, and it may be wiser to adopt a classification based on individual events rather than complete wear mechanisms. For the purposes of this discussion, however, I will follow the above classification.

The need for well-defined experiments on adhesion of various materials has already been vigorously emphasized. Adhesion is clearly the first event taking place in many wear processes. It is, however, important to realize that mere adhesion may not lead to a wear particle if the shear strength of the junction is less than that of the bulk material on either side of the interface. If the shear strength of the adhesive junction is greater than that of the bulk material on one side or the other of the interface, then transfer of material can take place. This transfer has been demonstrated for many different types of materials and often transfer of both harder and softer materials occurs. Experiments demonstrating the mode of deformation near junctions have been made by Cocks (ref. 49); they show that the junction in the case of metals does not break when sliding first starts,

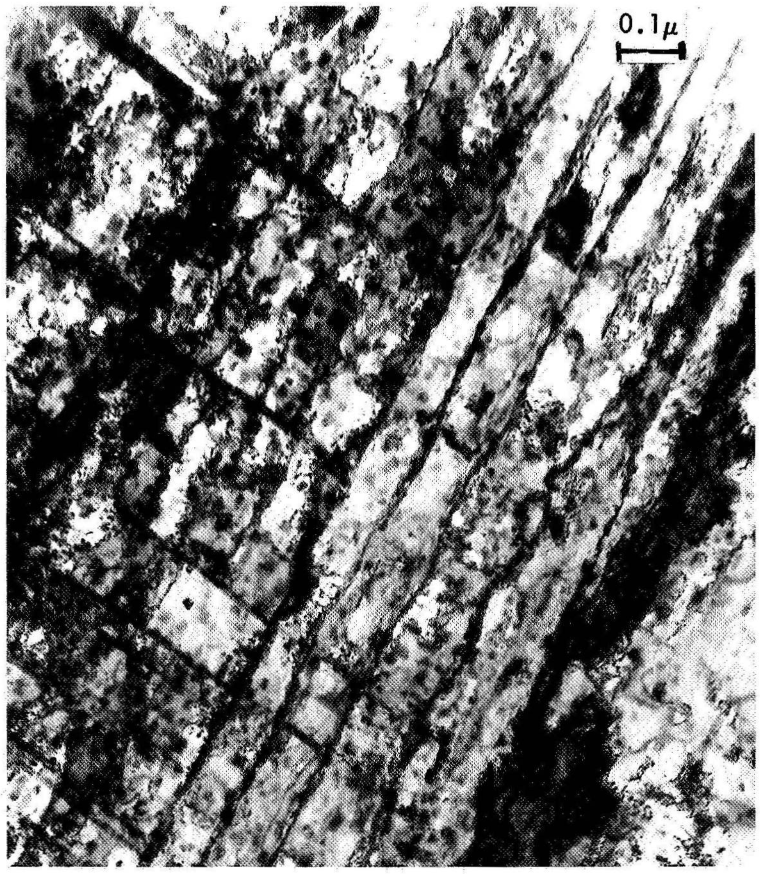


FIGURE 7.—[011] deformation track showing dislocation tangles and beginning of cell formation.

FIGURE 8.—[011] deformation track after heating showing sharpening up of cell structure.

but that plastic deformation of the material adjacent to the junction takes place, and sheared metal is then forced up from the surface. More detailed studies with controlled single crystal specimens of different orientations are highly desirable.

Even though adhesion can lead to material transfer from one surface to another as the first event of a wear process, it seems clear that adhesion in itself does not produce the characteristic loose particles of wear. How then are these produced? A number of complex mechanisms have been proposed.

It has been suggested in the case of metals that oxidation of the adherent fragments occurs; the oxide, being loosely bound, then comes off as wear fragments. There is no doubt that oxidation occurs in air, but it is not certain that the oxide will be loosely bound. Unless the temperatures generated in the sliding process are quite high, for most materials the thickness of oxide formed in a reasonable length of time may be under 10 000Å. Such films formed on clean surfaces in a dry environment are very adherent. If, on the other hand, the environment is one in which contaminants such as oils are adsorbed on the freshly sheared surface before oxidation occurs, the oxide may be very loosely held. If the temperature is high, a thick oxide may form and the residual stresses in the oxide may cause spalling. In certain corrosive environments loosely adherent hydroxides or hydrated oxides may form on certain materials. There is, however, no really good experimental evidence to substantiate the removal of

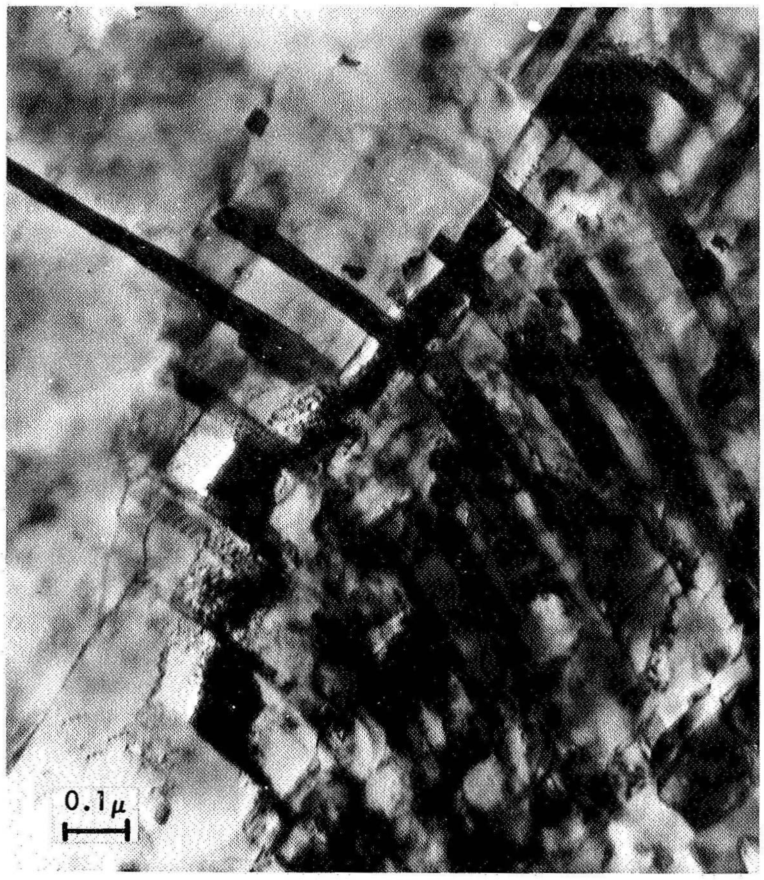


FIGURE 9.—[011] deformation track showing deformation faults and lattice rotation.

loose oxide as a general mechanism. The fact that loose oxide particles are often found on wear surfaces does not provide evidence of how they got there.

A possibility is that oxide particles are produced by a fatigue process in the oxide layer. As indicated previously, studies of the mechanical properties of oxide layers on metals deserve a high priority.

It has been suggested that wear particle formation is related to the ratio of the work of adhesion, or surface energy, to the penetration hardness. Rabinowicz (ref. 50) has presented evidence to substantiate this criterion, but the correlation for metals, as an example, is not very good. Applicability of this expression depends on the availability of reliable surface and interfacial energy data.

Another suggestion (ref. 51) is that wear particles are formed by a fatigue process in the substrate material. A considerable amount of evidence in support of this has been accumulated (ref. 47). Most of this evidence, however, is indirect and there seem to be very few metallographic studies which might confirm such a mechanism. Fatigue studies are terribly complex, but it will be necessary in the future to examine closely the deformation processes associated with solid junctions subjected to successive compressive and tensile stresses.

Corrosion and oxidation obviously play important roles in wear. Oxide formation on freshly exposed surfaces in many cases leads to a change from severe to mild wear. Apparently the role played by the oxide is in preventing the strong metal-to-metal adhesion of severe wear. The details of oxide formation are not firmly established even for well-controlled oxidation studies, and are not at all understood for surfaces undergoing wear (this is commonly called fretting corrosion). The presence of loose oxide particles on a wear surface may be due to fatigue in the oxide, to spalling, or to metal fragments formed by fatigue and subsequently oxidized on the surface. In fact, all of these processes may occur during a single wear experiment, and it will be very difficult to determine the relative importance of the different mechanisms. The actual wear may depend on the hardness of the oxide, but no evidence relating wear to the mechanical properties of the oxide is available.

Corrosion, and particularly stress corrosion and corrosion fatigue, may play important roles in the creation of wear fragments under special conditions. In these cases the fracture of the metal is enhanced by the presence of certain corrosive environments. Stress corrosion does not occur in pure metals and may be related to the presence of short range order in alloys. The cause of the localized chemical attack and the enhanced mechanical failure have been the subject of intensive investigation for many years, but the detailed mechanisms are obscure and still highly speculative.

Controlled experiments on wear in various chemical environments are obviously desirable. Particular attention should be paid to the structure and mechanical properties of surface films formed during these processes.

The presence between two surfaces of loose abrasive particles, whether they are oxide, dust, or hardened metallic fragments, can lead to a ploughing action which can remove material from one or both surfaces. A similar form of wear occurs when a rough hard surface slides over a softer surface. Obviously severe deformation of material below the surface can also occur, and with repeated application this may lead to fatigue failure. Abrasive wear may be made almost negligible at the beginning of a wear process, but as wear proceeds, the generation of loose particles may enhance abrasion. It is apparent that if the wear particles are sharp and jagged a ploughing or cutting action may be important, whereas if they are smooth-edged or rounded, subsurface deformation and fatigue may be more important. Abrasive wear depends primarily on the relative hardness of the abrasive body and the surface; it also depends on the crystallographic orientation in the surface. The elastic modulus is also important and it has been suggested that the ratio of the hardness to the elastic modulus may be the most relevant physical property, but this needs confirmation.

The necessary experiments have already been suggested in connection with the ploughing term. Electron microscope studies should be very important for revealing the detailed nature of the abrasive process, but only a few such investigations have been made. In particular, transmission electron microscope studies of samples after controlled abrasive wear should be revealing.

Analytical expressions for the so-called wear rate have been derived by a number of workers. These are usually expressions for the volume of material removed per unit sliding distance, or the ratio of the height (or depth) of the worn layer to the unit distance of sliding, and are not really rates. Archard has derived an equation for the magnitude of the wear, assuming that the unit event of importance is the contact of two asperities on opposing surfaces with a certain true area of contact. It is also assumed that the wear particles are equiaxed, or that the particle and junction diameters are equal. This leads to an equation of the form

$$\frac{V}{L}=KA,$$

where V is the worn volume, L is the sliding distance, and A is the true area of contact. The proportionality constant, K, has been interpreted as the probability that any given event will produce a wear particle, and experimental values of K range from about 10^{-2} to 10^{-7}. This interpretation has been questioned, however, and one of the major problems in wear today is the physical explanation of K.

On the assumption that the area of contact is determined by plastic flow of the asperities and that we can write

$$A=\frac{W}{p},$$

the wear equation can be expressed as

$$\frac{V}{L}=K\frac{W}{p}.$$

This shows the direct proportionality of wear to the load, W, which has been observed in many cases. It also shows the inverse proportionality of wear to the hardness which seems to hold for some materials, but which is not generally true. The independence of the apparent area of contact is also shown.

The assumption of equiaxed particles made in deriving the wear equation is questionable. In fact other equations have been derived on the basis of different assumptions for wear particle shape. The complexities of wear particle shape may indeed make an analytical description of wear essentially impossible. Other theories have been based on the removal of an oxide layer, assuming an oxide film thickness given by logarithmic or parabolic growth-rate equations. Such theories are subject to considerable doubt mainly because the oxidation process during wear is too complex to be represented by a simple equation. In particular oxidation will vary markedly from spot to spot as determined by transient temperatures, contamination effects, crystal orientation, and surface deformation; it is obviously not possible to assume a constant oxide thickness over even a small portion of the surface at any given time.

There is a great opportunity for fundamental studies. The complexity of wear, however, adds to the problem of designing definitive experiments. I believe, though, that very simple experiments can be designed similar to those done by the late Professor Allan T. Gwathmey on single crystal surfaces. These can confirm, with the use of modern electron optical tools, some of the basics of wear which are still highly speculative.

LUBRICATION

Rabinowicz (ref. 50) has defined a lubricant as "a substance which is capable of altering the nature of the surface interaction between contacting solids." A distinction is usually made between "fluid lubrication," where a thick film of gas or liquid completely separates two solids so that they do not touch, and "boundary lubrication," where much thinner films of solid, gas, or liquid are interposed in such a manner that solid-to-solid contact occurs at random points. I have included so called "solid lubricants" in the "boundary lubricant" category.

Boundary lubrication involves problems of surface film formation and their physical properties. These films may be partial monolayers of a gas, or they may be oxide films 1000A or more thick. They may consist of an ordered monolayer of a long-chain organic molecule, or they may be piled-up disoriented multilayers of an organic ester. The chemistry of lubricants is varied and will not be discussed here, except where reaction of the lubricant with the solid surface is involved. Surface films are formed in three ways: First, physical adsorption; second, chemisorption; and third, chemical reaction. These three processes have already been considered. In most practical lubricating systems for metals, oxide is present on the surface. The presence of different lubricants may influence oxide formation. Some lubricants may prevent oxygen from reaching the surface, and this in turn may lead to enhanced friction and severe wear if the oxide is necessary to prevent metal-to-metal contact. On the other hand, contaminants or lubricants may affect the adhesion of oxide to the surface so that a loose, easily removed oxide is formed which favors mild wear. The orientation of the oxide may be very important in many cases.

Although I have already discussed adsorption, I mentioned nothing about the adsorption of complex compounds such as organic long-chain molecules. It is frequently difficult to make a sharp distinction between physical and chemical adsorption, and more work is needed with complex molecules. The behavior of an adsorbed film could depend critically on how tightly bound or mobile the molecules are on the surface. In general, most polar organic molecules are bound very tightly to a metal surface, probably by chemisorption, but more quantitative data on the strength of the adhesive bond are badly needed.

There have been a number of studies on orientation and the crystallinity of long-chain organic compounds on metal surfaces, and there seems to be a good correlation of low friction with high orientation and crystallinity of the film. A sophisticated study by electron diffraction of the structure and orientation of long-chain fatty acids on single silver crystals has been carried out by Mathieson (ref. 52). This is exactly the type of study that we need more of to obtain fundamental data on boundary lubrication.

Organic or inorganic substances can chemically react with a solid surface to form a strongly adherent film. A typical example would be a soap formation by a fatty acid reacting with a metal surface. Chemical reactions with typical lubricants may, however, be complicated by their impure nature. This makes a fundamental analysis of mechanisms essentially impossible.

The nature of the thin films formed on solid surfaces is the first important question to be answered. What is the structure, the orientation, and the composition of the films? Once we have such data, it becomes necessary to determine the physical properties of these films.

Godfrey has emphasized the importance of three physical properties of boundary lubricants: melting point, shear strength, and hardness. There is good evidence that film shear strength determines friction in accordance with the theory that friction is a function of the ratio of the film shear strength to the plastic yield pressure of the underlying solid. Unfortunately, experimental data on the shear strength of thin film lubricants is rather meager. The shear picture is complicated, in the case of some lubricants such as fatty acids, by the fact that the molecules are bent rather than sheared during sliding. This definitely needs further study. The location of the shear plane is critical, and for most lubricant materials this is not known. It will be a function of the structure of the lubricant material, the strength of adhesion to the solid substrate, the strength of bonding between the lubricant molecules, and of the orientation with respect to the substrate. Such knowledge is essential to our understanding of boundary lubrication.

In general solid lubricant films are greatly superior to liquid films, because they apparently provide better separation of the asperities and prevent or minimize direct metal-to-metal contact. Considerable evidence supports the idea that lubricant failure occurs by melting. This correlation, however, has not been established for lubricants melting above about 150° C. More work is needed here, but it will be hampered at higher temperature by oxidation and decomposition.

Although many useful predictions can be based on our current knowledge of boundary lubrication, no comprehensive theory exists. We know that shear in the lubricant is necessary, but we have no knowledge of where it occurs, and no good data to base a prediction on. Melting points can be determined, but the actual melting points under conditions of high pressure are not generally known. It is difficult to derive general rules for boundary lubrication because each lubricant-substrate pair has its own specific properties. A great deal of basic work is needed.

CONCLUSION

It is apparent from the thorough in-depth reviews and discussions presented by the other speakers that considerable progress has been made in understanding the mechanisms of friction and wear. The general picture is not one, however, to generate overconfidence in our ability to predict the detailed behavior of a pair of sliding bodies. Amontons' laws seem to be followed generally, but they are only approximations for which the limits are not satisfactorily known. The great complexities of wear and lubrication have greatly limited the amount of basic data; our understanding in this area is not as good as in adhesion and friction. In the practical case, adhesion, friction, wear, and lubrication are all intimately tied together in very complex ways. Our only hope of understanding this situation is to untie some of the complex knots of interaction and study simpler systems one at a time.

It has not been possible in this paper to consider many important aspects of friction and wear. Virtually nothing has been said about nonmetals, but I think it can be safely said that the same type of information is needed for these materials as for metals. The fundamentals of adhesion, the nature of the deformation during sliding, and the effect of the environment all must be known, whatever the type materials concerned. It is clear that there is a marked scarcity of fundamental data and that carefully designed experiments on well characterized systems are essential.

A number of suggested areas for research have been given. The more important research projects are:

(1) Experimental studies of adhesion under clean conditions. (This should include studies of adhesion between different types of materials such as metals, polymers, and ceramic or inorganic materials.)

(2) Experimental studies of adhesion with controlled and characterized adsorbed species on the surfaces

(3) Theoretical studies of the effects of various adsorbed species on adhesion

(4) Continued theoretical and experimental studies of the adsorption of gases and complex molecules on clean surfaces

(5) Detailed experimental studies of the deformation processes associated with both adhesive bond breaking and ploughing (A knowledge of the dislocation distributions and interactions is essential.)

(6) Experimental studies of the mechanical properties of thin oxide films

(7) Experimental studies of the effect of environment on the mechanical properties of different solids (This is one of the most important areas of research, and includes stress corrosion, Rebinder effects, gas adsorption effects, and the mechanism of fracture in various environments.)

(8) Metallurgical studies involving the influence of structure, orien-

tation, alloy formation, surface energy, and work-hardening on a variety of materials under well-characterized conditions

DISCUSSIONS

G. E. Hollox and R. G. Lye (Research Institute for Advanced Studies, Martin Marietta Corporation, Baltimore, Maryland)

Because of the complex nature of friction and wear, the empirical characterizations have evolved in terms of parameters, such as the coefficient of friction, that can be readily measured in engineering practice. Yet, as Professor Lawless points out, these parameters represent the complex resultant of contributions from a variety of physical phenomena. Because changes in one of the contributions may lead to significant changes in the others also, the resultant is not usually a quantity that can be predicted with confidence. Nevertheless, a sufficient body of information has been accumulated to permit the design of equipment using empirical procedures and well-known values for the pertinent parameters. Unfortunately, however, this approach is less effective when unconventional materials are used in the construction, or when the apparatus must serve in novel, inhospitable environments. Consequently, the need is felt for greater understanding of the mechanisms in friction and wear and of the properties of contacting surfaces.

From the content of papers presented at this symposium, it is evident that friction and wear are influenced by mechanisms and properties which come under the headings of a variety of disciplines. Even when attention is restricted to the materials aspects of these phenomena, however, an extensive area of work remains to be considered. The scope of pertinent materials research as indicated by Professor Lawless and previous speakers may be summarized:

(1) Bulk characteristics:
  (a) Mechanical properties (elastic moduli, plastic deformation modes)
  (b) Thermal properties (thermal diffusivity, melting point and phase transformations, intrinsic and extrinsic diffusion coefficients)
(2) Surface characteristics:
  (a) Structure of clean and contaminated surfaces
  (b) Surface energy and its dependence on contamination
  (c) Chemical reactivity, physical and chemical adsorption

Professor Lawless has considered most of these but it may be profitable to emphasize some specific aspects concerned with the mechanical behavior and surface properties.

Mechanical Properties

Deformation of surfaces.—The behavior of solid surfaces under the stress applied by a moving stylus remains a considerable problem. Recent work, by Bowden and Brookes (ref. 53), for example, has attempted to relate the extent and character of the deformation and fracture along stylus tracks on single crystal specimens to the magnitude and distribution of resolved shear stresses generated by the stylus. Subsequent studies of this kind, under other conditions of temperature, surface contamination, and loading, will make significant contributions. It may also be profitable to include more detailed studies of material in the vicinity of the stylus track. For example, the macroscopic distribution of dislocations can be investigated by the Lang technique of X-ray diffraction, whereas local dislocation groupings can be examined using electron microscopy. The two approaches in combination should be of considerable value for interpreting the response of crystals to a moving stylus.

Deformation of asperities.—Although the behavior of a surface under a hard indenter has been studied in some detail, relatively less is known concerning asperities, which contribute importantly to the properties of real surfaces. In particular, there is the question of whether asperities exhibit properties that are characteristic of the bulk, or behave more nearly like single crystal whiskers. An adequate answer to this may be difficult to obtain, but it appears to be extremely important. An indication of the behavior of metal asperities is provided by the detailed topographic studies of aluminum surfaces discussed by Dr. Williamson. Because the outermost tips of the asperities deformed plastically before their roots exhibited obvious changes in shape, it appears likely that they possess the normal (work-hardened) properties of the bulk aluminum.

If the asperities do have properties essentially like those of the bulk, the familiar methods of the materials sciences can be used to establish pertinent values for the material parameters. Moreover, their deformation behavior should be amenable to experimental investigation and analysis by employing enlarged, idealized models of single asperities, with various mechanical constraints at their bases to represent the effects of the bulk and adjoining asperities. The approach followed by Rowe (ref. 54) may be adaptable to such purposes. Rowe employed tensile test specimens with a region of reduced diameter at their midsection and investigated the effects of prior work-hardening on their deformation and fracture under tensile loading. In a similar fashion, certain properties of junctions and asperities could be investigated under conditions similar to those encountered in friction and wear by studying the response of specimens to appropriate combinations of shear and compressive normal loads. By employing single crystals oriented in various ways, as Professor Lawless has suggested, the nature of the deformation could be studied as a function of the resolved shear stress on the principal slip planes. Moreover, specific effects of work-hardening could be investigated by examining the response of alloys such as Fe_3Si*, FeCo (refs. 55 and 56), or Cu_3Al (ref. 57), which can exist in the ordered (rapid work-hardening) or disordered (slow work-hardening) state at the same compositions. Certain effects of contamination could be studied by diffusing oxygen or other appropriate impurities into the region of reduced diameter prior to testing. In metals, the contaminated zone could be restricted effectively to the region of the simulated junction by using electrical resistance heating to increase the local temperature, and thus impurity diffusion coefficient at the reduced diameter. This technique could be used also to examine the influence of junction temperature on the deformation characteristics of model asperities.

Studies similar to those mentioned might also be used to investigate the contribution of brittle fracture emphasized by Byerlee (ref. 58) in his discussion of the friction of nonmetallic surfaces.

The behavior of these large idealized model asperities should be interpretable in terms of the properties of the bulk material. Thus, if their deformation characteristics and load bearing capabilities are sufficiently similar to those of smaller, conventional asperities, it may be possible to apply the results of these studies to the interpretation of real friction and wear phenomena. On the other hand, if gross discrepancies occur between the properties of small asperities and their large, idealized models, the nature and magnitude of the differences may be of value in deducing the origins of these discrepancies.

A possible contribution needing additional study is the role of surface diffusion at rubbing contacts. Mordike (ref. 59) has observed a rapid increase in the coefficient of friction between two rotating TiC discs at temperatures near 1400° C: close to that which Hollox (ref. 60) has suggested is sufficient to permit volume diffusion of the titanium atoms. Because bulk plastic deformation becomes important in TiC at much lower temperatures (near 900° C), Brookes (ref. 61) considered it unlikely that increased friction was the sole result of changes in the deformation behavior, and has suggested that the increased friction was a consequence of strengthening and growth of junctions by diffusion. The question cannot be considered resolved at present, however, because the mechanisms of deformation in TiC begin to change also in the temperature region near 1400° C (ref. 62). Thus, additional studies are needed to determine the origins of the increased friction, and particularly the effects of diffusion at rubbing surfaces.

***Lakso, G. E.; and Marcinkowski, M. J.: Contribution No. 1997, Ames Laboratory, U.S. A.E.C., (unpublished).**

Surface Characteristics

Theoretical studies.—Because surfaces influence the behavior of solids in a variety of ways, their properties currently are being investigated from many points of view. Even though many of these studies are not directly applicable to the problems of friction and wear, they suggest approaches which may be of value. Stern (ref. 63), for example, has considered the influence of the surface on the momentum distribution and wave functions of electrons in a metal. Now, because the electronic wave functions are determined by, and conversely, determine the equilibrium distribution of atoms in a crystal, it is expected that the distribution near the surface will be different from that in the bulk, and that the amplitude and direction of the atomic displacements near the surface will be closely related to the nature and degree of distortion of the wave functions. Thus, it appears possible that detailed studies of the electronic properties of solids near surfaces will provide valuable information regarding the structure and interfacial energies of clean surfaces.

In other studies, Bennett and Falicov (ref. 64), Grimley (ref. 65), and Gadzuk (refs. 66 and 67) have explored the electronic interactions responsible for the adsorption of individual atoms, or pairs of atoms, on metallic surfaces. Departing from most previous work on adsorption phenomena, they have included the electronic band structure in their representations of the metal. This approach is likely to yield a more realistic description of the electronic aspects of adsorption than is possible when the metal is considered as an aggregate of atoms. It should be exploited further to gain improved understanding of the adsorption processes involved in the interaction of lubricants and adhesives with solid surfaces.

A different aspect of the problem has been investigated by Bennett and Duke (refs. 68 and 69). Instead of considering the behavior of foreign atoms adsorbed on a solid surface, they examined the properties of an interface between two solids. Although they were interested primarily in the charge distribution and potential associated with a simple model of a bimetallic junction, appropriate extensions of their work may also have value for the understanding of adhesion between dissimilar metals.

The studies mentioned here represent significant advances in understanding some of the fundamental processes in friction and wear, even though the work may have been done with different purposes in mind. It seems appropriate, therefore, to extend such studies to include additional topics of specific interest to the phenomena of friction and wear.

Experimental studies.—As Professor Lawless mentioned, the friction and wear characteristics of real materials are markedly influenced by the environment. Because most metallic surfaces are covered with oxide films, however, the environment-sensitive mechanical behavior of these oxides may deserve special attention. Westwood, et al (ref. 70) have suggested that Rebinder effects (adsorption induced changes in the hardness of nonmetallic solids) may arise because chemisorption of appropriate molecular species is accompanied by changes in the state of ionization of point defects in a narrow region near the adsorbing surface, and possibly because of changes in the electronic charge on dislocations in this region.* Because the dislocation mobility in ionic solids is determined by the interactions between dislocations and point defects, changes in the state of ionization that modify these interactions will influence mobility, and thus surface hardness. Such effects are not expected to occur in materials with high electrical conductivity. Nevertheless they may have a prominent influence on the deformation characteristics of oxide-coated asperities, because the mechanical properties of the oxide may be altered by its environment and hence modify the interactions of dislocations with the metal-oxide interface. It appears possible, therefore, that an additional measure of control over friction and wear may be available when these effects are sufficiently understood; specific adsorbates can then be exploited to decrease (or increase) the hardness of surface oxide films as desired to satisfy particular operating conditions.

Attention should also be directed to a second influence of the environment, that is, changes in the composition, and thus mechanical properties, of the contacting surfaces by specific chemical reactions. In particular, Roberts and Owens (refs. 71 and 72) have achieved remarkable success in reducing friction and wear by employing iodine in charge transfer complexes to maintain lamellar metal diiodides on the rubbing surfaces of several appropriate metals; e.g., Ti, Co, stainless steels, and heat-resisting alloys of nickel and cobalt. The mechanical properties of the products formed in such reactions are not well known, but Roberts and Owens have attributed the low friction to the occurrence of planes of low shear strength in the lamellar diiodides because improved lubrication is not observed on metals that do not form iodides having this crystal structure.

Similar processes probably occur with many of the more conventional lubricants (e.g., formation of metal soaps by reaction with fatty acids), and presumably the low shear strengths of the products also contribute to the reduced friction obtained. The conventional lubricants, however, are not very effective with metals like titanium, whereas the addition of small concentrations of iodine charge trans-

---

*Westwood, A. R. C.; Goldheim, D. L.; and Lye, R. G.: to be published.

fer complexes to them, provides satisfactory lubrication even under high normal loads (ref. 72).

The observation (ref. 71) that sulfur-containing compounds failed to function as effective lubricants for titanium, even though TiS_2 has a lamellar crystal structure like that of TiI_2, indicates the need for more detailed investigations of both the kinetics of these reactions, and the material properties of the reaction products. Such studies can be expected to provide information for quantitatively applying this approach to a variety of operating conditions.

H. M. Davis (Army Research Office, Durham, North Carolina)

Being innocent to the study of friction and wear, I have been able to listen objectively and I have been greatly impressed by the high order of engineering revealed in these series of reviews on this fascinating and baffling subject. But the research reported is largely engineering research, in that it has been done on complex systems comprising poorly characterized materials, and affording a multiplicity of variables, some of which are uncontrolled and even unrecognized. Such an experiment has meaning for the specific system, but it rarely reveals the general principles being sought. It seems highly desirable, therefore, that the experienced investigators conceive experiments, apart from the practical systems of interest, whereby the significant variables may be isolated and reliably evaluated. I am impelled to add, with no desire to be offensive, that investigators of friction and wear ought not to exert any determined effort to avoid the intrusion of chemistry and metallurgy into their experiments.

M. E. Sikorski (Georgia Institute of Technology, Atlanta, Georgia)

I wish to describe briefly a combined technique for characterizing surfaces which is presently under study at the Georgia Institute of Technology. The approach considers the energy distributions of the secondary electrons emitted from a sample in addition to the LEED patterns obtained from the surfaces. LEED can be used to investigate the elastic scattering of low energy electrons and to get information about the structures of atomically clean single crystal surfaces or surfaces with ordered adsorbed gas layers. However, as was mentioned in Dr. Lawless' lecture, because of the dependence on periodic structures, the LEED patterns cannot always be used to detect surface contamination, and no information is given concerning the nature of a contaminant. The secondary electron energy distributions, on the other hand, result from inelastic processes which do not depend in detail on the structure of an adsorbed layer or on the ordering of crystallites in the substrate. These distributions can then be used to study surface contamination. Figure 10 shows an energy distribution curve of electrons for a clean tungsten (110) surface (ref. 73). There are three distinct groups of electrons contributing to the features observed in the curve of figure 10. The electrons in region I are the elastically reflected primaries. The electrons in region II are primaries which have undergone losses. Region III corresponds to the true secondaries. The small broad peaks that appear in region II and the subsidiary maxima contributing to the structure of the large secondary peak are related to the electronic properties of the material under study and to the contamination present on the surface. Studies have been made thus far of single crystal surfaces of tungsten (refs. 73 and 74), graphite and copper (ref. 74), and also aluminum and nickel.* Work on additional elemental metals and alloys is contemplated using single crystal, as well as polycrystalline specimens. It appears that the technique outlined above should be applicable to the characterization of surfaces in friction and wear experiments.

LECTURER'S CLOSURE

There seems to be very little left for me to say in closure. The discussion by Hollox and Lye is, I believe, quite valuable in that they have added considerable detail to many of the points concerning bulk and surface characteristics of materials. It seems quite clear that any improvement in our understanding of the complex processes of friction and wear will come only with detailed studies of the basic chemistry, metallurgy, and physics of the materials interactions. This has also been clearly stated by Dr. Davis in his discussion.

Basically, friction and wear involve the making and breaking of bonds between similar or dissimilar materials. Definitive experiments on adhesion are clearly a necessity as are refined studies of the deformation and fracture processes. In the usual friction and wear process, a very large number of variables may be involved; if any understanding is to be achieved, simpler systems must be studied initially so that the influence of the variables may be evaluated one at a time. It would seem most reasonable to start with the understanding of clean single crystal surfaces, and then to proceed to studies with controlled surface layers, such as adsorbed gases, oriented oxide films, or monomolecular hydrocarbon layers. I believe that this is a simpler method than to start with a "dirty" surface and to try to isolate the effects of the individual variables. There are some merits to starting with a complex

---

*Jordan, L. K.; and Scheibner, E. J.: Characteristic Energy Loss Spectra of Copper Crystals with Surfaces Described by LEED. Accepted for publication in Surface Science.

Private communications with G. F. Amelio, G. W. Simmons, and L. K. Jordan; Georgia Institute of Technology, 1967.

system and gradually simplifying it, but I believe this is most wasteful of research time and effort.

Modern research technology is certainly capable of handling even such a complex system as friction and wear. Electron microscopy techniques should be utilized in deformation studies to determine the details of dislocation distributions about stylus or wear tracks. The scanning electron microscope provides a powerful tool for fracture and wear studies and should be widely used. Basic studies of clean surfaces by diffraction techniques or field-ion techniques are needed along with studies of adsorption and chemical reaction on these surfaces. Modern techniques for detection and identification of surface impurities, such as the Auger spectroscopy technique mentioned by Sikorski, should be extremely useful in any clean surface studies.

Since most metal surfaces are covered with an oxide film, studies of the mechanical properties, and particularly the effects of environment on such properties, need careful detailed study.

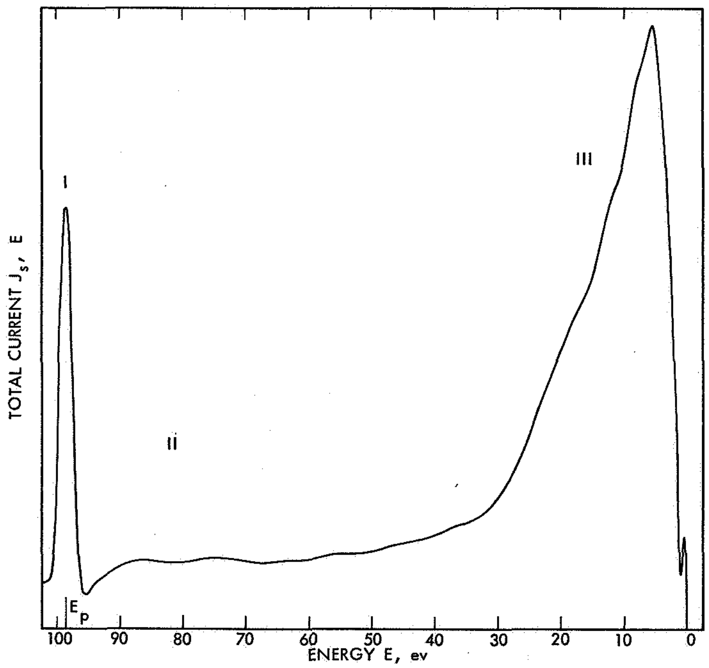


FIGURE 10.—Secondary electron energy distribution curve for the clean tungsten (110) surface. 98-volt primary electrons.

REFERENCES

1. Bowden, F. P.; and Tabor, D.: The Friction and Lubrication of Solids. Pt. II, Oxford Univ. Press, 1964.
2. Müller, E. W.: Field Ionization and Field Ion Microscopy. Advances in Electronics and Electron Physics, vol. 13, 1960, p. 83.
3. Davis, H. M.; and Lawless, K. R., editors: The Structure of Surfaces. Surface Science, vol. 8, 1967, p. 1.
4. Somorjai, G. A.: Note to "The Anomalous Surface Structures of Gold," D. F. Fedak and N. A. Gjostein, Surface Science, vol. 8, 1967, p. 98.
5. Lander, J. J.; and Morrison, J.: Structures of Clean Surfaces of Germanium and Silicon. J. Appl. Phys., vol. 34, 1963, p. 1403.
6. Simmons, G. W.; Mitchell, D. F.; and Lawless, K. R.: LEED and HEED Studies of the Interaction of Oxygen With Single Crystal Surfaces of Copper. Surface Science, vol. 8, 1967, p. 130.
7. Gibbs, J. W.: Collected Works. vol. I, Yale Univ. Press, 1948.
8. Herring, C.: The Use of Classical Macroscopic Concepts in Surface-Energy Problems. Structure and Properties of Solid Surfaces, R. Gomer and C. S. Smith, eds., Univ. Chicago Press, 1953, p. 5.
9. Mullins, W. W.: Solid Surface Morphologies Governed by Capillarity. Metal Surfaces, ASM, 1963, p. 17.
10. Cabrera, N.: The Equilibrium of Crystal Surfaces. Solid Surfaces, H. C. Gatos, ed., North Holland Publ. Co. (Amsterdam), 1964, p. 320.
11. Germer, L. H.: The Uncertainty Regarding Reconstructed Surfaces. Surface Science, vol. 5, 1966, p. 147.
12. Bauer, E.: The Study of the Degree of Order in Adsorption by Electron Diffraction. Absorption et Croissance Cristalline, C.N.R.S. (Paris), 1965, p. 20.
13. Lawless, K. R.: Oxidation of Metals. Energetics in Metallurgical Phenomena, vol. I, W. M. Mueller, ed., Gordon and Breach (New York), 1965, p. 345.
14. Fromhold, A. T., Jr.: Kinetics of Oxide Film Growth on Metal Crystals-I. J. Phys. Chem. Solids, vol. 24, 1963, p. 1081.
15. Gwathmey, A. T.; and Lawless, K. R.: The Influence of Crystal Orientation of the Oxidation of Metals. The Surface Chemistry of Metals and Semiconductors, John Wiley & Sons, 1960, p. 483.
16. Westwood, A. R. C.; and Stoloff, N. S., eds.: Environment-Sensitive Mechanical Behavior, Gordon and Breach (New York), 1966.
17. Westwood, A. R. C.: Environment-Sensitive Mechanical Behavior—Status and Problems. Environment-Sensitive Mechanical Behavior, A. R. C. Westwood and N. S. Stoloff, eds., Gordon and Breach (New York), 1966.
18. Westwood, J. H.; and Jorgenson, P. J.: Trans. Met. Soc., AIME, vol. 233, 1965, p. 425.
19. Müller, E. W.; McLane, S. B.; and Nishikawa, O.: In-Situ Manipulation of Field Ion Microscope Specimens. Electron Microscopy, vol. I, Maruzen Co. Ltd. (Tokyo), 1966, p. 235.
20. Greenwood, J. A.; and Williamson, J. B. P.: Contact of Nominally Flat Surfaces. Proc. Roy. Soc. (London), vol. A295, 1966, p. 300.
21. Bowden, F. P.; and Tabor, D.: Friction, Lubrication and Wear: A Survey of Work During the Last Decade. Brit. J. Apply. Phys., vol. 17, 1966, p. 1521.
22. Blok, H.: Theoretical Studies of Temperature Rise at Surfaces of Actual Contact Under Oilness Lubricating Conditions. Proc. General Discussion Lubrication and Lubricants, Inst. Mech. Engrs., 1937, p. 225.

23. DA VINCI, L.: The Notebooks of Leonardo da Vinci, E. MacCurdy, ed., Jonathan Cape (London), 1958.
24. AMONTONS, G.: De la Resistance Causee dans les Machines. Roy. Sci. (Paris), 1699, p. 206.
25. COULOMB, C. A.: Theorie des Machines Simples. Mem. Math. Phys. Acad. Sci.,vol. 10, 1785, p. 161.
26. DESAGULIERS, J.: Phil Trans., vol. 33, 1724, p. 345.
27. KELLER, D. V., JR.: Application of Recent Static Adhesion Data to the Adhesion Theory of Friction. Physical and Chemical Characteristics of Surfaces and Interfaces, Syracuse Univ. Press, 1967.
28. HOUWINK, R.; AND SALOMON, G.: Adhesion and Adhesives, Elsevier Publ. Co. (Amsterdam), 1965.
29. ELEY, D. D.: Adhesion. Oxford Univ. Press, 1961.
30. WEISS, P.: Adhesion and Cohesion. Elsevier Publ. Co. (Amsterdam), 1962.
31. KELLER, D. V., JR.: Adhesion Between Solid Metals. Wear, vol. 6, 1963, p. 353.
32. KELLER, D. V., JR.: On the Analysis of Metallic Adhesion Data. Adhesion or Cold Welding of Materials in Space Environments, ASTM, 1967.
33. KELLER, D. V., JR.: Effect of Contamination on the Adhesion of Metallic Couples in Ultra High Vacuum. J. Appl. Phys., vol. 38, 1967, p. 1896.
34. JOHNSON, K. I.; AND KELLER, D. V., JR.: Factors Affecting the Adhesion of Titanium and Molybdenum Couples. J. Vac. Sci. and Tech., vol. 4, 1967, p. 115.
35. BAILEY, A. I.: Friction and Adhesion of Clean and Contaminated Mica Surfaces. J. Appl. Phys., vol. 32, 1961, p. 1407.
36. BUCKLEY, D. H.: The Influence of Crystal Structure, Orientation and Solubility on the Adhesion and Sliding Friction on Various Metal Single Crystals in Vacuum (10^{-11} Torr). Adhesion or Cold Welding of Materials in Space Environments, ASTM, 1967, p. 248.
37. BOWDEN, F. P.; AND ROWE, G. W.: The Adhesion of Clean Metals. Proc. Roy. Soc. (London), vol. A233, 1955, p. 429.
38. TAYLOR, N. J.: LEED Study of the Epitaxy of Copper on the (110) Surface of Tungsten. Surface Science, vol. 4, 1966, p. 161.
39. BUCKLEY, D. H.; AND JOHNSON, R. L.: Marked Influence of Crystal Structure on Friction and Wear Characteristics of Cobalt and Cobalt-Base Alloys in Vacuum to 10^{-9} Torr. NASA TN D-2523, 1964.
40. SIKORSKI, M. E.: Correlation of the Coefficient of Adhesion with Various Physical and Mechanical Properties of Metals. J. Basic Eng., ASME, vol. 85D, 1963, p. 279.
41. ANDERSON, O. L.: The Role of Surface Shear Strains in the Adhesion of Metals, Part I. Wear, vol. 3, 1960, p. 253.
42. FOWKES, F. M.: Attractive Forces at Interfaces. Ind. and Eng. Chem., vol. 56, 1964, p. 40.
43. COURTNEY-PRATT, J. S.; AND EISNER, E.: The Effect of Tangential Force on the Contact of Metallic Bodies. Proc. Roy. Soc. (London), vol. A238, 1957, p. 529.
44. BAILEY, J. M.; AND GWATHMEY, A. T.: Friction and Surface Deformation During Sliding on a Single Crystal of Copper. ASLE Trans., vol. 5, 1962, p. 45.
45. STEIJN, R. P.: Friction and Wear of Single Crystals. Wear, vol. 7, 1964, p. 48.
46. DYER, L. D.: Rolling Friction on Single Crystals of Copper in the Plastic Range. Acta Met., vol. 9, 1961, p. 928.

47 Kragelskii, I. V.: Friction and Wear. Butterworths Sci. Publ. (Wash., D.C.), 1965.

48. Burwell, J. T.; and Strang, C. D.: On the Empirical Law of Adhesive Wear. J. Appl. Phys., vol. 27, 1952, p. 18.

49. Cocks, M.: Shearing of Junctions Between Metal Surfaces. Wear, vol. 9, 1966, p. 320.

50. Rabinowicz, E.: Friction and Wear of Materials. John Wiley & Sons, 1965.

51. Kerridge, M.; and Lancaster, J. K.: The Stages in a Process of Severe Metallic Wear. Proc. Roy. Soc. (London), vol. A236, 1956, p. 250.

52. Mathieson, R. T.: Orientation of Stearic Acid Monolayers on Silver Single Crystals. Nature, vol. 186, 1960, p. 301.

53. Bowden, F. P.; and Brookes, C. A.: Frictional Anisotropy in Nonmetallic Crystals. Proc. Roy. Soc. (London), vol. A295, 1966, p. 244.

54. Bowden, F. P.; and Tabor, D.: The Friction and Lubrication of Solids, Part II, Oxford Univ. Press, 1964, p. 100.

55. Marcinkowski, M. J.; and Chessin, H.: Relationship Between Flow Stress and Atomic Order in the FeCo Alloy. Phil. Mag., vol. 10, 1964, p. 837.

56. Marcinkowski, M. J.: Plastic Deformation in Iron-Cobalt Alloys, J. Inst. Metals, vol. 93, 1965, p. 476.

57. Davies, R. G.; and Stoloff, N. S.: Strain Hardening of Ordered F.C.C. Alloys. Phil. Mag., vol. 9, 1964, p. 349.

58. Byerlee, J. D.: Theory of Friction Based on Brittle Fracture. J. Appl. Phys., vol. 38, 1967, p. 2928.

59. Mordike, B. L.: The Frictional Properties of Carbides and Borides at High Temperatures. Wear, vol. 3, 1960, p. 374.

60. Hollox, G. E.; and Smallman, R. E.: Dislocation Loops in Titanium Carbide. Proc. Brit. Ceram. Soc., vol. 1, 1964, p. 211.

61. Brookes, C. A: Bulk Plastic Flow and the Frictional Properties of Titanium Carbide at High Temperatures. Wear, vol. 9, 1966, p. 103.

62. Hollox, G. E.; and Smallman, R. E.: Plastic Behavior of Titanium Carbide. J. Appl. Phys., vol. 37, 1966, p. 818.

63. Stern, E. A.: Electron States Near Boundaries. Phys. Rev., vol. 162, 1967, p. 565.

64. Bennett, A. J.; and Falicov, L. M.: Theory of the Electronic Configuration of a Metallic Surface-Adsorbate System. Phys. Rev., vol. 151, 1966, p. 512.

65. Grimley, T. B.: The Indirect Interaction Between Atoms or Molecules Adsorbed on Metals. Proc. Phys. Soc. (London), vol. 90, 1967, p. 751.

66. Gadzuk, J. W.: Nodal Hydrogenic Wave Functions of Impurities on Bounded-Electron-Gas Surfaces. Phys. Rev., vol. 154, 1967, p. 662.

67. Gadzuk, J. W.: Theory of Atom-Metal Interactions. I. Alkali Atom Adsorption, and Theory of Atom-Metal Interactions. II. One-Electron Transition Matrix Elements. Surface Science, vol. 6, 1967, pp. 133 and 159.

68. Bennett, A. J.; and Duke, C. B.: Self-Consistent-Field Model of Bimetallic Interfaces. I. Dipole Effects. Phys. Rev., vol. 160, 1967, p. 541.

69. Bennett, A. J.; and Duke, C. B.: Metallic Interfaces. II. Influence of the Exchange-Correlation and Lattice Potentials. Phys. Rev., vol. 162, 1967, p. 578.

70. Westwood, A. R. C.; Goldheim, D. L.; and Lye, R. G.: Rebinder Effects in MgO. Phil. Mag., vol. 16, 1967, p. 505.

71. Roberts, R. W.; and Owens, R. S.: Boundary Lubrication of Titanium-Titanium and Titanium-Steel. Wear, vol. 6, 1963, p. 444.

72. Owens, R. S.; Roberts, R. W.; and Barnes, W. J.: Iodine as an Extreme Pressure Lubricant Additive. Wear, vol. 9, 1966, p. 79.

73. Tharp, L. N.; and Scheibner, E. J.: Energy Spectra of Inelastically Scattered Electrons and LEED Studies of Tungsten. J. Appl. Phys., vol. 38, 1967, p. 3320.

74. Scheibner, E. J.; and Tharp, L. N.: Inelastic Scattering of Low Energy Electrons from Surfaces. Surface Science, vol. 8, 1967, p. 247.

Closing Remarks

P. M. KU
Southwest Research Institute
San Antonio, Texas

In the preceding pages, the phenomena of friction and wear, and their theoretical and practical implications, have been considered at length and from different viewpoints by various lecturers and discussors. Any further attempt on my part to evaluate the present level of understanding or to discuss the areas of needed research would therefore be quite superfluous. Rather, I propose to review the general conduct of the symposium in terms of the collective opinion of the participants and to outline the immediate plans for the symposium series.

As stated in the Preface, lubrication is a highly complex and interdisciplinary subject, and further advances in the science and technology of lubrication call for interdisciplinary cooperation. But this cannot come about without a common understanding of the problems involved, and common understanding of the problems cannot be achieved without the conscious application of interdisciplinary dialogue. The objective of the NASA lubrication symposium series was to promote interdisciplinary dialogue, it being assumed that actual interdisciplinary cooperation would ensue if the dialogue were effective and convincing.

It was recognized at the outset that the subject of lubrication is too broad to be covered effectively in a single meeting. Therefore, the decision was reached to approach the subject in a series of meetings, each relatively restricted in scope yet broad enough to permit meaningful interdisciplinary treatment. Friction and wear under unlubricated and boundary lubrication conditions were selected as the area of interest for the first such meeting. The program and conduct of this meeting were designed to facilitate and encourage interdisciplinary discussion.

The symposium had a total of 90 invited participants, composed of approximately 25 percent basic scientists, 50 percent lubrication research engineers, and 25 percent design and development engineers. Five of the participants contributed written discussions, but were not present.

TABULATION OF SURVEY RESULTS

At the conclusion of the symposium, a questionnaire was distributed and all attendees were asked for their considered opinions of the questions posed. Of the 85 persons present, 61 persons, or 72 percent, returned the questionnaire. The distribution of the respondents was 32 percent basic scientists, 52 percent lubrication research engineers, and 16 percent design and development engineers—a fair representation of all participants. The questions asked and the replies received are presented below.

(1) Do you think the concept of interdisciplinary approach to friction and wear (or to lubrication in general) is sound?

The replies were as follows:

[In percent]

	No response	*No*	*Perhaps*	*Yes*
Basic scientists	0	0	1	31
Lub. res. engrs	0	0	4	48
Des. and dev. engrs	0	0	2	14
	0	0	7	93

(2) Do you believe that meetings of this type can stimulate interdisciplinary research? Are there any other effective ways to stimulate interdisciplinary research?

The replies to the first part of the question were as follows:

[In percent]

	No response	*No*	*Perhaps*	*Yes*
Basic scientists	1	0	6	25
Lub. res. engrs	2	6	8	36
Des. and dev. engrs	2	2	0	12
	5	8	14	73

Typical replies to the second part of the question were as follows:

Incorporation of various disciplines in research groups
Exchange of scientists and engineers
Small discussion groups on specific topics
Publications devoted to interdisciplinary approach
Government contracts with interdisciplinary slant
Interdisciplinary academic degrees

(3) Do you believe that meetings of this type can stimulate interdisciplinary approach to design? Are there any other effective ways to stimulate interdisciplinary approach to design?

The replies to the first part of the question were as follows:

[In percent]

	No response	*No*	*Perhaps*	*Yes*
Basic scientists	14	11	1	6
Lub. res. engrs	4	21	7	20
Des. and dev. engrs	6	3	3	4
	24	35	11	30

The replies to the second part of the question were generally similar to those in response to the second part of the preceding question.

(4) Has the program of this symposium covered friction and wear adequately? If not, what other topics should have been included?

The replies to the first part of the question were as follows:

[In percent]

	No response	*No*	*Fair*	*Yes*
Basic scientists	3	0	8	21
Lub. res. engrs	0	16	17	19
Des. and dev. engrs	3	2	3	8
	6	18	28	48

Typical replies to the second part of the question were as follows:

Abrasive wear.
Fretting.
Friction and wear of nonmetals.
Friction and wear in electrical contacts.
Rolling friction and wear.
Solid lubricants.
Chemistry in lubrication.
Heavily loaded sliding systems.
Practical engineering problems.

(5) What is your opinion of the quality of the invited lectures?

The replies were as follows:

[In percent]

	No response	*Poor*	*Fair*	*Good*	*Excellent*
Basic scientists	2	0	3	11	16
Lub. res. engrs	1	0	3	21	27
Des. and dev. engrs	0	0	2	7	7
	3	0	8	39	50

(6) What is your opinion of the quality of the invited discussions?

The replies were as follows:

[In percent]

	No response	*Poor*	*Fair*	*Good*	*Excellent*
Basic scientists	3	0	7	8	14
Lub. res. engrs	1	0	11	20	20
Des. and dev. engrs	0	0	3	7	6
	4	0	21	35	40

(7) What is your opinion of the quality of the general discussions (i.e., those other than the invited ones)?

The replies were as follows:

[In percent]

	No response	*Poor*	*Fair*	*Good*	*Excellent*
Basic scientists	2	0	5	19	6
Lub. res. engrs	0	0	14	25	13
Des. and dev. engrs	0	0	1	9	6
	2	0	20	53	25

(8) Would you have liked to have a greater representation of basic scientists at the symposium?

The replies were as follows:

[In percent]

	No response	*No*	*Perhaps*	*Yes*
Basic scientists	4	22	2	4
Lub res. engrs	5	34	0	13
Des. and dev. engrs	3	12	0	1
	12	68	2	18

(9) Would you have liked to have a greater representation of design and development engineers at this symposium?

The replies were as follows:

[In percent]

	No response	*No*	*Perhaps*	*Yes*
Basic scientists	9	16	5	2
Lub. res. engrs	1	29	6	16
Des. and dev. engrs	2	3	4	7
	12	48	15	25

(10) Has the mailing of the *Proceedings Preprint* a month in advance been helpful to you?

The answer was unanimously in the affirmative.

(11) What is your opinion of the general conduct of this symposium? In what respects could the symposium have been improved?

The replies to the first part of the question were as follows:

[In percent]

	No response	*Poor*	*Fair*	*Good*	*Excellent*
Basic scientists	3	0	4	8	17
Lub. res. engrs	5	0	7	6	34
Des. and dev. engrs	2	0	3	3	8
	10	0	14	17	59

The replies to the second part of the question were typically as follows:

More time for general discussions.
More relaxed atmosphere.
Small discussion groups.
Panel discussions.
Provision for informal evening gatherings.

(12) Have you found this symposium of interest to you personally?

The answer was unanimously in the affirmative.

(13) Have you personally found this symposium helpful in the planning or execution of your research?

The replies were as follows:

[In percent]

	No response	*No*	*Perhaps*	*Yes*
Basic scientists	1	7	7	17
Lub. res. engrs	1	9	13	29
Des. and dev. engrs	16	0	0	0
	18	16	20	46

(14) Have you personally found this symposium helpful in your design and development work?

The replies were as follows:

[In percent]

	No response	*No*	*Perhaps*	*Yes*
Basic scientists	32	0	0	0
Lub. res. engrs	24	16	3	9
Des. and dev. engrs	6	5	2	3
	62	21	5	12

(15) Would you recommend that interdisciplinary symposia be held on other topics in the general area of lubrication science and technology? If so, please list the topics that you would like to see taken up, in your order of preference.

The replies to the first part of the question were as follows:

[In percent]

	No response	*No*	*Perhaps*	*Yes*
Basic scientists	7	2	2	21
Lub. res. engrs	7	5	2	38
Des. and dev. engrs	4	1	3	8
	18	8	7	67

In the order of preference, the suggested topics were as follows:

Liquid lubricants (including lubricant rheology, lubricant chemistry, synthetic lubricants, etc).
Friction and wear (unlubricated and boundary lubrication conditions).
Lubrication of concentrated contacts (elastohydrodynamic lubrication, lubrication of rolling-element bearings, gears, etc).
Solid lubricants and composites.
Hydrodynamic lubrication (including rotor dynamics).
Aerospace bearings (high-vacuum, high-temperature).
Electrical contacts.
Metalworking lubrication.

INTERPRETATION OF SURVEY RESULTS

In interpreting the results of the survey, it should be borne in mind that participation in this symposium was weighted toward those primarily active or interested in friction and wear under unlubricated and boundary lubrication conditions. For this reason, the preferred list of topics for future symposia was probably biased in favor of friction and wear, and against hydrodynamic lubrication. Moreover, rightly or wrongly, this symposium was primarily research oriented. Its purpose was to appraise and expand basic understanding, rather than to provide practical design guidelines. Therefore, in response to the question as to whether this symposium was helpful in research planning and execution, 66 percent of the respondents (basic scientists and lubrication research engineers) answered in the affirmative or qualified affirmative. However, for the question as to whether this symposium was helpful in design and development, only 17 percent of the respondents (lubrication research engineers and design and development engineers) gave favorable or qualified favorable replies. The fact that this symposium was considered helpful in research planning and execution was gratifying. The fact that it was not considered helpful in design and development was not surprising, but certainly regrettable. On the other hand, all respondents found this symposium of interest personally, suggesting that the long-range value of such meetings was appreciated by all.

The survey showed that the concept of interdisciplinary approach in lubrication was endorsed almost unanimously. It was felt that meetings of this type provide an effective means of stimulating interdisciplinary approach, but other means to compel interdisciplinary teamwork should not be overlooked.

A majority of the respondents recommended that additional interdisciplinary symposia be held. However, a more relaxed atmosphere and more time for general discussions were significantly favored. In this connection, it is interesting to note that, while the invited lectures received high ratings, the invited discussions were rated no higher in quality than the general discussions. This seems to suggest that, if the participants were carefully selected and were given adequate time to examine the lectures in advance, the practice of having invited discussions would have no special merit.

As to the desirable composition of the participants, the respondents apparently felt that the present distribution of basic scientists, lubrication research engineers, and design and development engineers was more or less satisfactory.

PLANS FOR FUTURE MEETINGS

The survey showed that the interdisciplinary lubrication symposium series serves a useful purpose and should be continued. The survey further revealed that the three most favored topics for future meetings were liquid lubricants, friction and wear, and lubrication of concentrated contacts.

The Steering Committee, after careful review of the survey results, recommended to NASA that consideration be given to holding two additional meetings, one in 1968 and one in 1969. This recommendation has been accepted by NASA. The two proposed meetings are as follows:

Interdisciplinary Workshop on Friction and Wear

A sequel to the symposium *Interdisciplinary Approach to Friction and Wear*, this workshop will be held on November 19 to 21, 1968, at NASA Lewis Research Center, Cleveland, Ohio. The participants will first meet in one gathering to select a number of key problems in friction and wear (both theoretical and practical), then convene in small discussion groups to consider these problems in depth and discuss possible methods of attack, and finally meet together to hear reports from the various discussion groups.

Interdisciplinary Approach to the Lubrication of Concentrated Contacts

This symposium will be held on July 15 to 17, 1969, at Rensselaer Polytechnic Institute, Troy, New York. The symposium will consider, among other topics, elastohydrodynamic lubrication, the critical temperature hypothesis, the structure and flow behavior of liquids, the structure of solids, the mechanism of contact fatigue, and the effects of materials, lubricants, and design. Each of the topics will be introduced by an invited lecture. However, there will be no invited discusions; this will then allow a greater proportion of time for general discussions.

The Steering Committee recognized the importance of an interdisciplinary study of the effects of liquid lubricants on lubrication. However, the subject is so broad and complex, and in many respects so controversial that, in the opinion of the Steering Committee, it could not be covered adequately in a single meeting. It occurred to the Steering Committee that some aspects of the subject, notably the structure and flow behavior of liquids, could be accommodated conveniently in the planned symposium on the lubrication of concentrated contacts. It was therefore decided that a broad study of liquid lubricants, or some facets of the subject, could be deferred until a later date.

U.S. GOVERNMENT PRINTING OFFICE: 1969 O—323-472

> *"The aeronautical and space activities of the United States shall be conducted so as to contribute . . . to the expansion of human knowledge of phenomena in the atmosphere and space. The Administration shall provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."*
>
> —NATIONAL AERONAUTICS AND SPACE ACT OF 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and technical information considered important, complete, and a lasting contribution to existing knowledge.

TECHNICAL NOTES: Information less broad in scope but nevertheless of importance as a contribution to existing knowledge.

TECHNICAL MEMORANDUMS: Information receiving limited distribution because of preliminary data, security classification, or other reasons.

CONTRACTOR REPORTS: Scientific and technical information generated under a NASA contract or grant and considered an important contribution to existing knowledge.

TECHNICAL TRANSLATIONS: Information published in a foreign language considered to merit NASA distribution in English.

SPECIAL PUBLICATIONS: Information derived from or of value to NASA activities. Publications include conference proceedings, monographs, data compilations, handbooks, sourcebooks, and special bibliographies.

TECHNOLOGY UTILIZATION PUBLICATIONS: Information on technology used by NASA that may be of particular interest in commercial and other non-aerospace applications. Publications include Tech Briefs, Technology Utilization Reports and Notes, and Technology Surveys.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION DIVISION
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
Washington, D.C. 20546