<table>
<thead>
<tr>
<th>TO:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approved for public release; distribution is unlimited.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FROM:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distribution authorized to U.S. Gov't. agencies and their contractors; Foreign Government Information; 1973. Other requests shall be referred to British Embassy, 3100 Massachusetts Avenue, NW, Washington, DC 20008.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AUTHORITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>British Embassy ltr, 12 Jun 1975</td>
</tr>
</tbody>
</table>
THIS REPORT HAS BEEN DELIMITED AND CLEARED FOR PUBLIC RELEASE UNDER DOD DIRECTIVE 5200.20 AND NO RESTRICTIONS ARE IMPOSED UPON ITS USE AND DISCLOSURE.

DISTRIBUTION STATEMENT A

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.
THE BACKGROUND TO CURRENT THEORIES OF SCUFFING

Edited by
P.B. MacPHERSON

Report of work done at Imperial College London for Westland Helicopters Ltd. Yeovil on Ministry of Defence Contracts, June 1972
RELEASE CONDITIONS FOR OVERSEAS DESTY

1. THIS INFORMATION IS RELEASED BY THE U.K. GOVERNMENT TO THE RECIPIENT GOVERNMENT FOR DEFENCE PURPOSES ONLY.

2. THIS INFORMATION MUST BE ACCOMPANIED AT SOME DEGREE OF SECURITY PROTECTION AS THAT ACCRUE TO IT BY THE U.K. GOVERNMENT.

3. THIS INFORMATION MAY BE DISCLOSED ON WITHIN THE DEFENCE DEPARTMENT TO THE RECIPIENT GOVERNMENT EXCEPT AS OTHER AUTHORIZED BY THE

4. THIS INFORMATION MAY BE SUBJECT TO PRIVATLY-OWNED RULES

© Controller HMSO London 1974
THE BACKGROUND TO CURRENT THEORIES OF SCUFFING

Edited by
P.B. MacPherson

Report of work done at Imperial College London for Westland Helicopters Ltd on Ministry of Defence Contracts

Summary
The Lubrication Laboratory at Imperial College is undertaking a series of investigations into scuffing phenomena on behalf of Westland Helicopters Ltd. The background to the work to be undertaken is set out in the present report in the form of a literature review. The review is divided under four headings:- Elastohydrodynamic lubrication in line contact; Lubrication breakdown under boundary conditions; Action of oil additives; Metallurgical aspects of scuffing.
Contents

Introduction 

Chapter 1. Elastohydrodynamic Lubrication in Line Contact
1.1. Introduction
1.2. A Survey of Theoretical Work
1.2.1. Line Contact
1.2.2. Point Contact
1.3. A Survey of Experimental Work
1.3.1. Measurement of Film Thickness & Shape
1.3.2. Measurement of Pressure & Temperature
1.4. Non-Newtonian Behaviour
1.5. Lubricant Starvation.
1.6. Partial Elastohydrodynamic Lubrication

References

Chapter 2. Lubrication Breakdown under Boundary Conditions.
2.1. Introduction
2.2. Boundary Lubrication
2.3. Lubricant Failure in Sliding Contacts

References

Chapter 3. The Action of Oil Additives.
3.1. Introduction to Extreme Pressure Additives
3.2. The use of Sulphur
3.3. Methods of Investigating Extreme Pressure Activity
3.4. Effect of Temperature on E.P. Films and their Lubricating Properties
3.5. The Effect of Sliding Load and Speed on E.P. Activity
3.6. Anti-wear and Anti-seizure Effects of E.P. Lubricants
3.7. Mechanisms of E.P. Action
3.8. Summary of E.P. Additive Action
3.9. Additive Interference
3.10. Literature Review of Additive Interference

References
Chapter 4. Metallurgical Aspects of Scuffing

4.1. State of the Art

4.2. Conclusions

References

92
101
102
Introduction and Editor's Comments

Without specialised knowledge it is difficult to plot a course through the mass of literature that has been published concerned either directly, or indirectly with scuffing. Contradictions are abundant and conclusions reached can be misleading in the extreme unless a careful analysis of the conditions on which they were based is made.

The Lubrication Laboratory at Imperial College has probably contributed more to the present day knowledge of the phenomenon of scuffing than any other single institution. This has been achieved by concentrating not on one aspect alone but by considering the overall pattern. The result of many years work has led to a reasonably comprehensive understanding of what happens to cause contacting surfaces to scuff.

In the course of the next few years reports will be issued on work at present being done at Imperial College for W.H.L. against M.O.D. contracts. It is thought that to make it easier for the non-specialist to understand better the background to this work a guide based on literature surveys could prove useful. The people in the Laboratory most directly concerned with the four main aspects of scuffing were each invited to contribute a chapter. In every case their reviews exclude reference to their recently completed and current work for this has not yet been published. For compactness, an elementary knowledge of the subject is assumed.

The collapse of the lubricating qualities of the oil film that leads to scuffing is a matter of chemistry. The circumstances that bring about this chemical failure are associated with the oil film thickness and thermal effects. Metallurgical considerations are equally important. These four subjects are each treated separately although clearly in practice they are inter related.

Editing as such has been kept down to a minimum and the guide can best be regarded as four independently written reviews each with its own list of references. The fact that the chapter on E.H.L. is longer than the others does not infer that it is the most important but simply that this subject has received much attention especially during the last decade.

Drs. Bailey, Spikes, Torrance and Wymer are all associated with M.O.D. contracts being worked on jointly by Westland Helicopters Limited and Imperial College. The work done by them in preparing the contents of this guide is appreciated.
CHAPTER 1

Elastohydrodynamic Lubrication in Line Contact

by

D.G. Wymer

1.1. Introduction.

The essential feature of an elastohydrodynamically lubricated contact is that the local pressures are sufficiently high to cause significant elastic deformation of the surfaces. The actual pressures involved depend on the elastic moduli of the surfaces and, in the case of soft materials such as rubber, these pressures need not be very high. The more usual case, however, involves high modulus materials, like steel, and here the pressures are necessarily high. Nonetheless, such pressures are easily obtained if the conformity is low, since the applied load is concentrated over a very small area. Typical examples include the contacts between gear teeth and in rolling element bearings. Under these conditions, the action of such high pressures on the lubricant is accompanied by a drastic increase in viscosity, and this has a profound effect on the behaviour of the contact.

Two elastic solids in nominal line contact can be represented by a single elastic cylinder on a rigid plane. For the two systems to be equivalent, the radius and elastic modulus of the cylinder are defined by the relations

\[
\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2}
\]

\[
\frac{1}{E'} = \frac{1}{2} \left( \frac{1-\sigma_1^2}{E_1} + \frac{1-\sigma_2^2}{E_2} \right)
\]

where \(R_1\) and \(R_2\) are the radii of curvature, \(E_1\) and \(E_2\) are the elastic moduli, and \(\sigma_1\) and \(\sigma_2\) are the Poisson's ratios of the two solids. \(R\) and \(E'\) are called the reduced radius and reduced elastic modulus respectively. The above discussion also applies to two solids in nominal point contact, whereupon the equivalent system is an ellipsoid on a plane. This form of representation greatly facilitates the treatment of elastohydrodynamic contacts.

The study of elastohydrodynamic lubrication is a relatively new subject but remarkable advances have been made in the last few
years, both theoretically and experimentally. The line contact problem has received by far the most attention because, by neglecting axial flow, it can be treated in two dimensions. This has resulted in a fairly complete theoretical analysis, which has been verified to a certain extent by experimental work. Since the point contact problem has to be treated in three dimensions, theoretical advances have been relatively few, and a complete solution is still awaited. Experimental work was, for a long time, also slow to develop. With the introduction of optical interferometry, however, the experimental side was given new impetus, and at the present time point contact experiments are far ahead of theory.

1.2. A Survey of Theoretical Work.

1.2.1. Line Contact.

Early interest in line contacts arose as a result of work on gear lubrication. The absence of wear on gear teeth led people to believe that they were lubricated by hydrodynamic film. The basis of hydrodynamic theory is Reynolds' equation which, in three-dimensional form, states that for an incompressible fluid between two moving surfaces

\[
\frac{\partial}{\partial x} \left( \frac{h^3}{\eta} \cdot \frac{\partial p}{\partial x} \right) + \frac{\partial}{\partial y} \left( \frac{h^3}{\eta} \cdot \frac{\partial p}{\partial y} \right) = 12U \frac{\partial h}{\partial x}
\]

where \( U \) is the mean surface speed in the \( x \)-direction and \( p \) is the pressure generated in a film of thickness \( h \) and viscosity \( \eta \). The derivation of this equation relied on several simplifying assumptions, as described, for instance, by Cameron (1). In the special case of infinitely long surfaces with no thickness variation in the \( y \)-direction, all derivatives with respect to \( y \) disappear, and on integrating the remaining expression, we get

\[
\frac{dp}{dx} = 12U \left( \frac{h-\bar{h}}{\bar{h}^3} \right)
\]

where \( \bar{h} \) is the local film thickness at the point of maximum pressure. In an attempt to verify the existence of a hydrodynamic film between gear teeth, Martin (2) in 1916 solved the Reynolds equation for the geometry of the equivalent cylinder on a plane,
and obtained the film thickness relation

\[
\frac{h_o}{R} = 4.896 \frac{U\eta_o}{W/L} \ldots \ldots \ldots \ldots (1.2)
\]

where \( h_o \) is the thickness on the line of centres.

Substituting typical values into equation 1.2 gave a film thickness of the order of 1 \( \mu \) in. This was much less than the heights of the machining marks on the gear teeth, and seemed to suggest that a hydrodynamic film could not therefore exist. This suggestion was refuted by later experimental findings, and it became clear that the theory must be modified under such conditions.

The effects of elastic distortion were included into the theory by Peppler in 1936 (3) and 1938 (4) again applying the result to gear teeth. Although the effect on film thickness was beneficial, the increase was still not enough. A similar result was obtained in 1941 by Meldahl (5). The effects of pressure-viscosity characteristics on rigid cylinder lubrication were examined by Gatcombe (6) in 1945, Hersey and Lowdenslager (7) and Blok (8) in 1950, and McEwan (9), Cameron (10) and Blok (11) in 1952, using various pressure-viscosity relationships. These all demonstrated an increase in film thickness over the Martin condition, but the assumption of rigid cylinders led to erroneous results.

A major advance was made by Ertel, whose work was published in 1949 by Grubin (12) and became known as Grubin's theory. For the first time, the effects of elastic distortion and pressure-dependent viscosity were considered together. The pressure-viscosity relationship assumed was the exponential law:

\[
\eta = \eta_o e^{xp}
\]

The deformation was assumed to be the same as in dry static contact as given by the theory of Hertz (13), thus knowing the inlet shape, the pressure curve in this region could be calculated from Reynolds' equation, which now became

\[
e^{xp} \frac{dp}{dx} = 12U \eta_o \left( \frac{h-h_i}{h} \right) \ldots \ldots \ldots (1.3)
\]
A reduced pressure, $q$, was defined such that

$$\frac{dq}{dx} = e^{-\alpha p} \cdot \frac{dp}{dx} \quad (1.4)$$

Hence equation 1.3 became

$$\frac{dq}{dx} = 12U \eta_0 \left( \frac{h - H}{h^3} \right)$$

and integrating equation 1.4 gave

$$q = \frac{1}{\alpha} \left( 1 - e^{-\alpha p} \right)$$

It was assumed that the pressure at the edge of the contact was sufficiently high to write

$$e^{-\alpha p} \approx 0$$

i.e. $$q \approx \frac{1}{\alpha}$$

The calculation could now proceed in a straightforward manner, and yielded the film thickness equation

$$\frac{h_0}{R} = 1.95 \left( \frac{e \eta_0 U}{R} \right)^{0.727} \left( \frac{W}{L} \right)^{-0.091} \quad \ldots (1.5)$$

Despite the assumptions in the derivation, equation 1.5 has since been found to be in reasonable basic agreement with experimental results. Ertel also discussed how the film shape and pressure distribution would be expected to differ in practice from that given by his simple model. His predictions were surprisingly accurate in that they included the presence of a film constriction near the exit and a pressure spike just prior to this constriction. He also considered the effects of lubricant compressibility and thermal processes on the film shape and pressure distribution.

Much of Ertel's qualitative discussion, including the presence of a pressure spike, was confirmed theoretically by Petrusevich (14) in 1951 when he obtained a common solution to the hydrodynamic and elasticity equations by a trial-and-error method. A numerical solution was presented by Weber and Saalfeld (15) in 1954, but although the effect of pressure-dependent viscosity was clearly shown, they only considered small elastic distortions. The deformations were therefore not near-Hertzian as in the case of Grubin or Petrusevich.

The most complete set of numerical solutions was developed by Dowson and Higginson. In the inlet region, they obtained the
pressure by direct integration of the Reynolds equation. A solution that also satisfied the elasticity equation was then found by an iterative procedure. This direct method was not suitable for the high pressure and outlet regions, however, due to the difficulty of obtaining a rapid convergence. This was overcome by solving, instead, the inverse hydrodynamic problem, obtaining a film shape from an assumed pressure distribution. The deformation of the surfaces resulting from this pressure distribution was calculated using elasticity theory, and was compared with the inverse hydrodynamic solution. The pressure distribution was then successively modified until the two results were in agreement.

Initial results published in 1959 by Dowson and Higginson (16) were restricted to one set of conditions only, covering a range of loads at just one speed. The pressure spike predicted by Petrusevich was not present in their solutions, although the general features of their results fitted in with expectations. In 1960 (17) they extended their results to include conditions of greater severity, with the aid of a digital computer. The pressure spike was now clearly shown under these new conditions. Typical results are shown in figure 1.1. In 1961 (18) they correlated their results, which covered a range of speeds, loads and elastic moduli, into a single empirical expression for the minimum film thickness at the exit constriction

$$h_{\text{min}} = 1.6 \alpha^{0.16} (\eta U)^{0.7} E^{0.3} R^{0.43} (\frac{W}{L})^{-0.13} \cdots \cdots (1.6)$$

It was shown by the same authors in 1966 (19) that the discrepancy with Grubin's equation was never more than 20%. The introduction of non-dimensional parameters considerably simplified the treatment and application of the results, and the following were suggested

Film thickness parameter: $H^* = \frac{h}{R}$; Speed parameter: $U^* = \frac{U \eta}{E'}$;
Load parameter: $W^* = \frac{W}{L E'}$; Materials parameter: $G^* = \frac{E'}{E}$

Equation 1.6 then became

$$H^*_{\text{min}} = 1.6 U^{0.7} W^{-0.13} G^{0.6} \cdots \cdots \cdots (1.7)$$

In 1962, Dowson, Higginson and Whitaker (20) considered the effects of speed more fully, and also made an allowance for the
Figure 1.1.
Pressure distributions and film shapes
(after Dowson and Higginson (17))
lubricant compressibility in their calculations. The speed was found to have by far the greatest influence. The effect of including compressibility was to introduce a slight modification to the film shape in the form of a small increase prior to the constriction, and to reduce the height of the pressure peak and to displace it downstream.

A similar numerical treatment was presented in 1961 by Archard, Gair and Hirst (21) in which they split the contact region into four zones and considered each one separately. They obtained results in close agreement with those of Dowson and Higginson, although the height of the pressure peak was slightly smaller. The disadvantage of the Dowson and Higginson method was that manual intervention was required in fixing the pressure curve. A fully-automatic direct procedure was ideally required, and this was attempted in 1962 by Stephenson and Osterle (22). However, they had difficulties in getting convergence at high loads, and as a result, only light loads were considered.

The incorporation of thermal effects into the theory was considered by Sternlicht, Lewis and Flynn (23) in 1961. In applying the energy equation to the calculations, however, they made the assumption that all the viscous heat was retained in the fluid. This assumption was far from the truth, and thus invalidated their results. In 1964, Cheng and Sternlicht (24) included the conduction of heat by the surfaces, and showed that thermal effects in the film had little effect on the film thickness. This confirmed experiments by Crook (25) using varying amounts of sliding in a disc machine. Dowson and Higginson's isothermal theory was therefore considered to be still essentially correct, including the presence of the pressure spike. Refinements to the thermal solution were made in 1965 by Cheng (26) and Dowson and Whitaker (27) by considering the effects of varying lubricant properties through the film as well as along it.

It was pointed out by Dowson (28) in 1967 that the film thickness equation derived in 1961 (equation 1.7) did not satisfy the requirements of dimensional analysis, and suggested that the computer solutions might be better represented by

\[ H^* = 2.65 U^{0.7} G^{0.54} W^{0.13} \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots (1.8) \]

The solutions of Dowson and Higginson cover the range of conditions most commonly found in practice, namely high elastic
modulus materials lubricated under heavy load by a fluid whose pressure-viscosity relationship resembles that of a mineral oil. In 1965, Dowson and Whitaker (29) presented a survey of isothermal solutions showing how the nature of the lubrication changes from the lightly-loaded "rigid" region covered by Martin's formula, through the moderately loaded region of Weber and Saalfeld (15), to the heavily-loaded "elastic" region of Dowson and Higginson. They showed the presence of an intermediate zone in which the surfaces could be considered as rigid, but the viscosity of the lubricant was enhanced by pressure. This situation does not occur for low modulus materials, however, since in this case, on steadily increasing the load, the elastic distortions take effect before any appreciable viscosity change. This elastic-isoviscous region has received theoretical study by Herrebrugh (30). In all, one can conveniently define four regimes of elastohydrodynamic lubrication, rigid/isoviscous, elastic/isoviscous, rigid/variable viscosity, and elastic/variable viscosity. The existence of these regimes was clarified in a paper by Johnson (31) by considering the theoretical results of a number of workers, and plotting them on a "map" in much the same way as one may plot the thermodynamic "state" of a fluid.

The choice of non-dimensional parameters played an important part in clarifying much of the work on surveying of theoretical solutions. Although Dowson and Higginson used four dimensionless parameters, it can be shown that only three are required fully to represent the problem. Several three-parameter sets have been suggested (32-35-28), each of them being justified on the grounds of bringing out a particular characteristic of elastohydrodynamic behaviour. Johnson (31), in demonstrating the existence of the four regimes used the parameters introduced by Theyse (35)

\[
\frac{L^0W}{L^2\eta_0UR}, \left(\frac{X^2W^3}{L^2\eta_0UR^3}\right)^\frac{3}{4}, \frac{W^2}{L^2\eta_0UR}\]

These appear somewhat confusing at first, but on closer inspection we see that they do have a physical meaning. On referring to the rigid-isoviscous equation of Martin (equation 1.2) we see that the first parameter is the ratio

\[
4.896 \times \text{actual film thickness} \over \text{"Martin" film thickness at same load and speed}
\]
The second parameter contains the pressure-viscosity coefficient but not the elastic modulus, and is therefore related to the rheological nature of the lubricant. The third parameter contains the elastic modulus but not the pressure-viscosity coefficient, and is therefore related to the material properties of the bounding surfaces. These last two parameters thus relate separately to the very characteristics which determine the regime of lubrication.

1.2.2. Point Contact.

As in the case of line contact, the earliest treatment of the point contact problem was confined to conditions of rigid surfaces and an isoviscous fluid. A particular solution of the two-dimensional Reynolds equation was obtained by Howlett (36) in 1946, as applied to a four ball machine, and a similar treatment for a sphere on a plane was given by Kapitza (37) in 1955. A solution was also obtained in 1958 by Korovchinskii (38). The pressure distribution was shown to be similar to the Sommerfeld solution for journal bearings, (the negative pressures were, of course, neglected in calculating the load capacity). This classical treatment yielded the film thickness equation

$$h = 13.6 \eta U R^2 W^{-2}$$

which, on substituting typical values, gave film thicknesses of the order of $10^{-8}$ in. Kapitza recognised the fact that this expression was only applicable at extremely light loads, and that, at higher loads, the pressure-viscosity effects of the lubricant had to be included. By equating the maximum pressure in the isoviscous case to $\frac{1}{\alpha}$, he derived a limiting film thickness at which independence of load was achieved. This was a first approximation to the full solution, in that it assumed no deformation, and could be expressed as

$$h_0 = 0.67 (\alpha \eta U)^{1/3} R^{1/4}$$

The first treatment to include the added effects of surface deformation was by Archard and Cowking (39) in 1965. They made the Grubin-type assumption of Hertzian deformation and, dividing
the inlet region into a number of elemental strips, applied the one-dimensional Reynolds equation to each strip. The reduction in pressure due to side leakage was assumed to be the same as in the undeformed case, and was accounted for by a side leakage factor. The dependence of film thickness on this side leakage factor was checked experimentally, and the agreement was found to be fair.

In the same paper, two semi-analytical solutions for a circular point contact were obtained using the two-dimensional Reynolds equation and assuming a Hertzian shape. The first solution was derived using a Kapitza-type analysis, yielding a limiting film thickness expression and a Sommerfeld type of pressure distribution. The second solution used the more realistic Grubin assumption that the reduced pressure at the edge of the Hertzian region was constant at \( \frac{1}{2a} \), and gave a pressure distribution similar to that derived by Grubin for line contact. Both solutions gave film thickness expressions in close agreement with that derived previously using a side leakage factor. Another solution based on Grubin's assumptions was derived in 1966 by Cameron and Gohar (4), using a hand relaxation technique in the inlet region.

The extension of the Kapitza theory to elliptical contact was made in 1968 by Snidle and Archard (41). A more refined analysis of elliptical contacts was made in 1969 by Cheng (42) using Grubin-type inlet conditions. In the special case of a circular point contact, the Cheng theory agreed favourably with that of Archard and Cowking. In 1970, a solution for circular point contacts was obtained by Wedeven, Evans and Cameron (43), again using Grubin assumptions. This result agreed very closely indeed with the circular point contact case of Cheng's theory.

All these solutions can be represented by the following expression for circular point contacts

\[
\frac{h}{R} = K \left( \frac{\eta U}{U_R} \right)^a \left( \frac{W}{E R^2} \right)^b \quad \ldots \quad (1.9)
\]

where the constants, \( K \), \( a \), and \( b \) are as presented in the following table
These results bear a close resemblance to the line contact case, and demonstrate that the effect of side leakage is not as severe as one might expect. Indeed, according to Cheng's theory, the film thickness in a circular point contact is only 18% less than that in a line contact, under conditions of comparable stress.

1.3. A Survey of Experimental Work.

1.3.1. Measurement of Film Thickness and Shape.

Early suspicions that there was often a substantial hydrodynamic film in concentrated counterformal contacts, under conditions where classical theory predicted little more than boundary lubrication, were first verified by the results of electrical resistance measurements. By applying a small voltage between the two bearing elements, low enough to avoid dielectric breakdown in the oil, measurement of the resistance showed whether or not there was metal-to-metal contact. In most cases, the expected periods of near-zero resistance were accompanied by significant periods of high resistance, indicating that there was indeed a hydrodynamic film, although frequently penetrated by asperities on the metal surfaces.

Resistance measurements in boundary lubrication studies were reported as long ago as 1940 (44) in a four ball machine, and in piston ring studies in 1942 (45) and 1946 (46). The method was also applied to thrust bearings by Brix (47) in 1947, and to a journal bearing in 1949 by Tudor (48). Of great interest at the time was the nature of the lubricating film between gear teeth, and the resistance technique was used to advantage here by Lane and Hughes (49) in 1952, showing that a hydrodynamic film could indeed be generated. This result was confirmed in later resistance experiments, using discs to simulate a gear tooth.
contact, by Cameron in 1954 (50), Lewicki in 1955 (51), Crook in 1957 (52) and El-Sisi and Shawki in 1958 (53).

Attempts to measure the film thickness by this method were unsuccessful, however, because the resistivity of the oil was very dependent on temperature and moisture content. El-Sisi and Shawki (54) stabilised the resistivity somewhat by adding 4% sodium petroleum sulfonate, but this still did not give dependable results. Attention was slowly diverting to other possible means of measurements. In 1955, Lewicki (51) attempted capacitance measurements between two lubricated discs, and showed, at least, the feasibility of such a method for determining thickness. The idea was taken up by Crook (55) in 1957 with a more sophisticated disc machine. The oil film capacitance was measured by using separate pads lightly riding on the oil film adhering to each disc, thus avoiding the problem of estimating the contact geometry between the discs. The transition from the rigid "Martin" condition to the elastohydrodynamic lubrication condition, typified by the diminishing load dependence, was clearly shown. The dependence of film thickness on speed, however, observed to be

\[ h \propto U^4 \]

was in poor agreement with Grubin's theory. The low speed index was probably the result, among other things, of not including the inlet viscosity variation.

In 1958, an attempt to measure film thickness was made by Sibley et al (56) by measuring the dielectric break-down voltage. The results were again largely dependent on the moisture content of the oil, however, and the method was not followed up. An extension of this method was found to have more success in the same year, Siripongse, Rogers and Cameron (57,58) showed that if sufficient voltage was applied across an oil film to cause dielectric breakdown, the subsequent voltage drop, at currents about \( \frac{1}{2} \) amp. depended only on the thickness of the film. They used this method to measure film thickness in a four ball machine, and in 1960, MacConochie and Cameron (59) applied the technique to gear teeth. This "voltage discharge" method was subsequently used on several occasions (60 - 63) but despite its applicability to the rough surfaces encountered in engineering components, the method fell into disuse, following doubts as to its validity and
the introduction of better methods. The main grounds for criticism were that the relatively heavy currents used (up to 5 amps.) were suspected of causing drastic changes in the contact region. These doubts prompted Dyson (64) in 1967 to examine the technique by making direct comparisons with capacitance measurements. He found that the agreement was not good and in particular he was unable to show the linear relationship between discharge voltage and film thickness as Cameron and co-workers had done. Moreover, he found that the relationship depended on the temperature and the slide/roll ratio, and he also encountered troubles with spark-erosion of the surfaces.

Crook (25) in 1961 presented new results from his disc machine, this time taking the inlet viscosity into account. He took the controlling temperature to be that of the disc surfaces, and his results yielded a relation of the form

\[ h \propto (\overline{U} \eta_s)^{0.5} \quad (1.10) \]

where \(\overline{U}\) was the mean surface speed and \(\eta_s\) was the oil viscosity at the skin temperature of the discs. In the same paper he described a new method of capacitance measurement. He measured the capacitance between the two discs and estimated the film thickness in the contact by assuming the shape of the deformed surfaces was Hertzian. The results agreed well with the previous ones using pads, except for a constant difference attributed to changes of dielectric constant and density of the oil under the extreme pressures in the contact.

Crook then extended the application of capacitance measurements even further by making the first ever measurements of film profile (65,66). To do this he replaced one of the steel discs by a glass one, and deposited a chromium electrode on its surface. As the leading edge of the electrode passed through the contact, its rate of change of capacitance with respect to the other disc was measured as a function of time on an oscilloscope. The film profile was shown to consist of a parallel region extending over much of the contact, with a constriction of about 10% thickness reduction over a much smaller region near the exit. Thus the essential characteristics predicted by theory were shown to exist.
The capacitance technique was also applied to point contacts in 1961 by Archard and Kirk (67) using a crossed cylinders machine. Because of the extra dimension in the geometry, they found it necessary to make the approximation that the shape of the gap outside the contact was independent of load. This was in addition, of course, to the assumption made by Crook that the film shape in the contact was parallel over a region the size of the equivalent Hertzian area. The accuracy of their results, therefore, must be eyed with perhaps a little suspicion. The empirical relationship derived from their measurements,

\[ h \propto \left( \frac{\sigma \eta_W}{R} \right) \left( \frac{W}{E^* R} \right)^{0.364} \left( \frac{\eta_W U}{R^2} \right)^{0.727} \]  

(1.11)
did at least show, however, that an elastohydrodynamic film could be generated in a point contact, and that it obeyed a similar relationship to that theoretically predicted for a line contact. This similarity was further demonstrated in later capacitance work by Archard and Cowking (39) on elliptical contacts where it was shown that the overall effect of the ellipticity on the film thickness was surprisingly small.

1961 also saw the advent of a completely new method of film thickness measurement. Sibley and Orcutt (68) passed a collimated beam of X-rays through the elliptical contact between two lubricated crowned discs, in the direction of rolling. By measuring the intensity of the transmitted radiation they were able to estimate the minimum film thickness in the contact. Their results fitted the following empirical relation

\[ \frac{h_{\text{min}}}{R} = 0.0573 \left( \frac{W}{E^* R} \right)^{-0.364} \left( \frac{\eta_W U}{R^2} \right)^{0.727} \] 

(1.12)

where \( W \) was the load per 2/3 width of the elliptical contact and \( E^* \) was defined as

\[ \frac{1}{E^*} = \frac{1}{E_1} \left( \frac{1 - \sigma^2}{E_1} + \frac{1 - \sigma^3}{E_3} \right) \]

This was in good agreement with the theoretical prediction for line contact, although the calculation of film thickness from transmitted intensity data was beset by numerous pitfalls, and it became necessary to apply several correction factors. A general
picture of the transverse profile was also presented, but the relatively large beam width of 0.030 in. precluded the observation of any local film variations.

The use of optical interferometry as a measure of film thickness was introduced by Kirk (69) in 1962 and Archard and Kirk (70) in 1963. Owing to the use of perspex for both surfaces, however, (again in a crossed cylinders machine) they were restricted to low contact pressures, and the role of the pressure-viscosity coefficient was not significant, therefore. Moreover, the extremely low reflectivity of perspex in oil meant that the interference colours produced with white light were very faint. Evidence of a constriction at each side of the contact was seen, however, and the general feasibility of the interferometric method demonstrated, at least.

At about the same time, Christensen (71,72) reported a variation of the method originally used by Crook in 1957, using pads riding on the oil film outside the contact. This method was essentially one of measuring the volume rate of flow through the contact, which was directly related to the pad displacement. To measure this displacement, Christensen used inductive transducers, instead of the capacitance method used by Crook. His results conformed closely to the empirical relation

\[ h \propto (\gamma U)^{0.63} W^{-0.2} \]  

The discrepancy between this and the theoretical expression for line contact (equation 1.6) could not be readily explained.

In 1963-6 Gohar and Cameron (73,40,74) reported the first successful use of interferometry to measure film thickness. Their studies were restricted, in the main, to point contact, but the use of a steel ball and a high refractive index glass plate as the bounding surfaces ensured that the fringe colours produced by white light were now bright enough to enable accurate measurements to be made. Since the elastic moduli of both surfaces were high, it was possible to generate pressures sufficiently high to cause drastic increase of oil viscosity, and so all the features of elastohydrodynamic lubrication were present. White light interferometry was mainly used for film profile measurements, and the introduction of monochromatic
interferometry allowed the measurement of central film thickness to be carried out over a greatly extended speed range. In later work (75) even higher loads were possible with the use of diamond or sapphire instead of glass.

A small amount of work was also done on line contact (using a cylindrical roller instead of a ball) including a qualitative study of the effect of end blending. Experimental difficulties led to the restriction of line contact studies to low loads, however, and the results were marred by the inability to obtain a good surface finish on the rollers. Up to this point, all their studies had been made in conditions of pure sliding, but a new experimental rig was later introduced (76) to allow measurements to be made in pure rolling. The experimental point contact film thicknesses showed some agreement with the theory of Archard and Cowking (equation 1.9). The line contact results were rather few in number, but some comparison with theory was possible, and showed a tendency to fall below theoretical predictions especially at the higher speeds.

Orcutt (77) in 1965, followed up Crook's idea of measuring film profiles using a capacitance technique. His results confirmed Crook's earlier findings, showing a constriction towards the outlet. The method was slightly different from that used by Crook in that the capacitance electrode was in the form of a painted platinum strip at right-angles to the direction of rolling. This strip was also used as a temperature transducer. In both cases, the sensitivity of the strip depended on it being very narrow compared with the contact width.

The lack of agreement between Crook's empirical results (equation 1.10) and the now well-established theory of Dowson and Higginson (16-18) prompted Dyson, Naylor and Wilson (78) to investigate further. In 1965 they published an extensive set of film thickness results obtained by a capacitance method. They used a disc machine similar to Crook's, and deduced the film thickness from the capacity between the discs, assuming, as Crook had done, that the deformations were Hertzian. They were able to describe their results by the following empirical expression

\[
\frac{h}{R} \propto (U \eta_0)^{0.65} \propto 0.56 W^{-0.15} \quad (1.14)
\]
This was the closest agreement with theory so far obtained for line contact. The discrepancy with Crook's results (equation 1.10) was significant, despite the fact that one of the test fluids was a sample of the oil used by Crook. The authors were unable to explain this discrepancy. The above empirical relationship, because of its good agreement with theory, has come to be accepted in preference to Crook's results, particularly so in view of the large number of lubricants used for its derivation. Knowing that the surface profile is certainly not Hertzian in shape, however, there must be doubts as to the accuracy of the method. The only technique which seemed to have any chance of removing these doubts was the optical method used by Gohar and Cameron. Up to the present time, however, no intensive line contact studies have been made using the interferometric technique.

Now that some degree of correlation had been obtained between film thickness measurements and theory in a line contact, interest was being concentrated more and more on other parameters such as pressure and temperature within the contact. These two in particular were thought to vary rather drastically throughout the contact, and in order to establish their dependence on film thickness, it became necessary to study the film profile more closely.

Kannel, Bell and Allen (79) in 1965 reported a variation of their X-ray transmission technique to allow film profile to be measured. They directed the X-ray beam through the contact in a direction perpendicular to the direction of rolling instead of parallel to it as before. To increase the resolution they reduced the beam width from 0.030 in. to 0.003 in. with a slit which could be traversed across the width of the contact. Although the characteristic film shape was observed, the accuracy was not high enough to permit the pressure distribution to be calculated with any reasonable degree of precision.

In 1967, Hamilton and Moore (80) combined a capacitance electrode, as used by Crook, with pressure and temperature transducers, so that film profiles could be superimposed on pressure and temperature profiles.

A further application of the capacitance technique was introduced by Vichard and Godet (81) when they measured the
film thickness between a cam and a tappet. Doubts on the validity of capacitance measurements were resurrected, however, by Astridge and Longfield (82) in 1968 when they obtained results with a point contact machine of high conformity. They concluded that the assumptions regarding the shape of the deformed surfaces, could, under certain conditions, give rise to large errors.

An interesting new method of measuring film thickness was reported by R. Cameron (83,84) in 1967. He measured the distance between two rolling discs using a variable magnetic reluctance technique. Reasonable agreement with theory was found, and the author describes how the method could be applied to gear teeth. The great advantage associated with the method appears to be its application to rough surfaces, since the measurement is not affected by metallic contact. No further work with this technique appears to have been published however.

A direct displacement technique using a distance transducer applied to a taper roller bearing was reported by Grafton (85) in 1968, but without much success. The method was later improved on by Jones and Crease (86), and quite good correlation with the Dowson and Higginson theory was obtained. The slight discrepancy was thought to be possibly due to an error in estimating the inlet viscosity.

The last five years have seen the emergence of optical interferometry as a very widely accepted method of film thickness measurement. Although, until now, it has been almost entirely restricted to point contact, our understanding of elastohydrodynamic lubrication has increased tremendously as a result of the development of the technique, and, unlike other forms of measurement, its accuracy has never been seriously challenged. Most of the work has been done at Imperial College by A. Cameron and co-workers. Following the earlier work of Gohar and Cameron, the optical system was greatly improved in 1967 by Poord et al (87-89) by replacing the high refractive index glass with a plate of crown glass having a thin layer of chromium on the side in contact with the ball. The variation of central film thickness with speed and load was studied for a wide range of fluids. Most of the fluids obeyed the following empirical
relationship

\[ \frac{h}{R} = 0.86 \left( \frac{U \eta_{0}}{E R} \right)^{0.63} \left( \frac{W}{E R^{2}} \right)^{-0.05} \left( \frac{\alpha E}{E} \right)^{0.6} \]  

although for some fluids, the index of the speed parameter differed slightly, but always in the range 0.55 to 0.7. The film profile was also accurately measured, showing the characteristic "horse shoe" shaped constriction first noticed by Gohar and Cameron. In most cases, the minimum film thickness occurred at the sides of the contact, and was quite sensitive to load. The thickness at the rear of the contact was, like the central region, very insensitive to load.

One disadvantage of Foord's system, however, was that the first bright fringe occurred at a thickness of about 1000 Å, and there were indications that an elastohydrodynamic film could exist at thicknesses well below this value. This problem was overcome by Westlake and Cameron (90-92) in a study of lubricant entrapments formed during the normal approach of a steel ball and a glass plate. (An interferometric study of these phenomena was also made by Dowson and Jones in 1967 (94,95).) By depositing a layer of silicon dioxide of controlled thickness over the semi-reflecting layer, the effective zero of the fringe system was shifted, so that films as thin as 100 Å could be observed and measured. They also improved the accuracy by using titanium dioxide as the semi-reflecting layer, to give multiple-beam fringes with correspondingly better definition, and by using a laser as a source. Further precision was achieved by replacing the TiO₂ with a multi-layer dielectric stack, and the technique was later extended to a rolling point contact rig with provision for running at temperatures up to 70°C (93,96). The thickness at the centre and at the rear constriction was measured for a wide variety of speeds, loads and test fluids, and the experimental results were correlated into two film thickness formulae. For the central film thickness they obtained

\[ \frac{h}{R} = K_{1} \left( \frac{U \eta_{0}}{E R} \right)^{a_{t}} \left( \frac{W}{E R^{2}} \right)^{b_{t}} \left( \frac{\alpha E}{E} \right)^{c_{t}} \]  

where \[ K_{1} = 62 \left( \frac{U \eta_{0}}{E R} \right)^{-0.063} \]
\[ a_1 = \frac{1}{30} \log \left( \frac{W}{E' R^2} \right) + 0.89 \]

\[ b_1 = \frac{1}{43} \log \left( \frac{U \eta_0}{E' R} \right) + 0.303 \]

\[ c_1 = 0.55 \]

and for the thickness at the rear constriction

\[ \frac{h}{R} = (0.075 \pm 0.005) \left( \frac{U \eta_0}{E' R} \right)^{0.77 \pm 0.04} \frac{0.14 \pm 0.01}{0.77 \pm 0.03} \]

Using an optical set-up similar to Ford's system Wedeven and Cameron (97, 98, 43) made the first detailed study of the effects of lubricant starvation in a point contact, showing that any expression for film thickness, to be completely general, must include a starvation parameter connected with the position of the inlet lubricant boundary. Unlike earlier optical work, the apparatus bore close resemblance to an actual ball thrust bearing, and the effects of gyroscopic ball spin were noticed. A short, qualitative study was also made of line contact (98, 43) using tapered rollers, but the surface finish of the rollers was rather poor for optical work, and the loads were not very high.

Snidle (99) has recently made a comparison of interferometric and capacitance measurements in point contact, using crossed cylinders of glass and steel. A thin layer of inconel applied to the glass cylinder was used both as a semi-reflector and a capacitor plate. In calculating film thickness from the capacitance measurements, he allowed for the change in shape of the surfaces outside the contact as the load varied, thus removing one of the major doubts associated with earlier point contact capacitance measurements. Although the agreement between the two methods was fair, a systematic difference in the speed index was apparent, particularly at high speeds and low loads, and was no doubt due, at least in part, to the assumption that the film was parallel in the contact.

Further work by Gohar (100) reported in 1970, included the effects of elastic modulus on the film thickness in a point
contact. Using a 220 fold range of elastic modulus he found that the results satisfied the expression for central film thickness:

$$\frac{h}{R} \propto E^{-0.1}$$

The dependence on speed and load was in agreement with previous findings, and could be expressed as

$$\frac{h}{R} \propto \left(\frac{U \gamma_o}{E'R} \right)^{0.7} \left(\frac{W}{E'R^2} \right)^{-0.11} \left(\alpha E' \right)^{0.49} \ldots \ldots \ldots (1.18)$$

The relation held for all materials combinations except for steel on perspex, which had a very low reduced modulus. In this case the speed parameter index was 0.66 and the proportionality constant was reduced by about 30%. In all cases, the minimum film thickness at the rear constriction was found to satisfy the relation

$$\frac{h}{R} \propto \left(\frac{U \gamma_o}{E'R} \right)^{0.78} \ldots \ldots \ldots (1.18)$$

which was in close agreement with the result of Westlake (equation 1.17).

The interferometric technique has recently been applied to an elliptical contact, using a ball and a grooved plate, by Thorp and Gohar. Initial work (101) was restricted to pure sliding, but further studies to include pure rolling are being continued.

1.3.2. Measurement of Pressure and Temperature.

The measurement of pressure and temperature in an elastohydrodynamic contact has always proved to be a difficult task, because of the small size of contact normally encountered. Two basic methods of overcoming this problem have emerged. The first is to scale up the size of the contact, and the second is to use miniature transducers in the form of evaporated metallic strips.

The main drawback associated with scaling up the contact was that the severity of the conditions was restricted, and it was
precisely this severity which characterised much of elasto-
hydrodynamic behavior. Higginson (102) in 1962 measured the
pressure distribution (in the direction of rolling) in the contact
between a rubber pad and a bronze disc, using a 0.02 in. diameter
pressure-tap. Although the results agreed with theoretical
predictions, the maximum pressure encountered was less than
301b/in² and so the pressure-viscosity effect on the lubricant was
absent. More realistic conditions were later obtainable by
Dowson and Longfield (103,104) using rigid materials and relying
on a large effective radius (78 in.) to produce a sufficiently
broad contact. The rubber pad was replaced by a conforming
bronze shoe, and by applying extremely high loads, pressures up
to 12,000 lb/in² were obtained. In addition to pressure
measurements, which again agreed with theory, temperature
measurements were also made using embedded thermocouples, and
showed a rise in temperature in the inlet zone, reaching a maximum
just before the outlet. As the load was increased, so did the
maximum temperature, and the peak moved nearer to the point of
maximum pressure. Knowing the local temperature and pressure
(and hence viscosity) it was possible to calculate the film
profile in the direction of rolling by solving the inverse
hydrodynamic problem. The film profile so obtained confirmed
the theoretical shape, clearly showing the exit constriction.
The pressure and temperature were also measured in the axial
direction (105) and the results indicated that side leakage was
negligible over most of the contact length.

The conditions were still not severe enough, however, for
one to expect the occurrence of the secondary pressure peak
predicted by theory. In 1965, Longfield (106) replaced the
bronze shoe with a steel one and as a result was able to
produce pressures of over 30,000 lb/in². The pressure spike
was still not observed, however, although he reported the
existence of a "rudimentary" peak near the exit under certain
conditions. By reversing the material combination (i.e. a
steel disc on a bronze shoe) it was found that the pressure
curve was altered in shape. Temperature measurements suggested
that this was a thermal conductivity effect.

The pressure distribution in a counterformal contact
was measured by Niemann and Gartner (107), using a cylinder
sliding on the side of a disc, and with a pressure-tap of 0.001 in.
width. Some of their results showed a maximum pressure in excess
of the Hertz value, but the shape of the distribution curve
seemed to be marred by the broadness of the pressure-tap. They
also measured pressure in the axial direction and confirmed the
findings of Dowson and Longfield that side leakage effects were
localised to the very ends of the contact.

A new means of pressure measurement was introduced in 1965
by Kannel, Bell and Allen (79). In the first part of their paper
they report that an attempt to calculate pressure curves from
their X-ray film profile measurements was largely unsuccessful.
It was found that the pressure curve was very sensitive to small
changes in film shape, and it seemed that the inaccuracies of the
X-ray method were therefore enough to frustrate the pressure
calculation. So they went on to measure the pressures directly,
and introduced the miniature transducer technique. Using a disc
machine with quartz discs, they measured the pressure acting on
a thin axial strip of manganin, evaporated onto one of the disc
surfaces, as it passed through the contact. The pressure was
measurable by virtue of the change in electrical resistance. This
method yielded pressure curves similar to theoretical predictions,
but some error was introduced by not accounting for the change in
resistance with temperature. A secondary pressure peak was not
observed, although its existance would in any case have been
masked to a certain extent by the finite width of the strip.

A manganin strip was also used by Cheng & Orcutt in 1965
(77,108) on a glass disc rolling against a steel one. The
temperature distribution was also measured, using a similar
strip of platinum, again from its change in resistance. Both
measurements were related to the film profile by recording the
capacitance between the strip and the steel disc. Knowledge of
the temperature distribution allowed the pressure measurements
to be corrected for temperature effects. Some discrepancies
between theory and experiment still remained, however, including
the absence of a pressure spike in the experimental results,
although a discontinuity in the pressure curve was evident
(see figure 1.2). The temperature distribution demonstrated
the effect of shear heating in the inlet and outlet zones, and
also showed that, under rolling conditions, there was significant
cooling in the central region due to conduction by the surfaces. Under sliding conditions this conduction was no doubt still operative, but was masked by shear heating (see figure 1.2.). The absolute temperatures measured, however, were significantly less than calculated theoretically (24,26). It was later suggested by Cheng and Orcutt (108) that better agreement with theory for temperature pressure and film thickness could be obtained by accounting for the viscoelastic properties of the lubricant. Indeed, it was implied by Fein (109) in 1967 that compressional viscoelasticity could cause the pressure spike to be reduced or removed altogether under certain conditions.

The first experimental verification of the existence of a secondary pressure peak was provided in 1965 by Kannel (110). The quartz discs used in earlier experiments were replaced by steel ones, allowing higher pressures to be generated, and the manganin strip was insulated from the steel by a layer of silica. Some electrical improvements were also made. Under certain conditions a truncated spike was clearly visible on the oscilloscope traces (see figure 1.3). The truncation was to be expected, since the strip was of comparable width to that of the spike. Calculation of the film profile gave only fair agreement with X-ray results, but the conditions were not quite identical and the X-ray method was possibly a little inaccurate. In later work by Kannel, Walowit, Bell and Allen (111), results were obtained for crowned discs (elliptical contact) using a manganin transducer with a narrow sensitive region to cover the centre of the contact only. Temperature measurements were also made, using a titanium transducer. The pressure spike was not as great as with cylindrical discs, but this was possibly due to variations in pressure in the axial direction. Using the experimental pressure distributions, the stresses below the disc surfaces were calculated, both for crowned and cylindrical discs.

Using a modified gauge design, Hamilton and Moore (80) in 1967 provided further proof of the existence of a secondary pressure maximum, and their temperature measurements, using a nickel gauge, were in general agreement with those of Orcutt (77). They also used Orcutt's technique of combining pressure and temperature measurements with capacitance film profile measurements, so that all three parameters could be inter-related.
Surface temperature: no sliding

Pressure distribution

Surface temperature: 22.5% sliding

Film thickness distribution

Figure 1.2.

Sample oscilloscope traces of temperature, pressure and film thickness measurements. The direction of motion is from left to right (after Cheng and Orcutt (108))
Temperature: 115°F  Rolling speed: 1400 ft/min

Figure 1.3.
Measured variation in pressure profile with disc loading (after Kannel [110])
More recently, measurements of pressure distributions have been deduced from photoelastic measurements of the subsurface stresses in one of the rolling pair. Klemz, Gohar and Cameron (112) in 1970, using a glass/steel combination in a disc machine, successfully employed this method, and showed, under certain conditions, a truncated pressure spike. Cope and Hains (113) using a glass/aluminium rolling pair, by observing isopachic fringes as well as the usual isochromatics, derived the inlet pressure sweep from the absolute stress values.


The sliding friction in an elastohydrodynamic contact is governed almost entirely by the conditions in the parallel region of the contact. Measurements of frictional forces have led to the belief that the rheological behaviour of the lubricant in this region is non-Newtonian, in that the viscosity appears to decrease with increasing shear rate. This is not a surprising result in view of the magnitude of the pressures and shear rates anticipated in this high-viscosity region.

In 1965, Dyson (114) gave a possible explanation of non-Newtonian behaviour of mineral oils in elastohydrodynamic contacts, in terms of viscoelasticity. To do this he made a comparison between the continuous shear results of Crook (115) and Smith (116), and the oscillatory shear results of Barlow and Lamb (117), basing the comparison on the viscoelastic model of Oldroyd (118). His treatment was not entirely successful in accounting for experimental friction data, however, but was a useful starting point.

The presence of non-Newtonian behaviour was not expected to have an appreciable effect on film thickness, because the generation of an elastohydrodynamic film was known to depend on the inlet conditions, where the pressures and shear rates were comparatively low. This inference was supported by the fact that the great majority of lubricants were found to generate films whose thicknesses agreed closely with the theories outlined in section 1.2., which, of course, were based on Newtonian assumptions.
Notable exceptions to this case, however, were observed with silicone fluids by Archard and Kirk (119), Dyson and Wilson (120), and Gupta et al (121), and also with polymer-blended mineral oils by Dyson et al (78), Hamilton and Robertson (122), and Foor et al (87). In all cases, the measured film thicknesses were less than those predicted by theory, sometimes by as much as 1 or 2 orders of magnitude. The results of Dyson and Wilson for silicones are shown in figure 1.4. It is seen that the lack of agreement increases with nominal viscosity. The discrepancies were immediately ascribed to non-Newtonian behaviour in the inlet region, since these fluids, unlike most, had been known to exhibit such behaviour in conventional viscometers at comparatively modest shear rates.

Dyson and Wilson went on to produce a new film thickness formula for these silicone fluids, again using the analogy between the behaviour of viscoelastic substances in continuous and oscillatory shear. Using information on the behaviour of silicones under high oscillatory shear rates, published by Barlow et al (123), they predicted the behaviour under continuous shear and used this to deduce the film thickness. The new equation agreed remarkably well with their experimental results, considering the many approximations made in the derivation.

1.5. Lubricant Starvation

The results of the Grubin analysis in section 1.2 indicate that the generation of an elastohydrodynamic film depends on the pressure build-up in the inlet region. Experimental results by Crook (25) confirmed this idea by showing that the film thickness was dependent on the temperature, and hence viscosity, in the inlet. In theory, the pressure build-up must commence at infinity, although for practical considerations one can assume that the inlet boundary is some finite distance away, since the bulk of the pressure generation takes place close to the edge of the contact. Thus, if there is a good supply of lubricant in the inlet region, one can assume that the film thickness is independent of the exact position of the inlet boundary.

Experimental evidence by Crook (55), Lauder (124) and Boness (125), however, suggests that when the lubricant supply is depleted
Comparison of experimental film thicknesses given by silicone fluids and predictions based on a Newtonian theory (after Dyson and Wilson (120))
in the inlet zone, the film thickness is less than expected, and that this is due to the inlet boundary being close to the edge of the contact, causing a late start to the pressure build-up. A complete film thickness expression therefore, should also include a term to allow for the influence of the inlet boundary position.

The inclusion of this effect into classical theory (rigid cylinders, isoviscous fluid) has been studied by several workers. Numerical solutions have been computed by Floberg (126), Dowson and Whitaker (29), Boness (125), and Dowson and Whomes (127), and a semi-analytical approach was reported by Sassenfeld and Walther (128). Most of these solutions have been concerned with the prediction of load capacity. A more useful form of presentation is in terms of film thickness compared with that in the infinite-boundary case. This was done in a semi-analytical solution by Wolveridge, Baglin and Archard (129), who also showed that their result was in agreement with previous computer solutions.

The extension of this approach to elastohydrodynamic conditions has been slow to develop. In 1967, Fein and Kreutz (130) investigated the theoretical effects of lubricant starvation in a hemispherical asperity model, and showed that the vast majority of pressure build-up occurred within a distance

\[ \frac{3(Rh_0)^{\frac{2}{5}}}{a^{\frac{6}{5}}} \]

ahead of the equivalent Hertzian contact (of radius 'a'). A computer solution for a starved line contact was produced by Orcutt and Cheng (131) in 1966, but only for one set of conditions. The semi-analytical classical solution of Wolveridge, Baglin and Archard (129) was also extended to elastohydrodynamic conditions and very good agreement with the numerical result of Orcutt and Cheng was shown.

Starvation in circular point contacts was studied experimentally by Wedeven (97, 98, 43) using optical interferometry to measure film thickness and inlet boundary position. He concluded that the onset of starvation occurred when the inlet boundary was at the position where the gap thickness was about
9 times the central film thickness. It was shown that this point occurred at a distance

$$\frac{3.52 (R \rho \omega)^{\frac{3}{2}}}{a^{\frac{3}{2}}}$$

from the edge of the equivalent Hertzian contact, thus agreeing very closely with the result of Fein and Kreutz. Wedeven combined his empirical results with the theoretical film thickness formula of Wedeven, Evans and Cameron (equation 1.9) to produce the following semi-empirical film thickness formula accounting for the inlet boundary position

$$\frac{b_s}{R} = 1.73 K_s \left( \frac{\alpha \nu_0 U}{R} \right)^{0.714} \left( \frac{W}{E R^2} \right)^{-0.048}$$

where $K_s$ is an empirical starvation parameter, defined as the ratio of the actual film thickness to the film thickness under fully flooded conditions. In reference 98 $K_s$ is given by

$$K_s = 1.24 \left( \frac{S}{S_f} \right)^{0.45} \text{ for } \frac{S}{S_f} < 0.62$$

where $S$ is the inlet distance, and $S_f$ is the minimum inlet distance required to obtain flooded conditions (both distances being measured from the edge of the equivalent Hertzian area). In reference 43 this expression is modified to

$$K_s = \left[ \frac{S}{S_f} \left( 2 - \frac{S}{S_f} \right) \right]^{\frac{1}{4}}$$

The computer solution of Wedeven, Evans and Cameron was extended to include various boundary positions in the inlet, and reasonable agreement with Wedeven's experimental curves was obtained. Computer solutions for the starved line contact case have recently been obtained by Dowson and Castle, and are awaiting publication (132).
1.6. Partial Elastohydrodynamic Lubrication.

A basic assumption made in the theoretical treatment of the elastohydrodynamic problem is that the bounding surfaces are geometrically perfect, and the load is entirely supported by the fluid film. In practice, of course, such surfaces are never encountered, but in a large number of cases the film thickness is significantly greater than the combined average heights of the surface irregularities, and so the rate of asperity-interaction is small enough to be neglected. In the low speed/viscosity range, however, electrical resistance measurements have shown that asperity collisions can be very numerous, and must therefore carry a significant part of the load. The effects of this on wear and fatigue life are obviously of profound importance when considering real bearings.

Experimental studies of this so-called partial elastohydrodynamic regime have been almost entirely based on resistance measurements between the bounding surfaces. The first major contribution to the subject was made by Purey (133) in 1961. Using a fixed ball sliding on a cylinder, he measured the inter-specimen resistance both instantaneously on an oscilloscope and as a time-average on a pen-recorder. The circuit is shown in figure 1.5. The resistance scale switch box provided a voltage of about 15 millivolts across the oil film, regardless of the resistance ratio selected, thus avoiding discharge through the film. The selection of various resistance ratios enabled the mid-scale point on the oscilloscope to represent a resistance of 10000, 1000, 100, or 10 ohms, enabling certain resistance ranges to be examined more closely. A typical instantaneous resistance trace oscillated rapidly between an almost infinite value and an almost zero value, showing that intermittent metallic contact was taking place. By assuming that the resistance was always at one of these two extremes, it was possible to estimate the "percentage contact time". The pen-recorder output provided a record of an entire experimental run, but care was needed in relating the average voltage to the percentage contact, since the relation depended on the resistance ratio employed. By measuring the percentage contact for various speeds, loads and viscosities, the transition from hydrodynamic conditions to boundary conditions...
Figure 1.5.

Circuit used in measurement of electrical resistance
(after Furey (133))
was studied, and showed that metallic contact was more prevalent than previously thought. In addition, it was possible to investigate the time history of metallic contact and friction, and the effect on this of various antiwear additives.

The nature of asperity contacts was more closely examined by Tallian et al (134, 135) using a similar circuit to Purey's connected to a rolling four ball machine. The voltage fluctuations were amplified and fed via a discriminator to an electronic counter, enabling the number of upward crossings past a set voltage level to be counted. Additionally, the fluctuations were recorded on FM tape, which, when replayed at a slower speed through a writing oscillograph, provided a detailed record of the instantaneous resistance trace on an expanded time scale. It was found that when a metallic contact was broken, the voltage rose in an exponential manner, with a time constant dependent on the capacitance of the system and the series circuit resistance. Thus if two contacts occurred in close succession, it was possible for the second to occur before the voltage had risen to its full height. This gave a false reading of average voltage, and, if the voltage rise was cut short below the discriminator level, also gave a false counter reading. A method of extrapolating to a zero time constant using known time constants was devised by Chiu et al (136) and was used to correct the results.

The amount of wear taking place was measured by a radio tracer technique, and it was shown that this was closely related to the extent of metallic contact. The surface microgeometry of the balls was measured using a surface profile recording instrument with a very fine stylus, and by treating these measurements statistically, it was possible to predict the contact time fractions and contact frequencies. These were then compared with experimental values. The computer contact frequency values were rather lower than the experimental values, but the contact time fractions were in good agreement. Using these latter results, the dependence of film thickness on speed was estimated and compared with the theory of Archard (equation 1.9). Although the relative dependence showed some agreement, the absolute values deduced experimentally were too high by a factor of about 3.

Christensen (137) in 1965 applied the resistance technique
to a line contact between two discs, in both rolling and sliding. The measurement circuit again included a discriminator and an electronic counter, and by counting clock pulses he was able to measure the duration of the metallic contacts. A correlation between the count rate and estimated film thickness was obtained, but only over a certain range. The duration of the contacts was studied in detail, and it was found that in pure rolling, some contacts lasted for most of their passage through the Hertzian region, although the majority were much shorter. For this, Christensen suggested that most contacts were made at the rear constriction. When sliding was introduced, the contact duration was typically much shorter, and using a simple model, an estimate of the area of contact was obtained. It seemed that, even under conditions of quite high metallic contact, the contact area was small compared with the hydrodynamic area.

In 1968, Tallian and McCool (138) made a study of individual asperity contacts using a ball rolling on a plane. By using highly viscous fluids they were able to generate films several times thicker than the r.m.s. roughness value at very low speeds. Under these conditions, the number of asperity interactions was so small that each one could be studied as an individual event, using the resistance technique as before. They found that when a contact occurred, the resistance dropped to a few thousand ohms in most cases, but sometimes to only hundreds of ohms. A few very short occurrences showed resistances of 100,000–200,000 ohms. The duration of most contacts was equivalent to a rolled-over distance of 0.1–0.2 times the Hertzian diameter, but many lasted throughout the whole Hertzian region. This was a similar result to that found by Christensen (137). What was very surprising, however, was that occasionally a contact lasted for a length of four to ten Hertzian diameters.

To explain these results, the authors, like Christensen, suggested that asperity contacts often remained throughout their passage through the Hertzian region, but that some occurred only at the rear constriction. They suggested that most interactions were elastic (inferring low slopes of asperities) but that occasional non-elastic asperity interactions resulted in local welding, thus explaining why extremely long contacts were sometimes found. These prolonged contacts were also thought to be possibly due to wear
particle formation. The constriction resistance of a small metallic point contact can be represented by the following expression (139,140)

\[ R = \frac{\rho}{2\pi r} \]

where \( \rho \) is the resistivity of the metal, and \( r \) is the radius of the contact. Substituting typical values for an asperity contact gave a resistance of only a few ohms, about three orders of magnitude less than observed in most cases. Tallian and McCool suggested that the difference was due to thin layers of high resistivity interposed between the metallic surfaces, which could be iron oxides or similar solids, or boundary layers of lubricant molecules adsorbed onto the metal surfaces. The resistivity of such layers is governed by the tunnel effect, and calculations showed that layers of this nature would have a resistance of several thousand ohms, just as observed experimentally. It was then suggested that the occasional low-resistance contacts were the result of penetration of the surface layer, a process which could subsequently produce local welding and wear particle formation.

Although the foregoing experimental work has yielded a better understanding of how asperities interact, it is still not clear how they actually deform in these interactions, and there is no good measure of mechanical forces generated. There is evidence to suggest that asperities do not have as great an effect in elastohydrodynamic contacts as expected. In particular, Westlake (93,96) has shown that elastohydrodynamic action appears to prevail at film thicknesses right down to about twice the combined c.l.a. roughness value, which is considerably less than the valley-to-valley distance. Such a result is not readily explicable, and serves to illustrate the need for more understanding of the problem.

Until recently, very little theoretical work has been reported. The effect of one-dimensional asperities in a hydrodynamic tilted pad bearing was considered by Christensen (141), both parallel to and perpendicular to the direction of sliding. Thompson and Bocchi (142) in 1971 calculated the load borne by asperities for a journal bearing and for a roller on a flat, by
combining the rough-surface theory of Greenwood and Williamson (143) and the elastohydrodynamic theory of Dowson and Higginson. They showed that some load was carried by asperities even at light loads, but was fairly constant up to some critical value, whereupon the asperity-load then increased rather rapidly.

Temperature, because of its profound effect on viscosity, was also shown to be important. A more complete analysis, based on the same two theories, has recently been formulated by Johnson, Greenwood and Poon (144), which showed that if the majority load was carried by the film, the effect of asperities on the bulk film thickness was small. This then enabled them to explain why $h_o/\sigma$ (the ratio of theoretical minimum film thickness to surface roughness) has been found by many experimentalists to be closely related to the rate of wear.
References


35. THEYSE, F.H. "Contact between rolling/sliding surfaces", Wear 1966 Vol.9, p.41.


CHAPTER 2

Lubrication Breakdown Under Boundary Conditions

by

M.W. Bailey

2.1. Introduction.

It has been realised for many centuries that if one surface slides upon another, friction is developed. In most cases this resistance to relative motion may be reduced by the introduction of a lubricant between the two load-carrying surfaces. Archeologists have found remains of a chariot dated at about 1400 B.C. with traces of grease still adhering to the axle, and a now famous Egyptian mural showing a slave applying lubricant to a moving block of stone is dated about 1900 B.C. (1). These illustrate that this knowledge is of very long standing. However, the study of friction and of the problems associated with lubricating moving surfaces has not been undertaken until much more recent times.

Amontons in the late seventeenth century and Coulomb in the early nineteenth century (2) were two among the earlier investigators and developed the basic law of friction, which states that the frictional force is proportional to the load. Since then interest in the study of lubricated and unlubricated frictional phenomena has increased. Now, lubrication technology - as it may be termed - has become an important branch of mechanical engineering science.

There may be considered to be two broad regimes of frictional behaviour in a simple system consisting of two surfaces moving relative to each other in the presence of a lubricant. At high speeds and low loads the two surfaces are completely separated by a film of lubricant whose fluid properties determine the frictional behaviour. The friction coefficient is very low - typically of the order of 0.001-0.01 - and wear of the surfaces is negligible. This is known as the hydrodynamic regime. The second regime is termed the boundary regime. This occurs when the conditions are more severe in that the unit loading is higher and the sliding speed lower. Here, the thickness of the film of lubricant is of the same order as the surface asperities and so there is no longer complete separation of the two surfaces. Some asperity contact is likely. The friction coefficient is much higher than the hydrodynamic case - of the order of 0.1 - and there may be very
considerable wear. The governing parameters are in this case the chemical characteristics of the surface materials and the lubricant and their interactions one with the other. In practice however, many mechanical systems operate in regions which are neither truly boundary nor hydrodynamic.

The work described in this chapter is an investigation into some aspects of failure or breakdown of lubricants operating under boundary lubrication conditions. Later, some of the literature in the field is briefly considered. Only the relevant references taken from the huge mass of available information are discussed.

2.2. Boundary Lubrication

One of the difficulties of boundary lubrication research was stated very clearly by Bristow in 1947 (3), ".... the investigator of boundary-friction phenomena is in the somewhat difficult position of trying to obtain fundamental information by means of something (friction) the mechanism of which is not understood". Since that time our knowledge and understanding has increased considerably, and also the present day investigator has many more tools at his disposal than merely measurement of friction. However, the basic problem is still the same, that of often having to interpret the results of experimental determinations of macro-phenomena (such as the friction coefficient) to gain fundamental understanding of the molecular processes involved.

One of the early workers in the field of boundary lubrication was Hardy (4). He carried out a considerable number of experiments using lubricants containing fatty acids. From the results it was proposed that boundary lubrication was made possible by the presence of an adsorbed layer of polar molecules on the surface. His model is shown in fig.2.1.

Further work was carried out following the introduction of the Bowden-Leben machine in 1939 (5). The machine has a hemispherical slider traversing a flat plate. Measurements of friction and temperature could be made. Bowden (6) modified Hardy's model to allow for the fact that in practice the sliding surfaces are never flat or smooth. This means that there must be some interaction between the surface irregularities or asperities under thin-film boundary conditions. Such interaction
Figure 2.1.
Hardy's Model

Figure 2.2.
Bowden's Model
is assumed to displace the adsorbed film and is shown in Fig. 2.2. This second model is obviously an advance over the first but it still does not fit all the experimental data. Adamson (7) much more recently has considered both these and their shortcomings and proposes a third which he shows to fit the data more closely. For Bowden's model to apply, the load has to be supported by an area \( A \), of which a proportion \( \alpha A \) consists of metal-metal contacts of shear strength \( S_m \) and a proportion \( (1-\alpha)A \) consisting of film-film interactions of shear strength \( S_f \). Hence the frictional force \( F \) may be expressed as

\[
F = A \left( \alpha S_m + (1-\alpha)S_f \right)
\]

or referring to Fig. 2.2.

\[
F = A_1 S_m + A_2 S_f.
\]

It is known that the proportion of metal-metal contact is very small and so most of the friction must be developed in the film. This leads to certain objections. For Amonton's law to hold for boundary friction, \( A_2 \) should be proportional to the applied load.

His proposal to explain this and other objections is shown in Fig. 2.3. Under light loads the situation is similar to that shown in Fig. 2.2. There is a very small proportion of the contact in actual metal-metal contact and the remainder is covered by a relatively ordered layer of polar molecules. As the load is increased the highest asperities deform plastically locally and much of the remainder of the contact suffers elastic deformation. Very few new metal-metal contacts are made. The elastically deformed surfaces pressurise the entrapped lubricant, the molecules of which become entangled. It may therefore be considered that under normal conditions, the relatively large area of elastically deformed surface pressurises the lubricant and supports much of the load and the entangled chains account for much of the friction. Adamson shows further how the model may be seen to agree quite well with experimental observations. The hypothesis of chain entanglement might be valid for molecules containing more than twenty carbon atoms but is inadequate for shorter chain lengths.
Returning to the earlier work, Tabor (8) in 1941 carried out experiments with the Bowden-Leban apparatus using various metal surfaces with solutions of fatty acids. These had previously been shown to be good boundary lubricants. He found that the measured transition temperature, at which the friction rose and the wear track became torn, was considerably above the bulk melting point of the particular fatty acid. It corresponded however quite closely with the melting point of the metal soap. This assumed that some chemical reaction had taken place. Further studies were made by Mentor and Tabor (9) and Sanders and Tabor (10). It became evident that for chemically unreactive surfaces – in this case platinum – the transition temperature was much lower and coincided with the temperature at which the fatty acid molecules became disoriented. Hence it was concluded that there were two distinct effects. One was with unreactive metals where the transition occurred when the adsorbed film became disoriented. The second was with reactive metals where a soap was formed. It was necessary for this to fail, possibly through melting, for lubricant breakdown to occur.

Frewing (11) carried out experiments also using the Bowden-Leben apparatus with solutions of long-chain acids and halides in white oil. He assumed that a transition from smooth sliding to stick-slip occurred when only a certain constant fraction of the available adsorption sites on the surface was occupied by the polar molecules. Above that critical fraction there is smooth motion, below that fraction the movement is stick-slip. He then applied the following argument.

The rate of build-up of the film may be taken as

\[ k_1 c (1 - \theta) \]

and the rate of desorption as

\[ k_2(\theta) \]

where

- \( k_1, k_2 \) are rate constants
- \( c \) is concentration
- \( \theta \) is proportion of available sites occupied.
Under conditions of equilibrium,

\[ k_1 \, c \,(1 - \theta) = k_2 \, \theta \]

so

\[ k = \frac{k_1}{k_2} = \frac{\theta}{c \,(1 - \theta)} \]

where \( K \) is equilibrium constant.

Applying the Van't Hoff isochore for the temperature dependence of \( K \)

\[ \frac{d(\log_2 K)}{dT} = -\frac{U}{RT^2} \]

so,

\[ \log_2 K = \frac{U}{RT} + \text{constant} \]

where \( U \) is heat of adsorption

\( R \) is Universal Gas Constant

\( T \) is absolute temperature.

Assuming that the transition occurs at a point where the proportion of occupied sites has decreased to a constant value say \( \theta = \theta_0 \)

then \[ \log_2 \left( \frac{\theta_0}{c \,(1 - \theta_0)} \right) = \frac{U}{R T_0} + \text{Constant} \]

where \( T_0 \) is transition temperature

and so,

\[ 2 \log_{10} C = -\frac{U}{R T_0} + \text{Constant}. \]

Hence, a plot of \( \log c \) versus \( 1/T \) should give a straight line, the gradient of which may be used to calculate \( U \).

Bigelow, Glass and Zisman (12) investigated the effects of concentration on transition temperature using a somewhat different technique. A carefully cleaned rectangular piece of platinum foil was dipped into solutions of long-chain polar compounds and then withdrawn. On withdrawal the solution was observed to roll off the surface leaving what appeared to be a dry surface, but was in fact an oleophobic monolayer of adsorbed polar molecules. The dipper was replaced and the temperature of the solution increased until the point at which the foil remained wetted on withdrawal.

This they suggested indicated the desorption of the monolayer. In the same way as Frewing they plotted \( \log c \) versus \( 1/T \) where \( T \) was...
now the temperature at which the foil wetted. They found linear relationships. The heats of adsorption were again calculated from the gradients of the curves and found to agree quite well with Frewing's results.

This work hence gave further support to the hypothesis that lubricant failure occurred when the surface film desorbed.

A considerable amount of work has since been carried out on the thermodynamics of the adsorption-desorption process. Askwith, Crouch and Cameron (13) published an extension of Frewing's treatment. Their analysis is as follows:

Assuming the adsorption-desorption process is in equilibrium (which also implies reversibility) then the equilibrium constant K may be given by

$$K = \frac{\text{(covered surface sites)}}{\text{(clean sites) (polar molecules in soln.)}}$$

If the ratio of covered sites to clean sites is \( \frac{\theta}{1-\theta} \) and the standard state for the free solute is 1 mole/1000 gm of solvent, the concentration is denoted by c. Applying the Langmuir isotherm,

$$K = \frac{\theta}{(1-\theta)c}$$

The standard free energy change is,

$$\Delta G^o = -RT \log_e K$$

and also for an isothermal process,

$$\Delta G = \Delta H - T \Delta S$$

where \( \Delta G \) is the change in Gibbs free energy, so

$$- \frac{\Delta H^o}{R} + \Delta S^o = R \left( \log_e \frac{\theta}{1-\theta} - \log_e c \right)$$

From this equation it can be seen that if the transition occurs at a constant value of \( \theta \), and \( \Delta S^o \) and \( \Delta H^o \) are independent of temperature, then a graph of \( \log_e \frac{c}{T} \) versus \( \frac{1}{T} \) should be a straight line. Good straight lines are obtained which adds justification to the assumptions.
From the plot of \( \log C \) vs. \( \frac{1}{T} \), the intercept can be measured. If \( \Theta \) is known, the standard state entropy change \( \Delta S^0 \) can be calculated. This was done by Askwith et al assuming, with small justification, that \( \Theta = 0.5 \) at scuffing. This standard entropy change was termed the entropy change of scuffing.

The work described above has given some insight into the possible action of chemically pure boundary lubricants. Much work has been undertaken in the investigation of the mode or modes of failure of practical lubricants. This has often been with a view to developing a method whereby the impending failure of a boundary lubricated mechanical system might be predicted. The following section describes some of the work in this extensive field.

2.3. Lubricant Failure in Sliding Contacts.

The failure of a lubricant as evidenced by scuffing is of great importance. If a criterion for incipient failure could be formulated with confidence it would be an enormous aid to the designer. It would then be possible to design safely up to the limit of the lubricant and of the materials used. Unfortunately this is not so at present. As was noted by Manton and O'Donoghue (14) quite recently, the only satisfactory and acceptable method of testing a lubricant is still that of a trial in the actual machine under the particular conditions of operation.

In 1937, Blok (15) suggested that the failure of a lubricant was due to the local temperature in the contact rising above a certain critical value. This value was characteristic of the metal-oil combination. He analysed the results of Almen (16), who had investigated the failure of spiral gears, and showed that the temperatures developed on the tooth faces were an important factor in seizure. The hypothesis of a failure (or critical) temperature was extended in 1939 (17). The critical temperature of a plain mineral oil was then considered to be almost independent of load, speed and bulk oil temperature - providing the bulk temperature was not so high as to cause chemical effects.

The temperature developed at the surface - the total temperature \( T_T \) - was the sum of the bulk or skin temperature \( T_S \) and the
temperature rise due to the friction developed in the contact termed the flash temperature $T_f$, 

$$T_T = T_s + T_f$$

The temperature $T_s$ may be measured by conventional means but measurement of the flash temperature is very difficult and in most cases impossible. The contact size is usually small and the temperature gradients are very great and so temperature measurements are not possible using the smallest available thermocouples. For this reason it is usually necessary to estimate the magnitude of the flash temperature theoretically.

Blok presented such an expression (18) and a few years afterwards Jaeger (19) published a more sophisticated theoretical treatment, and there have been others more recently.

Since Blok formulated his hypothesis of a single critical temperature for a particular metal-lubricant combination with plain mineral oils, much research work has been carried out attempting to test its validity in both practical and idealised cases. Gear designers are constantly investigating the possibility of a reliable design criterion, and it is for that reason that much work on the critical temperature hypothesis has been carried out by specialists in that field. Lane and Hughes (2) with a gear machine and straight mineral oils, found that within certain limits, scuffing of their gears could be predicted. They pointed out however that the empirical relation between scuffing load, speed, oil viscosity and operating temperature derived from their results was obtained using only one set of gears and would not necessarily apply to any other.

Kelley has produced several valuable papers on the topic of gear failure. In one of his earlier ones (21) he reports the results of tests using a gear machine and discusses the problems of applying the critical temperature hypothesis. He found a reasonable correlation between scuffing and a constant total temperature. One of the difficulties encountered was that of estimating the coefficient of friction. This has a very direct bearing on the magnitude of the flash temperature.

The problem of the estimation of important parameters is one
of considerable relevance as gear-machines and disc-machines have been used quite extensively for lubricant testing. These machines are well suited to the task of simulating gearbox conditions and have produced much very valuable information. They do however introduce some variables which cannot always be measured or estimated accurately. This renders the drawing of fundamental conclusions difficult. The problem of the application of the critical temperature hypothesis to gears is considered in an excellent review paper by Kelley and Lemanski (22). It describes some of the more recent work in the field of gear lubricant testing.

Benedict and Kelley (23) studied the effects of load, rolling velocity, sliding velocity and oil viscosity on the coefficient of friction using a roller machine and found very significant variations. However, as they indicate, in most work with gears or discs the coefficient of friction is assumed constant and an average value based upon experience is used. This of course leads to uncertainties in the estimation of flash temperature.

Fein, Rowe and Kreuz (24) published results of experiments using a pin-on-disc machine with lubricants consisting of solutions of fatty acids in pure hydrocarbon solvents and also a straight mineral lubricant. They measured the transition temperatures and detected a dependence upon load and speed. A linear relationship was established empirically between the reciprocal of the transition temperature and the logarithm of the load-speed ratio.

Fein (25) described further experiments using a four-ball machine and found a similar relationship to hold. It was also seen that the results from the four-ball machine could be considered to be an extension of the data taken on the pin-on-disc machine. Hence there existed a correlation between the two bench test machines. They did however appear to contradict the constant critical temperature hypothesis.

Meng (26), with a disc machine in which the two discs were of different steels so that the contact zone acted as a thermocouple, found that over the range of speeds used (sliding speeds of 0.25 - 1.75 m/sec) the temperature measured was substantially constant.
The scuffing experiments of Leach and Kelley (27), using a conventional disc machine, showed that for a given material-lubricant combination the failure temperature, as calculated from Blok's expression, was constant over a wide range of speed, load and operating temperature. They attempted to extend the work to include a few reactive additive lubricants, but without success.

In 1964 Ali and Thomas (28) published results of work on a disc machine using plain mineral oils where a constant failure temperature of the lubricants was exhibited above a sliding speed of 670 ft/min (3.4 m/sec). Below this speed the failure temperature increased slightly. However De Gruchy and Harrison (29) described tests on a gear rig the results of which O'Donoghue and Cameron (30) showed to exhibit a substantially constant failure temperature over a wide range of speed. The work of Shipley (31) is also cited who analysed some of the classical work of Mansion using an IAE gear machine. It was shown that there was a constant scuffing temperature for each of the three oils used.

Thus it may be seen that there is some difficulty with the critical temperature hypothesis, as different workers using different test machines produce seemingly contradictory results.

The problem of contradictory results from different test rigs has been long recognised. A sub-panel of the Institute of Petroleum attempted to correlate the results from two gear rigs, the IAE gear rig and the Ryder gear rig (32). These are both widely used in England and America for lubricant evaluation studies. Their findings were that in terms of scuffing load, under what were apparently identical situations, there was very little correlation. It was recognised however that some important factors of which there was no data from the standard tests had probably been overlooked. However a comparison between the results of the German gear rig, the PZG rig and those of the Ryder and IAE rigs was made by Niemann and his co-workers (33). When testing the scuffing resistance of six lubricants although the values of the scuffing loads obtained from each machine varied considerably, each machine graded them in the same order. This would seem to indicate that the value of scuffing load obtained from each machine was dependent on geometric factors and test procedure. Fowle (34) used the critical temperature
theory to attempt a correlation between the tests on the FZG and IAE rigs. He showed a reasonable correlation between the two rigs assuming a constant failure temperature for a given lubricant-material combination. In carrying out his comparison, some estimates were necessary (in particular that of bulk gear temperature) since measurements could not be taken. This again highlights the problems encountered in analysing the results published from gear and disc machines.

On the subject of correlation, several investigators have carried out tests to check the validity of the expressions derived by Blok. These estimate the gear tooth surface flash temperature and this component is very important. It may easily constitute the major part of the total temperature of the contact. Niemann and Lechner (35) used pairs of mating gears of different materials and utilised the thermoelectric effect in the contact - a similar technique to that used by Meng described earlier. There are limitations to this technique. High load capacity gear materials such as case-hardened steels produce very small thermoelectric effects. Combinations such as constantan-steel which produce large thermo-emfs are limited in their load carrying capacity. Also the temperature measured is some average of the range of temperatures existing in the contact zone. It was found that subject to these limitations, Blok’s flash temperature expressions agreed with the results obtained. A second very significant result of the investigations was that it was shown that gear scuffing was the result of some type of cold welding process since the melting points of any of the materials was not reached at any time.

Terauchi and Miyao (36) report work of a similar nature to Niemann and Lechner using only a constantan-steel pair of gears. They also conclude that Blok’s expression for the flash temperature of gear teeth was in broad agreement with the temperatures measured.

Matveevsky (37) has carried out much work using a four-ball machine and other bench test machines on the testing of the critical temperature hypothesis. He describes tests using four mineral non-additive lubricants at different sliding speeds. The results obtained using the four-ball machine, crossed cylinders and disc machine were compared. It was found that the total
temperature calculated just before failure was constant for a
given material-lubricant combination for each test machine.
There was also close agreement between the machines used. In an
earlier work (38) he described how transition temperatures had
been measured using a four-ball machine at a very slow sliding
speed (0.4 mm/sec). Tests were carried out at various ambient
temperatures until the temperature at which the friction rose
abruptly was found. The very slow sliding speed meant that
frictional heating of the specimens was negligible and so is
similar to a Bowden-Leben test. This rendered it unnecessary
for estimations of the flash temperature to be made since the
ambient temperature was exactly equal to the contact temperature.
It was observed in the later work that the value of the total
temperature calculated for the point contact machine just before
failure agreed quite closely with the transition temperature
measured in the slow-speed test. This suggested that the
failure in the high-speed tests was due to the same or at least
a similar phenomenon to that which caused failure in the slow
speed tests. Grew and Cameron (39) reported friction transition
temperature tests using solutions of the naturally occurring
surface active material from mineral oils in hydrocarbon
solvents. The surface active material was isolated by Groszek
and Palmer (40). The results of the test showed these solutions
to behave in exactly the same manner as solutions of fatty
acids. It was therefore concluded that the friction transition
phenomenon measured with mineral oils was due to the desorption
of the natural surface active material.

More work which supported the critical temperature
hypothesis was reported by O'Donoghue, Manton and Askwith (41)
using a disc machine and a wide range of plain mineral oils.
A revised technique was used. The discs were rotated at very low
speed so that the frictional heating was negligible. They were
fully immersed in the bath of lubricant which was slowly heated.
Using stainless steel discs, a sharp transition from low to high
friction was observed. No transition phenomenon was observed
over the temperature range available (up to 220°C) unless
stainless steel discs were used. It was assumed that the reason
for this was chemical attack on the disc surface by the oil,
this masked the transition effect.

Investigations which might cast doubt on the existence of a constant failure temperature criterion and which clearly indicate that much care has to be exercised in its application are those reported by Fein quite recently (42). As described earlier he had investigated the failure of mineral lubricants using the four-ball machine, pin-on-disc machine and disc machine. It had been concluded that there existed a relationship between the failure temperature and the load and speed. He now investigated the effects of different running-in and test procedures using a disc machine and non-reactive mineral oils. The results indicated that dependent upon the complete running history, so the magnitude of the failure temperature could vary by as much as 300°C. No metallurgical reasons for this variation in failure temperature were found. It was concluded that chemical reaction was the cause, even though the lubricants were considered non-reactive.

Fein and Kreuz (43) investigated the wear of hardened steel specimens using a four-ball machine with benzene and cyclohexane containing different concentrations of oxygen as lubricants. They found that with certain intermediate oxygen concentrations, a very low wear rate was measured, and an organic substance was formed on the rubbing surfaces. This was termed a 'friction polymer'. It was this substance which was covering the surface and reducing the rate of wear. They further showed that the formation of the polymer could be explained by catalysis. The freshly rubbed sites exposed clean metal surface. This metal surface acted as a catalyst for the subsequent chemical reaction.

Hermance and Egan (44) reported the formation of a dark amorphous substance which was formed on platinum rubbing surfaces in the presence of organic vapours. Their work is described by Chaikin (45). Chaikin investigated further the mechanisms of formation of the substance. It was concluded that the structureless product was formed by a combined process of catalysis and charged particle bombardment.

From this work it can be seen that lubricants which under normal conditions are assumed to be chemically unreactive, can, under the severe conditions encountered in sliding contacts be subject to some degree of chemical change. The presence of the
reaction products can modify the lubrication and friction behaviour as was shown by Fein et al (43).

Hence, it may be said that from the information published there would appear to be little doubt that excessive thermal conditions are usually the cause of failure under sliding conditions. Bathgate and Yates (46) recently applied three failure criteria to worm gears. The first was critical minimum film thickness, the second a surface fatigue criterion and the third the critical temperature criterion. They concluded that the third was the most viable. Leach and Kelley (27) showed how the film thickness just prior to scuffing varied over a range of seven to one. Early work by Ibrahim and Cameron (47) indicated that scoring often occurs after a sudden collapse of the oil film. This would all seem to point to lubricant failure being caused by excessive local temperatures.

The contradictions occur however in the evidence published in support of this failure temperature being independent of other variables. As has been shown, some investigators report constant failure temperatures, others do not. Blok recently published an extensive review (48) where he defended the case of the failure temperature of non-additive mineral lubricants being constant. He suggested many analytical and experimental refinements. It has been shown however that much evidence is available which clearly indicates non-constant failure temperatures. As was indicated by the discussion which followed the review, the basic need was for an understanding of the mechanisms of failure. Once such an understanding was achieved, the constancy or not of the failure temperature would be of secondary importance, if its value could be predicted.

From this review, three conclusions may be drawn.

(a) In most cases, the failure of a non-reactive mineral lubricant is due to excessive thermal conditions.

(b) With such a lubricant, this failure temperature is sometimes constant. In other cases it is not independent of other variables.

(c) Investigations are needed to establish failure mechanisms. This would in turn explain why under certain conditions lubricants fail at a constant temperature and under others they do not.
REFERENCES


8. TABOR, D. Desorption or 'Surface Melting' of Lubricant Films, Nature 147, 609 (1941).


43. FEIN, R.S., KREUZ, K.L. Chemistry of Boundary Lubrication of Steel by Hydrocarbons, Trans. Asle. 8 (1965).


45. CHAIKIN, S.W. On Friction Polymer, Wear. 10 (1967).


CHAPTER 3
The Action of Oil Additives
by
H.A. Spikes

3.1. Introduction to Extreme Pressure Additives.

Extreme pressure additives were first widely used in the 1930's, as the loads and frictional temperatures had become too high, in some applications, for the traditional mineral oils to withstand. It was then found that the addition of certain chemically reactive organic compounds enabled higher loads and temperatures to be used. These chemicals were therefore called "extreme pressure" additives.

Effective extreme pressure (or e.p.) additives usually contain one or more atom of an element that can react with metal surfaces under boundary lubrication conditions. These are the non-metals to the right of the periodic table, notably sulphur, chlorine, phosphorus and iodine.

It has been found that, in very general terms, an e.p. additive functions by reacting chemically with a rubbing metal surface to form an adhering film which has a lower shear strength than the metal itself. The film then acts as a solid lubricant and takes over the task of lubrication when metal to metal contact would normally occur.

A successful extreme pressure lubricant is designed to react only when temperatures are reached such that normal, oily boundary lubrication breaks down. This prevents continual corrosion of the metal surface under ordinary running conditions.

Although the overall mode of action of e.p. additives has been understood for a long time, only recently have experiments been done to investigate the exact mechanism of formation of the protecting layer. This is remarkable since appreciation of the mechanism would seem to be essential in the design of a successful lubricant.

3.2. The use of sulphur.

The sulphur containing extreme pressure additives have tended to become progressively more complicated in structure over the years. The simplest, and the first used, was sulphur itself, merely
dissolved in the lubricating oil. One of the earliest known references to lubricant additives is a patent of 1859 describing the addition of flowers of sulphur to a mineral oil. This is mentioned by Brown, (11). Since there are naturally occurring fractions of sulphur and sulphur compounds in mineral oils these must have been acting, unrecognised, as e.p. agents long before their value was appreciated.

The most commonly used sulphur additives today are sulphurized fats, olefins and terpenes Forbes, (2). Their detailed structure is not usually known, since they are manufactured by sulphurizing complex mixtures of naturally occurring terpenes, unsaturated fatty esters and similar compounds. Sulphur is introduced across the double bond.

Typical structures are shown by Ford (3) to be:

\[
\begin{align*}
\text{CH}_3 & - (\text{CH}_2)_{x} - \text{CH} - \text{CH} - (\text{CH}_2)_{y} - \text{CH} - \text{CH} - (\text{CH}_2)_{z} - \text{CH} \\
\text{S} & - \text{S}
\end{align*}
\]

a sulphurized terpene

\[
\begin{align*}
\text{CH}_3 & - (\text{CH}_2)_{x} - \text{CH} - \text{CH} - (\text{CH}_2)_{y} - \text{C} \\
\text{S} & - \text{S} - \text{R}
\end{align*}
\]

a sulphurized fatty ester

Also sometimes disulphides are used. These are especially amenable to experimental analysis, since they are pure chemicals and thus have a well-defined structure.

\[
\begin{align*}
\text{C}_6\text{H}_5 & - \text{CH}_2 \\
\text{S} & - \text{S} - \text{C}_6\text{H}_5
\end{align*}
\]

dibenzyl disulphide

3.3. Methods of investigating extreme pressure activity.

There have been two basic ways of investigating the action of sulphur e.p. agents. The first is the chemical approach. A large number of chemical techniques has been used to study the film produced on metals by sulphur additives, both in rubbing and static
conditions. Radiochemical methods have been particularly useful, with S-35 as the radioactive species, Manteuffel et al (4), Sakurai et al (5) and Campbell (6). Other methods which have been employed include electron probe microanalysis, Allum & Forbes (7), x-ray fluorescence, Bailey (8), electron diffraction, Godfrey (9), x-ray diffraction, Sakurai et al (10) and thermogravimetric analysis, Matveevsky (11) as well as straightforward chemical analysis, Frutton et al (12).

The other approach has been to observe the actual lubricating properties of various additives as a function of chemical structures, Allum and Ford (13), e.p. concentration, Forbes et al (14) and metal type (8).

Both these types of investigation have led to certain mechanisms being advanced as to the lubricating properties of sulphur additives. These mechanisms will be considered in section 3.7. Before then, some of the specific experiments investigating the effect of temperature, load and speed on e.p. activity will be described.

3.4. Effect of Temperature on e.p. films and their lubricating properties.

Although it has long been recognised that the function of an e.p. agent is to react with the metal surface under high temperature conditions, remarkably little systematic work has been done (or at least published) concerning the effect of temperature on e.p. reaction. This fundamental importance of temperature was recognised by Bowden (15) who suggested that the name "extreme pressure" agents would be better changed to "extreme temperature".

Frutton et al. (12), in 1946, showed that sulphur organics could form a sulphur-containing film at high temperatures. They mixed iron powder with e.p. solutions at 250°C and by subsequently adding acid were able to measure the extent of reaction from the volume of H₂S evolved.

Some years later, Elliot and co-workers (16) investigated the kinetics of reaction of sulphur and dibenzyl disulphide on steel at three temperatures. They used a radioactive technique and showed that reaction continued indefinitely with time within the limits of their experiments. At 100°C there was very little reaction but this increased at 150°C and 200°C. Dibenzyl disulphide
reacted less than sulphur at 150°C but gave a thicker film than sulphur at 200°C. Borsoff and Wagner (17) and Loeser and co-workers (18) carried out the same sort of experiments but under both static and rubbing conditions. Borsoff and Wagner studied the static reaction of dibenzyl disulphide on steel strip in the range 25 - 190°C. They found reaction rate increased with time and temperature. In similar experiments in a gear system they showed a constant film thickness was produced, which was between five and thirty monolayers deep. Loeser, Wiquist and Twiss (18) carried out static experiments to measure the reaction of zinckialkyl-dithiophosphates on cast iron. They obtained about three monolayers of sulphur at 100°C after twenty hours compared to twenty monolayers at 160°C. In dynamic loaded conditions twenty-five monolayers were formed in twenty hours at 100°C.

An elegant technique for measuring the reaction of sulphur e.p.'s at very high temperatures was devised by Barcroft (19). He measured the change in resistance due to corrosion in an e.p. reaction. From the changing resistance he obtained kinetics of reaction up to 700°C. Barcroft showed that the reaction rate obeyed the Arrhenius equation in the range of 500-700°C for dibenzyl disulphide on stainless steel.

Sakurai et al. (10) used Barcroft's technique to investigate further the rate laws governing the reaction of e.p. agents. They found that corrosion of iron and stainless steel usually followed a parabolic law. The exception to this was with dibenzyl disulphide on iron, which at high temperatures reacted abnormally quickly. This Sakurai ascribed to the formation of cracks in the surface of the metal and sulphide film which permitted further sulphur penetration. No such effects were observed with stainless steel. Typical thicknesses obtained for dibenzyl disulphide on stainless steel were 30nm at 410°C, 80nm at 560°C and 160nm at 650°C, all after ten minutes. On iron the films were generally about ten times thicker.

Manteuffel and co-workers (4), in 1964, investigated the effect of temperature and load on lubricating properties of various sulphur additives. For dibenzyl disulphide they observed a sharp increase in reaction in the 120°C region, where sulphur film thickness rose rapidly in both the rubbing and unscarred regions.
Dorinson and Broman (20) investigated the reaction of two aliphatic disulphides on iron powder in the temperature range 165-250°C. They showed that at least five monolayers of ferrous sulphide could be formed without altering the observed zero-order reaction rate.

Recently Matveevsky (11) has measured the reaction of various e.p. additives with iron powder using thermogravimetric analysis. He relates the degree of reactivity of each e.p. additive to its lubricating capacity.

A good deal of work has been done in the study of corrosion to investigate the action of extreme pressure type chemicals on pure metals such as iron and nickel. Thus Llopis, Arizmendi and co-workers (21) (22) (23) have looked into the effect of temperature on the reaction of sulphur, mercaptides and organic sulphones on iron and nickel. Generally speaking reaction rates always increased with increasing temperature although the various rate constants were different in different temperature ranges.

3.5. The effect of sliding load and speed on extreme pressure activity.

In general, increasing load and rubbing speed seems to produce a thicker e.p. film. Loeser et al. (18) showed that in the dynamically loaded case twenty-five monolayers of sulphur were produced on cast iron using zinc dialkyl dithiophosphate. This compared to only three monolayers under static conditions. The same effect was noted by Borsoff and Wagner (17) using dibenzyl disulphide in a gear system. The film depth increased from ten to thirty monolayers of sulphur in the range of load studied. Borsoff and Wagner noted that the actual film thickness at a given load reached a constant value with time shortly after the run was started.

Campbell (6) also investigated the reaction of radioactive dibenzyl disulphide on hypoid gears during running. He showed that a steady film thickness of 100nm was reached on the driving faces of gear teeth. On the non-pressure flank the thickness was about half this value.

Manteuffel and co-workers (4) measured the quantity of sulphur in both the loaded wear track and the unscarred metal surface at different loads in a 4-ball test rig. They found the sulphur content of the wear track to be about 100 times as great as
that of the unscarred region.

Barcroft and Daniel (24) calculated the thickness of sulphur e.p. films in hypoid gears. They used S-35 and found average film thicknesses of about 50nm. for an unscuffed gear surface, assuming 25% of sulphur in the film. This corresponds well with Borsoff's values (17). Barcroft and Daniel also measured the depth of penetration of reacted sulphur by progressively polishing away known thicknesses of the gear surface and measuring the radioactivity remaining. The limit of penetration under normal conditions proved to be about 700nm. For shock-loaded crown-wheel teeth the penetration was much deeper, however, and radioactive sulphur was detected at 5,000nm. depth.

Recently Allum and Forbes (7) have measured the film thickness of e.p. sulphur films in the wear track of a 4-ball rig using electron probe microanalysis. They estimated maximum film thickness under these, loaded, conditions as 300-500nm.

3.6. **Anti-wear and anti-seizure effects of extreme pressure lubricants.**

Davey and Edwards (25), in 1957, differentiated between two lubricating properties a sulphur additive could have. These were the production of a chemical film to prevent scuffing - an "anti-seizure" film, and the formation of a film on the bare metal or on the previously formed chemical film which reduced friction and wear, - an "anti-friction" or "anti-wear" film. The first of these could be measured in terms of the maximum load to seizure. The second was represented by the wear scar diameter before scuffing. Davey and Edwards found that different disulphides and monosulphides could have better or worse anti-wear and anti-seizure properties depending on their chemical structure.

Allum et al (13) (26) have extended this work and suggested a mechanism for the different effects, which will be discussed in the next section.
3.7. **Mechanisms of extreme pressure action.**

Prutton et al. (12) were the first to suggest a specific mechanism for the formation of sulphur extreme pressure films. They observed that dibenzyldisulphide was more reactive than phenyldisulphide and monosulphides and suggested the mechanism:

$$ R_2S_x + Fe \rightarrow FeS + R_2S_{x-1} $$

for disulphides although they could not ascertain whether the additive split up on the iron surface to give sulphur prior to the main reaction.

A good deal of work on metal–organic sulphide reaction mechanisms has been done in corrosion studies. The reaction of monosulphides on iron has been investigated by Titz (27) and Baladin (28). Titz and his colleagues suggested that mercaptan was formed in solution by thermal decomposition and this then reacted with iron. Baladin, however, could find no evidence of mercaptan and suggested that decomposition took place on the surface according to the sequence.

$$ Fe + C_{n}H_{2n} + 1SR \rightarrow C_{n}H_{2n} + Fe...H SR \rightarrow C_{n}H_{2n} + FeS + HR $$

Llopis et al. (21) investigated the surface reaction of a series of organic sulphides on iron. They showed that the speed of reaction was in the order

$$ R-S-R < R-S-S-R < R-S-H < R-S_{2}-R < R-S_{4}-R $$

in nitrogen but that air enhanced the reactivity of mono- and disulphides whilst supressing the mercaptan and higher sulphide reactions. Llopis proposed a reaction of disulphide via the radical sequence.

$$ R_{2}S_{2} \rightarrow 2RS \rightarrow 2Fe \ldots..SR \rightarrow FeS_{x} + R_{2}S $$

and he suggested that oxygen increased radical formation.

Davey and Edwards (25), using evidence obtained in lubrication tests, put forward a three stage process for the reaction of disulphides on steel. The first two involved the formation of an adsorbed layer, and a mercaptide layer.
Both these products, they suggested, protected the surface, like a fatty acid film, to reduce wear.

Finally under conditions of heavy loading and temperature these gave ferrous sulphide with good anti-seizure properties but only poor anti-wear properties.

This mechanism, which is very similar to Baladin's (28) has been greatly developed by Allum et al. (13) (26), who suggested that the efficiency of a given disulphide as an anti-wear or anti-seizure lubricant depended on its structure.

Allum and Ford (13) showed that the anti-seizure activity of different disulphides was related to the strength of the C-S bond. The weaker this bond, the more effective the additive in forming a sulphide film and thus preventing scuffing.

Later Allum and Forbes (26) demonstrated that anti-wear efficiency was related to the strength of the S-S bond, so that the weaker this bond the more effective the compound as an anti-wear additive. Thus for example, diphenyldisulphide is a poor anti-seizure but a good anti-wear additive relative to dibenzyldisulphide. This, as Allum and co-workers points out, fits in very well with the Davey and Edwards type model. If anti-wear requirements are the formation of mercaptide layers, this can be represented pictorially as:

\[
R-S-S-R + Fe \rightarrow Fe \cdot \cdot \cdot S_{\text{S-R}} \quad \text{A}
\]

\[
Fe \cdot \cdot \cdot S_{\text{S-R}} \rightarrow Fe_{\cdot \cdot \cdot S_{\text{S-R}}} \quad \text{B}
\]
for dibenzyl disulphide and diphenyl disulphide respectively. (In these diagrams, the arrows represent the movement of electrons). Obviously the mercaptan ion or radical intermediate is stabler in the phenyl case than the benzyl, being \( \alpha \) to the benzene ring, since it is stabilized thus: (ref. 29)

\[
\begin{align*}
\text{Ph} & \quad \quad \quad \quad \text{Ph} \\
\text{S} & \quad \quad \quad \quad \text{S}
\end{align*}
\]

(The ionic case has been drawn, but the reaction may go through a radical intermediate. Stabilization would occur in a similar fashion).

The reverse holds true for the formation of FeS. In this case the C-S bond will be cleaved e.g.

In this case the benzyl cation or radical would be stabilized by delocalisation, with the charge \( \alpha \) to the benzene ring, whereas the phenyl would not. Thus dibenzyl C-S bonds would break more easily. This explanation for the different lubricating effects of various disulphides is reasonable although it has been inferred primarily from the effect of chemical structure on lubrication with little other, supporting evidence.

So far only the mechanisms of sulphur film formation have been considered. It is also necessary to show how such films, once formed, serve to lubricate. There has been quite a lot of controversy in recent years over the exact composition of the anti-seizure film. Since Prutton (12) showed that sulphur was present in the layer, it had generally been assumed that the film consisted mainly of iron sulphides. The majority of film thicknesses mentioned earlier in this review were worked out from the actual sulphur content assuming such a structure. Some detailed analysis by Godfrey (9) in 1962, however, using x-ray and electron diffraction, showed that for some sulphur e.p. additives the main film constituents were oxides of iron.
Godfrey thus suggested that sulphur catalysed the formation of iron oxides and it was a combination of these with just a small quantity of sulphide which gave good anti-seizure properties.

Sakurai et al. (10) carried out similar experiments using X-ray diffraction and showed that whereas diphenyldisulphide gave mostly oxides at 490°C, dibenzyldisulphide produced mainly a sulphide film. The whole situation is thus very confused.

It is obvious that a sulphide layer by itself is insufficient to give good lubrication, as is often observed in the extreme pressure range. Greenhill (30) describes some experiments on the lubricating properties of artificially built up sulphide layers. He found that even when the optimum film thickness of 200nm. was reached, the coefficient of friction was still 0.4. For thinner sulphide films it was much higher. Greenhill also found that if fatty acid in paraffin oil were added to the sulphide film, the coefficient dropped to 0.1, even at temperatures where fatty acid and paraffin oil on the clean metal would not lubricate. Thus the combination palmitic acid and sulphide layer was better than the individual components, lubricating up to 200°C on copper and steel.

Sakurai et al. (31) demonstrated a similar effect. They showed that fatty acid and dibenzyl sulphide molecules were both more strongly adsorbed on sulphide layers than on normal metal oxide surfaces. Thus it seems likely that an extreme pressure film is not a simple sulphide layer. Not only may it contain a complex mixture of oxides, but there is perhaps also an associated physically or chemisorbed monolayer on top of it.

One further analysis of the mechanism of E.P. lubrication needs mentioning. This is a series of papers by Sakurai and co-workers who have developed a rate mechanism for the action of E.P. films in terms of the opposing effects of sulphide film formation and the film being abraded away (32) (5).

3.8. **Summary of E.P. Additive Action.**

The action of sulphur E.P. additives is, as the literature survey shows, extremely complex. Even for pure disulphides neither the mechanism of film formation nor the way this film then lubricates is fully established.

Increasing temperature always increases the film thickness
of the sulphur affected layer. There is a large variation, however, in the actual depth found by different authors.

Both loading and sliding increases the sulphur film depth. It is not clear whether this is due to enhanced reaction at higher frictional temperatures, to the generation of nascent metal, to the formation of cracks down which a chemical reaction can take place or merely to sulphur containing material being physically ground into the surface. A number of studies have shown that a constant film thickness is attained under sliding condition, as opposed to static conditions where film thickness seems to build up indefinitely with time.

There are no detailed studies of low temperature reaction of e.p. additives on steel. Borsoff and Wagner (17) have published the main work in this range but they do not fully describe the starting conditions of the reaction.


A lubricating oil may be quite a melange of chemicals. All are important and are there for a specific purpose. With so many different components great care has to be taken to ensure they they do not interfere with each other. Obvious possible interference effects are straightforward chemical reaction in solution between different additives, either cold or under hot running conditions. Such reactions are quite easy to identify and eliminate by chemical methods. More subtle solution interactions are possible, however, such as the formation of association complexes retaining the simple chemical properties of both constituents but lowering the activity of each.

Much more important are interference effects in specific chemical functions. Almost all boundary additives are designed to act as the metal surface, albeit in different ways and at different times. Since most of these additives are polar surface-active compounds the likelihood of interference as different additives compete for the surface of the metal is very great.

Indications that such interference or "antagonistic" effects do occur have been appearing more and more frequently in recent literature.
Literature Review of Additive Interference.

Almost all references to additive interference found, concerned competitive effects at the metal surface. These are less easy to predict than straightforward chemical interactions in solution and more difficult to circumvent.

The commonest type was the inhibition of extreme pressure activity due to the presence of polar surfactants. Smalheer and Maslin (33) in 1956, described some practical axle tests in which two types of boundary lubrication were observed, one operating at high speeds and the other at high torque. They found that those additives required to mitigate high torque conditions, carboxylic acids and esters, tended to reduce or negate the effectiveness of the sulphur e.p. lubricants added for high speed lubrication, and vice versa.

Okrent (34) in 1961 pointed out that oil-soluble soaps and surfactants having detergent properties lowered the efficiency of anti- wear additives in V-8 engine tests. Similarly Studt (35) showed that long chain acids and alcohols reduced the effectiveness of chlorine containing e.p. agents. Thus pure 1,1,1,3 dodecyltetrachloride in hexadecane solution had a higher failure load than when mixed with either lauryl alcohol or lauryl acid in the same solvent. Studt also noted that the fatty acid had a greater antagonistic effect than the alcohol and suggested that the polar additives reacted with the metal surface to form a protective layer against e.p. attack.

This approach was greatly extended by Grew and Cameron (36) who carried out a series of experiments to show that hexadecylamine interfered with the activity of diphenyldisulphide on steel by absorbing on the metal surface. Grew demonstrated that, using variable load, high speed 4-ball tests, mixtures of 0.3% wt. amine and 1.0% wt amine with 1.0% wt disulphide behaved in scuffing as though there were to disulphide present at all, in failing at quite low loads.

When the ambient temperature of the test was raised from 22°C to 100°C, the 1.0% wt amine + disulphide solution continued to show no e.p. activity. The 0.3% wt amine + disulphide solution, on the other hand, did not scuff within the load range of the test rig, so it was clear that disulphide was acting as an anti-seizure agent.
Grew concluded that normally the applied load raised the surface temperature until the adsorbed, amine film failed. Since the amine layer had prevented disulphide reaction, seizure took place as soon as the amine failure occurred. With 0.3% wt amine solution at 100°C the amine adsorbed film was not close-packed enough to prevent diphenyldisulphide attack. This meant that a sulphide layer could build up unhindered, with increasing temperature, to prevent seizure.

Grew (56) applied the same experimental technique to show that natural base-oil impurities could also lower the effectiveness of anti-wear agents. He compared the extreme pressure activities of diphenyldisulphide in light stock base oil and pure paraffin oil. Using variable load 4-ball tests he showed that the disulphide prevented scuffing more effectively in paraffin oil solution than in light stock, although the light stock on its own was a more efficient lubricant than pure paraffin oil. Grew also used the hot wire technique of Barcroft (19) to demonstrate that acetic acid corroded steel more rapidly in white oil than in light stock, presumably because of natural surfactant inhibitors in the impulse mineral oil.

Rounds (37) has compared the friction of steel using different base oils containing a variety of additives. He showed that the effectiveness of the additives generally increases as the base oil is progressively refined. Rounds also studied the interaction of the friction-modifying additive, chlorinated wax, with a series of oxidation inhibitors. From this work he suggested that base-oil oxidation of a limited nature may be an essential factor in additive response, although too much would be detrimental.

Interference effects at surfaces are not confined to those between an extreme pressure agent and a polar additive. Interference is also possible between a pair of different polar additives or between two e.p. agents.

Clayfield (38) has studied different mineral oils and additive solutions in terms of their spreading properties which, he shows, relate closely to their lubricating characteristics. He demonstrated interference effects between different base-oil mixtures and also showed that interaction between base stock and certain metallic detergent additives could occur, to produce a mixture with low spreading pressure.
Sprague and Cunningham (39) have shown that such interference can be quite selective. They noted that calcium phenate inhibitor cancelled the anti-friction behaviour of a 0.6% sperm oil solution but had only a small effect on a 0.6% saturated carboxylic acid solution.

Interference between different extreme pressure additives has been demonstrated by Sakurai and co-workers, (5), who showed that sulphur e.p. agents were suppressed by the addition of chlorine containing additives, dependent on the surface activity of the sulphur compound used.

These examples taken from the literature have mostly been surface antagonistic effects. Many other possible types of interference are possible, such as chemical reaction in solution or suppression of a required reaction by the quenching of an intermediate species. An interesting recent example of a possible type of interference that an additive formulator would have to consider is given in a recent paper by Ratnasamy and Fripiat (151), who show that thiophene and ethanethiol react chemically to decompose on the surface of MoS₂. This raises the question of whether solid MoS₂, a widely used chemical in lubricants, would adsorb other sulphur-containing additives. Such interference is not put forward as a particularly likely possibility, but merely as an illustration of the numerous difficulties that may arise in trying to prepare a non-interfering additive package.


3. FORD, J.F. "Lubricating Oil Additives - A Chemist's Eye View".


25. DAVEY, W. and EDWARDS, E.D. "The Extreme Pressure Properties of some Sulphides and Disulphides in Mineral Oils, as assessed by the 4-Ball Machine". Wear 1 (1957) p291.


34. OKRENT, E.H. "The Effect of Lubricant Viscosity and Composition on Engine Friction and Bearing Wear. II".

35. STUDT, P. "Zusammenhang zwischen Adsorbierbarkeit und Wirkungkeit von Hochdruckzatzen in Schmierolen" (in German). Erdol u Kohle, Erdgas, Petr. 21 (1968) p784.

36. GREW, W.J.S. and CAMERON A. "The Thermodynamics of Boundary Lubrication and Scuffing".


CHAPTER 4
Metallurgical Aspects of Scuffing
by
A.A. Torrence


When two iron or steel surfaces are rubbed together under dry, or very badly lubricated conditions, they undergo a remarkable transformation. Patches of the surface become greatly hardened, and when treated with ordinary etchants they appear structureless. They are usually left white or yellow by the etchant, and so are generally referred to as white phases, or, when they spread over most of the surface, as white layers. They have been observed in a number of different situations, and in the literature there are almost as many different theories as to their metallurgical structure. No theory seems to have been conclusively proved, which makes it difficult to decide what part the white layer plays in the situations where it occurs.

Several authors (3,10, 15-19, 21) have attempted to explain variations in wear rate by means of mechanisms involving the white layer, but such explanations have always been bedevilled by the uncertainty as to the white layer’s composition. This has led to a situation where some authors (3,10, 15-19) associate the formation of a white layer with protection of the rubbing surface and a reduction in wear rate, whereas others (2,21) believe that hard particles can break off the white layer and themselves promote scuffing and high wear.

This substance was first noticed by Atkins in 1927 (1). He was investigating causes of failure in cold drawn steel wires, and he observed that in some cases the wires were covered with a layer of material unetchable in nital (2% HNO₃ in C₂H₅OH). The layer was about 0.001” thick, and Atkins concluded that it was martensite formed by intensive surface heating of the steel wire when drawn through the die, followed by quenching by the bulk metal. This is, however, not really justified by Atkins’ observations since Nitrides or Carbides would have etched in the same way, and either might have been formed by reaction with the atmosphere or the lubricant.

A different theory was advanced by Schottky and Hiltenkamp
(2) to explain the formation of similar material which they found at places where two leaves of a leaf spring had rubbed together, and also on a broken turbine shaft. Apart from examining the material metallographically, they chemically analysed thin slices cut from it and conducted scratch hardness tests on it. Analysis showed that the Nitrogen content of the surface was as high as 0.04%, whereas the original material contained only about 0.006% Nitrogen. Scratch hardness tests showed that the white phase was harder than the substrate, and they concluded that this hardening was due partly to absorption of Nitrogen and partly to intense cold working. They also found extensive cracking in the surface layer and deduced that it must be brittle too. Cracks formed in this region had extended and led to flakes of metal being detached from the leaf spring; but in the turbine shaft they had caused complete fracture.

Mailänder and Dies (3) came to the same conclusion as to the nature of the white layer after conducting wear tests with a pin on ring machine. They used soft pins running on hard chromium rings, and by running tests unlubricated at a variety of loads, they established that at a certain load, wear rate fell dramatically, and a hardened skin was formed on the pin. Naturally they, unlike Schottky and Hiltenkamp, supposed that this hardened surface was beneficial. But they also found that a hard layer was formed in a Hydrogen atmosphere with a corresponding fall in wear rate, and in a reduced pressure atmosphere, a low wear rate was experienced throughout. So it is doubtful whether the absorption of Nitrogen is as important as they, and Schottky and Hiltenkamp considered it to be.

Yet another situation was investigated by Trent (4). During the early thirties it was realised (5) that the rate of breakage of the wire ropes used in colliery winding gear was excessive. Trent investigated failed wires from the ropes metallographically, and found that a large number of them had formed a non etching layer on their surfaces. This layer could be as much as 0.003" thick, and usually had a number of cracks in it. Sometimes the cracks would project downwards into the substrate, so he deduced that the layer was brittle and that cracks formed in it could be propagated through the wire causing it to fail. From the metallographic evidence he had amassed, he suggested that the layer
was martensite, but in the discussions which followed his paper it was pointed out that his evidence was inconclusive. He did not consider other explanations of the white layer, such as that outlined above, nor did he measure its hardness in any way.

Some confirmation for Trent's ideas came from Wrazej in 1945 (6). He investigated the driving bands of shells, and found that after they had been fired, a hard layer was formed on them. When he prepared micrographs of this layer he found that it was mainly white etching, but that here and there a patch of the original ferrite remained. From this he deduced that the $Ac_1$ temperature had been exceeded in the surface, but not the $Ac_3$ temperature, during the firing of the shell. So most of the original ferrite had been transformed to austenite, and when the shell emerged, was quenched to martensite. The remainder of the ferrite remained untransformed throughout. This is more convincing than the evidence of Trent, but still does not consider whether gases in the gun barrel can be absorbed, and play a part in the hardening of the surface.

Shortly after this Sakmann (7) reported the results of some tests he had done with a disc machine. He ran chromium steel (SAE 52100) against cast iron, and found that in some cases ridges formed on the surface of the discs. He examined these ridges metallographically, and observed that they had the characteristic white layer structure. Scratch hardness tests revealed them to be harder than the substrate, and tempering at various different temperatures caused the ridges to soften, and at $1400^\circ F$ they had reverted to the original steel structure. Sakmann concluded that the ridges were of a phase rich in carbides which had been formed by reaction of the steel with the carbon of the oil. This was really not backed up by any direct evidence. He had made no attempt to measure carbide content at any time by the usual intercept methods, and the fact that the original structure could be reproduced by tempering indicates that not much carbide formation could have taken place, if indeed any at all.

Clayton and Jenkins (8), working with a modified 4-ball tester, also found white phases in lubricated conditions. The test machine was modified to take 3 piston ring segments in place of the three fixed balls so that the upper ball rubbed on the bearing surfaces of the piston ring segments. Tests were run at constant
speed and various different loads, with an EN31 B.S. 960 (1% C 1% Cr) steel ball, and with cast iron segments. With the machine scuffed, white phases were observed on both ball and segments. Occasionally they were present on the cast iron when the machine was running at high loads, but had not scuffed. They were also found on scuffed EN31 balls from an ordinary 4 ball machine. The tempering behaviour of layers on EN31 was similar to that of α iron, but the layers on cast iron behaved differently. They showed no appreciable hardness loss below 500°C, but after 2 hours at 650°C they seemed to consist of a great deal of Fe₃C with some very heavily tempered martensite interspersed amongst it. Clayton and Jenkins advanced no definite theory to explain the formation of the white layer, but suggested that carburization by the oil might play a part.

A further contribution to the facts was made by Barwell (9). He observed at a scuffed surface, regions which etched to give a series of light and dark striations parallel to the surface and reaching a depth of about 0.001" below it. The hardness of these regions varied from 350Hv in the dark regions to 850Hv in the light regions. It was as though successive layers of material had been transferred to the surface hot, and then been quenched to give soft tempered structures in the dark striations, and hard quenched structures in the light ones. This of course implies that the hardness of the white layer is due to the martensite transformation.

In 1957 Welsh (10) made the first attempt to draw together all previous work on the subject and conducted systematic experiments into the nature and effects of these hard surface phases. He first ran tests with steel on a pin on ring machine dry at 1 kg load and 200 cm/sec speed. At low (0.12%) carbon contents the wear rate of the system was always high, but when the carbon content of the steel was increased to 0.34%, the wear rate decreased as time progressed to about 25mg/hr. As the carbon content was increased further, the time taken for a low wear rate to be established fell, until at 0.78% it was established after less than a minute's running. When specimens showing a low wear rate were examined metallographically, they were found to have a continuous hard skin over their surface, but specimens having a high wear rate had only hard patches on their surface. Welsh believed that the hard layer was giving the
surface protection in the same way that case hardening does.

He thought it was being formed by temperature flashes heating surface asperities to form austenite, and that these were quenched by the substrate to form martensite. This explained why steels of lower carbon content formed layers less easily. As the steel's carbon content rises, the ease with which it forms austenite also rises, so more is available to be quenched to martensite, giving a more continuous surface layer.

When similar steels were rubbed in lubricated conditions in a disc machine, they too would form white layers. Below 3 kg/cm contact load no wear was apparent, but above this level the discs would scuff for a time then revert to smooth running and would continue to do so up to 250 kg/cm contact load. After the discs had undergone transient scuffing the white layer had been formed, which still more convinced Welsh of its protective properties.

The only thing he could not explain with his theory of martensite formation was the very high hardness generated. This could be as much as 920 V.P.N., and was often higher than could be obtained by ordinary heat treatments. In some cases the difference was very large, perhaps 300 V.P.N. Welsh decided to simulate the effects of hot spots by sparking the steels he had used under various liquids and gases.

Sparking low carbon steel under paraffin or in air produced material similar in appearance to the white layer and of similar hardness, but sparking under water gave little hardening and a recrystallised structure. When however, the low carbon steel was first nitrided and then sparked under water, it did harden to the same value as that sparked in air. It was concluded that the hard material produced during sparking was a consequence not only of a martensitic transformation, but also of absorption of either carbon from oil, or nitrogen from the air. Because of the similarity between this and the white layer, it was supposed that they were more or less identical, and that the same processes were involved in their formation.

It is, however, difficult to accept Welsh's conclusion when certain other facts are considered. The most important is a thermodynamic one, that steel will only absorb appreciable quantities of nitrogen which is in a metastable condition or at very high pressures. This occurs in a spark, which is known to
dissociate nitrogen into atoms, in which state it can be readily absorbed by steel. In a rubbing contact very high pressures are generated in the metal (11), but the pressure of the gas surrounding it will remain low, so the amount of nitrogen which can be absorbed will be limited by the equilibrium between the gas at ambient pressures and steel. It is interesting to note that Schottky and Hiltenkamp (2) found that nitrogen was in fact absorbed in this sort of quantity, which is too low to account for the degree of hardening observed by Welsh.

Carburization is certainly a thermodynamic possibility, but it could only occur to the depth observed at very high temperatures. For carburization to occur to a depth of 0.001" in the time available (a few seconds), temperatures approaching the melting point would have to be maintained for most of that depth, so that the diffusion rate of carbon in the steel would become great enough to allow it. That high temperatures can be produced at a rubbing contact is well known (12), but it is doubtful that such temperatures could be maintained to such a great depth for any length of time (13). Most authors quote values of about $10^{-4}$ sec (e.g. 12, 13) for the duration of a high temperature flash between asperities; but even in accelerated liquid carburizing, 15 minutes at 900°C are needed to produce a case 0.003" thick (14). Even if carburization can occur, in lubricated sliding the high hardness of white layers produced in dry rubbing can still not be explained, especially as they have been produced both in vacuo by Welsh (10) himself, and in Hydrogen by Mailänder and Dies.

Welsh published a further paper (15) in which he investigated the dry wear of steels in a pin on ring machine over a wide range of load and speed. He found that at very low loads, mild wear occurred, but above a critical load, severe wear started. When the load was increased still more, another transition occurred and mild wear was re-established. At still higher loads, the wear rate of the pin fell sharply and that of the disc rose. Welsh ascribed these last two transitions as being caused by formation of a phase hardened white layer. At the lower load transition, the white layer forms, but becomes softened gradually with prolonged rubbing. Above the higher transition it maintains its hardness. By calculating flash temperatures of asperities, Welsh showed that the surface could become hot enough for phase hardening to occur.
More recently Eyre has published some papers (16,17,18) which follow on from Welsh's work. Using a pin on disc machine to study the dry wear of flake and nodular graphite cast irons (16,19) he found behaviour similar to that reported by Welsh for steels. The main difference was that the transition from mild to severe wear took place at much higher loads. At lower loads a continuous white layer was formed on the surface of the pin. Both pin and disc were covered with an oxide layer which was reddish at low loads, and became black as the load was increased. When the load reached a critical value, severe wear set in and was about 10^3 as fast as the mild wear. Under these conditions the wear track was of a plucked, metallic appearance, and white layer was hardly evident at all. Electron probe microanalysis (17) of the layer in mild wear indicated that it was an Fe/Si/C complex, and its tempering characteristics were consistent with this.

This was reiterated in another paper (18), and in addition it was reported that cast iron wear debris had been X-ray analysed. It had been found to contain only α-Fe, γ-Fe and Fe2C. White layers were also found in a steel during mild wear, but when these were tempered they softened at a lower temperature than those on cast iron. It was thought that they were probably martensite containing a fine dispersion of oxide or carbide.

Some further experiments (19) using plain carbon and chromium alloy steels indicated that not only is a white layer necessary for mild dry wear, but also a protective oxide layer. In all dry wear experiments with steels, it had been found that mild wear did not start at the beginning of a test, but after a certain distance had been slid. Chromium steels required a much longer sliding distance for the establishment of mild wear than did plain carbon steels. This was shown to be due to the difficulty with which the former steels oxidise at high temperatures, so that for mild wear an oxide layer supported by a hard substrate is necessary.

Since Eyre published his work, Kawamoto and Okabayashi have published an interesting paper (20) on investigations of the white layers formed on cast iron in a rod on ring apparatus. Tests were run in the air, and in vacuo and micro hardness measurements made at various depths below the surface. Where white layers were present, they were found to have a high hardness (1000Hv) which fell sharply at their boundary with the substrate to 400Hv. The
white layers were generally thicker in vacuo than in air, but gave similar X-ray diffraction patterns. These indicated that the white layer was a mixture of Fe₃C, γ Fe, Ferrite and Martensite in both cases, and wear debris gave similar patterns to it, indicating that the debris was formed mostly from the white layer.

When the white layer was tempered for 1 hr at 900°C, and etched, it had the appearance of very fine ledeburite, but too fine to be unequivocally identified as such. However, the X-ray diffraction patterns were consistent with this interpretation, so the authors assumed that the layer was in fact ledeburite. Now this can only be formed from a grey cast iron by melting it, then quenching it, so when lower wear rates were observed at higher coverages of white layer, they were attributed to local surface melting. This, it was argued, would absorb energy which otherwise would have gone to break off a piece of wear debris.

Some more work was done on cast iron by Rogers (21), but unlike most other investigators he looked into lubricated rubbing. By taking micrographs of a scuffed pin from a pin on disc machine and comparing them with some from a scuffed cast iron cylinder liner, he showed that the white layers from them appeared identical. They both tempered in the same way, softening from 630 Hv at room temperature to 300 Hv at 700°C. X-ray diffraction of the wear debris showed it to consist almost wholly of cementite, but debris from a dry wear test contained much Ferrite and Austenite as well. In this case, the white layer also looked different under a microscope, so it was thought that scuffing was a different wear process to dry wear, though the wear rates were very similar.

On this basis Rogers advanced a theory of scuffing, which he says occurs by the spalling of a white layer followed by abrasion by the hard fragments formed. Scuffing will only occur when sufficient carbon from the oil has reacted with the iron to form cementite.

Unfortunately this theory is open to a number of objections. One is that the white layers which Rogers found during scuffing have the same appearance as those found by Eyre during dry wear. Another is that though the wear debris may be Fe₃C, the white layer may be something else, which becomes carburized when detached as hot debris. A third is that there is no consideration given in Rogers' work as to whether the white layer is a consequence of scuffing, rather than
Scuffing being its consequence. Both interpretations can be placed upon it with equal ease, so at best it must be regarded as an unproved hypothesis.

Scuffing is, in fact, usually thought to be due to the removal of all protective lubricant films from a metal surface. This allows them to weld together, and gives rise to the definition that "scuffing is gross damage characterised by the formation of local welds between sliding surfaces" (22). Tabor showed (23) that fatty acids reacted with metals to form soaps which would protect the metals from seizure until they melted, and this was confirmed by his later work (24) (25). Frewing (26) (27) investigated a wide variety of additives in white oil, with a Bowden Leben machine and showed that the lubricant failed when the additive desorbed from the metal surface. This has been confirmed by a number of authors, amongst them Askwith, Cameron and Crouch (28), Cameron and Grew (29), Cameron and Crouch (30), and seems quite well established. All these authors predict that scuffing will occur at around 200°C contact temperature, as calculated by Blok's formula (31) for steels and mineral oils sliding together, and others (32) (33) have shown that this concept of a critical temperature does in fact work in a number of practical situations. The values of critical temperature determined by Blok's formula agree quite well with those from the Bowden Leben machine.

It seems, therefore, that carburization cannot be a primary cause of scuffing since desorption of boundary lubricants takes place at a very much lower temperature (200°C) than that needed for appreciable carburization (1100°C). Yet the carburization theory cannot be completely ruled out by this reasoning, as it is possible for very high temperature flashes to occur within a boundary lubricated contact of fairly low ambient temperature. This was shown by Bowden and Ridler (12) who used a constantan slider on a steel disc as the hot junction of a thermocouple, and were able to deduce the temperature at the surface from its e.m.f. However, as Jaegar pointed out (13), these temperatures are highly localised, and for mild steel a typical asperity 1 x 10^-4 cms in length will have a surface temperature which is 10 times that at a depth of 1.5 x 10^-4 cms, so even with these considerations Rogers' hypothesis seems unlikely, as temperatures needed for carburization will not occur to any depth.
4.2. **Conclusions.**

Running through all the above work is a vein of uncertainty as to the nature of the white layer. In fact Kawamoto and Okabayashi (20) have found five distinct theories of frictionally hardened layers in the literature. It has been suggested that they are strain hardened layers (34) **martensitic layers** (10), nitrided layers (2,3), carburized layers (7), or some kind of amorphous material (35). This uncertainty makes it difficult to predict its behaviour, and while it has been suggested to be beneficial in dry sliding (10) (15) (18), it has been thought both beneficial (10) and (21) deleterious in lubricated sliding.

The most powerful technique for investigating metallurgical structures is transmission electron microscopy, and this had not been applied to the problem of the white layer. The difficulties of using such a method to investigate such a small quantity of material are formidable, but it is possible to overcome them and to obtain some useful information, unobtainable by other methods.
REFERENCES


22. Memorandum on Definitions, Symbols and Units, Institution of Mechanical Engineers.


## DOCUMENT CONTROL SHEET
(Notes on completion overleaf)

Overall security classification of sheet .................................. **UNLIMITED** .................................................................

(As far as possible this sheet should contain only unclassified information. If is is necessary to enter classified information, the box concerned must be marked to indicate the classification eg (K), (C) or (S)).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BR 33386</td>
<td></td>
<td>S&amp;TMemo-5-73</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5. Originator's Code (if known)</th>
<th>6. Originator (Corporate Author) Name and Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>626700</td>
<td>London University Imperial College of Science &amp; Technology, UK</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5a. Sponsoring Agency's Code (if known)</th>
<th>6a. Sponsoring Agency (Contract Authority) Name and Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>824046</td>
<td>Procurement Executive, Min. of Def., UK</td>
</tr>
</tbody>
</table>

### 7. Title

**BACKGROUND TO CURRENT THEORIES OF SCUFFING**

### 7a. Title in Foreign Language (in the case of translations)

### 7b. Presented at (for conference papers). Title, place and data of conference

### 8. Author 1, Surname, initials

MacPherson, P.E.

### 9a. Author 2

### 9b. Authors 3, 4...

### 10. Date

6.1972

### 11. Contract Number

### 12. Period

### 13. Project

### 14. Other References

### 15. Distribution statement

Distribution controlled by MOD Technical Policy Authority via DRIC

### Descriptors (or keywords)

Scuffing, Scuff resistance, Surface properties, Lubricants, Lubricant additives, Reviews.

**Abstract**  The Lubrication Laboratory at Imperial College is undertaking a series of investigations into scuffing phenomena on behalf of Westland Helicopters Ltd. The background to the work to be undertaken is set out in the present report in the form of a literature review. The review is divided under four headings: - Elastohydrodynamic lubrication in line contact; Lubrication breakdown under boundary conditions; Action of oil additives; Metallurgical aspects of scuffing.
NOTES ON COMPLETION OF DOCUMENT CONTROL SHEET

This Document Process Sheet is designed specifically for MOD reports and reports produced by Contractors.

Boxes marked * below need be completed only if the information is readily available.

1. DRIC reference: Enter DRIC reference (BR number) if one has been assigned.

2. Originator's Reference: Enter the report number by which the document is identified by the originator of the report, in the form in which it appears on the cover.


4. Report Security Classification: Enter security classification or marking which limits the circulation of the report, or enter UNLIMITED when this applies.

5. Originator's Code: Code number for the DRIC-standardised form of the entry appearing in Box 6.

5a. Sponsoring Agency's Code: Code number for the DRIC-standardised form of the entry appearing in Box 6a.

6. Originator (corporate author): Enter name and location of the organisation preparing the report.

6a. Sponsoring Agency (Contract Authority): Enter the name of the monitoring MOD Branch or Establishment in the case of Contractor's reports. If an MOD report covers work funded by an outside agency, enter the name of that agency.

7. Title: Enter the complete report title in capital letters but omitting initial definite or indefinite articles. If the report covers a specific period, enter this after the title, eg (1.1.1972-31.3.1972).

7a. Title in Foreign Language: In the case of translations, enter the foreign language title (transliterated if necessary) and the translated English title in Box 7.

7b. Conference Papers: If 7 is the title of a paper presented at a Conference, or a Conference proceedings, enter the Conference Title, where it was held and the date.

8. Author 1: Enter the name of the first author, followed by his initials.

9a. Author 2: Enter the name of the second author, followed by his initials.

9b. Authors 3,4,...: Enter third and following authors' names.

10. Date: Enter the month (in figures) and the year of the report (Dec., 1969 is written 12.1969). If the report is undated but a period covered by the report is indicated, enter the date at the end of the period, eg 12.1969. Enter the inclusive number of pages in the report containing information, i.e. including appendices, tables and illustrations, and the total number of references cited.

11. Contract Number: Enter the number of the contract or grant under which the report was written.

12. Period: (always associated with the Contract Number). Only to be used for reports covering a specific period, eg quarterly, annual or final reports. Enter OR-1, AR, FR, as appropriate.

13. Project: Enter project name or number.

14. Other Reference: Enter any reference, other than those in Boxes 2 or 3, by which the report may be identified.

15. Distribution statement. Enter any limitations on the distribution of the document. If distribution is limited to particular groups eg MOD, MOD and its Contractors, etc. it should be stated. If the distribution is the responsibility of another authority eg a HQ Directorate (Technical Policy Authority) enter "Distribution controlled by MOD Technical Policy Authority". Enter "via DRIC" after "Authority" where appropriate and name the Technical Policy Authority if possible.

Descriptors: Any number of descriptors (or key-words) can be entered. If selected from a published thesaurus, eg The Thesaurus of Engineering and Scientific Terms (TEST), this should be indicated.

Abstract: The abstract should preferably not exceed 150 words, i.e. it can be considerably shorter than the Abstract to be provided on the Title Page of the Report. Information available in the report title need not be included in the abstract.