PITTING FAILURE OF GEARS

by

R. A. Onions

Thesis submitted to the University of Leicester

for the degree of Doctor of Philosophy

February 1973

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

DECLARATION

This thesis describes research carried out by the author whilst registered as a full time student at the University of Leicester. The contents are original unless stated otherwise or acknowledged and have not been submitted for a further degree in this, or any other, University.

R. A. Onions.

ABSTRACT

Failure due to pitting fatigue has been investigated under controlled laboratory conditions. The investigations used both a realistic laboratory test rig using $\frac{1}{2}$" face width gears and the geometrically simpler simulation of gears using a disc machine. The results obtained substantiated earlier work of Way and Dawson. The initiation and propagation mechanisms are generally considered to hold true. However, the gear tests showed that failure could occur much more readily than with discs and therefore the application of disc tests to gears must be viewed with caution. The results suggest a fundamental difference between the pitting behaviour of gears and discs.

The second part of the thesis is of a more theoretical nature. A theory of surface contact was developed along the lines of that by Greenwood and Williamson using a surface model developed by Whitehouse and Archard. These results show that a distribution of asperity curvatures increases the probability of plastic deformation. The plasticity index has been redefined in terms of a convenient two parameter defirition of surface topography. The theory has been applied to results obtained from a typical ground surface of hardened steel; when the anisotropy, which is part of such surfaces, is taken into account it is shown that only a small proportion of the contacting asperities are plastically deformed. The limitation of this form of model is discussed and a second approach is put forward using digital techniques. Theory has been developed to enable the contact of surface profiles to be simulated in a computer and the interference areas so formed have been related to the real Hertzian deformed areas of two rough surfaces. The approach is equally applicable to run-in surfaces which are not represented by existing models.

The implications of this work for future research are discussed; the need for a fuller understanding of partial and micro elastohydrodynamic lubrication by theory and experiment is stressed.

PREFACE

The work reported in this thesis was carried out partly in the Tribology Laboratory of the Department of Engineering of the University of Leicester and partly in the Industrial Lubricants Division of Thornton Research Centre.

The author is indebted to Shell International Petroleum Company for financial support throughout the project and also to Shell Research who made available equipment and financial aid for the experimental investigations of Chapter 3.

I also wish to thank Mr. S. Taylor for carrying out some of the experiments of Chapter 2 and Mrs. H. Sheppard and the Drawing Office staff for preparing the typescript and drawings for this thesis.

Finally the author would like to acknowledge the supervision given by Dr. J. F. Archard, to whom I give my sincere thanks for invaluable encouragement and helpful advice throughout the project.

CONTENTS

Symbols used in experimental and theoretical work

General

D $= \sigma/h_o$ ratio of surface roughness to oil film thickness.

E_1, E_2 Young's modulii of materials of surfaces 1 and 2.

E' Reduced modulus given by

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

G $= \alpha E'$

H_{min} $= h_o/R$

h_o Minimum lubricant film thickness

p_o Maximum Hertzian pressure

R Effective relative radius of contacting bodies given by

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2}$$

R_1, R_2 Radii of contacting bodies 1 and 2

U $= \dfrac{\eta_o \bar{u}}{E' R}$

$\bar{u}$ $= \frac{1}{2}(u_1 + u_2)$ line contact

u_1, u_2 Surface velocities of bodies 1 and 2

W $\equiv \dfrac{w}{E' R}$ line contact

w Load per unit length of face width

α Pressure exponent of viscosity $\left(\eta = \eta_o \exp(\alpha p) \right)$

η Viscosity

η_o Controlling viscosity

ν_1, ν_2 Poisson's ratios of materials of bodies 1 and 2.

Additional notation used in Chapter 4 with specific values where appropriate.

a	Radius of area of Hertzian contact
A	Total true area of contact
δA	Area of Hertzian contact for one asperity
$\mathcal{A}$	Apparent area of contact ($1 cm^2$)
C	Dimensionless asperity curvature
$\bar{C}^*$	Mean dimensionless curvature for all peaks
d	Normalized separation
E	Young's modulus (21×10^3 kg/mm^2)
E'	$= E/(1-\nu^2)$
g	Specific conductance ($8.35 \times 10^5\ \Omega^{-1}\ cm^{-1}$)
G	Total conductance
δG	Conductance of one asperity
H	Hardness
ℓ	Sampling interval (taken as $2.3\beta^*$ in the theory)
n	Number of contacts
N	Ratio of peaks to ordinates
$\bar{p}$	Mean real pressure ($=W/A$)
R	Radius of curvature of asperity
W	Total load
δW	Load borne by one asperity
y	Normalized co-ordinate (height/σ)
β^*	Correlation distance ($2.3\beta^* = 15\mu m$)
η	Density of asperities per unit area
ν	Poisson's ratio (0.3)
σ	Standard deviation of the height distribution ($0.5\mu m$)
ρ	Correlation between successive events

ψ $= \left(\frac{E'}{H}\right)\left(\frac{\sigma^*}{R}\right)^{\frac{1}{2}}$; Plasticity index (G & W)

ψ^* $= \left(\frac{E'}{H}\right)\left(\frac{\sigma}{\beta^*}\right)$; Plasticity index (W & A)

ω Compliance

ω_p Compliance corresponding to onset of plastic flow

Additional notation used in Chapter 5.

A One half the minor principle curvature of an asperity

A_{total} Total true area of contact

a Radius or major semi-axis of Hertzian contact

B One half the major principle curvature of an asperity

b Minor semi-axis of ellipse of Hertzian contact

D Half width of Hertzian contact zone (line contact)

$E(k')$ Complete elliptic integral of the second kind

$f(A/B)$ Function of A/B

H Hardness

$K(k')$ Complete elliptic integral of the first kind.

k Axial ratio of ellipse of Hertzian contact ($k \leq 1$)

$k'^2 = (1-k^2)$

L Half length of line contact for discs

l Length of one Talysurf profile

m Number of pairs of Talysurf profiles

N Number of profile traces per unit length

p_{mean} Mean Hertzian pressure

R Local relative radius of curvature for two contacting asperities.

W Load supported by two contacting asperities

W' Load per unit length at right angles to profiles

W_p Load borne by plastic contact

W_{total}	Total load borne by asperities
α	Half width of interference contact (perpendicular to profile for elliptic contacts)
α_{mean}	Mean value of summation of α
α_o	Radius or semi-major axis of ellipse of interference area
β	Half width of elliptic interference contact along the line of the profile.
β_o	Semi-minor axis of ellipse of interference
σ	Standard deviation of height distribution
ω	Depth of interference
ω_{mean}	Mean value of ω for an individual contact
ω_o	Compliance or maximum depth of interference
ω_p	Interference depths which would result in plastic deformation.

Symbols used only once are defined at the appropriate point of the text.

CHAPTER 1

1.1. General Introduction

In recent years one of the most striking developments in Tribology has been the emergence of the subject of elastohydrodynamic lubrication (e.h.l). In broad terms this area of Tribology has now reached the point where theory and experiment combine to provide a generally accepted body of knowledge which covers most of the range of conditions experienced in practice. However this knowledge has been concerned mainly with the lubrication of perfectly smooth surfaces, or of surfaces which, for all practical purposes, approach this ideal.

Unfortunately, the surfaces used in engineering practice are not smooth and one might assume that most failures of elastohydrodynamic lubrication arise from this fact. In his pioneer studies of pitting failure Dawson (1961) (1962) (1963) lent considerable support to this concept. He showed that the number of cycles to pitting failure was a function of the non-dimensional parameter

$$D = \frac{\text{combined surface roughness}}{\text{lubricant film thickness}} \qquad (1.1)$$

Of course, this work does not provide the full answers which are required; indeed, in his last paper, Dawson (1968) collected a great deal of microscopical and other related evidence as a contribution to a more detailed and complete account of the mechanism of pitting. Moreoever it should be remembered that Dawson's experiments were performed with a disc machine and there have been many comments in the literature which argue that the relationships between disc tests and gear behaviour are inadequate for a quantitative comparison.

There remains one other aspect of this work which requires exploration. The subject of surface topography, its examination and its

specification have shown marked developments in recent years. A significant part in these changes has been played by the development of digital techniques in surface topography. The output of profilometers can be presented in digital form and much information of interest can be provided by the subsequent analysis of this data by computer methods. The power of these techniques in tribological research has yet to be fully explored and relatively little use has been made of these methods in the study of lubricated systems. Clearly, these developments are relevant to our subject.

The work described in this thesis was planned against the broad background which has been outlined above. It was proposed to carry out most of the work in the Industrial Lubricants Division of Shell Research Ltd. at Thornton Research Centre. The original concept was that the project would commence with a study of pitting in gears using a large gear test rig. From the results of these tests and their comparison with disc machine results it was hoped that the work would develop according to its own logic.

In practice, for reasons outside the control of the author or his sponsors, the project has not taken this planned form. The test gears did not become available until two years after the commencement of the work. Therefore much of the work described below has been based upon a broad assessment of the literature rather than upon the closely knit development which had been planned. Whatever the faults of the author, some lack of coherence of this thesis must therefore be attributed to this background within which the work was performed.

Chapter 2 includes a general literature survey and a limited amount of work carried out on a disc machine, the results of which were intended to form a basis for comparison with those obtained from gears. The tests involved were carried out in order to substantiate the results published by Dawson, which were obtained with a different combination of materials

and the results could then be used in conjunction with those of Chapter 3 to enable comparison of disc and gear results. Chapter 3 describes the gear tests performed on a 5" centres test rig and provides a comprehensive set of results incorporating variations in surface finish, speed of rotation, and two oils of differing viscosity. The broad conclusion of this work is that Dawson's findings, based on disc experiments, are for the points investigated applicable to gears. The work suggests that the physical arguments of pitting failure are similar for both cases but that results obtained from discs can grossly overestimate the lives of gears. The work implies some fundamental difference between the two forms of testing and the conclusions suggest that the differences quite possibly stem from the differing kinematics of the two systems. When taken in conjunction with recently published information on in-service behaviour this work fills a gap in a broad spectrum of knowledge which successfully links engineering science with engineering practice and design.

As discussed above the next requirement is for more detailed knowledge. Chapter 4 therefore develops some recent work of Whitehouse and Archard (1970) on surface topography into a theory of surface contact. Such theories take the separation of the surfaces as the independent variable and the load, area, number of contacts and the electrical conductance as the dependent variables. Therefore this theory is applicable to both dry and lubricated contact.

The theory of Chapter 4 is markedly idealized. It is based upon the topography of freshly ground surfaces. Methods or theories of wider applicability are clearly desirable and, in particular, it would be of value to be able to deal with run-in surfaces. Chapter 5 therefore proposes a scheme based upon the direct use of profiles of surfaces, in digital form, and their use in the analysis of surface contact by

computer methods. Although this technique has not been fully developed it is described here as a suggestion for future developments.

Finally Chapter 6 surveys the work described in the thesis against its background, and indicates some areas where further work is required.

CHAPTER 2

EXPERIMENTAL INVESTIGATION OF PITTING FAILURE USING SPECIMENS OF SIMPLE FORM

2.1. Background

For the introduction of this chapter it is appropriate to digress for the moment to the general field of pitting phenomena. Much work has been carried out on this subject and reliable theories are still very elusive. It is generally accepted that pitting is a fatigue process; therefore the greatest concern is the amount of data required for analysis, at present much of the data already accessible may only be applicable to the specific conditions under which the tests were run.

There are generally two fields of investigation

a) That of rolling element bearing failure using hard materials, and

b) Gearing failure using relatively softer materials.

Investigations at the present can be broken into three sub-groups -

(i) Metallurgical - including heat treatment, material combination and grain structure.

(ii) Stress analysis - by running conditions or form of geometry.

(iii) Lubrication - involving film thickness, viscosity pressure coefficients and chemical action.

In group (i) Chesters (1958) using a disc machine has studied the failure of various material combinations. The driving disc was manufactured from various test materials and run against a case hardened steel disc. His results provided a family of S-N fatigue curves each requiring an increased number of stress cycles as the tensile strength of the test materials increased. In this set of results it was suggested that pitting first occurred when the maximum Hertzian surface pressure

was approximately equal to the tensile strength of the test material. Chesters (1963) amended this by concluding that the onset of pitting was likely to occur in the range 0.67 to 1.2 times the tensile strength depending on the material combination. This work also showed that plastic flow and strain ageing of the surface layers occurred due to the loading required to obtain pitting. Finally, throughout the combinations of materials used in these experiments, the type of heat treatment had a marked effect on the life of the specimens. Scott and Blackwell (1966) also carried out tests on material combination but in this case they used a rolling four ball lubricant tester and their results are therefore more applicable to bearings. A further result from their paper which would seem promising was the work carried out on the hardness of materials, where they found that for an En 31 ball combination there was an optimum hardness range to give a maximum fatigue life. From this the authors suggest that to reduce stress and plastic flow the largest possible contact area so obtained should be linked with the permissible amount of work hardening so as to avoid the increase in pitting which would occur. They conclude that the maximum fatigue life is obtained when the upper and lower balls are of similar hardness (i.e. race and balls in the practical case), or the rolling balls 10% harder and all should be in an optimum hardness range; this they point out is in agreement with the work of Zaretsky et al (two refs.1965). Another metallurgical aspect again applicable to bearings was fibre orientation. Scott (1963) has shown that this has a marked influence on the pitting fatigue life of balls, the results were not conclusive owing to the large amount of scatter obtained (see his Table 10.5, p.107) but broadly speaking it was shown that pitting was more likely to occur on or around the area of end grain. This view is also held by Bamberger (1969) and a diagram of failure density for fibre orientation is presented in his Fig.7 p.422. This problem of fibre orientation is also shown to be important in

bearing races, with flow parallel to the race being beneficial. While the materials are not those used in gearing the results are nevertheless relevant. With increased cold and hot rolling of gears, Ridgway (1970) shows that the amount of grain flow in the root of a rolled gear is considerable and that by such a production process the fatigue strength of the gears is increased.

Although comparisons have been made between gears and bearings there is obvious evidence that considerable differences do occur between the two types of testing. Tallian (1967/68) provides a comprehensive analysis of rolling contact failure in bearings, and includes details of the differences in the types of failure. There is also emphasis placed on the differences in the failure of balls compared with gears when the plastic ranges of the two are very different. The shakedown criterion of Merwin and Johnson (1963) for soft materials allows large cumulative plastic flow parallel to the rolling direction, this flow is apparently not observed in high hardness materials. Generally the characteristic shape of the pits found on gears and discs does not occur on balls where the pits may have a crater form, sharp edged and more or less flat bottomed. Because of the differences in ball and roller bearing results much of the published literature concerned with this situation has been left out of this discussion and only results which are relevant to both soft and hard materials are referred to.

The second group of investigations, on stress, has mainly been studied in a theoretical nature. Dowson, Higginson and Whitaker (1963) have analysed hydrodynamic lubrication in nominal rolling line contact. They used pressure distributions to calculate the stress distribution in the solid elements and stress contours were produced for varying values of U. An analysis of plastic deformation for an idealized elastic/plastic material has been carried out by Merwin and Johnson (1963) which predicts qualitatively the build up of residual stresses. They show

that the plastic flow is initiated at the position of the maximum unidirectional shear stress τ_{max}. The flow is initially at 45° to the surface and then due to residual stresses the flow either stops or continues until bulk plastic deformation occurs. The authors propose a failure criterion called the shake-down limit which relates the maximum Hertzian pressure p_o to the yield strength of the material k. The theory also predicts the cumulative displacement in the forward direction of motion. Johnson and Jefferies (1963) attempt to correlate life tests of a material in rolling contact with either a calculated or measured value (Hamilton (1963)) of forward flow δ and the layer thickness h. A final paper by Johnson (1963) presents some correlation of theory with experimental results from several sources of rolling contact fatigue. In the analysis he introduces the shake-down criterion (incorporating the influence of a lubricant film) for comparison of stress values at failure and found that many of the endurance limits were close to a value of 2 for p_o/k, which is much lower than the shake-down value of 4. His conclusion was that asperity interaction through the lubricant film, which will be discussed more fully later, could well be the prime cause of this discrepancy.

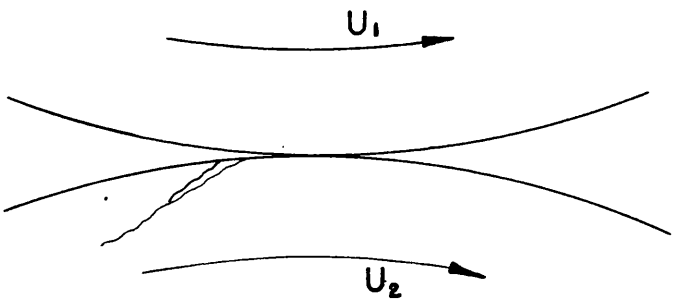
The third sub-group, the topics with which this thesis is mainly concerned, accounts for the influence of lubrication upon the fatigue failure in rolling contact. Of the literature available two authors Way and Dawson have carried out by far the most comprehensive and independent investigations of pitting failure with the softer form of materials.

As far back as 1935 Way investigated the influence of lubrication on pitting failure and found that pitting did not occur when there was no lubricant present, instead an oxide layer appeared on the disc tracks which tended to flake off as rolling continued. Tests run dry which then had lubricant applied failed earlier than those which ran initially

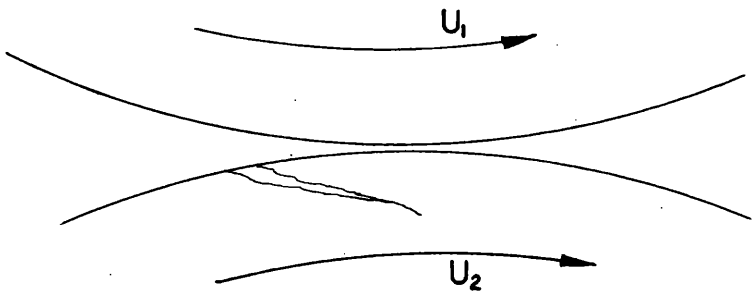
with a lubricant, these tests have also been verified by Dawson (1961). Way reversed the procedure and found that once cracks had appeared with a lubricant they would not progress further if the lubricant was removed. A further suggestion put forward by Way was concerned with the influence of the oil viscosity. It was found that with a very thick oil the pitting phenomenon was prevented, application of theory suggested that there was still a large amount of metallic contact through the oil film; unfortunately he did not have e.h.l. theory to call upon and subsequently Dawson (1962) verified this approach by postulating the D ratio which is defined in equation 2.1.1 below. Way did show experimentally that the film thickness had an effect on the pitting life and also that a rougher disc failed earlier than its smoother partner; Dawson (1962) gave more positive proof of Way's results. In one set of experiments using pairs of discs of equal hardness the load and film thickness were kept constant and the surface roughness varied. A straight line was obtained on a log-log plot of cycles to failure against sum of peak to valley average (P.V.A) surface roughness. A second set of experiments maintaining the surface roughness constant and varying the film thickness gave a very similar result. The combined presentation produced a single straight line and from this the now common D ratio was defined as

$$D = \frac{\text{Total initial surface roughness of the two discs}}{\text{oil film thickness}} \qquad (2.1.1)$$

Way, Dawson and many other authors have remarked on the characteristic shape that pits have in disc and roller tests and again Way (1937) put forward his proposals for this occurrence. He described the craters as fan shaped and showed that they, or the initial cracks, were normally inclined to the surface at apprpximately 15-25°; he suggested that they were initiated in sub-surface regions at a depth of about 0.001 ins. This was necessarily a tentative argument but he attempted his justification by observing small sub-surface cracks and equated the depths of these to



a) Propagation



b) Non propagating

FIG. 2.1.1a,b Method of crack propagation

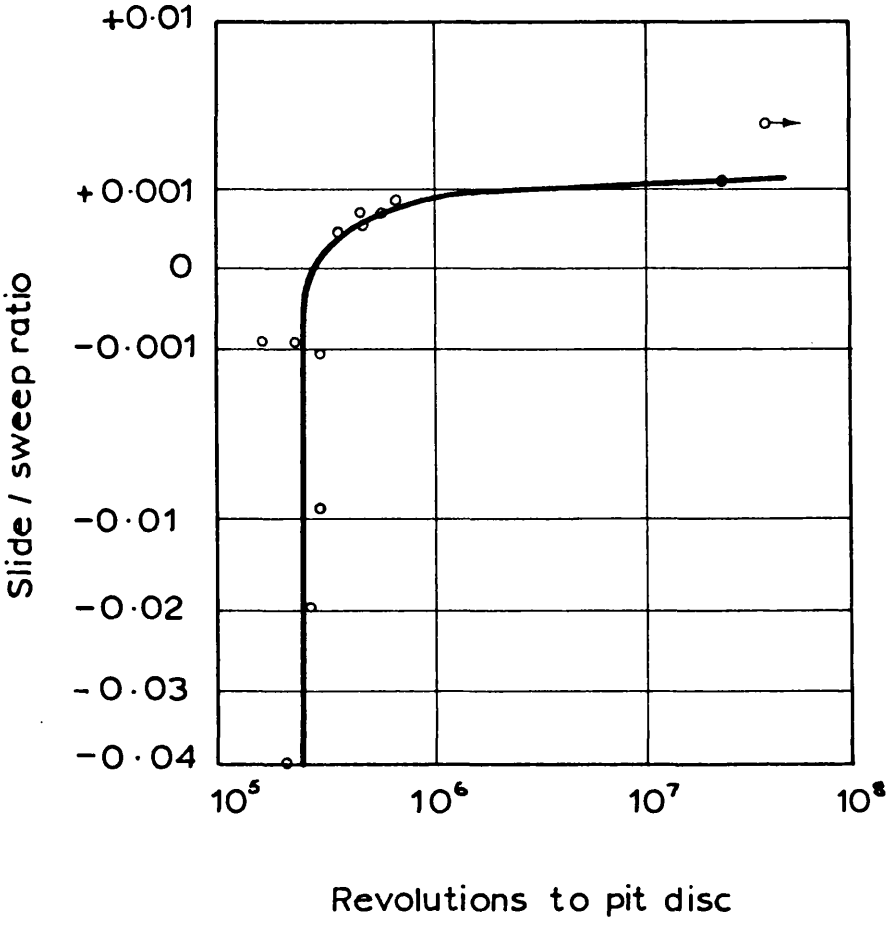


FIG. 2.1.2. Effect of slide sweep ratio on life to pitting.

values of the depth of maximum shearing stress calculated from a proposed surface model. Dawson (1968) corroborates these statements by showing further photo-micrographs of cracks running parallel to the surface and at a depth of 100μ" or less. Again Dawson supplies supporting theory tc suggest that these cracks could quite possibly be initiated by the stress associated with the surface asperities. Having proposed the method of crack initiation Way put forward the method of crack propagation and once again there is strong supporting experimental evidence from many authors indicating that he had the correct hypothesis.

Consider Fig. 2.1.1a in which two discs are rolling together, it is assumed that oil is already in the crack. As the crack approaches the contact area the opening would be closed thus trapping the entrained oil. As the crack progresses through the contact so the oil is forced down with increasing pressure and propagates the crack further into the metal. In Fig. 2.1.1b the condition is not fulfilled, the mouth of the crack does not enter the contact first, and so the oil would be forced out; therefore orientation of cracks is also an important factor in the propagation process. Dawson (1961) gave evidence for this effect when he showed that with positive slide/sweep ratio pitting did not occur Fig. 2.1.2. Further evidence for this is obtained from gears where almost all pitting occurs below the pitch line in the negative slide/ sweep region and only very occasionally do pits occur in the addendum or positive slide/sweep region. Chesters (1963) found that when the test material was the driven disc fewer stress cycles were required to initiate pitting than when it was the driver, this again is consistent with Fig. 2.1.2. Both Way (1940) and Dawson (1961) performed elegant tests to substantiate the propagation mechanism. They ran pairs of discs until arrow head cracks appeared, holes 0.02" diameter were then drilled to the tip of each crack. The cracks ceased to propagate due to the oil escape route now available, which provided ample proof of Way's mechanism.

In a second test Dawson (1961) produced cracks and then reversed the direction of rotation of the discs, again the cracks ceased to propagate due to the oil being forced out; on reverting to the original conditions the cracks again began to grow. Further work by Dawson (1963) produced S-N curves similar to those of Chesters (1958) and Martin and Cameron (1961). The latter authors found that by varying the viscosity of the lubricant they could produce a family of curves such that the greater the viscosity of the lubricant the higher the endurance stress of the specimens. Dawson (1963) reproduced this effect in his results, but also varied the D ratio not only by viscosity but also by varying surface finish. He used two values of D ratio and found that the life to pitting was very much greater for the lower D value. Further work by Dawson (1965) was carried out on the effect of the relative radius of curvature. He obtained results for relative radii of curvature of 1.2 and 10 inches and found that the greater value gave an increased fatigue life, along with this work he also carried out tests to determine the effect of forms of surface generation. He attempted to produce discs of similar finish to gears, but the results were not very comprehensive as only three successful tests were performed and the conclusions formed favoured a decreased pitting strength. Dawson put forward possible explanations of the results. To investigate the effect of machining stresses or asperity orientation he performed two tests, one with the machining marks in the circumferential direction and the other in the axial direction; both finishes were achieved with the use of emery paper, thus the residual stresses should have been the same. The two tests gave similar results which led Dawson to suggest that the machining process was the prime cause, the question still remains to be answered as very few results were obtained. Theory (Johnson, Greenwood and Poon (1972) and Christensen and Tonder (1971)) suggests that for small values of D the pitting life would in fact be increased for axially

finished discs as the oil film thickness would tend to be greater. To substantiate these results P.M.Ku (1972) recently provides some evidence for this being the case. His results show a drop of the order of 25% for coefficient of friction when using axially ground discs, which is consistent with the above mentioned theories. Larger values of D pose a more difficult problem as there is no theory available to predict a reliable value of film thickness when the surface roughness is comparable with, or larger than, the film thickness calculated from conventional e.h.l theory (which assumes perfectly smooth surfaces). Other work on the direction of machining marks has been carried out by Shotter (1958) who measured the frictional torque and electrical resistance between the discs. His conclusions were that polished surfaces gave lower coefficients of friction than rougher surfaces even after they had run in to an equivalent value. This problem is relevant to Chapter 5 where a method of analysis for run-in surfaces is presented. The problem may stem from defining a surface by one parameter R_a whilst proposals are now being made (Spragg and Whitehouse (1970/71)) for a two parameter model which takes into account the wavelength of the surface structure which, for a run-in surface must have increased due to the severe modification of the upper peaks of the asperities.

The final topic to be discussed is the lubricant and its related properties. Martin and Cameron (1961) produced S-N curves for different oils and plotted endurance stress against viscosity. They stated that for oils having a viscosity greater than 20 centi-stokes the endurance stress varied as the fourth root of the viscosity. Their S-N curves emphasised the point made earlier that the results obtained must apply to specific conditions on the curves so produced, otherwise conclusions become meaningless. The second conclusion they arrived at was that the time for pitting to occur was not necessarily solely a function of the viscosity but dependent on some other factor in the oil such as the

surface activity. This effect has been investigated by Galvin and Naylor (1965); they used a Wöhler rotating cantilever rig to perform fatigue tests, having different oils and additives as the surrounding medium. They used Medicinal White Oil as a standard and found that for the remaining oils tested two groups emerged. Group 1 consisted of oils that reduced the life of specimens working above the fatigue limit but had no effect on the lower fatigue limit itself. Group 2 oils reduced the level of the fatigue limit itself, as well as the fatigue life above the limit. The authors admit that the conditions in a bending fatigue test are different from those that exist in rolling contact but nevertheless they have put forward convincing evidence that chemical effects of various fluids can drastically reduce the fatigue limit of reactive metals.

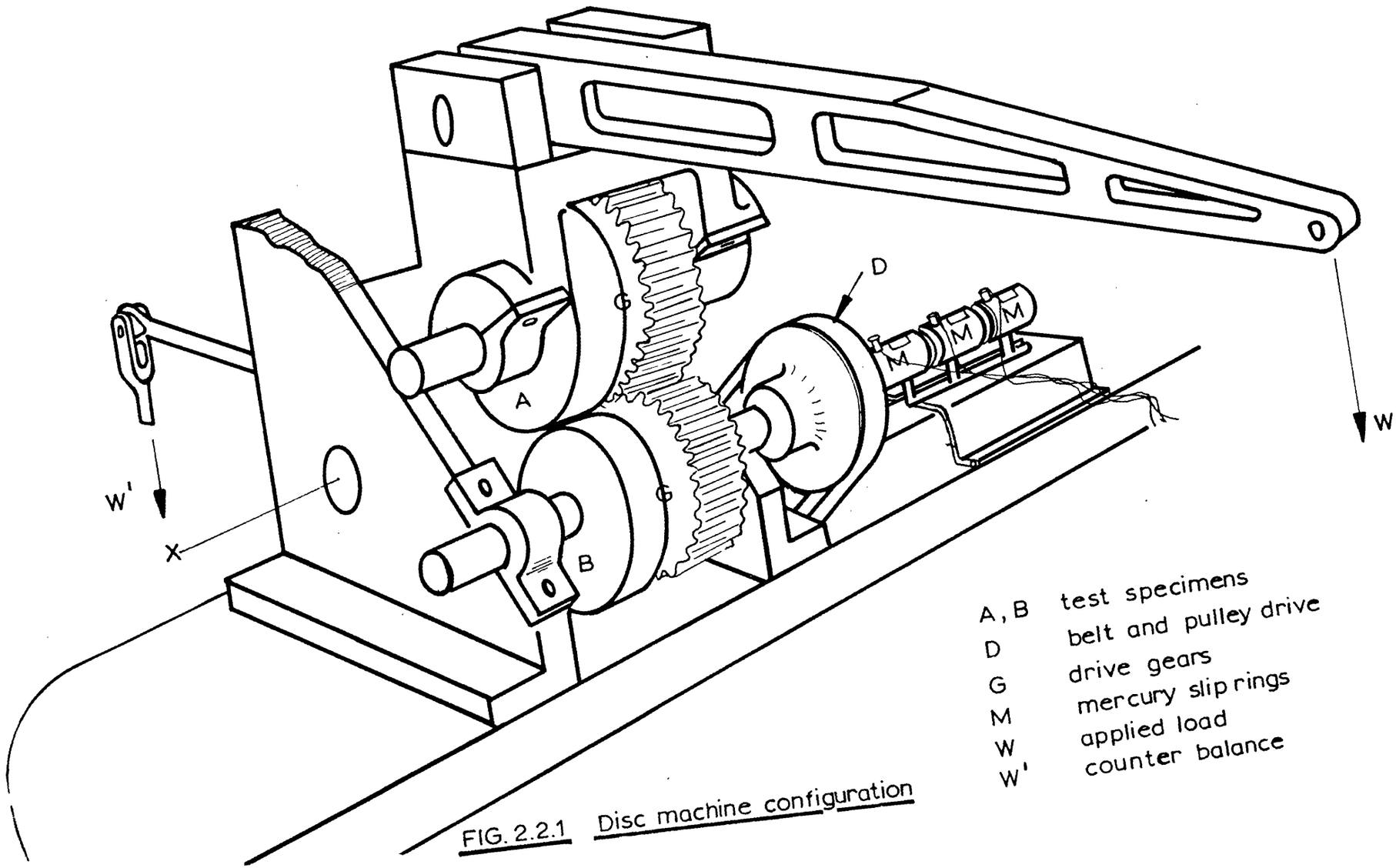
The last factor to be considered in this group is the possibility of predicting the oil film thickness. The results quoted later will use the equation derived by Dowson and Higginson (1961).

$$H_{min} = 1.6\, G^{0.6}\, U^{0.7}\, W^{-0.13} \qquad (2.1.2)$$

where H, W, U and G are dimensionless variables defined in the table of symbols.

Dawson and other workers have shown the validity of the D ratio and likewise in all cases have used the Dowson and Higginson formula or some equivalent. These results must be viewed with some reservation at large D values where the formulae are not, by the nature of their derivation, directly applicable.

This introduction has shown that pitting experiments must be carried out under clearly defined operating conditions as so many variables can influence the results. The survey has also shown that almost all pitting experiments with the softer gear materials have been performed with disc machines. An important part of this thesis is concerned with pitting tests on gears and these tests are discussed in the next chapter. As a



A, B test specimens
D belt and pulley drive
G drive gears
M mercury slip rings
W applied load
W' counter balance

FIG. 2.2.1 Disc machine configuration

preliminary to this work with gears this chapter describes a more limited range of tests with discs of the same material combination which is used in the gear tests.

The purpose of these disc tests are as follows

(a) To provide a direct comparison between pitting tests with discs and with gears.

(b) To provide a means of comparison with earlier work, such as that of Dawson.

(c) To investigate, where possible, some factors, such as the direction of machining, which may be influential in comparison between results obtained with discs and with gears.

2.2. The disc machine

The tests were performed on a Two Disc Pitting Wear machine simulating line contact. The geometric configuration of the discs in the machine is shown in Fig. 2.2.1, and the loading was facilitated by a lever system and spring balance. The lower shaft was driven by a half horse power a.c. motor via two pulleys and variation of speed was achieved by different pulley ratio. The upper test specimen was driven through a pair of gears mounted on the two shafts, variation of gear ratio and disc diameters allowed a predetermined value of slide/sweep ratio to be achieved by two different methods.

2.3. Experimental details

2.3.1. The discs

All tests were run with pairs of discs consisting of a case hardened En 34 mating disc running against a softer through hardened

	Pitting Specimen	Mating Specimen
Material	En 25	En 34
Heat treatment		Case hardened
Hardness V.P.N	300±10	800-850
Surface finish	Circumferential	Circumferential
	and axial	and axial
	0·25 – 1·8μm	0·25–1·8μm
Composition		
C	0·27 – 0·35	0.14 – 0·20
Si	0·16 – 0·35	0·10 – 0.35
Mn	0.50 – 0·70	0·30 – 0·60
S	0·050 max	0·050 max
P	0·050 max	0·050 max
Ni	2·30 – 2·80	1·50 – 2·06
Cr	0·50 – 0·80	——
Mo	0·40 – 0·70	0·20 – 0·30

Table 2.3.1

Material properties

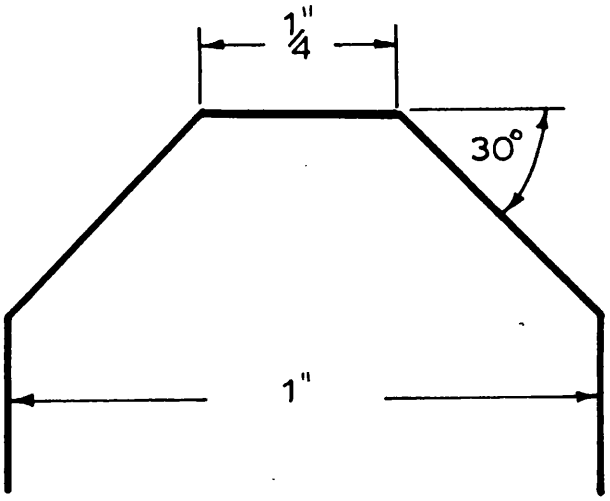


FIG. 2.3.1 Diagram of disc shape

En 25 test disc. Table 2.3.1 gives the properties of the steels.

The En 34 disc had a track width of 1" (25mm) whereas that of the En 25 had a narrower 0.25" (6.35mm) track width to enable the high Hertzian stresses required, to be obtained without overloading the rig. The disc also had a 30° chamfer to the edges as shown in Fig. 2.3.1. to avoid high edge stresses (Crook (1955)). Due to the smaller track width of the En 25 disc axial alignment was not necessary and the rig was self-aligning in the normal plane thus unevenly distributed loading was avoided.

Throughout the programme the ultimate aim was for comparison with gear rig results and thus the material combination was identical with that of the gears. As the manufacturing process for gears is expensive the total number required had to be kept to a minimum, therefore a case hardened steel, namely En 34, was chosen as the mating material as this was not likely to fail and could be used for further tests, always on the assumption that the surface finish would not alter appreciably. A second reason for the harder material was to compare evidence presented by Dawson (1962) that the surface roughness of this harder material was the controlling parameter for his D ratio.

2.3.2. Manufacture of discs

The disc surfaces were ground, one group circumferentially and the second group axially in order to compare simulated gear surfaces (axial) with normal disc surfaces (circumferential). One argument against disc representation is that the machining marks are orientated in a circumferential direction, which is in direct contrast to that of most gears where the marks are in the axial direction. Very little work has been carried out on tests using simulated gear finishes and to this end a simple and inexpensive technique was required to produce axial grinding marks on discs.

FIG. 2.3.2a Detailed photograph of axial grinding

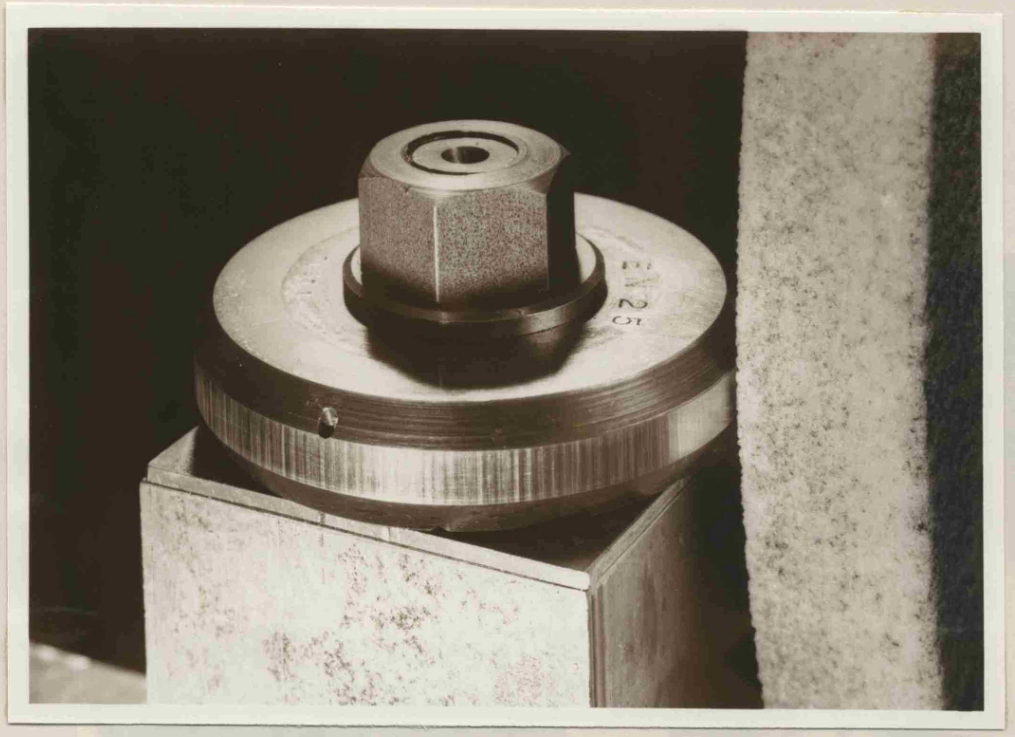


FIG. 2.3.2b Detail of disc and grinding wheel

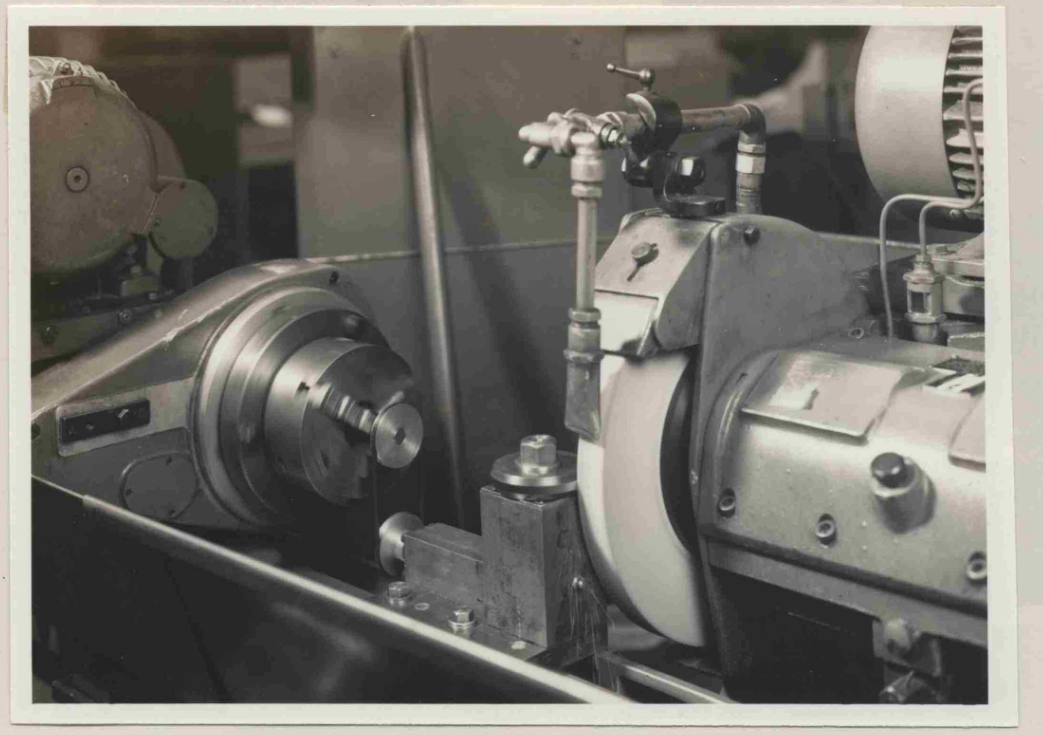


FIG. 2.3.2c General detail of axial grinding system

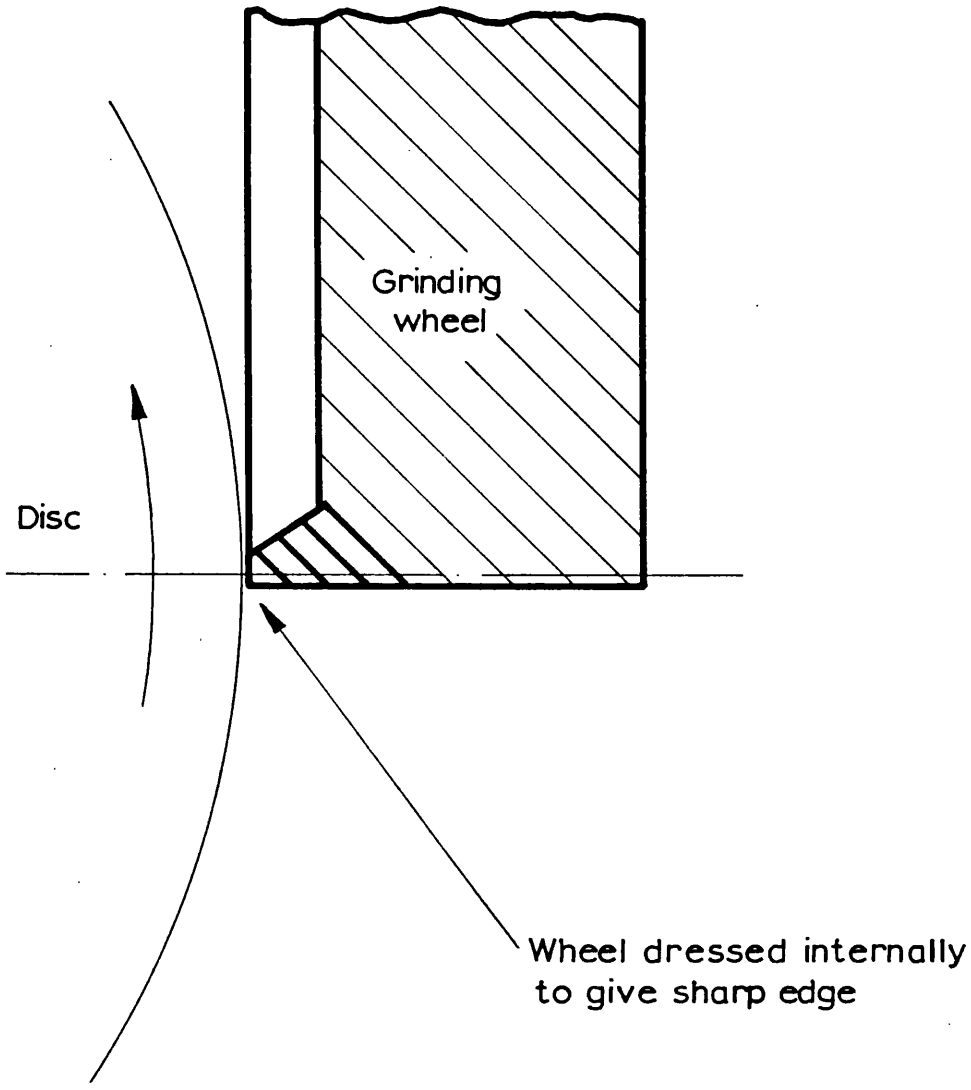
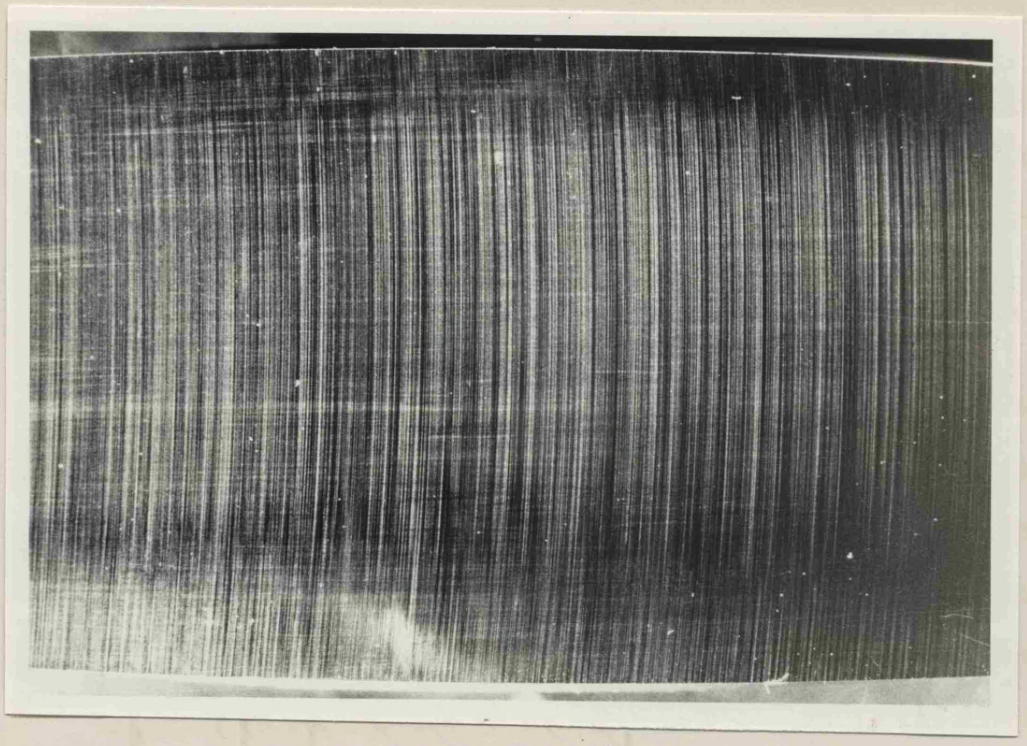
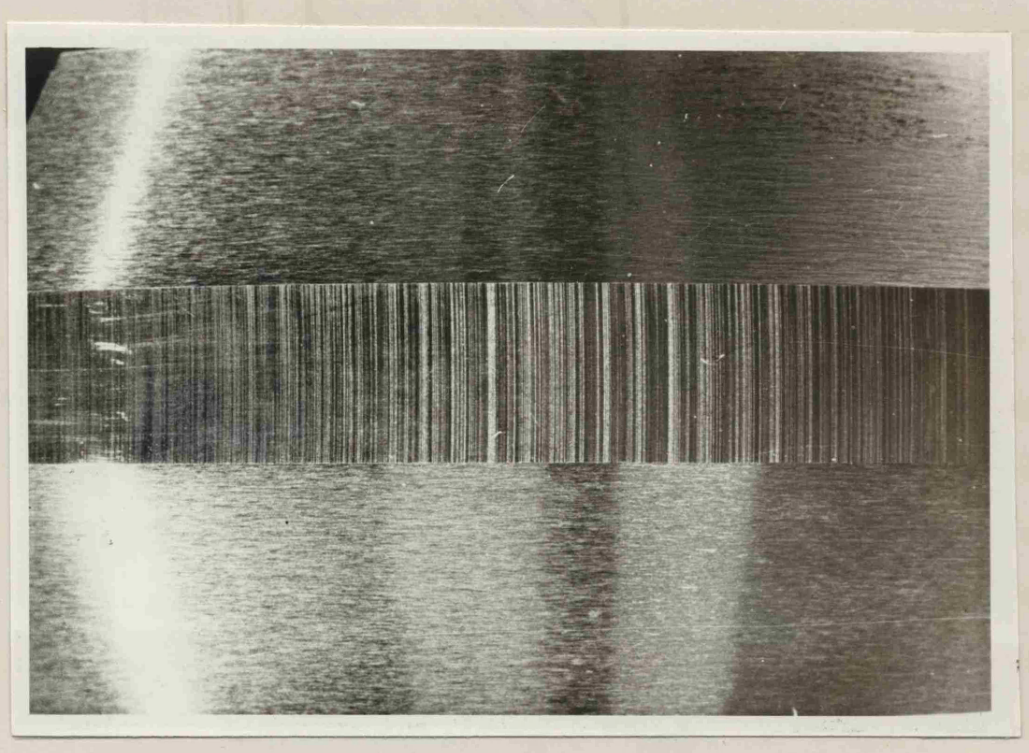


FIG. 2.3.3. Dressed grinding wheel for axial production



(a) En 34 disc



(b) En 25 disc

FIG. 2.3.4 Comparison of En 34 and En25 axially ground discs showing the small amount of curvature in the machining marks.

These discs were produced by using the side of the grinding wheel which, if large enough, would produce relatively straight lines across the track faces. Figs. 2.3.2abc, show detailed photographs of the grinding process and jig. The disc was mounted horizontally and rotated by means of a belt and pulley from the work head. The speed of rotation of the grinding wheel was kept constant and the speeds of the discs to be ground were 60 and 40 r.p.m. for the rough and smooth discs respectively. Before grinding the wheel was dressed as shown in exaggerated form in Fig. 2.3.3. Manufacture and development of the techniques involved was carried out in the workshops of the Engineering Department of the University of Leicester. * The ellipticity was checked by Talyrond and failed to show any large errors, the eccentricity of the bore to outside diameter was clocked by dial gauge which registered no greater error than that obtained by the normal grinding process of about 0.0001" (2.54μm).

Fig. 2.3.4 presents photographs of the final finish obtained on both En 34 and En 25 discs. The wider faced En 34 disc shows that the grinding marks do in fact take the form of small arcs but as the track width of the mating En 25 disc is only 0.25" (6.35mm) the curvature is negligible and the surface finish would seem to be identical in type to that of circumferentially ground discs. Cost of manufacture would not justify a more elaborate approach but work in America (P.M.Ku, private communication) has started on scuffing and related problems using axially ground discs. Their method is to grind in a similar fashion as one would for gears, this apparently is a very difficult and extremely expensive business.

---

* I am indebted to Mr. T. Pike for manufacture of the rig and production of the discs.

2.3.3. Running conditions

All tests were performed with the same applied load giving a maximum Hertzian compressive stress of 2×10^5 lbf/in^2 (1.37×10^6 kN/m^2).

A summary of the disc tests which have been performed is given in Tables 2.4.1 and 2.4.2a,b and these will be discussed in more detail shortly. At this stage it will be sufficient to explain that these tests are conveniently sub-divided as follows.

(i) A fairly comprehensive set of 10 tests (Table 2.4.1, tests 1.1 to 1.10) using axially ground discs and one to one gear drive. These tests will provide the main basis for a comparison between the disc and gear tests.

(ii) A more limited set of tests using a hunting tooth drive consisting of six tests with axially ground discs (Table 2.4.2a, tests 2.1a to 2.6a) and three tests with circumferentially ground discs (Table 2.4.2b, tests 2.1b to 2.3b). These tests were intended to provide some guidance on the possible influence of machining direction and the nature of the drive. Shotter (1961) and Dawson (1962) have both raised this point about the drive ratio having an effect, but the amount of work carried out is small and positive experimental verification has not yet been carried out. The change from one to one drive to hunting tooth drive would inevitably alter the slide/sweep ratio and this was kept constant by choosing suitable disc diameters. Values of diameters for the two drive ratios are given in Table 2.3.2. for a slide/sweep ratio of -0.037. This value was chosen from work presented by Dawson (1961) which showed that pitting of discs only occurs with a negative slide/sweep ratio (Fig. 2.1.2) and that for values between -0.001 and -0.04 there seemed to be little effect on the fatigue life, but for values of +0.001 to -0.001 (i.e. very close to the pitch line of gears) there was a very apparent effect and this region of slide/sweep ratios has been avoided.

The other variable parameters were viscosity, speed and surface

Disc	Diameter	Speed R.P.M.	Slide sweep ratio
En 25	2 · 96"	770 and 1500	− 0 · 027
En 34	3 · 04"	770 and 1500	+ 0 · 027
En 25	3 · 06"	720 and 1400	− 0 · 027
En 34	2 · 94"	770 and 1500	+ 0 · 027

Table 2 . 3 . 2

Disc diameters and speeds

	Viscosity Cp. 30° c	Viscosity Cp. 60°c	Viscosity Cp 100°c
Oil A	900	138	27
Oil B	145	33	9·2

Table 2 . 3 . 3.

Lubricant viscosity values

roughness. The viscosity was varied in two ways, firstly by controlling the inlet temperature and secondly by using two different oils. These oils had widely differing viscosities but were produced from the same base stock. As they were straight mineral oils (i.e. no additives) their chemical composition would be similar, their properties are given in Table 2.3.3. According to the work of Galvin and Naylor (1965) initiation and propagation of pre-pitting cracks should be similar for both oils. The bulk temperature of the discs was measured by an embedded iron/constantan thermocouple approximately $\frac{1}{8}$" (3.16mm) below the surface of the En 34 disc and the film thickness was calculated using these recorded values to determine the controlling viscosity. The surface velocity (Table 2.3.2) of the discs had two values for each group of tests. For the one to one drive this value would be slightly different than that of the hunting tooth ratio owing to the slight variation of diameters. The lower En 34 disc shaft was driven by the belt and had values of either 770 r.p.m or 1500 r.p.m. thus when using the hunting tooth drive the upper En 25 disc shaft would be rotating slower but this would be compensated for by the increase in diameter when calculating surface velocity values. Other than affecting the film thickness calculations slightly it was not thought that these differences would introduce any mechanical variations into the system. The third variable, surface roughness, was arranged such that there were three groups consisting of 10-12μ" (0.25 - 0.3μm), 25-30μ" (0.65 - 0.75μm) and 70-90μ" (1.8 - 2.3μm). Various combinations of these were then possible to run together.

2.4. Results of pitting tests

The ability to control the bulk temperature of the discs by controlling the inlet temperature was poor because the oil flow rate was too small, therefore the value was largely dependent on the rig characteristics

Test	En25 RPM	Temp °C	Film Thickness μm	Surface roughness μm				Oil
				σ_1	σ_2	σ_3	σ_4	
1·1	770	39	1·09	0·962	1·01	0·558	0·936	B
1·2	770	32	1·55	0·508	1·91	1·12	1·57	B
1·3	770	50	0·759	1·65	0·861	0·381	0·784	B
1·4	770	64	1·34	1·34	1·47	1·17	1·19	A
1·5	770	52	0·709	1·42	1·22	0·684	1·14	B
1·6	1500	59	0·962	0·356	0·482	0·254	0·508	B
1·7	1500	45	1·39	1·32	1·44	0·886	1·42	B
1·8	770	64	1·34	1·72	2·06	1·14	1·67	A
1·9	770	54	1·90	1·01	2·44	1·24	2·03	A
1·10	1500	70	0·634	1·65	1·12	0·551	1·01	B

σ_1 Initial surface roughness of En 25 disc
σ_2 " surface roughness of En 34 disc
σ_3 Final surface roughness of En 25 disc
σ_4 Final surface roughness of En 34 disc

Test	D Ratio				Cycles to pit $\times 10^5$
	D_1	D_2	D_3	D_4	
1.1	1.8	1.9	1.4	1.7	3·03
1.2	1.6	2·5	1.8	2.0	2·81
1.3	3.3	2·3	1·5	2·1	4·90
1.4	2·1	2·2	1·8	1.8	4.10
1.5	3·7	3·4	2, 6	3·2	1·38
1·6	0·87	1.00	0·80	1·07	>22·0
1·7	2·0	2·1	1·7	2·0	1.68
1·8	2·8	3.1	2·1	2·5	1·08
1·9	1·8	2·9	1·7	2·1	1·36
1·10	4.0	3·5	2·5	3·2	0·697

$$D_1 = \frac{\sigma_1 + \sigma_2}{h}$$

$$D_2 = \frac{2\ \sigma_2}{h}$$

$$D_3 = \frac{\sigma_3 + \sigma_4}{h}$$

$$D_4 = \frac{2\ \sigma_4}{h}$$

TABLE.2.4.1 Summary of tests: axially ground discs, one to one gear ratio.

Test	En 25 RPM	Temp °C	Film Thickness $h\mu m$	Surface roughness μm				Oil
				σ_1	σ_2	σ_3	σ_4	
2·1a	720	67	1·14	0·533	0·684	0·280	0·583	A
2·2a	720	70	0381	2·23	1·62	0·507	1·24	B
2·3a	1400	48	1·27	0·203	0·835	0·406	0·734	B
2·4a	1400	53	1.04	1.98	1·65	0·734	1·27	B
2.5a	720	41	1·01	0·305	0·280	0·254	0·280	B
2·6a	720	37	1·22	0·406	1·75	0·658	1·34	B

σ_1 Initial surface roughness of En 25 disc
σ_2 " " " of En 34 disc
σ_3 Final surface roughness of En 25 disc
σ_4 " " " of En 34 disc

Test	D ratio				Cycles to pit $\times 10^5$	Remarks
	D_1	D_2	D_3	D_4		
2·1a	1·1	1·2	0·76	1·0	13·4	-40μm on diam
2·2a	11	8·7	4·7	6·6	>0·65	-550μm on diam
2.3a	0·82	1·2	0·90	1·2	18·0	-120μm on diam
2.4a	3.5	3.2	1·9	2.8	>3.12	Gears bottomed
2.5a	0·57	0·60	0·53	0·55	>23	Zero wear
2.6a	1.8	2.9	1.6	2.2	>0·44	-370μm on diam

$$D_1 = \frac{\sigma_1 + \sigma_2}{h}$$

$$D_2 = \frac{2\ \sigma_2}{h}$$

$$D_3 = \frac{\sigma_3 + \sigma_4}{h}$$

$$D_4 = \frac{2\ \sigma_4}{h}$$

TABLE 2.4.2a. Summary of tests : axially ground discs, hunting tooth drive

Test	En 25 R.P.M	Temp °C	Film thickness $h/\mu m$	Surface roughness μm				Oil
				σ_1	σ_2	σ_3	σ_4	
2·1b	720	66	1·19	0·507	0·608	0·203	0·533	A
2·2b	1400	76	0·860	0·861	0·254	0·608	0·178	A
2·3b	1400	54	1·01	0·280	0·861	0·203	0·911	B

σ_1 Initial surface roughness of En 25 disc

σ_2 Initial surface roughness of En 34 disc

σ_3 Final surface roughness of En 25 disc

σ_4 Final surface roughness of En 34 disc

Test	D RATIO				Cycles to pit $\times 10^5$	REMARKS
	D_1	D_2	D_3	D_4		
2·1b	0·90	1·02	0·63	0·93	6·07	−25μm on diam
2·2b	1·3	0·59	0·90	0·41	> 24	−8μm on diam
2·3b	1·1	1·7	1·1	1·8	>5·14	Track mushroomed

$$D_1 = \frac{\sigma_1 + \sigma_2}{h_o}$$

$$D_2 = \frac{2\sigma_2}{h_o}$$

$$D_3 = \frac{\sigma_3 + \sigma_4}{h_o}$$

$$D_4 = \frac{2\sigma_4}{h_o}$$

TABLE. 2.4.2b. Circumferentially ground discs hunting tooth drive.

and the other pre-determined running conditions. This in itself was not too limiting and a temperature variation ranging from 30°C to 70°C was obtained, the resultant film thickness variation incorporating the other two controlling parameters was of the order of 25μ" (0.65μm) to 75μ" (2μm). The surface finish of the discs was measured by Talysurf with the radius attachment and a mean of four different measurements, taken at approximately 90° intervals around the circumference, was used as the initial value. This was then repeated after the tests had been completed and both mean values are recorded in Table 2.4.1. The film thickness was calculated from the Dowson and Higginson equation (1961), (equation 2.1.2). The value of the temperature used for this equation was that attained during steady state conditions and normally took approximately five minutes to reach. The predicted film thickness along with the values of surface roughness were used to provide four values of the D ratio as defined in Tables 2.4.1 and 2.4.2a,b. This may be compared with earlier work by Dawson (1962) who defined the D ratio in two ways as follows.

(1) The sum of the initial surface roughness of both the hard disc and the soft disc divided by the calculated oil film thickness.

(2) Twice the value of the initial surface roughness of the hard disc divided by the oil film thickness.

The two further definitions of D used here are basically similar to the first but the initial surface roughness values have been replaced by the final values determined after completion of the tests. A second minor difference in the definitions is that Dawson used a value of peak to valley average (P.V.A) measured from Talysurf traces, the results reported here use the old C.L.A value now called the roughness average Ra.

The discs in any one test were run together until enough pits had appeared to enable graphs to be plotted of number of pits visible to the naked eye on disc track against cycles run. If the discs had not

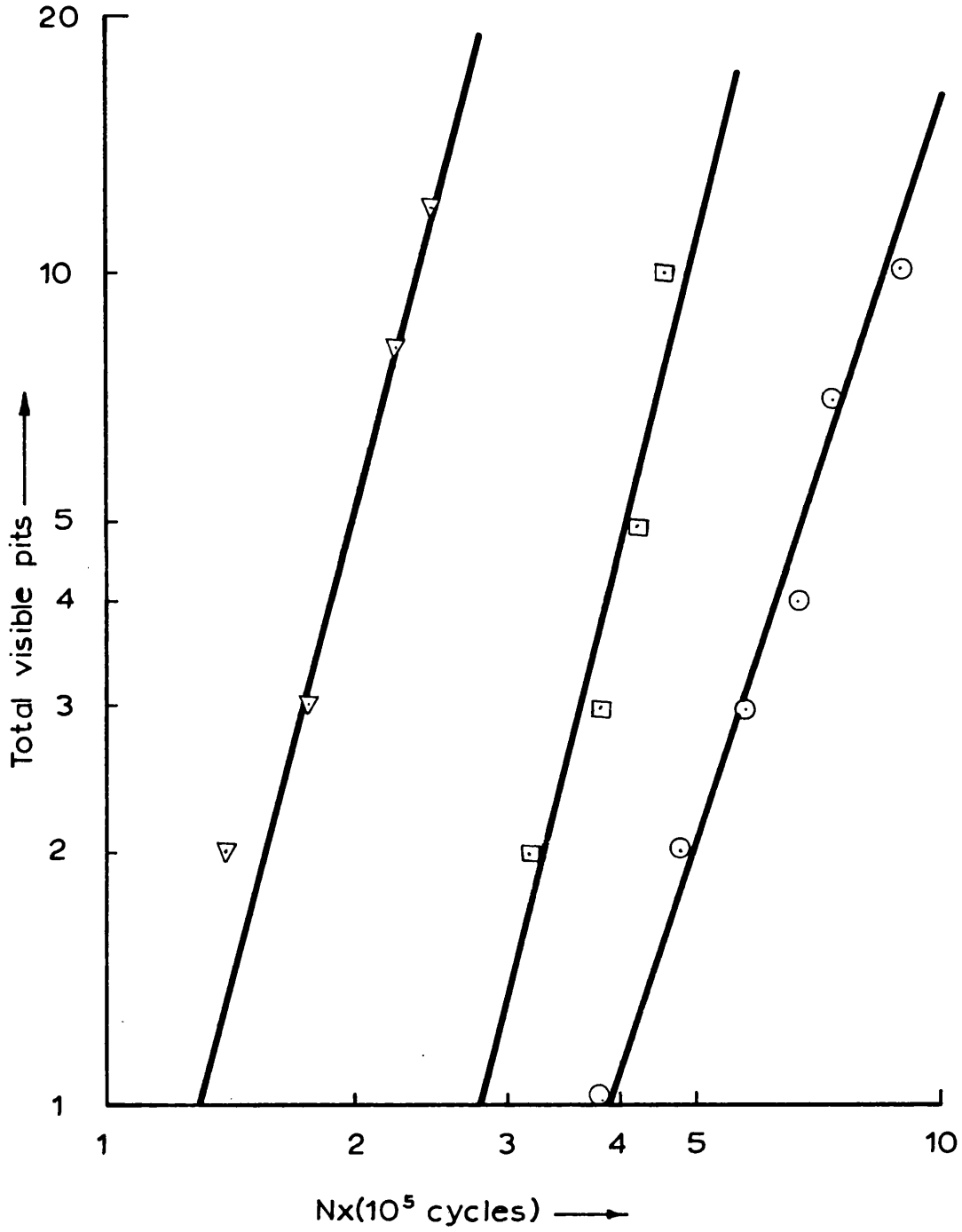


FIG. 2.4.1. Number of pits as a function of cycles run

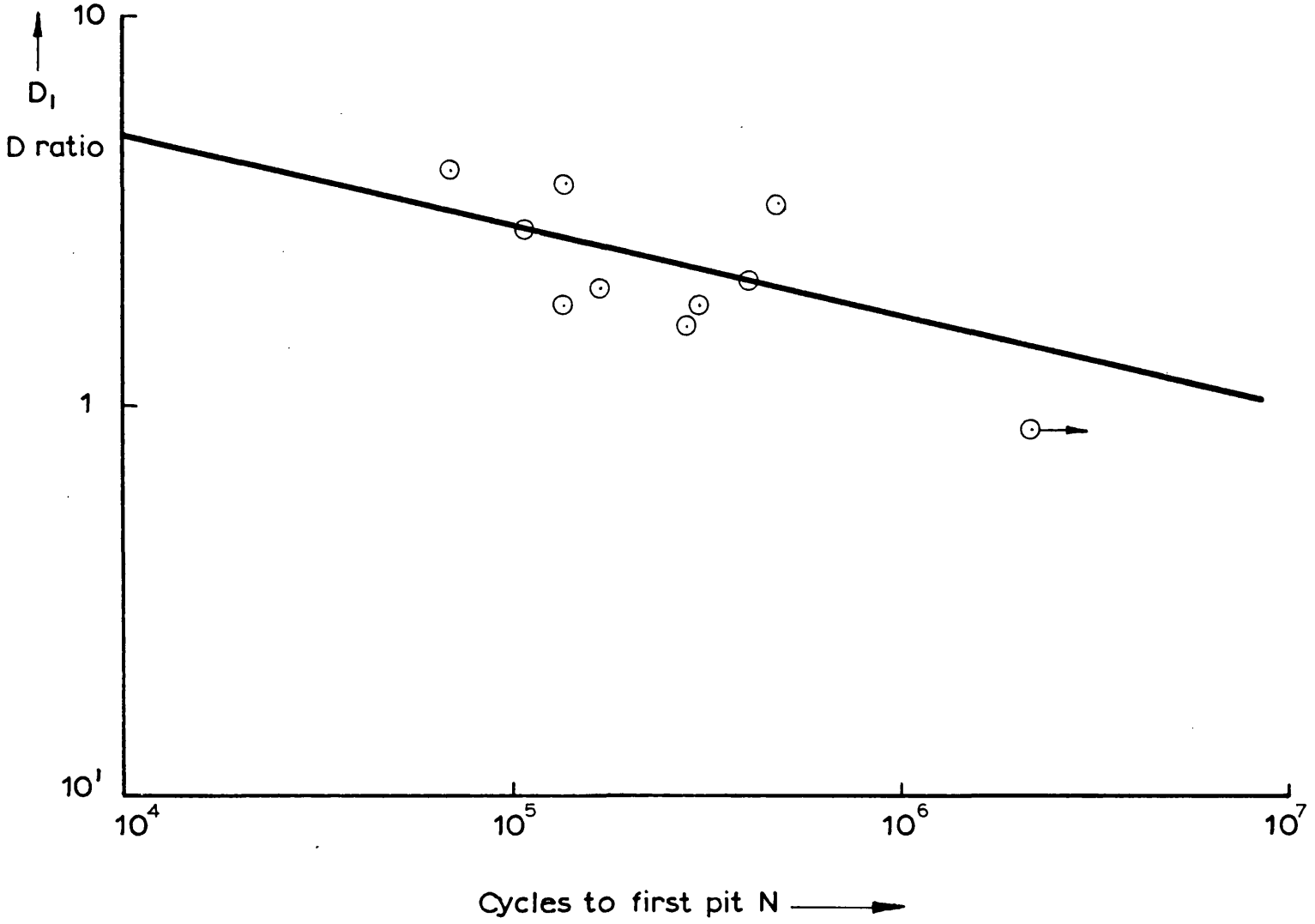


FIG. 2.4.2a D_1 as a function of cycles to first pit N

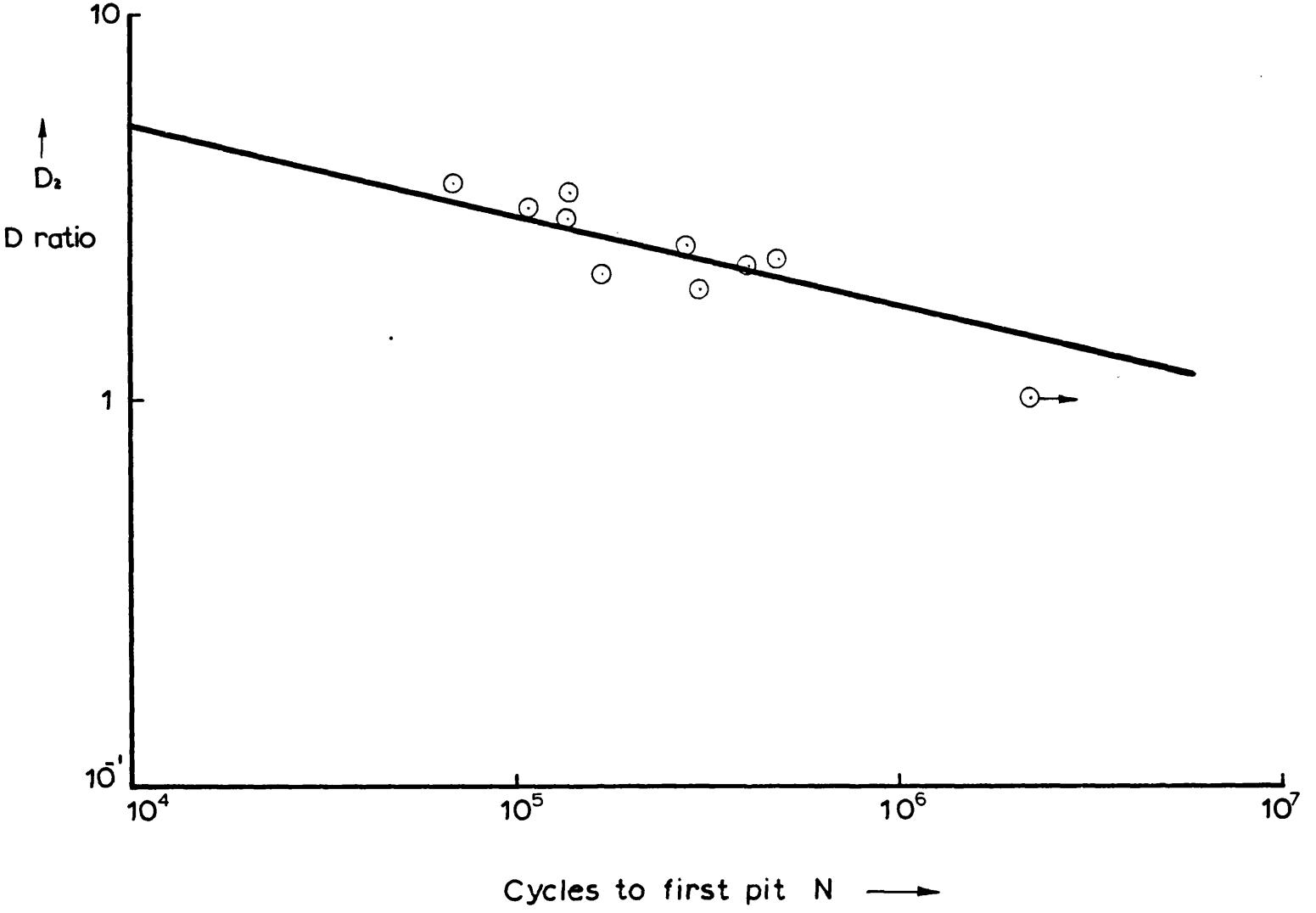


FIG. 2.4.2b D_2 as a function of cycles to first pit N

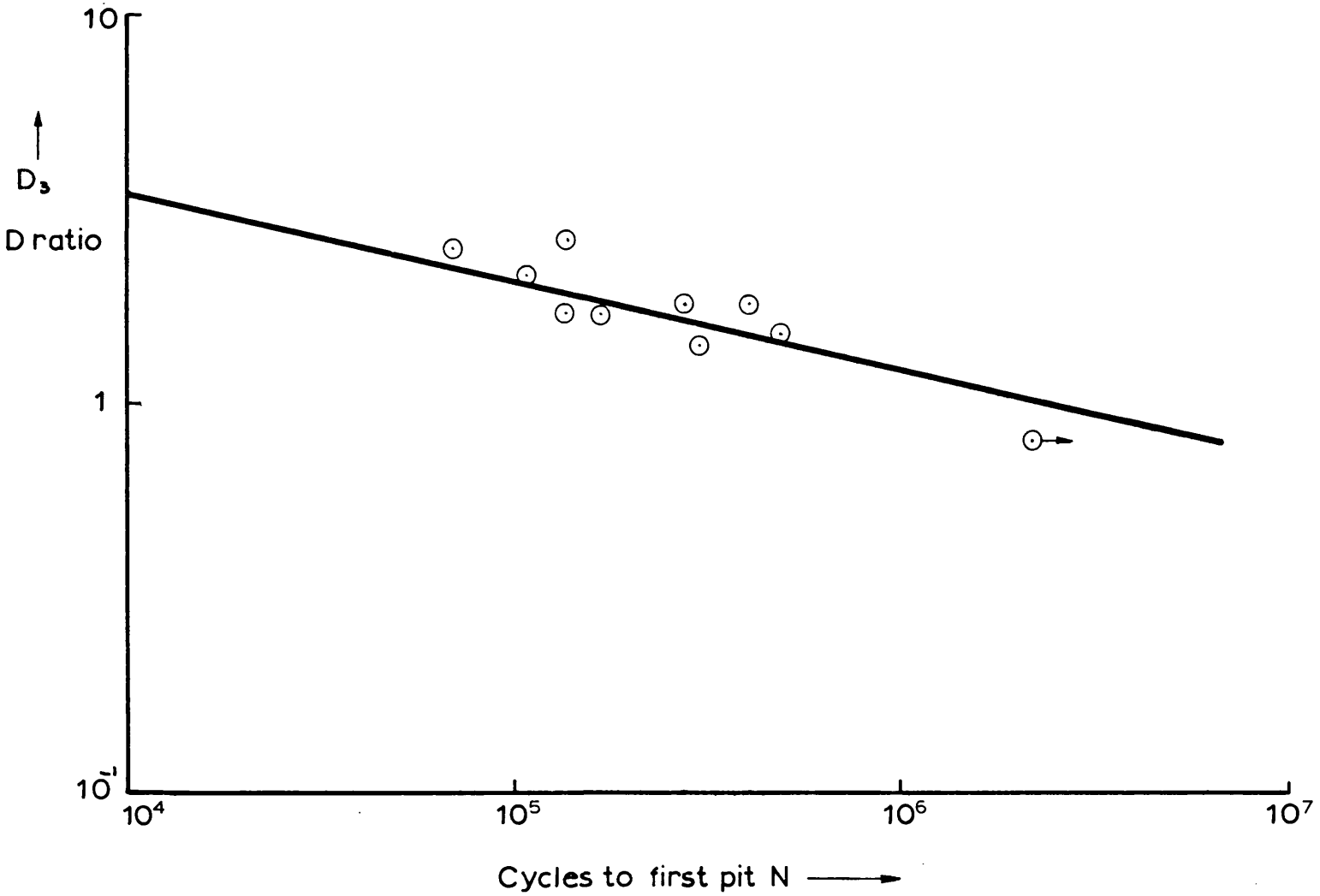


FIG.2.4.2c D_3 as a function of cycles to first pit N

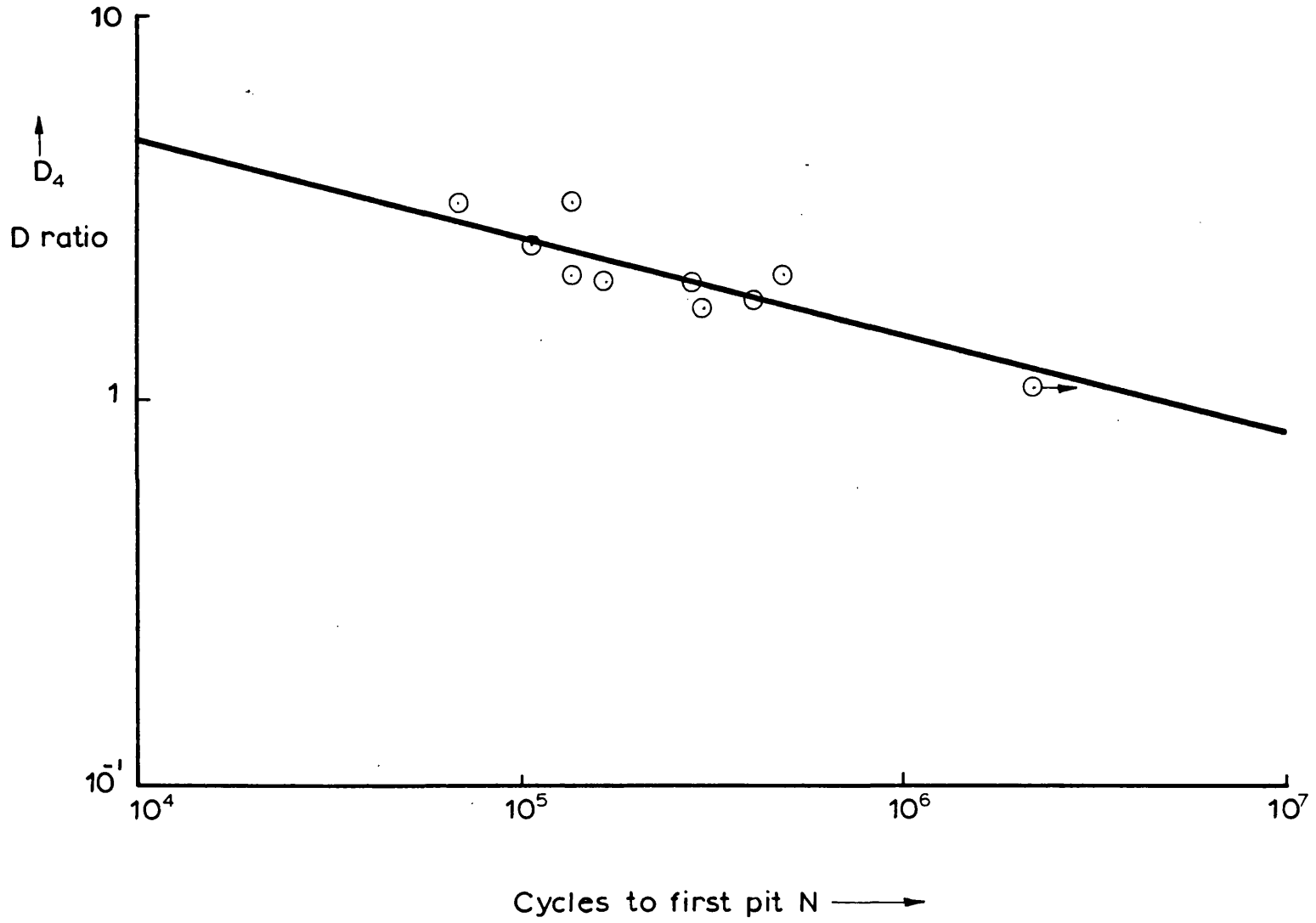


FIG. 2 .4 . 2d D_4 as a function of cycles to first pit N

Result of regression analysis

$D_1 = (\sigma_1 + \sigma_2) / h$

$\text{Log}_{10} D_1 = -0.232 \log_{10} N + 1.620$

Regression coefficient = 0.439

Standard error of regression = 0.179

$D_2 = 2\sigma_2 / h$

$\text{Log}_{10} D_2 = -0.236 \log_{10} N + 1.660$

regression coefficient = 0.727

Standard error of regression = 0.084

$D_3 = (\sigma_3 + \sigma_4) / h$

$\text{Log}_{10} D_3 = -0.217 \log_{10} N + 1.399$

regression coefficient = 0.688

Standard error of regression = 0.087

$D_4 = 2\sigma_4 / h$

$\text{Log}_{10} D_4 = -0.248 \log_{10} N + 1.665$

regression coefficient = 0.758

Standard error of regression = 0.081

TABLE. 2 . 4 .3.

pitted after 2 x 10^6 cycles the test was stopped due to lack of time. The graphs obtained for number of pits against cycles run were extrapolated back to obtain a value of cycles required to produce one pit, this point was set as the failure limit of the discs and examples of the graphs are provided in Fig. 2.4.1. The results obtained from the work described in this chapter did not yield as much information as was initially hoped. The ability to predict cycles run to first pit was in itself rather more difficult than expected, there being considerable scatter in some of the graphs of number of pits against cycles run and therefore a regression analysis was used to determine the significance of the results. Table 2.4.1 presents a summary of the ten tests performed with one to one drive; all these tests used axially ground discs. The cycles to first pit (N) obtained in these tests are plotted as a function of the four values of D in figures 2.4.2a,b,c,d. Four regression lines of $\log_{10}$ D on the values of $\log_{10}$ N have been calculated along with the correlation coefficient r and the standard error of estimate, these values are given in Table 2.4.3. As the Ra value of roughness varied around the disc tracks it was not thought worthwhile attempting to obtain a sensible value of P.V.A and so the calculated D values are somewhat lower than those of Dawson. Values roughly equivalent to those of Dawson may be obtained by multiplying the Ra value by a parameter K derived by Rubert (1959). An analysis was carried out using this approach to calculate D but the results of coefficient of regression and standard error of estimate were slightly worse than those already calculated and therefore these revised results are not quoted here.

Tables 2.4.2a and 2.4.2b present results obtained on the disc machine using the hunting tooth ratio of 29 to 31. Table 2.4.2b presents the results of discs ground in the conventional manner in a circumferential direction and Table 2.4.2a those results obtained for discs ground axially. It was not possible to carry out a regression analysis on the results of

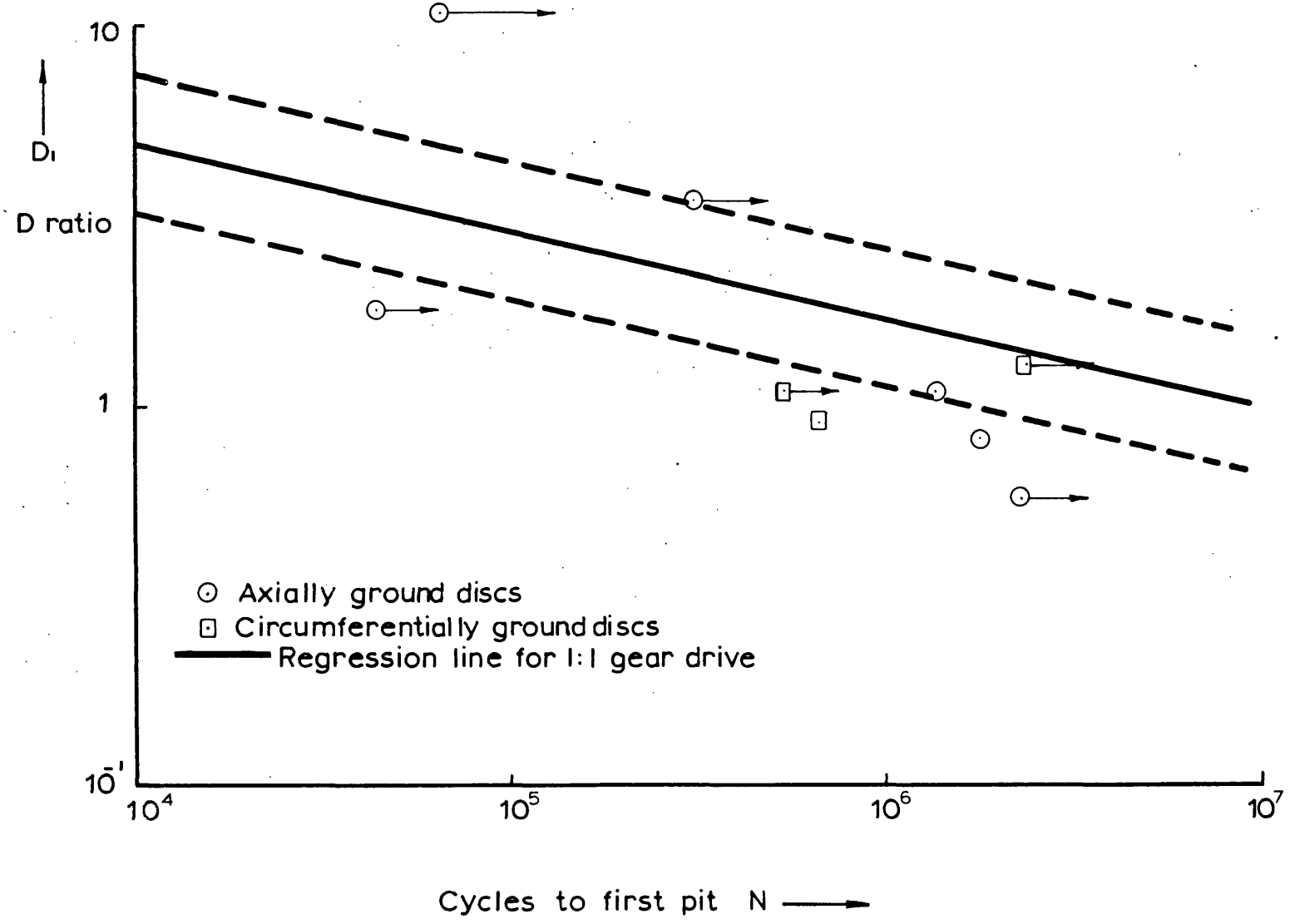

10
D_1
D ratio
1
10^{-1}
10^4
10^5
10^6
10^7
Axially ground discs
Circumferentially ground discs
Regression line for 1:1 gear drive
Cycles to first pit N

FIG. 2.4.3a

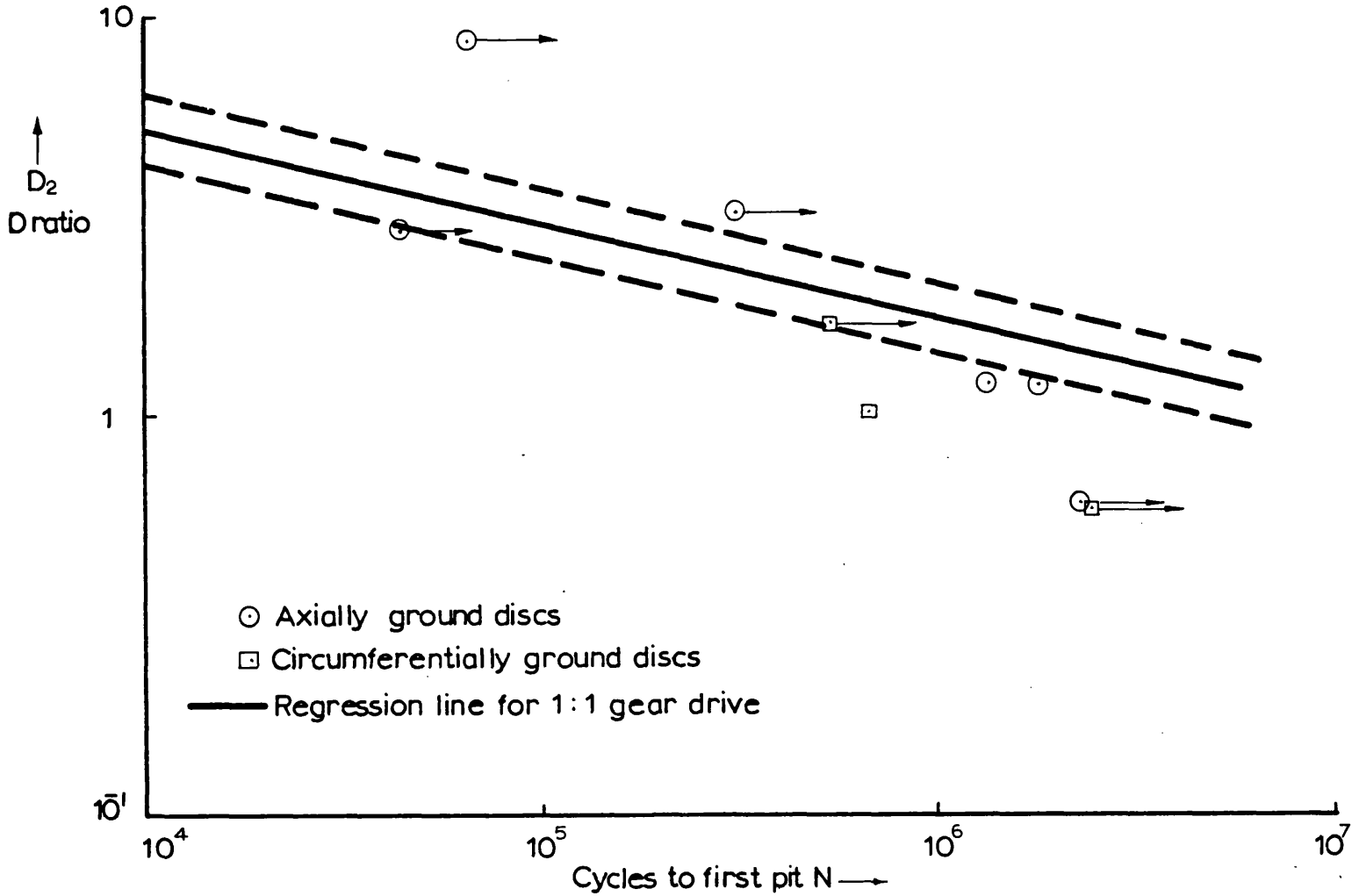

10
1
10^{-1}
D_2
D ratio
10^4
10^5
10^6
10^7
Cycles to first pit N
⊙ Axially ground discs
⊡ Circumferentially ground discs
Regression line for 1:1 gear drive

FIG. 2.4.3b

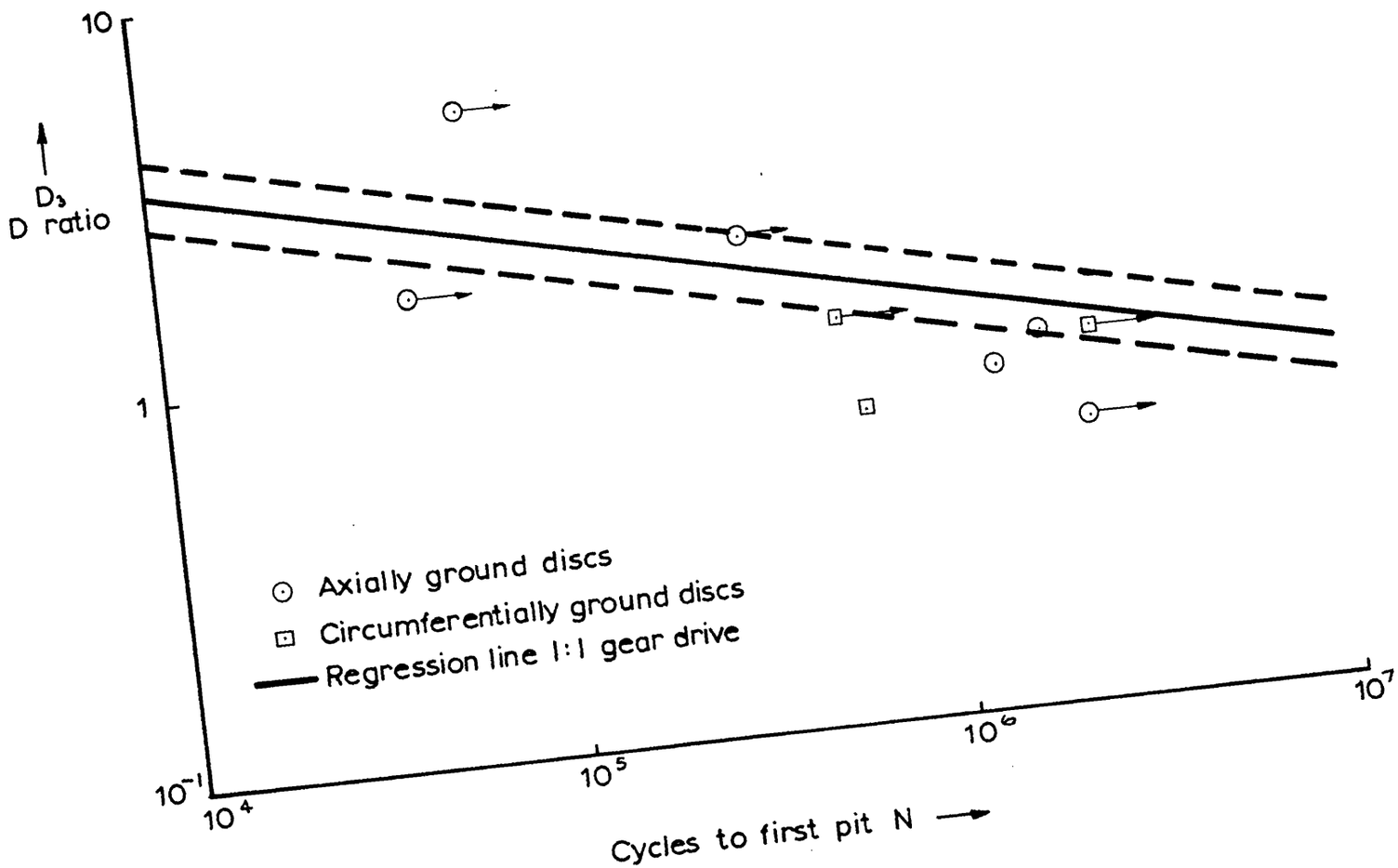

10
1
10^{-1}
D_3
D ratio
Axially ground discs
Circumferentially ground discs
Regression line 1:1 gear drive
10^4
10^5
10^6
10^7
Cycles to first pit N

FIG. 2.4.3c

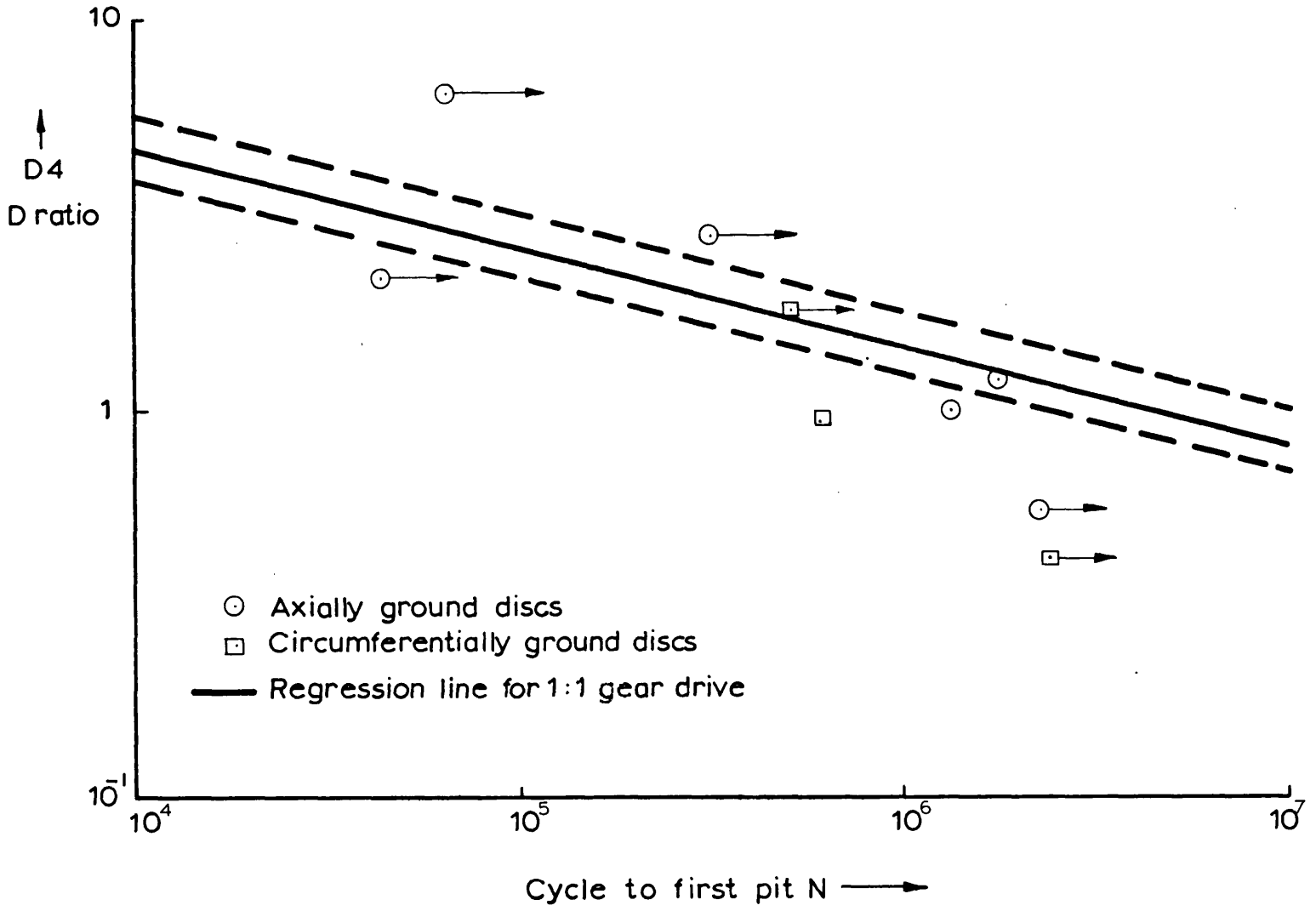

10
D4
D ratio
1
10^{-1}
10^4
10^5
10^6
10^7
Axially ground discs
Circumferentially ground discs
Regression line for 1:1 gear drive
Cycle to first pit N

FIG. 2.4.3d

FIG. 2.4.4. Photograph of typical shape pitting crack.

Tables 2.4.2a,b but for comparison the D ratios have been plotted as a function of either total cycles run (non-failed tests) or cycles to first pit (N) and are presented in figures 2.4.3a,b,c,d. Superimposed on these figures are the regression lines of the earlier results from Table 2.4.1, the broken lines representing the standard errors of estimate for the regression. Result 2.1b obtained from circumferentially ground discs lies well outside the standard error lines of all figures, and likewise for result 2.1a, obtained from axially ground discs, the third result 2.3a, is close or marginally inside the error lines. The other tests shown in these figures had no defined failure limit and are therefore marked with an arrow.

The disc machine used had a tendency to vibrate occasionally and it seemed that it may well have been a function of the surface roughness as it was on the rougher discs that this occurred. Talyrond traces on a selection of discs did not show any excessive undulations on the track suggesting macro-roughness and thus it could have been a function of the micro-roughness. These vibrations had to be tolerated as the system could not be damped by a dash pot. The maximum load variation before pitting was normally within the limits of $\pm$ 2 lbs to $\pm$ 5 lbs spring balance reading. As the tests progressed and the pits became more numerous this occasionally increased but not normally until 3 or 4 large pits were visible and a reasonable indication of pitting had been achieved; this was a natural consequence of the material being removed from the surface. The pits formed in a random fashion about the circumference and normally away from the edge of the track thus suggesting that the discs were loaded and aligned symmetrically. As reported by other authors the pre-pitting cracks and the pits themselves were of a characteristic fan or V shape an example of which is shown in figure 2.4.4.

The results obtained from a hunting tooth drive were very different to those of the one to one drive. Measurements of outside diameters of

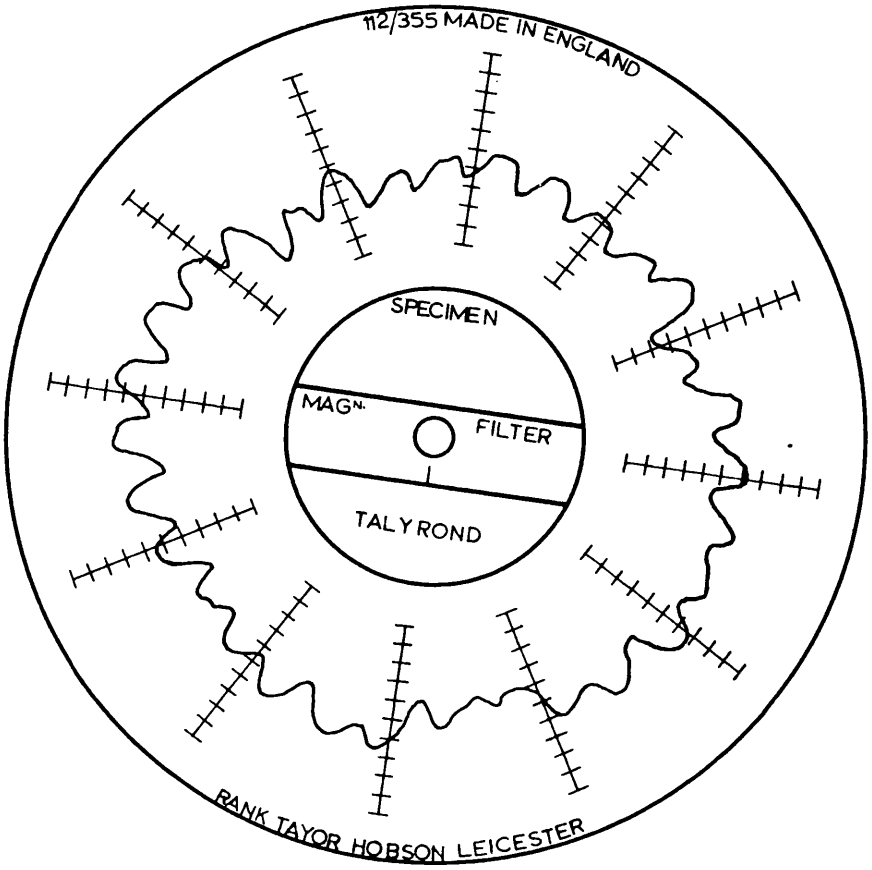
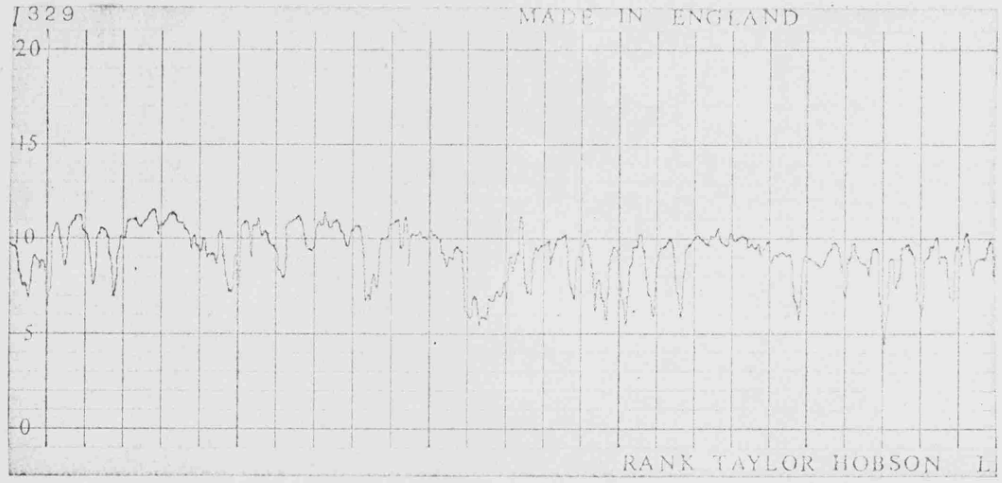


FIG. 2.4.5. Talyrond trace of En. 25 disc after running Drive ratio 29/31.
The figure shows 31 lobes on the disc surface. Magnification:- x 10^3

the axial one to one drive discs showed very little loss or wear and in the majority of tests were the same before running as after. In sharp contrast the softer discs on a hunting tooth drive nearly always showed considerable wear, to the extent of 10-15 x 10^{-3} ins. (250 - 380μm) loss on diameter. The remarks column of Tables 2.4.2a and 2.4.2b give some indication of this loss on diameter. This incidence of wear could not be ascribed to excessively high values of the D ratio alone since wear did not occur in the tests in which similar values of the initial D ratio (D_1 or D_2) were used with one to one drive. Thus there is strong evidence that this wear is associated with the use of a hunting tooth drive. Further support to this statement was the fact that certain of the En 25 discs developed flats on the test track which showed on the Talyrond as 31 lobes (Fig. 2.4.5); which coincides with the number of teeth on the driven gear mounted on the En 25 disc shaft. This was enhanced to a greater extent in one test where the loss of 30 x 10^{-3} ins. (750μm) on diameter finally caused the gears to bottom and gross damage to the addendun of both gears resulted. Until this test the gears had not bottomed and the system itself probably caused the generation of flats.

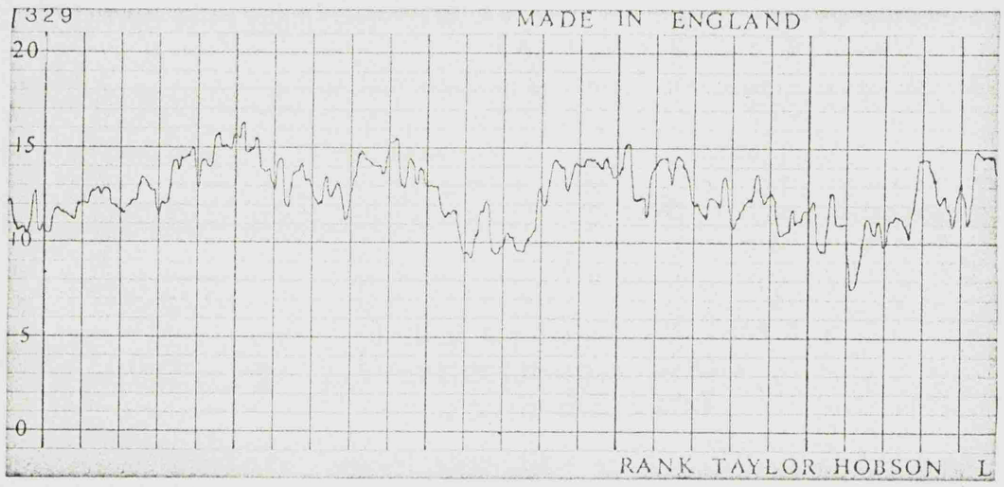
It is also significant to note that the tests stopped in Table 2.4.2a, without exception, were those where the initial or final surface roughness of the En 34 disc was greater than the predicted oil film thickness. There was only one result for circumferentially ground discs of Table 2.4.2b that was stopped due to a loss on diameter, and in this case the final roughness of the En 34 disc was only just less than the predicted oil film thickness. Thus, for a hunting tooth drive, the fact that the surface roughness of the harder disc is greater than the oil film has far more significance than when a one to one drive ratio is used.



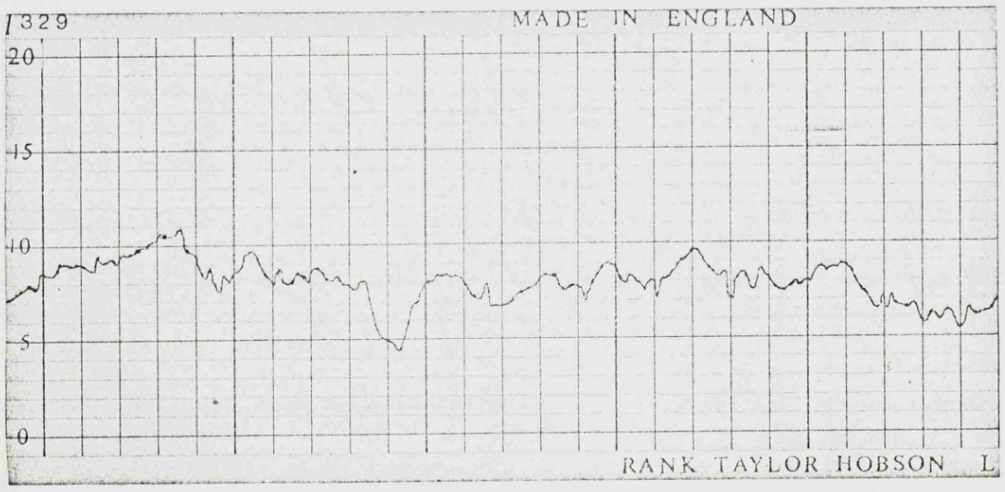
Initial trace En. 25 Test 1·2
Vertical mag. x 5000
Horizontal mag x 100

Final trace En. 25. Test 1·2
Vertical mag. x 2000
Horizontal mag x 100

FIG. 2.5.1a. Talysurf traces taken before & after running



Initial trace En. 25. Test 1·8
Vertical mag x 2000
Horizontal mag x 100



Final trace En 25 Test 1.8
Vertical mag. x 2000
Horizontal mag. x 100

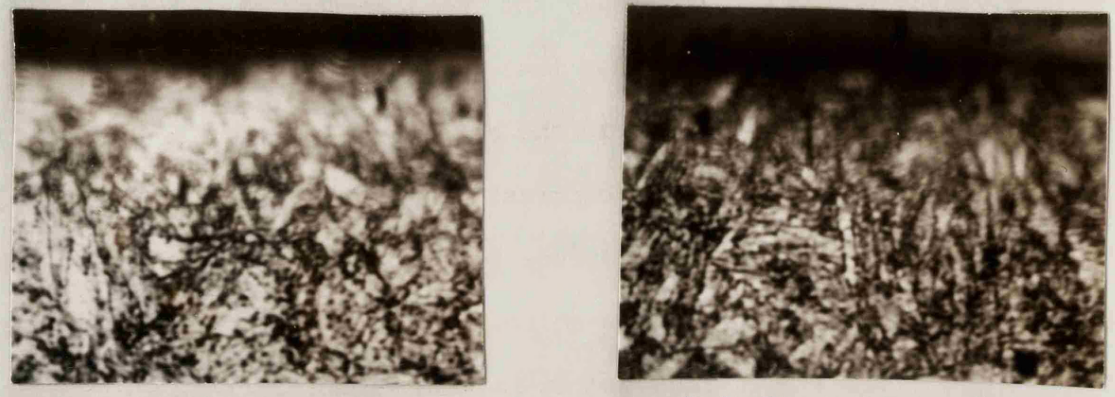
FIG. 2.5.1b Talysurf traces taken before & after running

2.5. Disc examination

2.5.1. Talysurf

Talysurf traces were taken of all discs before and after running and Table 2.4.1 gives results of the discs used in the first set of tests, axially ground and run with one to one gear ratio. It can be seen that in 9 out of 10 cases the En 34 mating disc showed some modification of the surface and a decrease in the Ra value; test 1.6 indicated a slight roughening but this was within the scatter of the results obtained before running. Two surfaces of the En 25 discs increased in roughness and in both situations it was a rough En 34 disc running against a smoother En 25 disc (tests 1.2 and 1.9), indication being that the surface of the En 25 disc tended to conform to that of the harder En 34. These two results substantiate those reported by Dawson (1962) which led him to the conclusion that the surface finish of the harder disc was the controlling parameter in the D ratio. Further consideration of Table 2.4.1 shows that test 1.8, where a rough En 34 disc ran against a slightly smoother En 25 disc, resulted in a 30% decrease in the Ra value for the softer disc after completion of the test. Therefore there is evidence that the surface finish of the harder disc is not necessarily the controlling parameter. Figs. 2.5.1a and 2.5.1b show two sets of Talysurf traces of En 25 discs before and after running. Fig. 2.5.1a shows how the smooth En 25 disc has been roughened by the En 34 mating disc (test 1.2) and Fig. 2.5.1b shows the reverse process of a rough En 25 disc being smoothed by a medium rough En 34 disc (test 1,8).

Tables 2.4.2a and 2.4.2b show the results obtained from the experiments with hunting tooth drive. Again in these Tables tests 2.3a and 2.6a show an increase of surface roughness, all the others decreased; however the validity of these results is not as sound as the tests of



(a) Axial (b) Circumferential

Mag. x 2,000

FIG. 2.5.2

Etched sections of un-run axially and circumferentially ground discs. Loss of focus due to curvature at the edge.

Mag. x 1,500

FIG. 2.5.3

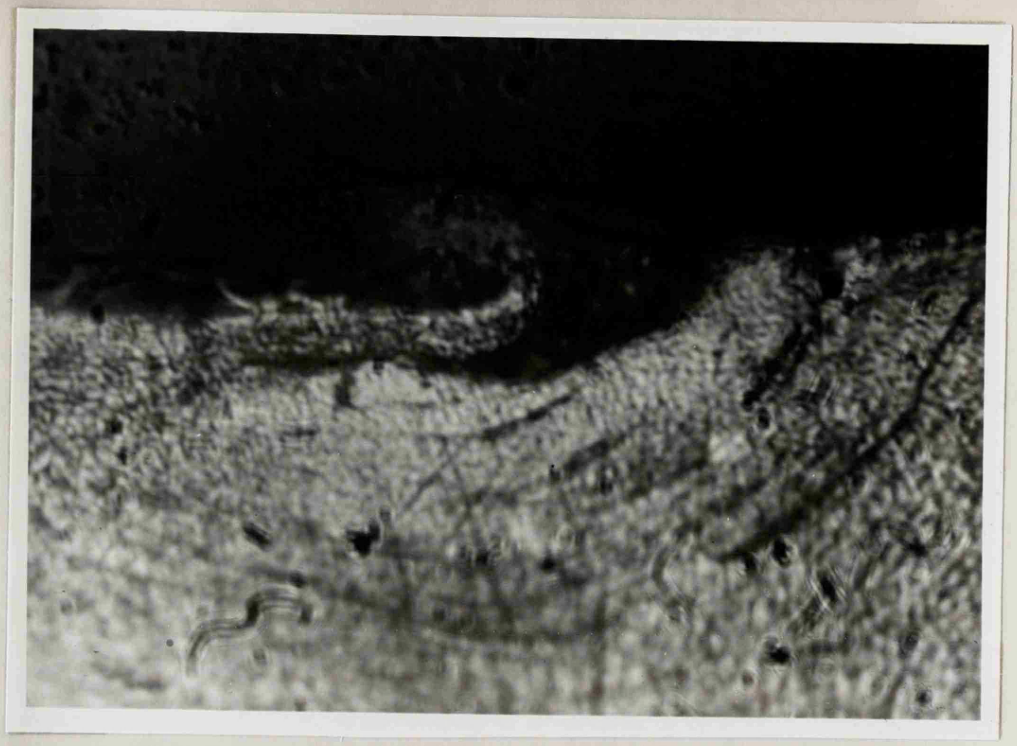
Sectioned disc showing sub-surface crack

Table 2.4.1 because the disc diameters had decreased and considerable wear had taken place. Nevertheless, they are in broad agreement with the previous set of results. Generally the final surface roughness of the softer En 25 disc is much less than that of the En 34 discs suggesting the softer disc does not conform to the harder to any great extent.

Although these results may contrast with those of Dawson they do not necessarily contradict his work because the method of asperity deformation is probably different. There are no rough circumferentially ground discs to compare with his experiments but it is quite possible that they would have altered in a similar fashion to those reported by Dawson. The orientation of the ellipsoidal asperities would be in the direction of motion and the harder surface would either cut or deform a groove in a softer disc along the direction of motion. This would not be the same action with axially orientated asperities. Axial asperities would quite probably cause a smearing action, due to the slide/sweep ratio, and thus have a tendency to smooth away the peaks of high asperities. When the softer surfaces were smooth and the harder rough, one would expect a slight roughening as the smooth surface is forced to conform but it would not be expected to increase to the same order of magnitude as that of the harder surface. One might also expect, as is the case in Fig. 2.5.1a, a longer wavelength structure of the surface due to the slide/sweep ratio.

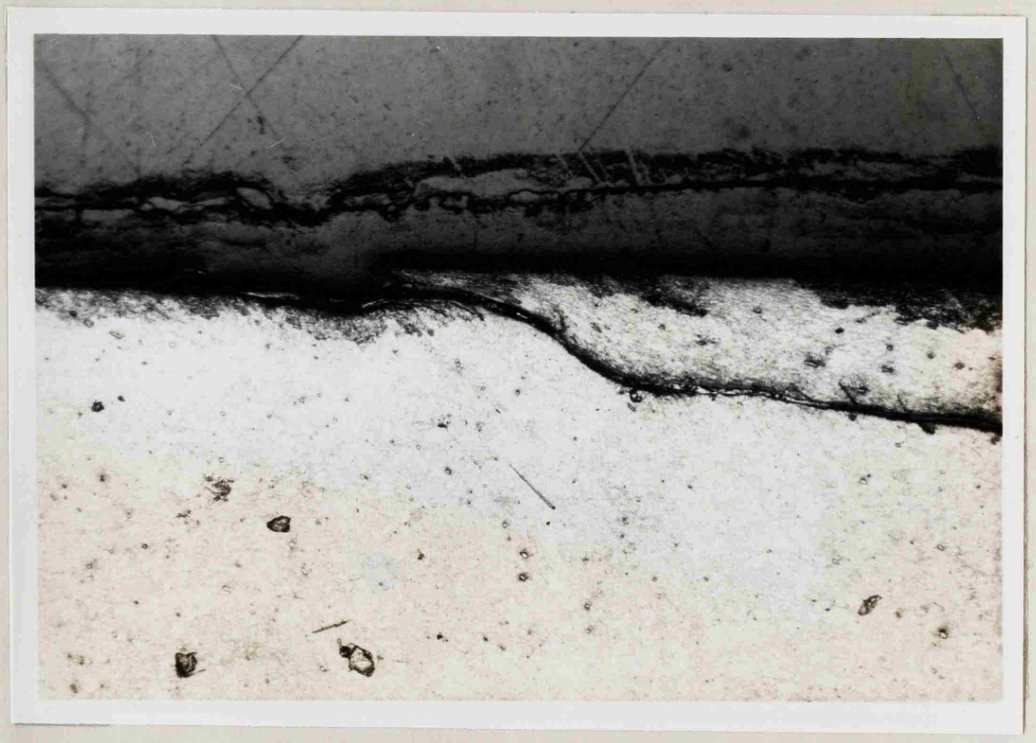
2.5.2. Metallographic sectioning

A small selection of the En 25 discs that ran were sectioned, along with two un-run discs one axially ground the second circumferentially ground. Figures 2.5.2a,b are photographs of sections on the un-run discs polished by $\frac{1}{4}$ micron diamond paste and given a light etch, these sections may act as a comparison. Fig. 2.5.3 shows subsurface cracks in a disc run to pitting failure and supports evidence of similar cracks observed by



Mag. x 2,000

FIG. 2.5.4 Etched section of En 25 disc showing strain lines



Mag. x 80

FIG. 2.5.5. Section of En 25 disc showing plastically deformed pitting crack.

Dawson (1968). They do not necessarily prove that their origin is below the surface for they could in fact be the tail end of a small surface originated crack, and it would be difficult to provide conclusive evidence to the contrary; however Fig. 2.5.4 does give further support to the probability of subsurface crack formation. This specimen was lightly etched with Fry's reagent to show any strain lines and there seems to have been a peculiar stress field produced in this area at a depth of the order of 200μ" (5×10^{-3}mm). Discussion of how this was caused is obviously speculative but the depth is comparable to that at which the maximum Hertzian stress of two interacting asperities would occur.

As mentioned earlier some of the discs formed flats or lobes on their tracks, one such disc has been sectioned (Fig. 2.5.5) and it does in fact show pitting cracks. These cracks have been smeared at the surface in such a way that the flake had been distorted along the surface in the direction of motion, whether the cracks would have continued to form pits is not known as the test was stopped due to excessive vibrations. The unusual feature of this test was the fact that in no other test had so many pre-pitting cracks been observed, normally if a crack was observed it progressed quite rapidly to a pit, usually before the second crack was visible.

2.6. Discussion and conclusions

Because the main purpose of these results was to provide support for the gear tests they are not as comprehensive as might have been hoped. Nevertheless they provide a number of insights into the mechanisms of pitting and give further support to Dawson's concept of the D ratio. The main conclusions are as follows.

(1) In agreement with Dawson's work the sum of the initial surface roughness does not seem to be the most appropriate way of expressing the influence of surface roughness. The regression analysis indicates that one of the other three definitions of D is to be preferred, but since the correlation is similar for all regressions there is no clear indication of the best choice. Dawson (1962) suggested that twice the initial surface roughness of the harder disc should be used, and for reasons of simplicity these results re-enforce that argument. On practical grounds this definition is to be preferred because the initial surface roughness is determined before running whilst the other values would not be available until the equipment had failed and therefore could not be incorporated in a design standard. Chapter 3 will provide further evidence of the D ratio and also help to clarify its definition.

(2) From the three results obtained for the hunting tooth drive, namely 2.1a, 2.3a and 2.1b, we may draw a tentative conclusion about the influence of machining orientation. There is some indication from these results that circumferentially ground discs fail before axially ground discs. This statement also has theoretical justification in the work of Christenson and Tonder (1971) and Johnson, Greenwood and Poon (1972) who analysed surface roughness and its influence upon predicted oil film thickness. They showed that surfaces having their asperities orientated in the axial direction cause an increase in the film thickness compared to that predicted for perfectly smooth surfaces, and conversely, circumferentially orientated asperities lead to a decrease in predicted film thickness. These results are not directly applicable as they are derived from relatively large oil films, whereas surface fatigue is encountered in regimes

where the D ratio is relatively large. Nevertheless their general predictions will probably hold true and as the validity of the Dowson and Higginson formula (equation 2.1.2) in this regime is equally questionable it does support the practical results presented here. P.M.Ku (1972) has also carried out work with axially ground discs and has evidence to show that for scuffing experiments.

(a) The axial discs have a greater failure load compared with circumferentially ground discs with equivalent surface roughness.

(b) The tractive forces from results using axially ground discs are very much reduced compared with tests using circumferentially ground discs.

(3) Again using the results from (2) above there is some evidence to suggest that the introduction of a hunting tooth drive promotes earlier pitting. This is contrary to a hypothesis put forward by Shotter (1961); more evidence is required for positive conclusions but the results do warrant further investigation. This point will be discussed in more detail in Chapter 6.

(4) The results marked with arrows, other than 2.5a and 2.2b shown in figures 2.4.3a,b,c and d, were stopped prematurely due to excessive wear, in some cases greater than 2×10^{-2} ins. (500μm) loss of diameter. Because of this wear, crack initiation was probably prevented; alternatively if cracks were initiated propagation may have been prevented by plastic flow in the surface. This is possibly the reason why some discs continued to run long after the predicted failure limit. These results suggest that severe wear has been introduced by using the hunting tooth drive and this implies a new factor which complicates the interpretation of pitting experiments. The mechanism has not been clearly

established; it does seem that it may be dependent upon some form of D ratio. The criterion for this wear seems likely to be dependent on the ratio of the Ra value of the hard disc to the oil film thickness. When this was greater than unity in the tests reported, wear took place. Further information about this mechanism is desirable.

(5) In broad terms the results support the conclusions of Way and Dawson on the mechanism of pit propagation and the roles of surface finish and film thickness.

Although some of the conclusions drawn from this chapter must be tentative the main purpose, which has been achieved, was to provide a basis for comparison with gear tests. The work carried out on gears is discussed in Chapter 3 and, in part, helps to clarify certain conclusions drawn here.

CHAPTER 3

EXPERIMENTAL INVESTIGATION OF PITTING FAILURE USING A LABORATORY GEAR TEST RIG

3.1. Introduction

Chapter 3 is a continuation of the work carried out in Chapter 2 for, at some stage it becomes necessary to determine whether the idealized running conditions of the disc machine are a good representation of results obtained in practice. Obviously the fundamental work discussed would be difficult and costly to carry out in the field, but most of the essentials of gear practice may be reproduced by using a laboratory gear rig, and in this way experimental conditions may be kept under defined limits.

Although there has been considerable work carried out in gear rigs the majority has been in the development field and no direct corroboration of the work by Way, Dawson and others is available in the literature to draw full comparisons. Way (1940) carried out some work on gears and drew similar conclusions to that of his earlier work but this was insufficient to yield as much information as Dawson's results would seem to suggest.

Dowson (1970) presented a paper on the role of lubrication in gear design and emphasised the fact that most of the theoretical and experimental work has been directed to an understanding of the problem and that much of the fundamental work has been established. What is required now is the ability to interpret these results such that use may be made of them in design. Some of this work is already incorporated in gear rating formulae but it is clear from the literature that there is a long way to go. Pitting and scuffing of gears has been discussed by Coleman (1967) in some detail and charts are presented showing the effect of certain controlled parameters, but most of his work discusses contact stress and no results are available for the effect of film thickness and surface

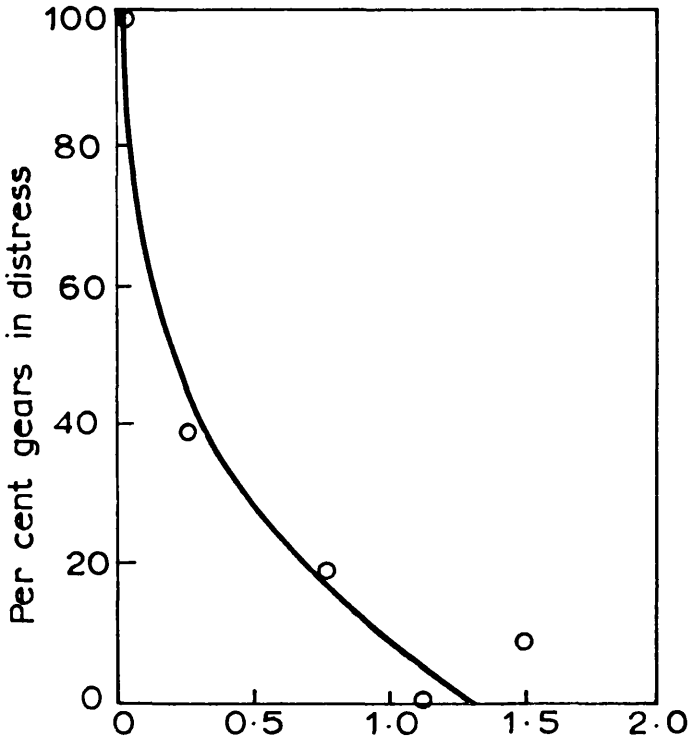


FIG. 3.1.1. Percent gears in distress vs. oil film thickness and surface finish

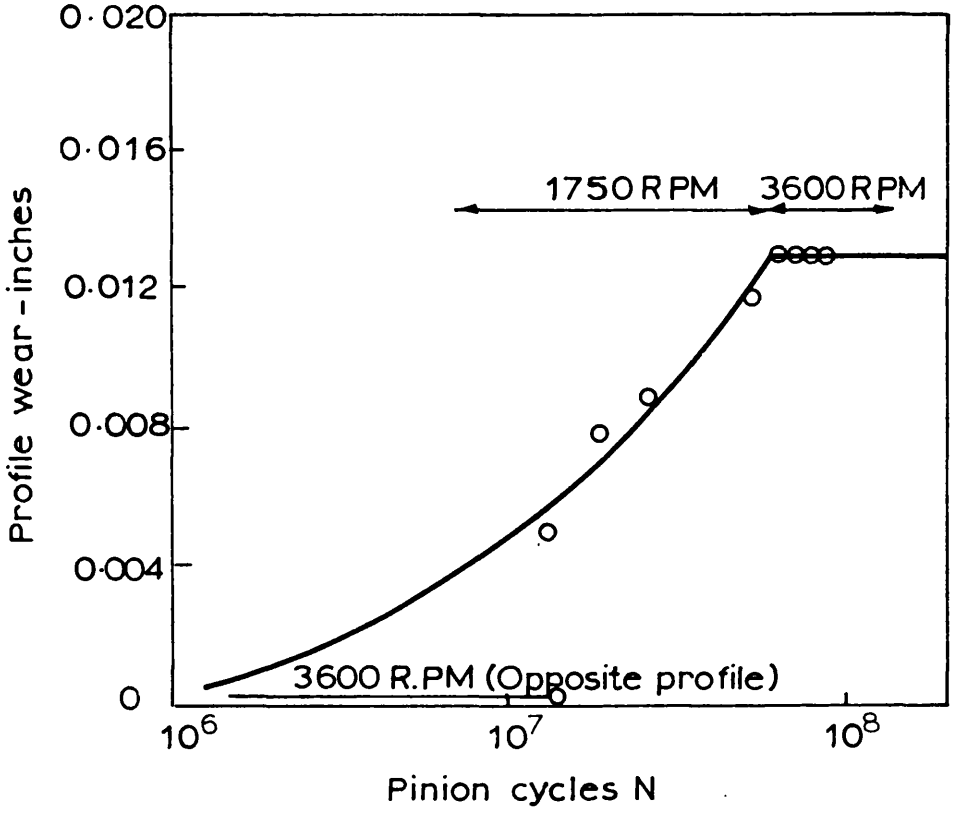


FIG. 3.1.2 Wear reduction due to increased oil film achieved by increase in rotational speed

roughness. With a similar approach Wellauer(1967) outlines the theoretical basis for gear strength and durability rating, and he also suggests areas where improvements may be made which at present are not taken into account by the theoretical equations. The A.G.M.A. rating formulae for durability of spur gears consists of four groups of parameters 1) material elastic properties, 2) load, 3) size and 4) stress distribution. He presents a discussion of the errors and factors that affect the equations but are not taken into account. One such factor which concerns this thesis is the evidence that gears operating under good lubricated conditions and surface finish have increased pitting resistance and that this bears some relationship to the pitch line velocity. Fig. 3.1.1 reproduced here from Wellauer (1967) shows the evidence which suggests that pitting failure of gears is very much dependent on the film thickness/surface roughness ratio and Fig. 3.1.2 shows how the wear due to an inadequate oil film was drastically reduced by increasing the pitch line velocity. Further work, carried out on an F.Z.G gear rig, was that of Niemann, Rettig and Bötch (1964-65) who discussed amongst other things oil temperature, nominal viscosity, surface roughness, speed and oil characteristics. They found that a fine finish was beneficial to fatigue life and certain running-in processes could help in this manner. Further corroboration of the D ratio was given by the fact that increasing the circumferential speed also increased the fatigue life with provisos that the flank error and roughness were small,with large errors they concluded that the effect could be reversed. Other evidence was provided by varying the oil temperature which produced a decrease in life for an increase in temperature and an even greater effect was obtained when the nominal viscosity was varied by the same order as that obtained by varying the temperature. This discrepancy could be partially due to the authors using the oil inlet spray temperature and not that of the gear, nevertheless these three points all have an effect on the film thickness

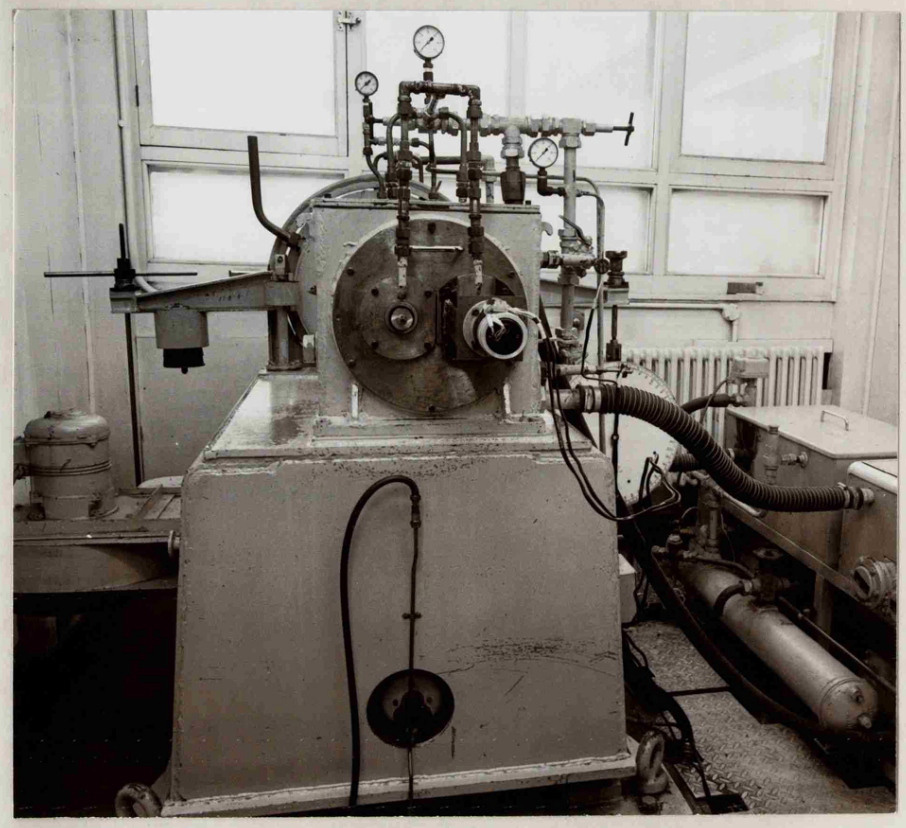


FIG. 3.2.1a

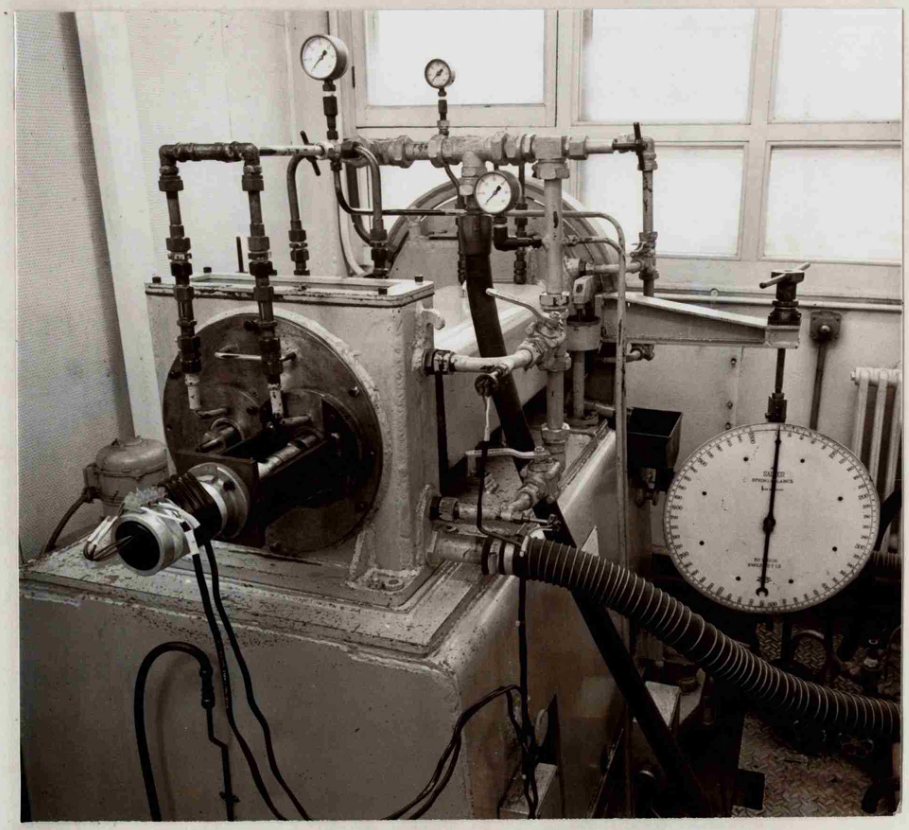


FIG. 3.2.1b

A.E.I. 5" Centres helical gear rig.

generated and therefore the D ratio. Shotter (1961) attempted to bring some order to work reported by Sykes (1959) and supplemented this with results he obtained from a gear rig. The overall conclusion was that drive ratio also plays a part in surface fatigue, the term asperity stress cycles was introduced which, by its name is the number of times one particular asperity may contact a second. He suggested that pitting was initiated by stresses in excess of the fatigue limit and that once all the higher asperities had been reduced, either by wear or fatigue, no further contact would occur and pitting would cease as all other stresses would be below the fatigue limit. A graph of number of pits as a function of asperity stress cycles was produced and gave strong support for this argument.

Most of the other work published in this field has given supporting evidence to that discussed above but in the majority of cases the results have been obtained from in-service equipment where many other factors may play important roles and analytical analysis is impossible.

The work carried out in this chapter was intended to bridge the gap between results obtained in the field, normally incorporating severe complications, and those of laboratory results usually obtained from high precision disc machines where the geometry and the dynamics of the systems are vastly simplified. Using a 5" centres gear rig it was intended to use similar parameter variables as Chapter 2 so that a direct comparison could be drawn from the two methods of investigation.

3.2. Description of Rig

The work of this chapter was carried out on a modified 5" centres Helical gear rig, Fig. 3.2.1 a,b, and designed on the power circulating principal similar to that of the more common I.A.E. $3\frac{1}{4}$" gear rig. Prior to this work the rig had been used for testing relatively large face

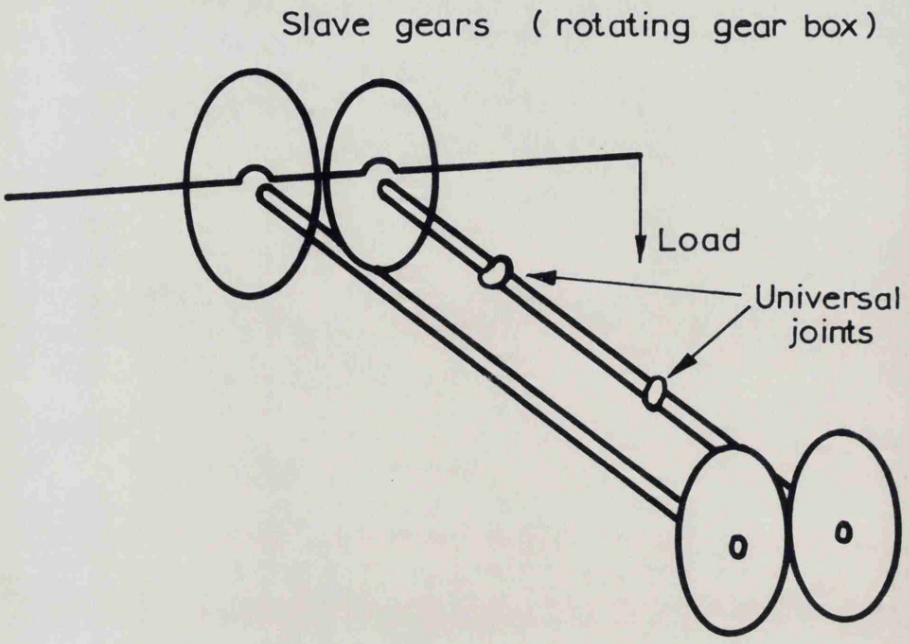


FIG. 3.2.2a

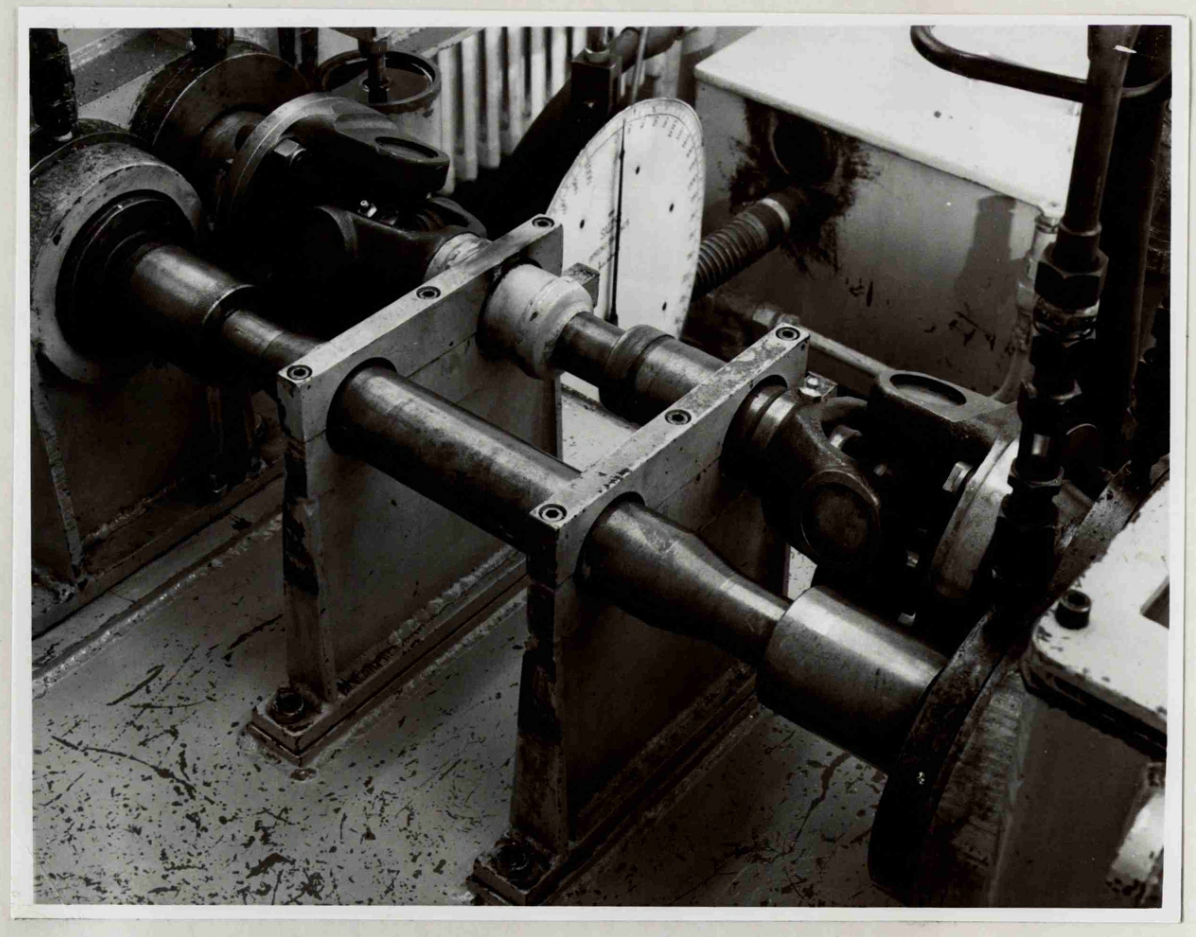


FIG. 3.2.2b

Loading arrangement for gear rig.

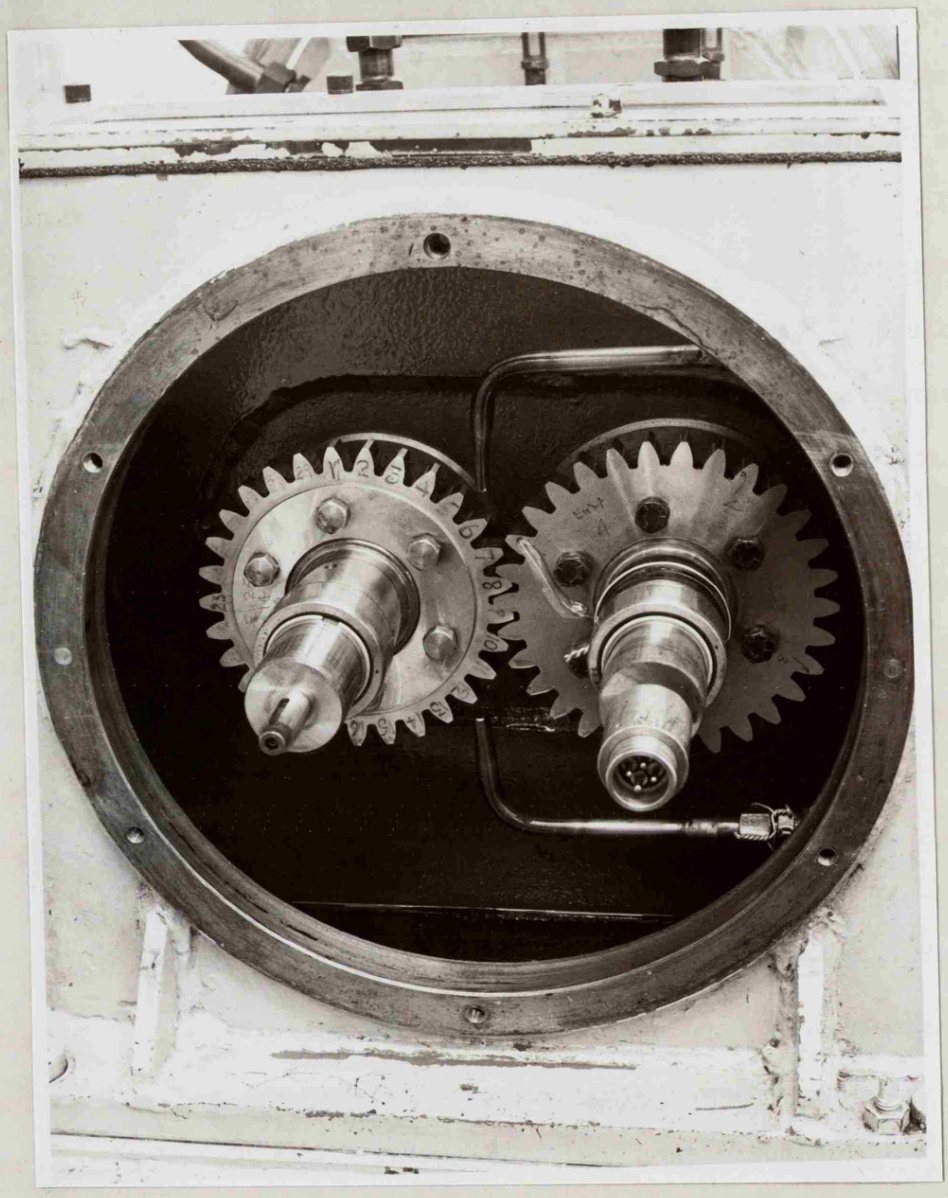


FIG. 3.2.3. Gear mountings and oil supply jets

width gears, the test end using 3" face width and the slave end 4" face width helical gears. The working face width of the gears used here was designed as $\frac{1}{2}$" (12.7mm) to enable the large Hertzian stress of 200×10^3 lbs/ins^2 (1.37×10^6 kN/m^2) to be obtained and consequently, to avoid material wastage new test end shafts were designed so that these gears could be bolted to a flange. The gears were loaded by locking a torque into the system by rotating the slave end gear box about one shaft whilst the second shaft incorporated two universal couplings, Fig. 3.2.2. The gear box was mounted on rollers and rotated clockwise or anticlockwise with the use of a spring balance, Fig. 3.2.1. Prior to loading the slave gear box was rotated away from the centre and the gears attached. When the gears were loaded the shafts were running in the central position. A useful advantage to this system, not attained in most other rigs, was that the load was applied to the gears whilst the rig was running at the specified speed; also any variation of loading, or large vibrations, would be registered on the spring balance.

Two separate oil sumps were used to supply the test and slave gears and in both cases the oil was also used to lubricate the support bearings of the shafts. The sumps were 5 and 10 gallons respectively and the oil in both cases passed through a magnetic filter and also a paper filter. The slave oil was circulated via a hand controlled water cooler. The test oil circulated via a cooler controlled by magnetic valve, which was in turn controlled by a temperature sensing device (Mini Ether) which used the outlet temperature of the cooler (measured by iron/constantan thermocouple) to switch the valve; by this method the oil supply temperature was kept at its required value to within one degree centigrade. The oil in both sumps was heated when necessary by two thermostatically controlled direct heating coils.

The supply of oil was provided by two fan shaped jets one on either entry side of the meshing gears, Fig. 3.2.3, both being permanently open.

	PINION	WHEEL
MATERIAL	EN 25	EN 34
NUMBER OF TEETH	29	30
PITCH CIRCLE DIAMETER	4·8333"	5·000"
BASE CIRCLE DIAMETER	4·5418"	4·6985"
DIAMETRAL PITCH	6	6
PRESSURE ANGLE	20°	20°
OUTSIDE DIAMETER	5·248"	5·418"
SIZE OVER 4 TEETH to give 0·010/0·015 BACKLASH	1·8150 1·8125	1·8173 1·8148

TABLE 3.3.1

Gear design specifications

This maintained symmetry of application and cooling when the direction of drive was reversed. The oil inlet temperature was measured by two iron-constantan thermocouples and recorded on a Honeywell chart, also recorded were the temperatures of the oil outlet from the coolers of both test and slave ends, and two temperatures of the En 34 gear obtained from embedded thermocouples below the surface of the back and front faces of two teeth.

Access to the test gears was permitted by removal of the end plate of the rig (Figure 3.2.3) which also contained the two end bearings of the test shafts. This whole end plate assembly was attached by six bolts, one of which was a locating stud thus permitting the assumption that the shafts and therefore the gears were realigned in the same manner after inspection and reassembly.

3.3. Experimental

3.3.1. Gear Specification

The material specifications of the En 25 and En 34 gear steels were the same as those for the disc materials and can be found in Chapter 2, Table 2.3.1. It was assumed that all physical properties of the steels were the same allowing for the fact that heat treatment was carried out in two separate operations, the hardness measurements made on selected specimens proved to be within the limits of scatter of results obtained for one batch.

The test specimens were ordinary spur gears with a small amount of tip relief, further details are given in Table 3.3.1. In a similar fashion as in the disc machine experiments, the En 34 gears had a greater face width to eliminate axial alignment problems but unlike the discs the En 25 gear did not have the 30° chamfer that helped to relieve edge stresses. The initial order for gears was 12 En 34 and 75 En 25 with

varying grades of roughness which, it was hoped, would be possible to group into 3 grades, those being 10 ± 2μ" Ra, 30 ± 5μ" Ra and 50 ± 10μ" Ra. Twelve only of the En 34 gears were ordered for two reasons, firstly it was hoped and later verified that due to their much greater hardness the surface roughness value Ra would not alter to any great extent and secondly this procedure reduced the cost of manufacturing, which would otherwise have been excessive. A decision was taken to use a hunting tooth ratio of 29 teeth for the pitting gears and 30 teeth for the En 34 mating gears. Basically the argument for the hunting tooth drive was that the errors would be shared by all teeth, whereas for a one to one ratio one particular tooth with an error would always mate with the same opposing tooth so that the propensity to pit in this situation may well have been affected, and scatter of results enhanced. Using the hunting tooth ratio it was thought that these errors would be distributed thus providing uniform pitting over the complete gear, making it possible to select only 4 teeth to be examined on each gear.

3.3.2. Rig Instrumentation

As with the surface finish it was intended to have 3 controlled bulk temperatures for the running conditions of the gears, which were measured by iron-constantan thermocouples embedded below the surface of the teeth. Two thermocouples were used in each test, one embedded below the forward face of a tooth and the second in the back face of a different tooth approximately 90° apart; as the gears were run in both directions two thermocouples were necessary and this would also give an approximate value of the temperature gradient across the teeth. The gears of En 34 were chosen to take these thermocouples as preliminary results using the En 25 gears showed a tendency for the surface to break away where the hole had been machined, suggesting that there was weakening of the material and

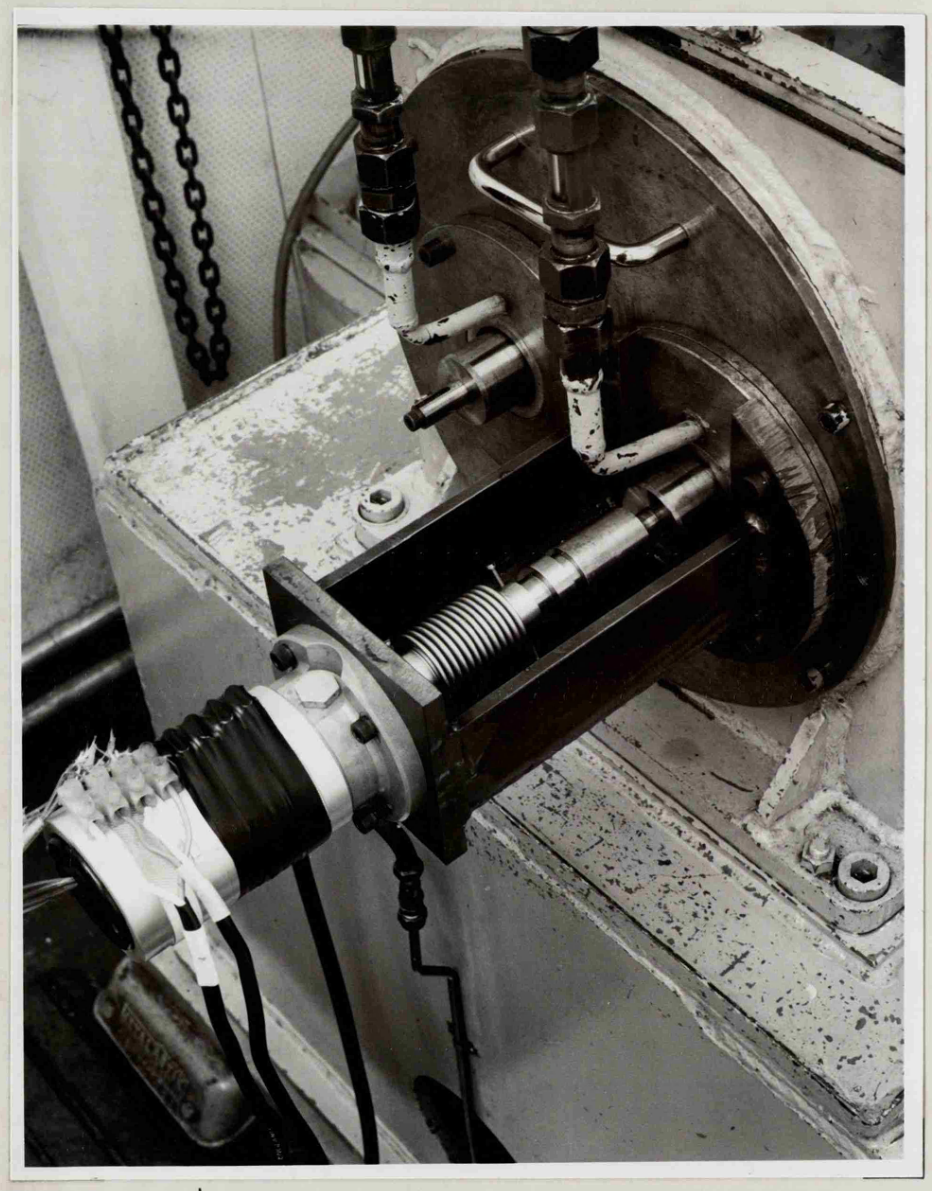


FIG. 3.3.1 Slip-ring assembly and mounting

interference with the sub-surface stresses, which was not acceptable for this form of testing. A secondary reason for measuring the En 34 gears was that of convenience, the test gear had to be removed from the shaft for inspection purposes and this would have resulted in a tedious job of disconnecting the thermocouple wires. The thermocouples consisted of iron-constantan wires welded together into a copper pellet, this was then inserted into a 0.086" diameter hole (whose centre was 0.125in(3.2mm) below the surface) with the aid of a hollow tool through which the wires were threaded, the tool was then given a sharp knock. This action peened the copper pellet and caused the expansion to make a firm hold, thus giving good thermal conductivity between the copper and steel, the remainder of the hole was then filled with Araldite preventing the leads from abrading on the edges.

The e.m.f generated was brought from the rotating shafts via an eight pole carbon brush slip-ring. To enable easy dismantling a four pin male-female plug was incorporated in the slip-ring assembly, the thermocouple wires from the gears came through the hollow shaft and terminated at the male side of the plug, Fig. 3.2.3. The female plug was attached to bellows to allow for minor eccentricity of the slip-ring assembly (Fig. 3.3.1), which was rigidly mounted onto the end plate of the test end gear box. The e.m.f was taken to a Honeywell chart recorder which gave direct readings in degrees centigrade. The bulk temperature of the running gears was normally controllable to within about 5°C by previously setting the inlet oil temperature to a known value determined by experience.

The speed of rotation of the shafts was measured by sensing reflected light with the aid of a photoelectric cell. The reflected signals were fed to a Racal Digital Frequency meter which displayed directly one tenth of the actual speed of rotation of the test gear shaft.

The remaining instrumentation on the rig consisted of safety cut-outs for loss of water pressure, oil pressure, or overheating and once

set these arrangements required no further attention.

3.3.3. Test Conditions

When the completed batch of gears was received it was not possible to group the finishes as initially hoped. The value of Ra ranged from 15 to 60μ" (0.33 - 1.5μm) and combinations throughout the range were used in a sequence that would allow the extreme and intermediate conditions to be covered. The other parameters to be varied were running speed, temperature and the oil, using the same approach as Chapter 2. The rig speeds used for the pinion were 500, 1200 and 3000 r.p.m. which gave pitch line surface velocities of 220, 530 and 1320ft/min respectively. The temperatures of the oils were chosen to be 30°C, 45°C and 70°C but in fact, as will be discussed later, these were not strictly adhered to in the tests. The oils used were the same as those of the disc machine work and the physical properties are given in Table 2.3.3. Various combinations of these variables were used for each test in such a way as to enable a wide variety of cross-over between running conditions and physical properties, resulting in a range of values of D where similar values were obtained by different combinations.

3.3.4. Running Procedure

Both gears were thoroughly cleaned in a light paraffinic solvent and allowed to dry. Talysurf traces of four selected gear teeth on both En 25 and En 34 specimens were taken and the gears mounted on their shafts. The oil was circulated and allowed to reach a steady temperature, a further half hour was then allowed and by this time the gears and test box were assumed to have reached a steady state. The rig was started with no load applied and brought to the correct speed, the gears were then allowed to

run for approximately one minute after which the load was applied steadily and continuously up to 330lbs, this plus the weight of the spring balance gave a maximum Hertzian stress of 200 x 10^3 lbs/in^2 (1.37 x 10^6 kN/m^2). Initially the rig was stopped every fifteen minutes by first unloading and then stopping the motor, the gears were viewed through an observation window and the rig was restarted with the same loading procedure. This practice was continued until experience was gained and it was possible to predict approximately when the gears would be about to fail. As mentioned earlier it was intended to examine only four teeth but owing to the large amount of scatter per tooth this was not possible and the front plate of the test gear box had to be dismantled, the gears removed from the shafts and each tooth examined. Only the En 25 pitting gears were removed and although a hunting tooth ratio was used it was convenient to note the exact location of the gear before removal. There was no necessity to remove the En 34 gear as surface fatigue did not occur on this specimen and previous tests had shown that the surface roughness remained very nearly constant. On removal of the En 25 gear it was again cleaned in the light paraffinic solvent and each tooth examined, the number of pits per tooth were noted in a table along with the length of run. Owing to the lack of time available the surface roughness was not measured after each interruption of the test and only final values, for both gears, were taken. The gear was replaced in its original position on the shaft, the rig assembled, the oil allowed to come to temperature and the previously described procedure was again followed. The test was considered complete when 5 or 6 values of the number of pits for number of cycles run was obtained, this gave enough information to plot total cycles run against total pits and allowed for extrapolation to determine cycles to first pit.

3.4. Results

3.4.1. Preliminary results of surface finish

Before the experimental work could be started it was necessary to examine a sample of both gear specimens to determine whether the stated surface finish requirements could, and had been met, by the manufacturer.

On receipt of three pairs of gears covering the range of finish required they were examined by Talysurf. Specifically the rough surface was of interest because the form of specification adopted was not that generally used; industrial specifications are more often a qualitative definition such as good or medium finish etc. and no reliable figures were available as to what values of Ra corresponded to these specifications.

The gears were set up in a jig and four Talysurf traces, axially displaced, were taken from root to tip using a radius arm attachment. This procedure was carried out for four marked faces of different teeth, two sets on forward running faces and two sets on the back faces. It was immediately noticed that the value of Ra varied considerably for the four traces taken on the rough gears, closer visual examination of the teeth revealed a symmetrical set of bands which resembled a herring bone pattern and, in the majority of cases, consisted of four such bands but occasionally the pattern consisted of eight. As the gear manufacturers were awaiting the results it was not possible to develop an elaborate technique for further examination by Talysurf and the existing gear holder had to be used. This provided movement of the gear in a perpendicular direction to the Talysurf head, thus enabling parallel traces to be obtained with approximately equal displacements. About twenty such traces were taken along the gear tooth faces from root to tip and the Ra values were plotted as a function of the order in which

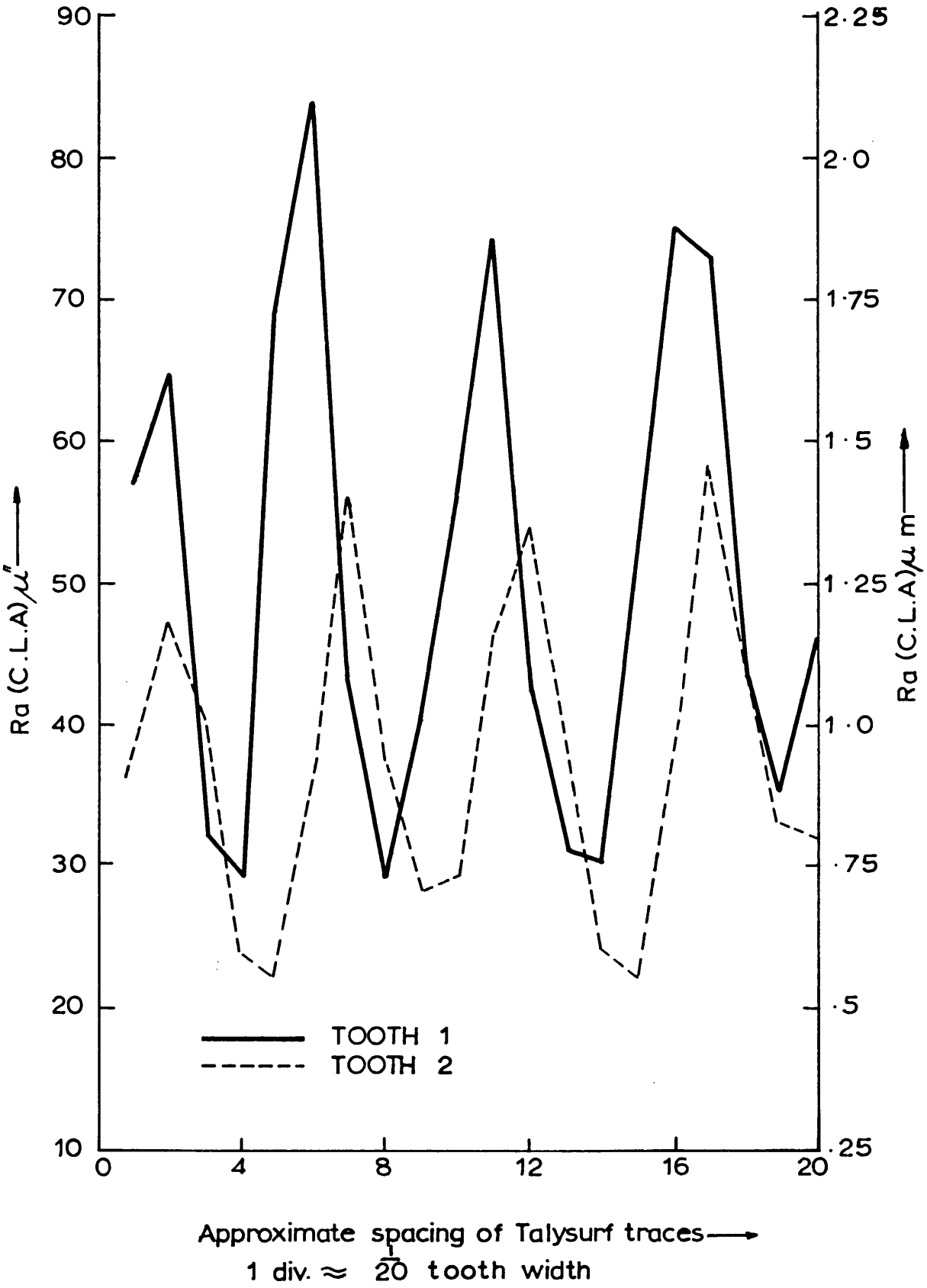
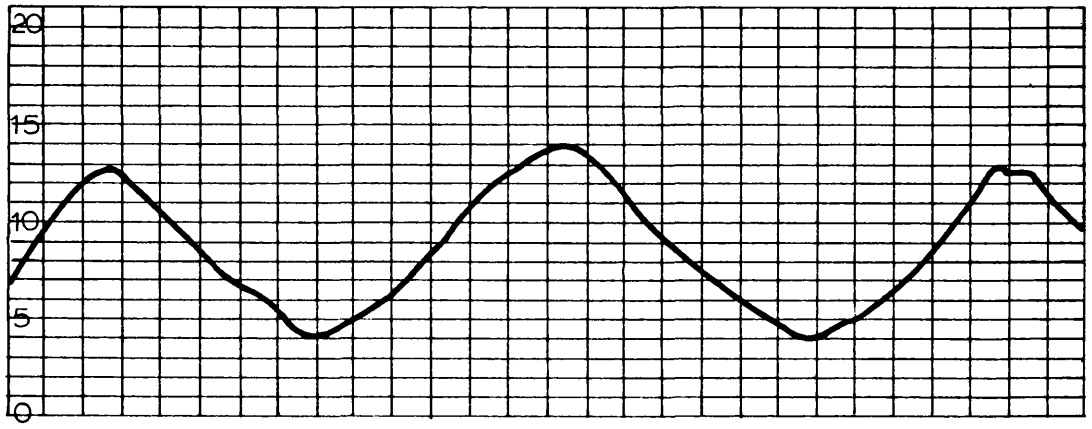
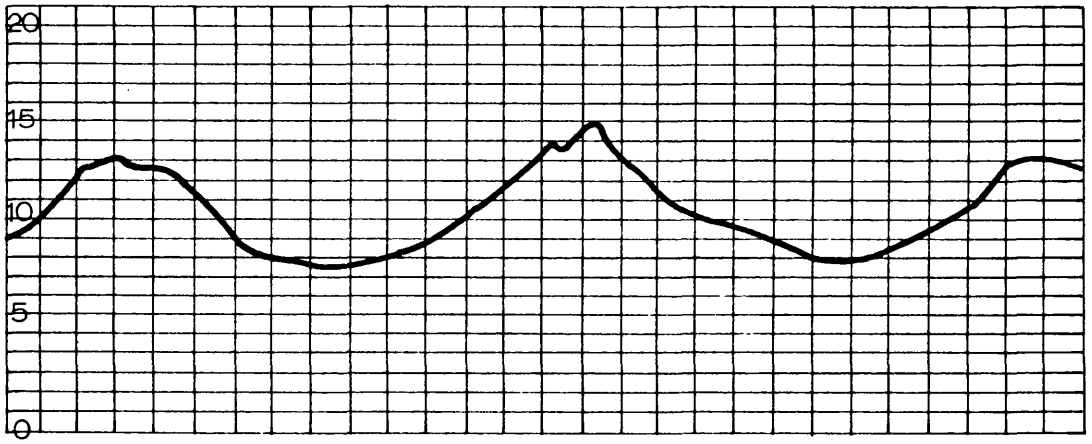


FIG. 3·4·1. Variation of Ra with spacing across teeth and also between individual teeth



TOOTH 1

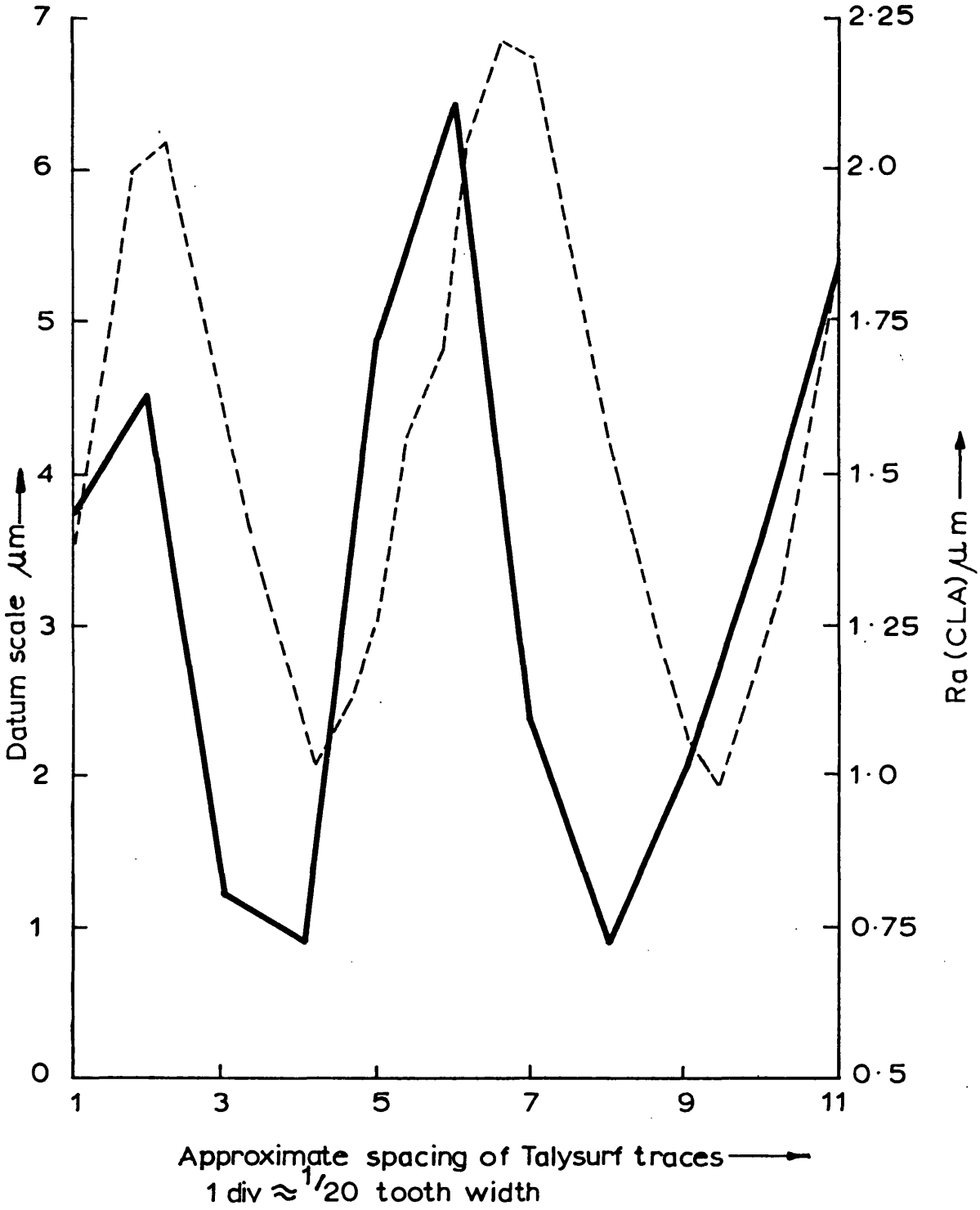


TOOTH 2

Talysurf traces across gear teeth using chatter arm

Magnification Horizontal x 20
Vertical x 5000

FIG. 3.4.2



Variation of Ra across tooth face

Replot of Talysurf trace across tooth face using arbitary datum

FIG. 3.4.3

Gear test results

Test	En 25 R.P.M	Temp °C	Film thickness μm	Surface roughness μm				Oil
				σ_1	σ_2	σ_3	σ_4	
1	1200	72	1.09	0.533	0.510	0.533	0.510	A
2	1200	72	1.09	0.510	0.480	0.457	0.510	A
3	1200	60	1.60	1.52	0.510	1.14	0.510	A
4	1200	60	1.60	0.510	0.510	0.406	0.510	A
5	1200	70	1.12	0.510	0.480	0.510	0.510	A
6	1200	72	1.09	0.510	0.510	0.432	0.510	A
7	1200	41	1.12	0.510	0.356	0.456	0.381	B
8	3000	67	0.86	1.70	0.305	0.709	0.381	B
9	3000	53	1.34	0.910	0.356	0.762	0.381	B
10	500	71	0.224	0.381	1.40			B
11	500	70	0.224	0.381	1.52			B
12	500	Thermocouples		0.330	0.987			B
13	500	broken		0.330	1.32			B
14	500	68	0.236	1.32	0.406	0.304	0.432	B
15	1200	50	1.47	1.06	0.356	0.533	0.381	B
16	1200	42	1.12	0.810	0.330	0.510	0.381	B
17	500	71	0.224	1.09	0.330	0.406	0.381	B
18	500	31	0.936	0.709	0.330	0.835	0.381	B
19	3000	44	6.34	0.356	0.356	1.19	0.406	A
20	3000	64	2.54	0.406	0.381	0.406	0.457	A
21	3000	45	5.83	0.356	0.381	0.356	0.457	A
22	3000	53	4.06	0.658	0.684	0.558	0.709	A
23	3000	58	3.3	0.381	0.684	0.432	0.709	A
24	500	72	0.608	0.406	0.684	0.583	0.709	A

σ_1 Initial surface roughness of En 25 gear

σ_2 Initial surface roughness of En 34 gear

σ_3 Final surface roughness of En 25 gear

σ_4 Final surface roughness of En 34 gear

TABLE 3.4.1

Gear test results

Test	D Ratio				Cycles to pit $\times 10^5$	Rate of pitting $\Delta N \times 10^4$
	D_1	D_2	D_3	D_4		
1	0·96	0·94	0·96	0·94	0·285	0·65
2	0·91	0·88	0·89	0·94	0·360	0·75
3	1·27	0·64	1·03	0·64	0·780	1·28
4	0·64	0·64	0·57	0·64	1·400	1·40
5	0·89	0·86	0·91	0·91	0·730	1·35
6	0·94	0·94	0·86	0·94	0·780	1·70
7	0·78	0·64	0·75	0·68	0·560	1·80
8	2·33	0·71	1·27	0·89	0·167	1·02
9	0·94	0·53	0·85	0·57	0·820	2·09
10						
11						
12						
13	GROSS PLASTIC DEFORMATION					
14	Tooth damaged when dismantled					
15	0·97	0·48	0·62	0·52	1·000	1·50
16	1·02	0·59	0·80	0·67	0·870	1·98
17	6·35	2·96	3·51	3·40	0·500	0·60
18	1·11	0·71	1·30	0·81	0·390	0·66
19	0·11	0·11	0·25	0·13	13·00	7·00
20	0·31	0·30	0·34	0·36	1·6	4·90
21	0·13	0·13	0·14	0·16	>23·0	
22	0·33	0·34	0·31	0·35	2·100	3·50
23	0·32	0·41	0·35	0·43	3·300	5·00
24	1·80	2·25	2·12	2·33	0·590	0·50

$$D_1 = \frac{\sigma_1 + \sigma_2}{h_o}$$

$$D_2 = \frac{2\sigma_2}{h_o}$$

$$D_3 = \frac{\sigma_3 + \sigma_4}{h_o}$$

$$D_4 = \frac{2\sigma_4}{h_o}$$

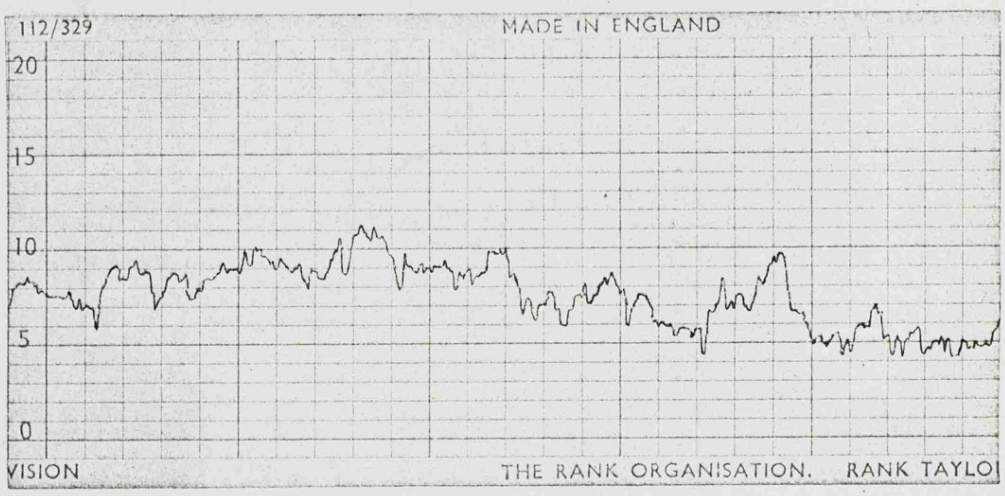
TABLE 3.4.2

they were obtained (Figure 3.4.1). The variation of Ra was clearly emphasised by this method which also showed the variation of surface finish for individual teeth. On the other hand the fine finish gears were much more consistent. The results obtained with the rough finish gears shown in Figure 3.4.1 suggest a sinusoidal pattern and axial traces of the gear teeth were taken using the Chatter Arm attachment which enabled a measure of the macro-roughness to be obtained, samples of such traces are shown in Figure 3.4.2. With this method it was only possible to obtain $\frac{1}{4}$inch traverse which was equivalent to half the face width of the gears, these values have been replotted using an arbitrary datum line and compared with the equivalent Ra values taken from root to tip across the equivalent area of the tooth and results are shown in Fig. 3.4.3 for one tooth. Even with these relatively crude methods there was an obvious correlation between the two results.

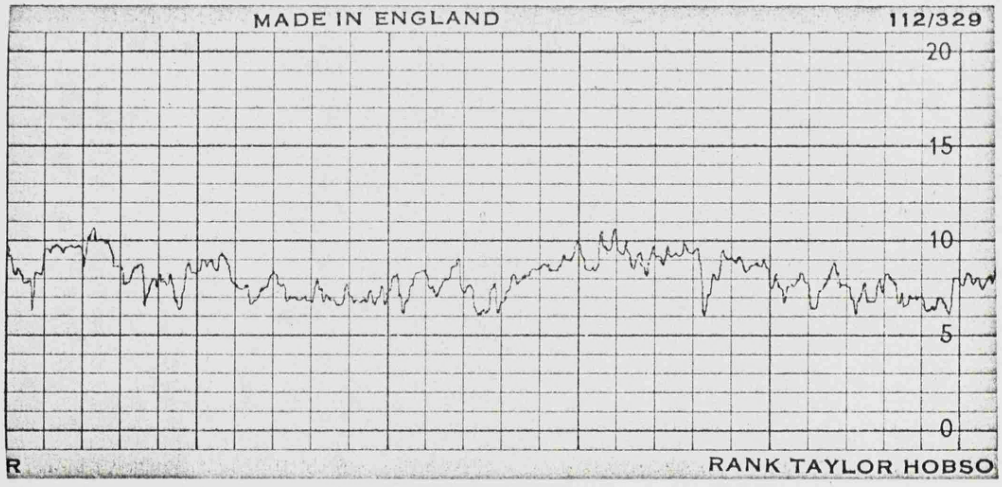
It became clear that for experimental work where surface finish may be of paramount importance these surfaces were unacceptable. Following discussions with the manufacturers it was found that this "herring bone" effect was a common occurrance for normal production gears and was assumed to be a peculiarity of the grinding machine, possibly the out-of-balance of the grinding wheel or stick slip in the carriage saddle of the machine. This suggestion seems to be born out by the fact that the final batch of gears, which were manufactured on a different machine, did not show these errors.

3.4.2. Experimental Results

Tables 3.4.1 and 3.4.2 present a complete record of all results performed on the A.E.I 5" centres Helical Gear Rig and the same procedure as Chapter 2 was adopted for the final analysis. Ra, the measure of their surface roughness was obtained by use of the Talysurf and radius

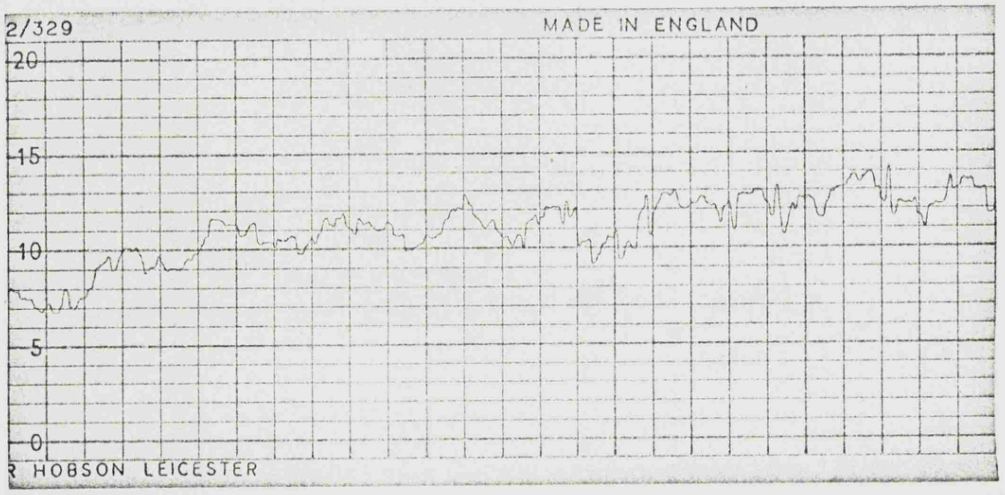


Trace of En 34 gear before running
Vertical magnification x 5,000
Horizontal magnification x 100
Ra = 0·5μm

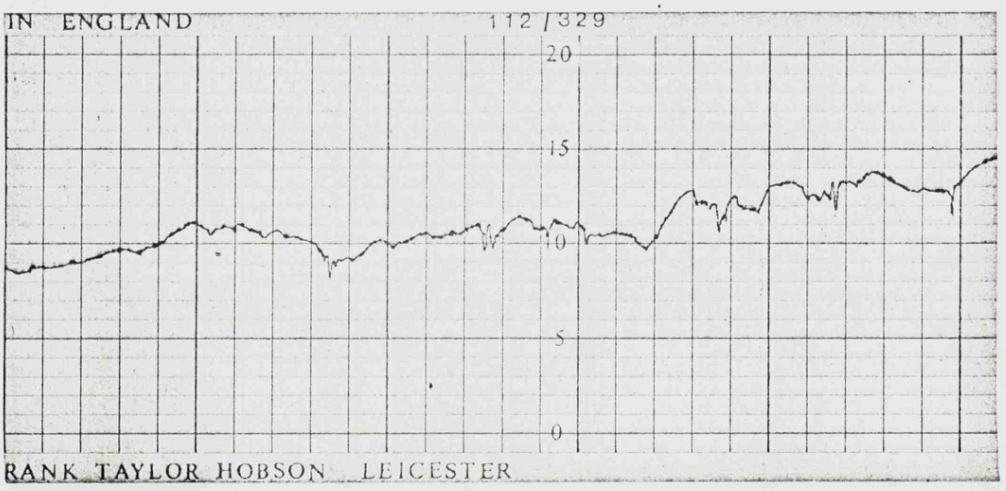


Trace of En 34 gear after $1 \cdot 5 \times 10^5$ cycles
Vertical magnification x 5,000
Horizontal magnification x 100
Ra = 0·5μm

FIG. 3.4.4a



Trace of En 25 gear before running
Vertical magnification x 5,000
Horizontal " x 100
Ra = 0·55 μm



Trace of En 25 gear after $1{\cdot}5 \times 10^5$ cycles
Vertical magnification x 5,000
Horizontal " x 100
Ra = 0·45 μm

FIG. 3.4.4b

Tooth No.	Number of pits on individual teeth during test.									
1	0						1	5	7	9
2	0					1	1	1	2	2
3	0						1	1	2	3
4	0					1	2	4	4	4
5	0			1	1	1	1	1	1	1
6	0						1	1	4	5
7	0					1	1	1	2	3
8	0						1	1	1	2
9	0									
10	0									
11	0							1	3	3
12	0					1	1	2	2	2
13	0			1	1	1	1	1	3	5
14	0					1	1	1	1	3
15	0						1	1	1	1
16	0								1	1
17	0									2
18	0							1	1	1
19	0									
20	0						1	2	3	4
21	0	1	1	1	1	2	3	3	3	4
22	0						1	1	1	1
23	0									
24	0									
25	0								1	1
26	0						1	1	2	4
27	0							1	3	4
28	0								1	2
29	0				1	1	3	3	5	7
Total	0	1	1	3	4	10	22	33	53	75
Cycles run $X10^5$	0·72	1·1	1·4	1·8	2·2	2·5	2·9	3·2	3·6	4·0

History of pitting process

Table 3.4.3

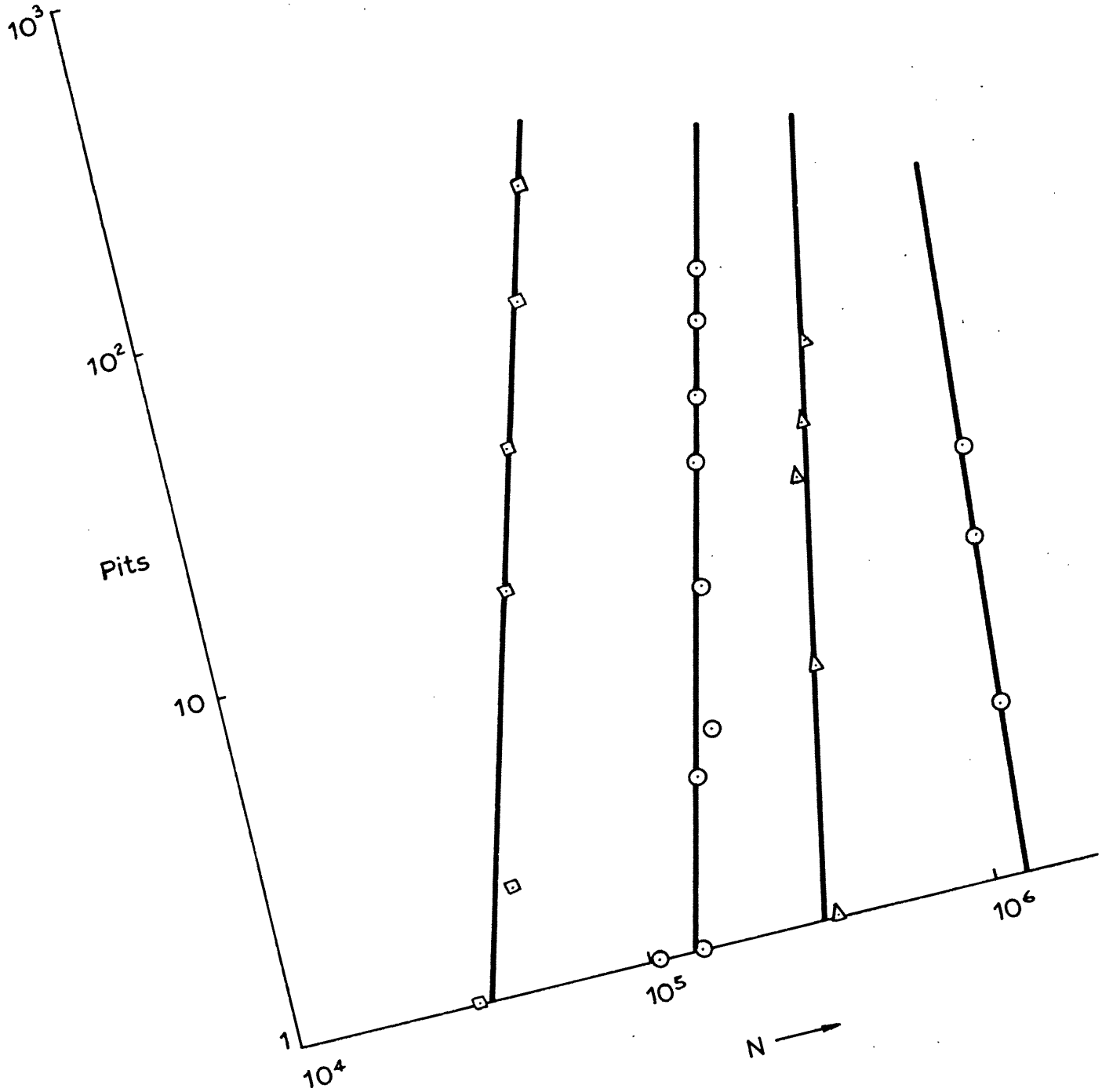


FIG. 3.4.5. Showing number of visible pits as a function of cycles run.

arm attachment, the trace being taken from root to tip along the profile. Sixteen values of surface roughness were obtained for each gear, four readings being taken on each of four faces, two of which were forward and two backward facing. This approach enabled the variation of roughness across the face of the teeth as well as from tooth to tooth to be determined. The working values of Ra were obtained by taking the mean of eight measurements and are reported in Table 3.4.1. Two pairs of Talysurf traces of En 34 and En 25 gears respectively are presented in Fig. 3.4.4 showing the surface detail before and after completion of the test.

Initially it was hoped that results would be obtained from the four teeth that had been Talysurfed but this was not possible, the pitting varied considerably from tooth to tooth and thus the total number of pits visible to the naked eye were counted on all teeth of the gears. The pits themselves were reasonably easy to pick out and no optical aid was necessary. In all tests the pits were found below and around the pitch line being of very similar form to those obtained in the disc machine work i.e. they were of classical shape but rather smaller. The arrow head pointed towards the tooth root and the pitted area was in the form of a band running the full face width, suggesting that errors due to misalignment were small. Table 3.4.3 presents the history of one such gear and allows the total number of pits to be determined for a given number of cycles, such values were plotted on log-log graph paper and the number of cycles required for the first pit was determined by extrapolation. Fig. 3.4.5 presents examples of these plots and it can be seen that in most cases they were straight lines presenting no difficulty in extrapolation.

To enable a value of the D ratio to be determined it was necessary to calculate the oil film thickness and this was obtained by use of the formula given by Dowson and Higginson (1966) (equation 2.1.2) which leads to a predicted value of film thickness at the pitch line.

Inserting the equivalent radii and velocity functions this equation becomes in dimensional terms

$$h = 1.6\alpha^{0.6} (E')^{0.03} \frac{(\ell \sin\psi)^{1.13}}{w^{0.13}} \left(\frac{\eta_o v N_1}{30}\right)^{0.7} \frac{R_g^{1.13}}{(R_g + 1)^{1.56}} \quad (3.4.1)$$

where $R_g = \frac{R_1}{R_2} = \frac{N_2}{N_1}$

N_1 and N_2 the gear and pinion speeds in rev/min.

$\ell = R_1 + R_2$ is the centre distance

R_1 and R_2 are the pitch radii

and ψ the pressure angle.

The temperature used in the calculations was that measured by embedded thermocouples after the system had reached a steady state and the values are tabulated in Table 3.4.1. The temperatures of the gears when running were reasonably controllable and a total variation of 30 to 70°C was achieved which, when the other physical values such as the two oil viscosities and the three speeds were incorporated, gave a range of film thickness at the pitch line of 0.22 to 6μm (10 - 250μ"). This led to a variation of D ratio of around 0.1 to 8 and as in Chapter 2 the four definitions of D were used and their values are given in Table 3.4.2.

Finally in all the tests reported there was no pitting above the pitch line and sectioning of four teeth showed no evidence of fatigue cracks towards the tips. Only a limited amount of work was done on this as discs had already been sectioned and the four teeth examined in this work yielded no information significantly different from that already obtained with discs.

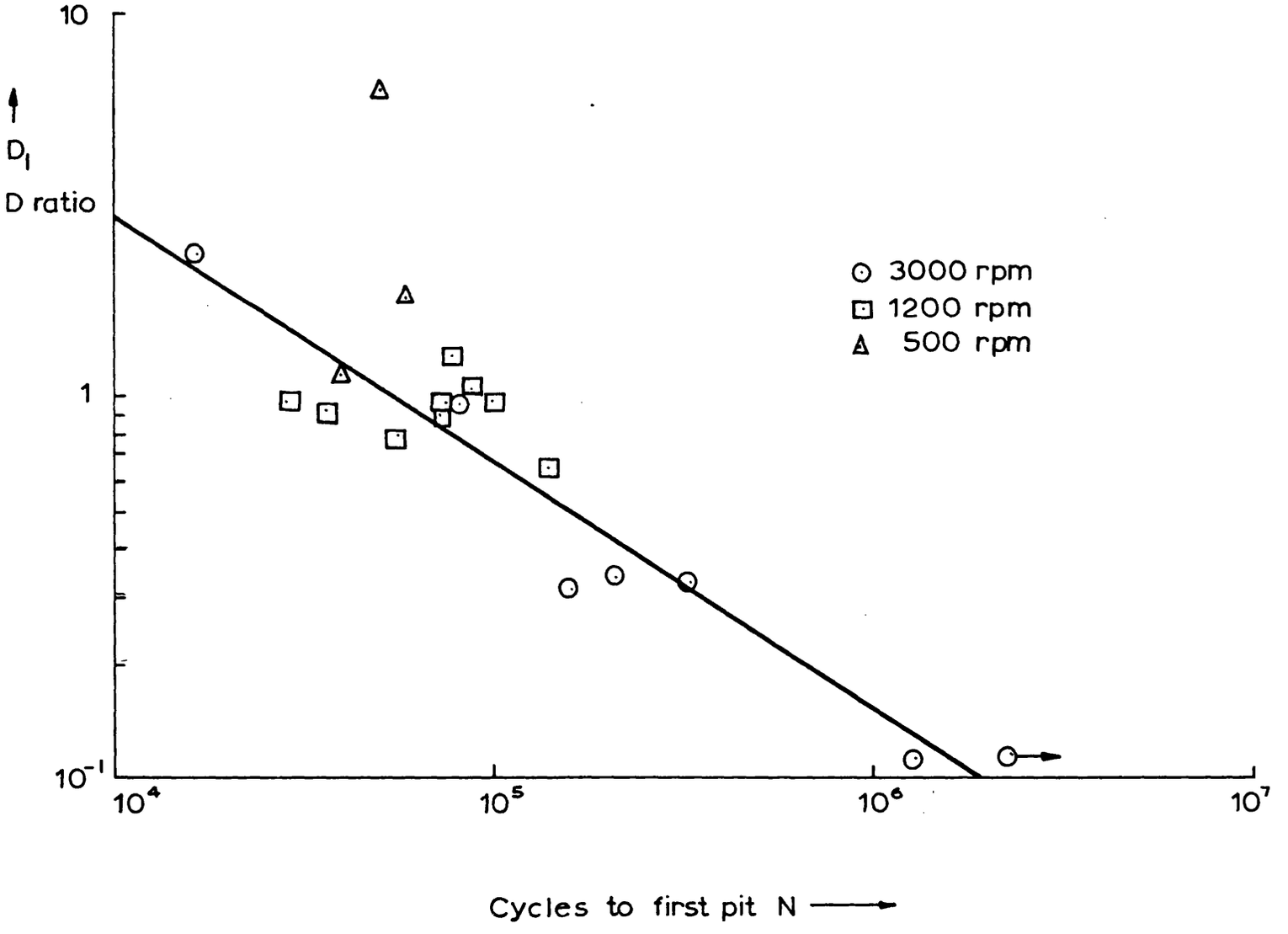


FIG. 3.5.1a. D_1 as a function of cycles to first pit N

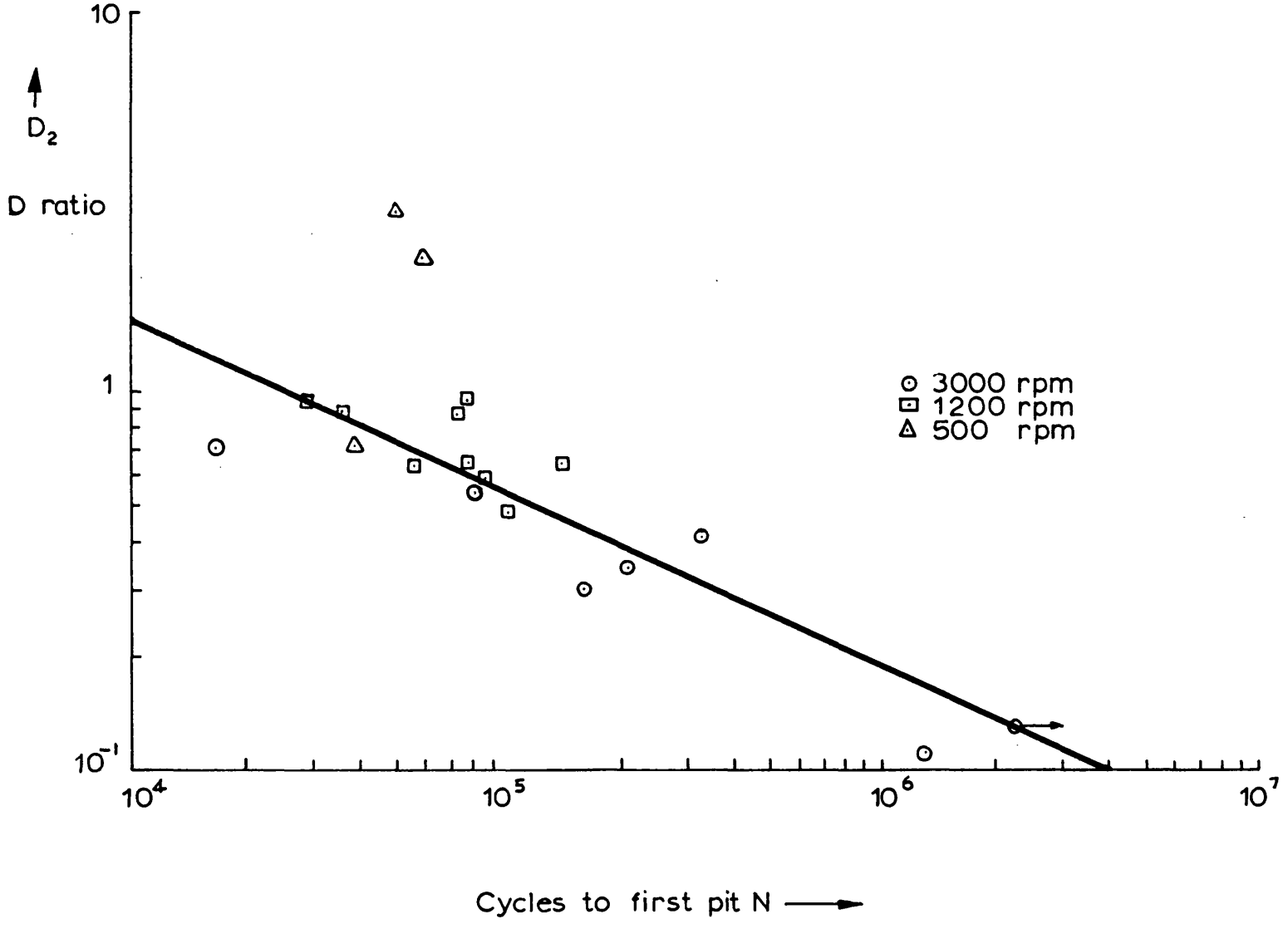


FIG. 3.5.1b D_2 as a function of cycles to first pit N

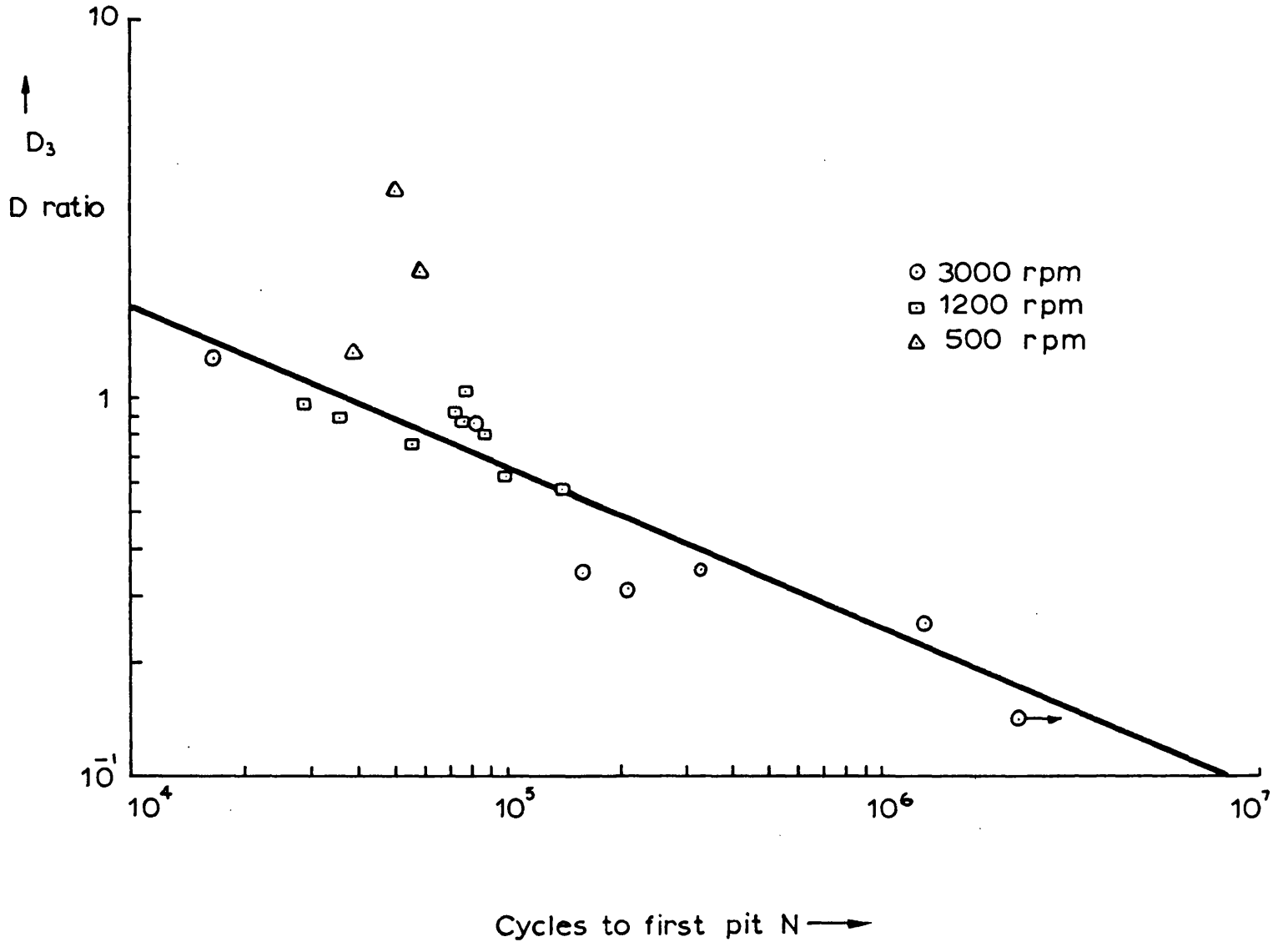


FIG. 3.5.1c. D_3 as a function of cycles to first pit N

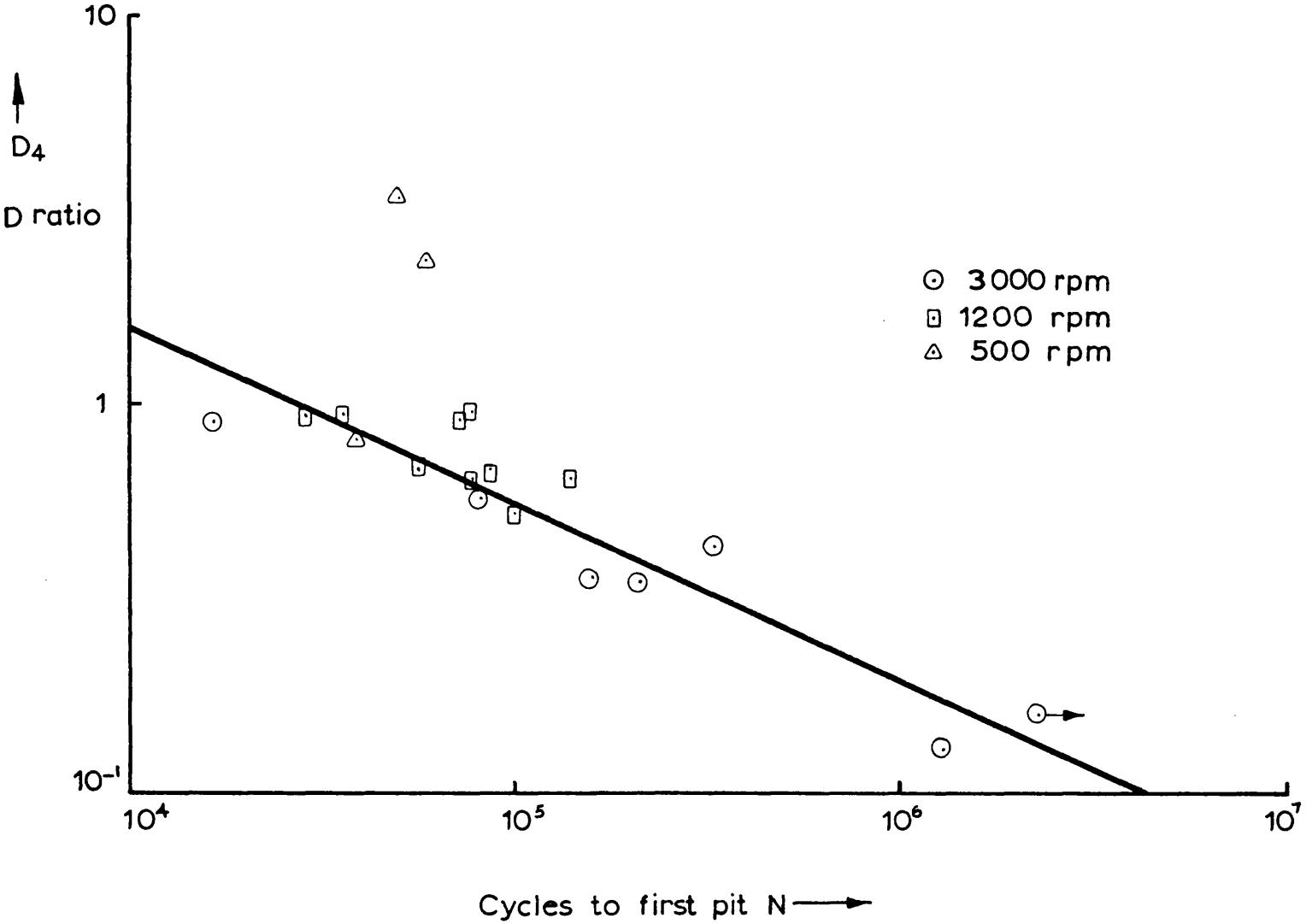


FIG. 3.5.1d. D_4 as a function of cycles to first pit N

Results of regression analysis

$D_1 = (\sigma_1 + \sigma_2)/h$

$\text{Log}_{10} D_1 = -0.643 \log_{10} N + 3.043$

Regression coefficient = 0.913

Standard error of estimate = 0.080

$D_2 = 2\sigma_2/h$

$\text{Log}_{10} D_2 = -0.453 \log_{10} N + 1.987$

Regression coefficient = 0.857

Standard error of estimate = 0.075

$D_3 = (\sigma_3 + \sigma_4)/h$

$\text{Log}_{10} D_3 = -0.427 \log_{10} N + 1.942$

Regression coefficient = 0.902

Standard error of estimate = 0.057

$D_4 = 2\sigma_4/h$

$\text{Log}_{10} D_4 = -0.452 \log_{10} N + 2.008$

Regression coefficient = 0.888

Standard error of estimate = 0.065

TABLE 3.5.1

3.5. Discussion of Gear Tests

The range of variables in this work was greater than that obtained in the experiments of Chapter 2; therefore the results merit closer examination. The graphs of rate of pitting are far more comprehensive and have much less scatter, as can be seen from Fig. 3.4.5. The values of the calculated D ratio were again plotted as a function of cycles to first pit and Figs. 3.5.1a,b,c,d present these results in graphical form. The figures show regression lines determined for results obtained at 3,000 and 1,200 r.p.m. only as the results from the 500 r.p.m. tests tend to deviate from this line quite considerably. This conclusion seems to be in agreement with work carried out by Dawson (1965) who found that for very slow speeds the life to pitting was very much greater than would be predicted by his other results and in fact were almost independent of the D ratio. This conclusion is borne out by the results obtained in this work but unfortunately only three results for 500 r.p.m. could be plotted as Tests 10-13 (which also ran at 500 r.p.m.) showed obvious signs of plastic deformation very early in their lives and were discontinued. It is significant that these were the only results that showed this tendency and clearly there is some factor occurring at lower speeds which is not taken into account by the D ratio. Table 3.5.1 presents the results of the regression analysis carried out for the four definitions of D. Once again there is corroboration of Dawson's work and his rejection of D_1 as a failure criterion; discrimination between the other three definitions is marginal but D_3 would appear to give better correlation and standard errors. Much more work is required to provide further positive information as to the definition of the D ratio. From a practical standpoint the results presented suggest that the D ratio should be defined as twice the initial surface roughness of the harder material divided by the oil film thickness. This would then allow a failure criterion to be determined

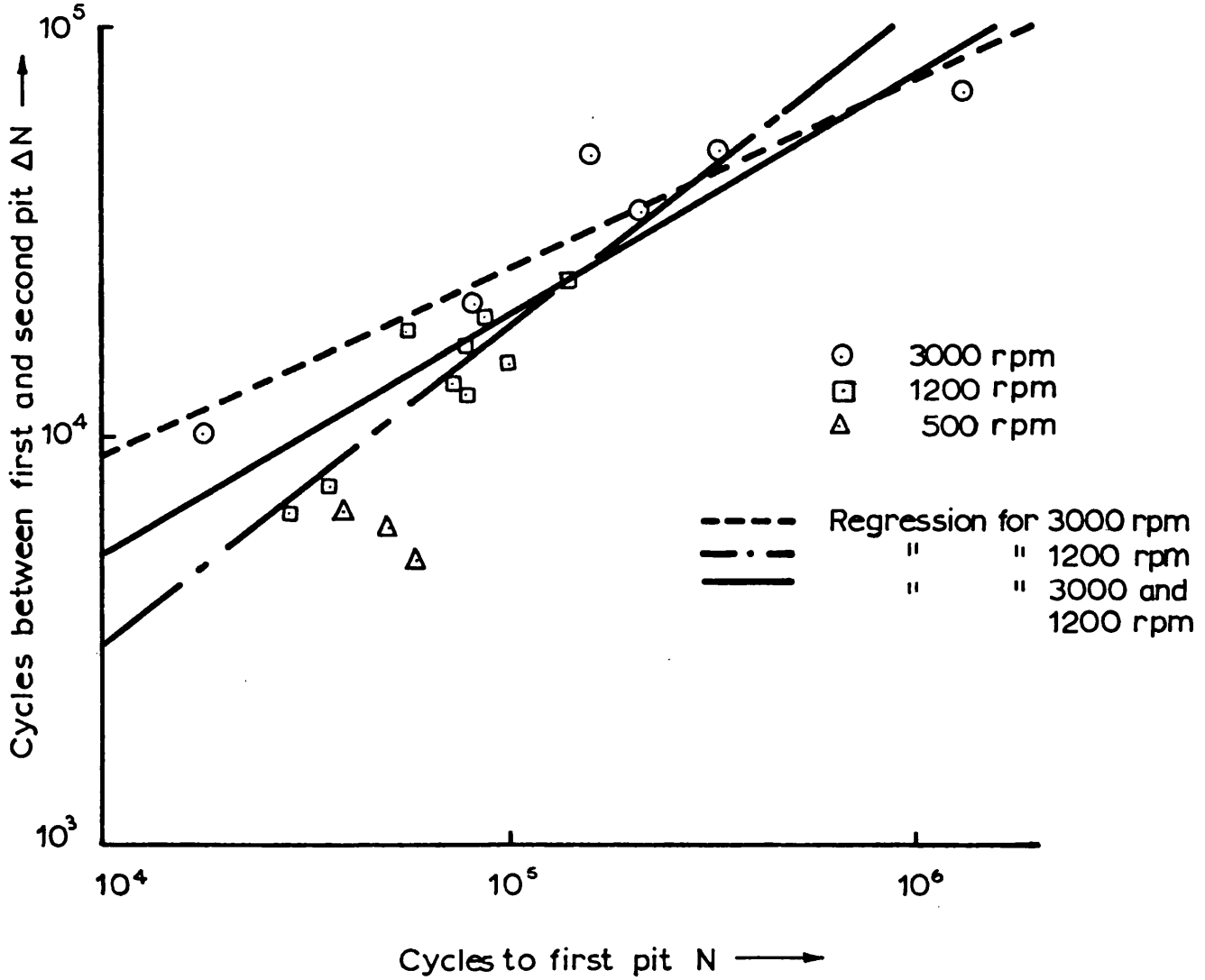

10^5
10^4
10^3
Cycles between first and second pit ΔN
10^4
10^5
10^6
Cycles to first pit N
3000 rpm
1200 rpm
500 rpm
Regression for 3000 rpm
" " 1200 rpm
" " 3000 and 1200 rpm

FIG. 3.5.2.

Results of regression analysis

Regression for 1200 and 3000 rpm combined

$\text{Log}_{10} \Delta N = 0.596 \log_{10} N + 1.322$

Regression coefficient = 0.901

Standard error of estimate = 0.079

Regression for 1200 rpm

$\text{Log}_{10} \Delta N = 0.774 \log_{10} N + 0.401$

Regression coefficient = 0.886

Standard error of estimate = 0.153

Regression for 3000 rpm

$\text{Log}_{10} \Delta N = 0.459 \log_{10} N + 2.122$

Regression coefficient = 0.948

Stanard error of estimate = 0.077

TABLE 3.5.2

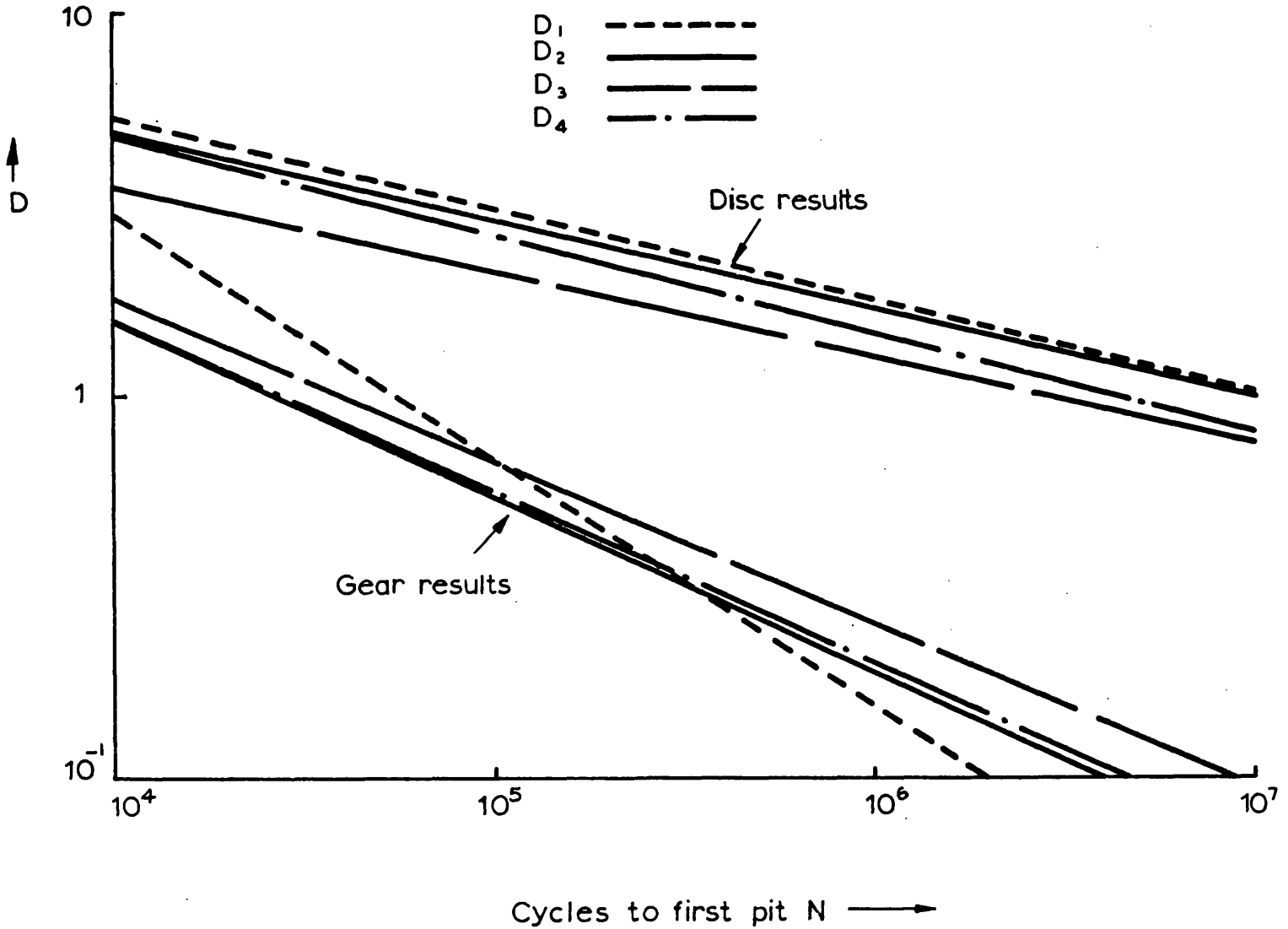


FIG. 3.6.1. Comparison of gear results to disc results

and applied before the failure mechanism has taken place; it would be of no consequence to the development engineer to define a criterion determined after failure has taken place but nevertheless D_3 could well be the criterion required for understanding the phenomenon of pitting fatigue failure.

Table 3.4.2 also provides values of the rate of pitting defined as the number of cycles required between the first and second pits, and these results are presented as a function of cycles required to first pit in Fig. 3.5.2. Again the 500 r.p.m. results depart from the general trend but in addition there seems a possibility that a further speed effect occurs here in that there are significant differences between the 1,200 and 3,000 r.p.m. results. Table 3.5.2 presents the equations for the regression analysis, where the 500 r.p.m. results are again disregarded., The results obtained for 1,200 and 3,000 r.p.m. combined were acceptable but visual observation suggested that the 3,000 r.p.m. results diverged from those of the 1,200 r.p.m. and therefore an analysis was carried out for the individual speeds. The results for 3,000 r.p.m. are comparable with the combined results but those obtained at 1,200 r.p.m. are of much lower statistical significance; this may be explained by the fact that the results obtained at 1,200 r.p.m. do not cover a wide enough range of N. There is a need for more results at 1,200 r.p.m. giving a greater range of N and thus allowing a more thorough analysis of the influence of speed.

3.6. Comparison of pitting tests with gears and discs

Fig. 3.6.1 represents on a single graph the variation of the number of cycles to pit (N) as a function of the D ratio for both gears and discs. Each set of results has been discussed in earlier sections of this thesis so that here, for the sake of clarity, the experimental

points have been omitted and the graphs merely present the regression lines from the statistical analysis. The figure shows N as a function of D_1 D_2 D_3 and D_4. The lines obtained from results using D_1 will be omitted from the comparison of results because, as discussed earlier, they show a markedly lower statistical significance. It can now be seen that the two groups of results are represented by two families of parallel lines. We may therefore treat the comparison in this discussion as that of two single straight lines.

The main features in Figure 3.6.1 which emphasise the difference between pitting in gears and discs are as follows.

a) The pitting life of gears is much shorter than the pitting life of discs, the basis of comparison being results obtained at the same value of the D ratio. As an example, the figure shows that a forecast based upon disc machine tests at a D ratio of unity, would overestimate the pitting life of gears by a factor of approximately one hundred.

b) The difference between the behaviour of discs and gears becomes progressively smaller as the D ratio is increased; as Fig. 3.6.1 shows, this arises from the fact that the pitting life of discs is even more critically dependent upon the D ratio than is the pitting life of gears.

Over the years differing views have been expressed upon the merits of the simulation of gear behaviour by disc machines or other laboratory equipment. A fairly comprehensive insight to this topic may be obtained by referring to two papers by Waight (1964, 1968). It is appropriate to confine this discussion to the context of the tests involved in this work and to seek the reasons for the divergence between the pitting behaviour of discs and gears.

Let us consider first the possible influence of the relative radius of curvature. For the tests in this thesis the discs had a value of 0.75" (16.5mm) and the gears 0.43" (10.5mm). Dawson (1965) provides some insight to this problem, using values of 1.2" (30mm) and 10" (254mm)

he found that, for the same values of the D ratio and maximum Hertzian stress, the increase in life, as determined for one pair of tests, was approximately five. If the small difference in radii of curvatures for the present work is considered in the light of the much larger change cited it seems clear that the divergence cannot be attributed to this cause, It is appropriate to introduce here the fact that Dawson's comparison involved, in a similar manner to the present work, differing drive ratios for the two experimental conditions. It is therefore desirable to give some consideration to the influence of these differing drive ratios.

The bulk of the work in Chapter 2 on disc tests incorporated a 1/1 drive whereas that of the gear rig was 29/30. The final results on the disc machine (Figures 2.4.3) suggested that the use of a hunting tooth reduced the life to failure if no wear took place on the test tracks. This was a tentative conclusion, but nevertheless the reduction in life could affect the comparison of discs and gears. However this reduction due to differing drive ratio was relatively small and it may be concluded therefore, from direct experimental evidence, that a variation in the drive ratio would have only a minor influence on the pitting lives. At this stage, as there is a conflict between the results quoted above and a theoretical approach, it is worthwhile discussing briefly the concept of asperity stress cycles (a.s.c) introduced by Shotter (1961) where

$$\text{a.s.c.} = (N/t) \times \text{time}$$

N being the r.p.m. of the wheel and t the number of teeth in the pinion. It would seem that the assumptions upon which this equation is based are open to doubt, further details are given in Chapter 6. Moreover this factor could not account for the difference in magnitudes of the pitting lives of gears and discs observed in these experiments. If Figure 3.6.1 was re-plotted using 'asperity stress cycles' to replace N, then we

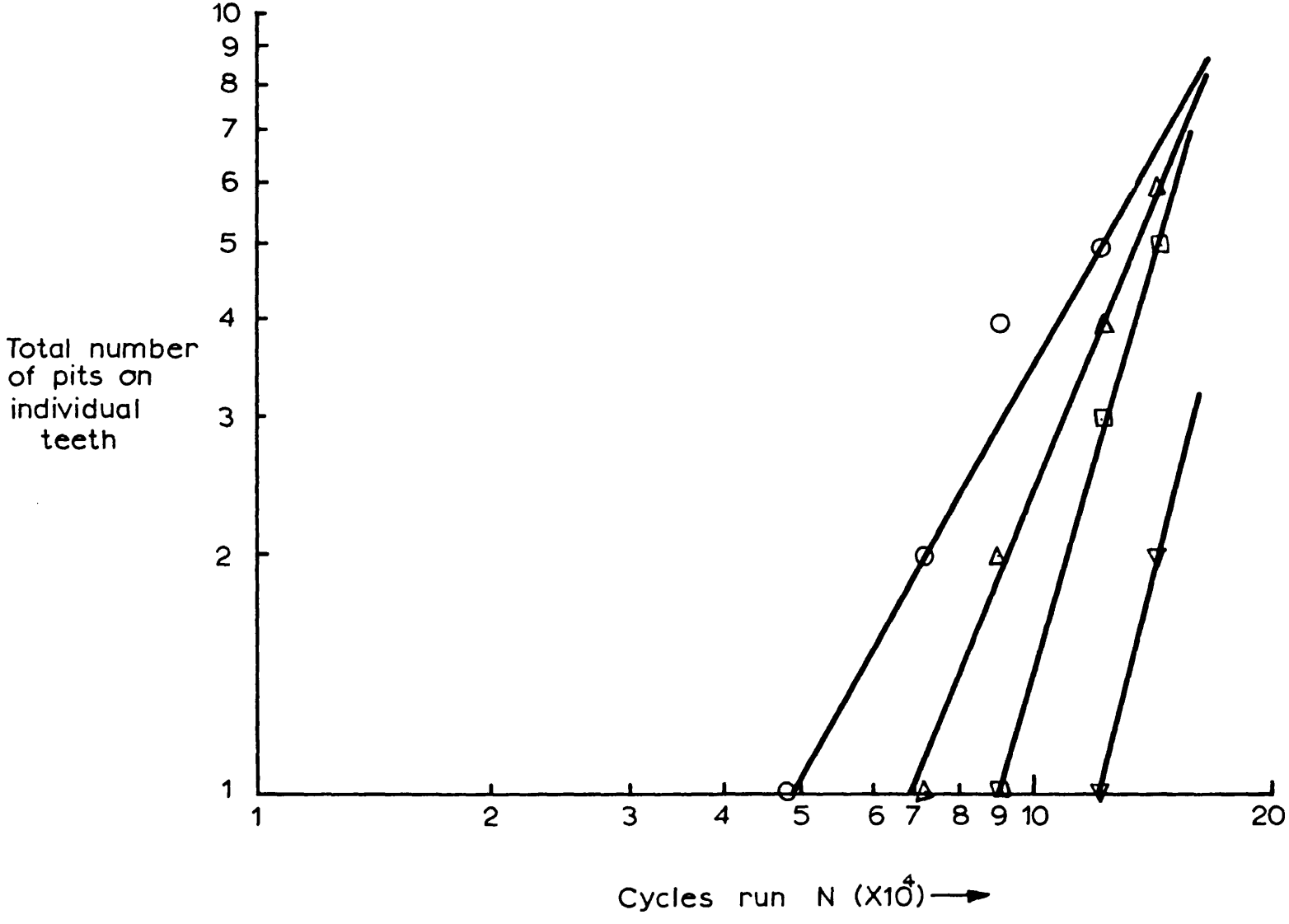


FIG. 3.6.2 Variation of cycles to first pit when using individual teeth from one gear

would find that the disc results would remain constant because for a 1/1 drive N equals 'a.s.c'. However, the results from gears would have a shorter calculated life as N is greater than 'a.s.c' for a hunting tooth ratio and thus the results would show an even greater divergence. In addition, this concept could not take into account the differences in the shapes of the two N versus D graphs.

As has been indicated earlier, there is some evidence that the character, and in particular the orientation, of the machined finish may have an influence upon lubrication and upon failure. For these reasons the orientation of machining in the bulk of the disc tests was made identical with that in the gear tests. It is possible that, because of the methods used, the character of the machined finish obtained for the two types of test differed. Nevertheless the relatively small differences in pitting behaviour produced in disc tests by the fairly radical change from axial to circumferential grinding provides fairly clear evidence that the divergence between discs and gears does not lie in any feature of the machined finish.

One feature of gear tests upon which no comment has been made is the relatively small scatter in the results; this is clearly revealed by the statistical analysis and arises, at least in part, from the fact that the gear test results represent the average behaviour for 29 teeth. This is illustrated by graphs (Fig. 3.6.2) showing the development of pitting with time for 4 individual teeth in a single experiment. The scatter in these graphs is clearly greater than in the corresponding figure for a whole gear shown in Fig. 3.4.5. To what extent do these differences between discs and gears arise from this type of consideration? To make a valid comparison between gear and disc tests we must also take account of the differences of the effective areas, which in practice will not be the total area of tooth contact through the meshing cycle. As an approximation we may take as the effective areas of the gears that which is

susceptible to pitting. This is confined to that area below the pitch line, in the negative slide/sweep region, see for example Fig. 2.1.2. It is also known that pitting for particular materials, is dependent on some critical Hertzian stress therefore we must neglect the area where tip relief and load sharing reduce markedly the propensity to pitting. Using these assumptions as a basis, the effective area of the gears is about $\frac{1}{4}$ of the total tooth faces and is about three times the disc area.

It is now permissible to make a very simple calculation which should make it possible to show whether these considerations will make any significant contribution to explaining the differences between the pitting behaviour in the two types of test. Using Fig. 3.6.1 we can obtain estimates of the pitting lives of discs and gears when operating under identical conditions, i.e. the same D ratio. These values can now be used in conjunction with Fig. 3.4.5., which shows the rate of development of pitting, to estimate the number of pits which would, in principle, exist on the gear surfaces if the gears operated for the pitting life of the discs. The figure arrived at is between 10^8 and 10^9; this is clearly unrealistic, and it becomes clear that the considerations of the last paragraph, therefore, have no significant influence upon this calculation and cannot be used to explain the differences between gears and discs. The argument is emphasised even more strongly if Fig. 3.6.2 is used for the rate of pitting where we now have an effective area of approximately one tenth that of the discs.

This discussion has made it quite clear that none of the factors which have been considered will account for the marked divergence between the pitting behaviour of discs and gears which has been observed in the tests reported here. Therefore it would appear that we must seek some more fundamental difference which is inherent in the nature of gears and discs. In particular, careful consideration must be given to those features of gears which are not reproduced in the simplified conditions of disc tests.

3.7. Discussion

For a given combination of materials (which has been maintained effectively constant in the experiments reported here) the two factors which have the greatest influence on pitting are the film thickness, expressed by the D ratio, and the stress. It is therefore appropriate to discuss each of these factors in turn.

No direct experimental evidence exists to show that the predictions of e.h.l theory apply to the conditions of gear teeth. Results presented by Ibrahaim and Cameron (1963), using the voltage discharge technique, may suggest that discrepancies do exist but there are widespread doubts about the validity of this method (Dyson 1966). Although Ibrahaim and Cameron claim excellent agreement between theory and experiment a closer examination of their results does suggest that they are generally lower than the predicted film thickness, and considerably lower than the experimental results using disc machines and measured by the capacitance technique. Secondly, the deviation from the predicted film thickness is greater for the thicker oil films. If these results are indeed correct then the gradient of the gear results in Fig. 3.6.1 would come more in line with that obtained for the disc results. Therefore there now exists a real need for some more universally acceptable experimental investigation of the lubrication of gears rather than discs. In the first instance it would be quite acceptable if such an investigation used very smooth gear surfaces. Whether this need might be met by the reluctance technique (Cameron and Gregory (1967/68)) is of considerable interest.

In addition there is now widespread interest in the extension of the existing knowledge of e.h.l. to take account of the roughness of the surfaces used in engineering practice. Johnson, Greenwood and Poon (1972) and Tallian (1972) have produced theories of "partial elastohydro-

dynamic lubrication". It has been argued by Archard (1973), that these theories apply only at low values of the D ratio and that at higher values (say D → 1) the whole concept of the conventional e.h.l. theory breaks down because the mean separation of the surfaces is increased above the value required for e.h.l. mechanisms to operate. At higher D ratios micro e.h.l mechanisms may become important. The argument implies that as the D ratio is increased it loses its direct physical significance but could still have an important value as a more empirical parameter; this is because the factors which are used to vary h in the calculated value of D (e.g. η_o and $\bar{U}$) would still have an influence upon the efficacy of the more complex lubrication mechanisms which are operative.

If these concepts are valid it may be significant to note that, as shown in Fig. 3.6.1 all the disc machine experiments are performed in one regime with D > 1 whilst nearly all the gear tests were performed in another regime, with D < 1. As shown in Figs. 3.5.1a,b,c,d the few gear tests performed with D > 1 indicate a departure from the behaviour of the rest of the tests.

Perhaps the major difference between discs and gears is to be found in the dynamic loads which are an inevitable consequence of the kinematics of gearing. Because pitting is critically dependent upon stress it seems conceivable that the differences shown in Fig. 3.6.1 arise from dynamic loading. If this is an important factor one would expect it to become more significant at higher speeds. Several authors have published data on dynamic loading of gear teeth and although they may not be in full agreement, the brief review given by Dudley (1962) suggests that this loading does increase with speed. Thus, for comparable D ratios, pitting life should fall with increasing speed. Now the shapes of the regression lines for the gear results shown in Fig. 3.6.1 are strongly biased by the few results obtained at low D values, all of which were obtained at the highest speed. Therefore an important feature of the dynamic loading is that it could well explain the differences in the

shapes of regression lines for the disc and gear tests. These conclusions gain some support from regression analysis of results for each speed considered alone. The gradients of these regression lines are nearer the values obtained from analysis of the results of disc tests, the pitting lines for 3,000 r.p.m. being less than the corresponding values for 1,200 r.p.m. The larger gradient obtained from the analysis of the combined results (Fig. 3.6.1) is certainly caused, at least in part, by the fact that all results for low D ratios were obtained at 3,000 r.p.m. Thus, within the inevitable limitations imposed by statistical analysis of a relatively small number of tests, it would appear that dynamic loading could provide a satisfactory explanation of the observed differences between gears and discs.

In comparing the results obtained with gears and discs, we have effectively eliminated certain factors as major causes of the discrepancies in behaviour. These factors include relative radius of curvature, character of surface topography, drive ratio, differences in effective area and statistical averaging. It seems likely that the divergence arises from more fundamental differences between the disc and gear tests and these differences are probably associated with the film thickness or with dynamic loads.

Although surface topography has been eliminated as a major factor in the comparison of disc and gear tests it remains true that its influence continues to be an important, and an imperfectly understood, factor in the initiation of pitting. Chapters 4 and 5 are therefore concerned with a more detailed investigation of surface topography.

CHAPTER 4

THE CONTACT OF SURFACES HAVING A RANDOM STRUCTURE

4.1. Introduction

Chapters 2 and 3 have emphasised the important role that surface topography plays in fatigue failure of rolling elements and an understanding of the details of surface contact is of great importance in many such aspects of tribology. This chapter intends to apply a model of random surfaces derived by Whitehouse and Archard (1970) to provide a theory of surface contact.

In earlier theories it was assumed that the true area of contact arose from plastic deformation of asperities (Bowden and Tabor (1954)), but in more recent times it has become more widely accepted that surface contact can often involve an appreciable proportion of contact between asperities which is entirely elastic. In this discussion, a point of critical importance is the relation between the total true area of contact, A, and the load, W, since from this relationship arise the laws of both friction and wear. If it be assumed that the deformation is plastic the true area of contact, A, is proportional to load, W, regardless of the nature of the surface topography and the consequent subdivision of the true area of contact. Archard (1957) first showed that, if multiple contact conditions were assumed, the relationship between area and load could be close to direct proportionality even if it were assumed that the deformation were entirely elastic. The model used consisted of superimposed asperities of differing scales of size; it was shown, however, that the question of physical significance in such theories was the way in which an increase in the load was divided between the enlargement of existing areas and the creation of new areas.

Greenwood and Williamson (1966) have discussed this same question using as their basis the evidence derived from digital analysis of profilometer records. They have shown that the distribution of heights on many surface profiles is close to a Gaussian form. Moreover, it was shown that the distribution of peaks on these same profiles was also Gaussian; the peak distribution was derived using three point analysis, a peak being defined when the central of three successive points in the digital presentation was the highest. Based upon this evidence they used, as a model representative of random surfaces a set of asperities having a Gaussian distribution. Their model was thus defined by three parameters; σ^*, the standard deviation of the asperity peak distribution, R, the radius of curvature of the asperities (assumed constant) and η, the density of asperities per unit area. Using this model Greenwood and Williamson (1966) showed that the relationship between A and W was close to direct proportionality even if it was assumed that the deformation was entirely elastic. They also used the model to calculate the probability of plastic deformation. It was shown that the proportion of the area of contact supported by plastic deformation was largely independent of the load but was critically dependent upon a plasticity index ψ given by

$$\psi = \left(\frac{E'}{H}\right)\left(\frac{\sigma^*}{R}\right)^{\frac{1}{2}} \qquad (4.1.1)$$

where H is the hardness of the material and $E' = E/(1 - \nu^2)$; E is the Young's modulus of the material and ν its Poisson's ratio.

In a more recent paper Whitehouse and Archard (1970) have provided a more rigorous analysis of random surfaces. A profile of such a surface was regarded as a random signal (Peklenik (1967-68)) represented by a height distribution and an autocorrelation function. The particular model used to represent the profile was a Gaussian distribution of heights and an exponential correlation function. It was shown that an important feature of this model was that all features of the surface

topography could be represented by two parameters; σ, the standard deviation of the height distribution, and β^*, the exponent of the exponential correlation function which was called the correlation distance. The characteristics of this type of profile of significance in surface contact (that is, the distributions of the heights and the curvatures of the peaks) derived from the theory were shown to be in good agreement with those derived from digital analysis of a ground surface.

In this chapter the results of Whitehouse and Archard (1970) will be used to derive a theory of surface contact. There are a number of features of this model which make such a theory desirable. First, it gives a distribution of peak curvatures which is dependent upon their height, the influence of this feature upon the relationship between area of contact and load is clearly of some significance in the development of ideas about the nature of surface contact. Second, as explained above, the model defines the surface profile in terms of two parameters; one (σ) is concerned with the height distribution and the other (β^*) can be regarded as being associated with the wavelength-structure. A recent development in the technology of surfaces is a proposal (Spragg and Whitehouse (1971)) for their characterisation in terms of two such parameters; therefore a theory of surface contact based upon similar principles must be of some relevance.

Before proceeding to a development of the theory of contact for such surfaces a brief outline of the model is presented; a more complete discussion is given by Whitehouse and Archard (1970) in their original paper.

4.2. The Model

The profile is defined by a height distribution and an autocorrelation function. The height distribution is Gaussian and is therefore

defined by the probability of finding an ordinate between h and (h + dh)

$$f(y) = (2\pi)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}y^2), \qquad (4.2.1)$$

where the height h has been expressed in the normalized form

$$y = h/\sigma \qquad (4.2.2)$$

Similarly the autocorrelation function is defined as

$$\rho(\beta) = \frac{1}{L} \int_{-\frac{1}{2}L}^{+\frac{1}{2}L} y(x) \cdot y(x+\beta)dx \qquad (4.2.3)$$

where y(x) is the height at a given coordinate x and $y(x + \beta)$ is the height at an adjacent coordinate $(x + \beta)$. In the model it is assumed that

$$\rho(\beta) = \exp(-\beta/\beta^*) \qquad (4.2.4)$$

where β^* is the correlation distance and is a measure of the rate at which the correlation between events on the profile declines as their spacing increases. When $\beta = 2.3\beta^*$ the correlation has declined to 0.1 and we shall take this spacing as that at which events on the profile have just reached the condition where they can be regarded as independent.

Using this model it is possible to calculate the characteristics of the profile which are derived from digital analysis. It is assumed that the profile is presented as a series of events separated by a sampling interval ℓ; i.e. there exists a correlation $\rho(= \exp(-\ell/\beta^*))$ between successive events. It can then be shown that the probability density function that any ordinate is a peak of curvature C at a height y is given by

$$f^*(y,C,\rho) = \frac{\exp(-\frac{1}{2}y^2)}{2\pi[2(1-\rho^2)]^{\frac{1}{2}}} \exp\left\{-\frac{[(1-\rho)y-\frac{1}{2}C]^2}{(1-\rho^2)}\right\} \operatorname{erf}\left[\frac{C}{2\sqrt{(1-\rho^2)}}\right] \qquad (4.2.5)$$

For the particular example, $\ell = 2.3\beta^*$, $\rho \to 0$, equation (4.2.5) can be written in the simplified form

$$f^*(y,C) = \frac{\exp(-\frac{1}{2}y^2)}{2\pi\sqrt{2}} \exp\left\{-(y - \tfrac{1}{2}C)^2\right\} \operatorname{erf}(\tfrac{1}{2}C) \quad (4.2.6)$$

In these equations the curvature is presented in non-dimensional form (see Whitehouse and Archard (1970)) and the radius of curvature of a peak is given by

$$R = \ell^2/C\sigma \quad (4.2.7)$$

One important feature of the theory is that equation (4.2.5) reveals that the characteristics of the profile, which emerge from digital analysis, are markedly dependent upon the sampling interval employed (i.e. upon the correlation, ρ, between successive events). This is the inevitable consequence of the fact that the profile consists of the random super-position of structures covering a wide range of wavelengths; the use of any given sampling interval ℓ reveals a structure which is mainly concerned with the wavelengths of this same order of size. It has been argued by Whitehouse and Archard (1970) that the main structure of the profile is best revealed, from three point analysis of a digital presentation, by using a sampling interval where successive events can be regarded as having just reached the situation where they can be taken as independent. This is achieved by putting $\ell = 2.3\beta^*$; then without undue error the simplified probability density function of equation (4.2.6) can be used.

We also need to obtain from the model a value of η, the density of asperities per unit area. Consider here only the case where the profile is sampled at intervals of $\ell = 2.3\beta^*$ and the events are effectively independent. Then, because the events are independent, there is an equal chance that any one (i.e. the central) of the three events is the highest. The ratio of peaks to ordinates in a profile is therefore one third; this conclusion has been fully justified elsewhere (Whitehouse and Archard (1970)). Similarly, we can consider the simplest possible

extension of the theory to three dimensions, the topography being defined by a grid of ordinates arranged in a simple coordinate system. A high point in three dimensions, a summit, is then defined in the simplest possible way when it is higher than its four nearest neighbours. Using the same simple argument as before the ratio of summits to ordinates is therefore one fifth and a simple estimate of the density of asperities per unit area is

$$\eta = \frac{1}{5}\left(\frac{1}{2.3\beta^*}\right)^2 \qquad (4.2.8)$$

A more complete discussion of the problems involved in the extension of the theory to three dimensions is given by Whitehouse (1971).

There are two significant ways in which the model of Whitehouse and Archard (W & A) differs from that of Greenwood and Williamson (G & W). First the distribution of peaks (regardless of their curvature) is not quite Gaussian but follows a distribution derived from the assumed Gaussian distribution of heights; for the main broad scale structure of the surface derived using a sampling interval of $2.3\beta^*$ the probability that any ordinate is a peak at height y is

$$f^*(y) = [1/4\sqrt{2\pi}]\ [1 + \mathrm{erf}(y/\sqrt{2})]^2\ \exp(-\tfrac{1}{2}y^2) \qquad (4.2.9)$$

Second, as explained above, the peak curvatures have a distribution which is dependent upon the height, the curvature of the highest peaks tending to have higher values than those at lower levels. In the theory which follows we shall require the probability density distributions of the peaks. To obtain these, equations (4.2.5) (4.2.6) and (4.2.9) must be normalized by the ratio of peaks to ordinates, N, which is given by

$$N = \frac{1}{\pi}\,\mathrm{Tan}^{-1}\left(\frac{3-\rho}{1+\rho}\right)^{\frac{1}{2}} \qquad (4.2.10)$$

For uncorrelated events (equations (4.2.6) (4.2.9)) N will be taken as $\frac{1}{3}$.

In what follows the results obtained using the W & A model will be

compared with those obtained by Greenwood and Williamson. It will be important to distinguish between the two significant features which have been discussed above, and therefore it is useful to introduce what will be termed an intermediate model; this will have an asperity peak height distribution given by Whitehouse and Archard (equation (4.2.9)) but the asperities will have a constant radius of curvature as in the Greenwood and Williamson model. For this (and when using the G & W model) the mean curvature for all peaks derived from the Whitehouse and Archard work will be used, i.e.

$$\bar{C}^* = (3 - \rho)\,(1 - \rho)^{\frac{1}{2}}/2N\sqrt{\pi} \qquad (4.2.11)$$

where N is the ratio of peaks to ordinates and is given by equation (4.2.10). For the main broad scale structure of the surface we can write $\rho \to 0$ and $N \to \frac{1}{3}$; thus

$$\bar{C}^* = 9/2\sqrt{\pi} \qquad (4.2.12)$$

whilst the radius of curvature is obtained from equation (4.2.7) with $\ell = 2.3\beta^*$.

In deriving the G & W theory in a form suitable for comparison with the W & A model it is appropriate, and most revealing, to choose a Gaussian distribution of asperity heights to give the closest accord with the W & A model. Since contact occurs at the highest asperities a fit is required at large positive values of y. Consideration of the peak distributions shows that this requirement is best met by using for the height distribution of <u>asperities</u>, three times the assumed Gaussian distribution of <u>heights</u>.

The distribution of peak heights derived by W & A is

$$\frac{1}{N} f^*(y) = 3\,\frac{1}{4\sqrt{2\pi}}\,[1 + \mathrm{erf}(y/\sqrt{2})]^2 \exp(-\tfrac{1}{2}y^2) \qquad (4.2.13)$$

$$= \frac{3}{\sqrt{2\pi}}\,\Phi^2(y)\exp(-\tfrac{1}{2}y^2)$$

The cumulative distribution corresponding to equation (4.2.13) is

$$F^*(y) = \frac{1}{N} \int_y^\infty f^*(t)dt$$

where t is a dummy variable for y

$$F^*(y) = \int_{t=y}^{t=\infty} 3\,\Phi^2(t)\,\frac{1}{\sqrt{2\pi}}\exp(-\tfrac{1}{2}t^2)dt$$

$$= [1 - \Phi^3(y)] \qquad (4.2.14)$$

This may be re-written in the convenient form

$$F^*(y) = [1 - \Phi(y)]\ [1 + \Phi(y) + \Phi^2(y)]$$

and since for large positive values of y

$$\Phi(y) \to 1$$

then in this same range of values of y

$$F^*(y) \simeq 3[1 - \Phi(y)] \qquad (4.2.15)$$

The corresponding distribution of <u>ordinate</u> heights from which equation (4.2.13) is derived is

$$f(y) = \frac{1}{\sqrt{2\pi}}\ \exp\ (-\tfrac{1}{2}y^2) \qquad (4.2.16)$$

whilst the cumulative distribution of ordinate heights is

$$F(y) = \int_y^\infty \exp(-\tfrac{1}{2}t^2)dt$$

$$= 1 - \Phi(y) \qquad (4.2.17)$$

Comparison of equations (4.2.15) and (4.2.17) shows that for large positive values of y

$$F^*(y) = 3\,F(y) \qquad (4.2.18)$$

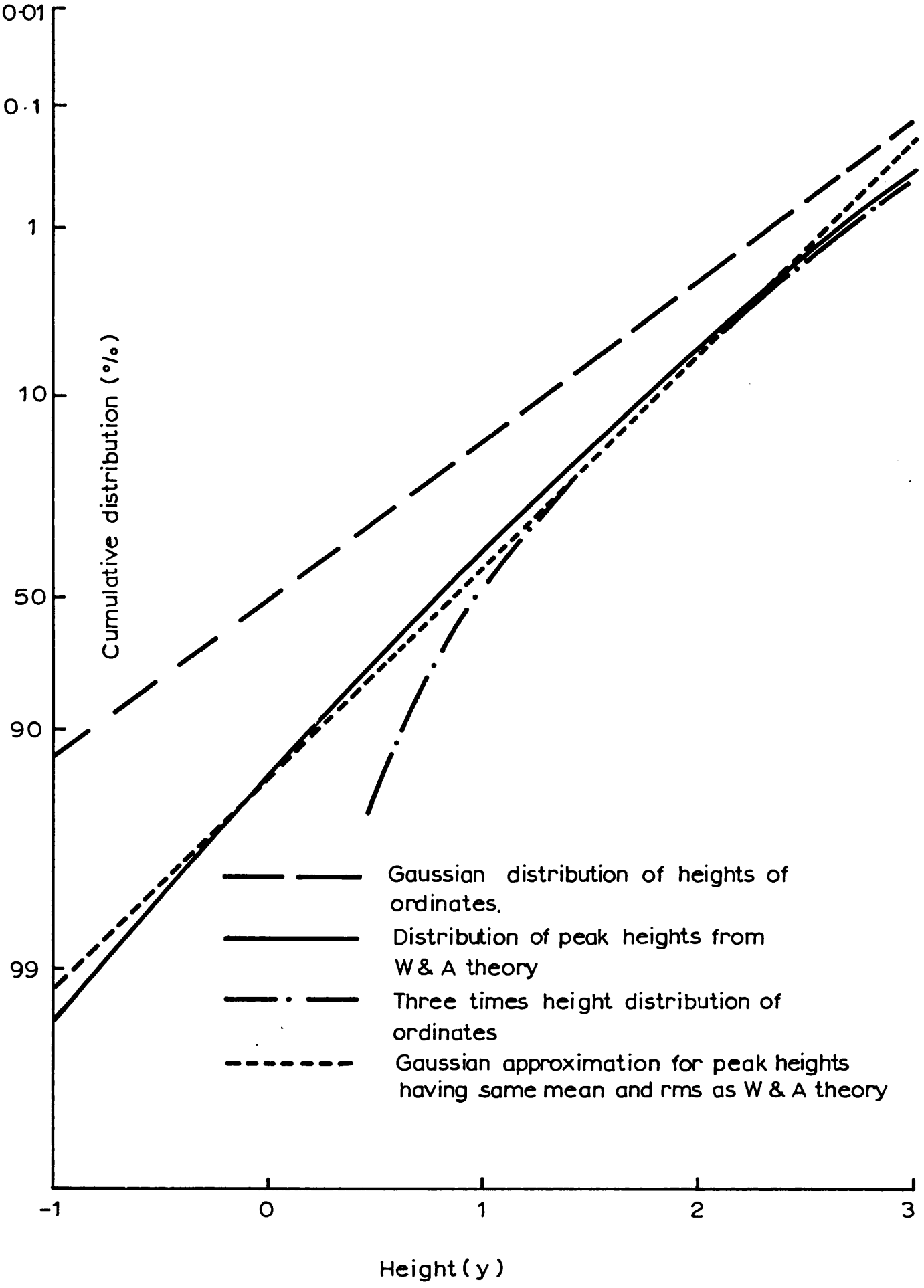


FIG. 4.2.1. Height distribution derived from theories

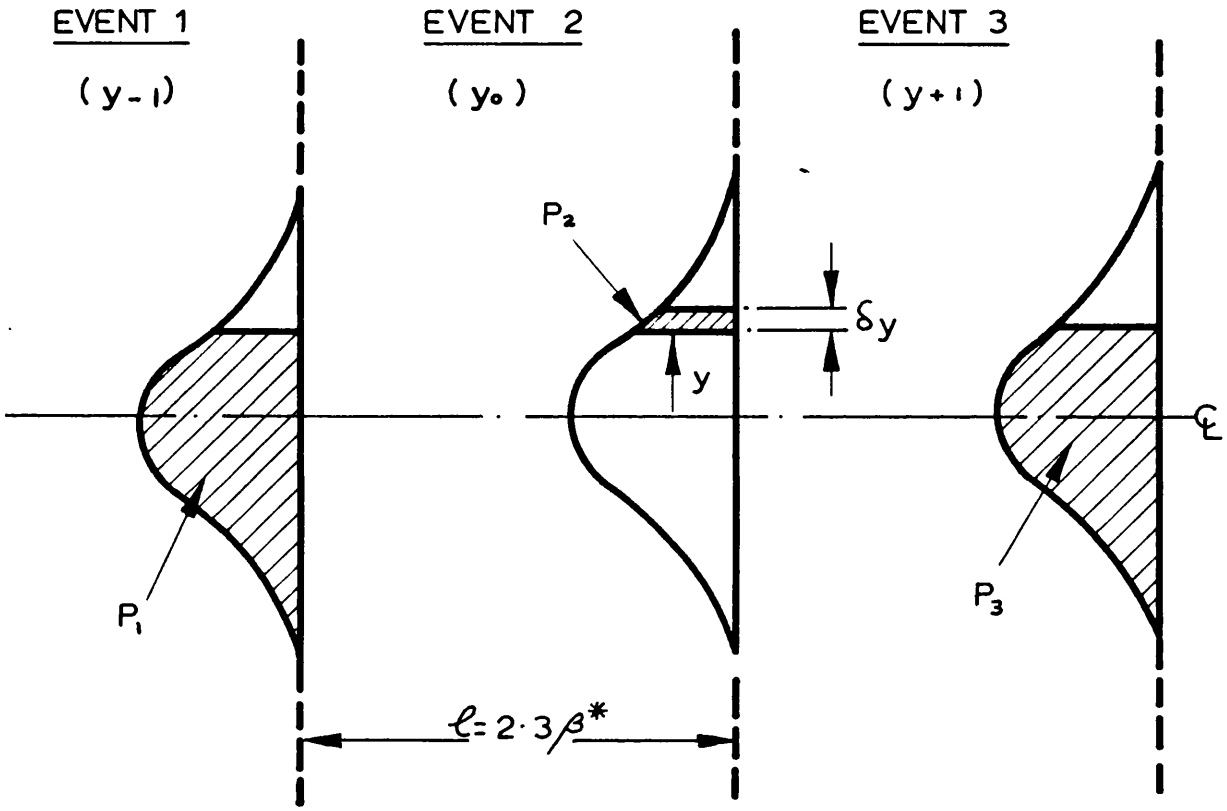


FIG. 4.2.2. Derivation of the peak height distribution

i.e. for large positive values of y the peak height distribution is, to a close approximation, three times the ordinate height distribution from which it is derived. Fig. 4.2.1 shows plots, on probability paper of the ordinate height distribution, F(y) (equation(4.2.17)), the peak height distribution, $F^*(y)$ (equation(4.2.14)), and the approximation for $F^*(y)$ (equation (4.2.15)). It will be seen that the approximation is very close for $y > 1.5$ and adequate for $y > 1.0$. It is shown in this chapter that the approximation, when incorporated into the theory of surface contact, gives results very close to the W & A peak height distribution for $y > 1$ (see the comparison between the G & W model using this Gaussian approximation and the intermediate model). The approximation of equation (4.2.18) has been found both when plotted on Fig. 4.2.1 or when incorporated into the G & W theory of surface contact, to give a good representation of the W & A peak distribution for values of $y > 1$. The reason for this is seen by considering the derivation of the W & A peak distribution for independent events which is shown in Fig. 4.2.2. The probability that the central event is a peak of height between y and (y + dy) is given by the product of the three shaded areas. Thus, when the central event occurs at large positive values of y the probability that it is a peak is virtually its own probability distribution. This, when normalized by the ratio of peaks to ordinates, $N = \frac{1}{3}$, gives a better fit for $y > 2$ than any other chosen Gaussian distribution.

4.3. Theory of surface contact

The analysis of contact developed below will follow the procedure adopted in most earlier theories. The model surface will be considered as pressed against a perfectly smooth rigid flat surface. The area of contact, A, the load, W, and the contact conductance, G, will be

derived as a function of the separation, d, of the surfaces. In this way the desired relationships between A and W and between G and W can be derived. The detailed derivation follows closely the theory of Greenwood and Williamson (1966), the separation, d, (normalized by σ) being the separation of the flat from the centre line of the rough model surface as a reference. In all the calculations which follow this reference line will be the centre line of the ordinate heights. In general (Whitehouse and Archard (1970)) this is not identical with the centre line of the distribution of peaks, or asperities, which, as might be expected, lies above the centre line of the ordinate distribution.

For a single asperity, of radius of curvature, R, the radius of the area of contact, a, the contributions to the area of contact, δA, to the load, δW, and to the conductance, δG, are given by the equations

$$\left.\begin{aligned} a &= (R\omega)^{\frac{1}{2}} && \text{a} \\ \delta A &= \pi R\omega && \text{b} \\ \delta W &= \tfrac{4}{3} E' R^{\frac{1}{2}} \omega^{\frac{3}{2}} && \text{c} \\ \delta G &= (R\omega)^{\frac{1}{2}} g && \text{d} \end{aligned}\right\} \qquad (4.3.1)$$

where ω is the compliance. $E' = \dfrac{E}{(1-\nu^2)}$, where E is Young's modulus and ν is Poisson's ratio; g is the specific electrical conductance. Then using the normalized version of equation (4.2.6) we may write the probable area of contact δA from a peak at height y of curvature C as

$$\text{prob}\,[\delta A]_{yC} = \pi R\omega \{f^*(y,C)/N\} \qquad (4.3.2)$$

where, as explained above, N can be taken as $\frac{1}{3}$.

The contact area, for all curvatures, from peaks at height y is obtained by integrating equation (4.3.2). Then using equations (4.2.2) and (4.2.7)

$$[\delta A_y]_{\text{mean}} = \pi(2.3\beta^*)^2 (y-d) \int_0^\infty \frac{f^*(y,C)}{NC}\, dC \qquad (4.3.3)$$

For a separation d we must consider all levels, from y = d to y = ∞ which contribute to the area of contact. Thus

$$A = \eta \pi (2.3\beta^*)^2 \mathcal{A} \int_d^\infty (y-d) \int_0^\infty \frac{f^*(y,C)}{NC} \, dC \, dy \qquad (4.3.4a)$$

The total number of asperities on the surface is $\eta \mathcal{A}$ where $\mathcal{A}$ is the apparent area.

From equations (4.3.1a) and (4.3.1d) similar expressions may be derived for the total load, W, and the conductance G. Thus

$$W = \frac{4}{3} \mathcal{A} E' (2.3\beta^*) \sigma \int_d^\infty (y-d)^{\frac{3}{2}} \int_0^\infty \frac{f^*(y,C)}{N\sqrt{C}} \, dC \, dy \qquad (4.3.5a)$$

$$G = 2\eta g \mathcal{A} (2.3\beta^*) \int_d^\infty (y-d)^{\frac{1}{2}} \int_0^\infty \frac{f^*(y,C)}{N\sqrt{C}} \, dC \, dy \qquad (4.3.6a)$$

These equations may conveniently be re-expressed in a form using the functions $F_A(d)$, $F_W(d)$, $F_G(d)$ which are tabulated in Table 4.3.1. Then using equation (4.2.8)

$$A = \pi \eta \mathcal{A} (2.3\beta^*)^2 F_A(d) = \frac{1}{5} \pi \mathcal{A} F_A(d) \qquad (4.3.4b)$$

$$W = \frac{4}{3} \eta \mathcal{A} E' (2.3\beta^*) \sigma F_W(d) = \frac{4 \mathcal{A} E' \sigma}{15(2.3\beta^*)} F_W(d) \qquad (4.3.5b)$$

$$G = 2 \eta g \mathcal{A} (2.3\beta^*) F_G(d) = \frac{2 \mathcal{A} g}{5(2.3\beta^*)} F_G(d) \qquad (4.3.6b)$$

and similarly the number of contacts, n, is given by

$$n = \eta F_n(d) = \frac{\mathcal{A}}{5(2.3\beta^*)^2} F_n(d) \qquad (4.3.7)$$

where

$$F_n(d) = \int_d^\infty \int_0^\infty \frac{f^*(y,C)}{N} \, dC \, dy \qquad (4.3.8)$$

Analytic forms for the integrals of equations (4.3.4), (4.3.5), (4.3.6), (4.3.7), (4.3.8) could not be found; however they can be evaluated by

TABLE 4.3.1.

d	$F_A(d)$	$F_W(d)$	$F_G(d)$	$F_n(d)$
0.0	0.3885	0.6120	5.451	2.6
0.2	0.3085	0.4617	4.557	2.4
0.4	0.2202	0.3381	3.684	2.1
0.6	0.1568	0.2403	2.868	1.8
0.8	0.1070	0.1647	2.154	1.5
1.0	0.7050×10^{-1}	0.1098	1.554	1.2
1.4	0.2760×10^{-1}	0.4392×10^{-1}	7.188×10^{-1}	6.7×10^{-1}
1.8	0.9390×10^{-2}	0.1536×10^{-1}	2.844×10^{-1}	3.1×10^{-1}
2.2	0.2805×10^{-2}	0.4695×10^{-2}	7.695×10^{-2}	1.2×10^{-1}
2.6	0.7360×10^{-3}	0.1266×10^{-2}	2.904×10^{-2}	3.9×10^{-2}
3.0	0.1710×10^{-3}	0.2988×10^{-3}	5.940×10^{-3}	1.0×10^{-2}
3.5	0.2295×10^{-4}	0.4083×10^{-4}	1.140×10^{-3}	1.5×10^{-3}
4.0	0.2493×10^{-5}	0.4494×10^{-5}	1.383×10^{-3}	1.8×10^{-4}

These results are tabulated with the normalising factor N incorporated

computer methods which are outlined in Appendix 1. Values of these integrals are listed in Table 4.3.1.

4.4. Plasticity Index

The plasticity index of Greenwood and Williamson (equation (4.1.1)) expresses the way in which the probability of plastic deformation depends both upon the topography of the surfaces and the mechanical properties of the material. It is based upon the assumption that all the asperities have the same radius of curvature, R, and that they have a Gaussian distribution of peak heights characterized by an rms value σ^*.

On the other hand the Whitehouse and Archard model characterises the surface in terms of the rms value of the height distribution, σ, and the correlation distance, β^*. A first approximate expression of the plasticity index in terms of the W & A model can be obtained by using the G & W concepts.

As explained above, the true peak height distribution (equation (4.2.9)) can be replaced by the ordinate height distribution and normalized by the ratio of peaks to ordinates, $N = \frac{1}{3}$. Also we can use the mean peak curvature $\bar{C}^*$; then using equations (4.2.7) and (4.2.12) we may write $R = \dfrac{2\sqrt{\pi}\ (2.3\beta^*)^2}{9\,\sigma}$

Thus
$$\psi = 0.69\left(\frac{E'}{H}\right)\left(\frac{\sigma}{\beta^*}\right) \qquad (4.4.1)$$

which is the simple expression of the G & W plasticity index[+] in terms of σ and β^*.

For the sake of simplicity and also for reasons of physical insight (which will be justified later) the plasticity index for the W & A model will be written as

$$\psi^* = \left(\frac{E'}{H}\right)\left(\frac{\sigma}{\beta^*}\right) \qquad (4.4.2)$$

[+] The result given in Table 2 of Whitehouse and Archard (1970) is in error.

We require, however, a more complete analysis of the probability of plastic deformation for the full W & A model. To a reasonable approximation the compliance corresponding to the onset of plastic flow is given by

$$\omega_p = R(H/E')^2$$

which, when normalized and using equation (4.2.7) gives

$$\omega_p^* = \frac{\omega_p}{\sigma} = \left(\frac{2.3\beta^*}{\sigma}\right)^2 \left(\frac{H}{E'}\right)^2 \cdot \frac{1}{C} \qquad (4.4.3)$$

The areas of contact formed by plastic deformation are those for which

$$(y - d) > \omega_p^* , \quad \text{or}$$

$$y > (d + \omega_p^*)$$

Thus, the probable area of contact (δA_p) formed by plastic flow is derived as before

$$\left\{\left[\delta A_p\right]_c\right\}_{mean} = \pi(2.3\beta^*)^2 \int_{d+\omega_p^*}^{\infty} (y-d) \frac{f^*(y,C)}{N\,C}\, dy$$

and using the same principles as in the derivation of equation (4.3.4) one obtains the total area of contact which involves plastic flow;

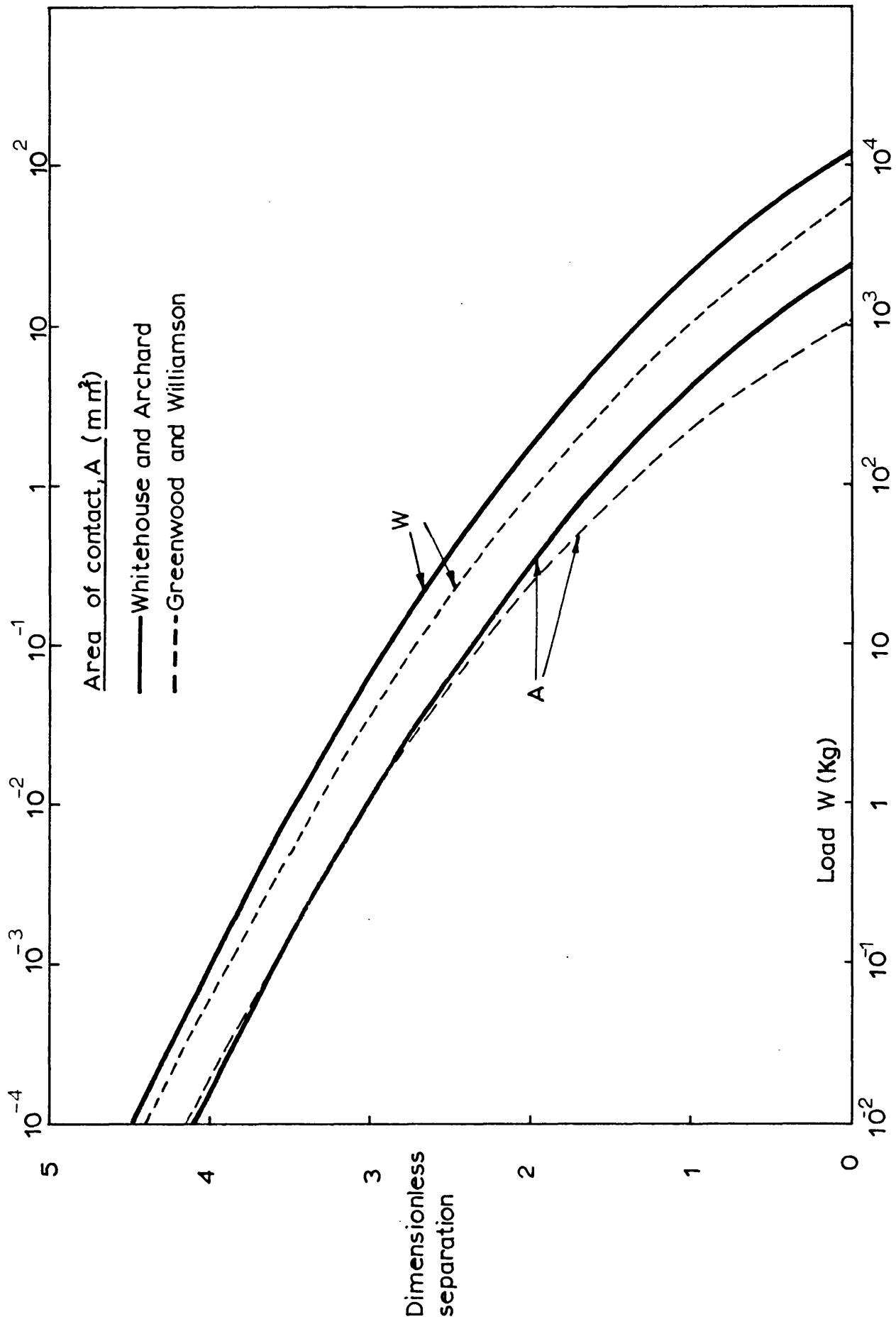
$$A_p = A\, \pi\, \eta\, (2.3\beta^*)^2 \int_0^{\infty}\int_{d+\omega_p^*}^{\infty} (y-d) \frac{f^*(y,C)}{N\,C}\, dy\, dC \qquad (4.4.4)$$

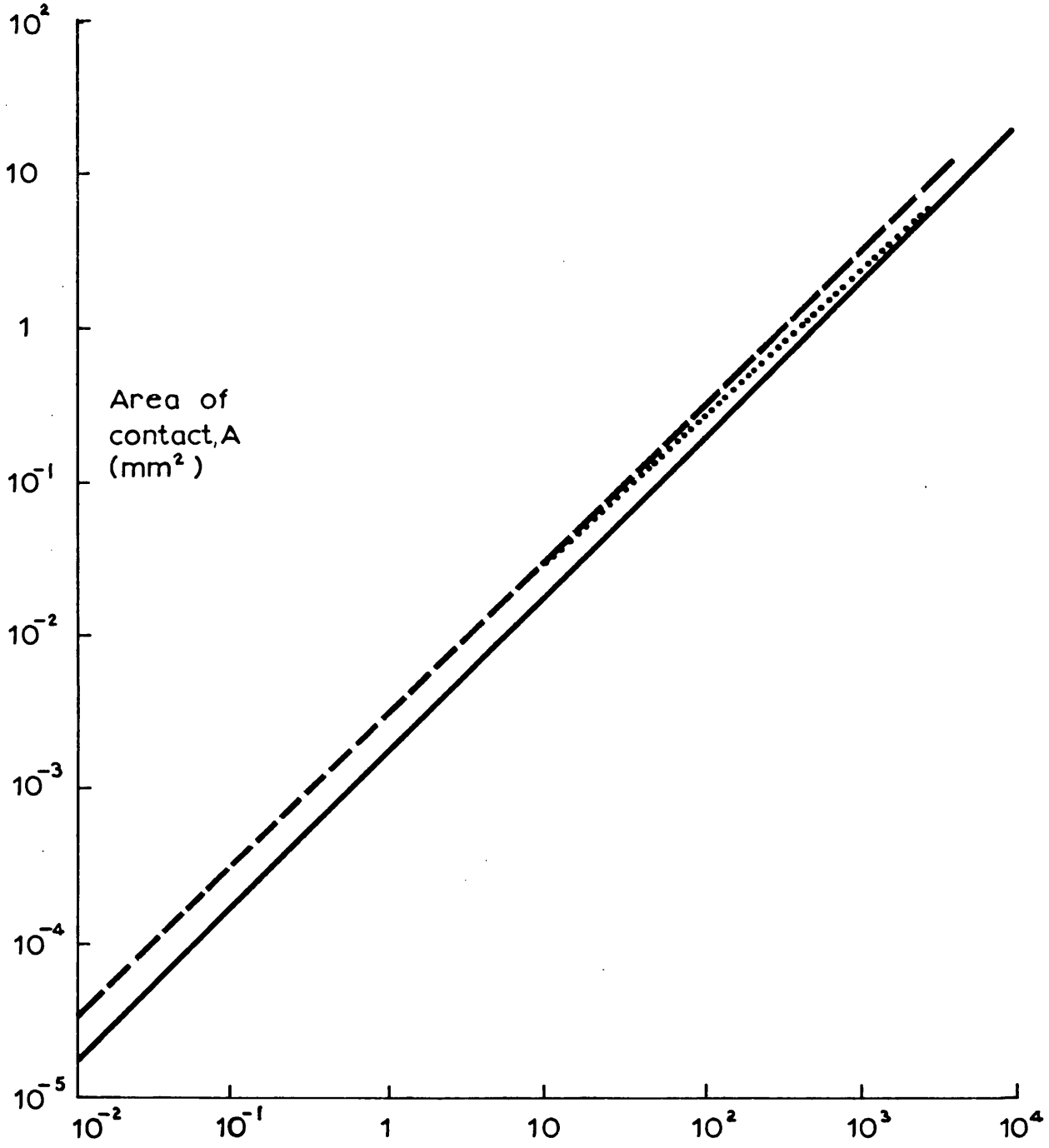
Equation (4.4.4), when compared with equation (4.3.4), gives a complete statement of the probability of plastic deformation with the W & A model. The integration of equation (4.4.4) by computer methods is outlined in Appendix 1. The results obtained are discussed below.

4.5. Results of the theory

The results of the theory are presented assuming steel surfaces

FIG. 4.5.1. Area of contact and load as a function of separation





Relation between area of contact and load

——— Whitehouse & Archard, $A = 10 cm^2$ or $1 cm^2$

– – – – Greenwood & Williamson, $A = 10 cm^2$

····· Greenwood & Williamson, $A = 1 cm^2$

FIG. 4.5.2

FIG. 4.5.3a Mean pressure at real area of contact

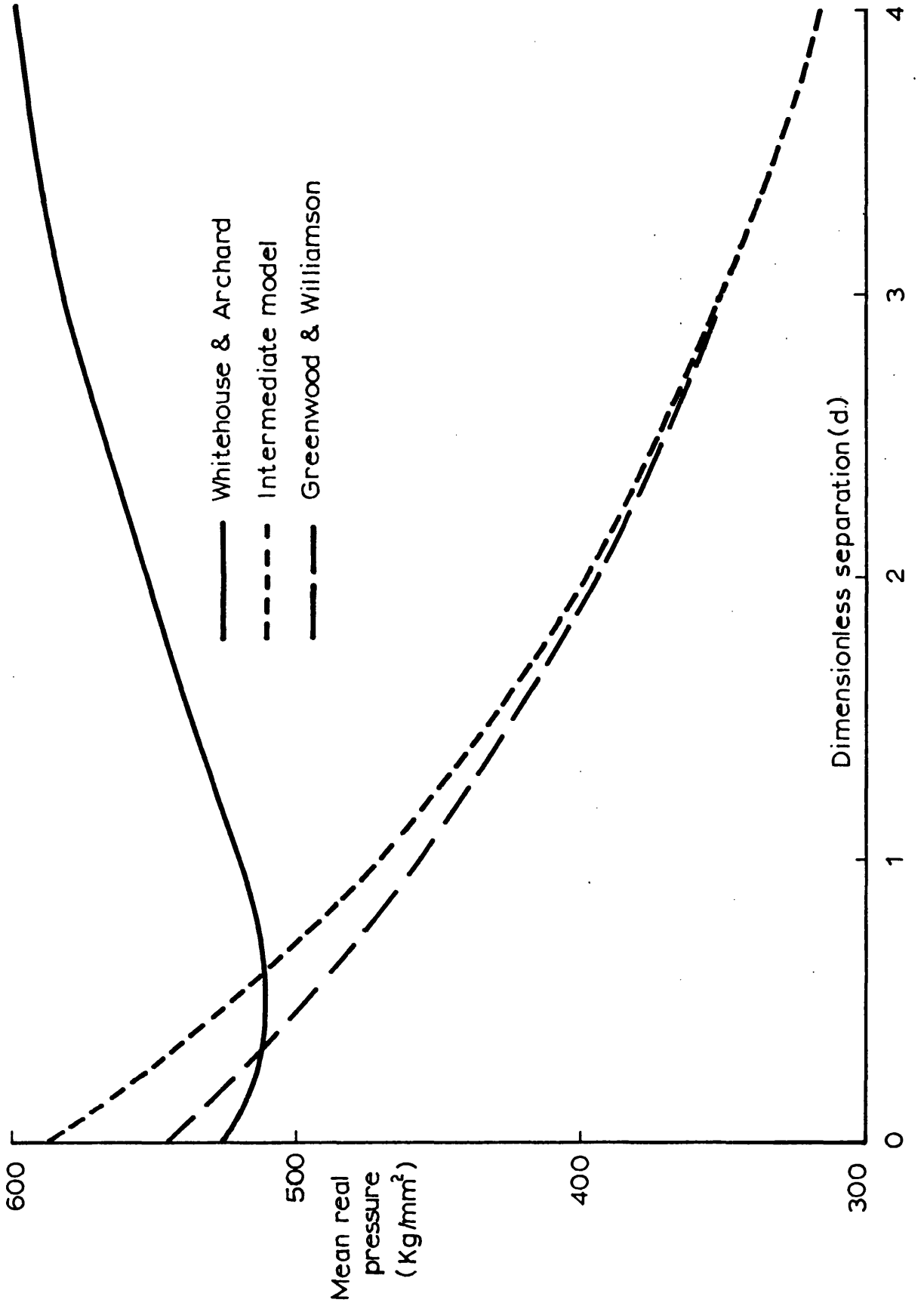
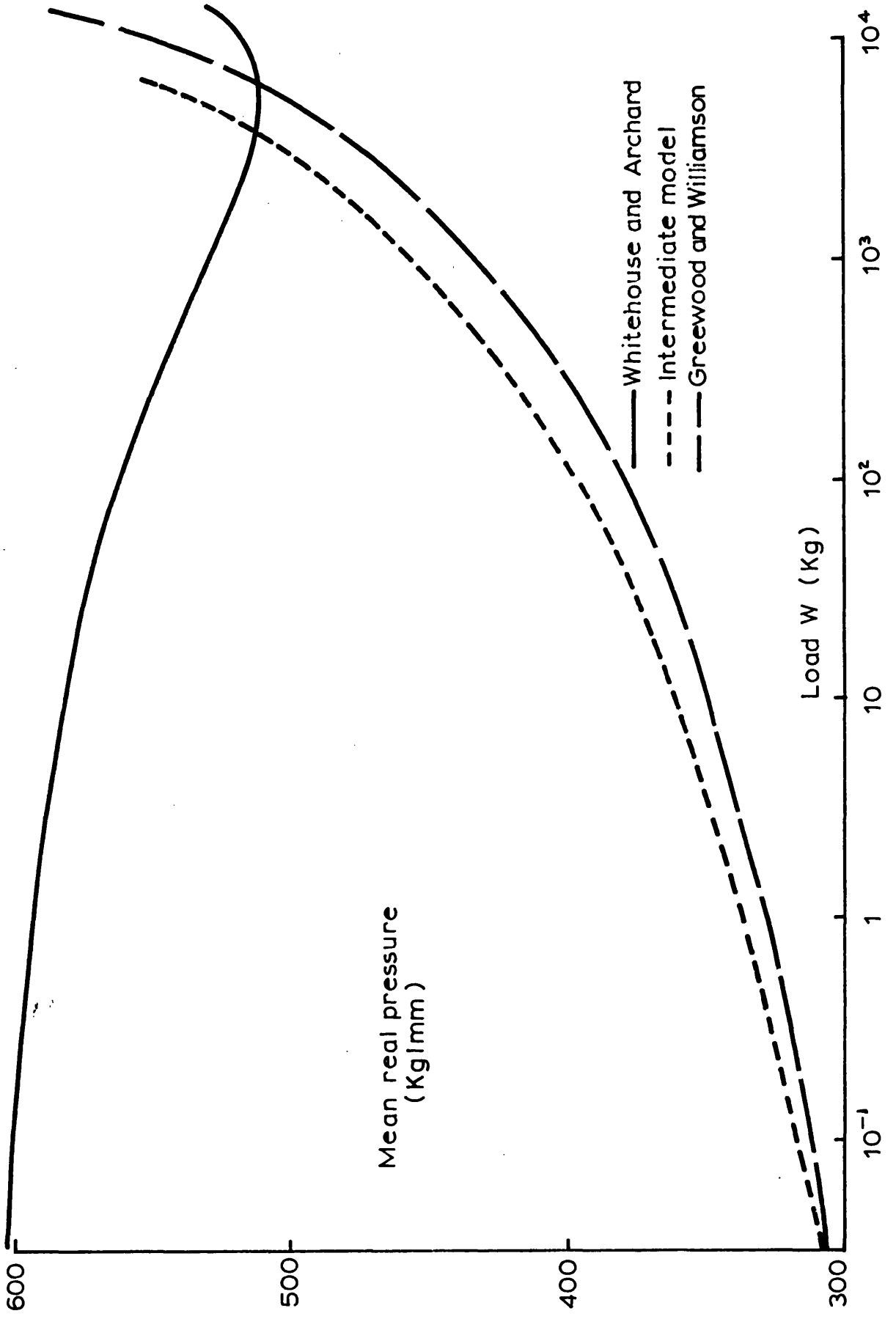


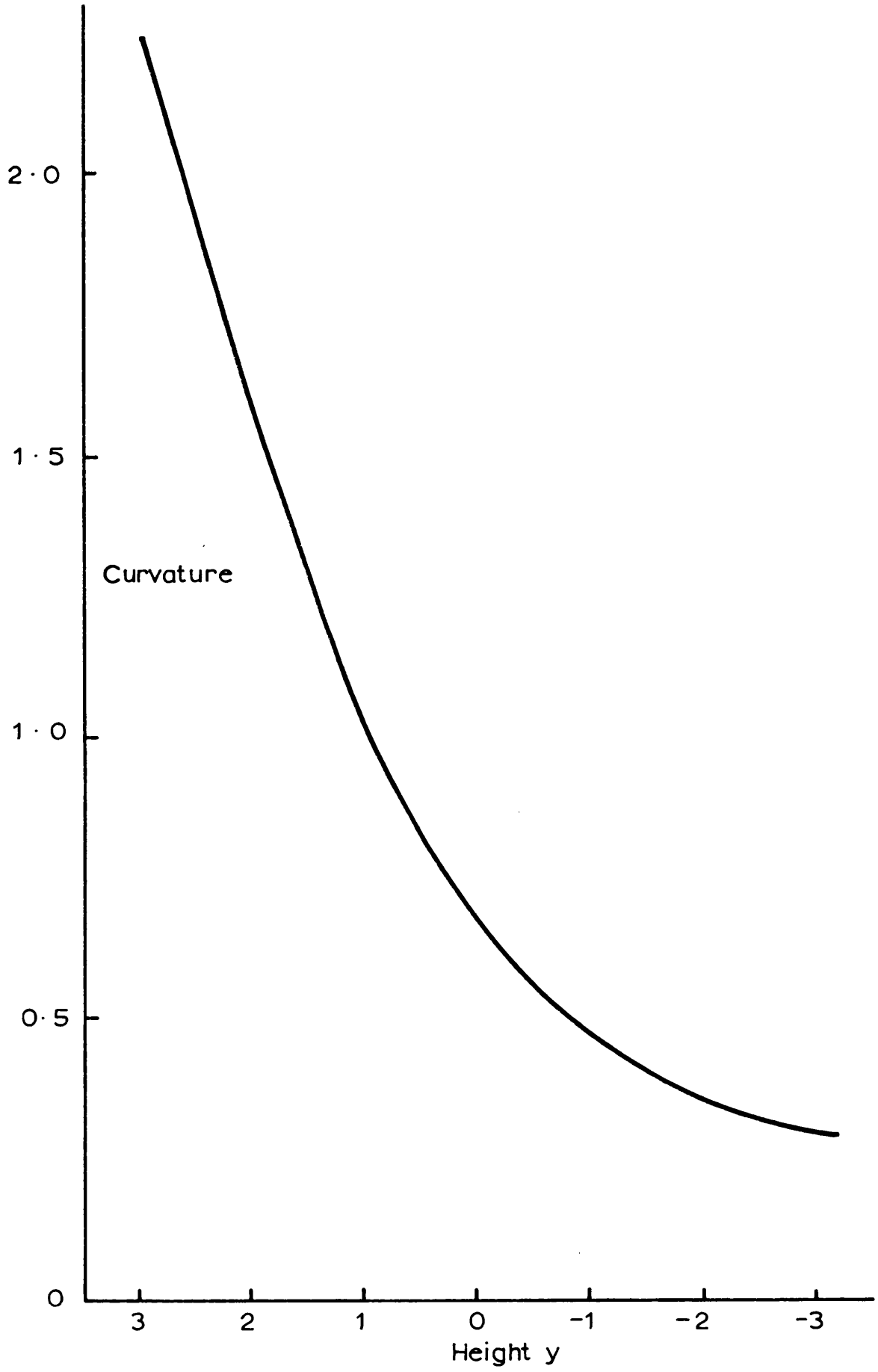
FIG. 4.5.3b Mean pressure at real area of contact as a function of load



having a surface topography similar to those found (Whitehouse and Archard (1970)) to be typical of ground surfaces. The important distinction here is that in the first instance it will be assumed that the surface topography is isotropic. Where appropriate, the results will be presented for the three models which have been described above. These are the original model of Greenwood and Williamson (1966), the model of Whitehouse and Archard (1970) and the intermediate model. The values of the constants used in these calculations are listed in the table of symbols.

Fig. 4.5.1 and Fig. 4.5.2 show the results which are derived directly from the theory. Fig. 4.5.1 shows the relationships between the area of contact and the dimensionless separation and between the load and the dimensionless separation. From these results one can derive relationships between area and load and these are shown in Fig. 4.5.2. Greenwood and Williamson showed that, with their model, a change in the apparent area, A, from $1cm^2$ to $10cm^2$ produced a small, but detectable difference in the area versus load relationship at heavy loads. This result is reproduced, for the assumed conditions, in Fig. 4.5.2. On the other hand, when the W & A model is used the change in A produces such a small change in the area/load graph that it is not detectable.

Of the graphs shown in Fig. 4.5.2, the W & A model produces an area versus load relationship which is closest to direct proportionality. This, and other aspects of the models, are more clearly established by considering the mean pressure over the true areas of contact, ($\bar{p} = W/A$) Fig. 4.5.3a shows $\bar{p}$ as a function of dimensionless separation and Fig. 4.5.3b shows $\bar{p}$ as a function of W. These figures reveal three major differences between the new W & A model and the old G & W model. First, in the W & A model the mean pressures are more nearly independent of load (or separation), this is in line with the area/load relationship discussed above. Second, the mean pressures fall slightly with the



Curvature of peaks as a function of height for the W&A model The curvature of each height is expressed as a ratio of the mean curvature of all peaks. The heigh(y) is normalized by the standard deviation, σ. The results are based upon the theory for $\rho = 0.1$

Fig. 4.5.4

increasing load, or with decreasing separation. Third, in the W & A model the mean pressures are significantly higher than those calculated from the G & W model; over the range of separations of most practical significance they are higher by between 20 and 80%.

The reason for these differences between the W & A and G & W models is demonstrated, quite clearly, by considering results for the intermediate model which are also shown in Figs. 4.5.3a and 4.5.3b. Both the intermediate model and the G & W model have asperities of constant curvature, the value being derived from the mean curvature ($\bar{C}^*$, equation (4.2.12)) of the W & A model. Significant differences between the G & W model and the intermediate model arise only at dimensionless separations between 0 and 1 when the differences between their asperity height distributions become apparent. On the other hand the results for the W & A model are significantly different from those of the other two models over the whole range of conditions, yet the only difference between W & A and the other two models is that it incorporates a distribution of asperity curvatures.

The significant feature of this distribution of curvatures is that as the height of the asperities increases their curvature increases (radius decreases). This point requires closer examination. Equations (4.2.5) and (4.2.6) show that curvature is, indeed, a function of the height, y. The point at issue is made more specific by taking the first moment of the distribution of equation (4.2.5), regarded as a function of C. In this way one can obtain the value of the mean peak curvature for all the peaks at a given height, y. Figure 4.5.4 shows the results of such a calculation for the case of interest : $\ell = 2.3\beta^*$, $\rho = 0.1$. It will be seen that the asperities of greatest significance in surface contact, that is those above the centre line, have curvatures ranging from values slightly below the overall average (at the centre line) to values twice the overall average (at y = 3).

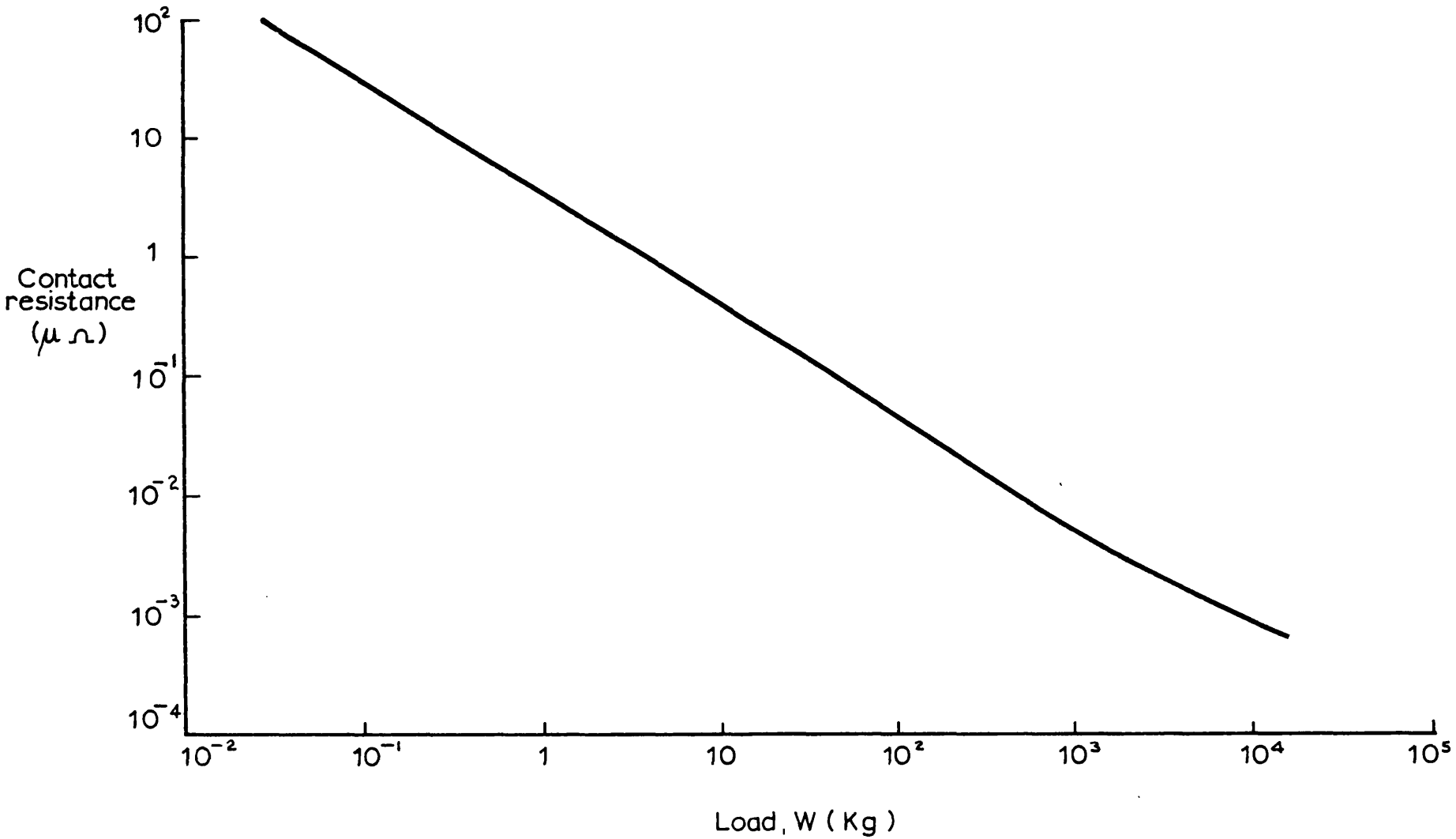


FIG. 4.5.5. Contact resistance as a function of load.

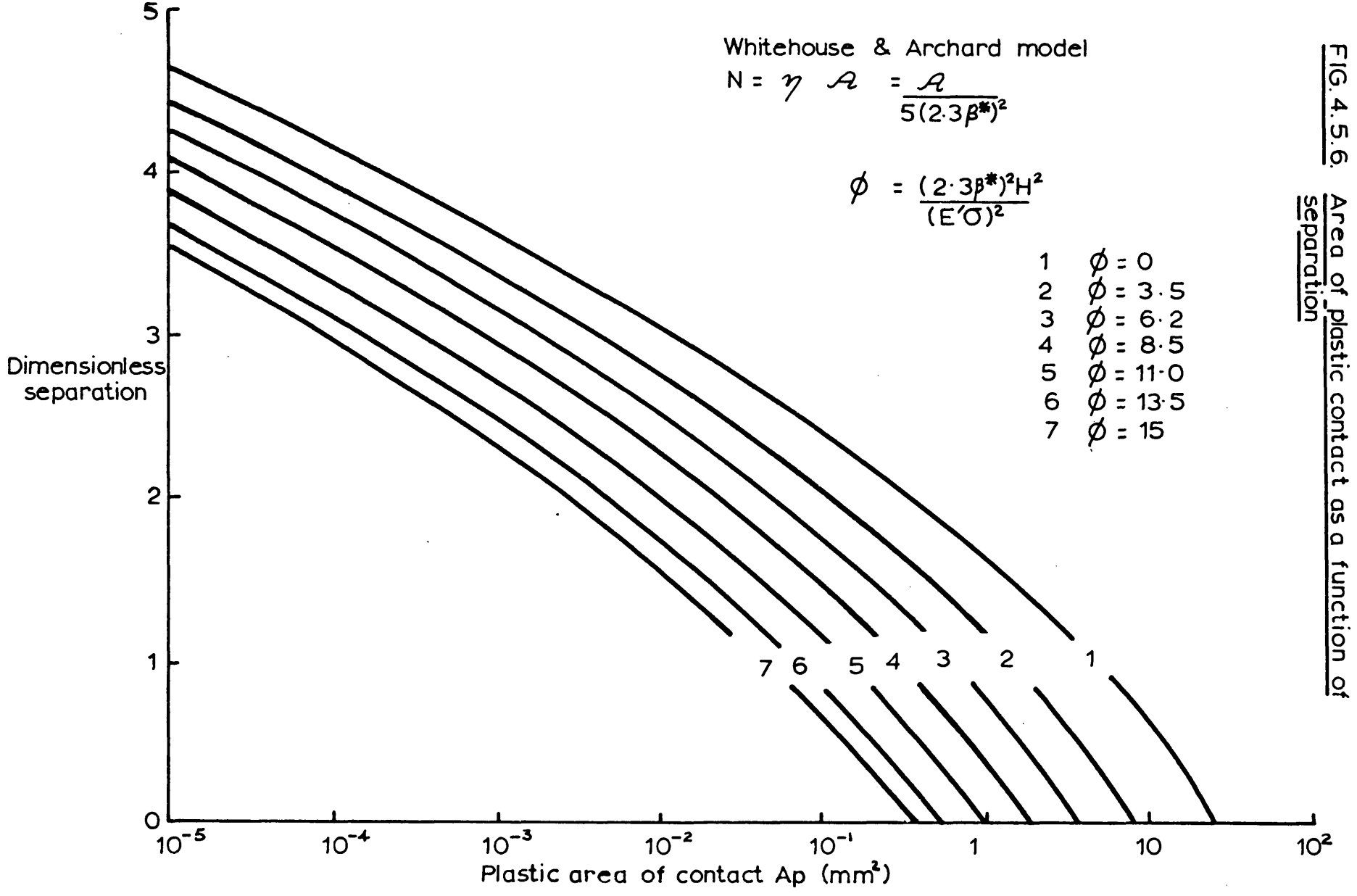
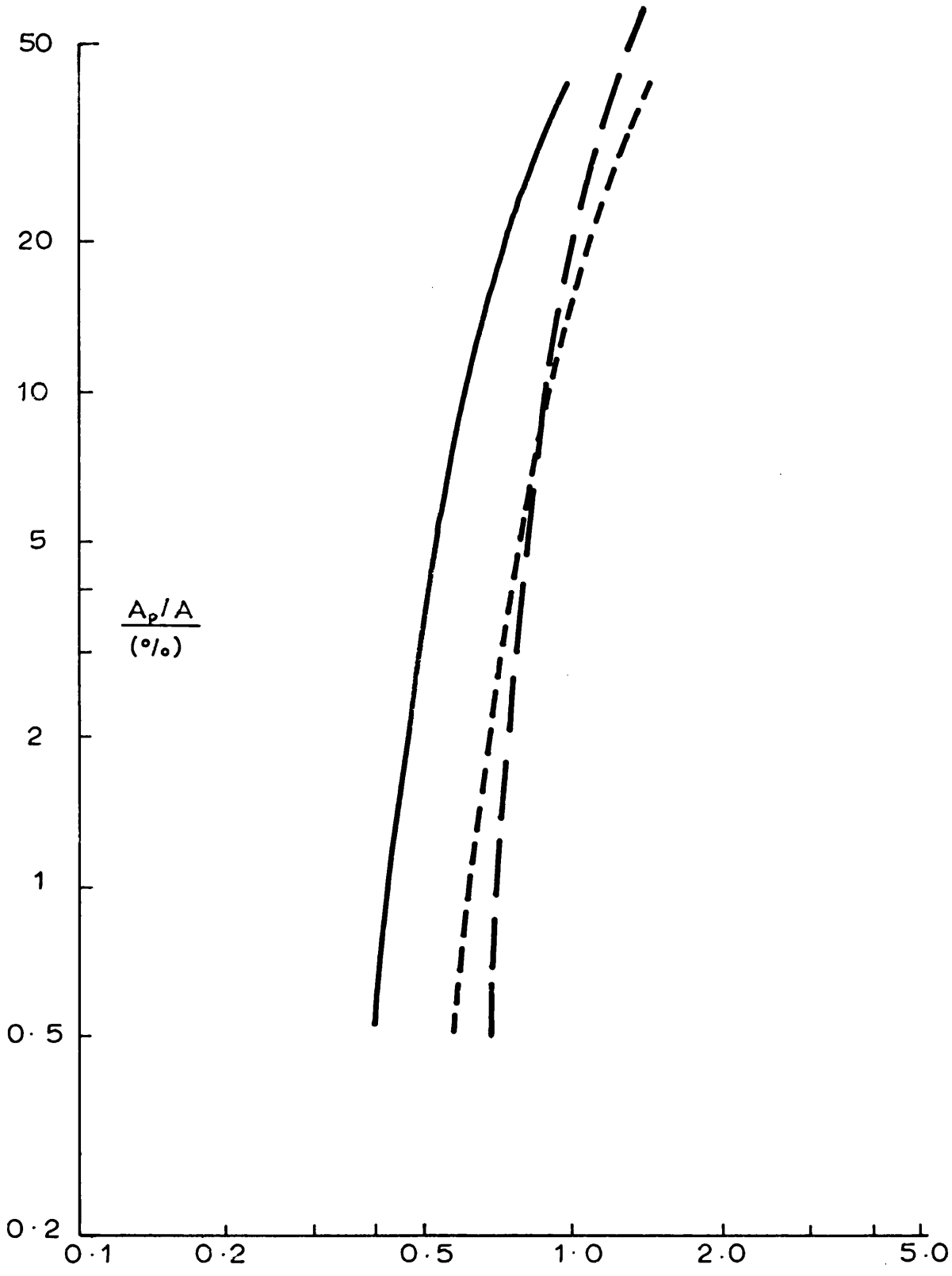


FIG. 4.5.6. Area of plastic contact as a function of separation

The consequence of this feature of the W & A model is that, as the surfaces are pressed together under load, the new asperities which are just being brought into contact have smaller curvatures (larger radii) than those already involved in the existing areas of contact. This produces exactly those consequences which were noted above. Thus when the highest asperities are involved the pressures are higher than those calculated using the mean curvature $\bar{C}^*$. Moreover, as the surfaces are pressed together the newly established areas of contact bear a greater proportion of the load than in the G & W model; this offsets the slight rise in mean contact pressure, $\bar{p}$, with increasing load which is a feature of the G & W model.

Having established the important physical principles involved in the differences between the W & A model and the G & W model, certain other results which can be derived from the models become understandable. Using equations (4.3.5) and (4.3.6) it is found that the relation between electrical conductance, G, and load, W, is to a very close approximation, $G \propto W^{-0.94}$ Fig.4.5.5; over the same range of loads the G & W model (1966) gives $G \propto W^{-0.90}$.

The differences between the two models is also important in plasticity calculations. Using equations (4.3.4) and (4.4.4) it is possible to calculate values of the total area of contact, A, and the area formed by plastic deformation, A_p, as a function of the dimensionless separation, d. Such values of A as a function of d have been shown in Fig. 4.5.1; similar values of A_p can be calculated for different values of the plasticity index as defined by equations (4.4.1) or (4.4.2). The result of these calculations is that the graphs of A_p run closely parallel to the graph of A versus d (Fig. 4.5.6). Thus the ratio A_p/A is virtually independent of the extent to which the surfaces have been pressed together; for a given value of the plasticity index, changes in A_p/A occur when the separation has values between 0 and 1 which corresponds



The influence of plasticity index upon the proportion of area of contact (A_p/A) which involves plastic flow.

—— W & A, ψ, eqn (19a) $W/A = 10^{-2}$ to 10^{2} Kg/cm^2

– – – W & A ψ_1^* eqn (19b) $W/A = 10^{-2}$ to 10^{2} Kg/cm^2

— — Greenwood & Williamson Table 1b $W/A = 1$ Kg/cm^2

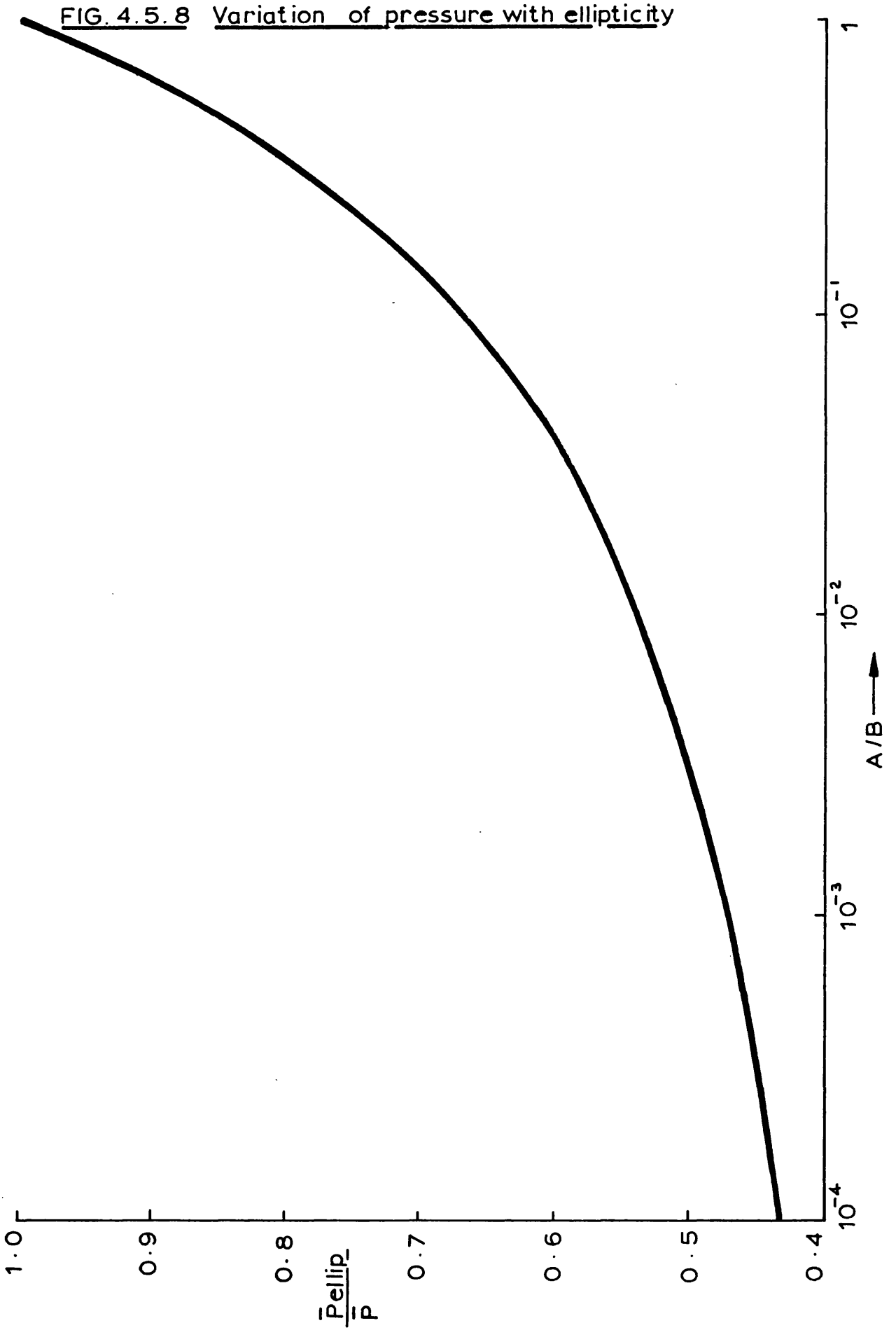
FIG. 4.5.7

to a very high nominal pressure of between 10 and 10^2 kg/mm^2. Thus, even more than is the case with the G & W model, the probability of plastic flow is virtually independent of the load and solely a function of the plasticity index.

However, the use of the W & A model changes the significance of the magnitude of the plasticity index. Use of the Greenwood and Williamson model shows that if ψ, as defined by equation (4.1.1), has a value greater than 1.0, plastic flow will occur even at trivial nominal pressures. On the other hand, if ψ were less than 0.6 plastic flow is most unlikely and only for values of ψ in the range 0.6 to 1.0 is the mode of deformation in doubt. The basis for these statements is shown in Fig. 4.5.7 in which values of A_p/A as a function of ψ are shown. We now derive a new expression for ψ based upon the W & A model using the mean value of the curvature for all peaks (equation (4.4.1)). Using this value of ψ in the full W & A calculations, as outlined above, shows that the significant range of ψ values is 0.35 to 0.60. Thus, as we would expect from the calculations of contact pressures shown in Figs. 4.5.3a and 4.5.3b, use of the W & A model, with its distribution of curvatures, results in a markedly increased probability of plastic flow. Likewise, in the range of ψ values where the mode of deformation is in doubt, use of the G & W model (with asperity curvatures based on mean values) results in a significant under-estimation of the probability of plastic deformation.

We have seen that the change from the G & W model to the W & A model changes the significance of ψ through the introduction of a distribution of asperity curvatures. For this reason, when using the W & A model, it seems appropriate to drop the numerical constant from equation (4.4.1) and use a new designation, ψ^*, (equation (4.4.2)). By a fortunate coincidence, as shown in Fig. 4.5.7 this restores the plasticity index to the significance which it had in the G & W model. If $\psi^* > 1.0$ plastic flow occurs even at trivial loads and if $\psi^* < 0.6$ plastic flow is most unlikely.

FIG. 4.5.8 Variation of pressure with ellipticity



The numerical example chosen to illustrate the theory produces results of considerable interest. Figures 4.5.3a and 4.5.3b show that the theory predicts contact pressures in the range 550-600 Kg/mm^2 for the full Whitehouse and Archard model and that these are very much greater than the simplified models for $y > 1$. For fully hardened steel ($H \approx 800\ Kg/mm^2$) the plasticity index ψ^* will have a value of approximately 2.2. Thus the contact of <u>isotropic</u> surfaces of the chosen characteristics would involve significant plastic deformation. It can be shown that introduction of anisotropy, typical of ground surfaces, will modify the contact pressure by a factor in the range 0.45 to 0.5. For an indication of the treatment of anisotropy, a detailed analysis is carried out in Chapter 5. It will be sufficient at this stage to quote only the equation of the mean contact pressure ($\bar{p}_{ellip}$) for ellipsoidal asperities as a ratio of the mean pressure ($\bar{p}$) obtained from spherical asperities.

$$\frac{\bar{p}_{ellip}}{\bar{p}} = \frac{\pi}{2\sqrt{2}} \left[\frac{(1 + A/B)}{E(k')\ K(k')} \right]^{\frac{1}{2}} \qquad (4.5.1)$$

where $A \leq B$ and A/B is the ratio of the principal curvatures. $K(k')$ and $E(k')$ are the complete elliptic integrals of the first and second kind respectively. A complete definition and explanation of these functions is given in Chapter 5. The influence of ellipsoidal asperities upon mean pressure as given by equation (4.5.1) is plotted in Fig. 4.5.8. The function shows that for A/B in the range 10^{-2} to 10^{-4}, which is typical of ground surfaces (Bell and Dyson (1972)) the elastic contact pressures would be reduced by a factor of 0.45 to 0.55 and consideration of its derivation suggests that the plasticity index would be similarly reduced. If Fig. 4.5.8 is now considered it would be expected that the contact of ground surfaces of fully hardened steel would show a small proportion of plastic flow. Moreover an examination of the definition of the plasticity index also suggests that plastic flow may not necessarily be

reduced by a decrease in roughness since this may well be accompanied by a proportional change in β^*. An indication of the variations of σ and β^* for practical engineering surfaces has been given by Whitehouse (1971). One further problem must be considered in the application of the theory; as stated earlier, the model assumed was that of a rough surface pressing against a perfectly smooth non-deformable plane. This is in fact the same situation as pressing one surface against its own mirror image and is totally unrealistic since it implies that the highest peaks oppose one another. On the other hand for two unsynchronised surfaces this would not be the case and as has been shown by Williamson (1968) the summits of one rough surface often contact points below summits on another. Because of this the value of the plasticity index should be reduced by a further factor of $1/\sqrt{2}$, for surfaces having the same surface roughness, providing even less plastic deformation. For a more detailed discussion of surface contact representation by use of planes and equivalent surfaces or two distinct rough surfaces one may refer to Greenwood and Tripp (1970/71).

Finally it is important to point out that the calculations of this chapter based upon the W & A model are concerned with the main structure of the surface topography. It is an important feature of the W & A model that other features of shorter wavelengths exist upon the surfaces. The presence of these structures will tend further to increase the pressures at the true areas of contact; it will also increase the probability of plastic flow, albeit over smaller regions of the surface. The relevance of these comments is shown by experiments (Whitehouse and Archard, (1970) ; Whitehouse (1971)), in which it has been shown that one of the significant effects of rubbing under conditions of boundary lubrication is largely to remove this smaller wavelength structure.

4.6. Discussion

A theory of surface contact has been developed for surfaces having a random structure, and has been compared with the earlier pioneer study of Greenwood and Williamson. It has been shown that the distribution of asperity curvatures increases the contact pressures and makes them more nearly independent of the load and apparent area. Also the plasticity index, which indicates the probability of plastic deformation, has been re-defined in terms of σ and β^*. The great merit is that the plasticity index is now expressed in terms of surface parameters which are more easily measured and which form part of proposals (Spragg and Whitehouse (1970/71)) for a future specification of surface topography.

The limitations of this model have also been discussed and the results have shown how significant differences can be obtained by using differing definitions of the surface model. It is of some interest to note that the W & A peak distribution has a mean value, $\bar{y}$, and a standard deviation, σ^*, given by

$$\bar{y}^* = 0.85\sigma\,, \quad \sigma^* = 0.75\sigma \qquad (4.6.1)$$

It is therefore tempting to employ a Gaussian distribution with the characteristics of equation (4.6.1) in a theory of surface contact. Such a procedure has been adopted by Johnson, Greenwood and Poon (1972) in their theory of partial elastohydrodynamic lubrication. This Gaussian distribution is a good approximation to equation (4.2.13) over the range $0 < y < 2.0$ as can be seen from Fig. 4.2.1. However it diverges rather markedly for larger values of y. At $d = 3\sigma$, it underestimates the area of contact, A, and the load, W, by the following factors

$$n' = 1.0n \; ; \; A' = 0.85A \; ; \; W' = 0.78W \qquad (4.6.2)$$

In these equations the prime indicates values derived from the Gaussian

distribution with the characteristics of equation (4.6.1) compared with those derived from equation (4.2.13). The differences between different asperity height distributions becomes of more significance in theories of contact through lubricant films where all the factors of equation (4.6.2) are expressed directly in terms of the separation of the surfaces. Therefore the work to follow, in Chapter 5, has been designed to obtain information about the contact of surfaces from their recorded profiles without any initial assumptions of the surface or its related parameters. Specifically Chapter 4 is mainly of use in analysing un-run surfaces which obey defined models; Chapter 5 is an attempt to remove this severe limitation.

CHAPTER 5

DIGITAL SIMULATION OF SURFACE CONTACT

5.1. Introduction

Chapter 4 considered a theory of surface contact using a defined model of the surface topography. To obtain verification of the applicability of the model used the surface geometry must be recorded for as many surfaces and production processes as possible, and statistically analysed. While undertaking this project as much information as possible must be obtained from the surfaces, thus we should have the requirements for an analysis of surface contact, assuming complete information about the asperities is known. We require to determine, by digital presentation and analysis of surface profiles the load borne by the solid contact either in the dry situation or the lubricated case, whichever regime the surfaces may be operating under. With an elastohydrodynamic film incorporated this should reveal the intermediate regime of lubrication under which many practical applications (e.g. gears, bearings etc.) operate, either during the initial stages of running in or throughout their lives. It is this region of Tribology where very little is known and this chapter hopes at least to provide a technique which would lead to a better physical understanding of the situation.

At present there are two important papers that have taken steps in this direction, those being Johnson, Greenwood and Poon (1972) and Bell and Dyson (1972). Both papers provide an analysis of asperity contact in elastohydrodynamic situations but they base their calculations on results derived from an assumed model of surface topography. The model used was that of Greenwood and Williamson (1966) but the experiments which they discuss are not necessarily performed with surfaces which conform to this particular model. However the authors do point

out that the analysis is applicable to any surface model that may be defined the limitation being in the unwieldy mathematics that would occur. In contrast the advantage of the method of analysis presented in this chapter is that it is applicable to any form of surface without recourse to definition.

Johnson, Greenwood and Poon showed that the asperity pressure is determined by the ratio of the theoretical film thickness to the combined surface roughness of the two surfaces (h_o/σ) and they conclude that an increase in load, which only has a small influence on film thickness, is carried by an increase of the fluid pressure. The theory was extended to include the plasticity problem and they found that the number of asperities that become plastic depends on the ratio of the film thickness to surface roughness, and also upon the plasticity index derived for the Greenwood and Williamson model. The theory is then used to compare experimental results of Tallian et al (1964) and Poon and Haines (1966/67) for the no contact time fraction τ as a function of h_o/σ and, having chosen suitable parameters, they show a reasonable fit for theory and experiment. Bell and Dyson in their analysis also used the model presented by Greenwood and Williamson but they remove the assumption of spherical asperities and extend the theory to ellipsoids. They assume that the manufacturing process produces circumferential grooves on the discs and that for a Hertzian theory long elliptical contacts are obtained owing to the asperity interactions. The theory developed is then related to practical results of frictional measurements obtained from a two disc machine. Considering the assumptions, which the authors admit were necessary in the theory in order not to overcomplicate the analysis, they obtained surprisingly good correlation with their practical results.

The intention of this chapter is in a sense to provide a new approach to the problem of surface contact without requiring the aid of

a mathematical model to describe the shape of the solid surface. The outline of the method is as follows.

1) As a first simple approach it will be assumed that the surfaces are covered by spherical asperities of varying radii of curvature, this will then be extended to ellipsoidal asperities such as would occur on circumferentially ground surfaces. The spherical asperities not only lend themselves to a simple mathematical argument but their use is further justified on the grounds that they enable a much greater physical insight to the problem, which for a first stage of development is very desirable.

2) For the physical argument it is assumed, at first, that a series of closely spaced Talysurf profiles are available giving what amounts to almost complete information about the surface asperities; the development of the theory commences from this assumption of full information and at a later stage it will be argued that such complete information is not required. All that would be needed is a sufficient number of profiles to give information which is statistically significant.

3) The profiles obtained are presented in digital form and thus any pair of profiles from opposing surfaces can be aligned, moved together and so indicate the regions of contact. This operation will be performed on the computer with the argument that when the gap between the surfaces is less than zero (i.e. negative separation) this will enable a value of "interference area" to be determined; this term should not be confused with the Hertzian deformed area, since the interference areas do not correspond to the areas that would be formed by solid contact. The areas of solid contact are smaller than the interference areas due to the deformation of the surfaces and so a relationship between these areas is derived in the theory.

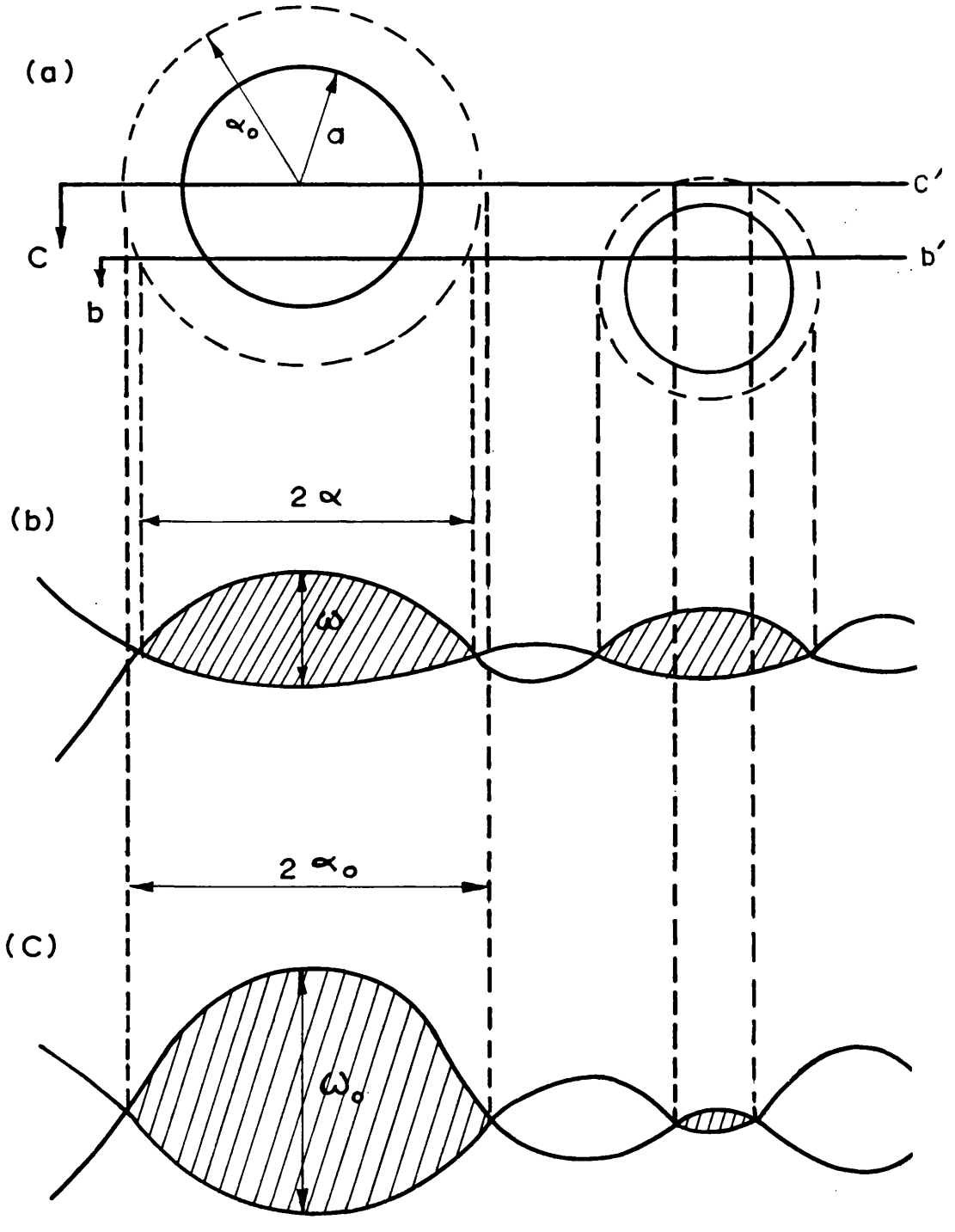


FIG. 5.2.1 Plan and sectional representation of surface showing interference widths α and depths ω Full circles in 5.2.1a represent areas of solid contact.

4) If there are 'n' profiles available from each surface then n^2 profiles will be available for comparison in the computer. At the onset it is assumed that the movement of the surfaces is at right angles to the direction of the profiles. All the n^2 profiles so obtained do not represent a situation which occurs at one moment in time; but, provided that all the n^2 pairs obtained can occur at some moment in time, all the information thus obtained is of significance in representing the time averaged behaviour.

5) As with the previous theory developed in Chapter 4 it will be assumed that the deformations of asperities are independent. This implies that each interference area examined is sufficiently far removed from all others such that its deformation arises solely from its own contact stresses.

5.2. Development of the theory

5.2.1. Theory of spherical asperities

Starting with the initial assumption that complete information is available for analysis, then we may consider two surfaces, consisting of an array of spherical asperities. Each surface is traversed by a series of Talysurf profiles separated by a spacing δy (where δy is very much less than any value of the Hertzian contact radius 'a' of any asperity being considered). The number of profiles per unit length on the surface will be

$$N = \frac{1}{\delta y}$$

A plan view of a surface is presented in Fig. 5.2.1a where two profile lines bb′ and cc′ are shown. The full circular lines represent the

areas of Hertzian contact and the broken lines the corresponding areas of interference derived from the surface profiles. Initially profiles of both surfaces are aligned and then brought together, Fig. 5.2.1b is one such pair under examination. The interference region for one contact will have a radius α and a depth of interference ω. It must of course be realised that in general the aligned profiles will not indicate the interference which corresponds to the central profile where α and ω will have their maximum values. Fig. 5.2.1c shows one such occurrance where the corresponding values of α and ω now take their maximum values of α_o and ω_o.

Since the asperities are spherical the local relative radius of curvature R of the surface may be defined by

$$R = \frac{\alpha^2}{2\omega} = \frac{\alpha_o^2}{2\omega_o} \qquad (5.2.1)$$

From Hertzian theory (Timoshenko and Goodier) it is easily shown that

$$a = \left(\frac{3\pi^2 WR}{4E'}\right)^{\frac{1}{3}}$$

$$\omega_o = \left(\frac{9\pi^4 W^2}{16 E'^2 R}\right)^{\frac{1}{3}}$$

gives $$a^2/\omega_o = R \qquad (5.2.2)$$

where $$\frac{1}{E'} = \frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2}$$ is the reduced elastic modules

and $$1/R = 1/R_1 + 1/R_2$$

where R_1 and R_2 are the radii of the two asperities, ν is Poisson's ratio and E is the elastic modulus of the respective bodies.

From equations (5.2.1) and (5.2.2) we have the important relationship between the radius of the Hertzian area of contact, a, and the maximum interference radius α_o,

$$\alpha_o = a\sqrt{2} \qquad (5.2.3)$$

The relationship between α_o the maximum interference half width and α_{mean} is given by

$$\alpha_{mean} = \frac{1}{4} \pi \alpha_o \tag{5.2.4}$$

Similarly it may be shown that the relation between the maximum interference depth ω_o and ω_{mean} is given by

$$\omega_{mean} = \frac{2}{3} \omega_o \tag{5.2.5}$$

Considering one asperity with a given region of contact defined by $A = \pi a^2$ this area would be indicated in the Talysurf profiles as an area of interference $A_i = 2A = \pi \alpha_o^2$

Applying the Hertzian equations the load W supported by solid contact may be written as

$$W = \frac{4}{3} \frac{a^3 E'}{R} \tag{5.2.6}$$

This region would be observed to occur on n successive profile pairs of separation δy

$$n = 2\alpha_o/\delta y = 2\sqrt{2}\, a/\delta y \tag{5.2.7}$$

Thus the load supported per interference event observed on the profiles is

$$\frac{W}{n} = \frac{4}{3} \frac{a^3 E'}{R} \frac{\delta y}{2\sqrt{2}\, a} = \frac{\sqrt{2}}{3} \frac{a^2 E'}{R} \delta y \tag{5.2.8}$$

using equations (5.2.1) and (5.2.3) the above equation may be rewritten as

$$\frac{W}{n} = \frac{\sqrt{2}}{3} \omega_o E' \delta y$$

and using equation (5.2.5) this then becomes

$$\frac{W}{n} = \frac{1}{\sqrt{2}} \omega_{mean} E' \delta y \tag{5.2.9}$$

For pairs of profiles spaced a distance δy apart equation (5.2.9) represents the contribution of each interference region on the profiles to the total load, it is not a physical description of what happens, it is

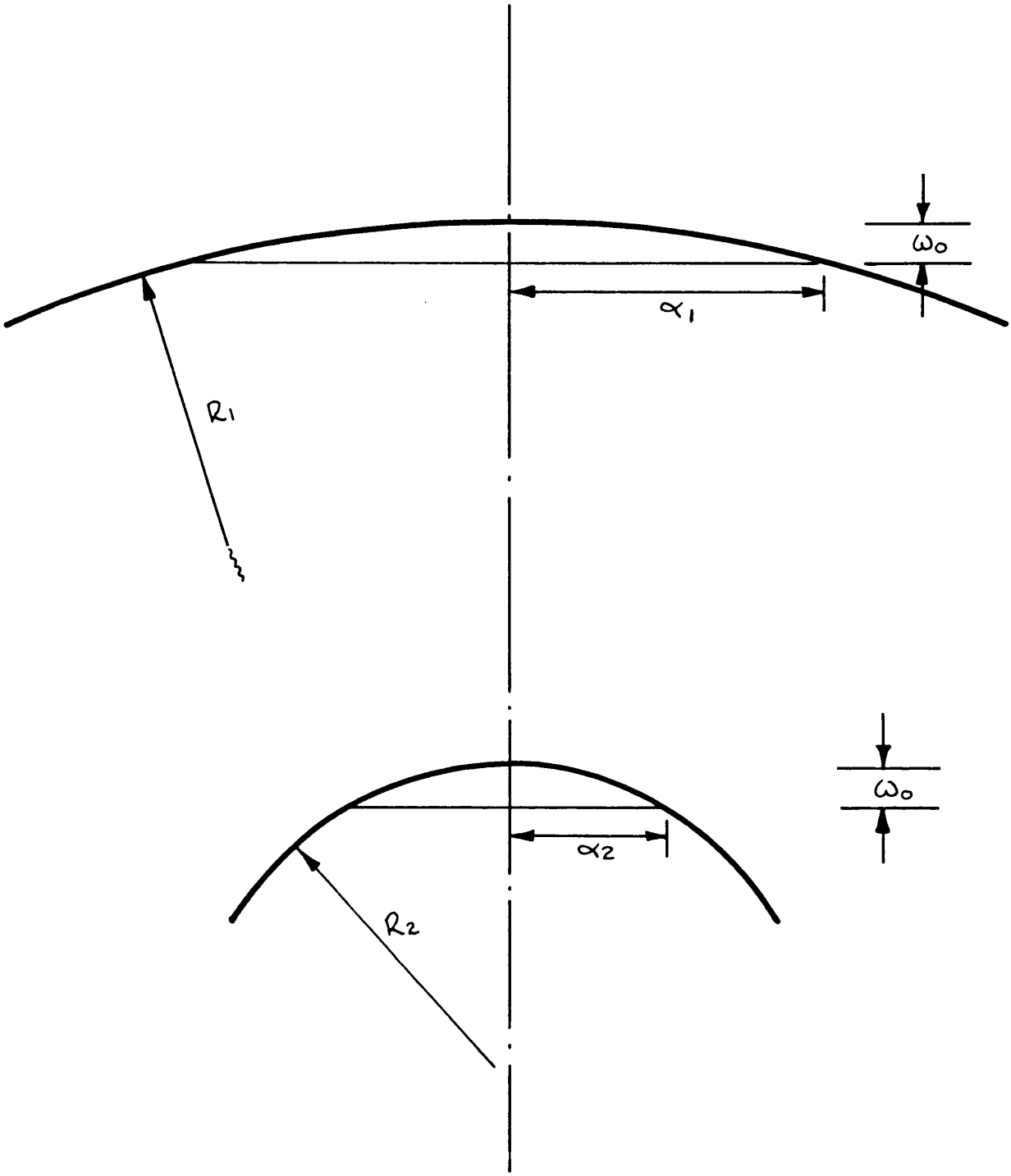


FIG. 5.2.2 Comparison of two interference regions both having the same interference depth. ω_o

the result of sharing the load W over the interference area if this were possible. As has been stated earlier, there are parts of the interference area that do not in fact form part of the area of contact, but even so the principle adopted is valid because the existence of an interference area automatically implies the existence of a contact area.

Equation (5.2.9) may be rewritten to indicate the total load for that contact area as

$$W = \frac{1}{\sqrt{2}} n \omega_{mean} E' \delta y$$

or may be transposed to give the contribution of an interference region on the profiles to the load per unit length at right angles to the profiles

$$W' = \frac{W}{n \delta y} = \frac{\omega_{mean} E'}{\sqrt{2}} \qquad (5.2.10)$$

This is a very neat and rather surprising result, it implies that in the proposed system the contribution of an interference region to the load supported per unit length is proportional to the depth ω of the interference region but independent of its width $2a$. In support of this argument consider the simple problem of Fig. 5.2.2 where there are two regions having the same interference depth ω_o but different interference widths a_1 and a_2.

Assume $a_1 = 2a_2$

Then from equation (5.2.1) $R_1 = 4R_2$

and using equation (5.2.6) $W_1 = 2W_2$

Thus the loads per unit length are

$$\frac{W_1}{2a_1} \text{ and } \frac{W_2}{2a_2}$$

which are equal.

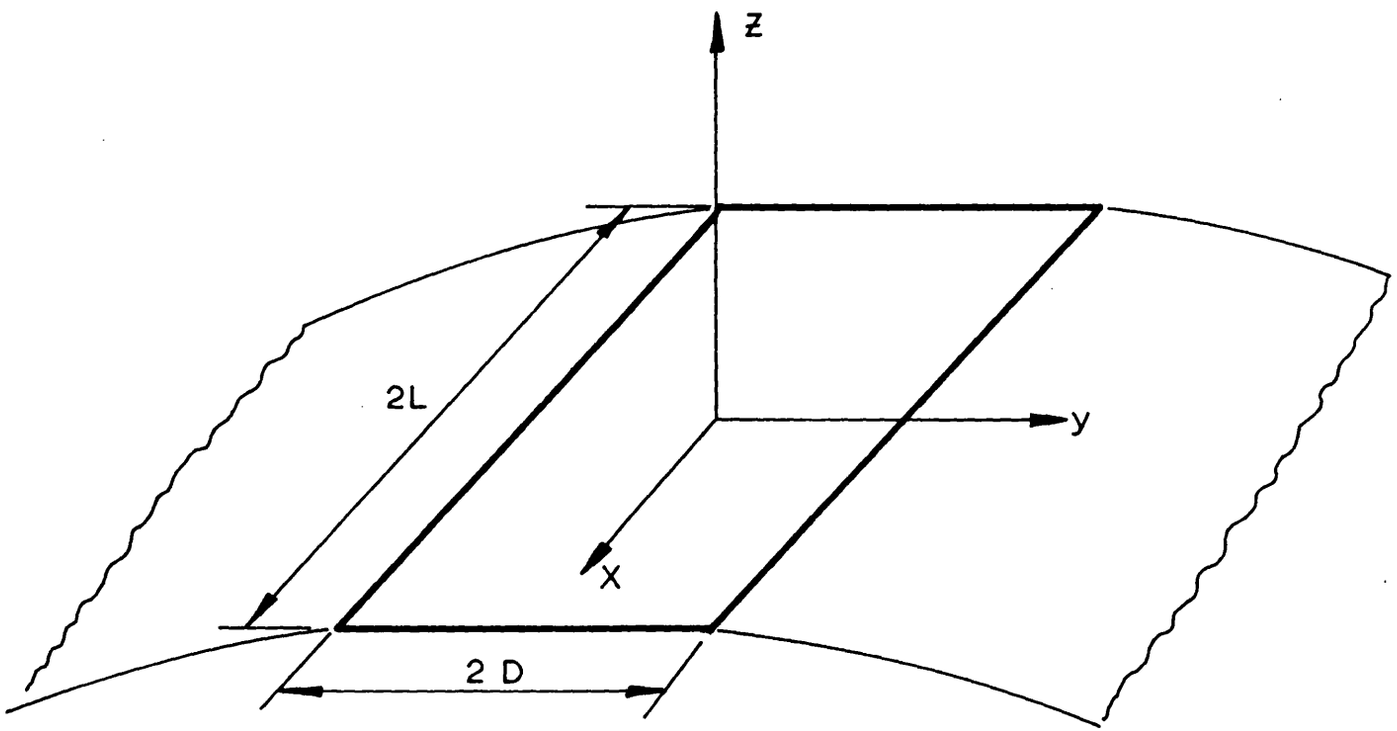


FIG. 5.2.3 Hertzian zone of an elasto hydrodynamic contact between discs.

5.2.2. Application to the disc machine problem.

Consider the Hertzian zone of an elastohydrodynamic contact between two discs, Fig. 5.2.3. The width of the discs is 2L and the width of the Hertzian band of contact is 2D. The motion takes place in the y-direction and the profiles of the surface are of length l (where $l < 2L$) and are in the x direction. Two more assumptions must now be taken into account before proceeding with the development of the theory.

1) The deformation (movement of the surfaces in the z direction) arising from asperity contact is small compared with the deformation arising from the main Hertzian pressure distribution.

2) The analysis may now proceed with the assumption that the asperity contacts can be considered equivalent to those between asperities on two flat surfaces of dimensions 2L x 2D.

Again the assumption that there are in existence sufficient Talysurf profiles to provide a good statistical average must for the present theory be taken into account. Let there be 'm' such pairs of profiles of length l; then the load per unit length in the y-direction corresponding to the total profile length $m^2 l$ may be derived from equation (5.2.10) which must be summed for all asperity contacts, i.e. it may be considered as summing all the interference depths for a single equivalent profile of length $m^2 l$, which gives the load per unit length in the y-direction

$$W'_{total} = \frac{1}{\sqrt{2}} E' \Sigma \omega \qquad (5.2.11)$$

where $\Sigma \omega$ is the sum of all interference depths observed by the computer.

Thus the estimate of total load supported by the area 2B x 2L is the load per unit area multiplied by 4D L and becomes

$$W = \frac{4 D L E'}{\sqrt{2}\, m^2 l} \Sigma \omega$$

$$= \frac{2\sqrt{2}\, D L E'}{m^2 l} \Sigma \omega \qquad (5.2.12)$$

5.2.3. Theory of plastic contact

Chapter 4 has shown that in many conditions the load supported by asperities involved plastic deformation and so the preceding theory is now extended to take this condition into account.

From Hertzian theory

$$p_{mean} = \frac{W}{\pi a^2}$$

$$\text{and} \quad p_o = \frac{3}{2} p_{mean}$$

$$\text{therefore} \quad p_o = \frac{3W}{2\pi a^2} \tag{5.2.13}$$

or from equation (5.2.6)

$$p_o = \frac{2 a E'}{\pi R}$$

and using equations (5.2.1) and (5.2.3) this becomes

$$p_o = \frac{2\sqrt{2}E'}{\pi} \frac{\omega_o}{a_o} < \frac{2\sqrt{2}E'}{\pi} \frac{\omega}{a}$$

The onset of plastic flow can be considered to have been reached when the maximum Hertzian pressure satisfies the equation

$$p_o > 0.6H$$

thus plastic flow has occurred when

$$\omega/a > 0.666H/E' \tag{5.2.14}$$

Therefore using the same arguments as earlier the total load supported by plastic deformation as detected by Talysurf is given by

$$W_p = \frac{2\sqrt{2}\, DL E'}{m^2 l} \Sigma \omega_p \tag{5.2.15}$$

where $\Sigma \omega_p$ is the sum of the values of ω where equation (5.2.14) holds for these particular regions of contact.

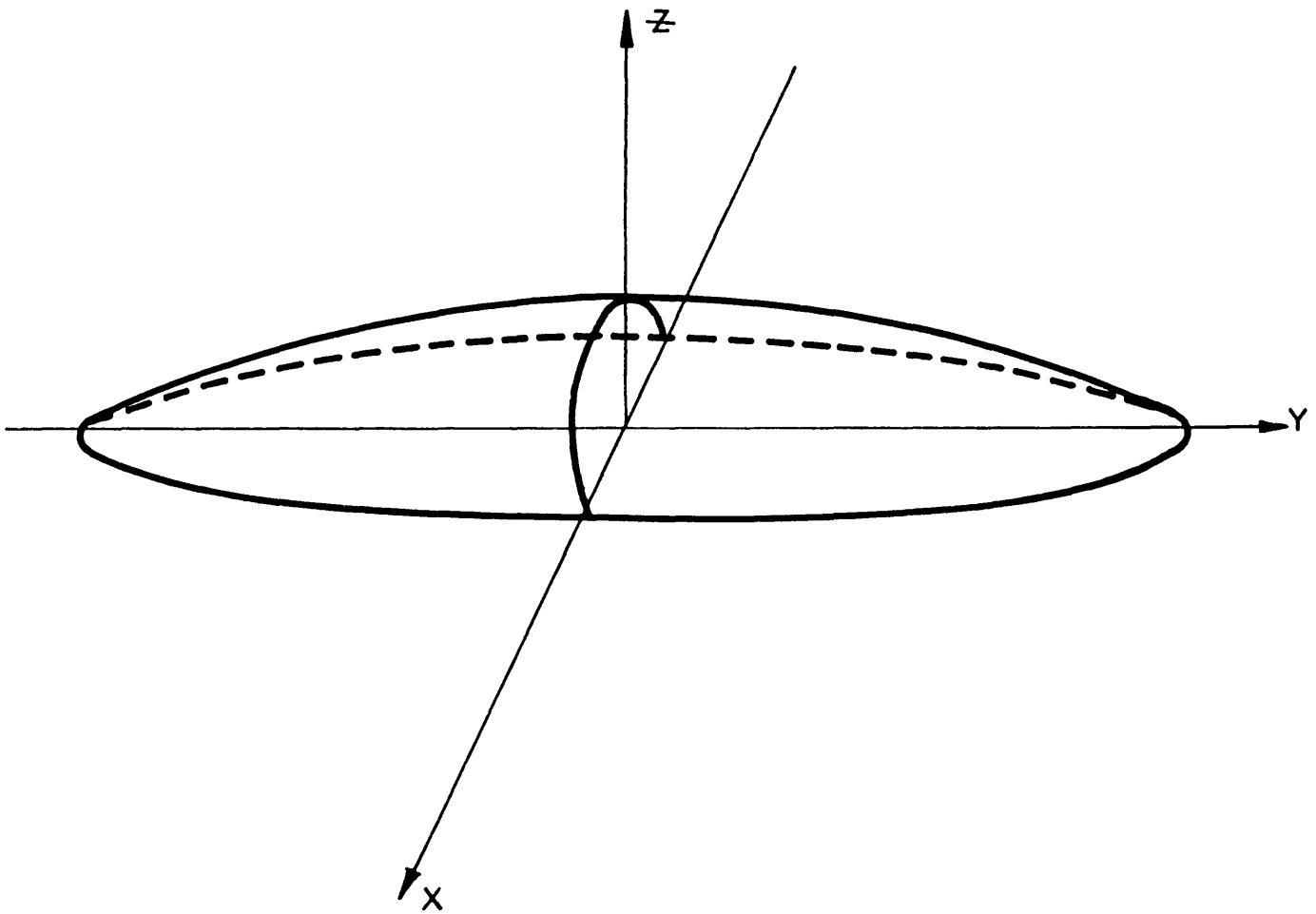
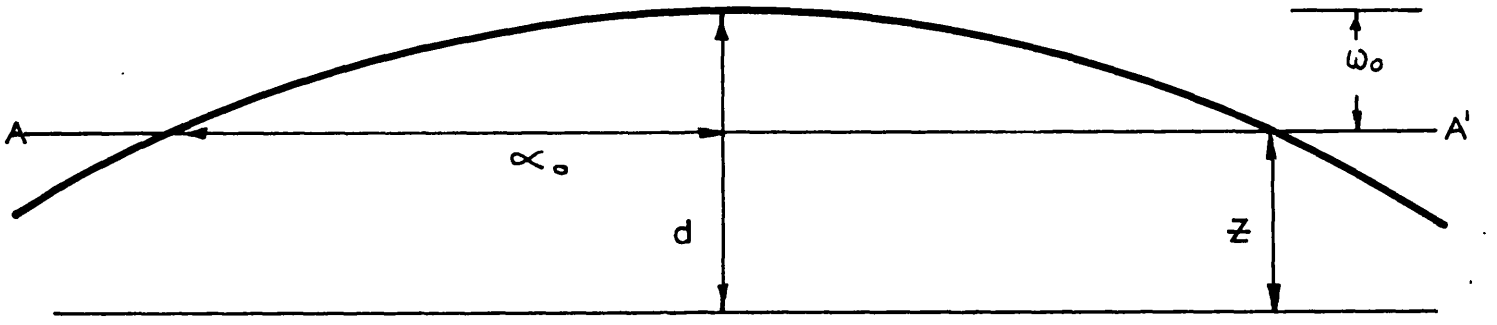
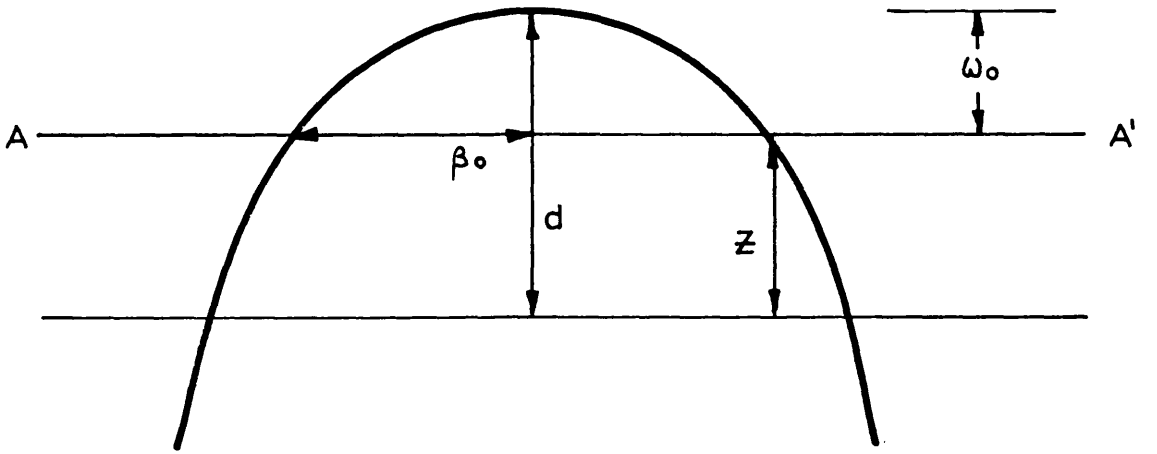


FIG. 5.2.4a Ellipsoidal asperity giving co-ordinates



(b)



(C)

FIG. 5.2.4. b & c b sectional view in Zy plane
c sectional view in Zx plane

5.2.4. Theory for ellipsoidal asperities

As an initial starting point the simplest case of a rough surface and smooth plane will be considered. The shape of a single asperity may then be described as

$$z = d - Ay^2 - Bx^2 \quad A < B \tag{5.2.16}$$

Figure 5.2.4 represents such a cross-section of an ellipsoidal asperity through its summit therefore α_o, β_o and ω_o are a maximum. The principal curvatures of the asperity are 2A and 2B (Thomas and Hoersch 1930) and d is the height of the summit above the mean level of the surface. If a smooth flat plane is now moved towards the ellipsoidal asperity to a height z above the mean plane of the rough surface we have an interference area created which, as in the previous theory, is not the same as the Hertzian area which would have been created by deformable bodies. It is required to find a relationship between these two areas so formed and the applied load which would cause them.

From the Hertzian theory for dry elastic contact presented in a more manageable form by Thomas and Hoersch (1930) the load W required to cause the deformation is given by

$$W = 2\pi E'(B + A)^{-\frac{1}{2}} (d - z)^{\frac{3}{2}} f_1(A/B) \quad (A/B < 1) \tag{5.2.17}$$

where $(d - z) = \omega$ the compliance of the asperity caused by the load W. As before E' is the reduced elastic modulus of the materials and

$$f_1(A/B) = (3k)^{-1} [E(k')]^{\frac{1}{2}} [K(k')]^{-\frac{3}{2}}, \tag{5.2.18}$$

where

$$K(k') = \int_0^{\pi/2} (1 - k'^2 \sin^2\theta)^{-\frac{1}{2}} \, d\theta$$

$$E(k') = \int_0^{\pi/2} (1 - k'^2 \sin^2\theta)^{\frac{1}{2}} \, d\theta$$

are the complete elliptic integrals of the first and second kind respectively and $k^2 = 1 - k'^2$ where k (≤ 1) is the axial ratio of the ellipse of contact. The equation relating it to the ratio A/B is given by

$$\frac{A}{B} = \frac{K(k') - E(k')}{k^{-2} E(k') - K(k')} \qquad (5.2.19)$$

If α_o and β_o are the major and minor semi-axes of the ellipse of interference, the equations for these values may be written as

$$\alpha_o^2 = \frac{\omega_o}{A}$$

$$\beta_o^2 = \frac{\omega_o}{B}$$

substituting for $\omega_o^{\frac{1}{2}}$ in equation (5.2.17)

$$W = \frac{2\pi E' \alpha_o A^{\frac{1}{2}} \omega_o}{(A + B)^{\frac{1}{2}}} f_1(A/B) \qquad (5.2.20)$$

or

$$W = 2\pi E' \alpha_o \omega_o f_2(A/B) \qquad (5.2.21)$$

where $f_2(A/B) = (A/B)^{\frac{1}{2}} (1 + A/B)^{-\frac{1}{2}} f_1(A/B)$

Repeating the earlier arguments this region of the ellipse will be observed to occur on n successive profiles of separation δy

$$n = \frac{2\alpha_o}{\delta y}$$

and the load supported per interference event will be

$$\frac{W}{n} = \delta y \, \pi \, E' \omega_o \, f_2 (A/B)$$

Also

$$\omega_{mean} = \frac{2}{3} \omega_o$$

and thus

$$\frac{W}{n} = \frac{3}{2} \delta y \pi E' \omega_{mean} f_2(A/B)$$

Therefore the load supported per unit width will be

$$W' = \frac{W}{n\,\delta y} = \frac{3}{2}\,\pi\,E'\,\omega_{mean}\,f_2(A/B) \qquad (5.2.22)$$

Comparison with equation (5.2.10) shows that the change from spherical to ellipsoidal asperities modifies the equation for the load solely by introducing a new constant of proportionality which depends only on the assumed shape of the asperities i.e. the length/width ratio.

5.2.5. Application to the disc machine

It is relevant to proceed in a similar fashion to Section 5.2.2 as the geometry thus described although not giving the same physical insight represents a better description of the surface of circumferentially ground discs which are in common use in the investigation of elastohydrodynamic phenomena.

As previously discussed consider two discs of width 2L and Hertzian band 2D. The motion is in the y direction and the profiles of length l (<2 L) are taken in the x direction. All previously mentioned assumptions hold true but for the fact that asperities are now ellipsoids and there are m pairs of traces.

The total load per unit length in the y direction corresponding to the profile length $m^2 l$ is

$$W' = \frac{3}{2}\,\pi\,E'\,f_2(A/B)\,\Sigma\omega \qquad (5.2.23)$$

where $\Sigma\omega$ is the sum of all the interference depths on the profiles.

Thus the total load supported by the area 2D x 2L is given by

$$W_{total} = \frac{3}{2}\,\frac{\pi\,E'\,f_2(A/B)\,\Sigma\omega}{\Sigma l} \times 4D\,L$$

$$W_{total} = \frac{6\pi D L E'\,f_2(A/B)}{m^2 l}\,\Sigma\omega \qquad (5.2.24)$$

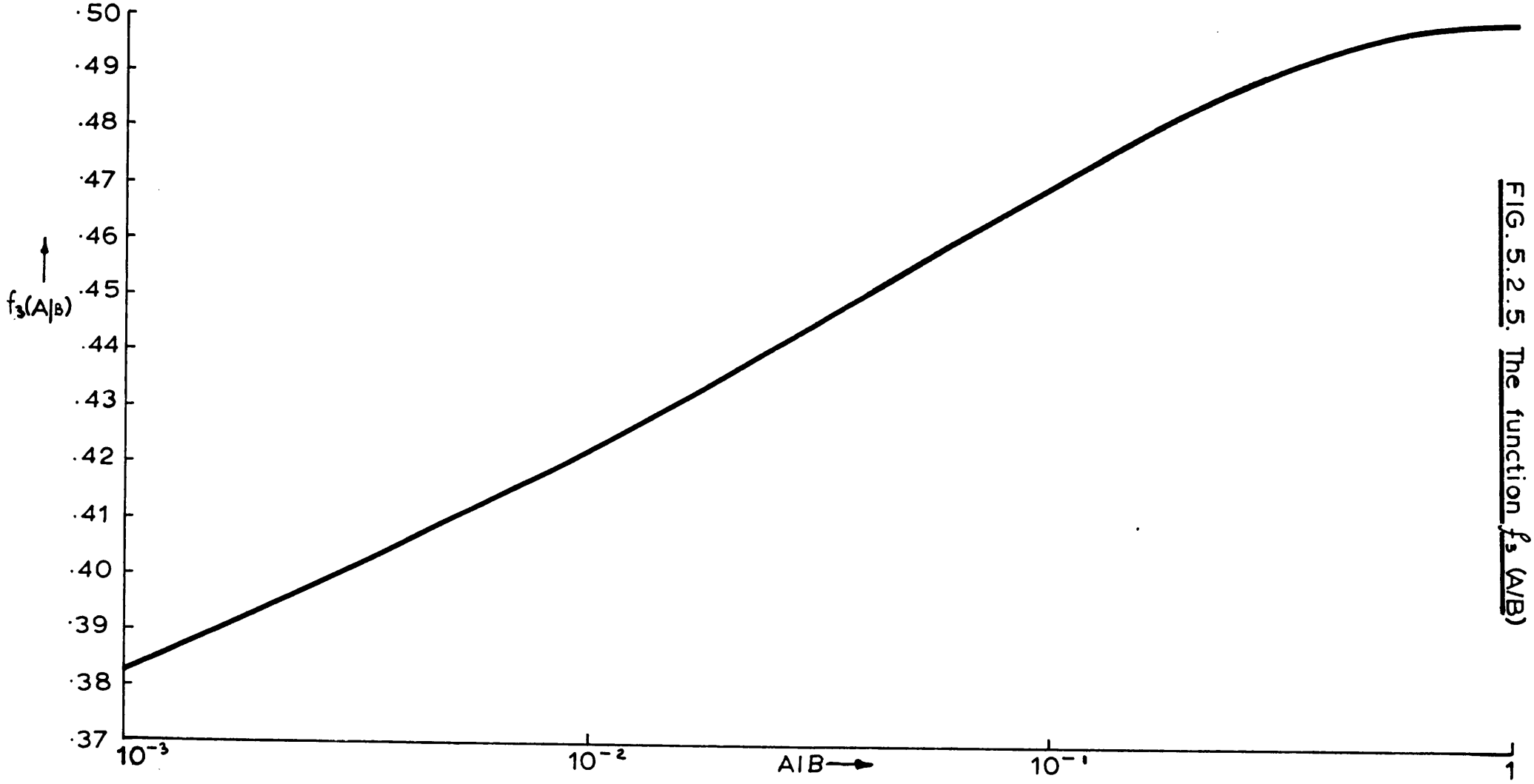


FIG. 5.2.5. The function f_3 (A/B)

This equation may now take an approximated form. The function $f_2(A/B)$ has been evaluated by Bell and Dyson (1972) for values of A/B from 10^{-4} to 1 and for the range from 5×10^{-2} to 5×10^{-4} (within which range the asperities on ground surfaces lie) this function may be replaced by a constant suggested by Bell and Dyson equal to 0.06. Thus the calculated load is not markedly dependent upon the assumed length/width ratio of the asperities. Then to an acceptable accuracy equation (5.2.24) may be written as

$$W_{total} = 1.13 \frac{DL E'}{m^2 l} \Sigma \omega \qquad (5.2.25)$$

In a similar fashion the analysis may be applied to determine the real area of contact which gives the equivalent equation of (5.2.24) as

$$A_{total} = \frac{8DL}{m^2 l} f_3(A/B) \Sigma \beta \qquad (5.2.26)$$

where $f_3(A/B) = k^{-1}[E(k')] [K(k')]^{-1} (A/B)^{\frac{1}{2}} (A/B + 1)^{-1}$

and is presented in graphical form in Figure 5.2.5.

5.2.6. Plastic contact for the ellipsoidal case

The argument for plastic contact is directly analogous to that used for the spherical case, what remains to be found are the values of semi-axes of interference and to apply the equations as before.

From the Hertzian theory equation (5.2.13) may be written for the more general case of elliptical contacts

$$\left.\begin{aligned} p_{mean} &= \frac{W}{\pi a b} \\ p_o &= \frac{3W}{2\pi a b} \end{aligned}\right\} \qquad (5.2.27)$$

Thomas and Hoersch provide the two basic equations for analysis of elliptic contacts

$$a = \left(\frac{3 W E(k')}{(A + B) E' 2\pi k^2} \right)^{\frac{1}{3}} \qquad (5.2.28)$$

The compliance ω_o is given by the equation

$$\omega_o = \frac{3W}{2a\pi E'} K(k') \qquad (5.2.29)$$

From equation (5.2.27)

$$W = \frac{2}{3} \frac{k^2 (A + B)\pi E' a^3}{E(k')} \qquad (5.2.30)$$

combining equations (5.2.29) and (5.2.30)

$$W = \frac{2}{3} \frac{k^2 \pi E' (A + B) a^3}{E(k')} = \frac{2}{3} \frac{\pi E' \omega_o a}{K(k')}$$

thus

$$\frac{a^2}{\omega_o} = \frac{E(k')}{k^2 (A + B) K(k')} \qquad (5.2.31)$$

we know

$$\frac{1}{A} = \frac{a_o^2}{\omega_o} = \frac{a^2}{\omega}$$

$$\frac{1}{B} = \frac{\beta_o^2}{\omega_o} = \frac{\beta^2}{\omega}$$

Therefore

$$\frac{a^2}{\beta_o^2} = \frac{E(k') B}{k^2 (A + B) K(k')}$$

or

$$a = \frac{\beta_o}{k} \left\{ \frac{E(k') B}{(A + B) K(k')} \right\}^{\frac{1}{2}} \qquad (5.2.32)$$

substituting equation (5.2.30) into (5.2.27)

$$p_o = \frac{k (A + B) E' a}{E(k')} \qquad (5.2.33)$$

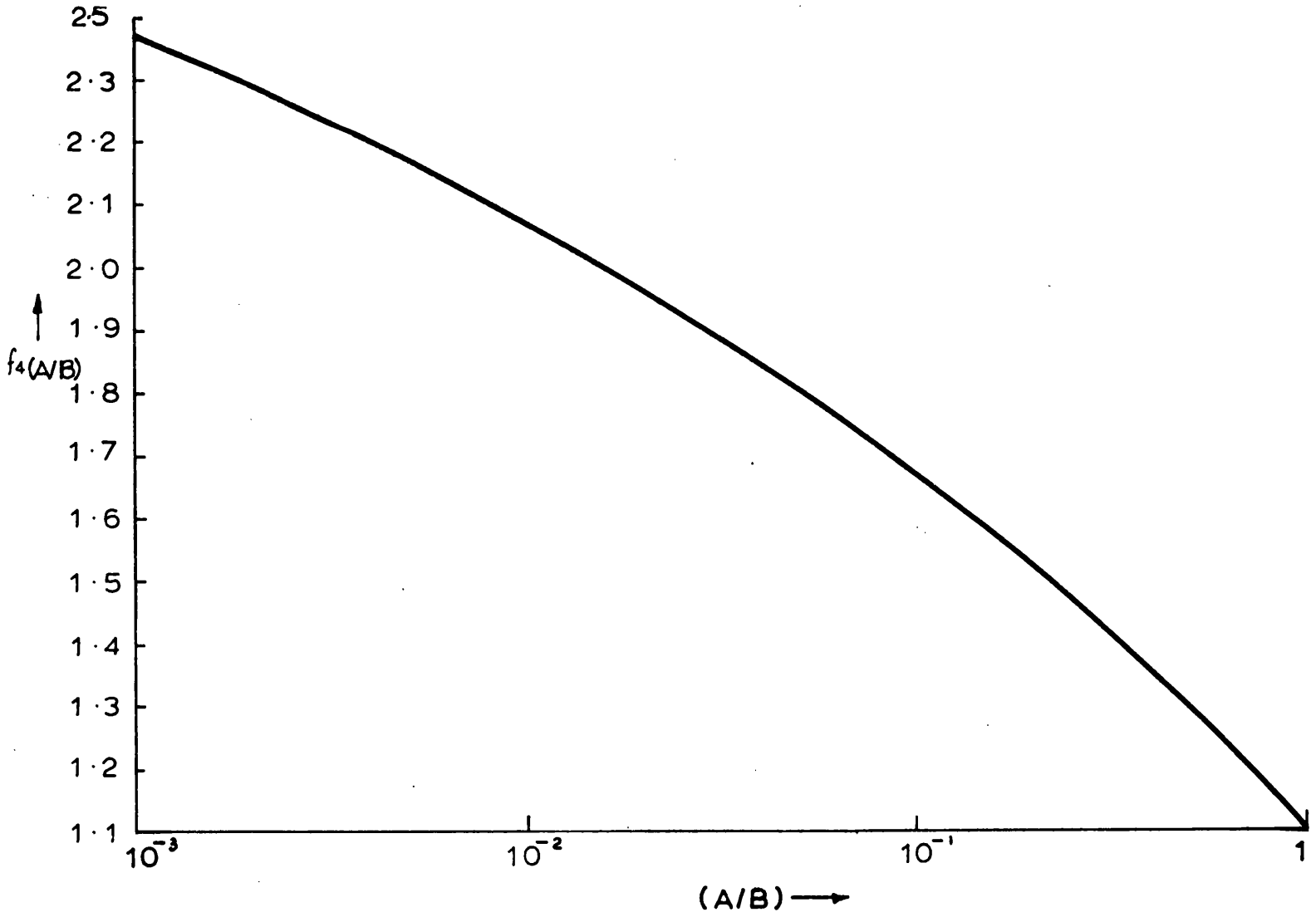


FIG. 5.2.6. The function $f_4(A/B)$

Combining equations (5.2.32) and (5.2.33)

$$p_o = \frac{\omega_o E'}{\beta_o} \left\{ \frac{(A + B)}{E(k') K(k') B} \right\}^{\frac{1}{2}}$$

It will be assumed that the onset of plastic flow is reached when the maximum Hertzian pressure satisfies the equation

$$p_o > 0.6H$$

This last statement is not strictly true, it is derived from the assumption of point contact i.e. from considerations of spheres. In the envisaged use of the theory the ellipticity may well rise to values more accurately regarded as "near-line contact", in which case the above statement may well not hold true. See for example Merwin and Johnson (1963) where the maximum Hertzian pressure for the onset of plasticity would be 0.65H for line contact. Even so if the assumption of constant ellipticity is used then this error simply reduces to a different constant being introduced into the final equation.

The condition for the onset of plasticity may now be written as

$$\frac{\omega_o}{\beta_o} > 0.6 \frac{H}{E'} \left\{ \frac{E(k') K(k')}{(1 + A/B)} \right\}^{\frac{1}{2}} \qquad (5.2.34)$$

Thus plastic deformation will have occurred if

$$\frac{\omega}{\beta} > 0.6 \frac{H}{E'} f_4 (A/B) \qquad (5.2.35)$$

where $f_4 (A/B) = \left\{ \frac{E(k') K(k')}{(1 + A/B)} \right\}^{\frac{1}{2}}$

The function $f_4 (A/B)$ has been evaluated for values of A/B from 1 to 10^{-3} and is shown in Figure 5.2.6. Using a value of A/B equal to one the result is the same as that derived for spherical asperities.

In the same manner as Section 5.2.3 we may now write down an equation for the total load supported by plastic deformation as determined from Talysurf traces.

$$W_p = \frac{3}{2} \pi E' f_2(A/B) \Sigma \omega_p \qquad (5.2.36)$$

where $\Sigma \omega_p$ is the sum of the values of compliance ω where equation (5.2.35) holds for those particular regions of contact.

5.3. Conclusions

This chapter has set out to give a greater physical understanding to the problem of surface contact by the direct use of profiles derived from surfaces used in experiments. It has provided a simple tool for the analysis of partial elastohydrodynamic lubrication with the aid of simple mathematics and a computer. The beauty of the approach is that no specific model is assumed and therefore it avoids the errors which arise from the fact that many surfaces, particularly those modified by running-in, do not fit the theoretical models. The technique described here is now being applied by Mr. R.T. Hunt to the analysis of surfaces derived from the experiments of Bell and Dyson (1972).

It has been shown that the load support from asperity contacts is directly proportional to the summation of all the asperity interference depths $\Sigma \omega$. It has also been shown that the load support involving plastic deformation can be similarly calculated by taking the summation only over asperity interactions of the required severity. The initial theory assumed spherical asperities but its extension to ellipsoids, such as occur on ground surfaces, has also been derived. The original conclusions of the first analysis still hold and the only additional assumption required is a mean value of the shape of the ellipsoids represented by the ratio (A/B).

The great merit of the approach is that it uses data from surfaces used in experimental work. Thus checks on some of the assumptions used in the theory can be derived from computer analysis. One obvious

question which arises from model analysis of two rough surfaces in contact is that of alignment of opposing asperities. Greenwood and Tripp (1971) pay great attention to the problems involved, realising that in general the highest points (summits) would not coincide. In the computer analysis of profiles presented here this problem does not arise because the analysis proceeds simply by observing the contact as it occurs, and in preliminary application of the method described in this chapter. Mr. R.T. Hunt has observed the existence of these misalignments as small tilts of the planes of contact to the mean lines through the contacting profiles. Also from this preliminary work it has become clear that sometimes, under heavy loads, two contacting regions combine to form one single region. By careful examination of such events means may be found to take account of this phenomenon.

The analysis at the present is possibly only applicable for pure rolling conditions as no account has been taken of sliding and the possible resultant wear. This problem is of course inherent in all surface contact models but Chapter 5 would seem to have the greatest possibility of modification to take this problem into account. Further development of the technique described here must obviously await the full development of the computer programmes, but as an example of its possible use, and power, consider the following experiment. Rough discs are run under lubricated conditions and fairly heavy loads such that marked running-in occurs with considerable modification of the surface profiles by plastic deformation of the asperities. The profiles of such surfaces are brought together in the computer and the analysis of this chapter is performed for a range of separations. It would be expected that as the separation is decreased there would, at first, be only elastic deformation of the asperities. As we pass through the separation at which the surfaces ran in practice there

should be a marked and fairly rapid increase in the proportion of plastic deformation. If such an effect was observed it would be strong evidence for assuming that this separation corresponded to that which occurred in the experiments.

These questions and problems in the method outlined here can only be answered after development of the computer programmes but it seems possible that the method described here may be a powerful tool for the investigation of mixed lubrication conditions.

CHAPTER 6

GENERAL DISCUSSION AND CONCLUSIONS

6.1. Experimental Work

In Chapters 2 and 3 the author has attempted to draw comparisons between results obtained for a geometrically simple disc machine simulating line contact and results obtained for a laboratory gear test rig. It can be claimed that in broad terms the conclusions reached by Dawson are also applicable to gears, but that disc machine predictions would grossly overestimate the life of gears.

The main conclusions drawn from this experimental work are listed below.

1) Both chapters substantiated the fact that pitting fatigue is clearly dependent on the surface roughness/film thickness parameter. However the correct definition of combined surface roughness is still not positively determined, but for engineering practice the most reliable definition would be twice the surface roughness of the harder element.

2) It was hoped that Chapter 2 would provide information about the influence the orientation of machining has on the life to pitting. Unfortunately this work was not carried through to completion, but the few results reported here suggest that the life to pitting of axially ground discs would in fact be somewhat greater than for those ground in the more conventional circumferential direction.

3) The disc machine work has shown that the drive ratio plays a much more important role than was first thought before the commencement of this work. There was a marked increase in the amount of wear taking place when the discs were driven by a hunting tooth ratio.

4) A comparison of the results of Chapters 2 and 3 showed that the lives of discs could overestimate those of gears by a factor of 10^2. [redacted] The generality of this discovery has not been established. A full discussion took into account the various physical parameters which could be controlled by design and it was concluded that there was an inherent difference between gears and discs. This reduced life was possibly attributed to doubts about the ability to predict the film thickness between gear teeth or the effects of dynamic loading.

5) The results of Chapter 3 show that speed influences pitting life independently of its influence on film thickness. At low speed (500 r.p.m) this effect is very apparent and similar to that reported by Dawson (1963) on discs. Statistical analysis also showed the possibility of a secondary speed effect for the other two running conditions again suggesting dynamic loading plays an important role.

The comparison of discs and gears has unfortunately been, for the bulk of results, drawn from two differing drive ratios. The argument of asperity stress cycles (Shotter (1961)) has already had some discussion and as stated earlier, it is the author's opinion that the explanation is an over-simplification of the problem. Briefly, the argument states that only when a complete cycle of tooth contacts has taken place will two opposing asperities again contact one another. Shotter proposes that the interactions which have the most severe stress conditions will result in pitting failure and these stress cycles will be a function of the drive ratio. The concept in itself is indeed valid for the determination of the number of fatigue cycles if one considers only two particular contacting asperities on opposing surfaces. What this argument overlooks is the possibility that one of these asperities comes into contact with other asperities. The resultant stresses may be less than the previous interaction but still sufficient to help in initiating fatigue. Secondly the hypothesis overlooks the possible influence of the material combination. This is most important when considering deformation. Consider

a case hardened specimen running against a relatively softer one, it may be fair to approximate the hard specimen to a non-deformable body. For simplicity let us now assume that the hard surface is rough and the soft surface smooth. Then for a one to one drive ratio, the highest asperity will have a greater probability than others of initiating pitting due to its higher contact stresses. Let us now introduce a hunting tooth drive with a similar combination of discs. In this situation the highest asperity will make more than one contact, dependent on the hunting ratio. If the stresses so obtained are above the critical limit then the probability of initiation of pitting will be even greater. We therefore have the situation where there are less individual stressing cycles per asperity interaction compared to cycles run, but we have a greater number of individual interactions which may initiate pitting. This over-simplified example illustrates that the concept of 'n.s.c' is inadequate. The problem is far more complex, it requires a statistical analysis and thus a surface model analysis may be a useful tool to help understand this problem. Chapters 4 and 5 are therefore relevant to this problem, they present two theoretical approaches to the understanding of surface contact of the type encountered in the experimental work of this thesis, and also other tribological failures of a similar nature.

6.2. Theoretical Analysis

Chapter 4 discussed the contact of random surfaces in terms of a general theory of such surfaces previously described by Whitehouse and Archard (1970). The surface was defined by two parameters σ the r.m.s value of the height distribution and β^* the correlation distance which is associated with the major wavelength of the structure. The theory is compared with the earlier study of Greenwood & Williamson (1966) which uses a three parameter model, and the influence of a distribution of

asperity curvatures is revealed. The concept of surface models and their limitations is discussed and from this a more general approach for computer analysis is developed in Chapter 5. The theory is the first approach for the analysis of such surfaces and their resultant modification throughout their running process. The main conclusions resulting from this work are as follows.

1) The primary influence of introducing a distribution of asperity curvatures is to increase the contact pressures and also to increase the probability of plastic flow.

2) The plasticity index, which indicates the probability of plastic deformation, can be re-expressed in terms of a two parameter model, namely σ and β^*. The great merit of this re-definition is that the plasticity index is now expressed in terms of surface parameters which are more easily measured, and which form part of proposals (Spragg and Whitehouse (1970/71)) for a future specification of surface topography.

3) An indication of anisotropy, for typical ground surfaces, is given in Chapter 4 and it results in the plasticity index falling in the region where the probability of plastic flow is quite small.

4) In Chapter 5 a basic theory for the analysis of surface contact, based upon computer simulation, has been provided in a form suited to any type of surface, and may therefore be used for surfaces modified by running in.

5) Work by R. T. Hunt (unpublished) suggests that a reliable estimate of load borne by solid contact can be obtained by this method.

6.3. Suggestions for Further Work

The work carried out in this thesis has opened up fields of investigation which now warrant considerable attention. A concise list of topics of the more immediate problems is presented below.

1) The work has shown that results obtained from discs will grossly overestimate the lives of gears, and that there is some fundamental difference between the two forms of specimen. Thus, pitting failure of gears, although more costly, can and must be investigated using a similar approach to that of discs.

2) The influence of asperity orientation in grinding may now be fully investigated with the aid of a cheap manufacturing process described in Chapter 2.

3) Some thought must be given to methods of determining reliable values of film thickness between gear teeth.

4) Using the facilities of (3) it may be possible to substantiate the effect the influence of speed has on the results other than the D ratio.

5) The concept of 'asperity stress cycles' should be given more consideration. A programme of disc machine tests with varying drive ratios should give some insight to this problem. This work may also benefit by the inclusion of resistance measurement techniques.

6) The theory of Chapter 5 is now at a stage where experimental results are required. Initially discs with zero sliding should be run and these results may then be compared with others obtained with a small amount of sliding incorporated. It may also be of use to include discs of unequal hardness.

7) One of the more important theoretical investigations should be the development of a theory for micro e.h.l which may be used in conjunction with the surface analysis models of Chapters 4 and 5.

At the initiation of this thesis most of the topics listed above were not apparent and it was inevitable that a complete investigation could not have been possible. Unfortunately throughout the project time was at a premium; the author could do no better than substantiate a small part of the subject and indicate the various directions that investigations must now go. Nevertheless the points above will, when

fully investigated, go a long way towards the understanding of pitting fatigue in gears; and hopefully a reliable theoretical model to help in the understanding of several forms of tribological failures will be forthcoming from Chapters 4 and 5.

APPENDIX I

METHODS OF INTEGRATION

We require to evaluate the integrals F_A, F_W, F_G and F_n defined by equations (4.3.4), (4.3.5), (4.3.6) and (4.3.8) of Chapter 4. The integral F_n is the cumulative distribution of asperity heights which has been given in equation (4.2.14).

Removing the constants and changing the order of integration for F_A leads to the form

$$I_A = \int_0^\infty \frac{\operatorname{erf}(\frac{1}{2}C)}{C} \int_d^\infty (y-d)\exp(-\tfrac{1}{2}y^2)\exp[-(y-\tfrac{1}{2}C)^2]\, dy\, dC \qquad (A1)$$

Let I be the inner integral, then, completing the square in the exponential term,

$$I = \int_d^\infty (y-d)\exp[-\tfrac{3}{2}(y^2-\tfrac{3}{3}yC+\tfrac{1}{3}C^2)]dy$$

$$= \exp(-\tfrac{1}{12}C^2)\int_d^\infty (y-d)\exp\{-\tfrac{3}{2}(y-\tfrac{1}{3}C)^2\}\, dy \qquad (A2)$$

Using the substitution $Z = \sqrt{\frac{3}{2}}(y-\frac{1}{3}C)$, $dz = \sqrt{\frac{3}{2}}\, dy$

$$I = \exp(-\tfrac{1}{12}C^2) \int_{\sqrt{\frac{3}{2}}(d-\frac{1}{3}C)}^\infty [\sqrt{\tfrac{2}{3}}\, z+\tfrac{1}{3}C-d]\sqrt{\tfrac{2}{3}}\exp(-z^2)dz$$

which yields

$$I = \tfrac{1}{3}\exp(-\tfrac{1}{12}C^2)\left\{\exp[-\tfrac{3}{2}(d-\tfrac{1}{3}C)^2] - \sqrt{\frac{3\pi}{2}}(d-\tfrac{1}{3}C)[1-\operatorname{erf}\sqrt{\tfrac{3}{2}}(d-\tfrac{1}{3}C)]\right\}$$

Hence

$$I_A = \int_0^\infty \frac{\operatorname{erf}(\frac{1}{2}C)\exp(-\frac{1}{12}C^2)}{3C}\left\{\exp[-\tfrac{3}{2}(d-\tfrac{1}{3}C)^2] - \sqrt{\frac{3\pi}{2}}(d-\tfrac{1}{3}C)\right.$$
$$\left.[1-\operatorname{erf}\sqrt{\tfrac{3}{2}}(d-\tfrac{1}{3}C)]\right\} dC \qquad (A3)$$

I_W and I_G cannot be reduced to a single integral but may be simplified by changing the order of integration and completing the square for the exponential term as above. The substitution

$$z = (y - d)$$

may then be made with the following results

$$I_W = \int_d^\infty (y-d)^{\frac{3}{2}} \int_0^\infty \frac{1}{\sqrt{C}} \exp(-\tfrac{1}{2}y^2)\exp[-(y-\tfrac{1}{2}C)^2]\operatorname{erf}(\tfrac{1}{2}C)\, dC\, dy$$

$$= \int_0^\infty \frac{1}{\sqrt{C}} \operatorname{erf}(\tfrac{1}{2}C)\exp(-\tfrac{1}{12}C^2) \int_0^\infty z^{\frac{3}{2}} \exp[-\tfrac{3}{2}(z+d-\tfrac{1}{3}C)^2]\, dz\, dC \qquad (A4)$$

$$I_G = \int_d^\infty (y-d)^{\frac{1}{2}} \int_0^\infty \frac{1}{\sqrt{C}} \exp(-\tfrac{1}{2}y^2)\exp[-(y-\tfrac{1}{2}C)^2]\operatorname{erf}(\tfrac{1}{2}C)\, dC\, dy$$

$$= \int_0^\infty \frac{1}{\sqrt{C}} \operatorname{erf}(\tfrac{1}{2}C)\exp(-\tfrac{1}{12}C^2) \int_0^\infty z^{\frac{1}{2}}\exp[-\tfrac{3}{2}(z+d-\tfrac{1}{3}C)^2]\, dz\, dC \qquad (A5)$$

For computation, equations (A3), (A4), (A5) are transformed to give limits from 0 to 1. I_A may be thus transformed by the use of

$$x = 1/(1+C) \qquad (A6)$$

to give

$$I_A = \int_0^1 \frac{1}{x^2} f(\frac{1}{x} - 1)\, dx$$

where $f(\frac{1}{x} - 1)$ is the integrand of equation (A3).

Similarly to transform equation (A4) we write

$$\zeta(C) = \frac{1}{\sqrt{C}} \operatorname{erf}(\tfrac{1}{2}C)\exp(-\tfrac{1}{12}C^2) \qquad (A7)$$

$$\psi(z,C) = z^{\frac{3}{2}} \exp[-\tfrac{3}{2}(z+d-\tfrac{1}{3}C)^2] \qquad (A8)$$

Then

$$I_W = \int_0^\infty \zeta(C) \int_0^\infty \psi(z,C)\, dz\, dC$$

and using the substitutions

$$x = 1/(1+C) \quad b = 1/(1+z)$$

we obtain the result

$$I_W = \int_0^1 \frac{1}{x^2}\, \zeta\left(\frac{1}{x} - 1\right) \int_0^1 \frac{1}{b^2}\, \psi\left\{\left(\frac{1}{b} - 1\right), \left(\frac{1}{x} - 1\right)\right\} db\, dx \qquad (A9)$$

I_G may be evaluated by similar methods replacing $z^{\frac{3}{2}}$ by $z^{\frac{1}{2}}$ in equation (A8).

The equation for the area of contact involving plastic flow is given by equation (4.4.4) Chapter 4 and may be reduced to the form

$$A_p = \frac{1}{5}\, \pi\, F_{Ap}(d)$$

where

$$F_{Ap} = \int_{d+\omega_p^*}^\infty (y-d) \int_0^\infty \frac{f^*(y,C)}{NC}\, dC,\, dy \qquad (A10)$$

which differs from the equation for A (equation (4.3.4)) by the lower limit of the outer integral. Using the earlier techniques we can write the required integral directly

$$I_{Ap} = \int_0^\infty \frac{1}{3C}\, \mathrm{erf}(\tfrac{1}{2}C).\exp(-\tfrac{1}{12}C^2)\left\{\exp[-\tfrac{3}{2}[d+\frac{\phi}{C}-\tfrac{1}{3}C]^2]\right.$$

$$\left. - \sqrt{\frac{3\pi}{2}}\,(d-\tfrac{1}{3}C)\,[1-\mathrm{erf}[\sqrt{\tfrac{3}{2}}(d+\frac{\phi}{C}-\tfrac{1}{3}C)]]\right\} dC \qquad (A11)$$

where $\frac{\phi}{C} = \omega^*_p$ and is defined by equation (4.4.3).

The integral was evaluated for values of ϕ = 3.5, 6.2, 8.5, 11.0. The results are shown in Fig. 4.5.6.

REFERENCES

ARCHARD, J.F. (1957) 'Elastic deformation and the laws of friction', Proc. Roy. Soc., A243, 190.

ARCHARD, J.F. (1973) 'Elastohydrodynamic lubrication of real surfaces', Tribology, In the press.

BAMBERGER, E.N. (1969) 'Effects of materials - metallurgical viewpoint', Interdisciplinary approach to the lubrication of concentrated contacts. Proceedings of a N.A.S.A. sponsored symposium, 409.

BOWDEN, F.P. and TABOR, D. (1954) 'Friction and lubrication of solids', Part I. (Oxford University Press).

BELL, J.C. and DYSON, A. (1972) 'Mixed friction in an elastohydrodynamic system', Proc. of symposium 'Elastohydrodynamic Lubrication'. Proc. I. Mech.E. 68.

CAMERON, R. and GREGORY, R.W. (1967/68) 'Measurement of oil film thickness between rolling discs using a variable reluctance technique', Proc. I. Mech.E. 182 Pt.3N, 24.

CHESTERS, W.T. (1958) 'Study of the surface fatigue behaviour of gear materials with specimen of simple form', Proceedings of the International Conference on Gearing, 91, (I. Mech. E., London).

CHESTERS, W.T. (1963) 'The effect of material combination on resistance to surface fatigue', Symposium on Fatigue in Rolling Contact, 86, (I. Mech. E., London).

CHRISTENSEN, H. and TONDER, K. (1971) 'The hydrodynamic lubrication of rough bearing surfaces of finite width', A.S.M.E. publication Paper No. 70-Lub-7.

COLEMAN, W. (1967) 'Bevel and hypoid gear surface durability: pitting and scuffing', Conference on Lubrication and Wear, Proc. Instn. Mech. Engrs., 182, Part 3A , 191.

CROOK, A.W. (1955) 'The elastic deformation of cylinders loaded in line contact', A.E.I. report A471 (Aldermaston, Oct. 1955).

DAWSON, P.H. (1961) 'The pitting of lubricated gear teeth and rollers', Power Transmission April and May.

DAWSON, P.H. (1962) 'The effect of metallic contact on the pitting of lubricated rolling surfaces', J. Mech. Engng. Sci., Vol. 4, (No.1), 16.

DAWSON, P.H. (1963) 'The effect of metallic contact and sliding on the shape of the S-N curve for pitting fatigue', Symposium on Fatigue in Rolling Contact, 41, (I. Mech. E., London).

DAWSON, P.H. (1965) 'Further experiments on the effect of metallic contact on the pitting of lubricated rolling surfaces', Symposium on Elastohydrodynamic Lubrication, Proc. I. Mech. E., 180, Pt.3B, 95.

DAWSON, P.H. (1968) 'Rolling contact fatigue crack initiation in a 0.3 per cent carbon steel', Proc. I. Mech. E., 183, Pt. 1, No. 4, 1.

DOWSON, D. (1970) 'The role of lubrication in gear design', Gearing in 1970, Proc. I. Mech. E., 184, Pt.30, 72.

DOWSON, D. and HIGGINSON, G.R. (1961) 'New roller-bearing lubrication formula', Engineering, London 192, 158.

DOWSON, D. and HIGGINSON, G.R. (1966) 'Elastohydrodynamic lubrication. The fundamentals of roller and gear lubrication', Pergamon Press.

DOWSON, D., HIGGINSON, G.R. and WHITAKER, A.V. (1963) 'Stress distribution in lubricated rolling contacts', Symposium on Fatigue in Rolling Contact, 66, (I. Mech. E., London).

DUDLEY, D.W. (1962) 'Gear Handbook - The design, manufacture and application of gears'. (McGraw-Hill).

DYSON, A. (1966) 'Investigation of the discharge - voltage method of measuring the thickness of oil films formed in a disc machine under conditions of elastohydrodynamic lubrication', Proc. I. Mech. E., 181, Pt. 1, 637.

GALVIN, G.D. and NAYLOR, H. (1965) 'Effects of lubricants on the fatigue of steel and other metals', Proc. I. Mech. E., 179, Pt. 3J, 56.

GREENWOOD, J.A. and TRIPP, J.H. (1970/71) 'The contact of two nominally flat rough surfaces', Proc. I. Mech. E., 185, 625.

GREENWOOD, J.A. and WILLIAMSON, J.B.P. (1966) 'Contact of nominally flat surfaces', Proc. Roy. Soc., A295, 300.

HAMILTON, G.M. (1963) 'Plastic flow in rollers loaded above the yield point', Symposium on Fatigue in Rolling Contact, 136, (I. Mech. E., London).

IBRAHIM, M. and CAMERON, A. (1963) 'Oil film thickness and the mechanism of scuffing in gear teeth', Proc. Lubrication and Wear Convention (Bournemouth) paper 20, 228, (I. Mech. E., London).

JOHNSON, K.L. (1963) 'Correlation of theory and experiment in research on fatigue in rolling contact', Symposium on Fatigue in Rolling Contact, 155, (I. Mech. E., London).

JOHNSON, K.L., GREENWOOD, J.A. and POON, S.Y. (1972) 'Asperity contact in elastohydrodynamic lubrication', Wear, 19, 91.

JOHNSON, K.L. and JEFFERIS, J.A. (1963) 'Plastic flow and residual stresses in rolling and sliding contact', Symposium on Fatigue in Rolling Contact, 54, (I. Mech. E., London).

KU, P.M. (1972) 'Private communication' Now to be found in publication by Staph, H.E., Ku, P.M. and Carper, H.J. (1972) 'Effect of surface roughness and surface texture on scuffing', Presented at the A.S.M.E. - A.G.M.A. Symposium on Gearing and Transmissions.

MARTIN, J.B. and CAMERON, A. (1961) 'Effect of oil on the pitting of rollers', Journal Mechanical Engineering Science, Vol. 3, No. 2, 148.

MERWIN, J.E. and JOHNSON, K.L. (1963) 'An analysis of plastic deformation in rolling contact', Symposium on Fatigue in Rolling Contact, 145, (I. Mech. E., London).

NIEMANN, G., RETTIG, H. and BÖTCH, H. (1964/65) 'The effect of different lubricants on pitting resistance of gears', Proc. I. Mech. E., 179, Pt. 3D, 192.

PEKLENIK, J. (1967/68) 'New developments in surface characterisation and measurements by means of random process analysis', Proc. I. Mech.E., 182, Pt. 3K, 108.

POON, S.Y. and HAINES, D.J. (1966/67) 'Frictional behaviour of lubricated rolling-contact elements', Proc. I. Mech. E., 181, 339.

RIDGWAY, D. (1970) 'Discussion of Paper 20 in Gearing In 1970', Proc. I. Mech. E., 184, Pt. 30, 208.

RUBERT, M.P. (1959) 'Confusion in measuring surface roughness', Engineering, London, 138, 393.

SCOTT, D. (1963) 'The effect of material properties, lubricant, and environment on rolling contact fatigue', Symposium on Fatigue in Rolling Contact, 103, (I. Mech. E., London).

SCOTT, D., and BLACKWELL, J. (1966) 'Study on the effect of material combination and hardness in rolling contact', Proc. I. Mech. E., 180, Pt. 3K, 32.

SHOTTER, B.A. (1958) 'Experiments with a disc machine to determine the possible influence of surface finish on gear tooth performance', Proceedings of the International Conference on Gearing, 102, (I. Mech. E., London).

SHOTTER, B.A. (1961) 'Gear tooth pitting and fatigue strength', Engineer, London, 212, 855.

SPRAGG, R.C. and WHITEHOUSE, D.J. (1970/71) 'A new unified approach to surface metrology', Proc. I. Mech. E., 185, 697.

SYKES, A. (1959) 'Allowable load on helical gears', Engineer, London, 208, 598.

TALLIAN, T.E. (1967/68) 'On competing failure modes in rolling contact', A.S.L.E. Transactions, 10, 418.

TALLIAN, T.E. (1972) 'The theory of partial elastohydrodynamic contacts', Wear, 21, 49.

TALLIAN, T.E., CHIU, Y.P., HUTTENLOCHER, D.F., KAMENSHINE, J.A., SIBLEY, L.B. and SINDLINGER, N.E. (1964) 'Lubricant films in rolling contact of rough surfaces', Trans. Amer. Soc. Lubr. Engrs. 7(No.2), 109.

THOMAS, H.R. and HOERSCH, V.A. (1930) 'Stresses due to the pressure of one elastic solid upon another', Bull. I ll.Univ. Engng. Exp. Stn. (No. 212).

TIMOSHENKO, S. and GOODIER, J.N. (1951) 'Theory of elasticity', (New York, McGraw Hill) pp.372-382.

WAIGHT, F.H. (1964) 'A study of fatigue pitting in gears', Proposed paper for submission to D.G.M.K. Annual Meeting 1964. Thornton Research Centre (Shell Research Ltd.) Report No. TRCP.1111.

WAIGHT, F.H. (1968) 'The basic principles of disc and gear testing machine design', Experimental Methods in Tribology, Proc. I. Mech. E., 182, Pt. 3G.

WAY, S. (1935) 'Pitting due to rolling contact', J. Appl. Mech. 57, A49.

WAY, S. (1937) 'Roller tests to determine pitting fatigue strength', 20th Semi-Annual Meeting of the A.G.M.A.

WAY, S. (1940) 'Westinghouse roller and gear pitting tests', 24th Annual Meeting of the A.G.M.A.

WELLAUER, E.J. (1967) 'A.G.M.A. experience in establishing co-ordinated gear rating standards', invited paper presented to the Semi-Int. Symposium,Japan Soc. Mech. Engrs.

WHITEHOUSE, D.J. (1971) 'The properties of random surfaces of significance in their contact', Ph.D. Thesis (The University of Leicester).

WHITEHOUSE, D.J. and ARCHARD, J.F. (1970) 'The properties of random surfaces of significance in their contact', Proc. Roy. Soc. A316, 97.

WILLIAMSON, J.B.P. (1968) 'Microtopography of surfaces', Proc. I. Mech. E., 182, Pt. 3K, 21.

ZARETSKY, E.V. and ANDERSON, W.J. (1969) 'Effect of materials - General background', Interdisciplinary approach to the lubrication of concentrated contacts, Proceedings of a N.A.S.A. sponsored symposium, 379.

ZARETSKY, E.V. PARKER, R.J. and ANDERSON, W.J. (1965)* 'Effect of component differential hardnesses on rolling contact fatigue and load capacity', N.A.S.A. tech. note D-2640, 1965 (March).

ZARETSKY, E.V., PARKER, R.J., ANDERSON, W.J. and MILLER, S.T. (1965)* 'Effect of component differential hardnesses on residual stress and rolling contact fatigue', N.A.S.A. tech. note D-2664, 1965 (March).

* The relevant points are covered by Zaretsky and Anderson (1969)